CHARACTERIZING THE LEARNING, SOCIOLOGY, AND IDENTITY EFFECTS OF PARTICIPATING IN THE DATA MINE

by

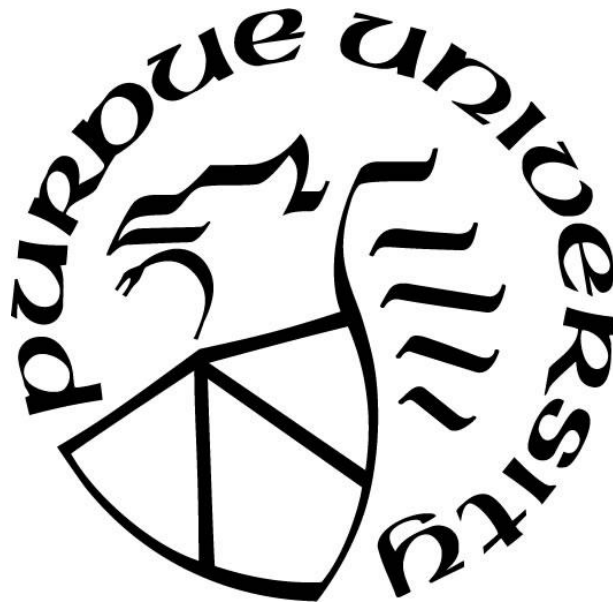**Aparajita Jaiswal**

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Computer and Information Technology

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Alejandra J. Magana, Chair

Department of Computer and Information Technology

Dr. Austin L. Toombs

Department of Computer Graphics Technology

Dr. Ida B. Ngambeki

Department of Computer and Information Technology

Dr. Mark D. Ward

Department of Statistics

Approved by:

Dr. Kathryne A. Newton

Dedicated to my mother Swapna Chatterjee

ACKNOWLEDGMENTS

I want to express my deepest gratitude to everyone who has helped me complete this dissertation. First of all, I would like to thank my advisor and mentor, Dr. Alejandra Magana, for constantly supporting and guiding me to become a better researcher and mentor. I am grateful for all your guidance and mentoring; without your help, this dissertation would not have been possible. I would also like to thank my committee members Dr. Mark Ward, Dr. Austin Toombs, and Dr. Ida Ngambeki, for their constant support and guidance. I am grateful to Dr. Mark Ward for providing me the opportunity to study The Data Mine, encouraging me constantly and patiently reviewing my work. I would also like to thank Dr. Austin Toombs for critically evaluating my dissertation and providing me necessary guidance to make it better. I am also thankful to Dr. Ida Ngambeki for her constant support and feedback.

I would also like to thank my significant others; I would not have achieved this without their encouragement and support. Firstly, I want to thank all my friends and colleagues in the ROCkETEd Lab. Thank you for being so supportive and helping me to learn so much. Specifically, I want to thank Joe, Paul, Tugba, and Devang; it has always been wonderful working with you all, and I have learned a lot from you. I also want to thank Paul, Ying Ying and Hector for being great friends. I would also take this opportunity to thank my Purdue Village friends, Pinaki, Surya, Gaurav, Romila, Richa, Salil, Manalee, Amit, and Bhavana, for being there for me in my all good and bad times. I am also grateful to my family for always believing in me. I want to thank my mother for always trusting my potential and encouraging me. I am also grateful to my father and brother for being there for me and believing in my decisions. I am also grateful to my in-law's family for being very supportive and encouraging. I also want to thank The Data Mine Staff for providing me with required resources for my study and my participants for sharing their experiences, and I learned a lot from them.

Lastly, I want to thank two of the most important persons in my life, my husband Prateek, for always being there for me, supporting my decisions, guiding, and encouraging me. Without your support, I would not have made it so far; thank you for everything! I would also like to thank my little one, Suchi, for being such a lovely and happy baby and giving me time to work on my dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The discipline of data science has gained substantial attention recently. This is mainly attributed to the technological advancement that led to an exponential increase in computing power and has made the generation and recording of enormous amounts of data possible on an everyday basis. It has become crucial for industries to wrangle, curate, and analyze data using data science techniques to make informed decisions. Making informed decisions is complex. Therefore, a trained data science workforce is required to analyze data on a real-time basis. The increasing demand for data science professionals has caused higher education institutions to develop courses and train students starting from the undergraduate level about the data science concepts and tools.

Despite the efforts from the institutions and national agency such as National Academies of Sciences, Engineering, and Medicine, it has been witnessed that there have been significant challenges in retaining and attracting students in the discipline of data science. The novice learners in data science are required to possess the skills of a programmer, a statistician, research skills, and non-technical skills such as communication and critical thinking. The undergraduate students do not possess all the required skills, which, in turn, creates a cognitive load for novice learners (Koby & Orit, 2020). Research suggests that improving the teaching and mentoring methodologies can improve retention for students from all demographic groups (Seymour, 2002). Previous studies (e.g., Hoffmann et al., 2002, Flynn, 2015; Lenning & Ebbers, 1999) have revealed that learning communities are effective in improving student retention, especially at the undergraduate level, as it helps students develop a sense of belonging, socialize, and form their own identities. Learning communities have been identified as *high impact practices* (Kuh, 2008) that helps to develop identities and sense of belonging, however to the best of our knowledge there are few studies that focus on the development of the psychosocial and cognitive skills of the students enrolled in a data science learning community.

To meet the demand for the future workforce and help undergraduate students develop data science skills, The Data Mine (TDM) at Purdue University has undertaken an initiative in the discipline of data science. The Data Mine is an interdisciplinary living-learning community that allows students from various disciplines to enroll and learn data science skills under the guidance of competent faculty and corporate mentors. The residential nature of the learning community allows the undergraduate students to live, learn and socialize with peers of similar interests and

develop a sense of belonging. The constant interaction with knowledgeable faculty and mentors in real-world projects allows novice learners to master data science skills and develop an identity. The study aims to characterize the effects of identity formation, socialization, and learning of the undergraduate students enrolled in The Data Mine and answer the following research question:

Quantitative: RQ 1: What are the perceptions of students regarding their identity formation, socialization opportunities, self-belief, and academic/intellectual development in The Data Mine?

Qualitative: Guiding RQ 2: How do students' participation in activities and interaction with peers, faculty, staff at The Data Mine contribute to becoming an experienced member of the learning community?

- **Sub-RQ 2(a):** What are the perceived benefits and challenges of participating in The Data Mine?

- **Sub-RQ 2(b):** How do students describe their levels of socialization and a sense of belonging within The Data Mine?

- **Sub-RQ 2(c):** How do students' participation and interaction in The Data Mine help them form their identity?

To approach the above research questions, we conducted a sequential explanatory mixed method study to understand the growth journey of students in terms of socialization, sense of belonging and identity formation. The data were collected in two phases: a quantitative survey study followed by qualitative semi-structured interviews. The quantitative data was analyzed using descriptive and inferential statistics, and qualitative data were analyzed using thematic analysis, followed by narrative analysis. The results of the quantitative and qualitative analysis demonstrated that learning in The Data Mine happened through interaction and socialization of the students with faculty, staff, and peers at The Data Mine. Students found multiple opportunities to learn and develop data science skills, such as working on real-world projects or working in groups. This continuous interaction with peers, faculty and staff at The Data Mine helped them to learn and develop identities. This study revealed that students did develop a data science identity, but the corporate partner TAs developed a leader identity along with the data science identity. In summary all students grew and served as mentor, guide, and role models for new incoming students.

CHAPTER 1: INTRODUCTION

The discipline of data science is gaining attention due to a drastic increase in the data generated on a real-time basis. It is important to procure, wrangle, curate, and analyze data using data science techniques to make informed decisions (De Veaux et al., 2017). The increasing demand for data science has also motivated higher education institutions to create and train a data science workforce. Therefore, higher education institutions have turned their focus to data science education and data science initiatives across the United States (Donoho, 2017; National Academies of Sciences, Engineering, and Medicine, 2018). Data science is an interdisciplinary field that requires the student to possess a foundation of mathematics/statistics, computer science, and knowledge of the specific discipline or domain that produces the data (Koby & Orit, 2020). In simple words, data science can be defined as the "science of planning for, acquisition, management, analysis of, and inference from data" (National Science Foundation, 2014, p. 4). The interdisciplinary nature of data science requires the students to possess an analytical frame of mind. The discipline of data science requires the students is to develop skills for identifying a research problem, exploring the data using mathematics or statistics principles, and arguing the findings of the analysis based on a theoretical foundation (Hardin, 2017). Therefore, the National Science Foundation has issued guidelines for the Higher Education Institutions for are creating interdisciplinary curriculum for undergraduate students that involves developing the key competencies such as "Computational and statistical thinking, Mathematical foundations, Model building and assessment, Algorithms and software foundation, Data curation, Knowledge transference—communication and responsibility" (DeVeaux et al., 2017, p. 6). Thus, engaging the undergraduate students from the first year will allow them to develop a strong foundational knowledge and apply data science skills to solve complex problems.

It is crucial to steer the undergraduate students in the right direction and engage their interest in a career-oriented field such as data science. The undergraduate students are novice learners who enter universities with great expectations and commitments. Especially, the first year of undergraduate education is critical for the students as they transition from their high school life into college life (Gergorić, 2020; Weidman et al., 2014b; Williford & Wadley, 2008). The prior studies (e.g., O'Keeffe, 2013; Singh, 2018; Williford & Wadley, 2008) have indicated the rough transition of students from high school to university education has led to dropouts. Another

important aspect that studies point out is that the first year is crucial. It helps the students develop a disciplinary identity and sense of belonging with context, peers, and faculty (Weidman et al., 2014b). Failure to develop identities (Williford & Wadley, 2008) and lack of socialization (O'Keeffe, 2013; Weidman et al., 2014b) results in dissatisfaction and finally leads to student attrition.

Prior research (e.g., Flynn et al., 2016; Hoffman et al., 2002; Lenning & Ebbers, 1999) has revealed that learning communities have proven effective in improving student retention. Especially at the undergraduate level, learning communities can help students develop a sense of belonging, socialize, and form their own identities. The studies conducted by (Arensdorf & Naylor-Tincknell, 2016; Solanki et al., 2019) have mentioned that there are limited studies that focus on the development of the psychosocial skills of the students enrolled in a learning community.

1.1   Problem Identification

Student retention in STEM is an ongoing issue for the higher education institutions (Carrino & Gerace, 2016). The factors that lead to attrition in STEM are: fewer opportunities to interact with faculty and peer, large lecture classrooms, strict grading system, course overload, independent assignment structure, etc. (Solanki et al., 2019). Learning communities have been identified as one of the high impact practices that helps to improve student retention by allowing students to interact with peers and faculty, providing them opportunities to socialize, and living and learning together (Kuh, 2008; Weidman, 1989; Weidman et al., 2014b). In general, it has been noted in Arensdorf and Naylor-Tincknell (2016) and Solanki et al. (2019) that there are very few studies on learning communities that focused on the impact of psycho-social factors to understand student experiences. Specifically, the study by Carrino and Gerace (2016) and Dagley-Falls et al. (2010) have pointed out that the majority of research in STEM learning communities measured the impact of being in the learning community on the student academic performance, graduation rates, and retention. However, there are very few studies conducted to understand the impact of participation in learning communities that leads to social interaction, identity formation, and implicit learning. Carrino and Gerace (2016) also mentioned in their study that "psychosocial factors have been under-researched in STEM education in favor of more cognitive and behavioral factors" (p. 3). Since psychosocial factors play an important role in improving student retention,

providing academic satisfaction, and improving the sense of belonging and socialization (Ekornes, 2021), it is crucial to understand the psychosocial factors affecting student experiences in STEM learning communities (Dagley-Falls et al., 2010). Moreover the prior studies by (e.g., Dagley et al., 2016; Magana et al., 2021; Solanki et al., 2019) have also identified that there are limited studies on discipline specific learning communities. Consequently, this study is focused on a data science learning community known as The Data Mine, that helps students to become data literate and develop data science skills. Also, prior studies on the domain of data science education (Hardin, 2017; Koby & Orit, 2020) have revealed that data science is an interdisciplinary field. It requires learners to possess various skills such as programming, critical thinking, and research skills in addition to the knowledge of relevant mathematical and statistical concepts. Students also need to have domain knowledge of the discipline that produces the data. Therefore, through this study we intend to understand how participating in The Data Mine is enabling students to develop interdisciplinary knowledge of data science and psychosocial skills such as identity formation, sense of belonging, socialization and transform into an experienced member of the learning community.

1.1.1 Scope

This study took place in The Data Mine at Purdue University. The Data Mine is a living learning community focused on helping students develop data science skills by immersing them in a collaborative learning environment. This study took place in between Fall of 2020 and Fall of 2021. The study involved one survey study (quantitative) and semi-structured in-depth interviews (qualitative) to understand the various opportunities that students get in The Data Mine, which help them to develop their identity, sense of belonging and learning and transform into a mature learner within the learning community. The quantitative study was a survey study. The intent of the study was to understand the perception of students regarding their identity and sense of belonging. The survey was conducted in the beginning of Fall 2020 and end of Fall 2020. In the first survey study in the Fall of 2020, 50 students participated in the study,. In the post survey, 25 students out of 50 students participated. The results of this study are presented in the Chapter 6 of the dissertation.

16

In the Fall of 2021, 19 students participated in the interview study. The results of the quantitative study helped us to design the interview questions. Semi-structured, in-depth interviews were conducted to understand the journey of students within the Data Mine and identify the various opportunities that participants got to develop into a mature learner of the learning community. Figure 1 below demonstrates the studies conducted for the purpose of this dissertation.

Quantitative Study

Pre-Survey (September 2020): 50 participants

Post-Survey (December 2020): 25 participants

Context: The Data Mine

Qualitative Study

Fall 2021: Data Collection- Semi-structured interviews

Average Length of interview: 1 hour 10 minutes

Participants: 19
Context: The Data Mine
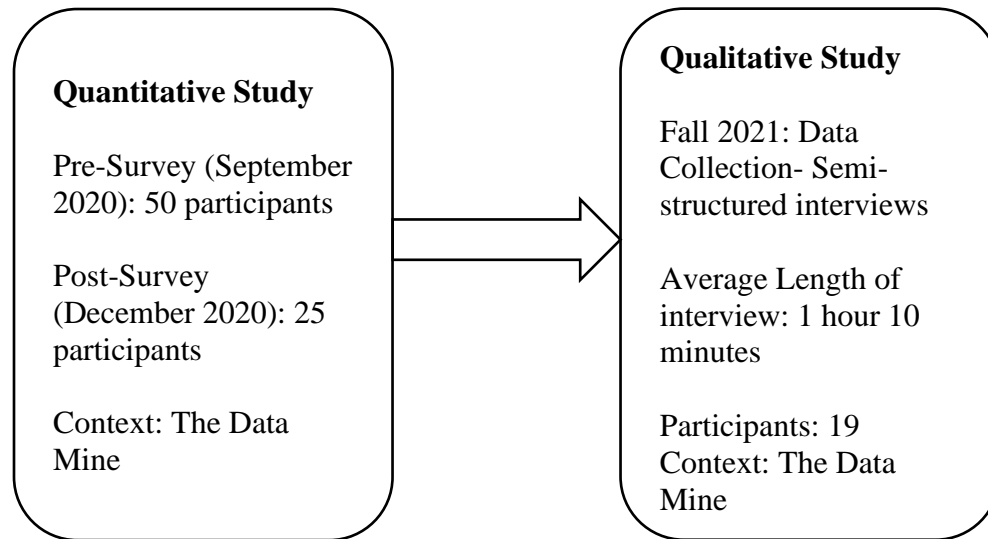
Figure 1: Timeline of the studies conducted

1.2   Significance

The data is integral for an informed decision-making process. The availability of computational tools to collect and disseminate data has exponentially increased data generation. The data is available on a real-time basis, and the organization must analyze the data instantly. Analyzing data on a real-time basis requires knowledge of statistics and computation skills. Statistics has been considered an important discipline that helps to bridge the "quantitative literacy gap" among undergraduate students (Deckard, 2017, p. 1). Computational thinking is considered the "third pillar of scientific inquiry" (PITAC, 2005, p. 1) because of its potential to solve complex problems using a multi-disciplinary lens. Data science effectively combines statistics with computational thinking to help in making informed decisions. The increasing popularity of data science and its ability to solve complex problems on a real-time basis has urged the premier institutions such as Massachusetts Institute of Technology, University of California, Berkeley,

University of Wisconsin, Madison (Donoho, 2017), and national agencies such as the National Academies of Sciences, Engineering, and Medicine to focus on data science education at the undergraduate level. According to the National Academies (2018), "to prepare their graduates for this new data-driven era, academic institutions should encourage the development of a basic understanding of data science in all undergraduates" (p. 1). The National Academies (2018) further stresses the importance of data science education based on the following rationale: (1) currently, a very small percentage of people engaged as data scientists, but by 2020 there will be approximately 2.7 million data-oriented jobs (Columbus, 2017; Gundlach & Ward, 2021; National Academies of Sciences, Engineering, and Medicine, 2018). Therefore, the knowledge of data science is crucial. (2) Students do not have clarity on the role of data scientists. They have a vague idea of what skills are required to become a data scientist. Therefore, data science education can point the students in the right direction. (3) Data science is highly interdisciplinary and is applied to every field from social sciences to core sciences. Therefore, the knowledge of data science is important for all educational domains.

Despite the efforts from the institutions and national agencies, there have been substantial challenges in retaining and attracting students for data science education. Some of the potential reasons that Koby and Orit (2020) have mentioned are: (1) the interdisciplinary nature of data science requires the students to possess a good understanding of statistics, mathematics, computer science, and domain knowledge. (2) Data science increases the cognitive load for novice learners as they are required to possess the skills of a programmer and a statistician to compute and analyze the results. (3) The data science discipline also requires the learners to possess non-technical skills such as critical thinking, effective communication to interpret and communicate the results of the algorithms, and (4) data scientists are required to possesses research skills, such as framing a research problem, understanding of research design, performing data analysis and results in interpretation. The research skills are taught especially at the graduate level, and students do not acquire these skills at the undergraduate level. 5) Lack of data science educators is another major issue that the discipline is facing.

Therefore, efforts must be directed towards improving students' learning experiences from the first year that promotes a "practice perspective" (Roth & McGinn, 1997, p. 92). From a practice perspective, the focus of learning is on participation in authentic data science experiences, where learning environments: (a) are personally meaningful to the learner to understand and develop data

science skills, (b) relate to the real-world problems and are conducted under the supervision of a faculty mentor, and (c) provide an opportunity to think in the modes of the data science discipline from multiple aspects.

1.3   Purpose of the study

The purpose of the study is to characterize the effects of identity, sense of belonging, and learning for the students enrolled in the residential learning community focused on data science education. Since the discipline of data science requires the learner to possess a variety of skills such as mathematics, computer science, research skills, communications skills, etc. (Koby & Orit, 2020). Thus, training students on all the required skills in one year is impossible. Therefore, we need to train students starting from the first year. Residential learning communities are well known for fostering student engagement by helping students experience a sense of belonging, develop identities, and promote long-term learning gains. Residential learning communities are also known as living-learning communities. They are the assimilation of academic and social mechanisms that provide meaning to student experiences. Experiences are vital for learning, and meaningful experiences occur within a specific context (Bobilya & Akey, 2002).

Learning communities provide a context that facilitates knowledge acquisition by immersing students in learning experiences (Jessup, 2015). The literature demonstrates numerous studies conducted on residential learning communities but lacks evidence of studies related to data science learning communities. Data science initiatives are on the rise across the United States. Therefore, efforts must be directed to improving the quality of students' learning experiences, particularly in introductory classes (Heroux & Allen, 2016), where serious learning gaps have been identified in students not formally trained in computation. Therefore, this study focuses on the students enrolled in a data science learning community known as The Data Mine and the researcher's intent to evaluate the effect of the residential learning community on the students' academic, social, and identity formation.

The learning communities are considered as a means to improve student retention (Pistilli, 2009). Prior studies have revealed that residential learning communities bring students of similar interests together to live and learn (Hurtado et al., 2020; Ujj, 2020). The integration of social and learning components allows the students to collaborate and engage in the collaborative construction of knowledge (Ujj, 2020). However, very few studies focus on the long-term impact

of participating in the learning community (Virtue et al., 2019), and very few studies have evaluated the effect of the residential learning community on computational or data science students. Also, the interdisciplinary nature of data science requires more research to understand the field and identify strategies to develop a data driven work force.

1.4   Research Questions

To meet the demand for the future workforce and help undergraduate students develop data science skills, The Data Mine (TDM) at Purdue University, has undertaken an initiative in the discipline of data science. The Data Mine is an interdisciplinary living-learning community that allows students from various disciplines to enroll and learn data science skills under the guidance of competent faculty and corporate mentors. The residential nature of the learning community allows the undergraduate students to live, learn and socialize with peers of similar interests and develop a sense of belonging. The constant interaction with knowledgeable faculty, mentors, and working on real-world projects will allow novice learners to master data science skills and develop an identity. The study aims to characterize the effects of identity formation, sociology, and learning on the first-year undergraduate students enrolled in The Data Mine and answer the following research questions:

Quantitative: RQ 1: What are the perceptions of students regarding their identity formation, socialization opportunities, self-belief, and academic/intellectual development in The Data Mine?

Qualitative: Guiding RQ 2: How do students' participation in activities and interaction with peers, faculty, staff at The Data Mine contribute to becoming an experienced member of the learning community?

- **Sub-RQ 2(a):** What are the perceived benefits and challenges of participating in The Data Mine?
- **Sub-RQ 2(b):** How do students describe their levels of socialization and a sense of belonging within The Data Mine?
- **Sub-RQ 2(c):** How do students' participation and interaction in The Data Mine help them form their identity?

We plan to investigate the following research questions through this study. The current study aims to collect both qualitative and quantitative data and employ a mixed-method research design to evaluate the outcomes in terms of psychosocial and cognitive aspects.

1.5   Assumptions

The following assumptions are inherent to the study:

1. Participants answered all the questions pertaining to survey and interview honestly.
2. Participants answered all the questions on the sign-up form for the interviews honestly.
3. Participants have limited knowledge of data science, and joining The Data Mine helped them to enhance their data science knowledge.

1.6   Limitations

The study has the following limitations:

1. The sample size for the quantitative study is small; therefore, a follow-up survey study with a larger sample size would be required to perform statistical analysis in the future.
2. Since the same sample of students were surveyed for the pre and post data collection and same questions were used, the students' pre-test responses may influence the post-test responses.
3. The majority of students in this study are from the corporate partner cohort, and therefore to understand the impact of other academic units in The Data Mine on student experiences, future studies on student experiences from other academic units are required.
4. We have interviewed just three non-STEM background students, but more students from non-STEM backgrounds must be interviewed to understand the experiences of students in terms of identity formation, socialization, and learning.

1.7   Delimitations

The delimitations for this study are:

1. This study has focused on the communities of practice framework to understand how students develop identities, a sense of belonging, and learn while engaged in the residential

learning community. The author acknowledges that there are other theories of identity and socialization, but they are not considered for this study.

2. The study did not use a control group to compare the effects of identity, sense of belonging, and learning on the students. The study's intent was not to compare the effect of learning community with traditional classrooms but to understand the affordances that learning communities create to help students develop identity, a sense of belonging, and learning.

1.8   Dissertation Structure

This dissertation consists of nine chapters. Chapter 1 focuses on the problem identification, scope, limitations, and delimitations of research. Chapter 2 and Chapter 3 are focused on the literature review and the theoretical framework. Chapter 4 describes The Data Mine in detail by highlighting its key elements and key members. Chapter 5 provides the overview of the research design and methods used in the study. Chapter 6 presents the results for the quantitative study. Chapter 7 and Chapter 8 present the results of qualitative study: Chapter 7 represents the results of the thematic analysis and Chapter 8 presents the results of the narrative analysis. Chapter 9 is focused on the discussion, implications, and conclusion.

1.9   Summary

The increasing importance of data science jobs in the industry has caused the Higher Education Institutions (HEIs) and National Agency such as National Academies of Sciences, Engineering, and Medicine to turn their focus towards the discipline of data science. Imparting data science education at the undergraduate level will help the HEIs fulfill the demand of the industries and help students develop computational and data science skills. The prior studies have demonstrated that attracting and retaining first-year students is a huge challenge faced by HEIs across the United States. HEIs must engage first-year students in pursuing career-oriented skills that will also help them to mitigate the retention challenges. The intervention strategies such as residential learning communities, and focusing on teaching and mentoring practices, have helped institutions to retain students. But to the best of our knowledge, we have found very limited studies that focused on data science education. In this study, we intend to focus on the impact of the data

science learning community in helping undergraduate students to develop psychosocial and cognitive skills.

CHAPTER 2:   LITERATURE REVIEW

2.1   Need for identity Formation and Sense of Belonging/Socialization among the undergraduate students

Student attrition among undergraduate students is a challenge that the United States is trying to address. The retention of undergraduate students has become a complicated problem for the entire nation (Lyons, 2012; Tinto, 2007). Hess (2019) mentioned that there are just 41% of students who graduate from college in four years and 59% in six years. Whannell and Whannell (2015) pointed out that first year of undergraduate studies is the most challenging phase for all new undergraduates. The transition from high school experiences to undergraduate life requires students to develop a self and social identity, and those who fail to develop the identity experience identity conflict, creating a void and disengagement that results in attrition (Whannell & Whannell, 2015). The feeling of isolation and disconnection that students experience (Singh, 2018) during their undergraduate education jeopardizes student success and retention (O'Keeffe, 2013). Williford and Wadley (2008) revealed that students who experience a high level of disconnection socially and academically demonstrate a high probability of dropping out of college. Liu (2010) has demonstrated a relationship between student retention and the feeling of disconnection. Liu (2010) argued that students who demonstrate a low level of connectedness or a high level of alienation have a higher probability of dropping out the college.

The prior studies have identified two factors such as the inability of students to develop identity (Whannell & Whannell, 2015) and failure to experience socialization and a sense of belonging (Hoffman et al., 2002; O'Keeffe, 2013) are key contributors that lead to student attrition. The higher education institutions need to direct their efforts in refining the quality of students' learning experiences, specifically, at the introductory level. In the fiscal year 2015-2016, approximately 3.9 million undergraduate students with debt dropped from college (Barshay, 2017). The attrition rate for first-year students in the United States is approximately 30 percent (Schneider, 2010, p. 30). A single college graduate can contribute $355,000 to tax revenue during their lifespan (Trostel, 2015). O'Keeffe (2013) and Schneider (2010) reported that the attrition of first-year students enrolled in a four-year degree program caused a loss of US $6.18 billion in subsidies. Furthermore, federal and state governments lost US$2.9 billion as a grant to the first-year students who dropped after their first year. The economic impact of higher graduation and

retention rate will lead to lower unemployment rates, higher tax generation, and college graduates leading a healthier life compared to high school graduates (Baum et al., 2010; Lyons, 2012).

2.2   Identity Formation and Sense of Belonging/ Socialization among the undergraduate students

Identity formation among undergraduate students is crucial (Hazari et al., 2020; Whannell & Whannell, 2015). The identity formation process helps the individual to develop a perception of self and how one is viewed or recognized by others (Godwin, 2016; Godwin et al., 2016; Hazari et al., 2020). Identity crisis is an important issue that undergraduate students experience when they move from their high school life and enter into undergraduate studies (Whannell & Whannell, 2015). The transition from high school to university education has been a challenge for novice students as they face difficulties to accommodate and connect with other unknown individuals (Whannell & Whannell, 2015). Prior studies (e.g., Mann, 2001; O'Keeffe, 2013; Pargetter, 2000; Whannell & Whannell, 2015) have indicated that students also experience: solitude, lack of stability, and dissatisfaction due to the rough transition. The feeling of solitude, instability, and disassociation has been recognized as important reasons leading to student attrition (Scanlon et al., 2007; Singh, 2018). Whannell and Whannell (2015) conducted a study to examine the relationship between student attrition and identity formation. The study revealed that the rough transition of a student from high school to undergraduate education inhibits identity formation among the students. The study further mentioned that there is a cyclical relationship between student's identity, social context, and the emotional commitment to identity. This cyclical relation among the three variables yields a positive result when students have positive experiences within the university context.

The first year of undergraduate studies is critical. The study by Scanlon et al. (2007) has revealed the reasons that lead to disassociation or solitude among the students: (1) The interaction between the faculty and student has declined due to large class sizes. The students and faculty have less time to socialize and interact with one another. (2) The universities are more concentrated on research and less focused on teaching activities. The increasing focus on research has also created a gap between the faculty and students in the first year, as freshmen students are marginally involved in research. (3) The university's campus size or geographical area also hinders the interaction of students and faculty members. (4) The consistent increasing cost of tuition has

25

compelled the students to work for longer hours, resulting in a lack of interaction between faculty and students.

One important point to note here is identity and socialization are inseparable constructs. A recent study by Hazari et al. (2020), revealed that the identity formation in terms of recognition, interest and performance, and competence are incomplete if students do not experience a sense of belonging. Since the identities develops in a social setting and not in silos, elements of social setting such as context and peer are critical for identity formation. Weidman et al. (2014a) conducted a study to understand the identity formation in the students through the process of socialization. The study used a model of undergraduate socialization proposed by Weidman (1989) to understand the association between identity formation among college students and the impact of socialization in developing student identities. Weidman's model sketches the identity of students in terms of the three forms of social interactions: (1) interpersonal social interaction that occurs with a peer and a faculty, (2) intrapersonal relations that related to learning such as attending classes or taking courses and (3) integration refers to getting adjusted to the college life and campus activities (Weidman et al., 2014a).

Separating the aspects of social interaction in the form of socialization and sense of belonging from identity construct is difficult. Therefore, it is important to study both constructs jointly to retain and attract students to higher education institutions. The studies by (e.g., Hoffman et al., 2002; Tinto, 1987, 2007) have revealed that the first-year college experience is critical in determining student retention. The psychosocial factors such as lack of socialization, belongingness, lack of social support majorly contribute towards student attrition. Studies by O'Keeffe (2013) and Whannell & Whannell (2015) revealed that students demonstrating a strong association with the institution have a lower probability of dropping out. The socialization or sense of belonging allows the student to develop an association with peers, faculty, and context (Morrow & Ackermann, 2012). The sense of belonging or socialization allows the student to accommodate the transition from high school to undergraduate life (Hazari et al., 2020). The factors such as the smooth transition of the student (Whannell & Whannell, 2015) and the development of a sense of belonging (O'Keeffe, 2013) reinforces the development of identity among the students (Hazari et al., 2020; Weidman et al., 2014a). Hausmann et al. (2009) defined a sense of belonging as "their psychological sense of identification and affiliation with the campus community" (p. 650). In their study, Hoffman et al. (2002) have suggested that to inculcate the sense of belonging or sense of

affiliation among the first-year undergraduate students, higher education institutions must develop first-year seminar courses or establish learning communities.

2.3   History of Living-Learning Communities

Living-learning Communities (LLC) are not a novel concept. The idea of LLC in the United States has been adopted from the residential college model followed at Oxford and Cambridge (Thelin, 2011; Thelin & Gasman, 2003; Ujj, 2020). The first residential education in the United States began at Harvard University in 1636; the concept was later adopted by other prestigious institutions such as the University of Chicago, Princeton, Yale (Dunn & Dean, 2013). The founding fathers at Harvard University considered that residing and learning at the same place has a positive impact on the character of a student (Dunn & Dean, 2013). The history of living-learning communities has a special mention for Alexander Meiklejohn's *Experimental College,* started in 1927 at the University of Wisconsin (Brubacher & Rudy, 1997; Dunn & Dean, 2013; Jessup-Anger, 2015). The Experimental College was based on the foundation to help students live and learn together (Stassen, 2003) and improve student-faculty interaction (Smith, 2001). The study by Smith (2001) portrayed the experimental college as an intervention to promote experiential learning. Smith (2001) mentioned in their work that:

> The Experimental College tried to build community and create a seamless interface between the living and learning environment. The pedagogy stressed active learning, seminars, and assignments that asked students to put the theory they studied into practice, a radical notion at the time. Teachers were seen as advisors and facilitators of learning rather than as distant figures on a lectern (p. 5)

The reason for the increase in enrollment can be attributed to G.I. Bill (Micomonaco, 2011). The living-learning communities regained their popularity in the 1960s, which led to an increase in college enrollment (Smith, 2001). Later again, in the 1980s, colleges demonstrated an increased emphasis on the socialization aspect to create meaningful learning outcomes. The focus on the socialization aspect brought a major shift that led to a resurgence of the learning community (Dunn & Dean, 2013). From that point onwards the learning communities became an integral part of higher education institutions.

2.4 Impact of Learning Communities on Undergraduate Students

The student retention and dropout rates of first-year students pose a huge challenge for higher education institutions. According to Tinto (1997)

> Often, social and academic concerns compete, causing students to feel torn between two worlds so that students have to choose one over the other. Learning communities students draw these two worlds together (p. 605).

Understanding the students' needs in the initial years of undergraduate life is crucial as these needs determine student retention (Taylor & Stamatoplos, 1999). The two important reasons that lead to student attrition are lack of identity formation (Whannell & Whannell, 2015) and lack of sense of belonging (O'Keeffe, 2013) or socialization (Weidman et al., 2014b). The studies by Flynn et al. (2016) and Hoffman et al. (2002) have revealed that learning communities have proved to be effective in improving student retention as they help students develop a sense of belonging, socialize and form identities. Learning communities are considered as assimilation of academic and social mechanisms that provide meaning to student learning experiences. Experiences are integral to learning (Bobilya & Akey, 2002), and these experiences do not occur in silos; they happen when students interact with the context, with their peers and faculty. The interaction with context peers and faculty allows the students to develop a connection with the university or the institution. The study conducted by (Flynn et al., 2016) demonstrated the impact of the learning community on students. The study was conducted on the first-year engineering students enrolled in a Mid Atlantic University. It focused on comparing the effect of the learning community on students enrolled in the engineering learning community and students who were not part of the engineering learning community. The results revealed that students who were part of the engineering learning community demonstrated a significantly higher level of sense of belonging with peers and context than students who were not part of the engineering learning community. The learning community allows students to collaborate and get actively involved in the process of knowledge construction and acquisition (Cross, 1998). These communities emphasize the social construction of knowledge, meaning; the learner and the mentor are collaboratively involved in the process of knowledge construction (Cross, 1998; Whipple, 1987). The shift from a teacher-centered approach to learner-centered and active learning led to the emergence of the constructivist approach of learning.

Studies by (e.g., Flynn et al., 2016; Kuh, 2008; Rocconi, 2011) have demonstrated that students who actively participate in the learning community demonstrate a higher level of academic performance and also retention. Flynn et al. (2016) also demonstrated that learning community students demonstrated higher GPAs and retention rates than non-learning community students. Another important aspect of the learning community is that it promotes student engagement. A study by Rocconi (2011) has indicated that participating in the learning community may not directly impact student learning gains but promotes student engagement, and student engagement acts as a mediator to promote learning gains. The factors that promote student engagement within the learning community are interaction with faculty, interaction with peers (Pike, 1999), and living in a residence hall (Inkelas et al., 2007). These factors result in higher academic achievement and student retention (Pike et al., 2008; Pike, 1999).

These studies establish that learning communities are effective mechanisms to improve student retention and promote student learning. In this study, we intend to understand the impact of the learning community from the perspective of undergraduate students enrolled in The Data Mine at Purdue University. The study intends to explore the interdisciplinary nature of the learning community in terms of identity formation, sense of belonging/socialization and learning. Considering the rise of data science initiatives across the United States and the increasing opportunity of data scientists for the next 70 years (National Academies of Sciences, Engineering, and Medicine, 2018), the institutions must understand the needs and retain the undergraduate students. Therefore, this study will act as a precursor in the area of data science education to understand the needs and characterize the effect of identity, sense of belonging/socialization, and learning among the undergraduate students enrolled in The Data Mine.

2.5   Benefits of Living-Learning Communities

Living-learning communities—also known as residential learning communities—allow students of similar interest to live and learn together (Bobilya & Akey, 2002). Living together in the residence hall on campus has proved to be an effective method of increasing socialization among students (Hurtado et al., 2020). Especially focusing on the undergraduate students residing on campus in residence halls and learning together with peers of similar interest allow them to collaborate and learn together. The residential learning communities are known to provide an active, cooperative environment that fosters a sense of community among the students (Cross,

1998; Rocconi, 2011). Living together on-campus helps the students overcome the sense of solitude and disassociation that they experience due to their transition from high school to undergraduate life (Weidman et al., 2014). Prior studies (Bobilya & Akey, 2002; Cross, 1998; Stassen, 2003) have revealed that Living Learning Communities (LLC) have been known to have a positive influence on the academic success and socialization of the students.

Learning communities have proved to be effective in improving student retention (Tinto, 2007) by mentoring students (Maltby et al., 2016), helping them develop a sense of belonging (Hoffman et al., 2002) and sense of identity (Maltby et al., 2016; Priest, 2012). The study by Brower and Inkelas (2010) demonstrated that students who lived in the residential learning communities for the first year demonstrated higher academic self-confidence and engagement after three years than their fellow students. Tinto et al. (1994) conducted a mixed-method study on the first-year learning community students. The resulting study revealed that students found their peer community extremely supportive, which helped them navigate any academic and social issues (Stassen, 2003). The study by Spanierman et al. (2013) demonstrated that students who lived in the learning communities' residence halls demonstrated a higher sense of belonging than students not living in the residence halls. Other studies by Hoffman et al. (2002) and Inkelas (2008) revealed the students living in the living learning communities exhibit a higher sense of belongingness than students who do not live in the residential learning community. Therefore, learning communities can serve as effective structure that can help to bring the students of similar interest together and help them develop psychosocial skills.

2.6   Summary

Learning communities have been recognized as an important intervention to improve student retention and serve as a bridge to integrating academic and social components of the higher education institutions. Participating in the learning communities has been recognized as beneficial for undergraduate students as it promotes socialization, a sense of belonging, and identity formation. Studies have also indicated that it is impossible to separate a sense of belonging and socialization from identity formation. Helping the students develop identities by engaging them in social activities leads to student retention and academic success.

CHAPTER 3: THEORETICAL FRAMEWORK

This study is conducted in an interdisciplinary learning community focused on data science education. The situated nature of the learning community allows the students of various disciplines to live and learn together. The focus of The Data Mine is to foster a sense of community and helping novice learners to develop data science skills to become expert data scientists or data science professionals in future. Therefore, the study will use situated learning in combination with communities of practice and legitimate peripheral participation as the theoretical lens to understand the impact of The Data Mine on the socialization, identity formation, and sense of belonging for the undergraduate students enrolled in the learning community.

3.1 Theoretical foundation of Situated Learning

Before taking a deep dive into the situated learning framework, it is crucial to understand the foundational theories that gave rise to situated learning theory. The theoretical foundation for the situated learning theory is Vygotsky's theory of social learning and John Dewey's theory of experiential learning. Vygotsky and Dewey recognized and argued the importance of social interaction for experiencing learning. Vygotsky's theory of social learning focused on the two important aspects: 1) learning occurs when individuals interact with peers of similar interest and that it is crucial for "full cognitive development" (Kohler, 2010; p. 20). 2) The existence of Zone of Proximal Development (ZPD), that helps the learner to acquire knowledge from a superior or under the guidance of the faculty or mentor (Kohler, 2010). Vygotsky (1978) defined the ZPD as "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). Social interaction helps the learners interact with other learners in a social environment that helps develop cognition by internalizing the imparted knowledge (Vygotsky, 1978). Also, being in the zone of proximal development will help the learners to interact with faculty, mentors or experienced peers and acquire a higher level of skills required to become experts in a particular field. The Data Mine is a residential learning community that provides the opportunity to interact socially with peers and learn under the guidance of the faculty or corporate mentor to develop data science skills. Another

important component that The Data Mine focuses on is creating meaningful learning experiences by immersing the students in a collaborative learning environment.

Dewey (1986) argued that learning cannot be experienced in the four walls of the classroom. To experience the true meaning of learning, the learner needs to immerse themselves into the learning environment and interact with peers. Dewey's theory of social and experiential learning lays the next level of foundation for the situated learning theory. Dewey's principle of continuity and principle of interaction are central to the idea of experiential learning (Giles Jr & Eyler, 1994). The principle of continuity states that experience is not a singular event but a chain of events. Also, the role of a mentor is crucial in guiding the learner to follow the correct sequence of events that could lead to learning gains. The principle of interaction claims that experiences result from the interaction between the learner and the context (Giles Jr & Eyler, 1994). The two principles proposed by John Dewey stress the importance of social interaction for learning to occur. Dewey (1986) stated that "I believe finally that education must be conceived as a continuing reconstruction of experience; that the process and the goal of education are one and the same thing" (p. 13).

The learner's social interaction with context, peers, or faculty also defines the situated nature of learning. From Dewey's perspective, school is a space to socially interact, learn together, and experience learning through progressive education. This active interaction between the learner and context allows the learner to search for the true meanings of knowledge and develop individual identity (Lindsay, 2001). Dewey (1986) viewed progressive education as a means to encourage student participation. Dewey (1986) argued that progressive education focuses not only on the participation of the student but it actually focuses on the active involvement of the student in the process of learning; traditional education fails to do so. Traditional education is teacher centered and has less focus on the student learning experiences.

From the perspective of Dewey and Vygotsky, one can deduce that the theory of social learning and philosophy of social and experiential learning laid the foundation for situated learning and communities of practice. Considering the two theories proposed by Dewey and Vygotsky, this study will further investigate students' experiences in The Data Mine, specifically, focusing on their journey from being a novice to expert members of The Data Mine, understanding opportunities for socialization / sense of belonging in The Data Mine, and identifying opportunities for learning new skills in The Data Mine.

3.2   Situated Learning

The concept of situated learning was proposed by Lave and Wenger (1991). The theory is based on the work conducted by Jean Lave with Vai and Gola tailors in Liberia. Lave (1996) conducted the study for five years, from 1973 to 1978, to understand the Tailor's Alley apprenticeship model in Liberia, West Africa. Lave (1996) observed that apprentices did not undergo any formal training in tailoring. They learned by observing their master and other peers. Some of the intriguing findings of the study were: 1) tailors were poor; they could afford just one apprentice at a time, but the master and apprentice worked hard to learn the art in the informal setting. 2) Around 85 percent of the apprentices who underwent the training mastered the skill and opened their tailor shops. 3) The practice of tailoring was socially situated in the Tailor's alley in Liberia. There was a very informal interaction between the apprentice and master. Apprentices interacted with their peer apprentices in the Tailor's Alley to learn from one another. Working regularly in a setting where every individual was immersed in mastering the art of tailoring through interaction with master and peers fostered situated learning. Since the context of the Tailor's Alley played a key role in helping the apprentices to develop the skill, it was impossible to separate the context (Tailor's Alley) from learning (mastering the tailoring skills). Lave also concluded that informal learning demonstrated by apprentice tailors in Liberia is extremely different from cognitive apprenticeship demonstrated in a US school, and both are correct in their own situated contexts. Lindsay (2011), in her study, mentioned another research that was conducted by Lave (1988) in California to understand the buying behavior and application of mathematics in a real-life situation. Lave conducted the "best buy" experiment to understand adults' calculation skills while purchasing groceries and how they perform the same calculation in a school setting. It was observed that buyers who made correct purchasing calculations in the grocery store failed to perform the same calculation in a school setting and vice-versa. The findings of the studies about Vai and Gola tailors of Liberia and the grocery shopping experiment revealed it is impossible to separate the learning from context, and hence the situated nature of the learning was revealed in both studies.

Situated learning theory focuses on the learning experiences that occur within a specific learning environment because all forms of learning are grounded within the *situation* (Greeno, 1997; Lave & Wenger, 1991; Zheng, 2010). The *situation* is also referred to as the *context* that holds the primary place in situated learning. Anderson et al. (1996; p. 6) emphasized the situated

nature of the learning that "all knowledge is specific to the situation in which the task is performed." The situated learning environment provides the context for the learners to interact and involve in collaborative knowledge construction with peers and mentors (Choi & Hannafin, 1995; Zheng, 2010). As mentioned earlier, the theoretical roots of the situated learning are grounded within the Vygotsky (1980) socio-constructivist theory. The socio-constructivist theory emphasizes that learners and society are inseparable components, and learning experiences are developed within the social context. Social context allows learners to interact with peers and mentors to create collaborative learning experiences (Simina, 2012; Vygotsky, 1980). The situated learning creates the zone of proximal development (ZPD) for the learners and mentors. ZPD allows the mentors to scaffold the knowledge of the learners within the authentic context that allows the learners to perform beyond the expected boundaries (Simina, 2012; Vygotsky, 1980). According to Lave (1991)

> ….learning is recognized as a social phenomenon constituted in the experienced, lived-in world, through legitimate peripheral participation in ongoing social practice; the process of changing knowledgeable skill is subsumed in processes of changing identity in and through membership in a community of practitioners; and mastery is an organizational, relational characteristic of communities of practice (p. 64)

The two important sub-theories in the situated learning theory are the community of practice and legitimate peripheral participation (Kirk & Macdonald, 1998); see Figure 2. Kirk and Macdonald (1998) emphasized developing an understanding of the two concepts. The notion of "community of practice (CoP)" emphasized the "learning partnership among people who find it useful to learn from and with each other about a particular domain. They use each other's experience of practice as a learning resource" (Wenger et al., 2011, p. 9). The communities of practice help determine the different possibilities of learning (Lave & Wenger, 1991) and contribute to developing the learner's identity (Kirk & Macdonald, 1998). The second notion of the theory, "legitimate peripheral participation," focuses on the participation of the novice learners within the communities of practice and their abilities to observe the peers develop relevant skills required to become expert learners (Kirk & Macdonald, 1998). The two notions of the community of practice and legitimate peripheral participation can be associated with the framework of learning communities at higher education institutions. The two notions represent learners' interaction with

other members such as peers and mentors in an integrated specific learning context. This opportunity of social interaction among the community of practice members allows them to engage in collaborative knowledge construction and develop their identities (Priest et al., 2016).
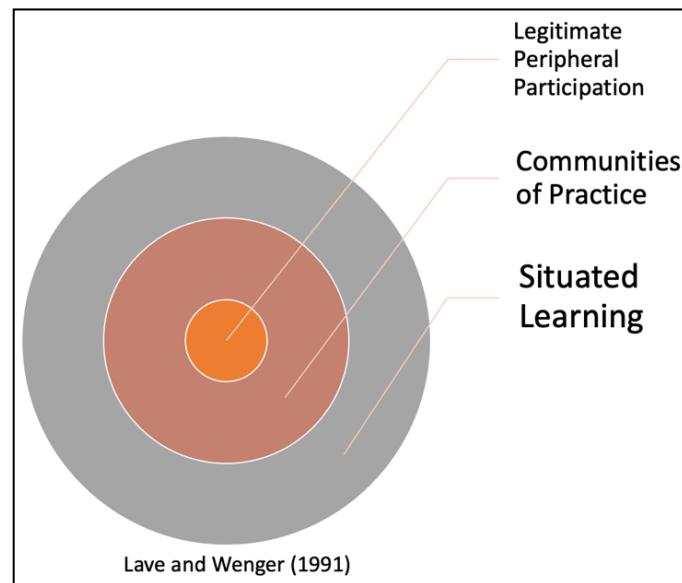
Figure 2: Theoretical Framework

The prior studies on situated cognition have revealed that context and social relationships are integral to student learning experiences. The study by Zheng (2010) indicated that students in situated learning demonstrate higher learning outcomes than students in the traditional classroom. The study also analyzed the association among the different variables of the situated learning. The results from the study demonstrated that "significant correlations were observed among authentic context, authentic activity, authentic assessment, collaboration, multiple perspectives, performance, and scaffolding" (Zheng, 2010; p. 479).

Priest et al. (2016) conducted a study to evaluate the social practices of first-year students in the learning community. The study used a situated learning framework to understand student participation. The results of the study indicated that students in the learning community found the context motivating to pursue the learning outcome. The learning community helped them to interact with others and create meaningful experiences.

3.3 Communities of Practice

The communities of practice framework was proposed by Lave and Wenger (1991). The researchers widely use this framework to understand the community members' interaction, growth, and identity formation in a situated setting. Wenger and Wenger (2015) have applied the communities of practice framework in several domains such as business organizations, government, education, professional associations, social sectors, and online communities. In this study, we intend to investigate identity formation, sense of belonging / socialization, and learning in a residential learning community through the lens of communities of practice. According to Wenger and Wenger (2015), a community of practice is not a group of friends or neighbors, but it is a group of individuals of similar interests who engage and work collaboratively to learn new skills in a specific context. In the words of Wenger and Wenger (2015), "communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" (p. 1).

To refer to a community as communities of practice, Wenger and Wenger (2015) have specified that it must possess the following characteristics: a domain, a community, and a practice. The *domain* refers to the area of activity or interest shared by the members of the community. The domain represents the common interest that members of the community share (Wenger, 2004). This study domain can be referred to as The Data Mine where students of similar interest collaborate to learn and apply data science skills. Wenger (2004) defined domain as

> The area of knowledge that brings the community together, gives it its identity, and defines the key issues that members need to address. A community of practice is not just a personal network: it is about something. Its identity is defined not just by a task, as it would be for a team, but by an "area" of knowledge that needs to be explored and developed (para. 13).

The *community* characteristic of the communities of practice framework refers to the interaction that occurs between the members of the community. The members share their interests, thoughts, collaborate, and learn together as a group under the guidance of a mentor or guide. Wenger and Wenger (2015) also stress the point that the community needs the participants to interact. For example, a group of students in an American school cannot be referred to as a community even though they may share many characteristics in common. To be called a community they need to interact with one another for making the learning happen. The community

aspect assures that none of the members operate in a silo rather than work as a team to acquire the desired skills. In The Data Mine, students work in a group to master the data science skills through interaction with faculty, mentors, or peers. In the words of Wenger (2004), the community is defined as

> The group of people for whom the domain is relevant, the quality of the relationships among members, and the definition of the boundary between the inside and the outside. A community of practice is not just a website or a library; it involves people who interact and who develop relationships that enable them to address problems and share knowledge. (para. 14)

The *practice* tenet of the communities of practice refers to the shared repertoire of ideas, experiences, stories, tools, methods that members of the community create as they interact with each other (Wenger & Wenger, 2015). The created artifacts are used by the members of the community to solve problems or challenges. In the case of The Data Mine, students work together in groups and they share their challenges and learning in person or through online discussion forums. They also apply data science concepts to create codes of data that could be used as a base to solve similar problems. Wenger (2004) defines practices as:

> The body of knowledge, methods, tools, stories, cases, documents, which members share and develop together. A community of practice is not merely a community of interest. It brings together practitioners who are involved in doing something. Over time, they accumulate practical knowledge in their domain, which makes a difference to their ability to act individually and collectively. (para. 15)

Participation is integral to communities of practice. The active participation on the part of members with the situated context inculcates a sense of community, allows to create learning artifacts through the act of practice, leads to identity formation and negotiations on finding the true meaning of experiences (Lave & Wenger, 1991; Lindsay, 2001). Indulging in the act of learning with peers and experts allows the members to mutually engage in the learning environment and develop a shared meaning of the learning experiences. The meaning of the learning experience is dynamic, they keep on changing as the goals of the joint enterprise change (Li et al., 2009).

The communities of practice establish a set of parameters to evaluate the membership of its participants in terms of mutual engagement, joint enterprise, and shared repertoire within the context (Smith et al., 2017; Wenger, 2011). Mutual engagement is achieved by setting the rules or

norms for the community members to operate within the communities of practice, it establishes the guidelines on what members are supposed to do as they engage in the practice (Smith et al., 2017). This mutual engagement also provides an identity to the participants as they mutually decide on the rules for the community of practice (Contreras Aguirre, 2019). The joint enterprise is the second parameter to evaluate the membership. The joint enterprise allows the members of the community to collaborate and define the purpose of the community of practice (Smith et al., 2017). The joint enterprise allows the members to come together and work towards a common goal (Li et al., 2009). The third criteria for evaluating the role of a member in the creation of a shared repertoire of tools, techniques, stories, ideas are based on the mutual agreement and purpose of the communities of practice (Smith et al., 2017). The shared repertoire binds the member of a community of practice together through a commonality they share in terms of resources that they developed (Contreras Aguirre, 2019). As the community of practice grows, the interaction among the members increases and that results in constant negotiation and renegotiations of ideas, thoughts, and perspectives to make the existence of the community more valuable and meaningful.

The negotiations of meaning within the communities of practice take place in two forms: participation and reification. Participation leads to interaction and collaboration that occurs among the members of the community whereas reification refers to the creation of new learning artifacts to support the members of the community (Smith et al., 2017). Furthermore, Smith et al. (2017) emphasized both participation and reification are equally important, they supported their argument with the help of instances. In instance 1, Smith et al. (2017), mentioned that when a student is facing some challenges in solving a problem, it is always a great idea to approach a competent peer who can solve the problem. Such an act will foster participation. In instance 2, they mentioned if a student is working on a problem and it is complicated, one can always write the solution and share the artifact with others for future use, this leads to reification. This is important to note that both the acts of participation and reification contributed to the process of meaning-making. The constant negotiations of meaning allow the members to develop new practices by creating a shared repertoire and improve interaction among the members (Lindsay, 2001). The negotiations and renegotiations with the meaning of learning experiences also help the identities of the member to evolve, as they progress towards becoming an expert member of the community.

In the context of higher education, the communities of practice are witnessed as an approach to improve student participation, collaboration within the educational settings. (Eshchar-Netz &

Vedder-Weiss, 2021; Fearon et al., 2012; Tomkin et al., 2019). The studies have demonstrated that participation is integral to communities of practice as it allows the members to actively participate and engage in social learning (McDonald & Cater-Steel, 2016). Active participation on the part of members within the situated context inculcates a sense of community, allows them to create learning artifacts through the act of practice, leads to identity formation and negotiations on finding the true meaning of experiences (Lave & Wenger, 1991; Lindsay, 2001). The constant negotiations of meanings allow the members to develop new practices by creating a shared repertoire and improve interaction among the members (Lindsay, 2001). The negotiations and renegotiations with the meaning of learning experiences also help the identities of the member to evolve, as they progress towards becoming an expert learner through legitimate peripheral participation within the communities of practice (Lave & Wenger, 1991).

3.4 Implementation of Communities of Practice in Higher Education

With the increasing focus on social learning, higher education institutions are inclining towards adopting novel approaches such as communities of practice (McDonald & Cater-Steel, 2016). Adopting communities of practice framework will help the instructors to follow active-learning approaches to encourage student engagement and improve participation (Fearon et al., 2012; Moulton et al., 2017; Orsmond et al., 2013; Tomkin et al., 2019). The other implications of participating in communities of practice are identity formation (Orsmond et al., 2013), leadership skills development (Green et al., 2017), improvement in student retention, and sense of belonging (Zammit et al., 2017).

Some of the prior studies demonstrated the need for establishing communities of practice within higher education institutions. The study conducted by Tomkin et al. (2019) discussed the implementation and outcomes of establishing communities of practice for faculties to improve teaching in STEM classrooms. The study revealed institutions are still focused on the old-fashioned method of imparting STEM instructions. Numerous seminars and workshops have been conducted to demonstrate the benefits of active learning strategies within the classroom setting, but faculties refrain from using those techniques especially in the large STEM classrooms. Therefore, the authors in the study proposed to create communities of practice, that allow the faculties from introductory STEM courses to collaborate and share their pedagogical techniques that would help to improve the teaching practices. The study used a mixed-method design to

perform the analysis. The study compared the teaching approaches for the faculties who participated in the communities of practice (COP Faculties) with those who were not members of the communities of practice (Non-COP Faculties). The study found that the faculties who participated in the communities of practice (COP Faculties) developed and implemented student-centered approaches in their classroom teaching, by engaging students in active learning and problem-solving. In contrast, Non-COP Faculties continued with their lecture-based teaching.

Since communities of practice is an effective method to impart social learning, Lawthom (2011) performed a study in the area of psychology that argued the effectiveness of the communities of practice on psychology students through a hypothesized case study. In this study, Lawthom (2011), hypothesized that universities in the United Kingdom (UK) can develop learning communities in the area of psychology by employing the principles of communities of practice. The author described the communities of practice as a framework that promotes learning through collaboration and social interaction. The study used the definitions provided by Lave and Wenger (1991) to describe the communities of practice framework. The author hypothesized that the outcome of participating in the communities of practice will help psychology students (novice learners) to develop the identity of *community psychologists* as they will work in a real context and interact with faculty and mentors from the area of community psychology. The author used a case study approach to hypothesize how undergraduate students will learn the skills possessed by *community psychologists* while working on real problems with experts and mentors. Lastly, the author discussed the impact of participation on novice learners, in terms of developing identities, mastering concepts of community psychology, and interacting with real patients.

Participating in communities of practice fosters collaboration among the participants and the interaction is explained by the mutual engagement between the mentor and mentee. A study conducted by Degn et al. (2018) used the lens of communities of practice to evaluate if the research groups at higher education institutions operate as non-hierarchical communities of practice or formal organizational structures such as teams. The study investigated four research groups, two in Netherlands and two in Denmark. Researchers interviewed and observed the participants and working atmosphere for one year. The study focused on the three structural characteristics of the communities of practice: the *domain, community, and practice*. The findings of the study revealed that research groups demonstrate the characteristics of the communities of practice as the mentor and mentee have an informal relationship, they work and set goals together, mentors help mentees

in developing identities that bring the groups together. The senior members of the group help the novice learners in identifying research problems, also groups focus on delegating and sharing responsibilities among the group members as each member is responsible for the group's success. Based on these findings, researchers concluded that high-performing research groups share more characteristics with communities of practice and less with formal organization entities, such as teams.

Table 1 below represents the studies discussed above demonstrating how communities of practice were characterized and the research methods used for documenting the outcomes of each study. From Table 1, one can infer that two studies (Lawthom, 2011; Tomkin et al., 2019) characterized communities of practice as a framework for social interaction and learning, the studies used definitions provided by Lave and Wenger (1991), to describe its characteristics. Whereas the study conducted by Degn et al. (2018) characterized the communities of practice based on the three structural elements: domain, community, and practice, also the three structural units informed the data analysis section of the study. Furthermore, to document the outcome of the communities of practice, qualitative methods were employed in the form of interviews, observations, and fieldwork for the two studies (Degn et al., 2018; Lawthom, 2011), whereas Tomkin et al. (2019) used a mixed-method approach.

Table 1: Represents Methods used for describing the COP* and documenting the Outcomes.

Study	Description of COP*	Outcomes evaluated	Research Methods for documenting the outcomes
Tomkin et al. (2019)	The COP was characterized as the platform for conducting regular meetings that allowed members (Faculties) to interact, participate and learn from one another by involving in collaborative activities.	The study was focused on evaluating the effectiveness of teaching practices for COP and non-COP Faculties.	Mixed method Design **Data Collection:** Classroom Observations based on COPUS protocol **Data Analysis**: Inferential Statistics
Lawthom (2011)	COP was viewed as a framework that promotes learning through collaboration and social interaction.	The study developed a hypothetical case to rationalize how participating in a Learning Community designed on basis of COP principles will help students **develop identities** of *community psychologists*	Case Study
Degn et al. (2018)	The COP was characterized as a group that brings together people of similar interests. The author further characterized the elements of COP: the domain, community, and practice.	The study focused to evaluate in what ways high-performing research groups shared characteristics of COP and how they differed from teams.	Ethnography **Data collection:** Observations, Interviews **Data Analysis:** Thematic analysis based on pre-defined criteria (*apriori*)

*COP- Communities of Practice
**LC- Learning Community

3.5   Identity formation within the Communities of Practice

One of the core outcomes of participation in communities of practice is developing an identity. The prior researches conducted by (e.g., Degn et al., 2018; Lawthom, 2011; Orsmond et al., 2013; Teeuwsen et al., 2014) revealed that identity formation is a critical outcome of communities of practice. Most of the studies focused on how participating in communities of practice can help novice learners to develop an identity and evolve as expert learners. Legitimate peripheral participation (LPP) is integral to communities of practice (Wenger, 1998); therefore it is important to understand how the dynamic process of identity formation occurs when individuals negotiate and re-negotiate their identities as they engage in new contexts (Carlone et al., 2015; Carlone & Johnson, 2007; Gee, 2001; Hazari et al., 2020). Identity is not just the process of realizing 'who am I'; it is also a process of determining 'how others see me' (Godwin et al., 2016). Therefore, communities of practice help learners develop identities helping the novice learners to move from peripheral participation to full participation by providing them opportunities to interact with other members of the community (Lave & Wenger, 1991).

Since the communities of practice are focused on a particular domain or discipline, it helps the learners to develop identity related to that domain or discipline (Lave, 2004; Pattison et al., 2020). But it is important to note that learners in the collaborative context such as in communities of practice, develop a socially situated disciplinary identity (Y. S. Kim & Merriam, 2010; Pattison et al., 2020; Xie et al., 2017). The communities of practice group together, the people of similar interest and provides opportunities for them to interact, participate and learn (Lave & Wenger, 1991). Since identities are dynamic, they are constantly negotiated within the social context. The previous study conducted by Godwin and Potvin (2017) has demonstrated how the concepts of critical agency theory were used to explain identity formation within communities of practice. Godwin and Potvin (2017) argued that critical agency theory examines the 'type' of person an individual is developing into and how experiences shape the identity of the learner within the disciplinary community. Indulging in experiences allows them to move from peripheral to full participation in the communities of practice. Furthermore, the study by Pattison et. al. (2020), cited the study conducted by Carlone and Johnson (2007) to demonstrate how the three constructs of disciplinary (science) identity, *performance, recognition, and competence* were operationalized in terms of situated identity. Carlone and Johnson (2007) described the *performance* as "social performances of relevant scientific practices—e.g., ways of talking and using tools". *Recognition*,

as "recognizing oneself and getting recognized by others as a 'science person'". *Competence*, as the "knowledge and understanding of science content" (p. 1191). The two studies by Godwin and Potvin (2017) and Carlone and Johnson (2007), provides evidence on how one can integrate the constructs of disciplinary identity within the context of communities of practice to understand the dynamic process of identity formation.

3.6   Socialization and Sense of belonging within Communities of Practice

The other important outcome of engaging in communities of practice is socialization and developing a sense of belonging (O'Keeffe, 2013). The community of practice brings the people of similar interest together and provide the opportunity to interact and learn from other members (Bobilya & Akey, 2002; Hurtado et al., 2020; Pike, 1999). Allowing the members to socialize and develop a sense of belonging is integral to communities of practice. The constant interaction between the newcomers with the old-timers allows the novice learners to move from periphery to center of the communities of practice (Lave & Wenger, 1991). The interaction of the novice learners with experts helps them develop an identity through constant participation, and this helps the members to develop a sense of belonging and commitment (Handley et al., 2006). Lave and Wenger (1991) argued that the collaboration and participation in the various activities allow the members to engage in a joint enterprise that helps to define a common goal and objectives for the communities of practice. The mutual engagement among the members provides the opportunity to define the rules of operation for the community of practice. Finally, once the goals and rules are defined, members engage themselves actively in the process of learning and create a shared repertoire of tools, techniques, stories, ideas, experiences as evidence of social learning within the communities of practice.

3.7   The intersection of communities of practice and learning communities in the Higher Education

Since the concept of communities of practice is versatile, it can be applied to learning communities in many forms. Learning communities in higher education is also an important intervention that has used the principles of the community of practice to foster participation among the members (Tinnell et al., 2019). The learning communities are platforms that promote learning by engaging the learners in a social environment (Bobilya & Akey, 2002), whereas communities

of practice perceived as a theoretical framework help to explain the interaction among the members that takes place within a learning community (Tomkin et al., 2019). Learning communities and communities of practice share many common features, for instance, they both focus on bringing the people of similar interest together and providing them a conducive environment to interact. In the subsequent paragraphs I will discuss two studies conducted in the context of learning communities and used communities of practice as their theoretical framework.

Priest et al. (2016) conducted a study to understand the experiences of participating in a learning community for first-year undergraduate students. The study used a lens of situated learning and communities of practice to explore the following aspects: "a) access and motivation (exploring how and why students entered into communities); b) meaning of participation (exploring students' perceptions of experiences of membership in communities); and c) learning trajectories (exploring how participation in a learning community influenced students' next steps as learners in general, as college students, or their professional aspirations)" (p. 364). The results of the study revealed that it was motivating for the first-year students to participate in the learning community, as it allowed them to collaborate with faculty and peers and engage in collaborative learning. It is important to note that learning communities provides a conducive environment to students and communities of framework acts as a lens to understand the experiences better from a theoretical standpoint.

Another study conducted by Tinnell et al. (2019) at the intersection of learning communities and communities of practice focused on how faculty members in the Faculty Learning Community (FLC) develop methods to improve their pedagogical approaches for teaching engineering courses. The concept of FLC was popularized by Milton D. Cox. The objective of a faculty learning community is to bring together six to twelve members and engage them in an active, collaborative environment and provide them the opportunity to improve their teaching practices (Cox & Richlin, 2004). The study by Tinnell et al. (2019) focused on a FLC that was created on the principles of communities of practice. The study demonstrated that for teaching engineering courses, the faculties across the HEIs use non-collaborative techniques that make the knowledge acquisition process for the learners boring and disinteresting. Therefore, HEIs need to help faculties learn and implement pedagogies that promote student collaboration. In this study, faculty members from the engineering department collaborated to form a Faculty Learning Community (FLC) based on the principles of communities of practice. The FLC met regularly and

shared their experiences regarding teaching practices that they followed and brainstormed together to improve the practices. Participation in the learning community allowed the faculties to learn from one another. The results of the study demonstrated that participating in the FLC helped the faculties not just improved their teaching strategies but also improved their research group collaborations.

3.8 Legitimate Peripheral Participation

Legitimate Peripheral Participation (LPP) is central to communities of practice. The concept of LPP helps to understand the relationship between a newcomer and an expert, in terms of identity formation, social interaction, knowledge, and practice in the communities of practice (Lave & Wenger, 1991). The LPP is a strategic region that is not completely inside or completely outside (Lave, 1991; Lave & Wenger, 1991; Wenger, 2004), it is the position where all the new members of the community start their journey. Lave and Wenger (1991) described LPP as follows:

> Legitimate Peripheral Participation provides a way to speak about the relations between newcomer and old-timers, and about activities, identities, artifacts. It concerns the process by which newcomer becomes a part of community of practice and knowledge. LPP is a depicter of engagement in a social practice that entails learning [through participation] as an integral constituent (Lave & Wenger, 1991; p. 29).

The new members are assumed to have limited learning experiences, as they interact with other members within the communities of practice, they develop their own identities and move from being novice learners to expert learners. Lave and Wenger (1991) argued the importance of LPP as learning happens when individuals interact with one another. The interaction with other members of the community allows the newcomer to developing the required skills and identity through full participation. Full participation is said to occur when "individuals adopt and perform the valued practices of that community and in so doing contribute their own experience and modify practice and shift values" (O'Donnell & Tobbell, 2007; p. 315). Wenger (1998) argued that the identity formation is the result of social learning that occurs in the communities of practice when members of the community interact with one another. Lave and Wenger (1991) further emphasized that "learning transforms who we are and what we can do, it is an experience of identity" (p. 215). Because according to Lave and Wenger learning just does not imply learning new skills or gaining

new knowledge rather it is a process that helps an individual *"to become a certain person or, conversely, to avoid becoming a certain person"* (Lave & Wenger, 1991; p. 215).

Lave and Wenger (1991) argued that the concept of identity cannot be separated from the social learning context. Therefore, focusing on the learning trajectories of the newcomer or novice learner will help to understand the identity formation of the learner within the situated context (Wenger, 1998). The learning trajectories define the current position of the newcomer and also the path the individual will take to become an expert in the field (Lave & Wenger, 1991). It is also important to note that newcomers need to adapt to the practices of the communities of practice by constant negotiation and renegotiations with other members of the community (Davis, 2006; O'Donnell & Tobbell, 2007). These constant negotiations and renegotiations help them to develop a perception of self that allows the newcomers to shape their identity (Davis, 2006).

The prior studies (Priest et al., 2016; Rehak, 2018; Ujj, 2020) have demonstrated that communities of practice and legitimate peripheral participation have been used as the theoretical lens to evaluate the impact of socially situated communities such as learning communities, and other professional communities. The combination of the two theoretical lenses helps the researchers to understand the interaction among members and track the learning trajectories of the members from being novice learners to expert members of the community. For instance, Teeuwsen et al. (2014) used the lens of legitimate peripheral participation to understand the identity formation of two part-time Ph.D. students working under the guidance of an experienced advisor. The study used an interpretive qualitative design to explore the growth of the two learners. The results of the study demonstrated that the constant interaction of the two part-time Ph.D. students with each other and with their advisors helped them to shift from peripheral participation to full participation within the communities of practice. The study focused on the struggles of part-time Ph.D. students concerning their identity formation and minimal opportunity for socialization. The study emphasized the initiative taken by the advisor, that helped the two part-time students to move from the periphery to the center of a learning trajectory. Susan as an advisor and experienced member of the research community provided an opportunity to her part-time Ph.D. students to get involved in a directed study course under her supervision. The course served as the platform for the two students to collaborate on a research project under the guidance of an experienced advisor and develop research skills, form identities, and experience a sense of belonging with the research community.

Another interesting study was conducted by O'Donnell and Tobbell (2007), in the context of the transition of adult learners to higher education. The study used legitimate peripheral participation (LPP) in the context of communities of practice to explore the learning, identity, and sense of belonging experienced by adult learners when they transition to undergraduate studies. The study interviewed 17 adult learners from different socio-economic backgrounds enrolled in a course "an introduction to university study for mature students". This course was mandatory for all adult learners who did not possess a formal high school qualification. The findings of the study demonstrated that the only route to move from the periphery to full participation is through engaging in the process of learning with other members of the community. The study revealed that the instructor taught skills such as essay writing to the adult learners as a part of the course, but learners learned those skills when they started engaging in the activities that required them to practically use those skills. The study also demonstrated that a shift of identity and sense of belonging only happened for those who were intrinsically motivated and were ready to engage in the social learning process.

3.9   Summary

The study intends to use the theoretical framework of situated learning along with two-sub frameworks communities of practice and legitimate peripheral participation. The objective is to understand the impact of social interactions of students with context, peers, and mentors within the learning community. Also, we intend to explore the ways that The Data Mine empowers these novice learners to develop data science skills to become experienced data science learners.

CHAPTER 4: THE DATA MINE

This study revolves around The Data Mine, a living learning community at Purdue University. The study intends to describe the journey of the students enrolled in The Data Mine. This chapter aims to explain the structure, key elements, role of students, staff, and faculty in The Data Mine. This chapter has been written after consulting the Interim Director of Integrative Data Science Initiative at Purdue University, and based on the content available in The Examples Book (Amstutz et al, 2021). This chapter starts with the history, highlights the growth of The Data Mine, focuses on the key elements and role of TAs, faculty, and staff in The Data Mine.

The Data Mine is an interdisciplinary learning community, established as a part of the Integrative Data Science Initiative, within Purdue University, with an intent to impart data science education to students from all disciplines. The Statistics Living Learning Community (STAT-LLC) was precursor of The Data Mine. The STAT-LLC was started in 2014 and lasted till 2019. The objective of the STAT-LLC was to help students develop statistical skills by immersing them into a living learning community environment. The STAT-LLC was restricted only to the sophomore students enrolled in any major at the Purdue University. Twenty students were recruited each year and were engaged in an active learning and research environment. The STAT-LLC turned out to be a big success. The success of STAT-LLC motivated the Principal Investigator of the STAT-LLC to pilot an extended and more inclusive model of a learning community.

In the year 2018-19, The Data Mine was piloted with 100 students. The Data Mine was a much larger living learning community than the STAT-LLC. The Data Mine was an opportunity for students from all domains (STEM and non-STEM) and all education level to live and learn data science skills. In the pilot year, The Data Mine just offered a one-credit hour seminar course in the fall and spring semester. The pilot program for The Data Mine was a great success and allowed it to expand from 100 to 600 students in their second year. Currently The Data Mine has approximately 800 students from various majors and education levels. The demographics of The Data Mine is demonstrated below in Figure 3 and Figure 4.
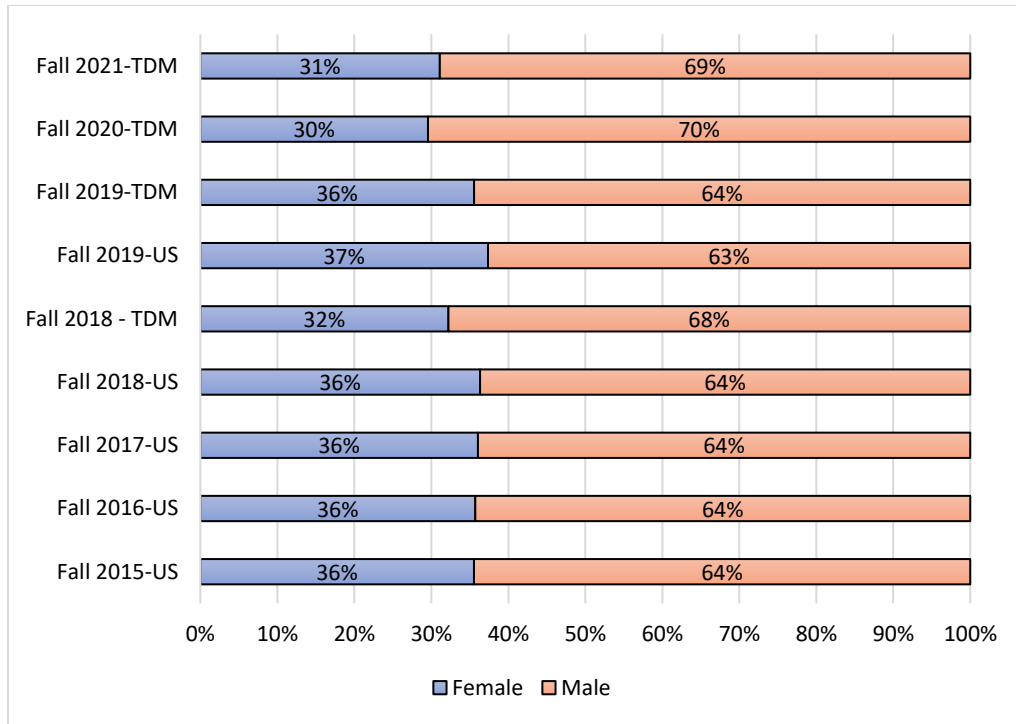
Figure 3 : Gender Demographics of The Data Mine students and STEM Students across the United States

Figure 3 shows a gender-based comparison of students enrolled in The Data Mine (TDM) from Fall of 2018 to Fall of 2021 with the students enrolled in an undergraduate four-year STEM degree course in the United States from Fall of 2015 to Fall of 2019 (National Center of Education Statistics, n.d.). From Figure 3, we can infer that the number of male and female students in The Data Mine is proportionate to the STEM enrollment in the undergraduate degree across the United States. Moreover, the actual number of female and male students has increased from 2018 to 2022 in The Data Mine, as per the data provided in Appendix A.

Figure 4 shows the comparison based on ethnicity/race/nationality with students enrolled in The Data Mine to that of the students enrolled in Purdue University (PU) (Purdue University, n.d.). It is evident from Figure 4 that the student population in The Data Mine is more diverse in comparison to the Purdue University enrollment, as white students are proportionally less in The Data Mine with better representation from other groups. However, the percentage of minority students in The Data Mine is proportionate to the demographics of Purdue University.
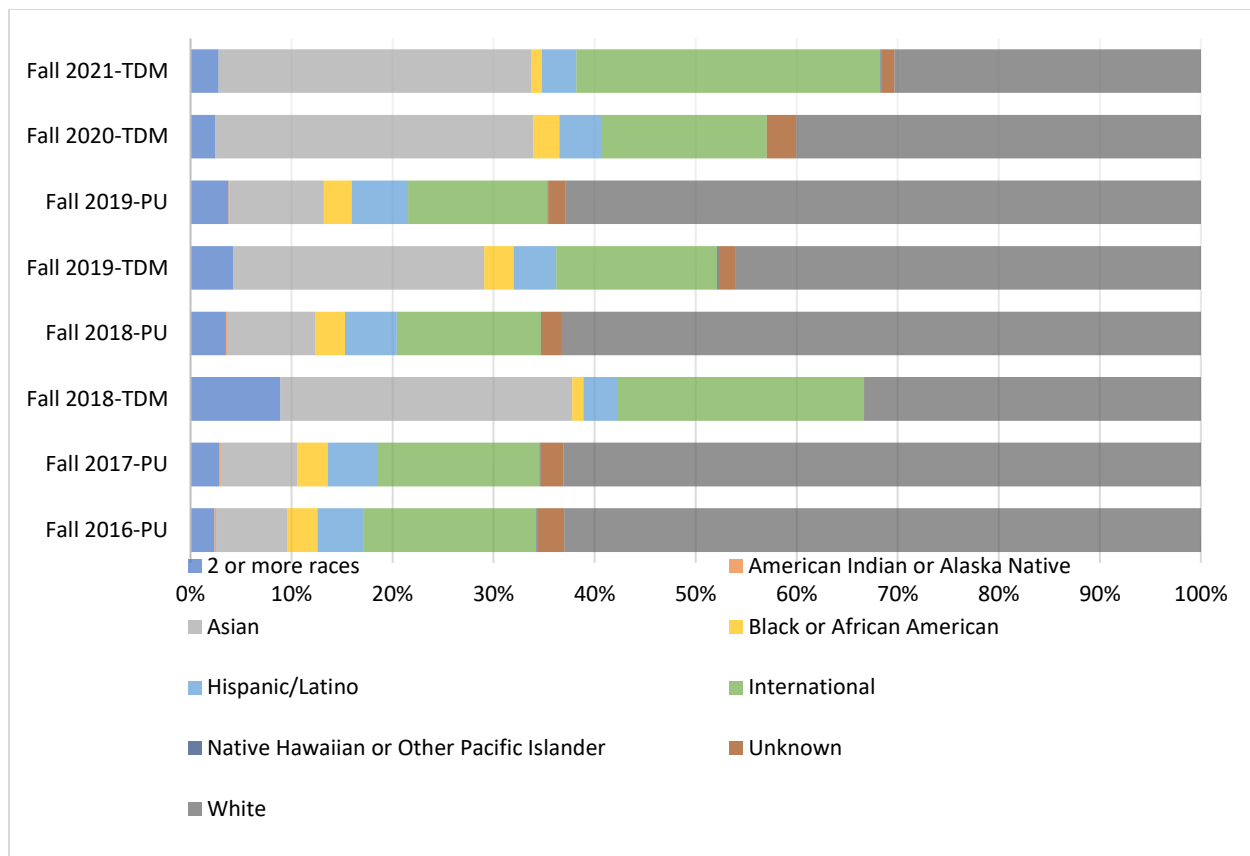
Figure 4: Comparison of the race/ethnicity/nationality of students in The Data Mine with that of Purdue University

4.1   Key Elements of The Data Mine

4.1.1   Data Mine Seminar Course

The Data Mine Seminar is a required course for all first semester students. The seminar course is 16 weeks long per semester, and is offered in the fall and spring semesters. Students meet weekly once for 50 minutes. It is a one-credit hour course that helps the students to develop programming and data science skills. Students learn to code in high level languages such as R, Python, SQL and Bash shell. Since the course follows an active project-based learning pedagogy, the weekly deliverable of the course is referred to as a 'project.' Students are given one week to work on their seminar project. The weekly projects allow the students to explore real world data sets such as Amazon music reviews, New York City taxicab rides, movie review data from Rotten Tomatoes, among others (Gundlach & Ward, 2021). The projects are designed in such a fashion that students incrementally develop the data science skills. The student quoted below demonstrates

that students found the seminar an enjoyable activity, that helped them to collaborate and learn at the same time.

> I definitely enjoyed the weekly Data Mine seminars; they were pretty chill but allowed me to pick up some new skills. I found the Data Mine environment to be cool; I was able to meet a lot of new people with similar interests.

The seminar course has three levels the STAT 19000 is the first level that has no pre-requisite, first year students from all majors are required to enroll in STAT 19000. The second level is STAT 29000 is specifically for the second-year students of The Data Mine. Also, students who are enrolled in STAT 19000 and believe that they have substantial prior knowledge of the data science and computer science can drop STAT 190000 within the first week of classes and enroll in STAT 29000. The third level is STAT 39000 is for the third-year students in The Data Mine. Enrolling in STAT 29000 and STAT 39000 is optional for all the second- and third-year students at The Data Mine. All these seminar classes are taught once a week every Monday. There are four time slots for the seminar class. Table 2 below represents timing and venue for classes. Starting in fall 2022, the course numbers will be offered under a new, permanent "TDM" course listing (which stands for "The Data Mine"), as opposed to the current, temporary "STAT" course numbers.

Table 2: Timing and Venue for the Seminar Course for Spring 2022

Timing	Venue
Monday- 8:30 AM	Hillenbrand Dining Court Atrium
Monday- 9:30 AM	Hillenbrand Dining Court Atrium
Monday- 10:30 AM	Hillenbrand Dining Court Atrium
Monday- 4:30 PM	Synchronous online on WebEx

The seminar course follows the flipped classroom format, as all the information about the project are available online to students on Thursday of every week. The class time is dedicated for problem solving. Faculty and TAs actively help students and answer any questions during the seminar. Students work on 14 projects during the whole semester, but only 10 best projects are counted towards the final grade. Each project is worth 10 points, and late work is not accepted in The Data Mine. Therefore, students are encouraged to work on their projects throughout the week and ask questions to TAs or faculty. TAs hold office hours throughout day and even during the

evening. Students are free to reach out to the TAs and get help. Also, the TAs are available on Piazza to answer any student queries. For the year 2021 Piazza had more than 11,000 posts, as many students use Piazza to ask questions. The TAs, faculty or any other student can respond to the queries on Piazza. A student quote below describes Piazza as a great platform to ask questions and the quick response that one receives after posting the query on Piazza.

> Piazza is a great platform to ask questions…. You also get you tend to get very quick responses. So, the excellent thing about it is that it's not just one TA looking at any particular thing, you're gonna have multiple people. So, whoever can answer the best or the fastest will always get back to you. So, if a student finds something that's particularly challenging in a project, what they will do is ask on Piazza.

4.1.2 Outside Events

The outside events allow the students to participate in the talks by industry and academic experts in the field of data science. These talks allow students to network with experts. The interaction with experts has helped students to find mentors and internship opportunities in the area of data science (Gundlach & Ward, 2021). The student quote below demonstrates that the outside events are motivating as well as open opportunities for the students

> The outside event I recently attended was on Monday. [Speaker] spoke a lot about possible opportunities for us data science students to get involved in. [Speaker] mentioned both spring and summer possibilities. This event was interesting because I got to hear about the way I could [use] my skills right away. Most of the other speakers inspire me for my future but to hear about different opportunities available to me right now is both exciting and invigorating.

Every student in The Data Mine is required to attend at least three outside events and submits reflection about each event. The three reflections contribute to 12% of the total grade. Each student reflection is minimum 1 page in length, double spaced and must have font-size 12. The reflection focuses on the following aspects: 1) what are the key take away of the event, 2) what did you learn from the event, 3) what new knowledge did you gain from the event, 4) what question do you have for the presenter.

Every student reflection is read by the faculty, staff, and/or TAs at The Data Mine. The student reflection about the outside events helps the faculty to:

1) understand the student experience and knowledge that they gained through participation. For instance, the quote below demonstrates that the participating in outside activity helped the student to identify the steps for conducting research. This demonstrates that students are interested in events that helps them to develop research skills.

> Throughout her talk, I was able to grab onto her advice about how to conduct your own research. She had numerous steps and potential action you must take while conducting research and how to act within this setting. The steps are listed below:1. Research Question, 2. Formulate questions to fill gaps, 3. Gather data from mostly the main sources of data, 4. Perform the analysis, 5. Draft report, create a graph, 6. Take notes about consumptions or constraints, 7. Iterate, Work with stakeholders, keep the people you're doing research for in good contact about what you're doing. With these steps, I will be able to apply them in my own experiences and practice on my personal relations with my research colleagues.

2) The reflections allow The Data Mine faculty to identify the student interest and select future speakers. For instance, the quote below is from student reflection and it describes the interest in the domain of coding, statistics, mathematics, and storytelling. This input from the students will help the faculty and staff to identify events that focus on communication and storytelling of data.

> I really enjoyed this talk for a number of reasons, but the biggest is that it reminds us we need skills outside of pure data science in order to become good data scientists. We need communication and storytelling skills in addition to the coding, statistics, and mathematics if we want to truly maximize our potential success as data scientists.

> Overall, the outside event is a wonderful opportunity to learn from the experts in the field and network with the professionals.

4.2   Structure of The Data Mine

The Data Mine can be broadly categorized into four units. 1) The General cohort, 2) Corporate Partners, 3) Research with Professor, 4) Academic Learning Community; see Figure 5.

```
                    ┌──────────────┐
                    │ The Data Mine │
                    └──────────────┘
        ┌───────────────┬──────────┴─────────┬────────────────┐
┌──────────────┐ ┌──────────────┐ ┌───────────────┐ ┌──────────────┐
│General Cohort│ │  Corporate   │ │ Research with │ │   Academic   │
│              │ │   Partners   │ │  Professor    │ │   Learning   │
│              │ │              │ │               │ │  Community   │
└──────────────┘ └──────────────┘ └───────────────┘ └──────────────┘
```
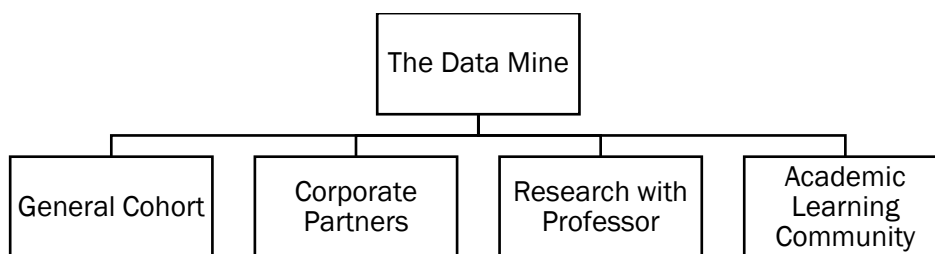
Figure 5: The Structure of The Data Mine

4.2.1   The General Cohort

The general cohort comprises of students who just register for the 1 credit hour The Data Mine seminar course and participate in at least 3 outside events in the fall and spring semester.

4.2.2   Corporate Partner

The Corporate Partner (CRP) cohort was started in the Fall of 2019. Students register in the corporate partner cohort in the Fall of each year and continue working with the corporate partner company till the Spring of following year. CRP is a 9-month commitment. Students enrolled in the CRP participate in the 1 credit hour seminar course and register for a 3-credit hour CRP course. Students in this cohort work on real world company project under the guidance of the corporate mentors. These projects are similar to capstone projects, but allow students to have these experiences earlier in the studies.

4.2.2.1   Corporate Partner Project Team

The students choose their team based on their interest in the project. The project teams are interdisciplinary and include students from all levels of education (freshman year to senior year). They have students from all backgrounds (liberal arts, management to computer science and data science). Each CRP team has a maximum 25 students. The CRP TAs and CRP mentors group those 25 students into a smaller team of 5-6 students. They basically break down the larger project objective into sub-objectives and each team works together to achieve the bigger goal. Each team is led by a CRP TA. There were 45 companies that were the part of The Corporate Partner cohort in the 2021-2022, but with the growing popularity of the Corporate Partner cohort we expect to have 70+ companies in the 2022-2023. The projects involve a wide range of methodologies from

machine learning to survey design. The projects cover varied topics. Some of the examples are: drug discovery, UX based projects, anomaly detection, data modeling market analysis etc. Students present their findings in a research symposium at the end of the spring semester. The student quote below describes the skills that the student gained while working on the CRP project.

> It [Corporate Project] was so broad and open-ended, which helped me problem solve and look at many different routes and aspects. It had a lot of pieces to it, like data retrieval, domain knowledge and research, statistical analysis, and machine learning, which I believe mimicked a lot of data science projects in the industry. It was certainly unlike any other project I have completed thus far at Purdue, even in my data science classes.

4.2.3 Key members of the Corporate Partnership
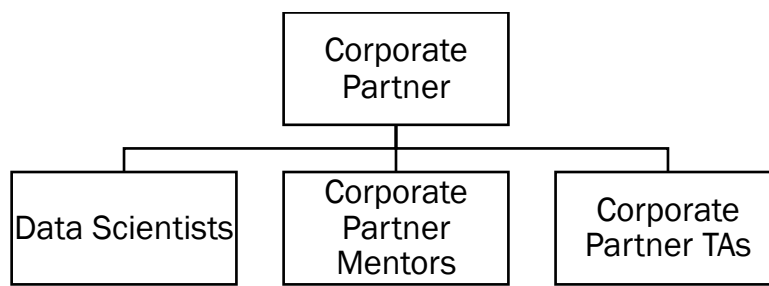
Figure 6: Key members of Corporate Partnership

The Figure 6 represents the three key members of the Corporate Partnerships.

1) **Data Scientists:** Data Scientists are the part of The Data Mine, and each CRP team is assigned a data scientist. The role of the data scientist is to serve as a subject-matter expert for the students. Therefore, when student teams work on CRP Projects, and they need help in troubleshooting any technical issues, they can set up a meeting with the data scientists. The data scientists are experts in dealing with problems related to machine learning, natural language processing, or data modeling. For complex problems, it is expected of the student teams to email the question to the data scientist prior to the meeting. Data scientists help the student to master new the data science concepts. Prior to the meeting with the data scientist, students are also expected to try the problems and then reach out to the data scientists with concrete questions.

2) **Corporate Partner Mentors:** Corporate Partner Mentors are the employees from the company who introduce the projects to the student teams. Their role is to provide the resources required by the students to achieve the project objectives. Student team and CRP Mentors meet online one time per week. The students are expected to report the progress of the project to the corporate partner mentors.

3) **Corporate Partner TAs:** Corporate Partner TAs (CRP TA) are the students who have been part of the corporate partners at least once. The CRP TAs are experienced individuals and serve as a liaison between the student teams and corporate partners. The CRP TAs are trained on the agile methodology, and they manage the team using the agile-scrum framework. The CRP TAs meet the team every week for 2 hours known as the lab hour, the team works on the project during those 2 hours. Also, in case team members have any questions, they can reach out to the CRP TA.

4.2.4   Agile Training for CRP TA

All CRP TAs undergo agile training. The agile is project management technique that was originally practiced in the software development industries. The agile framework comprises of a sub-framework known as scrum. Scrum is an effective approach of teamwork and is being practiced in higher education settings. The foundation of a scrum team are five scrum values, they are courage, focus, commitment, respect, and openness. Every team member abides by these five scrum values. Scrum has three key components sprints, roles and ceremonies. The *sprints and roles* will be discussed under sprints and roles heading and the *ceremonies* will be discussed under the monitoring progress heading.

4.2.4.1   Sprints and roles

In the Data Mine, product owners are the CRP mentors as they describe the product requirement to the development team. They also provide resources and direction to the team and communicate customer expectation. CRP mentors acts as a liaison between the stakeholders and customers. The development team is a team of students who work on the corporate partner project and deliver the required product. The scrum master is the CRP TA; they act as a liaison between the product owners (Corporate Partner mentors) and the development team (TDM Student team).

Their responsibility is to set clear and achievable goals, manage the project and keep the product owners updated on the progress.

Each CRP Project is a year-long project divided into two semesters. Each semester consists of 16 weeks, divided into 7 sprints. Each sprint lasts for approximately 2 weeks. The CRP TAs meet with the CRP mentors and decide on the product backlog. The product backlog is the list of tasks to be completed for the entire project. Based on the product backlog, the CRP TAs also work with their team members and create a sprint backlog. The sprint backlog helps the team to come up with the list of tasks that they plan to achieve for a two-week long sprint. Based on the sprint backlog, the scrum master and the development team decide on the tasks to be prioritized for that sprint. The also decide on the role allocation based on the skills and abilities of the team members. Finally assign the product owners their deliverables for that particular sprint.

4.2.4.2 Setting expectations and monitoring progress

The two crucial steps in in successfully completing the corporate partner projects are: setting expectations and monitoring the progress. Figure 7 below represents the elements of the CRP project.

1. **Sprint Planning** is the first step towards the CRP project. The sprint planning is conducted prior to starting of every sprint. The CRP mentors, scrum master and the development team meet before starting the sprint and decide on the deliverables. The development team members are assigned the ownership of tasks for the current sprint, and the scrum master monitors the progress on the task and keeps the product owners updated on the progress.
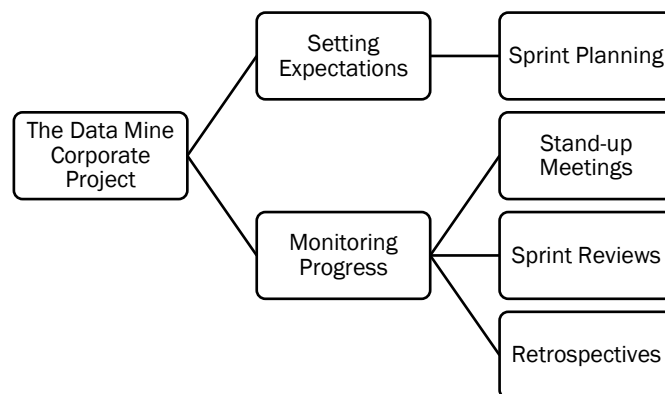
Figure 7: Elements of The Data Mine Corporate Partner Project

2. **Monitoring the progress** comprises of three *scrum ceremonies*, the stand-up meetings, sprint reviews and retrospectives.

- *Daily stand-up meeting:* In an industry setting the stand-up meetings are referred to as *daily stand-up meeting*, which happens for 15 minutes at the start of the day and end of the day. The objective of these stand-up meetings is to allow each team member of development to share their progress. Each member is required to answer the following three questions:

  - What did I do yesterday?
  - What will I do today?
  - Do I have any blockers?

  In the case of The Data Mine, student teams cannot have daily stand-up meetings, due to their academic schedule conflicts. Therefore, they have weekly standup meet ups for 15 minutes, where each team members responds to the three questions and updates their progress. The role of the scrum mater is to make a note of the blockers and help the team members to make smooth progress.

- *Sprint Reviews* are a second type of ceremony. In the Data Mine, the development team, scrum master and product owner meet once in every sprint. The development team demonstrates the progress they made on the project; that is, every member of the team provides a demo or showcases their models or analysis to the product owners and seeks their feedback. The scrum master annotates the feedback of the product owner and prepares a dynamic document and shares it with the development team and product owners.

  A CRP TA mentions her role as a Scrum master in the sprint reviews:

  So, during the weekly meetings with our Mentor, I take notes on them. So, I try to like, I have a big document going of notes for each meeting this semester, in case we need students miss it and he's students are late to it, or in case they want to go back and see what we talked about during that meeting. It also helps to like the end of the year when we're making our documentation guide for the next team up.

- *Retrospectives* in the Data Mine occurs once in each sprint, it allows the team members to reflect on the progress. The presence of the corporate mentors is not mandatory during the retrospective session; the scrum master handles this ceremony.

4.2.5 Tentative bi-weekly scrum schedule in The Data Mine

Since each sprint comprises of two weeks. Table 3 below represents the breakdown of the task for each week for a two-week sprint. The role of Corporate Partner mentor is crucial for sprint planning and sprint review only (Amstutz et al., 2021).

Table 3: Description of the Weekly Sprint in The Data Mine

Sprint	Day of the Week	Task
Week 1	Monday or Tuesday	Sprint Planning in presence of Corporate Partner
	Thursday or Friday	Weekly Stand ups and Working meeting
Week 2	Monday or Tuesday	Sprint Review Meeting in presence of Corporate Partner
	Thursday or Friday	Weekly Stand ups, retrospective and Working meeting

4.2.6 Weekly schedule of a Corporate Partner TA

The CRP TA plays a crucial role in planning, monitoring, and executing the corporate projects. Ever week they need to conduct a series of activities as they are represented in the Figure 8.

- *Team meetings:* Team Meetings are scheduled by the scrum master every week for 50 minutes. In these meetings development team meet the product owners. The important scrum ceremonies, sprint planning and sprint reviews were conducted during these meetings.
- *Team labs:* Team Labs are weekly sessions conducted by the scrum master. The lab sessions are 1 hour 50 minutes long, during the lab session development team and the scrum master meetings. The role of the TAs is to help the team members to feel comfortable. The TAs use some ice-breaker question to get the team talking and interacting with one another. TAs also conduct some team warm up activities such as Purdue Trivia, activities using Kahoot, etc.
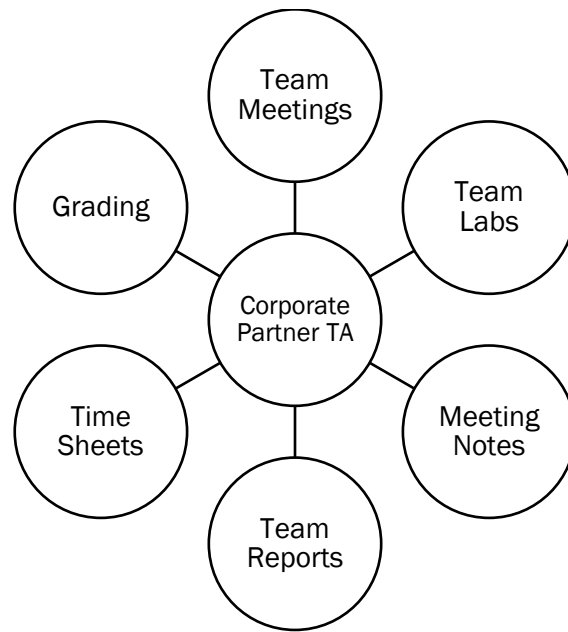
Figure 8: Roles and responsibility of the Corporate Partner TA

After the team members are comfortable, the TAs also help the team members with coding related questions or any other questions related to the project. For very technical questions the TA or team can reach out to The Data Mine staff. It is advisable to email the staff with the questions so that they can get the required answer during their session. The TAs also conduct the sprint retrospective session during the lab, where students mention about the aspects that are going well, aspects that need some attention, and also to highlight the roadblocks.

- *Creating and distributing meeting notes:* The TAs are also responsible for creating meeting notes after ever team meeting with the Corporate Partner mentor and after every lab meeting with the team. The TAs create the notes and distribute them among the team members using an online medium such as Slack, Google Docs, etc., within a day. The objective of the meeting notes is to remind the team about the required tasks to be completed, also it helps to keep a track of the progress of the work. The format for the meeting notes, as per Examples Book (2021) is given in Figure 9.

1. Introduction
    o Welcome conversation
    o Ice breaker
2. Sprint Planning
    o Data Engineering Team
        ▪ Executable 1
        ▪ Executable 2
        ▪ Executable 3
    o Data Visualization Team
        ▪ Executable 1
        ▪ Executable 2
        ▪ Executable 3
    o Data Science Team
        ▪ Executable 1
        ▪ Executable 2
        ▪ Executable 3
3. Question and Answer
4. Free Time to work
5. To Do for next meeting
    o Task 1
    o Task 2

Figure 9: Example of Meeting notes

- *Creating Team Reports***:** The TAs are responsible of submitting the team reports. The team reports are due for submission every two weeks. In the team report the TAs are expected to report on the progress of their team, based on some pre-defined questions. TAs also highlight the assistance that their teams need with respect to the corporate project. The report is read by The Data Mine staff and they provide feedback and help required by the team. The questions below represent the sample questions from The Example Book (Amstutz et al., 2021) that TAs answer as a part of team report:

1. Did you have both of your team meetings this week? If not, please explain. If yes, a simple "yes" will suffice.
2. What did your team do well this week, and why do you think it went well?
3. What did your team struggle with this week, and what would help that go better next week?
4. Who are your team's superstars (and why)?

5. Who struggled this week, and how do you plan to reach out to them?

- *Time Sheet:* TAs are required to keep the track of their responsibilities and update the time sheet regularly. The time needs to be completed and submitted along with the team report before 8:00 am on Monday. Additionally, TAs are also required to submit their number of hours worked on the Success Factor portal of Purdue University.
- *Grading:* The TAs are also expected to grade the sprint report of the team members. Rubrics for the grading are available on the Gradescope, TAs grade the sprint reports and post the grades on the Gradescope before Sunday night 11:59 pm.

4.2.7 Other Responsibilities of the CRP TAs
- *Team building Activities***:** Corporate Partner TAs are also responsible for engaging the students in the team building activities. The intent of the activities is to provide team members opportunity to socialize outside the academic setting. Some of the examples of the team building activities are: team dinners, online games, trivia nights, board game nights etc.
- *Creating Sub Teams:* The TAs are responsible to create sub-teams, to allow students of similar interest to work together. Studies have also demonstrated that managing small teams is easy. Small teams allow the members to interact with one another and are effective in achieving the desired objectives. The TA uses a questionnaire to divide the students of similar interest into same groups. Also, the TA is responsible to make sure that members in the sub-teams are comfortable working with each other. The quote from a CRP TA demonstrates how TAs act like a support system to students and execute the activities.

I mean I'm there as a support system. I tell them almost every meeting like uh. I'm here to make their lives and their experience in The Data Mine as fun as easy, as educational as possible, and so that that kind of goes from just being a person of support, using correct language and correct body language when addressing the team. That also means like accommodating students, and in some of their needs. So, you know, one thing that that my team does specifically is before the project even starts, I send out a survey and get to know their experiences and expectations. My main goal as TA is to make sure that everybody is learning what they want to learn. It would be an unfortunate situation where if someone had gone an entire semester doing parts of the project, that they weren't necessarily interested in, and so my goal is to make sure that educational experiences as is as tailored as possible

for everyone and then on top of that just kind of making sure everyone feels comfortable like I want the team to get to know each other we do. We do like fun little social events and gatherings and things, and Uhm, I think these are all great things at the team does that I know. Yeah, we were able to kind of bond together for sure.

Overall Corporate Partners cohort is a real-world experience for the students and The Data Mine makes the students equip to balance their academic and corporate project responsibilities.

4.3   Research with Professor

Research with a Professor is third unit of The Data Mine, which allows students interested in research to work on interdisciplinary data science projects. Currently The Data Mine has four research with professor opportunities: Computational Investigation of Living Systems, Human Development and Family Studies, and Public Health, Physics and Vertically Integrated Projects. All the four research projects have their own TAs who work with students and answer their questions. Students also get opportunity to work in groups on their research projects. A student from Research with Professor described his experience working with Physics research project:

> I have a couple friends who were in The Data Mine. Uh, and you know, I talked with them a lot. During the physics cohort, you know my group me talked almost every other day basically. And that TA we would talk with every week. Like I said, it's one of the best experiences I've had at Purdue. And so, I got to learn all kinds of things from these people and get to talk with people. Uh, and you're just getting to talk with people in those different fields. Those are really, really good experience.

Students in this unit work with the respective Professor and participate in the mandatory 1 credit hour seminar course and attend three outside events.

4.4   Academic learning communities

The fourth unit of The Data Mine comprises of 14 academic cohorts, they are: Actuarial Science, Agriculture, Analyzing Digital Gaming and Culture, Biology, Data in Health and Human Sciences, Data Visualization, Earth, Atmospheric, and Planetary Sciences, Human Development and Family Studies, and Public Health, Krannert, Nursing, Pharmacy, Psychological Sciences,

Regenstrief Center for Healthcare Engineering, and Statistics. There is at least one cohort from most colleges at Purdue University. Students from any college can join an academic cohort and take their courses. Faculties hold office hours in the Hillenbrand Hall. Some faculty members are Faculty Fellows who work with Resident Assistants to plan social events for the students. Students in these cohorts also get the opportunity to work on group projects and they learn interdisciplinary applications of data science.

4.5   Summary

In summary, The Data Mine is a unique living learning community on the Purdue University campus that allows students not just to work in a collaborative environment but also promotes socio-constructivist learning. Students are engaged in multiple events and activities, collaborate with peers and experts that help them to develop skills, create artifacts, experiences, stories, contribute to the body of knowledge, learnt from one another, and finally, develop their identities.

CHAPTER 5: METHODS

The intent of this study is to understand and identify the various opportunities that students get within The Data Mine to socialize, develop identity, and learn data science skills that helped them grow into mature learners. Therefore, I planned to answer the following research questions: **Quantitative: RQ 1:** What are the perceptions of students regarding their identity formation, socialization opportunities, self-belief, and academic/intellectual development in the Data Mine?

Qualitative: Guiding RQ 2: How do students' participation in activities and interaction with peers, faculty, staff at The Data Mine contribute to becoming an experienced member of the learning community?

- **Sub-RQ 2(a):** What are the perceived benefits and challenges of participating in The Data Mine?
- **Sub-RQ 2(b):** How do students describe their levels of socialization and a sense of belonging within The Data Mine?
- **Sub-RQ 2(c):** How do students' participation and interaction in The Data Mine help them form their identity?

5.1 Research Design

This study uses a mixed method research design. Mixed method design has been considered as an excellent approach to combine the qualitative and quantitative findings to answer research questions. Mixed method design is a holistic approach that collects, analyzes and interprets the findings of the study by combining both qualitative and quantitative approaches (Li et al., 2015). Combining the two method provides opportunity to the researcher to explore the research question in depth. Specifically for this study I used the sequential explanatory mixed method research design (Ivankova et al., 2006). The study started with an initial quantitative survey study followed by an in-depth qualitative study. The results from the quantitative study informed the design of the qualitative data study. The sequential explanatory mixed method design prioritizes the qualitative phase of the research design, and since the focus of this study is on the qualitative data collection and analysis, I found this research design appropriate for this study.

Prior studies by Jaiswal et al. (2021) have used a sequential explanatory mixed method approach to analyze student reflections. The objective was to understand the reflection scores assigned to each cluster of students. Conducting the qualitative inquiry for the student's reflection helped to identify the emergent themes.

Table 4 has been adapted from Ivankova et al. (2006) below represents the phases, sub-phases, procedure, and product that would emerge at every step of the study.

Table 4: Phases, procedures, and products of the sequential explanatory mixed-method design

Phase	Sub-Phase	Procedure	Product
PHASE 1 Quantitative Phase	Quantitative Data collection (RQ1)	A five-point Likert scale survey was used to collect data for students' perceived identity, sense of belonging, socialization, self-belief, academic/intellectual development	Pretest and posttest numeric data
	Quantitative data analysis	Data was analyzed using descriptive and inferential statistics	Measures of central tendency and spread of the pretest and posttest data (mean and standard deviation) and statistics for significant differences p-values and test statistic) and visualizations.
Intermediate Phase	Connecting Qualitative and Quantitative phases	Survey results were used to create the interview protocol and purposefully select participants from the population for conducting semi-structured interviews	An interview protocol and participant selection for the interview
PHASE 2: Qualitative Phase	Qualitative Data Collection Method (RQ2, RQ2(a), RQ2(b), RQ2(c))	Semi-structured interviews with students	Text Data (interview transcripts)
	Qualitative data analysis	i) To answer RQ2(a), RQ2(b), RQ2(c) thematic analysis was conducted to find emerging patterns in students' experiences	Codes, categories organized into themes, along with representative quotes describing students' experiences.
		ii) To answer RQ 2: Clustering the data and Narrative Inquiry	Student were clustered and representative students from each cluster were identified to write the narrative.
	Integration of the Quantitative and Qualitative results	Results from Quantitative analysis and Qualitative Inquiry was used to write the discussion section.	Discussion explaining the findings aligned with the theoretical framework, recommendations and implications for practice, and opportunities for future research

5.2   PHASE ONE: The Quantitative Study

An initial quantitative study was conducted to evaluate the perception of The Data Mine students regarding identity formation, socialization, sense of belonging, self-belief, and academic and intellectual development. To answer the following research question:

RQ 1: What are the perceptions of students regarding their identity, socialization opportunities, self-belief, and academic/intellectual development in the Data Mine?

The survey study was conducted in the beginning (September) and at the end (December) of Fall of 2020.

5.2.1   Context and Participants

Participation in the survey study was voluntary. Permission from IRB (IRB protocol number 2020-136) was obtained before conducting the study. Students were compensated $15 for participating in the study. The pre-post surveys were conducted in September and December of 2020. In the September of 2020, a total of 50 students from The Data Mine participated in the study. Furthermore, these students were followed up for a post-survey in December of 2020, in that 25 students participated in the study.

5.2.2   Procedure and Data Collection Method

The data was collected using the survey instrument that contained questions related to identity, socialization, sense of belonging, self-belief, and academic and intellectual development. The survey instrument used in the study was designed based on validated questionnaires. The reliability coefficient for each instrument used is mentioned in the subsequent paragraph.

The identity questions were derived from the questionnaire developed by Godwin (2016), as the instrument was validated and also demonstrated a high reliability coefficient (Cronbach alpha greater than 0.80). The questions were adapted for the current study to understand the perception of The Data Mine undergraduate students regarding identity formation. Godwin (2016), in her study, described three sub-constructs of identity formation: recognition, interest and performance/competence. For this study, recognition refers to what extent an individual perceives

oneself or how others (peers, parents, faculty) perceive the individual in the context of data science knowledge, skills, and abilities. Interest refers to the extent to which an individual enjoys learning the data science concepts and performance/competence is described as to what extent the individual is confident in applying the data science techniques and tools (Godwin, 2016). Since The Data Mine is a learning community, the survey also had questions for socialization and sense of belonging. The questions for sense of belonging were adopted from Chin et al. (1999), and socialization questions were adopted from Weidman and Stein (2003). The reliability coefficient for Chin et al. (1999) sense of belonging construct was Cronbach alpha = 0.95 and for the instrument by Weidman and Stein (2003) the Cronbach alpha for student to faculty (Faculty/staff/mentor) interaction was 0.64 and student to student (peer) interaction was 0.81. The purpose of the socialization construct was to understand the perception of socialization with The Data Mine student to student and student to faculty. The sense of belonging constructs consisted of three questions. The purpose of those questions was to understand to what extent students feel accepted in The Data Mine. The questions related to self-belief in data science perception were adopted from Scott and Ghinea (2014) to assess the data science self-concept, data science anxiety, and data science mindset of the students. All three constructs in Scott and Ghinea (2014) reported their composite reliability score greater than 0.70. The questions for academic and intellectual development were adopted from institutional integration scale (Pascarella & Terenzini, 1980). The Cronbach alpha score for academic and intellectual development scale was 0.74. Table 5 represents the sample questions from the survey. The complete survey is presented in Appendix B.

Table 5: Sample Survey Questions

Construct	Questions
Recognition	I see myself as a data science person
Interest	I enjoy learning data science concepts
Performance and Competence	I am confident that I can understand the data science concepts in class
Performance and Competence	Others ask me for help with data science concepts or assignments
Socialization Staff/Faculty/Mentor	Is there any professor or staff member in The Data Mine or the Data Science Certificate with whom you: - Sometimes engage in social conversation
Socialization Peer	Is there any student in The Data Mine or the Data Science Certificate with whom you: - Often discuss other topics of intellectual interest
Sense of Belonging	I feel a sense of belonging to my data science community (The Data Mine or the Data Science Certificate)
Data Science Self-Concept	I learn data science skills quickly
Data Science Fixed Mindset	I have a fixed level of data science aptitude, and not much can be done to change it
Data Science Anxiety	I often worry that it will be difficult for me to complete data science assignments
Academic/Intellectual Development	I am satisfied with the extent of my intellectual development as student in The Data Mine or the Data Science Certificate

5.2.3 Data Analysis Method

The quantitative data were analyzed using descriptive and inferential statistics. The descriptive statistics was used to describe the measures of central tendency for each construct. The descriptive statistics were calculated for all the students who participated in the survey study. There were 50 students in September and 25 students in December. Further *t*-test and McNemar test was used to test if there was any significant difference between the pre and post-test groups. It is important to note that inferential statistics was computed only for the sample of 25 students who participated in both pre-test and post-test surveys. The t-test was used to test the significance of all the 5-point Likert scale responses, whereas the McNemar test was used to test the significance in case of Yes/No questions.

The findings of the quantitative analysis were grouped into two categories: (1) Descriptive Statistics for the 50 Pretest students and 25 Posttest students, and (2) Descriptive and inferential

statistics results for 25 students participated in both pre and posttest. The results for each category are represented in the Chapter 6.

5.3   Phase Two: A Qualitative Inquiry

5.3.1   Rationale and Significance of conducting the Qualitative Inquiry

This qualitative study intends to understand in detail the factors and identify the opportunities that led to increase in socializing, sense of belonging, self-concept, and identities as students progressed through the semester in The Data Mine. The qualitative study will allow us to conduct an in-depth analysis of student experiences in The Data Mine. The qualitative research method is natural and interpretive as it is used in the studies that are conducted in natural settings (Denzin & Lincoln, 2011) and emphasizes the experiences of the participants within the research context (Banks, 2018). The qualitative approach helps the researcher to explain the *how* of the phenomenon by accepting the fact that there are multiple ways to explain the same phenomenon (Jones, 1995). Moreover, it allows the researchers to conduct in-depth analysis and discover the implicit meanings that explain the phenomenon (K. Kelly et al., 2006). Merriam (2002) argued that no setting in this world is constant, and the social interaction among the individuals with the context creates new meanings and provides multiple perspectives. This leads to the fact that it is hard to measure the impact of any phenomenon that constantly changes with time and people; therefore, conducting the qualitative study by in such cases is the most appropriate methodology (Merriam, 2002).

Qualitative research must be focused on a purposefully selected sample that provides a detailed description of the phenomenon (Banks, 2018; Creswell & Creswell, 2017). The detailed description of the phenomenon allows the researcher to conduct an in-depth analysis of the experiences and also track the changes that occurred over time (Merriam, 2002). The flexible nature of qualitative research allows the researcher to collect data using various sources, such as observation, interviews, reviewing documents or artifacts, to identify the true meaning and gain insights into the phenomenon of interest (Creswell & Creswell, 2017; Merriam, 2002). Since the researcher is the instrument for the qualitative research (Creswell & Creswell, 2017), meaning researcher is responsible for both qualitative data collection and analysis therefore it is crucial for the researcher to acknowledge those biases and implement the checks such as the member checking

and peer debriefing. The objective of this study is to understand how students in The Data Mine grow from being a newcomer to an experienced member of the learning community, and what opportunities The Data Mine provides them to grow into an experienced member. Therefore, I plan to use qualitative methods to answer the following research questions:

Guiding RQ 2: How do students' participation in activities and interaction with peers, faculty, staff at The Data Mine contribute to becoming an experienced member of the learning community?

- **Sub-RQ 2 (a):** What are the perceived benefits and challenges of participating in the Data Mine?
- **Sub-RQ 2(b):** How do students describe their levels of socialization and a sense of belonging within The Data Mine?
- **Sub-RQ 2 (c)** How do students' participation and interaction in The Data Mine help them form their identity?

5.3.2 Epistemological Approach, Researcher's Role, and Positionality

The epistemological approach for this study is social constructivism. The social constructivist approach acknowledges the impact of the social context on learning; therefore, in the context of The Data Mine, we believe that learning occurs due to the social interaction among the members of the learning community. Social constructivism is an appropriate epistemological approach for this study, as my objective is to understand the perception of students enrolled in a living-learning community. Social constructivism will also serve as a lens for conducting the qualitative analysis. The objective of the study confirms with the underlying assumption of social constructivism that individuals have a complete understanding of the world in which they reside (Creswell & Creswell, 2017). Therefore, the researcher relies on the participants' opinions about the phenomenon under study. The researcher's role, in this case, is to ask open-ended questions and interpret the meanings as described by the participants (Creswell & Creswell, 2017).

5.3.3 Researcher's Positionality

In the case of qualitative research, it is crucial to clarify the researcher's positionality. The positionality of research refers to the extent to which the researcher is involved with the topic and the relationship between the researcher and participants (Ortega-Alvarez, 2019). Since clarifying

the positionality is crucial, I would like to clarify my involvement with the topic. As a researcher, I am not directly involved with the topic, but I always was inclined towards the field of data science. I completed my MBA in 2013 and joined the Information Technology (IT) industry. In my corporate tenure of five years, I realized that data-driven decision-making is crucial for making sound and informed decisions. During my tenure as Senior Manager Operation Excellence, I observed that fresh undergraduate students found it difficult to make data-driven decisions. The knowledge of statistics and computing was shallow. This made me think numerous times, how can higher education institutions spark interest among undergraduate students to develop data-driven competencies. After considering multiple options, I decided to pursue a Ph.D. and focus my interest on data science education.

Also, to clarify, I do not have any direct involvement with the participants, as the participation in the study was voluntary. Interested students signed up for the study using a link, and students who met my purposeful sampling criteria were contacted for the interview study. But as a researcher my aim was to help students to provide a detailed account of their perceptions and experiences, therefore before starting the formal interview process, I started with a casual conversation that helped to establish a rapport and make them comfortable. I also informed them of their rights regarding answering the question or withdrawing from the interview process at any point in time. I also explained to them how their responses will help in improving the instructional strategies and future student experiences at The Data Mine. For my part, I made sure that I acknowledge the biases while narrating the story that my participants told me during the interview.

5.3.4 Participants and Procedure

The participants for this study are the students enrolled in The Data Mine. The participants for this study were purposefully selected based on the following criteria that the students must have completed at least one year in The Data Mine. The Data Mine staff were contacted to help me with recruitment of the students for the interview study. Staff helped us identify the students based on our criteria. The email containing the details of the study and a link to a short survey with sign up link was sent to all the eligible students. The participation in the study was voluntary. Student responses to the sign-up survey were reviewed, and those who met our participant selection criteria were selected as the participants for the study. I interviewed 19 students; the average length of each interview was 1 hour 10 minutes. Each participant was compensated with

a $15 Amazon gift card for every 30 minutes of participation. Before starting the interview process, participants were asked to provide their consent and respond to a short demographic survey. I also let the students know their rights, for instance, they can choose not to answer a question during the interview process or withdraw from the interview at any point in time and, they can ask me any questions during the interview. Table 6 below demonstrates the demographics of the students for the qualitative study.

Table 6: Demographics of Qualitative Study Participants

Participant	Gender	Number of years in The Data Mine	Cohort	STEM or Non-STEM Major	TA or Not TA
Finley	Male	1.5 Years	General Cohort, Corporate Partners	STEM	Seminar TA
Arrow	Male	3.5 Years	Corporate Partners, General Cohort	STEM	Seminar TA
River	Male	3 Years	Research with Professor, Data Visualization, Statistics	STEM	Not a TA
Reef	Male	3 Years	Corporate Partners	STEM	Not a TA
Tommie	Female	2 Years	Corporate Partners	STEM	Corporate Partners TA
Jackie	Male	3.5 Years	Corporate Partners	STEM	Corporate Partners TA
Marion	Male	3.5 Years	Corporate Partners	STEM	Corporate Partners TA
Jamie	Female	3 Years	Corporate Partners	STEM	Corporate Partners TA
Morgan	Female	2 Years	Corporate Partners	STEM	Corporate Partners TA
Mickey	Female	1.5 Years	Corporate Partners	STEM	Not a TA
Kendall	Female	3 Years	Corporate Partners	STEM	Not a TA
Robbie	Female	3 Years	Corporate Partners	STEM	Corporate Partners TA
Quinn	Male	3 Years	Corporate Partners, Actuarial Science	STEM	Corporate Partners TA

Table 6 continued

Orion	Male	3 Years	Corporate Partners	STEM	Corporate Partners TA
Riley	Female	2 Years	Corporate Partners	Non-STEM	Corporate Partners TA
Frankie	Female	2 Years	Corporate Partners	STEM	Corporate Partners TA
Skylar	Female	2 Years	Corporate Partners	Non-STEM	Corporate Partners TA
Rio	Female	2 Years	Corporate Partners	STEM	Corporate Partners TA
Harley	Female	1.5 Years	Corporate Partners	Non-STEM	Not a TA

Table 6 represents the background information for all 19 students who participated in the interview process. 8 male students and 11 female students participated in the interview. All the students had more than 1 year of experience in The Data Mine. Eighteen of the students were in the Corporate Partners program, and 1 student was the part of the Research with a Professor cohort. The majority of students were from the STEM background, and 3 students were from non-STEM backgrounds.

5.3.5 Procedures and Data Collection

In-depth open-ended interviews are considered as an excellent method to deep dive into any topic and search for implicit meanings, and also, they are more appropriate to conduct an interpretive inquiry. For this study, data was collected by conducting semi-structured interviews lasting between 60 to 90 minutes. Semi-structured interviews allow the researcher to ask open-ended one-to-one questions to the participants, to gain an understanding of the phenomenon under study (Mueller, 2019). These interviews serve as an excellent method to understand the experiences of the participants with the phenomenon and report the implicit meanings, as participants get an opportunity to be spontaneous and descriptive (Doody & Noonan, 2013; Figgou & Pavlopoulos, 2015). To conduct a meaningful interview session, it is important that interview questions should be clear and must follow a logical order (Mueller, 2019).

The interview questions for this study are based on a pre-designed interview protocol; see Appendix C. The interview protocol has undergone multiple iterations based on the feedback

received by the experts in the field of qualitative inquiry. The interviews were conducted using Microsoft Teams, and each interview session was recorded for conducting the analysis. Prior to the start of the interview, students provided their consent. The interview began with few casual open-ended questions to establish a rapport with the participant. Gradually I asked questions to understand students' perception with regards to various opportunities that The Data Mine provided for learning, socialization, identity formation, sense of belonging, and development of data science skills. The conversation of the interview focused on two criteria: 1) to understand how students perceive themselves after spending a substantial amount of time and what were the benefits and challenges of participating in The Data Mine, and 2) opportunities provided to students in terms of mastering technical and non-technical skills, socialization, and growing as an individual. I also asked follow-up questions in case the responses of the student were brief. Post conducting the interviews, I transcribed the interviews and analyzed them using qualitative research software NVivo 12.

5.3.6 Data Analysis

For analyzing the results for this study, I used two data analysis methods. The initial data analysis was focused on conducting a thematic analysis, followed by narrative analysis of students interview responses.

5.3.6.1 Thematic Analysis

The thematic analysis helped me to determine the patterns in the data set. I used the steps described by Braun and Clark (2012). I conducted an inductive thematic analysis to identify the patterns that demonstrated commonality or differences among the participants' experiences in the Data Mine. Thematic analysis is a "standalone, theoretically-unbound" qualitative research method used in the area of education research (Braun & Clarke, 2006; Ortega-Alvarez, 2019). Braun and Clarke (2012) defined thematic analysis as a "method for systematically identifying, organizing, and offering insight into patterns of meaning across a data set" (p. 57). The focus of the method is to analyze the entire data set to identify the common patterns across data. The themes are the implicit or explicit patterns that emerge during the analysis; they help to describe the salient features of the phenomena under study (Joffe, 2012).

Furthermore, Joffe (2012) also highlights that the pattern of themes depends on the type of approach that the researcher followed for conducting the analysis. There are two methods of conducting thematic analysis: deductive analysis and inductive analysis. In the case of deductive logic, the researcher conducts the analysis based on an existing theoretical framework, whereas in the case of inductive logic, analysis is performed on the raw data, which allows the researcher to identify the naturally emerging patterns from the dataset (Braun & Clarke, 2006, 2021; Joffe, 2012). The inductive thematic analysis allows the researcher to go beyond the explicitly detectable facts, it helps to analyze the implicit meanings in the data set. I used thematic analysis for my study for the following reasons, as these reasons align well with my study: 1) thematic analysis is considered appropriate for the social constructivism epistemology. 2) Thematic analysis is a commonly used method to analyze interviews and has been commonly used to analyze semi-structured interviews (Joffe, 2012).

Since I performed an inductive thematic analysis, I followed the steps outlined below to conduct the analysis. The steps for conducting the thematic analysis are as follows:

1) *Getting yourself familiar with the data:* The first step was to get familiar with the data collected. This means that as a researcher must spend time going through the data collected multiple times and make notes or transcribe them. Going through the data carefully and consciously allowed me to develop an understanding and take note of the emerging hidden meanings that served as a baseline to answer the potential research questions.

2) *Generating the initial codes*: In this step, I assigned codes to the data. The codes can be semantic or latent. Assigning codes to a set of statements helps me to interpret the meaning of the data. Coding allowed me to categorize the data by providing meaningful labels.

3) *Searching for the prospective themes:* In the third step, I started looking for patterns in the data in the form of themes. The themes are more descriptive than the codes that allowed me to answer the research questions by making meaningful interpretations of the data. In this phase, I reviewed the codes and merged the similar codes to produce a single theme.

4) *Reviewing the generated themes:* This step is complex because at this stage I needed to make some important decisions on what themes to keep and which ones to discard. From a researcher perspective I evaluated each theme based on some criteria and made sure that selected themes answer the research question. While reviewing the themes, the researcher needs to go over the data once more to confirm that the selected themes capture the meaning

of the data collected. In this I reviewed the themes and have also merged some smaller themes into a meaningful theme or also broken down a broader theme into smaller themes based on the requirements of the research question.

5) *Defining the themes:* After the themes were selected, in this step I defined the specific themes in the form of "topic summaries or overviews of things said by participants in relation to a particular topic" (Braun & Clarke, 2021; p. 6). In this phase I also selected quotes or extracts that represent the meaning for each theme. For this phase, I defined each theme and selected appropriate student quotes supporting the meaning of the theme.

6) *Writing the final report:* In the last step, I organized the themes in logical sequence along with quotes to answer the research question meaningfully. The goal of the report "is to provide a compelling story about your data based on your analysis" (Braun & Clarke, 2012, p. 69). I logically arranged the findings, quotes, and reported them in my dissertation.

In this study, I used inductive thematic analysis followed by narrative analysis. The inductive thematic analysis helped me to identify the patterns that are common and different among students.

5.3.6.2   Narrative Analysis

For conducting the narrative analysis, students were clustered based on hierarchical clustering, and representative narrative profiles of students were created for each cluster, based on their interview responses.

The narratives are an excellent form of story-telling to illustrate the human experiences in the world (Connelly & Clandinin, 1990). Since the objective of my study is to understand how students enrolled in The Data Mine form identities, develop a sense of belonging, and socialize with other members, therefore following the narrative analysis approach will allow me to capture the experiences of each participant in the study and express it in the form a story. Since time immemorial humans know the art of storytelling; through this art they express their feelings and "experience the world" (Connelly & Clandinin, 1990, p. 2).

Narrative analysis is an analytical method that has a very broad umbrella. It depends on how the researcher wants to use the analytical lens of the narrative analysis (Elliott, 2005; Phoenix et al., 2010). Narrative analysis is defined by Riessman (2008), "Narrative analysis refers to a family of methods for interpreting texts that have in common a storied form" (p. 11). The text can be in the oral, written, or visual form that helps the researcher to interpret how people perceive the

world around them. Studies by (e.g., Ortega-Alvarez, 2019; Petty et al., 2018) have used narrative analysis to analyze the interview data to understand the experience of people with their world or actions. One important note is that narrative analysis considers the factors such as context and time while constructing the narrative profile of the participants. Phoenix et al. (2010) discussed how narrative analysis can be done in two ways. One method is as a story analyst, the researcher recruits participants, collects data, creates stories, and then analyzes them to understand the experiences of the people in a particular context. The other method to perform the narrative analysis is by being a storyteller. In this case the research follows the similar steps: recruiting participants, collecting data, and in this case the method of analysis is the story, as stories itself has an analytical lens, and the participant uses that lens when he describes his experience within a context.

Since, as a method, narrative analysis has been used to capture the experiences of the participants, it has been used in the field of education to capture the stories and experiences of the instructors and students. For example, Ortega-Alvarez (2019) used deductive thematic analysis followed by narrative analysis to capture the conception of the Columbian Engineering teachers regarding improving teaching practices. Based on the prior studies and my intention to collect data using semi-structured interviews to understand the perception of students in a socially situated environment, I believe that narrative analysis is an appropriate analysis method to answer the guiding research question of this study.

Narrative Analysis was used to create the narrative profile of the students based on the students' interviews. To analyze the stories told by the participants and then "restory" them into a logical order, Creswell suggested a logical order for restorying: "gathering stories, analyzing them for key elements of the story (e.g., time, place, plot, and scene), and then rewriting the stories to place them within a chronological sequence" (Creswell, 2007, p. 56). The focus on reorganization of the stories in chronological order is that the participants may not tell the story in a chronological sequence, therefore the researcher could arrange the story in chronological order while restorying. Chronology can take many forms. It can be in the sequence of start, middle, and end or maybe detailing the past, present, and future. Apart from chronology, the researcher should also focus on the emerging themes and make sure to cover them in the plot of the story.

Specifically, for this study, I followed the steps delineated by (Emden, 1998; Kelly & Howie, 2007; Petty, 2017) for conducting the thematic analysis. The narrative analysis was conducted based on the following six steps; these steps are also represented in Figure 10:

```
┌─────────────────────────────────────────────────────────────────────────┐
│       Reading the interview transcripts multiple times to understand the content │
└─────────────────────────────────────────────────────────────────────────┘
                                    ▼
┌─────────────────────────────────────────────────────────────────────────┐
│          Deleting all the interview questions and comments of the interviewer │
└─────────────────────────────────────────────────────────────────────────┘
                                    ▼
┌─────────────────────────────────────────────────────────────────────────┐
│         Deleting all the words that undermine the key idea of the experiencesf │
└─────────────────────────────────────────────────────────────────────────┘
                                    ▼
┌─────────────────────────────────────────────────────────────────────────┐
│                     Re-reading the clean transcripts again │
└─────────────────────────────────────────────────────────────────────────┘
                                    ▼
┌─────────────────────────────────────────────────────────────────────────┐
│   Repeating steps 3 and 4. Focusing on the series of events/themes that emerge from the │
│                                transcript │
└─────────────────────────────────────────────────────────────────────────┘
                                    ▼
┌─────────────────────────────────────────────────────────────────────────┐
│             Combining the event and themes to generate a coherent story │
└─────────────────────────────────────────────────────────────────────────┘
```

Figure 10: Steps for conducting Narrative Analysis

1. *Reading the interview transcripts multiple times to understand the content*: It is crucial for the researcher to go through the interview transcript multiple times as it helps the interviewer to understand the experiences of the participant and helps to make a note of the important events.

2. *Deleting all the interview questions and comments of the interviewer:* This step is crucial as it helps to get a better understanding of the flow of ideas and experiences.

3. *Deleting all the words that undermines the key idea of the experiences:* Deleting the unnecessary words that undermine the key ideas is important as it hinders the researcher to understand the flow of events or experiences.

4. *Re-reading the clean transcripts again:* Reading the transcript after removing the words gives a clear flow of events and experiences.

5. *Repeating step 3 and 4. Focusing on the series of events/themes that emerge from the transcript:* The steps 3 and 4 are repeated multiple times to identify the broad themes that emerges from the interview transcripts.

6. *Combining the event and themes to generate a coherent story:* In the last step, the themes and events are combined to create a coherent story about the participant.

Following the steps mentioned in Figure 10 helped me to conduct my narrative analysis in an organized fashion. Therefore, we can conclude that "the narrative study tells the story of individuals unfolding in a chronology of their experiences, set within their personal, social, and historical context, and including the important themes in those lived experiences" (Creswell, 2007, p. 57).

5.3.7 Trustworthiness

Establishing the trustworthiness of the research findings is crucial in any type of research. It is important to note that establishing the trustworthiness of qualitative research is different from proving the validity and reliability of the quantitative research (Shenton, 2004). To prove the trustworthiness of the qualitative research the researcher must meet the following criteria proposed by Guba (1981): credibility, transferability, dependability, and confirmability (Shenton, 2004). *Credibility* corresponds to the internal validity in the case of quantitative research. In the case of qualitative research, credibility refers to the following standard operating procedures to conduct the qualitative research (Connelly, 2016; Shenton, 2004). The adoption of correct procedures to carry out research will also establish a trust that the findings of the study are consistent with the reality (Guba, 1981; Shenton, 2004). For this study to establish the credibility of the research I followed the standard procedures delineated by Creswell (2007) for conducting the narrative inquiry and steps outlined by Braun and Clark (2012) to conduct the thematic analysis.

Transferability corresponds to the external validity in the case of positivist research. Transferability refers to what extent the findings of the study can be applied to implement in another context. In the case of qualitative research "researchers support the study's transferability with a rich, detailed description of the context, location, and people studied, and by being transparent about analysis and trustworthiness. Researchers need to provide a vivid picture that will inform and resonant with readers" (Connelly, 2016, p. 436). In this study, I intend to provide a vivid description of The Data Mine as a learning community and opportunities provided to students in The Data Mine that help them to form identities, learn data science skills, experience a

sense of belonging and socialization. Moreover, I interviewed students to understand their perception of The Data Mine and the opportunities provided.

Dependability corresponds to reliability in the case of quantitative research. To establish dependability in the case of qualitative research, the researcher needs to provide a detailed description of the research design and process so that the study can serve as a model for others conducting similar studies (Shenton, 2004). Shenton further stressed that the detailed explanation of research design and methods will help the researchers and readers to develop an in-depth understanding of the methods employed and their effectiveness. To meet the dependability criteria, I provided a detailed explanation of the qualitative research design that I used for this study. Also, I provide details on how I intend to collect, analyze, and present the findings of the study.

Confirmability refers to the objectivity of the researcher. Confirmability is the extent to which findings of the study are consistent and ensures repeatability (Connelly, 2016). In the case of qualitative research, it is important to ensure that the findings of the study are the representation of the experiences and stories of the participants and not the researcher. There are multiple methods to ensure confirmability of the research, such as triangulation, interrater reliability, member checking, peer debriefing etc. To ensure the reliability for this study, I performed member checking. For the member checking the narrative with student quotes were send to participants to confirm it the narrative resonates their described experiences. Peer-debriefing was conducted with the help of a peer (Guba & Lincoln, 1989; Lincoln & Guba, 1985; Spall, 1998). For the purpose of peer debriefing, I met a doctoral student with a strong research method background and also an expert in education research methods from my lab. The peers were not part of The Data Mine project. Regular meetings and extensive discussion were conducted around the research questions and research methods. For the data analysis, peers and I independently coded two interviews and discussed results. Based on the discussion, I revised the code book and re-did the coding. The final themes that emerged were discussed, and similar themes were merged. Furthermore, I wrote the summaries for the narratives for the two interviews and discussed the results with the peers to seek their feedback. Conducting the peer-debrief for this study helped me identify and focus on the parts of the story that needed more attention. It also helped me to identify some aspects that I understated, or points that were exaggerated during narrative analysis. It also allowed me to review the themes and come up with final themes for my results section.

5.3.8 Ethical Considerations

Ethical considerations are mandatory in case of human subject research as it ensures protection to the participants under study. Since this study involves human participants the permission from Institutional Review Board was taken before conducting the survey study in the September of 2020. For conducting the interview study another IRB approval was taken in the July of 2021.

Quantitative Survey Study before taking part in the survey study, students signed an IRB form that served as their consent for taking part in the study. The consent form provided the details of the study, the potential risks/benefits, and compensation. Prior to the study, it was communicated clearly through the recruitment email and consent form that participation in the survey was voluntary and students can leave the study at any point in time. Also, only the complete responses will be considered for compensation and data analysis, whereas incomplete responses will be dropped from the study. Steps were taken to maintain the privacy of participants and ensuring the confidentiality of responses, such as students were identified before conducting the data analysis and all information is stored electronically in an appropriate location to avoid any data breach.

Interview Study before conducting the interviews permission from the IRB was obtained. Students provided consent by agreeing or disagreeing to participate in the study. The consent form provided the details of the study, the potential risks/benefits, clauses for compensation, and student rights, such as voluntary participation, withdrawal procedure, etc. After obtaining the consent, the interviews were conducted, and students were compensated. Since the interviews were conducted and recorded via Microsoft Teams, all the audio files will be deleted after the data is analyzed.

CHAPTER 6: STUDENT PERCEPTION OF IDENTITY, SOCIALIZATION OPPORTUNITIES, SELF-CONCEPT, AND ACADEMIC/ INTELLECTUAL DEVELOPMENT

In this chapter I intend to present the results of the first phase of the mixed method design. An initial quantitative study was conducted to answer the following research question:

RQ 1: What are the perceptions of students regarding their identity, socialization opportunities, self-belief, and academic/intellectual development in The Data Mine?

The findings of the quantitative analysis were grouped into two categories: (I) Results for the 50 Pretest students and 25 Posttest students, and (II) Results for 25 students who participated in both pre and posttest.

6.1   Results for 50 Pretest and 25 Posttest students

In this section I present the result for first category that is for 50 pre-test and 25 post-test students. The results are presented in the form of visualization. The bar charts clearly demonstrate the increase in the perception for all the constructs.

6.1.1   Perception of Identity Formation

The bar chart below presents the visualization results for the three sub-constructs of identity formation. The analysis of the bar plot revealed that the students demonstrated a gain in perceptions for all the three sub-constructs of identity as they progressed through the semester. Figure 11 shows the mean and standard deviation for the recognition, interest, and performance/competence for the students' pre-and post-survey responses.
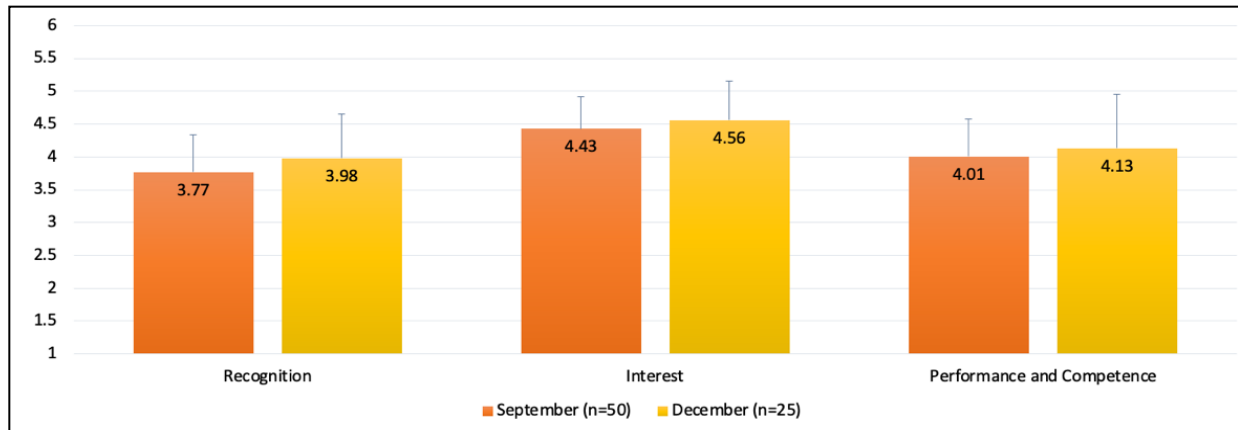
Figure 11: Perception of Identity Formation among The Data Mine students

6.1.2 Perception of Socialization

The socialization survey consisted of Yes or No questions. The survey was used to evaluate the following (1) perception for faculty/staff mentor in The Data Mine and (2) perception of peers in The Data Mine. The survey conducted in September (pre-test) of 2020 revealed that students demonstrated a lower level of socialization with mentor/faculty/staff, and moderate level of socialization with the peers. Whereas the survey conducted in the December (post-test) 2020 revealed an overall increase in the socialization perception for both constructs. This means that students found more opportunities to socialize with faculty/staff/mentor, and peers in The Data Mine, refer to Figure 12.
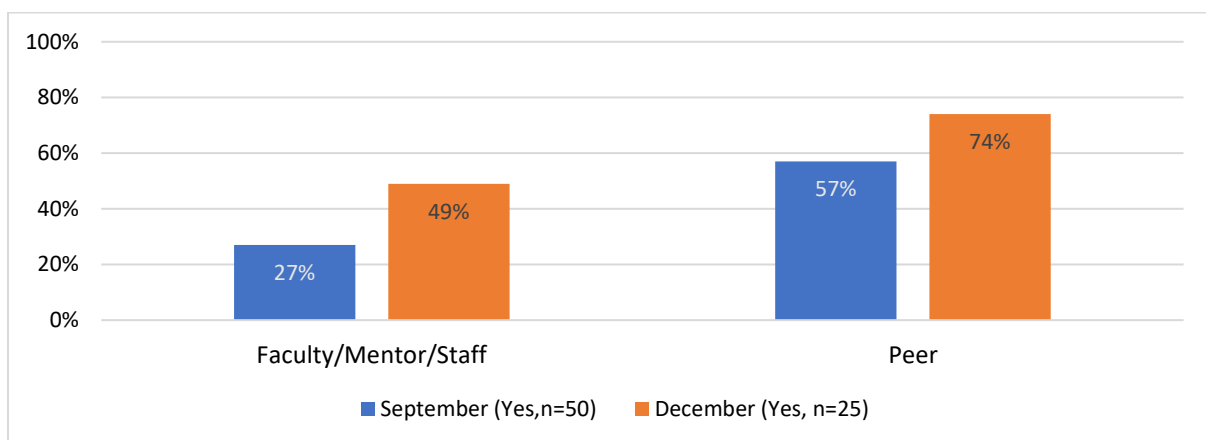
Figure 12: Perception of Socialization among The Data Mine students

6.1.3 Perception of Sense of Belonging

The bar plot revealed that students in The Data Mine demonstrated an increase in the perception of a sense of belonging as they progressed from September 2020 to December 2020, refer Figure 13. Sense of belonging refers to the perception of students on how accepted or included they feel in The Data Mine.
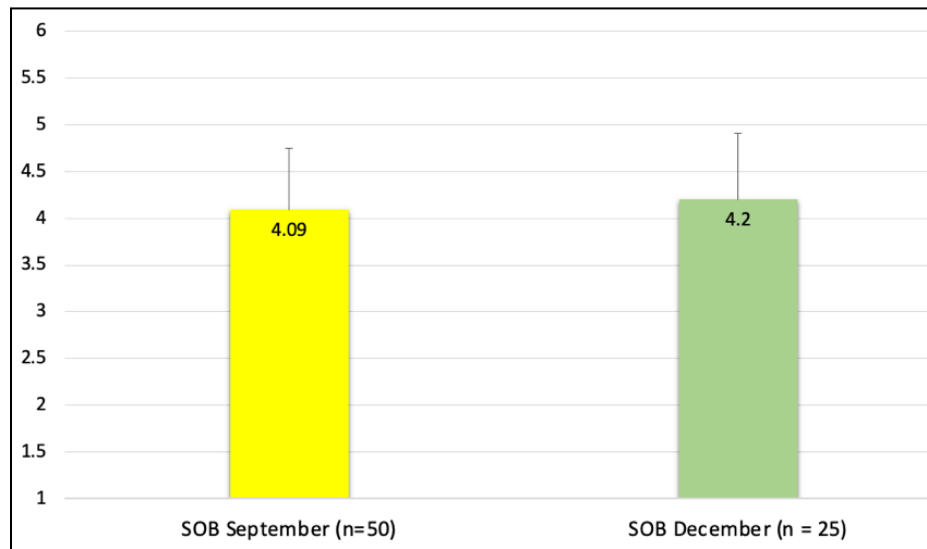
Figure 13: Perception of Sense of Belonging among The Data Mine students

6.1.4 Perception for Data Science Self-Belief

The perception for data science self-belief is comprised of three sub-constructs: data science self-concept, data science fixed mindset and data science anxiety. The results for all three constructs are presented below:

1) **Data Science Self-Concept:** The self-concept of students increased as students progressed from September to December. Self-concept refers to the one's perception about their behavior, abilities and distinctive characteristics (Bailey, 2003). Self-concept is shaped through interaction with others and contributes towards identity formation. Increase in self-concept demonstrates that students felt an increase in their self-worth. Refer Figure 14 to see the increase in the self-concept.
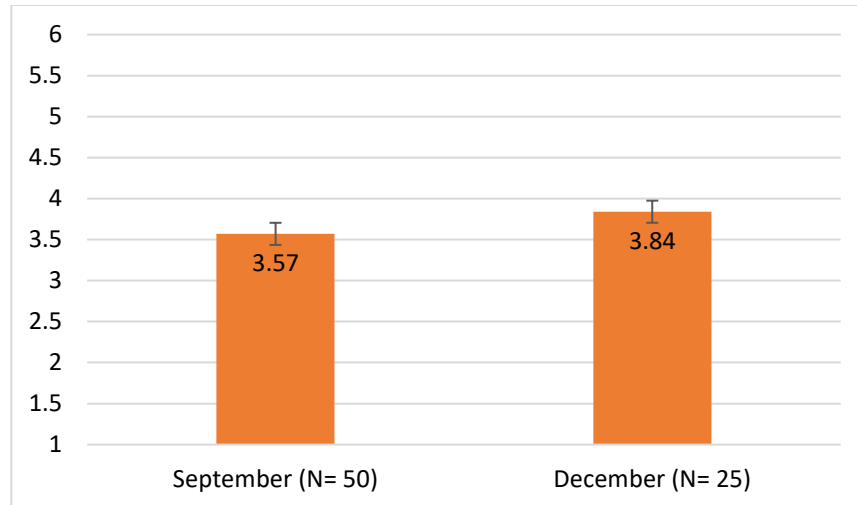
Figure 14: Perception of Self-concept among The Data Mine students

2) **Data Science Fixed-Mindset** The data science fixed mindset of students decreased as students progressed from September to December. Fixed mindset refers to the perception of students where they believe that they cannot change much about their knowledge or learning ability (Scott & Ghinea, 2014). A low average for fixed mindset demonstrates that a majority of students believed that they could learn new skills and increase their knowledge or improve their learning ability. Figure 15 represent the decline in fixed mindset among The Data Mine students.
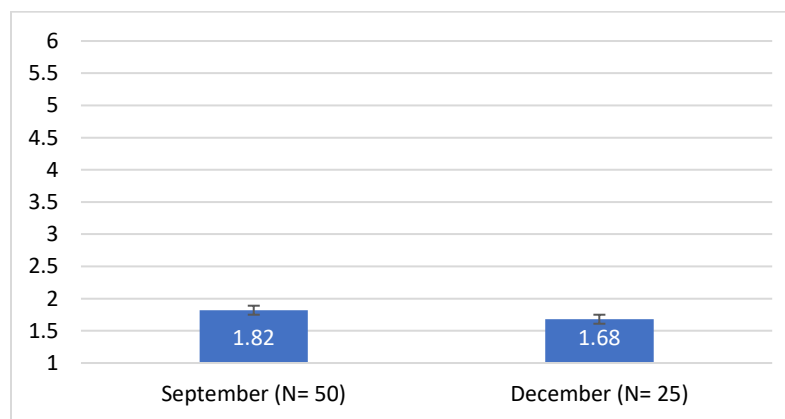
Figure 15: Perception of Fixed Mindset among The Data Mine students

3) **Data Science Anxiety:** The data science anxiety of students decreased as students progressed from September to December. Anxiety in the context of data science refers to the situations where student express issues with trouble shooting the errors, feels intimidated or helpless in solving a data science problem. Decrease in anxiety among the students means that students felt confident in approaching data science problems (Scott & Ghinea, 2014), refer Figure 16.
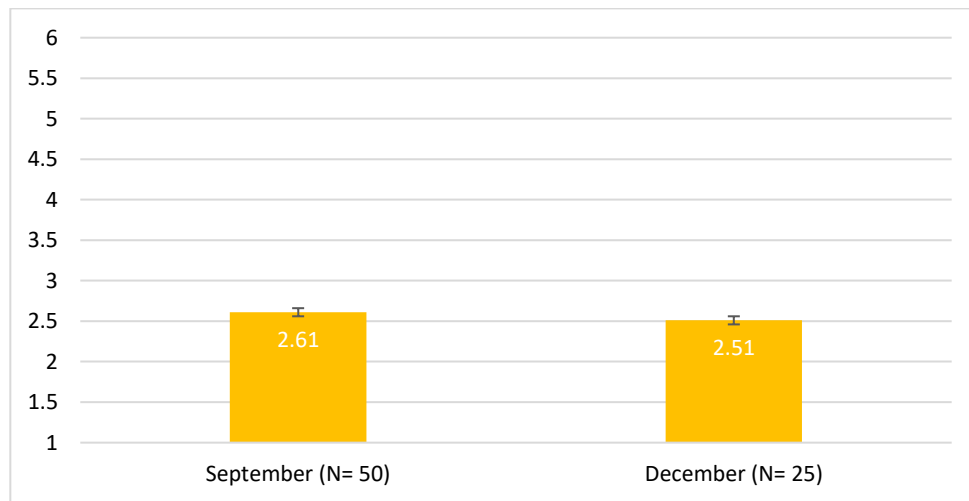
Figure 16: Perception of Data Science Anxiety among The Data Mine students

6.1.5 Perception of Academic & Intellectual Development

The perception of academic and intellectual development increased as students progressed from September to December. This means that students considered themselves to be academically and intellectually confident as they progressed through the semester. Figure 17 demonstrates the increase in academic and intellectual development.
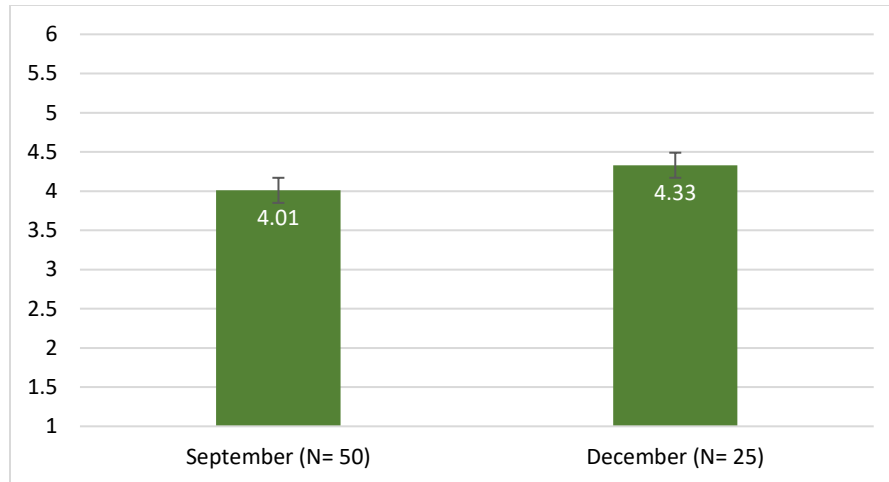
Figure 17: Perception of Academic and Intellectual Development among The Data Mine students

6.2 Results for 25 Pre and Posttest Students

The results for 25 students who participated in pre-test and post-test are presented in the paragraphs below. We present the results in the form of descriptive and inferential statistics. The result of the study reveals that there was an increase in perception for all positive constructs and decrease in the perception for two negative constructs such as data science fixed mindset and data science anxiety. To test the significance between the pre-test and post-test results, the paired sample t-test was conducted for all constructs, except socialization. McNemar paired sample test was conducted for socialization construct as the data was binary (Yes/No responses) in nature (Ciechalski et al., 2002). The result for each construct is presented below:

6.2.1 Perception of Identity Formation

The results of descriptive analysis revealed that students demonstrated an increase in the perception for all the three constructs of identity formation: recognition, interest, and performance/competence. The results from Table 7 for the paired t-test demonstrated that there was a statistically significant increase only in the perception of recognition in posttest (M= 3.98, SD = 0.67) from the pretest (M= 3.73, SD = 0.60); t (24) = 2.08, p = 0.04.

Table 7: Perception of Identity Formation

	September N= 25		December N=25				
	Mean	SD	Mean	SD	Gains	t	p
Recognition	3.73	0.6	3.98	0.67	0.25	2.08*	0.04
Interest	4.48	0.5	4.56	0.59	0.08	1.11	0.27
Performance and Competence	4.02	0.64	4.13	0.83	0.11	0.85	0.39

6.2.2 Perception of Socialization

The results of analysis revealed that students demonstrated an increase in the perception for both the constructs of socialization, they are, *Socialization with mentor/faculty/staff in The Data Mine and Socialization with peers in The Data Mine.* Since the data was binary in nature, a McNemar paired sample test was performed to understand if there is any significant difference between the pretest, posttest results for *Socialization with mentor/faculty/staff in The Data Mine and Socialization with peers in The Data Mine.* Although there was an increase in perception for socialization, the results of the McNemar paired sample test revealed that there was no significant difference in the perception of students for socialization with mentor/staff/faculty or peer.

a) Perception regarding socialization with faculty/staff/mentor in The Data Mine

From the contingency Table 8 we observe that number of students who agreed that they got opportunities to socialize with Faculty/Staff/mentors at The Data Mine increased as they progressed through the semester. A McNemar paired sample test was conducted to see if the change in the perception of students were significant (Ciechalski et al., 2002). The test revealed that there is no significant difference ($X (1) = 3.13$, p=0.07).

Table 8: Student perception of socialization with faculty/staff/mentor

	December (YES)	December (NO)	Total
September (YES)	4	1	5
September (NO)	7	13	20
Total	11	14	25

b) *Perception regarding Socialization with peer in The Data Mine*

From the contingency Table 9 we observe that number of students who agreed that they got opportunities to socialize with the peers at The Data Mine increased as they progressed through the semester. The results of the McNemar paired sample test revealed that there is no significant increase in the perception change ($X (1) = 0.17$, $p=0.683$)

Table 9: Student perception of socialization with Peers

	December (YES)	December (NO)	Total
September (YES)	14	2	16
September (NO)	4	5	9
Total	18	7	25

6.2.3 Perception of Sense of Belonging

The results of analysis revealed that students demonstrated an increase in the perception for the sense of belonging. The results of the paired t-test demonstrated that the increase was not significant, refer Table 10.

Table 10: Perception for sense of belonging

	September N= 25		December N=25				
	Mean	SD	Mean	SD	Gains	*t*	*p*
Sense of Belonging	4.13	0.65	4.2	0.71	0.07	0.75	0.46

6.2.4 Perception of Self-Belief in Data Science

The results for the three constructs of the student self-belief are demonstrated in Table 11.

a) **Data Science Self-Concept:** The results of descriptive analysis revealed that students demonstrated an increase in the perception for self-concept. The results of the paired t-test demonstrated that there was a statistically significant increase in the perception of self-concept in posttest (M= 3.55, SD = 0.71) from the pretest (M= 3.84, SD = 0.80); *t (24)* = 2.04, *p* = 0.02.

b) **Data Science Fixed-Mindset:** The results of descriptive analysis revealed that students demonstrated a decrease in the perception for data science fixed mindset, but the decrease was not significant. Results of the t-test in presented in Table 11.

c) **Data Science Anxiety:** The results of descriptive analysis revealed that students demonstrated a decrease in the perception for data science anxiety, but the decrease was not significant. Results of the t-test in presented in Table 11.

Table 11: Perception of Student Self-Belief in Data Science

	September N= 25		December N=25				
	Mean	SD	Mean	SD	Gains	*t*	*p*
Data Science Self Concept	3.55	0.71	3.84	0.8	0.29	2.40*	0.02
Data Science Fixed Mindset	1.74	0.72	1.68	0.66	-0.06	0.41	0.68
Data Science Anxiety	2.69	0.83	2.51	0.87	-0.18	1.7	0.1

6.2.5 Perception of Academic & Intellectual Development

The results of descriptive analysis revealed that students demonstrated an increase in the perception for academic and intellectual development. The results of the paired t-test demonstrated in Table 12, which shows that there was a statistically significant increase in the perception for academic and intellectual development in posttest (M= 3.99, SD = 0.76) from the pretest (M= 4.33, SD = 0.57); t (24) = 3.16, p = 0.004.

Table 12: Perception of Academic and Intellectual development

	September N= 25		December N=25				
	Mean	SD	Mean	SD	Gains	*t*	*p*
Perceived Academic and Intellectual Development	3.99	0.76	4.33	0.57	0.34	3.16*	0.004

6.3 Summary

The result of the quantitative study demonstrated that there was increase in the perception of identity, socialization opportunities, sense of belonging, self-belief and academic/intellectual

development as students progressed from September to December. Also, there was a statistically significant increase for the following sub-constructs: recognition, self-concept, and academic/intellectual development. The overall analysis revealed that students benefited from participating in The Data Mine but identifying the factors that led to the increase in identity, socialization, sense of belonging, self-belief and academic/intellectual development is crucial. Also, developing a deeper understanding of student experiences is required. Therefore, conducting a qualitative inquiry and interviewing students will allow us to understand various opportunities in The Data Mine that helped them develop the above-mentioned psychosocial skills.

CHAPTER 7: STUDENT PERCEPTION OF BENEFITS, CHALLENGES, SENSE OF BELONGING AND IDENTITY IN THE DATA MINE (THEMATIC ANALYSIS)

This chapter intends to answer the following three sub-research questions and present the results of the thematic analysis in detail.

- **Sub-RQ 2(a):** What are the perceived benefits and challenges of participating in The Data Mine?

- **Sub-RQ 2(b):** How do students describe their levels of sense of belonging and opportunities for socialization within The Data Mine?

- **Sub-RQ 2(c):** How do students' participation and interaction in The Data Mine help them form their identity?

7.1 Student Perception of Benefits and Challenges

This section describes the student experiences in The Data Mine in terms of the benefits and challenges they experienced as a member of learning community. We intend to answer the following research question **Sub-RQ 2(a):** What are the perceived benefits and challenges of participating in The Data Mine?

7.1.1 Themes for Benefits in The Data Mine

The student perception of the benefits of The Data Mine describes the various opportunities, advantages, and resources that The Data Mine provided to students. It is divided into three broad themes. Each theme is described below and with their relevant codes and representative student quotes.

Theme 1: *The Data Mine provided multiple resources, communication channels, and opportunities to participate in events to develop data science skills.*

Theme 1 demonstrated that students in The Data Mine were provided various resources such as TAs, staff, faculty mentors and graduate students they helped the undergraduate students to develop data science skills. Students were offered technology-based resources such as the Examples Book (Amstutz et al.,2021). The Data Mine also established communication channels,

including Piazza, which allowed the students to ask questions related to data science or share any updates. The Data Mine also provided opportunities to develop data science skills by allowing students to participate in the seminar course and outside events.

Table 13: Codes for Theme 1

Codes	Students	Frequency
TAs were helpful and knowledgeable	19	42
Seminar course helped to develop data science skills	19	36
Faculty and Staff were supportive and knowledgeable	17	31
Dr. W is a wonderful Mentor	16	27
Outside events helped to understand the application of data science	16	18
Piazza platform to ask questions and seek help	12	14
Graduate Students were helpful	10	16
Helpful Examples Book	7	7

Table 13 demonstrates the codes that led to Theme 1. The various resources, medium for constant communication and participation in events was very beneficial to students as it helped them to develop data science skills and promoted collaboration. The paragraphs below describe each code that led to formation of the theme 1.

1. **TAs were helpful and knowledgeable:** The students acknowledged that the TAs were a great resource in The Data Mine as they were helpful and knowledgeable. The students mentioned both the TAs in the Seminar Course and Corporate Partners were very competent to handle student queries. Quinn described the seminar course TAs as knowledgeable and helpful as they were always present to answer questions during the seminar class, conducted office hours and also resolved queries on Piazza. Quinn also believes that the TAs helped him to develop data science skills and answer any questions he had either regarding the seminar project or coding.

   TAs were always there to answer our questions. They had office hours pretty much available throughout the week. You can approach them with any problems. They used to answer queries on Piazza. I used that a lot, as I got quick responses. TAs helped me a lot to develop my data science skills because anytime I had any type of questions or was missing any part of code, they were always there to fulfil it and that definitely contributed to my data science learning.

Reef also found TAs and faculty very helpful as they resolved his queries during the class. Reef also approached the TAs after the seminar if he had any questions about the project problems.

TAs and Dr. W were extremely helpful during the seminar class. Uhm, the TAs guided us and helped us to understand the problem during the seminar. Even after the seminars I use to go to TAs if I could not solve any problem.

Riley speaks about her experience when she joined The Data Mine. She was from a non-STEM background and not well versed with data science concepts. She used to approach her corporate partner TA with questions, and she found him very accommodating, supportive, and patient.

Uhm, my TA in the corporate partners was very helpful. That is one best thing about being in The Data Mine, TAs do not expect you know all the stuff and they don't get mad when you don't. They are very helpful, even if you are like, at times I feel nervous to ask questions, but I always felt they are willing to answer any questions you have.

Morgan described her corporate partner TA as knowledgeable and helpful. She approached her corporate partners TA when she had any questions, or even for non-academic advice. Morgan always found the TA very supportive, easy to talk to and friendly. Morgan and the TA developed a good interpersonal relationship.

Ah, my TA definitely answered a lot of my questions about The Data Mine as a whole and also just about project expectations or what I needed to prepare for every week. Uhm, and she just was a really good resource. If I ever had any questions about anything. I would tell her all the time if I was worried about not being able to contribute to something or if I was pulling a lot of weight of this certain part of the project. We definitely do well develop a good relationship and I caught up with her last week. We're still friends.

2. **Seminar course helped to develop data science skills:** The students mentioned that they found the seminar course a great foundational resource that helped them to develop data science skills. Students also described seminar course as a perfect opportunity to learn and interact with peers. Frankie described the seminar class and the learning environment collaborative and cool, as it allowed her to talk to her peers, have lunch and work at the

same time. Seminar class helped Frankie to ask questions to TAs and Dr. W during the class time.

We had a weekly seminar class in the dining court of Hillenbrand. I went to that every week. Dr. W would teach us some new data science skills or language or what not. We also did projects on that every week. Seminars were unique, it was kind of a cool new way, and it was lot easier to collaborate with other students, TAs, Dr. W, because it was very casual setting, we were eating and learning together.

Harley is a non-STEM major, and she found seminar class important as she learnt multiple ways to solve same problem from her peers and Dr. W.

We had seminar class every Monday, it was interesting to see what problems other people were getting stuck on and different ways people went about solving those problem. It was actually great to see that students found their way to solve problems. And then Dr. W would catch up on that and say, yes you can actually do it this way. It was really helpful to know so many different ways to solve the same problem.

Rio joined The Data Mine during the COVID-19 pandemic, so she has attended both virtual and this year she is attending in-person seminar class. She found both virtual and in-person seminar classes interactive and collaborative. She likes how the class helped the students to learn various ways to solve problem. She found Dr. W actively involved in helping students to answer any questions that they might have, and students helped each other to solve problems.

So, I was involved in seminar both this year and last year. Last year it was all virtual because of COVID, but even then, it was pretty interactive, and we would just join like a WebEx session with Dr. W and which we could like work individually on our projects. And then if we had any questions, we could ask him directly and he would answer and that was really beneficial because it gave you like a certain amount of time to work on a project that involved like data science and coding. And then there was help immediately if you needed it. And then there's also moments where other students in the session could chime in and give their two cents. Uh, so despite it being like on a virtual platform, it was still pretty interactive.

In the Fall of 2021, the class was in-person, happening at the Hillenbrand dining court atrium. Rio loves the in-person model of the class as she finds it fascinating to see how

students are immersed in a problem-solving environment. Dr. W and TAs are always present to solve any problems or answer any question during the seminar. Rio attends the seminar class along with her friends and she loves to bounce ideas back and forth with her friends while working on the problems.

This year I have had it in Hillenbrand at the dining court atrium and I really like it because it kind of puts you in, uhm, like an environment where everyone's kind of working on the same thing towards the same goal. I have a few friends of Data Mine, so I'll go to the seminar together and will sit together and work through the project. Uh, and then if we have any questions, both Dr. W and all of the TAs on hand, so it's just a really nice collaborative environment where we can work with each other, and kind of like bounce ideas off each other.

Orion found that the seminar course helps the students to incrementally build the knowledge. For example, in the first semester, the class focused on helping the students to develop a basic understanding of programming and statistical analysis. In the second semester, the class focused on the application of the learned programming skills to solve real problems. He discussed how the seminar course projects in the first semester were easy, such as creating simple graphs and visualization, and it got very creative in the second year. It involved creative and critical thinking as the intent was to help students apply the knowledge that they gained in their first semester.

You take, so it goes from 100 level course to a 200-level course to three, and so you keep learning new things. So, since I've done it for two years, the first year they just taught you, like I said the tooling so it the takeaway was mostly how to do stuff at the start. So, if you're given this data set how your how are you going to query it in an order? What commands you use? How to create plots? Like it was more applied, and it taught you how to write the code and implement it. That was the focus in the second semester. It moved towards software development. Building data frames on panels and then trying to understand so building more so not simple bar plots and scatter plots going beyond that to understanding, how to make a heat map. Then they started teaching us what kinds of plots to use to derive certain insights. So, they gave us a full data set and they and they gave us a problem statement and then asked us it was starting to be more creative. So, they gave a data set and they said, let's say you are an airline administrator or manager and if you had like an airline data set, and you want to come and make a new product or and make a new promotion offer. What would you suggest would be a good promotion offer from this data set so then you could use different distributions? Then, what different types of analysis you can do? So it's a matter of using everything you learned and then applying it, and then trying to derive insights on your own.

3. **Faculty and Staff were supportive and knowledgeable:** The students found the faculty and staff at The Data Mine were very helpful, supportive, and knowledgeable. This includes all the faculty members and staff, including the Data Scientists of the Data Mine. The students mentioned about multiple academic, moral, or emotional support they received from faculty and staff at The Data Mine. Finley described faculty and staff at The Data Mine as proactive in answering students query and clarifying any doubts. He could not be on campus during his first semester and he was connecting remotely from his home country, but he can never feel that there was any delay from the part of The Data Mine faculty or staff in getting back to his emails.

Faculty and staff at The Data Mine are very helpful. I know that I can email Dr. W. I know that I can mail Ms. M. You know any one of them or any other staff and I'll have a reply within pretty much the next hour. You know, I sometimes reached out to them with just general issues, such as when you're signing paperwork, or may be when you're working on projects. Obviously, I've mailed them at odd hours because in the fall semester I was in my home country, so you know they weren't necessarily awake. But I don't think I've ever had to wait for more than like 3 or 4 hours for a reply. And that is like the worst-case scenario.

On the same lines Harley shared her experience; she found The Data Mine faculty and staff care for their students. She shared an instance where Dr. W reached out to her and asked her how she was doing.

As far as The Data Mine faculty and staff are concerned, they are amazing. They care for their students. For example, this summer also Dr. W reached out me asking, hey how are you doing? I was wondering what's going on with you. This demonstrates how much faculty or staff care about the students.

Jackie is TA for Corporate Partnerships, and he shared his perspective. He has always found Dr. W very supportive, and he also appreciates Ms. M who manages the corporate partnerships. Jackie says that it is not necessary to reach out to The Data Mine faculty and staff with an academic question only, you can talk to them about any issue that you are facing.

In Data Mine if you are stuck on a problem you can go to Dr. W or other staff. Ms. M has been a great resource manages the Corporate Partnerships. As a TA I go to her for Corporate Partner related question. Um the best part at The Data Mine is

you can reach faculty or staff with any questions. It need not be academic, it could be professional advice, and they have a huge network, and they will work with you to solve your problem.

4. **Dr. W is a wonderful Mentor**: The students described Dr. W, as a great mentor. The students mentioned about their interaction with Dr. W and expressed how caring and emphatic he is as a mentor. Arrow describes Dr. W as an excellent person, as he cares for his students very much. He sits down with his students to solve any problems or guide them.

Dr. W is a great mentor. I mean he's just an excellent guy who definitely prioritizes students and he cares about them a lot. If you need some mentorship or, given if you have very specific technical questions, he's willing to sit down with you and work with you to solve the problem.

Orion considers Dr. W, as his advisor. He has worked with Dr. W since he joined The Data Mine, and he has found him very supportive and helpful. He has reached out to Dr. W with any coding or data analysis related questions and Dr. W has always advised him. Orion also loves that Dr. W is very quick in answering students query and he acknowledges that mostly Professors are busy, but Dr. W prioritizes his students and gets back to them as early as possible.

Uh, definitely working with Dr. W helped me a lot because he would provide us a lot of advice. So, he is the one that gave me a lot of advice on how to write code, data analysis and what exactly I can use. Above it so he was more of my advisor throughout The Data Mine. Uh, experience I would say and also, he was one of the most welcoming professors I've met at Purdue, and I'm surprised at how quick his responses are on email. Normally by email a professor takes at least two or three days. Understandable that if they would get 100 or even hundreds of emails every day. Dr. W is never like, he's never delayed on my emails for more than a day, ever. It's been at least two or three hours and I would get a response for him instantly. So that was an amazing mentor that I had met because of The Data Mine.

5. **Outside events helped to understand the application of data science:** Students mentioned about their participation in the outside events. They found outside events motivating and attending these events helped students to understand the application of data science in real world. Jackie mentioned that he attended the outside events and found them

great. Since he loves sports and numbers, he specifically mentioned about an outside event that focused on sports analytics, and he found the event very cool, as he was able to see a very different application of the data science. Jackie said:

Other events they had outside events is what they call them that we had to go to every year. They'd have speakers come and it was all just about different data science applications. I mean kind of like mini lectures, I guess. So, there was a guy he was big into like sports analytics. I mean I love sports. I love messing with numbers. I actually do a lot of sports data, uh, so that was a perfect fit for me and to go and see someone else talk about their experiences, their findings, you know, different sports that they find cool to look at, things like that from a numbers perspective, as it's really cool to see all these different lectures. But it's really cool to learn from people who have, you know, done it their whole career and had great success with what they do.

Quinn also found the outside events to be inspirational and a good way to understand the application of data science in real world. He mentioned about an event where a speaker from the Purdue Co-Rec came and spoke about how they analyzed the swipe in and swipe out data to predict the number of visitors for each semester.

Outside events were actually really cool. There was one in particular I remember the speaker from COREC came and visited and he kind of just like talked about some data analysis they did based on user swipe ins. So, like what kinds of people are swiping in, meaning like whether are they on campus or off campus say for a semester or like throughout the year? When do people tend to come more? There was just interesting to see like what kinds of data analysis they did. That one stuck out to me.

Also, Quinn was able to attend outside events pertaining to data ethics; he finds those events very valuable.

Uh, it was cool to see like there was a lot of um ethics focused outside events. They were very cool. It is an issue that we should talk about, and it was cool because they choose people to speak about it in terms of data and systematic issues in the data science field. I think that choosing to have outside events and people speak to us about it was super cool.

Orion was also on the same lines. He found outside events valuable as they helped to learn the application of data science. He recalled an event where the Purdue basketball coach

visited The Data Mine and he spoke about how their teams use sports analytics to understand the game. Orion was fascinated to see the application of data science in the field of sports.

Uh, so there was an event where we got to meet up with the Purdue basketball team coach. In that event I got to talk to Purdue basketball coach, and he mentioned how they used a lot of data analysis. They analyzed their draft picks, like they would analyze from just the heights of the players, their agility and then their results in terms of their shooting percentage, their interval percentage, and so on. So, the numbers they spoke and then every single decision they made was based on the data. Playing the statistics and understanding how to use data related to that I felt was very interesting. It was even a sport where I thought it was a lot more about talent. And you know, hard work. There's a lot of numbers involved in it as well, so the way they assign players to which position has a lot of data involved in it. And understanding even other teams' defensive strategies, their offensive strategies are also data driven.

6. **Graduate Students were helpful:** The students emphasized on the importance of graduate students in their teams and described in what ways having graduate students was helpful for them. Frankie mentioned that in her team there were three graduate students, and they were very helpful to the entire team as they were well versed with statistics and data science concepts. Frankie mentions that even though she was the TA, she depended a lot on the graduate students as they were like mentors for all the undergraduate students in the team.

There were three graduate students on my team last year. They were great. I would also consider them really good mentors to me as well, even though I was kind of leading the group and I was that TA like they definitely carried a lot of the project just because they had a lot more experience with data science and statistics. Because last year, you know, I'm currently in a regression class and a time series class, but I was not in those last year and we use a lot of those concepts last year. So, I definitely leaned a lot on the graduate students who had, you know, taken classes in that before. So, they were very very helpful, and they were not only helpful to me, but the students as well. And they acted as great mentors to the other undergraduate students too.

Orion also found the graduate students friendly and helpful. He mentioned that it is difficult to establish contact with graduate students as they attend different classes but being in The Data Mine, Orion was able to meet graduate students and work with them.

That was I would say a very good experience because you don't get to meet a lot of grad students and the two grad students on the team were actually very friendly and we're very open to help. So that worked out really well as well. So, I connected with them very closely too.

Kendall also agreed that having graduate students was very helpful. She spoke about her Corporate Partners project experience. There were two graduate students on her team. According to Kendall even though they were not designated team leads but one of the graduate students played a very important role of managing the team and getting the work done.

There were two grad students. One of them was in nursing and one of them was in consumer science, so they were very knowledgeable in like our field and the two of them were able to help us put together our presentations and stuff. And then I would also say that one of the graduate students was a leader in a way, because she would often kind of put our presentations together and make sure everybody was on the same page.

7. **The Piazza platform promotes interaction among students and faculty:** Piazza is an online collaborative communication platform that is used by The Data Mine faculty, students, and staff. Students mentioned they used of Piazza for asking questions, seeking clarifications, sharing information, or replying to others. Jamie mentions her experience with Piazza from academic standpoint and she found Piazza a wonderful platform to ask questions and clarifications.

Piazza is a great platform to ask and answer questions. From an academic standpoint it is a great place to ask. Dr. W or TAs generally answer the questions, but any students can also answer the questions.

Just like Jamie, Reef also found Piazza a wonderful platform to ask questions. Reef mentioned that Piazza was so extensively used by students that his 98% of questions were already asked by some other student on Piazza. He also liked that posting can be anonymous on Piazza, so students did not hesitate to ask questions.

We have Piazza, which is like a question answer forum. Uh, so a lot of problems or questions were asked on there, and so it was a very good resource that I could use. 98% of the time my questions were probably already addressed on Piazza. Students

could post notes and I think one good thing about Piazza is like it can be posted anonymously.

Kendall also agreed that Piazza was great platform She mentioned that interaction on Piazza was extensively high. Students asked questions and other students, TAs or Dr. W would respond to the queries. She also gave an example, let's assume, a student asked a question regarding a problem and later a student was successful in solving that problem, then they would get back to other students on Piazza and share their approach to that problem. So, Kendall found that Piazza was also a good platform to learn from one another.

Piazza is a great platform to ask questions. Uhm, I remember Dr. W gave us like a number of interactions. It was something crazy too. It was a lot. I can't remember exactly what it was. You also get you tend to get very quick responses so. The excellent thing about it is that it's not just one TA looking at any particular thing, you're gonna have multiple people. So, whoever can answer the best or the fastest will always get back to you. So, if a student finds something that's particularly challenging in a project, what they will do is ask on Piazza. And sometimes, even on Piazza, you might not figure it out and then they will come back later on and say, hey, if you're still having this problem, this is what I did that fix it. So, you got a lot of help from other students in that way as well.

8. **Helpful Examples Book:** The Examples Book (2021) is an online resource for students in The Data Mine that has solutions and videos related to various data science problems. Students mentioned about their experience in using the Examples Book and also described in what ways it was helpful to them while working on seminar projects. Quinn finds it a great resource and very helpful, as every problem in the Examples Book was supported by a video and student could view the video and solve similar problems. It also contains great examples and makes the coding easier for the students.

Uh, the [Examples Book] that every single problem comes with a video to on how to go through each problem comes with a video made by Dr. W on how to go about solving it or something. It really just provides a lot of good examples in each coding language, so I think the projects are very good.

Rio also found Examples Book an excellent resource. She has used Examples Book when she gets stuck on coding. She also gives an example that if a student wants to apply a loop

function and does not remember how to do it. They can refer the Examples Book and see how it can be applied.

Uh, but they also have something called the Examples Book, which provides a lot of information on various functions within different languages and also provides like examples. For a reference in case like you don't know how for loop works or how you put a nested for loop. For instance, you can look at the examples book and you can like kind of see how it was applied to a different data set and in different question and you can try to translate that. Also Dr. W makes a lot of videos walking through the projects and kind of explaining them. So, if you're feeling really stuck on you can watch that and kind of understand how he's doing it in the approach that he's taking, I'd say that they focus more on like the projects focus more on applying your knowledge and maybe your newly learned knowledge rather than finishing up and like getting a good grade on them.

In summary, we can say that students identified participating in events in The Data Mine was very beneficial. Theme 1 demonstrated that the various resources that students acknowledged contributed to developing their data science skills. The human resources such as the faculty, staff, TAs, and graduate students played a vital role in helping students develop data science skills. It was also noticed that students found Examples Book valuable and Piazza an excellent platform to interact with other peers and ask any questions they had. Piazza was also used by the faculty and staff at The Data Mine to post any updates. Participation in the events such as outside events and seminar course was mandatory for all students who joined The Data Mine. Students found participating in seminar course an excellent opportunity to develop the foundational data science knowledge. Seminar course engaged the students in an active learning environment that allowed them to interact with peers, ask any questions to the faculty and TAs and learn data science skills. On the other hand, outside events allowed students to attend talks delivered by experts from industry and academia. During the outside events, students got to learn about the application of data science in various disciplines and career opportunities in data science. Overall, The Data Mine provided ample opportunities for the students to learn from senior members, peers and participate in events and develop data science skills.

Theme 2: *The Corporate Partners at The Data Mine provided the real-world corporate experience, excellent mentorship, and was a unique opportunity for students.*

Table 14 represents the codes for benefits of participating the corporate partnerships. Students compared the corporate partner project experience to working in a real world set up. They also found the mentors very supportive and knowledgeable to answer their questions and easy to talk to. Some students also mentioned that being the part of corporate partnership was one of the best experiences for them at Purdue.

Table 14: Codes for Theme 2

Codes	Students	Frequency
Corporate Partners provided real world experience	14	31
Corporate Partners Mentors are knowledgeable and supportive	12	26
Corporate Partnerships is best experience at Purdue	12	12

Theme 2 demonstrates that students found the corporate partners project very similar to one that they would get in a real-world setting. They also loved the mentors as they were knowledgeable and supportive to students. The experience with corporate partnerships was a unique experience for the Purdue Students as they worked on real-world problems. The paragraphs below discuss each code that led to the creation of theme 2.

1. **Corporate Partners provided real world experience:** Students described their experiences with the corporate partners projects and found them similar to real world corporate experiences. Frankie described her corporate partnership experiences as real-world industry experience. Also, she found the corporate project a great experience as she had the opportunity to work on real data set.

   I joined the corporate partners cohort of The Data Mine because I wanted to work with the company since I didn't have an internship my first summer, so I kind of wanted some company real world experience. Um corporate project was a great experience we worked on real world data set.

   Harley describes her corporate project allowed her to develop algorithms for the corporate partner company and she implemented the code in a real-world setting and that was indeed a valuable experience for Harley.

   By joining the Data Mine and working with Corporate Partner company, I actually got to like work on the actual like algorithms and see how the process works

107

internally of trying to fix them. Try and identify bias and sort of data mining. So yeah, that was really cool.

Morgan describes corporate partners as a unique experience, as she finds it a blend of professional and research experience. She mentions that a corporate partnership is an excellent opportunity to learn and work in a real-world setting.

In Corporate Partnership you're learning while also working at the same time, but you have an expectation to be professional and engage in the project in a corporate style. It's a really unique experience, so you kind of learn a lot about what it's like to be a professional while also getting educational experience and learning what it means to be a researcher. So, it is a real-world experience.

2. **Corporate Partners Mentors are knowledgeable and supportive:** Students described the corporate partner mentors as knowledgeable, helpful, supportive, and understanding. Skylar described her corporate partner mentor as a wonderful and very friendly person. Skylar mentioned that her mentor cared for her students, and she was very knowledgeable at the same time. Interacting with corporate partner mentor helped Skylar to develop multiple research skills.

My corporate partner mentor is great. She is great mentor and friend. She's very responsive and she does a lot of things for her students. Not because it's like out of necessity she does it anyways because she cares a lot about her students. I learned a lot from her about writing a research paper and conducting research and using like statistical modeling inside of research in order to gain aggregate results.

Marion mentioned that corporate partner mentor taught him professionalism and team management techniques. He was amazed to see how the mentor emulates company values and that was something he felt is important.

My corporate partner mentor is someone that from [company name] had taught me a lot about what it meant to be a professional and how to lead a team and what it means to. I mean [company name] is a very traditional, a very conservative and strict pharmaceutical company and our corporate partners emulate that, and they teach us what it means to be a professional in a very large and powerful company. So that kind of rubbed off on me. And I'm certainly valuable or I'm certainly grateful for those valuable lessons.

Orion described the corporate partner mentor as very supportive, as she taught him team management, project management skills. Also always motivated him to move forward, she also shared some of her contacts whom Orion can reach out to discuss his interest.

She [Corporate Partner Mentor] was again very supportive. So, it was less about the technical side of things and more about working with the corporate partner company, like she's the one who taught me how to organize and manage the team. I like you know, maybe dividing the tasks and there were some students who were feeling overwhelmed. And so, I like, ask, uh, she provided me advice on how I can tackle this issue, how to make and assign the tasks in a manner that everyone can contribute, and so on. So, she definitely helped me develop the project management skills and also, she has always been helping me in moving forward. She has always been very helpful in networking. She would also give me contacts about of people in the company who would who has worked in the fields that I'm interested in, and I can ask them about their experience.

Robbie described her corporate partner mentors were very supportive and helpful. They helped students to identify their interest and provided all the resources required by the students to work on the projects. Also, they provided help related to any questions that students had.

They gave us the projects that we were going to work on and then they were just like figure out which projects we guys are interested in. What we want to do, going beyond with that. And they're also there as like a help and support. So, like if we have any questions about the project or any questions about the sort of tools that we're using because we're using tools that they actually use as well, so just making sure that get any help that we need for the project.

3. **Corporate Partnerships is best experience at Purdue:** The students mentioned that participating in Corporate Partnerships was the most amazing and a valuable experience at Purdue. Riley is a non-STEM major, and she mentioned that she was overwhelmed a bit when she joined the Corporate Partnerships as she did not have any previous background of coding. But she learnt a lot through her corporate partner projects and considers it one as one of the amazing experiences that she had at Purdue, and she also thinks that she will not get such experience anywhere else.

I was just overwhelmed with a little at first with all like the new things that was learning. I didn't know GitHub; I didn't know coding. It was like a lot to learn it

first and it like was like undergraduate research which I have never even attempted. So yeah, it was overwhelming, but I think I adapted to it well and it was overall an amazing experience. I don't think I honestly would have got it anywhere else here at Purdue.

Quinn described his experience with the Corporate Partners as a cool and real-life experience. He calls it a unique and valuable opportunity as if he was working in a real-world company.

I think that the corporate partners specifically are cool because you get to work with a real-life company. Uhm, feel like that's a pretty unique opportunity that I got at Purdue. And I think that the uniqueness of getting to work with a company like one on one and with the team is valuable.

In summary, theme 2 mentions the benefits that students experienced due to their participation in the corporate partner cohort of The Data Mine. Students found that working with corporate partners was like working in a real-world corporate company. Students were using a real-world data set and were involved in solving actual problems. Students have described their corporate partner mentors as helpful, supportive, and knowledgeable. The corporate mentors constantly guided the student teams at The Data Mine, and they also provided all the required resources to the student teams for completing their projects. Students found their experience working with the corporate partner company, unique, as they worked on real-world data set with a real-world company.

Theme 3: *The Data Mine provided opportunities for internship, research and developing networking skills.*

The Table 15 represent the codes for theme 3. From the Table 15 we can infer that more than half of the students reported that The Data Mine provided the opportunities to network and build connections, internship options and opportunities to work on research projects.

Table 15: Codes for Theme 3

Codes	Students	Frequency
Networking with mentors/speakers/other professionals	15	18
Internship Opportunities	14	17
Research Opportunity	14	14

Theme 3 demonstrates that students in The Data Mine were provided multiple options for internships and opportunities to master professional and research skills. The codes that led to creation of theme 3 are discussed below.

1. **Networking with mentors/speakers/ other professionals:** Students have mentioned that they had opportunities to meet and network with mentors, speakers, and other professionals at The Data Mine. Jamie described her experience as a vice president of The Data Mine Advisory Board (DMAB). So as a part of DMAB Jamie organized outside events in The Data Mine that gave her opportunities to talk to the speakers before the event and have lunch with them. So, this helped Jamie to network with the speakers better and learn about the field of data science.

   As a part of Data Mine Advisory Board, we organized outside events. We got opportunity to talk and network with speakers. So, we just meet up with the person like 20 minutes before, set up the projector and all those equipment that they needed me to speak with them beforehand. You know, get them like ice breaking session sort of and then you know filled any questions that students asked during or after the speaking event. We used to talk to them, and we had like dinners or lunches with them and ask any questions related we had in our mind.

   Morgan talks about the networking opportunities in The Data Mine. She mentioned that, as a student of The Data Mine, she was able to talk to multiple speakers who came to The Data Mine and learn about their life and get a holistic perspective about the data science application in the real world. Those meetings helped Morgan to establish connections and network.

   As a Data Mine student, we kind of go to events talked and interacted with different speakers, listen to different perspectives on things that aren't necessarily coding. So, I think it was really helpful to go to lectures or go to events held by professionals or corporate partners and talk to them about what their day-to-day life was like or

111

talk about the issues in the industry. Or just get a more holistic perspective on the work we're doing and build some connections.

2. **Internship Opportunities:** Students described the internship opportunities that they got through The Data Mine. Marion mentions about the incident where he landed into an internship because of his corporate partner mentor. He talked about and incident where he felt that he was ahead of coursework, and he can go ahead and do an internship or a Co-op. Since he was a corporate partner TA and The Data Mine had already provided him networking opportunities, he contacted the corporate mentor and asked him if the company has any Co-op or internship opportunity. The rapport with the corporate mentor that he established through The Data Mine helped him to get a co-op.

I mean, pretty much kind of what it was I was going through my academic studies, and I realized that I was ahead on my coursework. I realized that I had extra time within my schedule to still graduate on time, but also have an opportunity to search for a Co-op during this past spring semester of 2021. Uhm, and so I thought that it would be extremely valuable for me to do, and so I had reached out to my mentor at Corporate Partner company as he knew my skills. I mean, I said hey, is there anything that you know about and then, how can I get involved? And through those processes, I got an interview and I had met one of my mentor's peers and we had worked on a project together for six months. And all this happened just because I was with The Data Mine, working as a Corporate Partners TA for that company.

Quinn mentioned about his internship Dr. W helped him to get. He mentioned an incident where Dr. W had sent him an email about an internship, and he applied to it and got that. He is thankful to Dr. W for letting him know about the internship opportunity.

My internship this past summer was the application for it was sent out by Dr. W, so like I just kind of progressively moved up. I worked with a company called [Research institute's name]. They are a Research Institute based in Indianapolis. It was a virtual internship. Uhm, and The Data Mine helped me get the internship.

Riley's internship experience was different. She discussed how Ms. M helped Riley to get the internship with big giant company and it was a very exciting experience for Riley.

So, this summer, Ms. M asked me if I was involved or if I was going to go have an internship or anything like that this summer and I when I told her no, she said I think we have an opportunity for you. Internship was all remote and it's a big, big

pretty big program. During this summer and I was part of the data engineering team for project we were doing, which was extremely exciting.

3. **Research Opportunity:** Students mentioned about the research opportunities that they got through The Data Mine. This was first encounter with research for Skylar. The Data Mine helped her to find an interesting research opportunity. She is enjoying her research experience and finds it very rewarding.

Because of Data Mine I found a research opportunity and now the Professor I work with is probably like one of the closest faculty members to me and then she asked me to assist her class, which happened to be one of the corporate partner programs. This year inside data mine, so now I'm a TA for the corporate partners program as well.

Finley also mentions that he too got a research opportunity in Fall of 2020. It is important to note that Finley worked from his home country for the Fall of 2020. He found research experience very valuable as he learnt statistical modeling. He considers that the research experience will be very helpful for him in future, and it makes his resume much more impactful.

In the fall 2020 semester I worked with Dr. W and another professor. Uh, the research was with the statistics department and the College of Nursing on a research study that that analyze the general and mental health of Purdue students during the COVID pandemic using these two questionnaires. I was a team leader of the Data Analytics team so some of my responsibilities included, migration of data, cleaning data using R, imputing data using R, running linear regressions, multiple regressions, analyzing data or things like that, I think that was a great addition to my resume.

In summary, theme 3 demonstrated students' opportunities at The Data Mine to develop themselves professionally. Being in The Data Mine, students found multiple opportunities to develop networking skills, research skills, and came across various internship options. Students developed networking skills by interacting with speakers during outside events and by interacting with their corporate partner mentors. These interactions with speakers and mentors helped students learn about various career opportunities in data science and helped them get interviews for internships and co-ops. Students have also mentioned that faculty and staff at The Data Mine posted various internship opportunities on Piazza or reached out to students directly. Lastly,

students have mentioned they developed research skills by participating in corporate partner projects and presenting their work in a research symposium. Also, by engaging with faculties on various research projects.

7.1.2 Themes for Challenges in The Data Mine

The student perception of the challenges in The Data Mine is divided into two broad themes. Each theme is mentioned below and with their relevant codes and representative student quotes. It is important to note that the proportion of the challenges are much less than the benefits reported by the students. The subsequent paragraphs describe each theme for challenge.

Theme 4: *The corporate partners presented challenges related to delay in data sharing, changing the project scope, and mentors being busy.*

The Table 16 represent the code for the challenges experienced by the students with corporate partners. Even though students mostly appreciated the Corporate Partners they did experience some challenges.

Table 16: Codes for Theme 4

Codes	Students	Frequency
Corporate Partners did not provide data on time	3	3
Corporate Partners Project changed suddenly	1	1
Corporate Partners were Busy	1	1

Theme 4 demonstrates the challenges presented by corporate partners and the frequency of codes demonstrate that the intensity of challenge is much less than the benefits that students experienced in The Data Mine. The codes for challenge that led to formation of Theme 4 is explained in detail with representative student quotes in the subsequent paragraphs.

1. **Corporate Partners did not provide data on time:** The students mentioned about the instances where they did not receive the data from corporate partners on time and there were delays in overall project. Mickey describes her experience with corporate partners as boring as their team was constantly writing code, but they did not get the actual data from the corporate partners in the Fall semester. She mentioned that the idea was to allow

114

students to relax in Fall and do actual work in Spring. Mickey did not like the approach, so she dropped out of that corporate partner project.

We had only been working on developing like the codes for that we didn't really do any work with actual data sets. Uh, which is part of why I left because I just got kind of bored. I don't really know, to be honest, but I think there was a lot of this idea that in fall you can just goof off and then spring will start actually working and so they gave us a lot of busy work for fall and I was only there for fall. I dropped it after that I was like there's no point, but I think they were just sort of waiting for spring to give us work.

Jamie had a good experience with corporate partners but there were some issues initially as the Corporate Partner company acquired another company and due to some legal paperwork, the corporate partner could not share the data on time.

Um, first time it wasn't as we didn't get the data until like November due to some legal issues because the corporate partner company was just bought another company and there were some delays, um, so we didn't get the data and that delayed the overall project.

Frankie also had a good experience with corporate partners, but she also reported that there was a delay in getting the data from corporate partners. Corporate Partners could not share the data till October so the team could not get started with the project.

I believe we didn't get the data that we were supposed to work with until September or something. So, we joined this in August and we're taking it for class credit, and we didn't really get started on the project until maybe September or October, just because everything was very slow starting out, so there were just little things.

2. **Corporate Partners Project changed suddenly:** In one case student reported that the corporate partner project changed in its scope without consulting the student team. Kendall was the only student who discussed about the sudden change in the project. Although Kendall is very much happy about being the part of corporate partners team, but she discussed how the change in corporate partner project affected the team's motivation.

It was kind of difficult. We came back from the Christmas break our project has changed so much it was almost like an entirely new project and I think at that point

a lot of people kind of lost motivation so towards the end it was difficult to kind of pull everyone together and work together on it.

3. **Corporate Partners Mentors Busy:** One student mentioned that corporate partner mentors were busy and did have time to guide the students on the project. Arrow is the only student who mentioned that the corporate partner mentor was busy, even though the project was very interesting he could not work on the project due lack of guidance.

In my sophomore year, I worked with a different corporate partner. The project was very very interesting. But I don't think it panned out well. Because the lead and mentor was a very busy guy and eventually it didn't work out.

In summary, theme 4 identified challenges related to the corporate partners' cohort of The Data Mine. Students reported three types of challenges. The first one is related to delays in data sharing. Three students reported this challenge, and they mentioned that they received data from the corporate partners very late, which led to a delay in the overall project and, in some cases, student drop-off from projects. The second type of challenge was related to a change in project scope, which just one student reported. The student mentioned that the corporate partner had changed the entire project after the student team came back in the spring semester, which led to demotivation in the team. The third problem was related to corporate mentors being busy. In one case, the mentor was a very busy person. He could not devote time and guide the student team, which led to some dissatisfaction.

Theme 5: *The non-STEM students initially felt intimidated and sometimes they feel less confident when stuck on a data science problem.*

The Table 17 represent the codes demonstrating the challenges experienced by the non-stem background students. They felt intimated initially as they did not have prior coding background. They also feel less confident if they get stuck on data science problems and they are unable to figure out the solution themselves. Theme 5 specifically captures the challenges of the non-STEM students. The codes that led to formation of theme 5 are explained in detail with representative student quotes in the subsequent paragraphs.

Table 17: Codes for Theme 5

Codes	Students	Frequency
Felt intimidated Initially	3	5
Feel less confident when get stuck while solving a data science problem	3	3

1. **Felt Intimidated initially:** All three non-STEM background students mentioned that they felt intimidated after joining The Data Mine. Harley mentioned that she felt intimidated initially as she did not have a strong coding background and The Data Mine was all about coding and she found the projects a bit high level, so she felt bit intimidated.

> Yeah, when I joined the Data Mine, I was honestly a I'm not gonna lie. I was pretty intimidated because only prior joining Data Mine I had very little bit of experience with R and I noticed that we also needed Python which I had never touched and the projects we were given it seems pretty high level for a college, almost specially for me being an outsider. So, I was really intimidated initially.

Riley also experienced imposter syndrome initially as she also did not have prior coding background.

> Initially I had a big impostor syndrome where like I don't know how to code and what am I doing like.

Skylar on the same line felt that most of the students in The Data Mine was from data science or data analytics background. So, she felt a pressure working with them and felt little intimidated.

> Because a lot of people in the cohort are also data science or analytics majors. So sometimes you feel this sort of tension not being the same major. Uh oh, I need to. I feel like I need to catch up. And so, it's a lot of like self-learning that you have to do in order to make yourself feel like you know, not make it, not feel like you're in impostor syndrome, or feel like you don't really fit well.

2. **Feel less confident when get stuck while solving a data science problem:** The non-STEM major students mentioned that they get stuck sometimes while solving a data science problem and their confidence level goes down if they were not able to figure out the next

steps or trouble shoot the issue. Harley mentioned about the situations when she sometimes got stuck on code or code did not work for her but worked for other students. She used to find it difficult to troubleshoot the problem.

> I get stuck sometimes. Sometimes code didn't work for me. I did like exactly what they [other coders] did and I completely modified it to my script and it's like still crashing. Uh and always, I mean, as I was just so lost.

In summary, theme 5 captures the experiences of the three non-STEM students. All three students reported two significant challenges. The first one was that the non-STEM students felt intimated when they initially joined The Data Mine as they had never coded before. They also felt that most students in The Data Mine were from STEM backgrounds and had prior coding skills. The second challenge that students mentioned was feeling lost and losing confidence when unable to solve a given data science problem. All three students felt lost and less confident when they were stuck while solving a data science problem or unable to troubleshoot the errors.

7.2 Student perceptions of socialization and sense of belonging at The Data Mine

This section highlights the themes for socialization and sense of belonging. The section intends to focus on the opportunities of interaction, collaboration, and socialization that The Data Mine provided to the students, as well as the extent to which students felt comfortable in interacting with peers, faculty, and staff at The Data Mine. Therefore, in this section we plan to answer the following research question **Sub-RQ 2 (b):** How do students describe their levels of sense of belonging and opportunities for socialization within The Data Mine? Based on our analysis, the themes for socialization and sense of belonging can be classified into two broad categories they are, learning through socialization and experiencing the sense of belonging. Themes 6 focuses on the learning through socialization aspect of The Data Mine, and theme 7 focuses on how students felt accepted and experienced a sense of belonging in The Data Mine.

7.2.1 Themes for Sense of Belonging and Socialization
Theme 6: *Interacting with diverse peer groups, working on corporate project, meeting like-minded people allowed student to collaborate, socialize and learn from each other.*

The Table 18 represent the codes that demonstrates learning through collaboration and socialization opportunities in The Data Mine. The students in The Data Mine found various opportunities to learn together with their peers such as living in Hillenbrand Hall and working together on problems, working in interdisciplinary groups on corporate partner projects, and peers providing multiple perspectives to solve the same problem.

Table 18: Codes for Theme 6

Code	Students	Frequency
Worked as group on problems and learnt from one another	19	70
Got multiple perspective to solve same problem through interaction with students from various backgrounds	19	35
Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor	15	24
Socialized with my peers and team	13	25
Lived in Hillenbrand and met people with similar interest	13	35

Theme 6 and Table 18 represent the multiple opportunities that students got in The Data Mine to interact and collaborate with peers and learn from one another. Each code that led to creation of theme 6 is explained in detail with corresponding student quote.

1. **Worked as group on problems and learnt from one another:** The students mentioned about their interaction with peers while working on a problem and how that led to learning from one another. Marion mentions that he has learnt a lot by interacting with his peers at The Data Mine. He has made great friends and had got chance to learn from some of the brightest minds. He gave an example, that he learnt front end development from one of his peers as he did not know how to code in React Native (a front-end development software) but since that was a requirement of the project, a senior in the team guided him.

Uhm, I mean it. It's so huge. I mean, a lot of the learnings that I think I've had come from discussions that I have with my friends with the people around me with other Data Mine students. Also, I was learning from really, really, really bright students around me. The Data Mine has so many great students that that I'm able to kind of pick their brain and learn from and so. Uh, I just try to be a sponge and be available for learning opportunities, I guess. I think one thing that comes right off the bat, at least in like a technical sense, is some front-end development stuff in React native. I mean, that it was my sophomore year. I had never really touched any front-end development. Uh, software is and so. It was I was working with a senior on the

119

team, and he had kind of given me an opportunity to learn under him a little bit and so from a technical sense he gave me foundational learning in into React native.

Reef refers working in group as a process. He said that he never had an experience of working with group for such a long period, generally, the group experiences in class projects are just for a semester. So, working in group for such a long period in The Data Mine helped him to learn new concepts, identify different approaches to solve the similar problems and made him realize that his approach is not always the best approach to the problem.

Ah. It's [group work] definitely a process, I think, um because we're like so many different personalities were so many different backgrounds. And it's interesting in like what your interest is, isn't the same as someone else is, and how you approach something isn't the same as someone else. So that team working experience is something that I hadn't experienced a lot in college. I mean especially for like such, an extended duration of time, and so that was that was a good experience. But it was definitely like a learning experience as well. For me, like being in a group helped me to grow. Help me to realize my way is definitely not always the best way, and so just like confirming that. And yeah, having that experience was great.

2. **Got multiple perspective to solve same problem through interaction with students from various backgrounds:** Students have mentioned that working with peers in group projects, seminar problems or corporate projects that helped them identify new methods to approach the problem. Jackie found it unique to work with students from various backgrounds. He said that working with different students helped him to understand the problems from various lenses. It was wonderful to know multiple methods to solve the problem. He further added since his corporate project was research based there was no single answer to the question. Therefore, looking at the problem from various perspectives and seeking inputs from students of different background is very helpful.

There was the cool thing was there was a lot of different backgrounds. Well, but I mean it was all different. We had a couple of mechanical. We had aero, astro engineers, we had computer science majors, so it's a lot of different backgrounds, from an academic perspective. Uh, and I think that that helps to come up with unique solutions, especially for the project. It was more like research based, so I think that actually worked better to have a more diverse team because you know everyone is looking at the problem through their own lens. Uh, and when you're doing research, you're really just trying to find the best solution, right? There's not

necessarily one correct answer. It's just the best you can find. Uh, and that was that worked really well with having a different, diverse backgrounds, so we had lots of different academic backgrounds on the team.

Harley is a non-STEM major, and she faced some problems when it came to coding. But she found that her challenges were not unique many students even from STEM background faced similar issues. She also mentioned that students during the seminar class would reach out to Dr. W and ask similar questions and Harley would learn a new method to approach the problem.

Yeah, exactly, and it's like since it's the very beginning. I also I noticed that some of the problems I was facing was also faced by a lot of other people, so it wasn't like I always had to personally interact with him [Dr. W] to get it solved because someone else would get to him and ask him how to do this and I could just like bounce from there.

Morgan called it fun to work with peers on seminar projects. She found it exciting to see how her peers approached the problem, what functions they used and how she solved the problem. Also, it was always interesting for her to learn new approaches to solve the problems.

I think it's a ton of fun. It's always fun to collaborate with your peers on project and just to see how they would approach it versus how I would approach it. I think it's a good time. Definitely, it's very interesting to see if there are different functions that they would use, or maybe just different methods or ways of approaching it. I think you learn a lot by working with other people.

3. **Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor**: Students described how they worked in interdisciplinary team for corporate projects. The teams were comprised of students from different majors or different level of education such as freshman, sophomore, junior or senior. Orion described his experience of working with an interdisciplinary team. His team was comprised of students from all levels of undergraduate education and included two graduate students. He described his project as highly technical and the role of the senior students such as graduate students and junior and senior undergraduate students was critical, as they had most of the technical skills.

Freshman students were new to programming, so the team allocated them simpler tasks to execute and helped them to learn complex tasks.

Our team had three or four juniors and two seniors and two grad students. Many of them were freshmen or sophomores. So, since this was such a big task, there were freshmen and sophomores, they were completely new to programming. For the project we need students having experience in creating mobile applications, creating APIs, create dashboard and they don't have any experience with that. So luckily since there was varying experience, we had to have the senior members responsible for more of the more technically challenging tasks and we provided the smaller tasks for the freshmen and sophomores and also gave them opportunity to learn from the senior member.

Morgan describes her team as diverse. They had students from various levels, but she also describes how students were distributed geographically. With respect to major, students were from computer science and data science backgrounds, but there were also students from economics, finance, and hospitality management. She also appreciated the help that they received from corporate partners and TA, they acknowledged the fact that students were from various backgrounds and varied competencies. Morgan also describes how tasks was distributed among the students based on their competencies. The technical work was assigned to students from technical backgrounds, whereas qualitative work was assigned to students from non-stem and non-coding backgrounds.

Yeah, so last year or team was split up pretty evenly between undergraduates and graduate students. I think we had about five undergraduate students, ranging from freshman to juniors, I believe, and then we also had master's students and PhD students as well. In terms of demographic, I think it was pretty spread out. We are all from different places. I know I was in Iowa. Team members were in Colorado, California. There were also people on campus. So, some traveling through the year. Uh, and then I think most of us came in with a CS or data science major. Uhm, but there were also people studying economics, finance, hospitality management is just a really diverse set of people, I think. TA and mentors were really understanding of the fact that we didn't all start in the same place and so they helped us kind of manage our expectations and figure out how we could come to the same place as a team. Uhm, I will say by the end of the first semester we were pretty much all in the same place and a lot of it's difficult to say because our project was not necessarily 100% about coding. It was a lot of qualitative factors that we were looking at two and so having the diversity in majors, especially having the non-stem majors helped to contribute a lot to our understanding of the project, from a non-coding background as well.

4. **Lived in Hillenbrand and met people with similar interest:** Since prior to COVID-19 living in Hillenbrand was a requirement for being in The Data Mine so many students lived in the Hillenbrand Hall and even many students live there currently as well. Students shared their experiences of living and learning in Hillenbrand Hall. Jamie lived in Hillenbrand Hall, and she described how she would just grab a meal, sit in the lobby, and meet other students. Every Friday she used to meet her friends in the lobby of Hillenbrand, discuss ideas, solve problems, or discuss the logic of solving the problems. Also, she says it was a great experience living in Hillenbrand Hall, as every student in The Data Mine lived next to each other so if she had any question, she could reach out to students living next door.

We all live in Hillenbrand, So I would just grab a meal, sit and then like kind of meet with some people if I saw them. You know, we can just work on the project and then someone would help me like walking around and then you know that was like the environment. Every Friday we'd get together in the sub lobbies of each floor. Uhm, we just dealing bounce off ideas. Uh, on how to work on the project or just even corporate partners. We could not talk about the specific data because of the confidentiality agreement, but we were just kind of sitting like work through the logic of it because we can discuss that. So that was really interesting and how we could just kind of, you know, knock on a next door and like you could find someone who would be working on the same problem as you.

Orion found it fascinating that a majority of students in The Data Mine lived in Hillenbrand and students belonged to several majors but shared the common interest in data science. Therefore, Orion found it easy to start a conversation with any students and make friends. Also, all students lived in Hillenbrand Hall, attended the seminar class, and were able to grab lunch and dinner together. That provided them multiple opportunities to interact and create a close bond.

So, it worked out was since everyone was interested in data science, their career goals were also very similar. So, they would be CS, DS majors and also the non-STEM majors also were very interested in this field, so they were very tech savvy as well so. Uhm, everyone was, I would say like we of course we related on other ways as well. Like we, just it gives us an initial conversation starter. I would say to start off with. OK, cool we're both interested like a career goal or similar like you're interested in tech, you know interested in mobile applications? Interested in what's the newest software available? So that's how I would say it started off the conversation and then of course once you build a certain connection. Then you can always take it from there. You become closer in the sense you could relate on other

things as well, like sports, and also, we got lunch and dinner together in the seminar. It used to happen that it was in the dining courts of Hillenbrand, so we used to eat together as well. So, in a way, since we used to see each other more often, we would also get comfortable and get to know each other.

5. **Socialized with my peers and team:** The students described about various socialization opportunities they had in the Hillenbrand Hall, or with their teams, friends, and peers in The Data Mine. Robbie says that The Data Mine is a built-in community where one can easily find a person to talk to, study with and be a friend. Since she lived in the Hillenbrand and there were many events that used to take place on the floor or throughout the residence hall, which allowed students to communicate and socialize.

So, I think the fact that there's already like a built-in community, and I think that really helps when you're trying to find friends or when you're trying to find like study groups or just people to talk to. We also do like floor events, and there's Phoenix Club, which they also do like dorm room wide events. So, I think that having those built-in sorts of systems for events in communication and help out and socialization.

Marion also mentioned many social events that he participated in, such as ice-cream socials. Since he is a corporate partner TA, he also organizes events for his team to socialize.

There were social events in The Data Mine, like ice cream nights or things like that or maybe just kind of meeting up with friends from down the hall. That was certainly a lot more frequent. Now the social aspect kind of comes just as a TA like organizing team events and things.

Mickey mentioned about the virtual events conducted by The Data Mine Advisory Board. She participated in those events last year and she enjoyed those events. Also, she is the part of The Data Mine Advisory Board this year, so they were planning to host in-person events in the fall of 2021.

So, The Data Mine Advisory Board did a lot of social events, which I thought was really fun. They did in person as well previously, but obviously last semesters it was virtual because of COVID. But this semester they were planning a lot of in person, but they did like trivia nights and game nights and like things like that essentially, and like also study sessions and group events. So, it's partially social,

partially academic. I went to a lot of the social ones like the trivia night in the game night. I really enjoyed it.

In summary, theme 6 focused on learning through socialization. It was important to note that learning in The Data Mine happened through socialization. Students got various opportunities to interact with peers in The Data Mine. They worked on corporate projects, lived in Hillenbrand Hall, met in the seminar class, participated in social activities, etc. Every interaction led to some form of learning. Corporate projects were excellent opportunities to learn from peers of different discipline and get new perspectives to solve problems. The seminar course allowed the students to learn data science skills in an active learning environment. Living in Hillenbrand Hall helped the students' meet peers of similar interest, work together on projects, and bounce ideas. It was important to note that the learning at The Data Mine was integrated into the context of the learning community and the activities allowed students to learn from one another and grow together.

Theme 7: *Students felt a sense of belonging and acceptance in The Data Mine as they established connections with faculty, staff and peers.*

The Table 19 represents the codes that demonstrates the inclusive nature of The Data Mine. From the Table 19 it can be inferred that The Data Mine fosters a judgment free environment where the students, staff and faculty feel comfortable to interact and built connections.

Table 19: Codes for Theme 7

Codes	Students	Frequency
I feel included	15	33
Felt a sense of community or belonging	15	30
Faculty and Staff made me feel connected	15	27
Felt connected with peers and Made Friends	15	25

Theme 7 and Table 19 represents the ways that The Data Mine used to foster a conducive and judgement free environment. It provided an opportunity to every member to express themselves and establish new connections. Each code that led to the creation of theme 7 is explained in detail with a corresponding student quote.

1. **I feel included:** The students mentioned that they feel accepted and included. Also, students mentioned that staff, faculty, or peers made them feel included and accepted in The Data Mine. Robbie found The Data Mine very inclusive because she found it fun to live in Hillenbrand Hall with other Data Mine students. She also met Dr. W and Ms. E, regularly, since they had an open-door policy, students used to drop by their office to greet them and interacted with them for not just academic reasons but even on a personal basis.

   Data Mine is very inclusive. We all got together or same floor. You know, I met Dr. W on Monday, Tuesday. Sometimes he'd be in the office at 10:00 PM. It is very easy to socialize with anyone in Data Mine. Just on the way to class I use to drop by Ms. E's office and her door was always open. I just talked to her about stuff not always you know related to Data Mine. It's pretty good.

   Mickey was initially very closed off, and she did not interact with students initially when she joined Purdue, but after being in The Data Mine, she has started collaborating with other students, reaching them out for any kind of questions, and has also learnt to communicate professionally with others.

   So, I think from me like initially I focused a lot on like group efforts and just on like very very closed off circles like I would only message that TAs or only message Dr. W and ask for help. But now I think I'm a little more comfortable with everyone. So, like I can even message like the people I know who are in the class and be like hey like did you understand how to do this? I feel more comfortable even just going to a few of the TA that I know, and I think in general I've gone from just asking questions to also just being more socially involved as well. So, I was really, really sheltered during high school. I was very closed off and in my own space, Data Mine I think was a big factor in learning to communicate with other people professionally and I think that really was beneficial to me.

   Orion felt that faculties and staff at the Data Mine have made him feel included because they care about the students, they are very supportive, and they provide a conducive environment to students to learn and excel.

   I feel I do feel very welcome in the Data Mine, like Ms. M, Dr. W, uh, Ms. H, everyone is very welcoming. They've been very supportive. trying to understand our problems, helping us to like to work with team members, but also helping us to learn more as well.

Kendall feels that she is actively included in The Data Mine. She has lived in Hillenbrand Hall. She made many friends and they worked on problem and socialized together. She also feels that Dr. W and Ms. E have always provided her opportunities to excel. Therefore, she feels a sense of belonging at The Data Mine.

I would say, uh, I feel like, I'm actively included which I think is important. I made friends freshman year through The Data Mine classes through being on the same floor etc. And to this day we still interact with each other and actively include each other in different things, so that is one way I would say I feel like a sense of belonging in The Data Mine. And then another way that I think it shows up is in the fact that Dr. W and Ms. E, always also actively include me and lookout for opportunities for me and make sure that I am not only just doing well as a student, or even just as a person, but that I have a lot of resources at my disposal, so I would say in that way I'm also actively included in The Data Mine. As opposed to just being like OK, I belong or I'm similar to everyone else which is of course it's also true because we're all you know, pursuing the same class work and things like that.

2. **Felt a sense of community or belonging:** Students felt a sense of community or belonging being at The Data Mine, or living in Hillenbrand Hall or while engaging with students, staff and faculty in various activities and events in The Data Mine. Frankie enjoys being in The Data Mine. She found a sense of community and acceptance in the Data Mine. She found the staff and faculty very caring. Living in Hillenbrand Hall allowed her to talk to meet peers of similar interest, work in teams and learn from one another. She also mentioned that she made some great friends at The Data Mine.

Uhm, I really enjoyed being a member of it and I felt like it was kind of a community too. I think, especially because the faculty are very caring, you know, they care and make sure they wanna make sure that you're having a good time and you're learning a lot and meeting people in addition to you. Know when I lived in Hillenbrand in that dorm with that pretty much everybody in The Data Mine lived in my sophomore year that very much felt like a community. Because, you know, I see people who were on my teams or people who were in the dining court seminar. With me I'd see them in the hallways. Or, you know, I'd see them walking. You know two and from Hillenbrand or in the lobby. So, it definitely felt like a community there.And you know, I made friends with the people that were on my team. You know the people who were on my team sophomore year like I still talked to like I just got lunch with one of them the other day. And so, I definitely feel like I made a lot of friends through that, and it very much did have a community aspect.

Marion also felt a sense of community at The Data Mine. He mentioned that there were students from various backgrounds and majors, but everyone shared a common interest in data science and that bonded them together. He made some great friends who did not just bond academically but also shared similar interests, beyond academics.

Ah, I felt the sense of community, pretty much. I mean, we were all we were all pretty academically focused, for sure. Uhm, uh, two of my roommates were data science majors and two of us were statistics majors. A lot of the students on the hall were data science, statistics, math. You will see students that are from all over campus that are interested in data science and analysis. So, I mean there's students from Polytechnic, from Krannert, from uh College of liberal arts. There are just so many students that are from all over. So, I would say for us it was just like our academic or academic goals were pretty aligned or professional goals were pretty aligned. And also, we just became good friends because of our common interests in data science, sports and games. Uh, I don't know. Yeah, we just we just kind of fit very well beyond just academics.

3. **Faculty and Staff made me feel connected:** Students mentioned that they felt faculty and staff were helpful, supportive, friendly, and made them feel connected to The Data Mine. Kendall described emphasized that Dr. W and Ms. E were very supportive and caring. She mentioned Dr. W was her faculty fellow and she use to meet him every day. She also mentioned that Ms. E made the students feel connected by remembering names of 800 students and made herself available for all weekly meetings. Kendall specifically mentioned that Ms. E sent her a e-card when Kendall was not well and, she keeps in touch with students even though she is not on the staff of The Data Mine anymore.

In terms of faculty Dr. W, definitely was somebody I saw all the time, particularly because he was my faculty fellow so he would come to all of our floor events. And also Ms. E, she was always there at the weekly meetings, and she always tried super hard to remember all 800 of our names and then. For one of my outside events that I wrote, she reached out to me, and she said she kind of enjoyed the way I've wrote it. I guess it's very genuine in my response and we had been following up with each other ever since then. Uh, and I had gotten uh, kind of sick. she would always check up on me and make sure I was okay, and she sent me like a sweet little E-card when I was not doing well. So she definitely looks out for all of her students and keeps in touch even though she is no longer in The Data Mine right now.

Robbie described her experience with Dr. W and Ms. M. She mentioned that Dr. W identified new opportunities for Robbie in the field of data science. She also found Ms. M

very helpful as she was always available to students with any questions related to corporate partnerships. Robbie found both Dr. W and Ms. M friendly and easy to talk to.

Dr. W, he has just opened up a lot of opportunities for me into like data science fields and data science jobs and then also there's Ms. M, she is the head of managing for corporate partners, so she's also a communication point that we will use for asking questions about what to do for corporate partners or meeting times and what other programs to use for. Like when we go through our weekly meetings and just suggestions about what we should do. Overall, they are very friendly and easy to talk to.

Frankie found Dr. W and Ms. M very helpful, as they always made her felt connected and included in The Data Mine. They always checked with students to see how they were doing and addressed any concerns that they had.

I think also the faculty and staff at Data Mine are really good about you know checking in with their students as well. Like I never felt super far removed from anything, you know Dr W or Ms. M like they always make very clear that they were here for us and everything. So, I think you know I felt included and accepted.

Morgan described all her interactions as great, with faculty and staff at Data Mine. She found them very helpful and kind people. The most important aspect that Morgan mentioned was that the faculty and staff made all students feel connected, irrespective of their background and competencies.

I think all my interactions were great with faculty and staff at Data Mine. They were very kind people, very supportive and understanding of what your background is and they work really hard to make sure that The Data Mine is inclusive of everyone, regardless of major where you come from, what your grade is, they just do a great job in making it a good learning community.

4. **Felt connected with peers and Made Friends:** The students mentioned that their peers in The Data Mine made them feel connected and the interaction helped them to make good friends. They described their bonding with one another and sharing similar interest. Frankie mentioned that she felt connected with her peers, as they met regularly for the project, worked together, experienced similar challenges, and solved them together.

I would say my peers especially made me feel connected. I think I kind of bonded the most with them, especially because we were doing the projects together and we were meeting, you know with each other two or three times a week. So, we saw each other the most. So, I feel like they kind of made me feel the most connected and you know we were with each other in the same boat and everything.

Jackie felt very connected with his Co-TA in corporate partners. He was a wonderful friend to Jackie as they shared similar interests related to academics and sports. Jackie felt good when he discussed his problems related to projects or even personal problems. They bonded very well.

Definitely the other Corporate Partner TA. He's definitely someone I connected with quite a bit. Uh, like I said, super similar interests. We just we just went out golfing, I think two or three weeks ago. So, it's just it's nice to have someone from a different background, but really similar person overall to go to with issues both on the projects because we're both working for Corporate Partner or you know, just about academics or personal decisions for careers. Things like that, yeah, we just love to talk about things like that so that's cool.

Rio had made friends in The Data Mine. They are from the Pharmacy major, but Rio did not know them prior to joining The Data Mine. Rio is a corporate partner TA; therefore she had the opportunity to meet different students and make friends. She loves to interact with them and to answer any question of her team members. Also, in this process of team handling, she feels connected and made friends.

I have a few friends that I made in Data Mine, and I also have a few friends from pharmacy, but also, I have some friends who were in pharmacy but I didn't know them until I interacted with them in The Data Mine. I was the TA for our corporate partner cohort both this year and last year. So, I got to interact with a lot of people on a more like individual, one to one level as well as in a group setting. So, a lot of the ways that I made friends was setting up like office hours and meeting up with other students who are in my team to help them develop their skills and help do the tasks that we were assigned. And this just gave me like an opportunity to get to know them for who they are and talked to them and kind of form a friendship that works in both the work environment and also more of a social environment as well.

In summary, theme 7 illustrates that participating in The Data Mine helped the students establish strong bonds with faculty, staff, and peers. Faculty and staff provided various opportunities to students, such as internship opportunities, networking opportunities, engaging

130

them in research, and demonstrated empathy and emotional support. Students felt very connected to the faculty and staff and felt comfortable sharing their thoughts and ideas and seeking input from them. Students also found their peers very helpful and understanding. They felt the environment of The Data Mine was judgment-free and very acceptable. Therefore, students experienced a sense of community and belonging within The Data Mine.

7.3   Perception of identity formation in The Data Mine

In this section I intend to focus on how students recognized themselves after spending substantial time in The Data Mine. Also, I want to understand in what ways the interaction with members and participating in various events helped the students to develop their identity. Therefore, I plan to answer the third **Sub-RQ 2(c):** How do students' participation and interaction in The Data Mine help them form their identity? The inductive thematic analysis revealed that students are in the process of developing their identities. These identities are being developed through constant interaction with the members of learning community and by participating in various events in The Data Mine. It was observed that all students were developing data science identities, but the students who were Corporate Partners TAs (CRP TAs) in The Data Mine were developing the data science identity but also demonstrated a leadership identity. The CRP TAs manage a team of students and act as a bridge between corporate partners and The Data Mine. Therefore, the result of this section is divided into two sections: 1) Theme for data science identity, 2) Theme for leadership identity. Furthermore, each section has a broad overarching theme followed by sub-themes.

7.3.1   Theme for Data Science Identity

Theme for Data Science identity: *Students were in the process of developing the data science identity, they felt competent in applying their data science skills to solve real-world problems and pursue a data-oriented career in future.*

The paragraphs below demonstrate the sub-themes for the data science identity. The sub-themes are divided into five categories: *1) How did I see myself prior to joining The Data Mine? 2) How do I see myself now? 3) How do others see me? 4) How do I want to see myself in the future? 5) What skill/competencies did I develop by participating in events and interacting with*

131

peers/faculty/staff in The Data Mine? Each category has sub-themes followed by qualitative codes and representative student quotes.

Category 1: How did I see myself prior to joining The Data Mine?

Sub-Theme 1: *Prior to joining The Data Mine, the majority of students recognized themselves as novices, and only a few had prior coding/programming background.*

The Table 20 represents the codes for sub-theme 1. From Table 20, it could be inferred that students' skills prior joining The Data Mine could be categorized into two types, the first one being a novice learner and second type is students with moderate to high level of coding/programming skills.

Table 20: Codes for Sub-Theme 1

Codes	Reference	Frequency
I was a novice learner	14	20
I had prior programming/coding knowledge	5	5

Sub-Theme 1 and Table 20 represent perception of students about their data science skills prior to joining The Data Mine. The two codes that led to the creation of sub-theme 1 are explained in detail with corresponding student quotes. But it is important to note that the majority of students identified themselves as novices in terms of data science.

1. **I was a novice learner:** The majority of students mentioned that they had very limited to no knowledge of data science, statistics, or coding competencies prior to joining The Data Mine. Marion mentions that before joining The Data Mine, he did not have much idea about the data science concepts as he did not have any technical background in high school. He thought that joining The Data Mine will be an exciting opportunity and that will help him learn data science skills.

   My idea of what the field of data science might look like was pretty blank. Um, like I didn't really know exactly what to expect, and I guess like I knew that I was going to be coming in learning some data science stuff I didn't really know what that was, I was not a technical student in high school, I didn't really have much technical background, so before coming in as a freshman I kind of just took it in and I just thought that the Data Mine be an opportunity for me to learn some data science skills.

Skylar also mentioned that she did not have an data science or analytics background before joining The Data Mine. She learned data analysis, data science concepts and coding in R, at The Data Mine.

I completely learned R from like scratch. Uhm, uh, and it was I think was just a great experience for me to start with data science and analytics, specially, since I didn't really have a background before I and I kind of just went in.

Harley mentioned that she had very minimum coding skills before joining The Data Mine. At The Data Mine she learned to code in R and Python. She also developed the skill to look for the right resources, and she felt competent to use her skills to work on outside projects.

I had very minimum programming skills. But after the data mine, I would I actually learned how to do R and I also learned how to fill in my skill gaps in both R and Python. So, I sort of noticed OK. There's like a lot I can do by. One thing I did learn was how to search up the problem and sort of like go on GitHub and look at someone else's code and put that in mine and modify that. And yeah, so I thought that was like really amazing that I could just do that and to this day, I find that really helpful because even after The Data Mine, I sort of work on some personal stats projects. And I like anytime I don't know anything I know how to like get help like go look that up and then come back to my script and sort of implement that in my script.

Jamie had a very limited Python programming skills prior joining The Data Mine and she did not know to code in R, data science skills, or web scraping. She learned all these skills through The Data Mine seminars and working with corporate partners.

Yeah, I only knew Python and that was very rudimentary, but high school Python not a lot, just syntax. And like you know logics and stuff, so I had no data science skills in terms of like learning R, analyzing data. You know like scraping data, none of that. I just knew nothing about it, and this is all like learned through Data Mine seminars or through like my project with corporate partners.

2. **I had some prior programming/coding knowledge:** Five students mentioned that they had moderate to high level of programming or coding knowledge prior to joining The Data Mine. Arrow had prior coding knowledge, he always loved to build algorithms, participating in hackathons. But joining The Data Mine helped him to learn more skills such programming in R.

I just loved building algorithms, that sort of like what I like doing it. Specially like I like participating in hackathons. I knew programming as someone who comes from like a more CS background like I have this bias towards Python development and Data Mine seminars were an excellent way to familiarize myself with like R tools.

River also had prior coding knowledge. He had coded in Python and had some knowledge of R programming from his first-year statistics class.

Um, I had coded in Python before. And I've done like some web scraping stuff before. But other than that, uh, oh, and I had also, I think in my first statistics class, I learnt some R but not much so.

Category 2: How do I see myself now?

Sub-Theme 2: *After spending more than one year in The Data Mine, the students recognized themselves as data science people, or data scientists or a combination of both, based on their data science knowledge and competencies.*

The Table 21 represents the codes for sub-theme 2. The majority of students recognized themselves as data science people, as they still believe that they do not possess all the required skills to be identified a data scientist. One student felt competent with his skills and identified himself as a data scientist, and one student believed that he is a combination of both.

Table 21: Codes for Sub-Theme 2

Codes	Reference	Frequency
I am a Data Science Person	17	18
I am a Data scientist and a Data science person	1	2
I am a Data Scientist	1	1

Sub-Theme 2 and Table 21 represent perception of students about their data science skills, after spending at least one year at The Data Mine. Each code that led to the creation of sub-theme 2 is explained in detail with a corresponding student quote.

1. **I am a Data Science Person:** The majority of students described that they are in the process of developing data science skills, they have started to feel competent in using them, but they have yet to identify themselves as data scientists, as they feel that they need to develop a lot more skills and gain actual experience. Therefore, they feel comfortable in

134

calling themselves a data science person. Morgan mentioned that she has developed data science skills and grew in terms of coding skills. Therefore she is comfortable calling herself a data science person. Morgan also believes that she needs more real-world experiences such as internships, learning new languages and developing hard skills to call herself a data scientist.

But in terms of data science, I would definitely not call myself a data scientist by any means. I think I have a lot to learn, a lot more experience to have, so I would just say I'm someone that's interested and learning about coding. So, I would call myself something a data science person. I wouldn't call myself a data scientist until I've gotten more of maybe internship experience or just learning more languages or more skills or just actual hard skills, I guess.

Frankie also believes that she is a data science person. She still feels that she has lot more to learn even though she has developed a lot of skills in college and in The Data Mine. She feels that she is not confident as a data scientist should be when working on a company project.

A data science person. Uhm, I really, you know. Although I do feel like I do have a lot more skills since coming to college and doing The Data Mine. I don't feel confident in like being a full-on data scientist or being, you know, responsible for you know certain objectives that a company like hey build this model and tell us you know what to do next. I would not feel confident doing that. I feel like there's so many you know, little things. But I feel like a lot of the data science skills I've learned in college at The Data Mine have been helpful, but very like surface level and not getting fully into like diagnostics and models and stuff.

2. **I am a data scientist and a data science person:** Finley is the only student who identified himself as a data science person and a data scientist. He believes that he loves coding in high level languages such as R and Python and enjoys data science learning, so he is a data science person. But at the same time, he is working on some complex machine learning projects outside of The Data Mine and also has done internships with reputed companies where he applied his data science skills, so he feels that he is a data scientist too.

Ah, ah, I would think of myself as I think on both the data scientists and a data science person, let me tell you why, because I think I'm a data science person because I really enjoy data science. But after coming into college and this is mostly because of the Data Mine. Ah, I think I've found myself to be more of a data science

person because, you know, I don't like some of the low- level languages such as C that we have to deal with in Computer science. I like Python. I like a high-level language where I don't have to get involved with the complexity of the system. So yeah, I would say I'm a data science person because I really enjoy data science now, even more than CS. Also, I've taken up some projects Uh, within machine learning outside of college, which are really fun, and I would. I would call myself as a data scientist. I think you get to call yourself a data scientist when you actually worked. You know, in data science and that is something I've done with the research and then with the [Company Name] job as well. So, I would say I'm both a data science person and a data scientist.

3. **I am Data Scientist:** Arrow perceives himself as a data scientist as he has done prior internships, has adequate data science knowledge. He also believes that given any data set he can collaborate and use his data science skills to solve the problem like any data scientist does.

I'm definitely an aspiring data scientist. I would say I'm a data scientist. Just because I do everything that a data scientist does, so I would say I am a data scientist. Such a general term. Now there's definitely so many roles within data science. But I sort of feel given a data set or any data science problem I can collaborate like a data scientist to solve that, and I would say that I definitely have the skills required by the data scientist.

Category 3: How do others see me?

Sub-Theme 3: *The majority of students also felt that their data science skills are recognized by their peers and faculty in The Data Mine.*

The majority of students in The Data Mine agreed that their faculty and peers recognized their data science skills. Table 22 represents the code below:

Table 22: Code for Sub-Theme 3

Code	Reference	Frequency
My peers and faculty recognize my Data Science skills	15	15

Sub-Theme 3 and the Table 22 represent the perception of students about how they are seen by others in The Data Mine because of their data science skills. One code that led to creation of sub-theme 3 is explained in detail with a corresponding student quote.

4. **My peers and faculty recognize my Data Science skills:** The majority of students believed that their faculty or peers recognize their data science skills. Quinn believes that peers and faculty recognize his skills. Faculty recognize his skills because they grade student projects, so they know how each student is performing. Also he mentioned that Dr. W offered him opportunities for internships and jobs. This showcases that they know his abilities and potential. Peers too recognize his skills because he led a team as a TA, he was also competent in answering his students queries and teaching them when required.

> Uhm, I think so because, faculty recognize my skills. I think that they do because, uh, we they like grade our projects so clearly, they know like that. I have the skill that I have, and I also think that the faculty specifically has recognized it because I've they've offered me like internships and jobs and they understand the level at which I work and like what my aspirations are, so I would definitely say they recognize my skill. Also, for my peers I teach them and answer their queries so I would definitely say they too recognize my skills.

Frankie too believes that the staff and faculty recognize her skills. At the same time, she believes that faculty and staff at The Data Mine are very supportive, and they want students to succeed too, they always give recognition to students. Also, specifically for Frankie, she believes that faculty and staff at Data Mine trust her and recognize that she has developed a bundle of skills at The Data Mine.

> I think they do, yeah. I think just the whole culture at The Data Mine is just very supportive of like learning and I always felt that the people there like kind of believed in me. Whether it was, you know, peers or faculty or other mentors, they always kind of believed in me and, told me I could do whatever I wanted to do and that I learned a lot. And so, I think it has just that whole culture of learning and believing. And you know other people, I think really, it makes me feel like they trust me with that kind of stuff and believe that I have learned a lot of skills.

Robbie too believes that faculty and peers recognize her data science skills. She mentions that since she is taking a third-year statistics class (STAT 390), that makes faculty believe her competency. She also mentioned instances where faculty offered her various opportunities within the Data Mine. Also, as Robbie is a TA, she believes that her peers recognize her skills as she helps them when they work in-person on a corporate partner project.

I think so and I just think that the fact that they have confidence in me, you know, I'm in the STAT 390 class which is the third-year class. Also, they let me know the various opportunities just like saying hey, these opportunities exist in data mine, incase if you are interested. I think that that's kind of enough for me to know that they think that I have the abilities. I think my peers too know my skills. Um, sometimes people will ask for help on projects when we're working together in the in-person meetings, and I'm happy to help them with it so.

Category 4: How do I want to see me in the future?

Sub-Theme 4: *Students enjoy learning data science as they believe acquiring knowledge of data science is important, and therefore they plan to pursue a data-oriented career.*

The Table 23 represents the codes that justify the reason why students are interested to learn data science skills at The Data Mine.

Table 23: Codes for Sub-Theme 4

Codes	Reference	Frequency
I plan to pursue a data-oriented career	19	28
Knowledge of Data Science is important to me	19	20
I enjoy learning Data Science	14	19

Sub-Theme 4 and Table 23 represent students' interest in data science that they developed after joining The Data Mine. Three codes that led to creation of sub-theme 4 are explained in detail with a corresponding student quote.

1. **I plan to pursue a data related career:** The students mentioned that they want to pursue a career related to data science, data analytics, or any career that emphasizes the application of data. Some students also mentioned that they want to go to Graduate school to study courses related to data science and then pursue a future career in a data-oriented field. Rio mentioned that she wants to enter the field of bioinformatics, specifically focused on computational drug design. She also mentioned that she was unaware about this field prior to joining The Data Mine. Previously she wanted to go into biotechnology and become a research scientist as she did not have any idea about the computational aspect in drug discovery.

Uh, so as of now, I'd really like to enter the field of bioinformatics. Uh, maybe with like a focus on Computational aided drug design. But that has evolved through The Data Mine like I didn't know that was a field until I started with The Data Mine. I would say before I joined The Data Mine I wanted to work in biotechnology as like a research scientist as I didn't have any idea about the computational background of the like drug discovery things and all these things.

Harley is a non-STEM major student. Before joining The Data Mine, Harley wanted to attend Graduate School to study Psychology, but after joining The Data Mine her perspective and interest changed. The Data Mine helped her to overcome the fear of statistics and coding. She was able to apply to an Information Systems program at a large public university and she got accepted to it. She never thought that she would pursue an Information Systems major. She was also happy that the interview committee was incredibly impressed by her data science acumen. They offered her a scholarship as well. In the future she wants to pursue a data-oriented career.

I was just planning on going through Graduate School for psychology. Since that was my interest, and I was confident that I could do that. But my plans changed, so what happened was I got into The Data Mine, and I noticed that was actually enjoying programming and stats and methods a lot more than I thought I would, and it shifted my Graduate School decisions. So, I actually applied and got accepted to an information systems program. I can never imagine myself doing this. But after being in The Data Mine I realized I want to do something that's in the territory of data science and actually actively involved in that. I applied to like the Information systems program at [Large Public University]. They were actually impressed with my projects, and they had like two rounds of interviews, in the first round they asked interview questions revolving like statistics, basic methods, programming how it could go like a sort of a kind of technical or quantitative questions, and I was actually able to answer them. They were really impressed with my answers that how I was able to sort of think through that. And then I got invited to the final round. Then it was only shortly after they told me they were like so impressed they not only just like accepted me early on during the summer but also gave me like a scholarship over 50%. I don't think this would have happened if I hadn't been in The Data Mine because my fear of programming and stats would still be there. I guess eventually end up working in sort of in consulting but work that involves data science and being able to work with a lot of data to arrive at solutions.

2. **Knowledge of Data Science is important to me:** All the students mentioned that the knowledge of Data Science is important to them as it helped them get summer internships

and help them to apply their knowledge to real world settings. Finley believes that the knowledge of data science is extremely important to him. He mentions about the incident when he got his first summer job just after his freshman year, he felt that the only reason he got that job because of the data science skills he developed.

Absolutely. Like I said before, data science can be used in any industry, new word, so you know, having, uh, both practical and theoretical knowledge of it can truly help from you in your career.

I got a job over the summer.
Uh because you know it was the first time that I got a job and that was because of my skills in data science that I developed here at The Data Mine. And you know, obviously it was a very proud moment for me. Uh, getting a software engineering internship as a freshman, so you know that was the first time I think I felt that wow my data science knowledge has really helped me out.

Reef also believes that the knowledge of data science is important to him as he feels that he will get a competitive advantage during the hiring process. He is a computer science major but also has data science skills that will set him apart and are valued by many employers.

So, I think data science is important for like a general knowledge. So, I think it would be very useful with getting hired with like as an advantage for the hiring process. Because data science is something that I would know and that a lot of companies are probably looking at, and so it would show a certain amount of flexibility that I have with not just computer engineering, but because I have this experience in data science and knowledge of working with data.

Rio describes that, prior to joining The Data Mine, she did not understand the value of data science, statistics, or computer science courses, as she was a pharmacy major. She mentions an instance in which, during her freshman year, she took a statistics course, but she could not understand the usefulness of statistics to her major. When she joined The Data Mine last year, she understood the value of data science, computer science and statistics to pharmacy and drug discovery. Now she believes that the knowledge of data science and computing is very important to her.

I do, yeah. Uh, um specifically within the areas that I want to go into, so whether that's in like pharmaceuticals or I've been kind of interested in maybe going into bioinformatics. I think that Pharmacy and drug discovery is becoming more computer based than it has ever been. Just because the amount of data that we're collecting on into even have a working knowledge of that, I think is very valuable. For instance, I know like I took statistics class for my major when I was a freshman. And I remember saying when I was a freshman, I was like I don't see why this is applicable to my major because it's pharmacy. Uhm, but then last year it really hit me that like statistics and data science and computer science all play this really big role in drug discovery. Uh, and even like statistics within clinical trials, data science within clinical trials and creating new therapeutics. It has a really, really big role.

3. **I enjoy learning Data Science:** The majority of students mentioned that they enjoy learning data science skills at The Data Mine and find it fascinating to apply their skills to solve real world problems. Finley mentions that he enjoys learning data science. He enjoys data science learning most when he can see some results from application of skills. For instance, he likes to train machine learning models and he finds it satisfying to see when his models are able to do accurate predictions.

Absolutely I enjoy learning Data Science. I think it's all it's, you know. One of my favorite things to do like training model is really fun and when you see the whatever accuracy metric you're using and you see that going up, uh, with every block with every iteration of the training set you know that is really satisfying and really fun. And it's something that I've enjoyed doing.

Mickey too enjoys learning data science. She compares acquiring data science skills to learning website designing. She finds website design a very redundant task, whereas learning and applying data science skills are very fascinating. She enjoys working on the seminar projects in The Data Mine; she finds them not very easy nor too hard, but rather, they are appropriate to help students to think critically and solve them.

Yeah, definitely, I really do. I always knew that I wanted to do something with coding, so I think it's really nice to be able to do it in a more productive way. In my opinion, more productive way than just designing a website. I actually quite enjoy some of the projects. I think that they're in the sweet spot of there, just hard enough where you actually have to work and just easy enough that you'll actually be able to solve it. I feel like it's a really good opportunity and the nice thing is we're allowed to like Google things, and I think that's a big part of why I like it. Is that you can just Google things and then learn what you didn't know how to do.

Category 5: What skills/competencies did I develop by participating in events and interacting with peers/faculty/staff in the Data Mine?

Sub-Theme 5 (a): *Students developed data science skills and felt competent in applying their skills to solve real problems.*

The Table 24 represents the codes demonstrating the competencies of students at The Data Mine.

Table 24: Codes for Sub-Theme 5 (a)

Codes	Reference	Frequency
I developed Data Science skills	19	70
I feel competent in applying Data Science skills	19	37
I can apply my Data Science skills to solve real world problems	17	45
I feel competent over my peers	11	15
I did self-learning	7	10
I Mastered new Data Science skills	5	14

Sub-Theme 5(a) and Table 24 represent the data science skills that they developed after joining The Data Mine. Each code that led to the creation of sub-theme 5(a) is explained in detail with a corresponding student quote.

1. **I developed Data Science skills:** The students described how participating in various events in The Data Mine helped them to develop Data Science skills and broaden their knowledge of data science. Morgan describes when she joined The Data Mine, she came in as a novice with very little background of coding and no data science skills. The Data Mine helped her to develop her skills exponentially. She also mentions how participating in the seminar course and corporate partners helped her to develop coding and statistical analysis skills. Moreover, the corporate partners helped her to develop data collection skills and to work with qualitative data.

   I say the Data Mine has helped me with my data science skills like exponentially. I came in with very very little coding background and approximately zero data science skills. I mean I knew a little bit about statistics and a little bit about programming, but not nearly as much as I know now, after having done the seminar classes and just getting that real world experience through the project. Yeah, so first semester. They typically focus on are just very foundational skills and second

semester they move a little bit to Python And then through the project my corporate partners project I learned a lot about machine learning ANOVA, Random Forest, just different statistical models.

I also developed data collection skills Uh, we were trying our best to figure out a way to convert those qualitative data sets to more quantitative datasets just to get some sort of analysis. We interviewed a lot of employees at Corporate Partner to figure out how they gather that data, what they did with the data. Uh, and kind of what how employees felt after they gather that data. Just kind of looking at it from a larger perspective and just the numbers we tried to answer the big research question.

Rio is a pharmacy student, and she joined The Data Mine to learn coding skills. Prior to joining The Data Mine, Rio did not have any programming or machine learning knowledge, as none of the coursework in College of Pharmacy required her to code. She also mentions that, after participating in the seminar course and the Corporate Partners program, she learnt to code in Python, Bash, etc., and she also learnt machine learning skills.

I just saw it as a really good opportunity to develop my skills in an area that I had no experience in. Yeah, so uh, I joined the corporate partner cohort with The Data Mine. Uh, between working with corporate partners and also in like STAT 190 and 290 this semester. I really like fulfilled my goal of learning how to code, learning how to kind of think in a programming manner and like to manipulate data. Yeah, so I actually had no working knowledge of like Python or any machine learning setting before I joined the Data Mine always found it kind of interesting, but none of my coursework in their college of pharmacy ever required you to have a computer science class in order to relieve you time to take a computer science class. Data Mine has taught me like techniques in Python, I've learned how to code in our work with bash or the command line and various machine learning skills. So mostly we focused on neural networks and random forests. Ah, especially when it comes to like machine learning and kind of like understanding how algorithms work on and how like hyperparameter optimization works.

2. **I feel competent in applying Data Science skills:** The students mentioned that they felt competent in applying their data science skills to real world data sets, to their course work, or during corporate projects. Finley mentioned that he felt competent when he was able to use his knowledge of data science to explore and analyze real world data sets. He started working on data sets on Kaggle and applied the concepts such as linear regression and

clustering. During that time, he found a difference between being a student and evolving into a person who proficiently applies data science skills.

So, I think that particular moment was when I started applying my data science skills to projects outside the class. Because class projects, everyone else do, every student has to do, and more students end up doing that. But I think the point where I realized that wow, I have some ability to be able to actually analyze data, When I started using Kaggle.
Yeah, OK, you know, so I would read and a lot of datasets from Kaggle. Or do a data visualization with and then run some basic statistical analysis. The models like linear regression, K-means nearest neighbors, grid search, CV, stuff like that. So yeah, I would say that the point the switch over from me being or just a student to someone who actually applies will data science skills. Or you know to real world datasets and stuff like that, and that was a point where I would say that I was proficient.

Jamie mentioned that she felt that she was growing into a competent learner in multiple ways. She mentioned about her last internship experience, which was focused on data science. She felt that she was able to understand most of the problems herself and required less mentorship. Also, Jamie found that she was evolving in terms of knowledge and skills as she moved from first year of seminar course to second year. In the second she required less guidance on solving the problems and felt competent.

Yeah, I think my last data science internship with a company really helped because I could understand most of the stuff and look up the content and understand it without really requiring like help from a mentor. Guiding me as much as that happened like the first year of The Data Mine. In the second year I think I could understand a lot of it and figure out this stuff on my own.

3. **I can apply my Data Science skills to solve real world problems:** The students mentioned about how they applied their data science skills to solve real world problems. Also, students mentioned that they understood the application of data science in other disciplines. Jackie described a situation in which he was doing an internship in his sophomore year at a manufacturing plant. He found that there were no metrics to capture and report the productivity of the plant. He found that none of the engineers at the plant had a concrete answer about how well the plant did. So, he helped the plant to develop a tool to track the performance and productivity on daily basis. He was very happy that he

was able to use his data analytics skills that he developed at The Data Mine with his engineering skills to come up with the tool that solved a real problem.

When the company I was working for was a manufacturing company in Michigan and they basically didn't have a very good costing system and they also didn't have a good way of answering, for instance, how well did the plant run yesterday? They had no way to answer that, so it was kind of just hearsay, right? Like you know, I thought we did okay, but there's no metric to prove it, so that doesn't really help. It's not really viable to do that, especially at a large scale.
Uh, and so actually that's sophomore year, or that that internship I spent the summer working on basically developing a like a production performance analytics model essentially and up method to evaluate plant performance.
And that was also kind of, you know, when I realized I was like, you know I can, I can do this for a living. Like this is because it's a perfect medium for me of combining engineering and the work is engineering specific but also being able to apply data analytics to that and draw conclusions that its typical engineer wouldn't be able to draw.

Mickey mentions about a competition that she participated in, during her time in The Data Mine. The competition required her to work on real data sets related to healthcare and also to identify and merge real data sets, to come up with a unique solution for a problem. For the first time, Mickey was independently working on real world data sets, and she was able to manage and merge different data sets and come with a unique solution. She also mentioned that she felt very competent while working on the assignment.

Essentially, we're just given one data set from I think like a hospital. They gave us a data set and they were like; you find something incredible in this data, but you can use outside data and so we ended up on the learning how to develop like actually find resources from the US government. Find resources from reputable sources, get resources of GitHub, and actually compile them. Change formatting, rearranged them, visualize them together and that was something that I'd never like merge datasets before. We had to collect data sets and that really taught me a lot. What I learnt was how to use government data sets and I had never used it before, so that was the first time I worked with fairly large data and actually manage like merge datasets and those types of things, so I knew that I'd been able it because I was able to apply it in that time. I knew that I was like fairly competent with it.

4. **I feel competent over my peers:** The students expressed that they have developed skills in The Data Mine that make them feel competent over their peers. Frankie believes that she has more knowledge than her peers in terms of coding and analysis. She mentions that in a

statistics class that she is working alongside peers who have a very little knowledge of coding in R or Python or how to conduct statistical analysis such as developing regression models.

I have a little bit more knowledge than my peers of the technical aspects of it. Some of my statistics class people have like a very, very basic understanding of R or Python or whatever we're using, but a lot of times they don't kind of know the more advanced thing or like data manipulation, especially they kind of know just how to plug in data until like a you know linear regression model or something, or an ARIMA model, but they don't really know fully what the models are doing in the background I think or how to code more advanced stuff.

Marion believes that he has developed so many skills at a very young age by participating in The Data Mine. He also believes that the skills he has developed make him feel more competent over his peers who are in computer science or data science majors. He believes that many individuals in the computer science or data science majors may not have the skill set that he has.

Uhm, yeah, I suppose so. I mean, I think that there are skills that I've learned in the Data Mine that students in the data science major, or maybe computer science major don't necessarily have, and that's not a fault of their own. It's just that I had this unique experience where I was exposed to specific types of software that maybe that in your typical course load you might not necessarily get exposed to until maybe senior year, or maybe when you're in your profession. So, I think that that's certainly an advantage. I mean, you just get exposed to so much and, and especially at a young age. I mean the Data Mine does so good have of getting younger underclassmen into technology very quickly. So, I mean that's certainly valuable as well.

5. **I did self-learning:** This code captures the instances where students mentioned that they were able set their own learning goals, monitor progress, and engage in self-learning activities to enhance their knowledge. Orion mentioned about the corporate partners project that required a lot of self-learning. He mentioned that he felt confident in doing the self-learning, as he had seniors in the team who were very helpful. Every student was engaged in learning based on their capabilities, and senior members of the team were always there to offer help. Orion also mentions that self-learning is easy and fast when you have someone who can guide you or help you troubleshoot the problem.

146

So, of course there was a lot of self-learning. Actually, in this process so uh, definitely since this. There were senior members who worked with it before. There was some, so it the whole process went like everyone would learn it on their own. Look at YouTube tutorials online and try to complete it. And if they faced any bugs or had any issues then we would have another senior member, helped them get through it. I would say that's the biggest thing about so of course, everyone in the software field will say that you can do everything on your own. Everything is available online. You can just do an online course or see a YouTube tutorial and you can make it make an app yourself. But the things that people don't talk about is the issues you face. So, let say you follow these online tutorials, and you face a bug. There's no one to help you, so you can go into an endless spiral of Googling what that debugging error is, but you end up with more problems than you started with. Instead, if you have another person who's actually had experience working with that kind of problem, or working with that, he might know instantly how to solve that issue, or what that might be related to so that that's what this project gave. So, people were in a way promoted to self-learn because you always had help in the back.

Rio also did a lot of self-learning to improve her knowledge of data science. When she joined The Data Mine, she was not just a student but a TA too. Since she was leading a team, she needed to be well-informed, and so she did self-study, watched YouTube videos, and reached out to professors and corporate mentors. She also took a certificate course on Coursera on machine learning, as she was using machine learning concepts in her corporate partner project.

Uh, last year it was kind of hard because I came in and I was also a TA, but I had no background experience, so it took a lot of like self-studies as I have class. I've watched a lot of like YouTube videos explaining different techniques. I've read a lot of research papers. I've talked to a couple of professors like I've got talked to Dr. W. I talked to Corporate Partner Mentors for clarifications. I'm even somewhat talked to one other Professor because he's also involved in The Data Mine. Uh, over the summer to kind of like work on improving my knowledge in machine learning I took on a certificate course on Coursera for introduction to machine learning.

Tommie also engaged in self-learning as she is a visual learner. Also, she learns best by practicing the concepts so she would explore various software packages, run some mock programs, apply the concepts that she learnt and finally apply them to her project.

For me on the project, I'm very much a hands-on or visual learner, so the way I learned it best was by doing it and practicing it. So, I would, you know, load in my

package and then I would play around with it so I could learn. This is how this function works to really apply it then to my project.

6. **I mastered new Data Science skills:** Some students who already had prior knowledge of data science or coding skills prior to joining of The Data Mine added more skills or new skills to their portfolio. Arrow came from a computer science background, so he knew to code in Python, but he had never coded in R. Also, the statistical data analysis was new to him. As a part of The Data Mine, he learnt how to conduct statistical data analysis and that made him comfortable while talking to the statisticians and communicate his ideas better.

Yeah, I would say from the technical perspective like there's this huge focus on R programming. In in The Data Mine, and as someone who comes from like a more Computer Science background like I have this bias towards Python development, and this is an excellent way to familiarize myself with like R tools. Being more familiar with data wrangling and stuff like that is very helpful.

Uh, also I feel like that sort of made me more comfortable, like talking to statisticians. It's not just like working with other computer scientist, so yes.

River is a chemical engineering, mathematics, and statistics major. He had coded prior coming to the Data Mine. But after joining The Data Mine he learnt to code and to perform data analysis better. For instance, he mentioned that he had learned to code in Python and Bash. He also learnt the basics of web scrapping, parsing JSON files, data visualization and machine learning skills. He feels that being a chemical engineering, mathematics and statistics major, he would not have learnt about particle physics. But being part of the Physics cohort, he learned about quantum adiabatic computing and neural networks.

Yeah, I mean, I've certainly learned a lot of skills about being able to work and manipulate data. I've gotten to see a wide I think the whole kind of paradigm of Data Mine is like more of a breadth and depth, and so I've been able to see a very large breadth of different tools and techniques that exist out there solve different problems, which I think has been invaluable. So, it's, uh, it's all kinds of things I have gotten to see all kinds of stuff I have been here, that's kind of why I'm still doing it. Because there's always new stuff. You get to learn. So, working with data in Python working with data in like the command line or like bash or you know Unix, type stuff. Using those types of commands. Basics of web scraping. Learned basics of like serialized data formats like XML and JSON and how to parse those types of files. So, then we you know during the physics project I learn more about

quantum adiabatic computing. We learn more about neural networks and in particular like generative adversarial neural networks. I learned about using Tableau and the data visualization class. It gives you a nice overview of a lot of different topics.

Tommie also had prior programming knowledge, but she was not an expert in applying the disciplinary knowledge to the actual context. The Data Mine not just helped her to enhance her coding skills but also helped her to learn new concepts such as natural language processing and machine learning, and apply them to solve real world problems.

I learned how to code in R as well as enhance my skills coding in Python And currently I'm working to learn Unix as well, which is another coding language. I worked on with corporate partners I had to introduce a lot of new skills to me. So, I learned about packages with like natural language, machine learning. So, it was applying these super like big concepts to our particular projects, so you had to learn about them and then you had to apply them.

Sub-Theme 5(b): *Students developed skills such as research, mentoring, and professional skills.*

The codes in the Table 25 below demonstrate the non-technical, professional and research skills that students developed as the part of The Data Mine. Also, it is important to note that since the discipline of data science is interdisciplinary, the research and non-technical skills are equally important for the students (Koby & Orit, 2020). Sub-theme 5(b) and Table 25 represent non-technical and professional skills that students developed after joining The Data Mine. Each code that led to creation of sub-theme 5(b) is explained in detail with a corresponding student quote.

Table 25: Codes for Sub-Theme 5 (b)

Code	Reference	Frequency
I developed research skills	12	18
I developed Professional Skills	10	13
I developed Mentoring Skills	10	12

1. **I developed research skills:** The students described how students got various research opportunities at The Data Mine and developed research skills. Frankie mentioned about her experience of working on the Corporate Partners project. She mentioned that, apart from learning technical skills, she also learnt research skills. Prior to joining The Data Mine, she had never explored research papers and never worked on research projects. Now

149

she has developed research skills, so she can write abstracts, create research posters, analyze data, and present her research to a wider audience.

So, I learned a lot about that and besides the technical things too, I learned how to kind of put together like a research paper and poster. I had never done any like technical kind of research before, so I learned a lot about that as well. And, you know, writing abstracts and background information and I learned a lot about the specific background information on the projects. You know the first year we did stuff related to like weather and crop yield, and different biophysical parameters and then the second year we looked at more HR and absenteeism and I didn't really know anything about those, and I think a lot of, or a big part of data science is, learning your background information so you're actually know what the data is telling you. Or you can interpret things correctly.

Skylar said working on research project with a Professor at The Data Mine helped her develop research skills. Skylar mentions that she developed a wide range of data science skills, from conducting analysis to writing a manuscript.

Yeah, I learned a lot about writing a research paper and conducting research and using like statistical modeling inside of research in order to gain aggregate results. So, all of what I've learned from my research on was because of her [mentor]. So, I can only really speak on this on this side of research. The research I've done and so I like the process of like acquiring data, we focus mostly on the process of acquiring it and then cleaning it and then analyzing it based on different like statistic modeling. Yeah, I like it was like linear regression. We did like odds ratios and things like that.

2. **I developed Professional Skills:** The students described the soft skills such as communication, teamwork abilities, presentation skills, people management skills, leadership skills, and other professional skills that they developed at The Data Mine. Mickey mentions that she was very closed off prior to joining The Data Mine. She also took a course on campus to improve her communication and presentation skill, but that was not very helpful. The biggest skill that she learnt was to collaborate and work with other students in a professional setting such as corporate partners. She also developed presentation skills and communicating her ideas to others.

Yeah, so like I said, I think collaboration in a professional setting was the biggest thing that I learned. I think also just learning how to present to people because I wasn't very good at presentations. I did take [course name] but that doesn't really

teach you very much. So, learning how to present to an employer learning how to present to clients that sort of thing.

Rio mentioned that she learnt a bouquet of professional skills, starting from leadership skills, as she was managing a team. Then she leant communication skills, as it was very different to communicate and manage team in a virtual setting. She also learnt people management skills, such as strategies to work with students from different backgrounds.

I'd say I learned a lot of leadership abilities or leadership skills. Communication was a lot harder because everyone was virtual. Also, students were in different time zones. And so, I kind of learned how to communicate really effectively last year also learned in this year, like how to cater to different audiences so not everyone on the team comes from the same background, has the same experiences, understands things the same way. So, kind of learning how to work with each different type of person my interaction with students really helped, that's like a skill that I could carry on with me, so the top there would be like leadership, communication and just interacting with different backgrounds.

3. **I developed Mentoring Skills:** The students described their experience of working with peers and junior members at The Data Mine and developing mentoring skills. Some students also mentioned specific opportunities that they got through The Data Mine where they mentored students. River mentioned about an opportunity that he got through The Data Mine. During a summer he came across an internship opportunity posted by Dr. W on Piazza. He worked with a chemical engineering Professor to develop a data science curriculum for undergraduate students in the chemical engineering. He was also required to mentor those students for four weeks based on the curriculum. Therefore, he developed mentoring skills while working with the undergraduate chemical engineering students.

Last summer, I was like a mentor for a chemical engineering data science project. And part of my role was mentoring and developing the curriculum that they were going to learn because they needed to learn all the skills, they were going to use in the project within the first two weeks. So, they were chemical engineering students, so they were like junior and senior chemical engineering students, but they didn't have a background in data science.

Arrow likes to mentor his peers as he is a Head TA for seminar course. Also, he finds that mentoring helps him to improve his communication skills as it allows him to explain his ideas and concepts to others.

Um, definitely, I feel like I'm doing a lot more mentoring as head TA. And being able to explain tough concepts to people were trying to understand things that sort of like that. That's just improving my communication skills.

The theme and subthemes for data science identity demonstrated that all the nineteen students who participated in the interview study demonstrated a data science identity, meaning they perceived themselves as a data science person, data scientist, or both. The majority of students acknowledged that they were novice learners before joining The Data Mine, but after joining The Data Mine, they developed considerable skills, such as coding skills, research skills, communication skills, professional skills, etc. Students also acknowledged that their peers, faculty, and staff have started recognizing their data science skills. All the students demonstrated an interest in learning and applying data science skills and wanted to pursue a data-oriented career in the future.

7.3.2 Theme for Leader Identity

This section focuses on the perception of students who were the TA for Corporate Partners (CRP). The CRP TAs are constantly interacting with faculty, staff, and peers at The Data Mine, helping their peers to trouble shoot problems, guiding them and coordinating with corporate teams. This constant engagement of TAs with various members and in their role as a peer leader have led them to also form a *leader identity*.

The result of this section comprises one of the broad, overarching themes that demonstrates the leader identity perception of Corporate Partners TAs. The broad theme is further broken down into categories, and each category has sub-themes. The categories for leader identity are: *1) Who am I? 2) I enjoy my role and interacting with my Co-TAs, 3) Roles and responsibilities as Corporate Partner TAs.*

Theme for leader identity: *The corporate partner TAs agreed that they were developing a leader identity along with a data science identity, as they were enjoying their role and felt competent in solving student queries and managing a team.*

The paragraphs below demonstrate the corresponding categories, sub-themes, qualitative and student quotes that fall under the broad theme.

Category 1: Who am I?

Sub-Theme 6: *All the corporate partner TAs perceived themselves as an emerging leader.*

The codes in the Table 26 represent how the corporate partner TAs recognized themselves. Sub-Theme 6 and Table 26 represent how Corporate Partners TAs recognized themselves at The Data Mine. The code that led to creation of sub-theme 6 is explained in detail with a corresponding student quote.

Table 26: Code for Sub-Theme 6

Code	Reference	Frequency
I am a leader	12	12

1. **I am a leader:** The corporate partner and seminar TAs identified themselves as growing leaders. They describe how leadership opportunities at The Data Mine helped them to develop leadership skills. Jamie is thankful that she was the TA for the Corporate Partners team. She was able to teach technical skills and lead a team. She also mentioned that school does not necessarily teach the leadership skills; rather they come from experience.

   I'm really thankful I did [TA for Corporate Partners] because just the experience you get from, you know, teaching kind of like leading the team. I guess it's a leadership skill that you don't necessarily get in school, so it it's so cool to just be able to learn from kind of better.

   Quinn also mentioned that being a leader was a great and novel experience for him as he managed a team, facilitated meetings, and motivated his team members.

   So, it was pretty novel experience for me. Uhm, I learned that I really like leading people. I think getting to like facilitate meetings and Uhm, manage people and help people speak up when they're not speaking up enough or indicating. And I learned a lot about myself and the way I like to lead and kind of what makes a good team,

153

such as like having good communication outside of meetings and during meetings and kind of inspiring people to want to work on what you're working on.

Category 2: I enjoy my role and interacting with my Co-TAs

Sub-Theme 7: *The Corporate Partner TAs mentioned that they enjoy being TAs and interacting with and learning from other TAs.*

Sub-Theme 7 and Table 27 represent that all the CRP TAs enjoy their role and learning from other TAs. The code that led to creation of sub-theme 7 is explained in detail with a corresponding student quote.

Table 27: Code for Sub-Theme 7

Code	Reference	Frequency
I enjoy being a TA	12	12
I enjoy interacting and learning from other TAs	12	12

1. **I enjoy being a TA:** CRP TAs mentioned that they love being a TA and enjoy their role. Riley mentioned that she enjoys her role as she is engaged in research and learning something new every day. She also gets the opportunity to interact with other faculty and staff at The Data Mine and to plan activities for her corporate partner meetings.

   I loved being a student. It was great because you're really involved. It's not a normal class where you sit back, and you just take notes and take an exam. And that's what I love about The Data Mine even as a TA, uhm, that like you know, it's not just that I'm not taking exams and learning that way because you're doing like real research, which is amazing, so but being in that role is different, but it's like so I love it so much more because I do get to talk more with the mentors from Corporate Partner and more with Ms. M and Dr. W and other administrators within the data mine, which is great.

2. **I enjoy interacting with and learning from other TAs:** Student TAs mentioned that they interacted with and learned from other TAs. Tommie mentioned that, as a TA, she got several opportunities to learn from other TAs. She mentioned about a time when all TAs in The Data Mine met and shared their experiences and learnt from one another. Tommie attended one of those meetings that allowed her to talk to multiple TAs—both new and experienced ones—and learn from their experiences.

154

Uh, for the students they have The Data Mine corporate partners has basically put together like a time where all the TAs can come together, sort of like as a gathering and we can talk and sort of like learn more about each other's experiences to improve our own. So, the one of the meetings I was able to go to, we talked like as a group. Basically, about our experiences and then we did sort of like a speed dating setting sort of idea where you talked to someone across from you, and then you know you move down and then you talked to the next person. So, I was able to talk to a large variety of people and then I was also able to talk to like some of the more experienced mentors who are coming back.

Marion mentioned that he learnt many non-technical skills from his fellow TAs. He attends the TA meetings every week and he can talk to his peers and learn from their experiences, and he also implements them in into action. He considers this knowledge sharing to be extremely valuable.

I guess from a non-technical sense just like. How to approach a leadership opportunity like I think you know The Data Mine has a slew of TAs. Now we have, you know 45-50 days and we have meetings every week as a TA group and I think just having conversations with my peers is so valuable to hear their experiences and hear the solutions that they're implementing that I can take from them and implements into my team is extremely, extremely valuable.

Category 3: Roles and responsibilities for Corporate Partner TAs

Sub-Theme 8 (a): *The Corporate Partner TAs helped students by answering any technical or non-technical questions in the lab meetings or during office hours.*

Sub-Theme 8 (a) and Table 28 indicate that all TAs responded to other students to clarify any questions or doubts that students had regarding any project or assignment, and conducted office hours to provide more assistance. The codes that led to the creation of sub-theme 8 (a) are explained in detail with a corresponding student quote.

Table 28: Codes for Sub-Theme 8 (a)

Code	Reference	Frequency
I helped students by responding to their queries	12	14
I conducted office hours to help students with their queries	12	14

1. **I helped students by responding to their queries**: The CRP TAs mentioned that they helped students and answered their questions (technical or non-technical) or taught them a

new concept or skills. Rio mentioned that her Corporate Partner has a GitHub site, and students were asked to go utilize it during the first few weeks. Since the content had lot of pharmacological terms that were very new for computer science students, Rio annotated the added comments to the notebook. She also created a video for her team to understand the basic concepts of machine learning and the data set that they were using for the project.

> Specifically, our corporate partner company has a tutorial site on GitHub. Uh, I had the students work through that for the first couple of weeks, but it was kind of confusing 'cause it uses various different biological and pharmacological terms, that was difficult for the CS students. So, I've gone through like in my free time, and all annotate that, or add comments to the notebook. I made a video on going through one of the tutorials that focused on machine learning and kind of explaining, like how it works. What each function means, what each column in our data set means and why that's important, so the goal.

Jackie mentions that his team has lot of freshman students, and they have questions not just related to project but about Purdue in general. Therefore, he makes himself available to answer their questions and guide them.

> I know a lot of the kids I had on my team last year were freshmen. So, to have someone at Purdue who's gone through things, whether it be The Data Mine, whether it just be classes, you know, just someone you can reach out to for any kind of question. Whether it be a corporate partners related or not, I think is really valuable to the kids, so I try to make myself as available as possible so they can always reach out with any questions.

2. **I conducted office hours to help students with their queries:** All CRP TAs mentioned that they conducted office hours either in-person or virtually to help students by answering their queries. Riley mentions that conducting office hours in The Data Mine is very flexible. Therefore, she conducted an office hour at 11:00 pm in the night and students came. She specifically mentioned that she also hosted small office hours in case any students needed help. In one case she conducted an in-person office hour where she helped three students on a project and submitting the project to GitHub.

> It's great about this The Data Mine, where I don't have to like help students at specific times, I make my own office hours like one time I had them in 11:00 PM and students came so it was very fun like that so. Like I mentioned as a TA I hosted some small office hours getting things done like I helped three students and I helped

one of them for a long time because we were troubleshooting, and I eventually got all of the projects submitted onto Git.

Rio mentioned that she too conducted office hours that allowed her to meet students and helped them to develop technical skills. She found that conducting office hours helped her to know her team members better and she made friends.

I made friends was setting up like office hours and meeting up with other students who are in my team to help them develop their skills and help do the tasks that we were assigned. And this just gave me like an opportunity to get to know them for who they are and talked to them I generally host office hours Wednesdays and then we have TA meetings once a week, so those are Wednesdays from 4:00 to 5:00 PM

Sub-theme 8 (b): *The corporate partner TAs managed teams, conducted team bonding activities to promote student engagement and served as a channel between students and corporate mentors.*

Sub-Theme 8 (b) and Table 29 indicate that all corporate partner TAs managed teams, conducted team building exercises and acted as a bridge between corporate partners and the student team. The codes that led to creation of sub-theme 8 (b) are explained in detail with a corresponding student quote.

Table 29: Codes for Sub-Theme 8(b)

Code	Reference	Frequency
Managed Teams	12	14
Served as a Liaison between The Data Mine and Corporate Partners	12	12
Promoted student engagement through Team bonding activities	12	12

1. **Managed Teams:** The corporate partner TAs described how they developed team management skills. Riley mentioned that she loves to be a TA and she has developed team management skills. Riley and her Co-TA work together to manage the team. They plan lab times, conduct ice-breaking activities, and get assignments done during the lab session. She makes sure that the team meets their deadlines in a timely fashion.

I love it [team management] so much more because we get to plan things like lab times and you know, we do also a little icebreaking activity typically, every time which is good. Uh, and yeah, planning stuff like that just behind the scenes is what

157

I like. And also, I like to say that I have leader qualities, so I also learnt like getting small assignments done, making sure the team has it the deadlines in their calendar and notes are being written and stuff like that so.

Jackie too is a Corporate Partner TA, and he also manages a team. Managing the team has improved his leadership skills. He mentions that the best thing he has learnt is how to manage a team of diverse people. He has also learnt how to handle the team in a virtual setting. He mentions about last year when everything was online due to COVID-19 and it was hard for a TA to manage a team, so he started organizing one on one sessions to check-in with students and keep track of their progress.

Definitely learnt more leadership skills. Uh, I manage a bunch of students, but they can be any background they can be you know, any age level, they can be freshman. They can be grad students so watch people kind of learn and grow over the semesters or over the whole year I guess is really cool, but I definitely learned kind of what I want to say. You know, kind of keeping people online. Uh, because it's easy for people to slack off. You know once it was not checking in on them. So last year with COVID, we didn't actually have any in person meetings, so it's difficult to kind o, keep people accountable virtually so I try to set up one on one to check in with the kids. Throughout the week I make sure people are making progress, because you know if it's virtual, it's super easy. Just turn your camera off, turn your microphone off and go take a nap during a meeting, whatever.

Tommy also managed a team. She was responsible for meeting the team twice per week. One meeting was virtual using Microsoft Teams and second meeting was in-person, which was a 2-hour lab session. She helped her team to adopt the agile framework. She specifically mentions about her first lab, in which her team developed survey questions for data collection.

Yes, so we meet once a week on teams for about an hour and then we meet in person for a lab for about 2 hours each week. Yeah, so I've come help facilitate them through a couple activities. We've talked about agile; we did an activity with that to sort of introduce what that method is to the students. That was actually our first lab, but every lab since we always are doing something in terms of how we are developing our questions for our surveys and data collection.

2. **Served as a Liaison between The Data Mine and Corporate Partners:** The corporate partner TAs described ways that they acted as a liaison between The Data Mine and

Corporate Partners. Marion mentioned that his role as TA is not focused on doing the work but getting the work done. Therefore, he is responsible for managing the team, dividing the work into sprints, ensuring smooth execution of each sprint, and finally acting as a bridge between The Data Mine student team and the Corporate Partners.

 I'd say the main difference as a TA is where we are removed from the physical development of the application in some sense, so our role is now less of a developer, and we are more of a leader where we are expected to keep everything on track. And so, the way we do that is we work within the agile methodology. Uhm, I'm responsible for conducting Sprint events and making sure that the development team, which is the group of students are, are on track with their user stories and their task ownership. And then I'm also kind of like that bridge between the corporate partner mentor and the students. Uhm, like that a role is meant to be that communicator between both. So, to be kind of like a resource for each of those groups.

3. **Promoted student engagement through Team bonding activities:** The corporate partner TAs described the steps that they take to improve student engagement in their teams. For instance, they conduct some team bonding activities such as icebreakers, going out for lunch, or other social activities with their team members. Rio mentions that she loves to do the ice-breaker activity to help her team members to interact with one another. For instance, during the first 10 minutes, she will post a list of questions on the screen and ask students to share their views. Also, she has conducted events such as speed networking where students get to talk to other students in a short time frame. Rio believes that these icebreakers are an excellent way to boost team motivation.

Yeah, so usually like when students are just walking in. Since December I have been in BM building, but there are a lot of students that come from like the other side of campus because they have a class right before lab. So generally, for like the first like 10 minutes or so, I'll put up like a question on the screen or like a list of questions on so for instance. It's like it could be any flavor Jellybean which flavor would you be? What's your favorite Donut? What's your favorite song? Uhm, but then also try to do like a game for the first 10 to 15 minutes where they can interact with each other and like because it's on a Friday afternoon is when our labs are so everyone kind of dragging by the end of the week. So doing ice-breakers kind of gets the energy flowing. So, the first meeting that we did we had like a speed networking thing, so everyone talked to each other for one minute. I had a bunch of questions on the screen, and everyone got to know like their names where they're

from. Also, I had questions. It's like this is a pop tart, a calzone or ravioli. Or also like playing like Ninja, so that's more of an interactive game, so go outside and then I'll have them get in like their groups. And then we'll play a few rounds and enjoy. And so those are like kind of the Ice Breakers that I like to do.

Tommie mentions about the instance where their corporate mentor is coming to meet the team and after lab the whole team is going on a team lunch.

Oh yeah, so our mentor is actually coming out next week. We're super excited. But she's supposed come out for our lab, so will do lab and then we are going to lunch together.

The theme for leader identity was particular to the corporate partner TAs (CRP TAs). All CRP TAs demonstrated a data science identity and a leader identity. All CRP TAs were responsible for managing corporate partner teams at The Data Mine and acting as a liaison between the team and the corporate partners. The CRP TAs in The Data Mine underwent thorough scrum-agile training, and as a TA, their role is to follow the scrum principles to manage their teams. All the CRP TAs have a variety of roles and responsibilities, such as conducting team meetings, team labs, creating meeting notes, circulating them to team members, generating team reports, grading, conducting office hours, responding to any technical or non-technical student questions, and keeping the corporate mentors updated on the project progress.

7.4   Summary

This chapter explored three sub-research questions. Their results helped us to understand the benefits and challenges of participating in The Data Mine. It was observed that students identified more benefits and few challenges. With respect to socialization, students identified multiple opportunities to socialize and described their peers, faculty, and staff as very interactive and accepting. All students mentioned that participating in The Data Mine helped them to develop a data science identity. Another intriguing finding was that students who were the Teaching Assistants (TAs) for corporate partners also demonstrated an emerging leader identity.

CHAPTER 8: GROWING INTO AN EXPERIENCED MEMBER OF THE DATA MINE – NARRATIVE ANALYSIS

In this chapter I intend to answer the overarching guiding research question of our study. **Guiding RQ 2:** How do students' participation in activities and interaction with peers, faculty, and staff at The Data Mine contribute to becoming an experienced member of the learning community?

The question above will also take into consideration the analysis for the three sub-research questions conducted in Chapter 7. In this chapter we first conduct an agglomerative hierarchical clustering using complete linkage (Cheng & Shwe, 2019) and then select the representative students from each cluster for narrative analysis. Hierarchical clustering is one of the most widely used methods for clustering in education studies (Jaiswal, Lyon, Zhang, et al., 2021; Jaiswal, Karabiyik, Thomas, et al., 2021; Psaromiligkos et al., 2009). The study by Jaiswal, Lyon, Zhang, et al. (2021) used hierarchical clustering to cluster students based on their reflection score. The clustering helped to divide the students into two categories of active and inactive reflectors. Further thematic analysis was conducted on three representative students from each cluster. The study by Ortega-Alvarez (2019) used hierarchical clustering to group teachers who demonstrated similar views of teaching. The clustering algorithm created four groups, and the author selected representative students from each cluster to conduct the narrative analysis. This dissertation study draws on the study by Ortega-Alvarez (2019), and therefore I conducted agglomerative clustering (a type of hierarchical clustering) based on the code frequencies and further conducted a narrative analysis. Agglomerative clustering is a widely used method for user profiling (Balcan et al., 2014; Yin, 2018). Cobo et al. (2012) used agglomerative clustering to cluster learners' profiles based on their participation in an online learning platform. Similarly, Yin (2018) used agglomerative clustering followed by K-means to group students with similar behavioral profiles. Specifically for this study, agglomerative clustering using complete linkage was conducted to produce tighter clusters (Ridwan & Retnawati, 2021). This means that students who are very similar to each other were grouped together. The codes generated for each student and the frequency of each corresponding code served as input features for the clustering algorithm. The clustering algorithm produced five clusters, as shown in Figure 18. Each color in the Figure 18 represents a cluster. From the dendrogram we can interpret that the student at the left most end of the dendrogram

(Finley) has different experiences, characteristics, skills, etc., from the students at the left most end of the dendrogram (Harley, Riley, Skylar, Frankie, and Rio). It means that students who share common characteristics are closer to one another (Ridwan & Retnawati, 2021).

Based on the results of the clustering, I intend to identify the common codes for clusters 2 to 5. For cluster 1, I have considered all the codes, as there is just one student. Therefore, the result for each cluster includes the table of codes that summarizes the characteristics of that cluster based on the qualitative analysis, followed by a narrative of students in each category.

The narrative analysis is an excellent method to retell the story and experiences of the students enrolled in The Data Mine. The intent of conducting the narrative analysis is to present the stories of the representative students in each cluster. The stories will also help to illuminate how the experiences of each student in each cluster are different from the experiences of students in the other clusters. To conduct the narrative analysis I used the step delineated by Emden (1998); Kelly and Howie (2007) and Petty et al. (2018). Students from each cluster were purposefully chosen; an exception was Finley, as he was the only student in cluster 1. The intent was to select students that demonstrated a wide range of experiences within The Data Mine. For Cluster 1, since Finley was the only student, therefore, he was selected for narrative analysis. In Cluster 2, Arrow and River were selected. The reason for selecting Arrow was that he has been in The Data Mine for the last 3 years. He has demonstrated a substantial amount of growth within The Data Mine. He started as a student, and he is now a head TA for the seminar courses. I selected River because he also had been in The Data Mine for the past three years. He is the only student out of nineteen students interviewed who has participated in the *research with a professor* cohort, lived in Hillenbrand Hall, and grew personally and professionally within the Data Mine. For Cluster 3, I chose Marion, as he was a part of The Data Mine from its inception, and he did not have any coding background prior to coming to Purdue. He grew exponentially at The Data Mine by developing and honing leadership skills, data science skills, and people management skills. Marion has also been the TA for the Corporate Partners Team for the past two years and has also done co-ops as a Data Engineer. For Cluster 4, I selected Kendall, as she spent almost three years in The Data Mine and grew from a novice learner to a mature learner. She describes the challenges and benefits that she faced as a member of Corporate Partners Team, benefits of living in Hillenbrand, and her interactions with faculty and peers. She also mentioned about the benefit of mentoring that she received from faculty and staff in the Data Mine. One important point to note here is that

Kendall is not a TA, but as a student she got multiple opportunities to grow into an experienced member of the learning community. Lastly, for Cluster 5, I chose Riley as she is a non-stem major, since she joined The Data Mine with no coding knowledge, but she grew into a mature learner. She has developed coding skills and professional skills at The Data Mine. Also, she is currently a Corporate Partner TA.
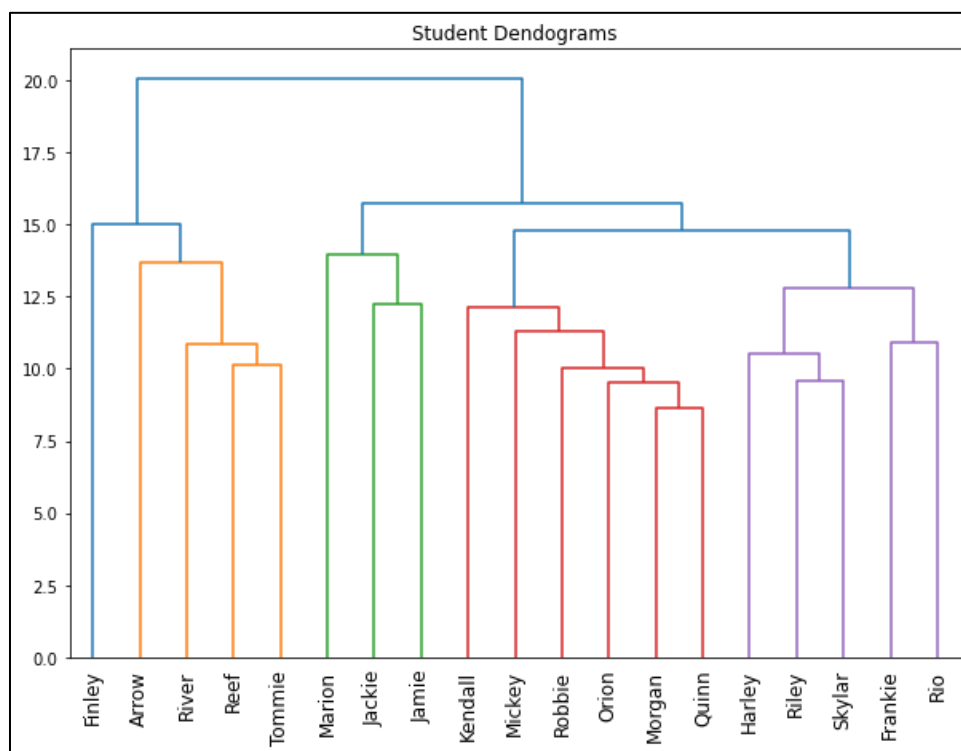
Figure 18: Student Clusters

8.1 Cluster 1: An International Student grew into a Seminar TA

The cluster 1 consists of just one student, Finley. Finley is an international student, joined The Data Mine in the Fall of 2020. He could not join the Data Mine Corporate Partnerships program as he was taking course remotely from his home country for the Fall 2020 semester. He started in The Data Mine as a general cohort student but found multiple opportunities to grow and learn. In the spring of 2021, he was on campus and was able to attend in-person classes. He found multiple opportunities to interact with faculty, staff, and peers in The Data Mine and work on research projects. He also got an opportunity to become a TA for a seminar course. In the Fall of 2021, he was able to join the Corporate Partners Team and work on a company project. It is

noteworthy that being a remote student for a semester and starting in a general cohort, The Data Mine offered him multiple opportunities to grow and learn. Table 30 below illustrates the codes identified from Finley's interview.

Table 30: Codes for Cluster 1

Category	Codes
Perception of Benefits/Challenges	Faculty and Staff were supportive and knowledgeable
	Dr. W is a wonderful Mentor
	TAs were helpful and knowledgeable
	Seminar course helped to develop data science skills
	Internship Opportunities
	Outside events helped to understand the application of data science
Perception of Sense of Belonging and Socialization	Got multiple perspective to solve same problem through interaction with students from various backgrounds
	Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor
	Lived in Hillenbrand and met people with similar interest
	I felt included
	Faculty and Staff made me feel connected
Perception of identity	I am a data scientist and a data science person
	I developed Data Science skills
	I mastered new Data Science skills
	I feel competent in applying Data Science skills
	I developed research skills
	I can apply my Data Science skills to solve real-world problems
	I enjoy learning Data Science
	Knowledge of data science is important to me
	I want to pursue a data-oriented career in future
	I had prior programming/coding knowledge

From Table 30, we can interpret that Finley found The Data Mine a very beneficial and valuable experience. Irrespective of participating virtually for a year, Finley found the faculty, staff, and TAs very helpful and knowledgeable. Also, he participated in various activities such as seminars and outside events and got an internship opportunity through The Data Mine. After Finley was on campus, he started living in Hillenbrand Hall, participated in a corporate partners project, and found multiple opportunities to interact with peers and learn from them. The interaction with peer, faculty, and participating in various events helped Finley to develop many new skills, including data science skills and research skills, and apply them in a real-world context.

The subsequent paragraphs present the journey of Finley in The Data Mine.

8.1.1 Finley

Finley is a sophomore in The Data Mine. He is double majoring in computer science and data science. Finley is an international student, and he attended his freshman year online from his home country. Finley is interested in machine learning. He is interested in machine learning at the intersection of data science and computer science; therefore, he chose computer science and data science as his majors. Initially, Finley wanted to join the computer science learning community located at Earhart Hall. Since he was on the waitlist, he decided to (instead) join the Data Mine. He considered choosing The Data Mine to be one of the best decisions of his life. Before joining The Data Mine, Finley went through a few videos about The Data Mine, and he was fascinated by the instructional style of the seminar course. Finley loved how the seminar classes were taught; it was a hands-on experience as no lecturing was involved. He found those classes a perfect opportunity to build on his data science skills; at that point, he decided that The Data Mine would be the right choice for him. Finley wanted to join the corporate partners Team of The Data Mine but since he was in his home country, he could not enroll in this team for his freshmen year, so he went ahead with general cohort. Since Finley is on campus in his sophomore year, he is a part of Corporate Partners Team and TA for the STAT 190 seminar course. Finley mentioned that The Data Mine exceeded all the expectations he had before joining it.

> I'm really glad that I joined Data Mine. A lot of the things that I've done like research, my corporate project that I am working currently and, uh, TA, my job or are completely because of The Data Mine and their opportunities. I wouldn't have

165

had these opportunities and experiences without The Data Mine. So, you know, I think not getting into the computer science learning community was kind of a blessing in disguise.

Even though he was in his home country, he was never deprived of opportunities. He mentioned a research project that he participated in the Fall of 2020, and that was his first encounter with research. He worked on a research study that analyzed the impact of COVID-19 on the general and mental health of university students. He served as a team leader for the data analytics team, where his responsibility included migration of data, cleaning data using R, imputing data using R, and creating statistical models. He mentioned that this project was an excellent learning opportunity for him as he was applying the skills that he developed, and he is very thankful to Dr. W for his mentorship on the whole project. He describes Dr. W as an amicable and cheerful person. Finley says,

> Ah, you know, working with Dr. W, was a lot of fun. I think Dr. W, without a doubt is my favorite professor on campus. He's just, really funny, really cheerful, always willing to help if you mail him even at 11:00 PM. He's going to respond in like 15 minutes with all the information you could need.

The most interesting point that he mentioned about Dr. W is that Dr. W writes very personalized emails, and his email writing style has changed the way Finley writes his email. He further adds that Dr. W is not just an excellent mentor but a wonderful person who cares about his students.

Finley also discussed other participation opportunities in The Data Mine, such as the seminar class, which he found very useful, even though he had prior programming background. He found that every seminar project had some purpose. The projects helped the students develop the data science knowledge incrementally. For example, in the first week, students learned fundamental mathematical functions such as doing calculations like 5+5 and print function, and gradually the complexity of projects increased. For example, in the assignment for Week 8, students were supposed to use all the skills they have developed through Week 1 to 8. Finley also emphasized the programming aspect, and he said students start with R and move ahead and learn Python, Bash, and SQL. Finley describes these seminar classes as very different from traditional computer science or statistics classes because he says that at the end of the semester, when students look back, they realize that they have developed so many skills. Finley also participated in virtual

outside events; he found these events a perfect avenue to learn more about the data science application in a real-world setting. He also finds that participating in the outside events allow the students to network with the speakers. In one of the events, the speaker mentioned the internship opportunity available in their lab, Finley applied for that, but he couldn't join because of delayed paperwork issues. He also mentioned another outside event that stood out to him. The speaker for this event was a data scientist working with San Francisco 49ers. He talked about the application of data science in sports. Since Finley loves sports, this event helped him learn a lot about sports analytics and its application. Finley says:

> It was a speaker from San Francisco 49ers, and he did sports analytics. And you know, I'm big into sports or, I watch a lot of sports and sports analytics is something that I haven't had too much experience with. But now it seems extremely interesting, and it's something that I would love to do so or hearing his talk on, you know, sports analytics was really interesting, and it opened up the possibility of actually getting into sports analytics as a career.

In Fall 2021, Finley was part of the corporate partners team, and he is already excited that he will be working with a company. It has been just two weeks since Finley and the team have started collaborating on the project. They have been engaged in some small group activities, which he found fun, and he is excited to work and learn with his corporate partner team. Finley describes his first two weeks with the corporate partners team, where he got the opportunity to meet his team members and to work together on some activities:

> I'm in corporate partners, which is very collaborative. So, in our first week we had to split into teams of two. I think we were about 15 or were about 15 or 16 of us. So, we were split into teams of like 7-8 people and introduced us to the agile methodology, which they use in the corporate partnerships' cohort. In corporate partners they made us make a Data Mine building out of Legos, and you know they had a bunch of objectives and requirements that were, you know, sticky notes on a whiteboard and working with 7-8 other people building this project. Thinking of ideas was, it was a lot of fun. It's I think it's been one of the highlights of this semester. Most of our corporate partners activities have been collaborative, so yesterday, for example, we had a class where we had three research papers that we had to read, and we were split into teams of five people each, and each group was assigned one research paper. We had to go through the paper and then present the paper to the rest of the class, so you know the collaborative aspect and in corporate partners has certainly been. It's definitely a lot of fun and I'm looking forward to working with my teammates even more in the future.

After spending more than a year in The Data Mine, Finley considers himself a data scientist and data science person. He perceives himself as a data scientist when he compares his skills with other sophomores. He gives an example of a coding class in which he was the first student out of 200 students to complete the given code. Also, he is working on various data science and machine learning industry projects. Moreover, he also got an internship with a reputed company through The Data Mine. In his internship, his role is to improve prognostic centric algorithms written in R to Python and then deploy them in a PySpark environment. But he perceives himself as a data science person as well because he is currently in the phase of developing data science skills and is also very passionate about data science. Finley also mentioned that he feels very competent in using the skills he developed at The Data Mine in other computer science classes and helped other students in the seminar class, as he is a TA for STAT 190. As a TA, his responsibilities were to help students during the seminar class, conduct office hours, and grade projects. Finley considers himself very fortunate that he joined The Data Mine. He says he grew exponentially in just one year. The opportunities he got in and through The Data Mine are unmatched, and he is looking forward to his corporate partner experience and working with his peers on the corporate partner team. Therefore, he is very excited to work in a team and to explore the real-world data. He is looking forward to a meaningful experience.

8.2 Cluster 2: Programmers developed data science, professional, and research skills

Cluster 2 consists of four students Arrow, River, Reef, and Tommie. The commonality among these four students is that none are novice learners. All four students had prior programming backgrounds, some had moderate coding skills, and some were experts. Participating in The Data Mine gave each one multiple opportunities, to develop a portfolio of data science, professional, and research skills. Students worked in groups either on corporate partners projects or on research projects where they learnt multiple ways to approach a problem and guided other students in problem solving. All these students lived in the Hillenbrand Hall and had a great 'live and learn' experience at The Data Mine. Table 31 below illustrates the common codes that were identified from the four interviews.

Table 31: Common Codes for Cluster 2

Category	Common Codes
Perception of Benefits/Challenges	TAs were helpful and knowledgeable
	Seminar Course helped to develop Data Science skills
	Faculty and Staff were supportive and knowledgeable
	Outside events helped to understand the application of data science
	Piazza platform to ask questions and seek help
Perception of Sense of Belonging and Socialization	Got multiple perspective to solve same problem through interaction with students from various backgrounds
	Worked together in groups on problems and learnt from one another
	Lived in Hillenbrand and met people with similar interest
Perception of identity	I developed Data Science skills
	I can apply my Data Science skills to solve real-world problems
	I mastered new Data Science skills
	I feel competent in applying DS skills
	I plan to pursue a data-oriented career
	Knowledge of Data Science is important to me
	I developed research skills
	I helped my peers to solve data science problems
	My peers and faculty recognize my Data Science skills
	I had prior programming/coding knowledge

From Table 31 we can interpret that these students did not recognize themselves as novices; rather, they agreed that they had data science or coding skills prior to joining The Data Mine. However, they also agreed they mastered new technical skills after joining The Data Mine, as they were able to participate in various events such as the seminar course and outside events. Since the students of this group had prior coding knowledge, they helped their peers in learning data science skills. Therefore, they also believe that their peers and faculty recognize their data science skills.

The students accepted that the knowledge of data science is important to them, and they intend to work in a data-oriented field. Students agreed that they find themselves competent in applying their skills to the real-world settings. Students also mentioned that they learnt a lot through interaction with the TAs, staff, and faculty as they were very helpful and knowledgeable. All the students in this group lived in the Hillenbrand Hall; therefore, they met people of similar interests and got multiple opportunity to learn together.

In the subsequent paragraphs I have described the journeys of two students Arrow and River in The Data Mine.

8.2.1   Arrow

Arrow is a senior year student in The Data Mine. He is a computer science, data science, and statistics triple major. He chose computer science and data science as his majors because he is passionate about creating algorithms and participating in hackathons. He combines his computer science and data science knowledge to solve complex problems. Also, he believes that data science is a growing field, and it offers him multiple opportunities to apply his artificial intelligence and machine learning knowledge to the real-world problems.

Arrow joined The Data Mine in his freshmen year. He was always passionate about data science. As a freshman, he had to choose a university residence, Arrow found Hillenbrand Hall a perfect fit. Living in Hillenbrand Hall and participating in The Data Mine gave him the opportunity to meet people of similar interests. Arrow finds the connections that he has built in Hillenbrand are very valuable and long lasting.

> I'm a data science major and so, as you know, as being in a data science major like I was just interested in the field already. But also, I just wanted to be surrounded by other residents who are also passionate about the same thing. I would say in The Data Mine is where I met my core friends in college. So, I would say like those connections persisted throughout.

As a part of The Data Mine, he participated in various events. One of them was the seminar course. He found the seminar a great way to get a good head start on the data science projects. Although Arrow had a prior programming background, he found the seminar beneficial as it helped him to learn to code in R and master new statistical skills. Arrow described the seminars as a perfect opportunity to meet and collaborate with like-minded people. Since he had a strong

computational background, he used to help his peers sitting next to him during the seminar, those interactions during the seminar class helped him make some great friends in The Data Mine. As far as developing the technical skills were concerned, Arrow already knew to program in C and Python, but the seminar helped him learn R and Bash. Also, he learned data wrangling, and he believes that understanding data wrangling is crucial. He has often witnessed that data wrangling is mostly taken for granted, but it is a crucial data science skill, and The Data Mine seminars helped him learn that. Arrow also mentioned that working with other students during the seminar helped him become an effective listener and communicate his ideas better. He also acknowledged the role of TAs in the seminar class. Arrow described the seminar TA as one of the important resources to The Data Mine students. He found the TAs very approachable, helpful, and knowledgeable.

> In my freshman year we didn't have TAs in the Data Mine for corporate partners. We only had them for seminars. And so, collaborating with the TAs was always very, very helpful. Uhm, especially going to seminars and being able to ask serious question and especially if you get stuck, they would just nudge in the right direction. We could definitely finish it properly.

Arrow has been a part of The Data Mine since its inception, he has also participated in other activities such as outside events and found them extremely important. He mentioned a specific outside event where a female artist with a strong computational background visited The Data Mine to give a lecture. In her lecture, she described her approach to blending computational skills with artistic skills to create paintings. The artist used Generative Adversarial Networks (GANS) to create artwork. Arrow found this specific outside event very inspiring as he is conducting research in GANS, and he was fascinated to see the interdisciplinary application of GANS. Arrow describes his experience as follows:

> There was this artist that was traveling. I think she stopped by Purdue, and she came to the Data Mine to give a talk, and her work was very interesting because she wasn't a normal artist. She actually had a very computational background, and she used GANS, which are a type of deep learning model to generate art and the way she does that is by training her models on the data set. And then see what sort of new images can be generated, and so I found that was very cool because GANS can be used for a lot of creative tasks and just seeing some of the results that she produced and some of the artwork was kind of inspiring to see the true potential of GANS. Honestly, I would say some of the research that I'm doing today also revolves around GANS and so that was kind of cool to see the power of that.

Arrow was also part of the corporate partners team in his freshman and sophomore years and later moved to the general cohort for his junior and senior years. The corporate project he was involved in focused on anomaly detection on time series data, which allowed Arrow to use machine learning and AI skills that he had developed in the Data Mine. He found corporate partners to be an excellent experience, as the corporate project allowed him to work on a real data set and to use his data science knowledge to solve a real-world problem. Arrow also found the project experience unique as he learned new skills that were not taught in a regular college classroom to a freshman student.

> Yeah, corporate partners is really cool. I think working with companies is a great way to get some real-world experience without like having to like to chase an internship down, and its sort of like balance is your academic work with like real life. I think so. I worked with the corporate partner company. My project was really interesting because I was able to apply a lot of the unsupervised machine learning stuff, but I was very passionate about the data set that they provided. So, I definitely learned so much with that experience. Especially because this project was focusing on anomaly detection on time series data, and that's not something that you learn in your freshmen year of college.

The corporate mentors were very knowledgeable, approachable, and extremely helpful. The corporate project promoted collaboration and allowed him to learn from others and get multiple perspectives to solve a similar problem. The team was very interdisciplinary—the team was comprised of students from computer science, data science, statistics, and biology majors. Arrow says, "It's definitely fun to work with people with different backgrounds, as everyone brings an interesting idea to the table". He also mentioned that since the data was related to heart disease, the inputs from biology students were extremely valuable. The students from computer science, data science, and statistics helped with the computational part of the problem. This kind of collaborative real-world learning experience was entirely new for Arrow, and he found this experience was way different from a regular classroom experience. In his sophomore year, Arrow worked with another company, and he found the project very interesting, but he could not get the data on time and found the corporate mentors were busy, so he felt a bit demotivated. In his junior year, he decided to move to the general cohort and continue with the seminar course as he wanted to focus on more on his junior year coursework.

Currently, he is the Head TA of the seminar course. As a Head TA, he not only helps students during the seminar, but he also coordinates and helps other TAs if they are stuck with some questions. He also creates rubrics for grading seminar projects, and answers Piazza questions. As the Head TA, he developed mentoring, communication skills, people management, and technical skills. After spending a substantial number of years in The Data Mine, Arrow perceives himself as a data scientist as he finds himself competent to apply his data science skills during his internships and corporate partner projects. He also could collaborate with others from various disciplines and work together on problems as a data scientist does. Also, he possesses all the skills required for an entry-level data scientist position.

Arrow believes that joining The Data Mine and living in Hillenbrand Hall was one of the best experiences that he had at Purdue. He made good friends, learned from his peers, connected with faculty and staff at The Data Mine, and grew as an individual. Arrow says:

> As a member of The Data Mine, uh, I feel good to be part of a community that you know could shape my future. I made a lot of friends. I loved working with them, especially as they can bring something to the table so. I learned about like variety of technical skills from my friends, including like you know, SQL, Python. I would say we have a very healthy social interaction with faculty and staff in the Data Mine and with peers and TAs in Hillenbrand Hall and in a very very good environment.

Arrow believes that the opportunities that he got at The Data Mine helped him grow as a person. He agrees that he knew how to code, but The Data Mine gave him opportunities to apply that knowledge to solve a real-world problem. He also developed professional, networking and mentoring skills at The Data Mine, and he believes that those skills are hard to develop as a student. Arrow shares a strong bond with The Data Mine and values the connection that he has built with faculty, staff, and peers at The Data Mine.

8.2.2   River

River is a senior year student in The Data Mine. He is a triple major in chemical engineering, mathematics, and statistics. River started as a Chemical Engineering major with a minor in Mathematics at Purdue. After joining The Data Mine, he developed an interest in statistics and thought if he completes a few more courses, he could obtain a major both in Mathematics and Statistics; therefore, he decided to do a triple major. River did not join The Data Mine in his first

year. Since he lived in Hillenbrand Hall, he met many students in The Data Mine, and he learned about The Data Mine from them. River decided to join The Data Mine and found it an exciting opportunity to learn programming and data science skills that would add to his profile and will make him suitable candidate for a job. River says:

> Uh, so I was really interested in learning more about programming and working with data, and because it's so important and like you know for industry jobs and I thought it would be a set of skills that if I developed, it would make me more hirable. So yeah, so I thought it would be a good thing to join.

River started his journey with the *Research with a Professor* cohort of The Data Mine. He got to participate in various events and described his Data Mine experience as an exceptional experience as he learned to analyze data and gained an in-depth knowledge of data wrangling, modeling, and interpretation. He agrees that he had prior programming knowledge. Still, The Data Mine helped him to develop a broader and deeper knowledge of the domain and learn numerous methods to approach data science problems. He says:

> Data Mine gives you a nice overview of a lot of different topics. It's seeing a lot of different ways of doing the same thing, like, uh, or just a lot of different tools and techniques that exist. So, everything from learning SQL to learning, or how to write a bash or shell script or may be learning code in Python versus R.

River found the seminar course to be an excellent opportunity to learn various programming skills. He learned R, Python, Bash, and SQL from his seminar classes. He loved the in-person seminar classes, as they were more enjoyable, and students met in the dining court of the Hillenbrand Hall to work on problems in an active learning environment. He found that those seminar classes were an excellent platform to ask questions as the TAs and Dr. W were always available to respond to student queries. River says:

> So, I think the nice thing about the seminars, I think it's twofold. One, it's just nice to see a lot of other people working on the same thing as you. You know four or five like TAs and Dr. W all time ready, so immediately if you have a question, you can immediately ask it and resolve the issue.

Apart from the TAs and Dr. W, River found his peers in the Hillenbrand Hall extremely helpful for assisting him with any question related to seminar assignments. He emphasized that it was easy to meet peers and chat with them and get the problems solved. River says:

> We would just talk through some of the seminar problems with the peers. Or like you know, maybe so such and such person is really struggling with number 2 and so you tell them. Oh, look up this function so we just sit up when we had dinner on the floor. So, we'd have dinner and then we just talked about if there were any problems anybody had.

River also participated in the outside events and found them inspiring, as those events allowed him to learn more about how to approach a real-world problem. He mentioned about three events that he found highly insightful. The first one was the talk by Vint Cerf. The second event that he liked was a lecture by the head of the American Statistical Association, where the speaker spoke about the importance of p-values. The third event that he stood out to him was a talk by an artist who used Generative Adversarial Networks (GANS) to create art forms. River says:

> So, I think, especially with like attending those outside events, I think it's definitely broadened my understanding of what is included in data science. And also, it's just giving me a glimpse at all kinds of different fields that are useful in the data sciences, right?

River considered the outside event a perfect opportunity to learn and know more about the field of data science. This is one of the reasons, he believes, that he was very lucky to join The Data Mine. Although, it was mandatory for each student in The Data Mine to attend at least three outside events, River went to more than the required number of outside events every semester as he found them exciting and interesting.

River participated in three different cohorts in The Data Mine. For the first year, he was a part of the *Research with a Professor* cohort and worked on a Physics research project; in the second year he joined the Data Visualization cohort; and in the Fall of 2021, he joined the Statistics cohort. River had a wonderful experience in his Physics research project, as he met a great group of people, and he considers it one of the best experiences at Purdue. He says:

> Yeah, I think that was probably one of the best group projects and group experiences I've had while I've been here at Purdue. Since we lived in the same

building, we could just be like, hey, it's a Saturday. Do you all want to work on this? And everyone would agree.

River and his team members worked on two different projects. For the first project, they created a Generative Adversarial Network to predict a specific distribution of some particles. Then in the next one, they used the quantum adiabatic computer to work on a clustering problem. They also had a peer mentor for the group, he was a graduate student in Physics. River described him very accommodating and knowledgeable. River and group met the peer mentor every Friday and post meeting they all went out for lunch to socialize and get to know each other beyond the project. The group River worked with was very interdisciplinary. There were people from engineering, computer science, and history. River acknowledges that a student from the history major was not great at programming, but he had excellent writing and proofreading skills. So, the group made him responsible for writing formal reports, weekly reports, creating and updating presentations, and creating the poster for the research symposium. River believes that

> If you are interested in learning data science, no matter what your background is, you'll succeed in The Data Mine. The Data Mine is very much like you get in or you get out what you put in type of program.

In the second year, River was in the Data Visualization cohort. It was a decent experience, but he did not find this cohort very organized. He got an opportunity to work on a group project where they had to select a dataset and tell a story through visualizations. River did not find the project very interesting, but the project helped him to learn Tableau. He says:

> The data visualization project was much smaller in scope than the physics project. It took place over a week or two rather than for the duration of the semester, so I didn't get as close with the people in that group. We had to make a case for the implementation of a data visualization class using sets of data we could find. The project was somewhat ill-posed in that there weren't incredibly detailed instructions, but we were able to put something together and bonded over the struggle of finding good data for the presentation. We had to make extensive use of Tableau during the project, so I gained a lot of skills in that software while working on it.

River is a part of the statistics cohort this year, and he is looking forward to that experience. Overall, participating in the cohorts and seminar helped River develop many data science skills.

During the interview, some of the skills that he mentioned were R, Python, and Bash, methods to parse XML and JSON files, machine learning, data visualization and software like Tableau, survival analysis, design of experiments, etc. River considers himself a data science person, as he has the required competencies to apply his data science skills to solve real problems, but he also agrees that he needs more real-world experience to call himself a data scientist. River wants to make a career in the field of computational chemical engineering, River says:

> The sensors in the petroleum industry collect a huge amount of data every single day, and such industries require competent people to analyze that data on a real-time basis make informed decisions.

He mentioned that finding a computationally sound chemical engineer is not easy in industry, but since he has developed those competencies, he can be a great fit. River also mentions various opportunities at The Data Mine helped him to grow professionally. He mentions an incident that he was scrolling through Piazza. He found a summer job posting by Dr. W that mentioned that a Professor from Purdue's Chemical Engineering department was looking for an individual who could help them design a course on programming for Chemical Engineering undergraduate students. River realized he was able to put together a curriculum for the students also identify resources that would be helpful for them. River described his experience as follows:

> Last summer, I was like a mentor for a chemical engineering data science project. And part of my role was mentoring and developing the curriculum that they were going to learn because they needed to learn all the skills, they were going to use in the project within the first two weeks. So, they were chemical engineering students, so they were like junior and senior students, but they didn't have a background in data science. So, they're highly competent in chemical engineering and stuff and so, we have that common language to talk about the thing. And then it was because they were trying to. They were doing a project in chemistry, and so, I was just basically there to help teach them, and then once they learn and also, just act as a resource for them to ask questions too.

River had a wonderful experience being a part of The Data Mine. He felt very included made some great friends in The Data Mine. They learned from each other all the time. River finds the Dr. W and staff at The Data Mine were great. Especially he mentions Dr. W and says, "Dr. W, is kind of like you know he's kind of the figurehead and I think his enthusiasm is kind of

infectious." He also had terrific roommates who were in The Data Mine, and he loved living in the Hillenbrand Hall, he says:

> I think that gives you a sense of belonging in that you know the person you're rooming with, the person in the room, two doors down they're all working on data science projects too. So even if you know nothing about them, you can. You know that there's one thing in common that you have an interest in data science.

Participating in the Data Mine was an amazing experience for River. He grew exponentially and acquired both technical and professional skills. Also, The Data Mine helped him identify his career goal and equipped him to pursue a data-oriented career in the near future.

8.3   Cluster 3: Novice Learners became Corporate Partner TAs

Cluster 3 consists of three students Marion, Jackie, and Jamie. The commonality in these three students is that they all joined The Data Mine in their freshman year and have spent more than 3 years in The Data Mine. They all started as students in the first year of the corporate partners cohort and are now corporate partner Tas, leading and managing their own teams. They all agreed that they had very limited knowledge of data science or coding before joining The Data Mine but participating in various events, seminars, corporate partners, interacting with peers, faculty and staff have helped them to develop a wide variety of data science and professional skills. Overall, they agreed that they grew into experienced members of the learning community.

Table 32 represents the common themes in each category that applied to all the students in Cluster 3. From Table 32, it can be interpreted that for the identity category, students recognized themselves as novice learners, data science people and emerging leaders. They mentioned that they developed data science skills, professional skills, also team management skills as TAs. Students also found that participating in The Data Mine was very beneficial to them as they participated in the Corporate Partnerships, outside events, and the seminar course. They also developed profession skills such as communication, presentation, and people skills. Faculty, staff, and TAs were supportive and interacted with them, helped them to develop data science skills. Students also mentioned that staff and faculty at The Data Mine were very friendly and made them feel included. They also got opportunities to work with peers together on problems in groups, in Hillenbrand

Hall, and were able to learn from one another. As TAs, they also socialized with their team members.

Table 32: Common Codes for Cluster 3

Category	Common Codes
Perception of Benefits/Challenges	Faculty and Staff were supportive and knowledgeable
	Seminar Course helped to develop data science skills
	Networking with mentors/other professionals
	Dr. W is a wonderful Mentor
	TAs were helpful and knowledgeable
	Outside events helped to understand the application of data science
	Corporate Partners Mentors are knowledgeable and supportive
	Corporate Partners provided real-world experience
	Corporate Partnerships is the best experience at Purdue
Perception of Sense of Belonging and Socialization	Worked together in groups on problems and learnt from one another
	Faculty and Staff made me feel connected
	Felt a sense of community or belonging
	Got multiple perspective to solve same problem through interaction with students from various backgrounds
	Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor
	Felt connected with peers and Made Friends
	Socialized with my peers and team
	Lived in Hillenbrand and met people with similar interest
	I felt included
Perception of Identity	I was a novice learner
	I am a Data Science Person
	My peers and faculty recognize my Data Science skills
	I developed Data Science skills
	I developed Professional Skills
	I can apply my Data Science skills to solve real-world problems
	I feel competent in applying Data Science skills
	I plan to pursue a data-oriented career

	Table 32 continued
Perceptions of Leader identity	I am a leader
	I helped student by responding to their queries
	Managed Team
	Served as a Liaison between The Data Mine and Corporate Partnerships
	I conducted office hours to help students with their queries
	I enjoy interacting and learning from other TAs
	I enjoy being a TA

In the subsequent paragraphs I have described the journey of a Corporate Partner TA Marion in The Data Mine.

8.3.1 Marion

Marion is a senior year student at The Data Mine. He is a major in Economics and Statistics. He mentioned that the year he started at Purdue was the first year of the data science program, so he was skeptical about choosing data science as his major. Since he heard many positive points about data science as a field in his high school, so he decided to join The Data Mine in his freshman year. When Marion joined The Data Mine, he had no clue what he would learn. He did not have any coding background, since he loved numbers. He thought he would learn to do some data analysis, as it looked like a good opportunity. Marion was part of the pilot program of The Data Mine. He further adds in The Data Mine pilot program, there were 100 students who all lived in Hillenbrand Hall. Multiple cohorts did not exist back then; all students just participated in the seminar course and worked on the weekly projects. In the very next year that is fall of 2019, he says, "The Data Mine just took off." There were hundreds of students with multiple cohorts, and the Corporate Partnerships cohort was introduced, which turned out to be the biggest success for The Data Mine. Marion joined the Corporate Partnerships as a student in his second year and is now working as a TA for the Corporate Partnerships cohort. Marion mentions that initially when he started his journey at Purdue, he thought of doing economics research in the future, but after being in The Data Mine for four years, he has gained so much of knowledge and data science is so interdisciplinary that he believes he is no longer limited to economics; rather, he can also go into fields like data engineering. Marion credits The Data Mine with transforming him and his career

aspirations. He has learned so much in these four years that he would have never learned if he had not joined The Data Mine. He mentions that he was shy, timid, and a novice before joining The Data Mine, and now he finds himself an experienced and confident professional.

Marion describes his Data Mine journey as an experiential learning opportunity. His first encounter with the data science concepts occurred in the seminar course. He calls it a direct and targeted learning experience, especially for novice learners, that laid the foundation of data science for him. Prior to joining The Data Mine, he had never programmed in any language. He recalls his first day in the seminar class when he found himself clueless. Marion described his experience as follows

> I remember my first day sitting in seminar. Dr. W gave a brief introduction about what The Data Mine is, and then we were right into the first project. And I remember Dr. W working through some bash commands and kind of just doing some really basic analysis within Bash and was navigating through like Google scholars, HPC like Terminal, Dooley, I guess and so. Uhm, I remember looking at that thinking about what is going on. I was so lost I had no clue what was happening. It was way over my head. It was something that I had never experienced and so there were a lot of times where it was like that where I didn't know exactly what was going on and so it was very common for me to be working together in groups and with peers around me. I was very fortunate to have, you know, great roommates that were also interested in the things that I was interested in and so we got the chance to do some great work together. The TAs were also there as a resource for us to ask questions, and of course, Dr. W would always be around as well.

The seminar course helped him understand the logic behind writing codes in R, Python, SQL, and Bash. He learned to look for the relevant resources on GitHub, Google scholar, and other data science platforms to complete the assigned projects. Marion says that his experience in The Data Mine is very different from the students who have joined recently. Marion considered himself an experienced member of The Data Mine, he has seen the learning community growing. He recalls the pilot year when all 100 students, a handful of TA and Dr. W gathered in the dining court, students grabbed their dinners, and worked on the data science problems. In the second year of The Data Mine, as the number of students and TAs increased, seminar class and projects were more organized, the class provided a hands-on experience on solving data science problems. Marion mentions it was delightful to see that every student in the dining court was sitting, enjoying lunch while working on a data science problem. When things went online due to COVID-19, he missed the social interaction during the seminar.

Marion joined the corporate partnerships in 2019 and described working on corporate partner project as an outstanding experience. The corporate partner project was more of a research experience where he worked in a team under the guidance of corporate mentors to solve some real-world problems. In 2019, the Corporate Partners program was a small initiative with a few companies and a handful of students. Marion was one of the students to join corporate partner project in 2019. He found the experience extraordinarily professional and real-life. It was a unique experience for him where he worked on a research project, but in an industry setting. He worked in an interdisciplinary team under the guidance of corporate mentors. Marion worked with a reputed pharmaceutical company where the project was to create a data pipeline for clinical trial data acquisition. The corporate partner project was a great learning opportunity for Marion. He mastered skills related to data acquisition, data architecture, data engineering, visualization, and full-stack development. Marion says that the corporate partner project was a teamwork experience, where every student in the team played a crucial role. Team members divided the task, learned, and taught one another. When they needed any help, they reached out to corporate partner mentors. In 2019 there was no corporate partner TA, so students faced some challenges in reaching out to the corporate partner mentors; therefore, it was an overwhelming situation for some students. The next year, The Data Mine decided to appoint corporate partner TAs. Marion got the opportunity join the first batch of corporate partner TAs. He found it an exciting learning opportunity that helped him to develop data science skills, leadership skills, professional skills, team management skill, and networking skills. Marion's networking skills helped him to develop a good interpersonal relationship with corporate partner mentors and he got a chance to Co-Op with the corporate partner company in the summer of 2020. Being a corporate partner TA, Marion developed professional skills. He was now responsible for the team management, guiding the team members, and was serving as a liaison between the corporate partner mentors and The Data Mine team. Since the corporate project was focused on full-stack development, Marion adopted the agile framework, and he says,

> And so, the way we do that [my team] is we work within the agile methodology. Uhm, and so, uh, we [TAs] are trained on agile specifics. TAs are kind of the drivers for implementing that methodology into our teams. Now some teams implemented agile at a lesser level, but under my personal preferences we go full on agile, and we implement it to a very strong degree, at least on my team because I think it is the easiest way for us to understand and to stay consistent. In our in our development cycles so my role really is kind of is almost kind of like emulating a

scrum master. That's kind of my role as a TA. So, our role is now less of a developer, and we are more as a leader we keep everything on track, and so the way we do that is we work within the agile methodology. Uhm, I'm responsible for conducting Sprint events and making sure that the development team, which is the group of students are on track with their user stories and their task ownership. And then I'm also kind of like that bridge between the corporate partner mentor and the students. Uhm, like that a role is meant to be that communicator between both. So, to be kind of like a resource for each of those groups.

Marion mentioned that the team he is handling as a TA is very interdisciplinary, students are from various backgrounds, but each person brings something unique to the project. For example, students with management backgrounds helped the team create an effective presentation and write the reports. Also, as a TA, he made sure to assign his team members the tasks they were interested in doing. Since Marion has developed a good number of technical skills working on the corporate partner project, he can now guide the students on where to look for resources or write a specific part of code. Marion laughs and mentions that before joining The Data Mine, he hardly knew any of these technical terms, but he still believes that he is still learning and there is so much to learn in data science. Marion believes that he is a data science person and not a data scientist, as he is still learning. He acknowledges that he does not have the competencies of a data scientist as there is much more to learn. He gave an example that with the coding languages such as Python and R, a data scientist can apply those coding skills in multiple ways, but Marion has yet to learn those techniques.

> I mean, right now I would say I'm a data science person. Uhm. I'm certainly not a professional like I have so much to learn when it comes to the data science profession. I have so much to learn it in statistics and that's my undergrad degree. Like I graduate in a few months, and I have so much to learn in the field of statistics, let alone data science and the computational sciences. So, I would say that The Data Mine has really taught me that there's just so many opportunities outside. For example, like simple Python or R. I mean, there's just so many different things that a data scientist can do. And I I'm certainly not there yet. I think that anybody who is an undergrad gonna be data science people, not data scientists just yet, I'm certainly not a professional.

Marion feels confident in applying his data science skills to other classes that he takes. He wants to pursue a career in a data-related field. Before joining The Data Mine, he knew he wanted to pursue a career related to mathematics, statistics, and computer science, but he was not

sure what that would be. But after spending four years in The Data Mine, he knows what career choices he can make.

> Before joining The Data Mine, I knew that I wanted to do something with math, statistics, computers. I didn't know what that was until I had joined The Data Mine and I had learned about all this stuff. Now I can certainly see myself in aid like data science developer role. Whether that's like as a data engineer, a pure data scientist, maybe a machine learning engineer. I think data engineering is certainly a good fit with some of my full-stack experience and data architecture experience. But also like as its statistics major, ML engineer and someone who's performing analysis. I'm sure he's also a good fit so I could see myself in that sort of role.

Marion believes The Data Mine as a learning community is an excellent place for any novice learner to develop data science skills. He finds that every staff and faculty in The Data Mine wants the students to succeed. The Data Mine fosters an "incredibly inclusive environment" and provides an opportunity to learn and socialize with like-minded people. He also mentions he met extraordinary other young individuals when he joined The Data Mine, and now they are best friends. He believes that the relationship that he established at The Data Mine will go a long way, as it allowed him to meet people who were just like him when they all started in The Data Mine. They all faced similar challenges, encountered them together, learned from one another, grew together as matured learners, and are now guiding other young students. Therefore, for Marion, the time spent with The Data Mine is an unforgettable journey. The memories and relationships that he has created in the Data Mine will be cherished forever.

8.4   Cluster 4: All novice programmers developed technical and professional skills

Cluster 4 consists of six students: Kendall, Mickey, Robbie, Orion, Morgan, and Quinn. The commonality among these students is that they all are STEM majors, and all were novice learners prior joining The Data Mine. After joining The Data Mine, they got multiple opportunities to develop their data science skills. They all worked in groups, developing teamwork skills. Four of them joined corporate partners as students, developed team management skills, and became TAs for corporate partners. Two students—Mickey and Kendall—were not TAs but they too developed multiple skills and they had a great experience at The Data Mine. Mickey was a closed off person who spoke and interacted less, but after joining The Data Mine she started to communicate and interact with other members, she feels confident in sharing her ideas in a group setting, and she is

also a member of the Data Mine Advisory Board. Kendall has been in The Data Mine during the past three years. She lives in Hillenbrand Hall and had wonderful opportunities to learn from faculty, peers, and TAs in The Data Mine.

Table 33: Common Codes for Cluster 4

Category	Common Codes
Perception of Benefits/Challenges	Dr. W is a wonderful Mentor
	TAs were helpful and knowledgeable
	Faculty and Staff were supportive and knowledgeable
	Seminar Course helped to develop data science skills
	Graduate Students were helpful
	Outside events helped to understand the application of Data Science
	Corporate Partners provided real-world experience
	Corporate Partners Mentors are knowledgeable and supportive
Perception of Sense of Belonging and Socialization	Worked together in groups on problems and learnt from one another
	Felt a sense of community or belonging
	Felt connected with peers and Made Friends
	Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor
	Got multiple perspective to solve same problem through interaction with students from various backgrounds
	Socialized with my peers and team
	Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor
	I felt Included

Table 33 continued.

Perception of Identity	I was a novice learner
	I am a Data Science Person
	My peers and faculty recognize my Data Science skills
	I can apply my Data Science skills to solve real-world problems
	I developed Data Science skills
	I feel competent in applying Data Science skills
	I plan to pursue a data-oriented career
	I enjoy learning Data Science
	Knowledge of Data Science is important to me

Table 33 represents the common codes for Cluster 4 students. From Table 33, it can be inferred that students were novice learners prior to joining The Data Mine. But the top codes highlight that student reported themselves as Data Science persons as they developed data science skills, are competent in applying data science concepts to solve real and interdisciplinary problems, and plan to pursue a data-related career. All the students in this clusters mentioned about the mentoring they received by Dr. W and that helped them to learn data science skills. They also got opportunity to interact with the TAs, graduate students, staff, and they found them knowledgeable and helpful. All students described seminar courses and outside events as important; these helped them not just to develop data science skills but also to learn more about the field. Since all the students in this cluster were in the corporate partners cohort, the majority of them had a great experience. Students also described that they felt included in The Data Mine and got multiple opportunities to learn from one another, felt a sense of belonging, and made friends within The Data Mine. In the subsequent paragraphs I have described the journey of Kendall in The Data Mine.

8.4.1 Kendall

Kendall is a junior year student in The Data Mine. She joined The Data Mine in her freshmen year and is majoring in Data Science and Statistics. Kendall started her Purdue journey with a Data Science major, and later she figured that if she enrolled in a few more classes, she

could get an Applied Statistics major, so she decided to double major. Before joining Purdue, Kendall was browsing for some courses related to data science, and The Data Mine accidentally came up as a search result. Kendall was fascinated by the description of The Data Mine and was excited about the learning community aspect, and she thought this was an excellent opportunity for her. She says:

> So actually, when I was researching data science because it wasn't something I knew very much about, and Purdue Data Mine popped up as like a research result. To be very honest, when I was like Oh my goodness, it sounds like a great opportunity. I thought I should do it. And then I looked at myself and I realized, the kind of mentorship that I would receive, and the fact that it would be an introduction to coding for me because I had never coded before, so that's really what encouraged me to join and then other motivating factors was the learning community, I was excited to have a community of people who would have been having the same interests I did, so that was kind of a second reason, primarily, because I really wanted that introduction to coding.

Kendall started her The Data Mine journey with the Statistics cohort, as she felt that she was more a numbers person. She perceived the statistics cohort to be easy and focused on solving questions related to statistical concepts such as measures of central tendency, normality, etc. She was surprised by the learning opportunities that the seminar classes provided her. She says, "the classes went beyond my expectations, and I learnt so much." She learned to code in R and Python in her classes, and she believes that this is something unique, as compared to a traditional undergraduate statistics course. She agrees that it was a little challenging for her as she did not have prior coding background, but she loved how Dr. W allowed students to play around with data and TAs were available to answer any questions. Since all the seminar projects were based on real data, Kendall describes one of the seminar projects based on an Amazon music review. She compared reviews about the singers from the United States and reviews about the singers from Trinidad and Tobago. She found that reviews written by people from Trinidad and Tobago were longer, and people demonstrated a love for music. Since she found the project interesting, she did some text mining to deep dive more into the music reviews. She felt that the seminar class was a great introduction to data science for any novice learner. The weekly project helped her think like a data scientist and develop data science skills. Kendall believes that the skills she has developed in her seminar course have helped her to get a research assistantship in a psychology lab. Kendall proudly says: "I love to use the skills that I developed at The Data Mine in my research lab."

Kendall mentioned that, prior to the COVID-19 pandemic, the seminar class used to take place in Hillenbrand dining court, where a large group of students were seen working on the weekly projects. She described how even after the seminar was over, students used to meet each other in Hillenbrand, and they discussed and solved data science problems. Students also participated in the floor events such as karaoke night, movie night, ice-cream socials and socialized with other students. Apart from the seminar course, Kendall also participated in the outside events at The Data Mine. She found the outside events very inspirational and motivating, as the events helped her understand the application of data science in a real-world setting. She explicitly mentioned an outside event that demonstrated the application of data science in the health care industry. She mentioned participating in the outside events inspired her career goals, she says:

> I thought that [outside event] was just incredible, it was like a year ago of pandemic and I still think about it and that really kind of gave me a lot of inspiration for what I want to do and what I want to at least look into and the kind of work. I think it is fascinating and I would enjoy doing later on in my career.

Kendall described changes that took place in The Data Mine during the COVID-19. She describes that the seminar class went online, students used to meet on zoom, Dr. W used to help the students, but students no longer worked together in groups as they met before COVID-19. Since she was so used to working with her friends on seminar problems, her group made it a rule to meet online over lunch and work the seminar projects. Kendall also mentioned that Dr. W also came up with an Examples Book which was very helpful for the students in completing their seminar projects. She also mentioned that Piazza was a great way to ask questions or share any updates during the pandemic. The outside events also took place online, but the good thing was that they were recorded, so if someone missed it, they could watch it later.

Kendall also joined Corporate Partnerships and worked with a leading food and beverage company in her sophomore year. She was a part of an interdisciplinary team comprising of undergraduate and graduate students from diverse backgrounds. A Corporate Partner TA led the team and there were two corporate partner mentors. The project in the first semester was very broad, that required the team had to web scrape various social media websites to collect customer reviews and then, using Natural Language Processing, students were supposed to analyze the data to propose some recommendations. Since the project goal was comprehensive and it was challenging to collect data from various websites, the corporate partner mentors suggested

focusing on Amazon reviews. Kendall also mentioned about an incident when the corporate partners changed the project scope at the end of Fall semester and student team felt very demotivated. But, overall, Kendall found the experience very rewarding as she was able to work in a team and learn from one another. She mentioned that the Corporate Partner TA was very efficient very knowledgeable in computer science and data science disciplines. Also, he did a great job in managing the team and answering the questions. She found that having graduate students on the team was extremely valuable as she learned a lot from them. Kendall mentions that one of the graduate students had excellent writing and presenting skills. She used to work in advance and prepare a presentation on behalf of the team and present it to the corporate mentors. Kendall considers the graduate student an informal leader for the team as she took a lot of ownership of the project.

Kendall believes that The Data Mine helped her to grow as a person, and she says that:

I definitely grew up in a lot of different ways, so there's the most obvious like, well I have the skill set that I didn't have before, that is the coding skills, so there's growth in that sense. But I also think. I've also learned mentorship in a kind of way because I've worked on a team of 20 people. I feel confident in mentoring people because we have like this big Data Mine discord right now that was made during the pandemic, that it doesn't get used as much as Piazza but There were some freshmen who are asking questions about like what The Data Mine is like, before they entered and I felt comfortable enough to kind of reach out to them and say hey, this is what The Data Mine does. Kind of try to give them some guidance in that regard, so I do think I've also learned some kind of mentorship or leadership skills as   well.
So that's another way I grew, and that in terms of like clarifying my professional path or uhm, presentation skills or the soft skills that I developed. I definitely grew a lot.

Even after learning so much, Kendall believes that she has more to learn, and The Data Mine gave her a good head start. She perceives herself as a data science person because she is now competent in cleaning, merging, analyzing data sets using data science techniques. The seminar classes and Dr. W have helped her develop a curiosity with data and a passion for finding solutions. However, there is still a lot more to explore, and she is confident to develop those skills in near future and develop a data-oriented career. Kendall wants to join the health care industry and work in computational biology. She gives credit to the outside event that demonstrated the use of data

science to solve health care problems. Post attending that event Kendall was very much fascinated by the application of data science in healthcare.

Kendall believes that being a part of The Data Mine is one of the best experiences at Purdue. Kendall feels that The Data Mine is like her home. She says, "I not just only belong here but I feel actively included." Kendall made some best friends in The Data Mine; she says that faculty and staff in The Data Mine know the needs of each student personally; she gives her example:

> I feel like a sense of belonging in The Data Mine it is evident in the fact that Dr. W and Ms. E, have always and also actively include me and lookout for opportunities for me and make sure that I am not only just doing well as a student, or even just as a person, but that I have a lot of resources at my disposal, so I would say in that way I'm also actively included in the Data Mine.

Kendall believes that the interactive and collaborative aspects of Data Mine sets it apart from any other traditional classroom in Purdue. The students learn so much from their peers just through interaction. Kendall agrees that COVID-19 has impacted the social aspect of The Data Mine. Still, faculty and staff at The Data Mine are trying their best to bring in new opportunities that facilitate collaboration and interaction.

8.5 Cluster 5: All girls grew into experienced members of The Data Mine

Cluster 5 consists of five students: Harley, Riley, Skylar, Frankie, and Rio. The commonality among these students is that they all are female students. Cluster 5 is the only cluster that includes three students from non-stem majors. It is noteworthy that all students in The Data Mine get equal opportunity irrespective of their background. In this cluster four students were corporate partner TAs, and two out of four students were from non-STEM backgrounds. Also, all the students in this clusters were novice learners and each of them developed a variety of technical, non-technical and professional skills and want to pursue a data-oriented career. Table 34 represents the common codes for Cluster 5.

Table 34: Common Codes for Cluster 5

Category	Common Codes
Perception of Benefits/Challenges	Faculty and Staff were supportive and knowledgeable
	Seminar course helped to develop data science skills
	TAs were helpful and knowledgeable
	Corporate Partners provided real-world experience
	Corporate Partners Mentors were wonderful
Perception of Sense of Belonging and Socialization	Worked together in groups on problems and learnt from one another
	Got multiple perspective to solve same problem through interaction with students from various backgrounds
	Felt a sense of community or belonging
	I felt included
	Worked in Interdisciplinary Teams for Corporate Projects and learnt division of labor
Perception of Identity	I was a novice learner
	I am a Data Science Person
	I Developed Data Science skills
	I can apply my Data Science skills to solve real-world problems
	I Developed Professional Skills
	I Developed research skills
	Knowledge of Data Science is important to me
	I plan to pursue a data-oriented career
	I Enjoy learning data science

Table 34 represents the common codes for the students in Cluster 5. The uniqueness about this cluster is that it has no computer science, statistics, data science, or engineering background students. Three students are from non-STEM backgrounds as well. From Table 34, it is evident that all the students identified themselves as novice learners before joining The Data Mine. It is great to see how all the students reported themselves as data science people and have developed data science skills, research skills, and professional skills after joining The Data Mine. The

students also mentioned the support they received from faculty, TAs, and staff at TDM and how participating in the seminar and outside events helped them develop data science skills. Students also got opportunities to work in teams and described faculty, peers, staff as very accepting. Since all the students were part of corporate partnerships, they all found corporate projects similar to real-world experiences and described the corporate mentors knowledgeable and helpful. In the subsequent paragraphs I have described the journey of Riley in The Data Mine.

8.5.1   Riley

Riley is a junior year student in The Data Mine. She joined The Data Mine in her sophomore year. Riley is majoring in finance. She recalls an event when she was going through her Outlook mailbox, and suddenly, she came across an email regarding The Data Mine. Riley considers college to be a perfect place to explore and participate in various activities. Joining The Data Mine was challenging for her, but she thought it was something she should do, as she wanted to go beyond Excel sheets and learn something new. She also thought that participating in The Data Mine, specifically in Corporate Partners, would help her learn computational and people skills that will differentiate her from her peers in finance. After joining The Data Mine, she says that *"The Data Mine was not just about coding it was much more than that."* Riley described the seminar classes as the building block for mastering data science concepts. Since she joined The Data Mine during the pandemic, the classes were conducted online. She described Dr. W as a great mentor and faculty member who makes data science easy for even those students who have never coded. Dr. W made the seminar classes very straightforward for all students to follow along. Riley also used the Examples Book (2021) while working on seminar projects and she described it as a very beneficial resource. In Fall 2021, once the classes went back to normal and were in-person, Riley felt it was a different experience. The in-person seminar class took place in the dining court of Hillenbrand Hall. The TAs and Dr. W guided the students and answered students' questions; it was way different from how it happened online. Since Riley did not have prior coding knowledge, she experienced a few challenges with coding. Still, she found TAs and Dr. W very approachable to discuss any problem and get that sorted. She says,

> Every challenge is very specific. Like everyone knows of coding. So, I guess yeah like you said, like in general troubleshooting. But also, like beginning to know some programming language and moving to another is difficult. So, like even as a second-

year student in The Data Mine it was a little adapting, but they're [Dr. W and TAs] all very helpful and Dr. W and the team make the projects themselves so he's very like, you know he knows what he's talking about. Dr. W could be very descriptive and like, okay, you didn't do this, or like oh so you have to do this. He actually helped me this Monday. Yesterday with something like trivial, but obviously it made a big difference in the project. So just like small things like that to add up, and if you do need help and you got it, they are very happy for you. They're just here to help anyone that needs, which is great.

Riley joined The Data Mine intending to be a part of the corporate partnerships. She worked with a reputed company. She finds that experience as one of the best experiences of her life. Riley learnt so much from her Corporate Partnerships cohort. She mentioned that mentors were supportive and knowledgeable. Riley finds corporate partnerships very well structured. She started as a student, but later on, she got an opportunity to become a Co-TA for the corporate partnerships team. She also finds her Co-TA Harry very knowledgeable, and she learned a lot from him. Riley found the corporate partner experience enriching, and it is very different from a regular college classroom. Riley described her experience of being a student and now being a TA for corporate partners. Riley says

I loved being a student. It was great because you're really involved. It's not a normal class where you sit back, and you just take notes and take an exam. And that's what I love about The Data Mine even as a TA. Uhm, that. Like you know, it's not just. I'm not just taking exams and learning that way because you're doing like real research, which is amazing, so but being in that role is different, but it's like so I love it so much more because I do get to talk more with the mentors and from Corporate Partner Company and more with Ms. M and Dr. W and other administrators within The Data Mine, which is great.

Riley developed plenty of technical skills by working on corporate projects, such as data visualization and Python programming. She said that data visualization was the core of the project. Therefore, Riley learned to use specific software explicitly created for research by the corporate partner company. She also learned how to look for the right resources, for example, the use of GitHub. Riley mentions that since she did not have prior coding knowledge, she sometimes feels intimidated when it comes to programming, but she agrees that she grew into a mature learner within a year, and she feels that she has skills that other finance majors don't have. Riley says

I would say I have become more confident. And it's a big change because last year I had no clue what I was doing. So, in terms of coding, uh, I'm not crazy confident. I'm not crazy like skillful in that set up, but I think I always like to look on the bright side and think of I mean compared to other maybe other finance students or business students. I'm learning more, so if anything, I should be like you know, oh, wow, this is good for me. Even if I'm just learning small things and stuff like that, so yeah.

Being a TA was an excellent opportunity for Riley, as she developed people skills, team management skills, networking skills and mentoring skills. She also found that being a TA for corporate partners was challenging. There were no formal assignments or exams to evaluate student performance, so sometimes it was challenging to keep the students motivated on the project, and sometimes students procrastinated on the work. As a TA, she is answerable to the corporate partners; therefore, one of her roles as TA is to keep the team's motivation high, be well planned and drive the team along with her Co-TA to achieve the desired goals. She also feels that, being a TA, her responsibility is also to learn the required software for the project, and she spends a substantial amount of time learning Python and answering questions that the team might have. As a TA, Riley has felt that she has developed corporate skills. She said,

I really feel like not to say I'm working a nine to five or anything like that, but I really feel like we're part of the company almost. This is how I would want a job to be, and I think it's exciting to think about the future like that because working with people from the corporate partner company and getting all this organizing stuff. It really does help me in the sense that like yes, other leadership positions have helped me. Like being a leader and being charged with things and be organized and anything like that.

Riley also developed leadership skills by working on the projects with the corporate partner. She finds leading a team is not easy, as taught in the textbook. It requires a lot of effort and commitment from the leader's side. As a Co-TA, Riley believes that being organized as a leader is essential, so she described her week as follows: the team met twice in the week, Tuesday, and Thursday. Tuesday is an online meeting with the corporate partner mentors where the team presents the work to the mentors, and mentors provide feedback and answer any questions that teams have regarding the project. On Thursday, just the team meets, it is a two-hour lab session where students work on the project, ask questions, and interact with one another. It was in-person in the Fall of 2021, so students had a great time exploring new concepts, working in teams, and

learning from one another. Apart from these meetings, the team has a Slack channel where they interact with one another in case of any questions or provide feedback or updates on the project. Riley and his Co-TA are also responsible for answering Slack messages and emails regarding the project. Riley feels that in The Data Mine she is learning every single moment; she feels that being a TA does not mean that she will stop learning.

> I am a student who keeps learning even though I do have a leadership position, that does not mean like I am all knowing. I think I just mentioned that to the students in my corporate partners team. I have learned that from Dr. W himself. He's always saying that I am learning, so he's learning himself and it's important to hear that as a student, I think because like sometimes you like, get caught up and like oh my teacher knows everything, or this professor is a genius. And then so when he mentions that he is learning, you feel more comfortable and feel that learning is lifelong process.

Riley mentioned that after spending a substantial amount of time at The Data Mine. She feels that she has developed a bouquet of technical and professional skills. She finds herself very different from other finance majors. She jokingly said that she was thinking of changing her major from finance to data science at one-point last year. Now, when she looks at her resume, she finds fewer finance skills and more data science skills; this is something she never thought of in her life. She credits her growth in terms of data science and leadership skills to Dr. W and Ms. M, and she says they are great mentors who have guided her always. Today Riley feels confident in applying data science skills is because of Dr. W's mentorship, and she got the TA opportunity for corporate partners because of Ms. M. They both have helped her in every possible way to become a better person. She says,

> I've only known them [Ms. M and Dr. W] for two years, but they're always so welcoming and they don't necessarily feel like you know like a boss, or like someone who's like very in charge of everything, and you never see them, something like that. They really do feel like they are a part of the team, and they want to help you and you are just like them, at the same level and that makes me like blows my mind.

Riley reflected on her journey from when she started in The Data Mine, and she mentioned that if somebody had asked her to rate her knowledge of data science skills, she would have given a zero or one. But now, Riley feels that The Data Mine has made her feel so competent that she

can give herself a six or seven for data science skills. Riley laughed and said that she never thought that she would do a data engineering internship in her life, and she did one last summer. She worked with a renowned company last summer on a data engineering project, which helped her to build a robust technical portfolio. Riley believed that prior to joining The Data Mine, she described herself as a finance major and finance person, but now she has other aspects added to her identity. She identifies herself as a finance person who knows data science and preferably wants to be called a data science person and an emerging leader, as she is managing the corporate partners team. Riley never took a Statistics class before joining The Data Mine, but she is taking one now, and she feels very competent in solving statistics problems using R, as she learned R in her first year of seminar. Riley also mentioned that being from a non-STEM background, she sometimes experiences imposter syndrome, especially when she is stuck while solving a problem. She finds her peers in The Data Mine very helpful as they always help her to learn new concepts and help her incase if she is stuck. Riley is very happy about her progress as a non-STEM major and she loves learning new concepts so currently she is learning Bash, as she believes that there is a lot to learn regarding data science. Riley feels that knowledge of data science is essential to her major. She uses data visualization tools to visualize the data, and she feels that knowledge of machine learning is crucial if someone is working with stock market data. Riley wants to pursue a career that involves finance and data science. She said, "finding a finance major with data science skills is a rare combination, and industries would love to hire a person with such a rare background."

Riley is grateful to The Data Mine for providing her with great opportunities to progress in her undergraduate life. She loved the inclusive environment at The Data Mine, and she never felt that she was an outlier as she was not a data science or computer science major. The faculty, staff, and peers embraced her like one of them. She made some great friends at The Data Mine; one of them is her Co-TA, Harry. She learned a lot from him; he is highly knowledgeable and a great mentor. She believes that any student interested in data science joining The Data Mine will develop data science skills, just being in The Data Mine, participating in the seminar, and learning from peers and faculty. She also mentioned that being in The Data Mine is one the best experience she ever had at Purdue.

8.6   Summary

The intent of this chapter was to understand the growth journey of the students in The Data Mine. Since the results of the thematic analysis revealed that many students shared similar growth stories therefore, I wanted to group the students with similar experiences together. To perform the grouping, the hierarchical clustering was used and the clustering algorithm returned five clusters. The first cluster consisted of one student Finley, an international student who joined The Data Mine in the Fall of 2020. He perceived himself as a mediocre level of programmer. But he did not have any research or data science skills prior to joining The Data Mine. The Data Mine allowed Finely to develop data science and research skills, find internship opportunities, and become a seminar TA. These opportunities and participation helped Finley to grow exponentially in The Data Mine. Cluster two consisted of four students, and the commonality among the four students was that all the students had a prior coding background. All the students develop many skills such as data science, research, and networking skills. All these students lived in Hillenbrand Hall. They got enough opportunity to socialize with peers, faculty and staff. Since they had prior coding skills they also served as mentor and helped peers around them with coding problems. Moreover, two students from this cluster evolved into student TAs. Arrow served as the Head TA for the seminar course, and Tommie is a corporate partner TA. The third cluster consisted of three students; all of these students have been in The Data Mine for more than three years. Marion was part of The Data Mine from its inception year. Marion, Jackie, and Jamie started The Data Mine as novice learners with no data science or programming background. They all got multiple opportunities to grow, develop skills and evolve into corporate partner TAs. Jamie also got opportunities to serve as President and Vice-President of the Data Mine Advisory Board. Cluster four consisted of six students, and all six students identified themselves as novice learners before joining The Data Mine. Out of six students, four became corporate partner TAs they are Robbie, Orion, Quinn, and Morgan. Whereas Kendall and Mickey were not corporate partners TAs but they too grew interms of data science and programming knowledge. Each student in cluster four acknowledged that they got multiple opportunities to grow and develop into mature learners. The fifth cluster consisted of five students. All the students in this cluster identified themselves as female and novice learners. Out of the five students, three were from the non-STEM background they were Harley, Skylar, and Riley. They all mentioned that they felt intimated initially when they joined The Data Mine as they did not have prior coding knowledge. But it was important to note that all the three students

developed coding and data science knowledge. Skylar and Riley became corporate partner TAs, whereas Harley got admit from a reputed public university in Information Systems program. Rio and Frankie were the other two students in cluster five. They both are from STEM background, but they too identified themselves as novice learners and found multiple opportunities to grow, learn and evolve as corporate partner TAs.

Students from each cluster demonstrated a growth as they progressed in The Data Mine. The story of their journey and experiences in The Data Mine reveals that they did not just grow in terms of data science skills, but they developed skills, such as research skills, mentoring experience, networking skills, and leadership skills. The bouquet of skills that they developed make students well equipped for any industry job and graduate school. The majority of students felt that The Data Mine was way different from a traditional classroom experience, and therefore they described it as an exceptional and best experience that they had at Purdue University.

CHAPTER 9: DISCUSSION, IMPLICATIONS AND CONCLUSION

This chapter discusses the results of this dissertation study. The study followed a sequential explanatory mixed methods design to answer the following research questions:

Quantitative: RQ 1: What are students' perceptions regarding their identity formation, socialization opportunities, self-belief, and academic/intellectual development in The Data Mine?

Qualitative: Guiding RQ 2: How do students' participation in activities and interaction with peers, faculty, staff at The Data Mine contribute to becoming an experienced member of the learning community?

- **Sub-RQ 2(a):** What are the perceived benefits and challenges of participating in the Data Mine?

- **Sub-RQ 2(b):** How do students describe their levels of socialization and a sense of belonging within The Data Mine?

- **Sub-RQ 2(c):** How do students' participation and interaction in The Data Mine help them form their identity?

Aligned with an explanatory mixed-method design, the quantitative study was exploratory, whereas the qualitative study was the center of the study. The quantitative study intended to understand students' perceptions regarding their identity formation, socialization opportunities, self-belief, and academic/intellectual development as they progressed through The Data Mine. To conduct the study, a pre-post survey was conducted. The study's overall results revealed that students increased the perception of data science identity formation, socialization, sense of belonging, self-concept, and academic/intellectual development. Students also demonstrated decreased data science anxiety and a data science fixed mindset. The quantitative study results were very motivating, which led to the second study to understand the underlying factors that led the students to report an increase or decrease in the constructs mentioned above. For this, I conducted a qualitative study. I created an interview protocol and purposefully selected the participants to conduct in-depth semi-structured interviews based on the results from the quantitative study. The average length of each interview was 1 hour 10 minutes. The study

intended to first answer the sub-research questions (i.e., Sub-RQ2 a, b, and c) by conducting thematic analysis.

Further narrative analysis was conducted to answer the guiding research question (i.e., Guiding RQ2). I plan to discuss the study's findings, combining the results that I obtained from a quantitative and qualitative study. Therefore, in this chapter, the discussion is divided into the following sub-sections: 1) perceptions of benefits, 2) perceptions of challenges, 3) perceptions of socialization and sense of belonging, 4) perceptions of identity formation, 5) growing into an experienced member of the learning community.

9.1   Perception of Benefits

The result of the study demonstrated that students found participating in The Data Mine a valuable experience, and they reported several benefits over challenges. The thematic analysis for benefits of participating in The Data Mine helped me to identify three themes, that were identified from the students' interviews. The first theme focused on the various opportunities The Data Mine provided to students to develop data science skills. The students, during the interview, mentioned the role of mentors, faculty, peer-TAs, graduate students in helping them develop data science skills. They mentioned how interacting with these experts and reaching out to them with any questions helped them develop data science skills. Studies by (Crisp et al., 2017; Gutierrez, 2012) have demonstrated that mentoring is critical, especially for undergraduate students. It allows them to develop professional and leadership skills, improve academic performance, and lead to retention. In the context of the learning community, senior members or mentors played a vital role. Being in a living and learning environment, they (1) helped the students to grow and develop various skills, (2) provided an opportunity to interact and collaborate, and (3) offered emotional support to mentees as and when required (Leidenfrost et al., 2014). The study conducted by (Hessenauer & Law, 2017) demonstrated that students in the learning community benefitted from the mentor-mentee relationship, faculty mentors and peer mentors helped the students to develop professional and research skills, guided them throughout their undergraduate journey, provided them career guidance and developed a strong bond, and that helped students to develop a sense of belonging.

The results of Hessenauer and Law's study align well with the context of The Data Mine. The mentorship occurs through interaction and collaboration; The Data Mine provides multiple

opportunities to the students to interact and collaborate with faculty, peer-TAs, graduate students, and staff in The Data Mine. The Data Mine has also implemented Piazza for students, faculty, TAs, and staff to collaborate, ask questions, share updates or challenges and foster interaction and collaboration. The mentorship provided by the faculty, peer-TAs, graduate students, and staff is informal, meaning that it occurred when a student reached out informally in-person or posted a query on Piazza to a peer-TA, graduate student, faculty, or staff in The Data Mine in case of any questions. Interaction among the students and peer-TA, graduate student, faculty or staff is frequent and common at The Data Mine. It is important to note that interaction is key to the informal relationship as the continuous interaction with the mentor allows the mentee to build trust, and constant communication helps them better understand (Mullen, 2007).

The Data Mine also provided multiple opportunities to participate in various events such as seminar classes, outside events, and working on real corporate projects. Participation in these events helped the students to develop data science skills. The seminar class is mandatory for all the first-year students in The Data Mine; the objective of the seminar class is to lay the foundation for data science. The study by Jaijairam (2016) revealed that seminar courses play a vital role, especially for the first-year students, as it (a) offered content in a specific area, (b) helped students to develop the foundational skills related to that particular area of interest, and (c) improved retention as it helped students to meet other first-year students and develop a sense of belonging. The seminar class at The Data Mine followed an active learning form of instruction where students were engaged in the problem-solving environment while working alongside their peers, faculty, and TAs (Hoffman et al., 2002; Jaijairam, 2016). Engaging with other students in an active learning environment allowed the students to meet students of similar interest in The Data Mine. Similar results were observed in Jaijairam's (2016) study that attending the seminar class allowed students to meet students of similar interests. Working in an active learning environment also helped the students in The Data Mine to develop a higher level of cognitive thinking and divergent thinking skills. The study by Gergorić (2020) revealed that engaging students in active learning allows them to critically think, reflect and communicate their ideas to peers and faculty.

The second theme of benefits focused on the corporate partner cohort of The Data Mine. The students worked in teams on corporate partner projects on real-world data. The corporate partnership is an experiential learning opportunity for the students to work for a corporate company and solve an authentic problem. Engaging students in experiential project-based learning and

working in groups helped them share their perspectives and ideas and learn complex concepts quickly by engaging with their peers. Similar results were observed in the study conducted by Fini et al., (2018). Project-based learning allowed the students in The Data Mine to apply their disciplinary knowledge to solve real-world problems. Along similar lines, the study by Asonitou (2013) demonstrated the importance of project-based learning in helping students translate their disciplinary knowledge to real-world knowledge. Working on and solving real problems motivated the students by making them interested in the discipline and improving their confidence in applying their skills (Blumenfeld et al., 1991). Each corporate partner team was led by peer mentors, which provided the peer mentor opportunity to develop leadership skills, professional skills, people management skills, and team management skills. Students worked in teams in the corporate partner project. They developed data science skills, but they also got the opportunity to develop other essential skills such as teamwork, communication, critical thinking, listening, and presentation skills (Jaiswal, Lyon, Magana, et al., 2021). It is important to note that The Data Mine corporate partner team followed the agile scrum framework. The peer mentor, also known as the corporate partner TA, took the role of the scrum master, the students were the development team, and the corporate partners were the product owners (Examples Book, 2021). Applying scrum principles to manage a team was an effective and efficient approach. For example, a study by Karabiyik et al. (2020) conducted in a higher education setting demonstrated that the scrum master played a vital role in managing a scrum team. The study was conducted in an undergraduate level classroom where students worked in teams and followed the agile scrum framework. The study results suggested that if the scrum master was efficient in leading the team, the team demonstrated a high team effectiveness that allowed them to set practical goals, assign roles, develop interpersonal relationships, and design efficient group processes. Although the application of the agile scrum framework was limited in a higher education setting, prior studies (Jurado-Navas & Munoz-Luna, 2017; Karabiyik et al., 2020) identified great benefits in implementing the scrum framework and principles for effective team management.

In the third theme of benefits, students mentioned the internships, networking, and research opportunities The Data Mine provided. Offering internship opportunities to the students aids the universities in addressing the issue of skill gap (Hora et al., 2017). Internships are mutually beneficial to undergraduate students and employers as they help the undergraduate students to apply their disciplinary skills in a practical context under the supervision of an expert, and it aids

employers to find inexpensive resources and potential future employees (Callanan & Benzing, 2004; Hora et al., 2017). Studies by (Beard & Morton, 1998; Binder et al., 2015; Cook et al., 2004; Hora et al., 2017) described the numerous benefits of doing an internship. Students who do internships (1) have a higher chance of getting hired than students who did not do internships, (2) are more competent to apply their theoretical knowledge to a practical context, (3) develop soft skills such as interpersonal, communication, and leadership skills, and (4) demonstrate a higher level of job satisfaction and have a solid professional network. Students in The Data Mine had multiple opportunities to do internships, and faculty and staff at The Data Mine made students aware of various internships opportunities available in the market. The Data Mine also provided research opportunities. Engaging undergraduate students in the research is one of the most impactful pedagogical practices (Baron et al., 2020). Undergraduate students involved in research are most likely to graduate on time and have fewer chances of dropping out of the program (Seymour et al., 2004). Also, students who participate in research develop research skills that demonstrate academic success and personal growth (Kuh, 2001). The Data Mine provided various research opportunities. For instance, during the interview, students mentioned multiple research opportunities. The first one was with corporate partners; every student in the corporate partner cohort researched the problem assigned. They conducted a literature review to understand the background. The second type of opportunity was that students in The Data Mine could also join Research with a Professor-led cohort and engage with faculty at Purdue to conduct research. For example, one student was part of a Physics research project. He was a Chemical Engineering graduate and was engaged in particle physics research. The third research opportunity that students mentioned was that few students came across a research opportunity on the Piazza forum. Allowing the students to participate in research The Data Mine helped them develop a research portfolio and motivated them to pursue a research-oriented career. The Data Mine also provided networking opportunities to the students by connecting with their corporate mentors or networking with the speakers during or after the outside events. The capability to network or connectedness capability was an essential element that contributed to the students getting hired (Bridgstock, 2017). When provided with the opportunity to network, students need to interact actively and collaborate with their professional connections and draw value from their social capital (English et al., 2021). Social capital, in this case, is the network of professionals that students have created by utilizing the networking opportunities in The Data Mine. Students mentioned how the network

created through The Data Mine helped them get internships and co-ops during the interviews. The students have found The Data Mine very different from a traditional classroom. It allows them not only to develop data science skills but also to develop skills such as research, networking, and teamwork that will make them strong candidates for industry careers.

9.2 Perception of Challenges

The overall result of the analysis revealed that students experienced multiple opportunities that helped them gain knowledge and develop several technical, non-technical, and professional skills by participating in The Data Mine. Still, few students reported some challenges with a few aspects of The Data Mine. The two categories for challenges were: (1) challenges with corporate partner cohort and (2) challenges faced by non-STEM background students. Challenges with corporate partners were related to data sharing and some mentorship concerns. A few students reported that the corporate partners could not provide the data on time.

In some cases, corporate partner mentors were busy and could not guide the student groups; these challenges delayed the project and led to demotivation in students. A study conducted by (Jaiswal, Lyon, Magana, et al., 2021) also identified similar challenges with the corporate partner cohort of The Data Mine. Due to internal or legal challenges, corporate partners did not provide the data on time. Since The Data Mine follows a scrum framework for corporate projects and the corporate partner is also the stakeholder, studies (Drury et al., 2012; Hajdarevic, 2018) have revealed that conflicting priorities of the stakeholder in the scrum team is one of the critical reasons that leads to poor decision making and inefficiency in a scrum team. Looking carefully at the second challenge reported by students, corporate partners were busy, mentors had time constraints, and could not attend to mentees' needs. This is an example of unproductive mentoring. It impacts the mentee's motivation level adversely (Eby & Allen, 2002). Even though very few students reported challenges related to corporate partners, The Data Mine needs to address these challenges that impacted student motivation.

Non-STEM background students reported the other type of challenge. Three female students from non-STEM backgrounds reported feeling intimidated when they joined The Data Mine as they did not have a prior coding background. They also felt less confident when they got stuck while solving a data science problem. Studies by (Betz & Hackett, 1983; Pajares & Miller, 1994; Walker, 2018) have demonstrated that despite having the mathematical or computational

aptitude, women students tend to feel less confident in math or related computational fields. It is important to note that pursuing a STEM-related career is not restricted to only STEM majors. Even a student from a non-STEM background may feel competent and demonstrate interest to pursue a STEM career (Su et al., 2009; Wang & Degol, 2017). Therefore the faculty must motivate students and help them develop self-efficacy to study and work in STEM (Grata, 2019). In The Data Mine, faculty and staff keep the students motivated and provide them equal opportunities irrespective of their academic background.

The discussion regarding the benefits and challenges perception the key takeaway was that students identified more benefits participating in The Data Mine, leading to a holistic development. However, students faced some challenges which needed to be addressed.

9.3   Perception of Socialization and Sense of Belonging

The qualitative and quantitative results for socialization, sense of belonging, and academic/intellectual development demonstrated that students agreed that they got multiple opportunities to socialize with peers, faculty, and staff, which improved their sense of belonging and academic/intellectual development.

The quantitative study results demonstrated that students experienced an increase in socialization with both faculty/staff/mentor and peers in The Data Mine. It was noticed that the students reported a higher level of interaction with peers than with a faculty/staff/mentor in the initial survey; this result is consistent with the conduced by Weidman and Stein (2003). One of the core reasons for a higher level of interaction with peers could be that students lived and learned with their peers in The Data Mine, and when this survey study was conducted, living in Hillenbrand Hall was a requirement for all The Data Mine students (Hurtado et al., 2020). Therefore, it can be assumed that students found sufficient opportunities to connect and interact with their peers (Hurtado et al., 2020; Jaiswal, Lyon, Magana, et al., 2021; Schneiderman, 2020). It is also noteworthy as students progressed through the semester, there was an increase in the socialization with the faculty/staff/mentor at The Data Mine (Hoffman et al., 2002). This reveals that students got opportunities to interact formally and informally with faculty/staff/mentors at The Data Mine (Weidman & Stein, 2003).

Similar results were obtained regarding the perceptions of sense of belonging. It was observed that students demonstrated an increase in the perception of their sense of belonging as

they progressed in The Data Mine. This reveals that students felt accepted, valued, and included in The Data Mine. Sense of belonging is very closely connected to the interaction of students with their peers and faculty/staff/mentors in an institutional setting (Goodenew, 1993). Studies (e.g., Osterman, 2010; Schneiderman, 2020) have revealed that if the students have a good level of interaction with their peers and faculty, they demonstrate a high sense of belonging. Sense of belonging is also closely associated with identity formation, as students can't form identity if they do not experience a sense of belonging in the institution (Hazari et al., 2020). From the results, we can interpret that The Data Mine provided various opportunities to students to connect and interact with faculty/staff/mentors and peers.

Moreover, it was also observed that there was a significant increase in academic and intellectual development perceptions. An increase in academic and intellectual development happen when students are provided an opportunity to socialize with peers or faculty or engaged in the academic and non-academic events (Kuh, 2001; Pascarella & Terenzini, 1980). The perceptions of academic and intellectual development focus on how students from diverse backgrounds, experiences, and characteristics integrate themselves into the academic and social values of the institution (Pascarella & Terenzini, 1980). Studies by (e.g., Pascarella & Terenzini, 1980; Tight, 2020) have demonstrated that student retention is closely associated with academic and intellectual development. The students who demonstrate a higher level of academic and intellectual development are more likely to engage with the institution and less likely to drop out (Tight, 2020). Studies (O'Keeffe, 2013; Tinto, 2007; J. C. Weidman & Stein, 2003) have demonstrated that student retention is closely associated with socialization opportunities and sense of belonging; therefore, a significant increase in the academic and intellectual development of the students reveals that students from diverse backgrounds felt accepted and integrated within The Data Mine as they found multiple opportunities to socialize and develop a sense of belonging.

The result of the quantitative study is consistent with the results of our qualitative study. Students agreed that they got more opportunities to socialize as they were progressing through The Data Mine. The qualitative study helped us identify the exact opportunities that led to increased socialization, sense of belonging, or academic/intellectual development by interacting with people of diverse backgrounds. Two themes were identified for socialization and a sense of belonging: 1) focused on the interaction with peers in The Data Mine, and 2) focused on the factors leading to sense of belonging within The Data Mine. Regarding the theme for interaction with peers, students

mentioned various opportunities to interact and collaborate as learning community members. Students found their collaboration with peers very valuable as that allowed them to work in their respective groups and learn from one another. Working as a group leads to collaboration and interaction among the group members (Hammar Chiriac, 2014). Studies by Jaiswal, Thomas, Karabiyik, et al. (2021), Jaiswal, Karabiyik, Thomas, et al. (2021), and Karabiyik et al. (2020) have identified various benefits for students working in groups. Some of the benefits are that students learn to set smart goals, assign and monitor roles, develop interpersonal relationships, and learn skills such as communication, critical thinking, and teamwork. Working in a group fosters creativity and learning from one another. Communication, collaboration, creativity, and critical thinking are four essential 21st-century skills students developed through the group work (Canez, 2018). The Data Mine provided several opportunities for students to collaborate and engage in learning, such as living in Hillenbrand Hall and working on corporate projects. Students have shared their experiences working with students of diverse backgrounds in interdisciplinary teams and learning new approaches to solve problems. Working with people from diverse backgrounds improves team performance (Bowers et al., 2000) and promotes creativity. The study by Stahl & Maznevski (2021) revealed that working in a diverse team helped the team members develop various ideas and discover new problem-solving styles, which led to creativity and critical thinking, resulting in high team performance. Having culturally diverse colleagues in a team allows the members to acknowledge and respect the cultural differences among the team members and create a fusion team (Crotty & Brett, 2012). Crotty and Brett further argued that a fusion team could be perceived as a fusion cuisine that lets team members from diverse backgrounds share their unique perspectives and creatively solve problems.

Similarly, working with students from different disciplines also fosters creativity and engagement that improves academic and intellectual development (Pascarella & Terenzini, 1980). Students got to meet and work with other students from different majors. Students mentioned that working with a student from a different major is always very useful as they bring a different perspective to the problems. Wellmon et al. (2009) conducted a study where he engaged students from four disciplines: education, clinical psychology, physical therapy, and social work, in an interdisciplinary learning experience. The study results revealed that students could understand the importance of team members from other disciplines. They were able to understand the overlap of disciplinary knowledge while working on the same task. Also, students acknowledged the benefits

of interdisciplinary collaboration. Therefore, we can determine from our analysis that the corporate partner projects were the perfect opportunity for students to work in interdisciplinary teams at The Data Mine.

Lastly, since The Data Mine is a living-learning community, it allowed the students to collaborate, socialize, interact and learn while residing in Hillenbrand Hall. Students mentioned positive experiences living in Hillenbrand Hall, such as meeting people of similar interest, making new friends, participating in study groups, socializing with staff, faculty, and peers (Hurtado et al., 2020). Studies have revealed that living and learning experiences (Hurtado et al., 2020; Jaiswal, Karabiyik, Thomas, et al., 2021) positively affect students and help them socialize, develop strong bonds with faculty and peers, experience a sense of belonging, and improve retention. Adams (2014) conducted a study and compared the benefits that students experience in a Living Learning Community (LLC) with non-LLC students. The result of our study is similar to the results obtained by Adams (2014). In the study, Adams (2014) demonstrated that LLC students experienced the transition to their college life in a more accessible way, as they found a friendly and helpful peer group. LLC students also formed close bonds with faculty as they interacted and socialized with the faculty members beyond classroom activities. LLC students also felt that they have more accessibility to academic resources and get more academic support, because peers of similar interest and faculty surround them to guide them. Lastly, LLC students also felt privileged as they participated in many special events that helped them develop their skills, and they were first to be made aware of such events. At the same time, the non-LLC counterparts did not experience any such benefit. Therefore, living in Hillenbrand Hall also helped students experience similar benefits and socialization opportunities.

The second theme focused on the factors that helped students experience a sense of belonging in The Data Mine. The analysis of the student interviews revealed that they found The Data Mine environment very inclusive and judgment-free. Students found themselves comfortable in expressing their thoughts and feelings. They also found the staff and faculty empathetic and supportive (Hausmann et al., 2009; Hoffman et al., 2002). They felt a connection with staff, faculty, and peers at The Data Mine and experienced a sense of community. Sense of belonging has always been considered an essential factor in promoting student retention and identity formation (Hausmann et al., 2009). A sense of belonging is only experienced when students feel connected with faculty, environment, and peers in the institution (Weidman & Stein, 2003).

Socialization with members of the learning community and participation in the activities help the students feel connected with the context and members (Weidman, 1989; Weidman et al., 2014). Making the students feel included allowed them to express themselves freely. This cultivated a sense of belonging and made them feel like integral members of the community (Morrow & Ackermann, 2012).

The communities of practice theory and principles apply to learning communities (Magana et al., 2021). They bring people of similar interests together and allow them to interact and collaborate by providing a judgment-free environment (Jaiswal, Lyon, Perera, et al., 2021). Communities of practice foster a sense of belonging (Lave, 2004; Lave & Wenger, 1991). They allow the members to socialize regularly, engage in activities they are interested in, and become experienced community members (Lave & Wenger, 1991). A study conducted by Magana et al. (2021) also demonstrated that the sense of belonging of the students increased as they were spending more time with the members of the learning community. This finding aligns well with the results of our study, as all the students interviewed have spent at least more than a year in The Data Mine. The learning communities are considered high-impact initiatives that foster a sense of belonging (Cole et al., 2017). Our study's sense of belonging result aligns with a study conducted by Hoffman et al. (2002). In the study by Hoffman et al. (2002), they revealed that students found their peers to be very supportive and approachable as they shared the same classes, and it was easy to approach them for any notes or seek their inputs on any assignment questions.

The study by Hoffman et al. (2002) further argued that students found the environment very inclusive and judgment-free. Since they all knew each other, students felt free to express their opinions and ask questions. In the Data Mine, we found a similar situation described in Hoffman et al. (2002); students felt connected with their peers. They all participated in seminar courses and attended outside events. Most of them were part of the corporate partner team and lived in Hillenbrand Hall. The regular interaction with the peers made the students feel comfortable, less hesitant in expressing their opinions, and able to develop close bonds.

The study by Hoffman et al. (2002) also discussed the role of faculty in helping students develop a sense of belonging in a learning community. The themes for students' interaction with faculty revealed that students found the faculty compassionate, friendly, and approachable. Students also felt that the faculty knew them personally and cared for them. The students felt connected and free to reach out to them in personal or professional guidance. Similar results were

observed in the case of The Data Mine: students described faculty were approachable, friendly, empathetic. Students also felt connected and comfortable sharing their concerns or seeking opinions. It is also important to note that students felt comfortable not just with the faculty but also with the staff at The Data Mine. Staff at The Data Mine equally cared for students, ensuring that students felt comfortable, and provided them various opportunities to grow and progress. Based on the themes for socialization and a sense of belonging, we can conclude that The Data Mine created an environment where every learning community member felt comfortable, connected, and fostered learning through interaction and collaboration with peers, faculty, and staff in The Data Mine. Based on the themes for socialization and sense of belonging, we can infer that The Data Mine created an environment that made students feel comfortable, connected, and learning occurred through interaction and collaboration with peers, faculty, and staff.

9.4   Perception of Identity formation

This section focuses on the students' identity formation in The Data Mine. For the quantitative survey study, the instrument by Godwin (2016) was adapted in the context of data science. A further qualitative study was conducted to understand the process of identity development in The Data Mine. The survey study results demonstrated an increase in perception of data science identity as students progressed in The Data Mine. The data science identity survey comprised three elements: recognition, interest, and performance/competence. Recognition refers to how an individual sees oneself and how others perceive the individual (Godwin, 2016; Godwin et al., 2016). The results of our study demonstrated that there was a significant increase in the recognition aspect of identity. Godwin (2016) has considered recognition central to identity formation. She further argues that how others recognize an individual is crucial, allowing the individual to develop their perception of themselves.

Therefore, a significant increase in recognition was demonstrated by students. That means students have started identifying themselves as data science person or data scientist or both. Students felt that their peers and faculty at The Data Mine recognized their data science skills. The results also reveal that there was also an increase in the other two sub-constructs of identity: interest and performance/competence. Even though the increase in interest and performance/competence was not significant, we can interpret (from the increase) that students were interested in learning data science concepts from the beginning and were developing data-driven competencies at The

Data Mine. Godwin (2016) associated the increase in interest with the student motivation and performance/competence with students' self-belief. This suggests that if there is an increase in the interest in data science, it also reveals that students feel motivated while working on a data science problem. An increase in performance/competence reveals that students feel competent in applying their data science skills.

Also, students believed that they are genuinely associated with data science. Since the increase in performance and competence is closely associated with self-belief, the survey instrument contained a few questions related to data science self-belief. The data science self-belief focuses significantly on how students perceive themselves regarding data science knowledge and skills (Scott & Ghinea, 2014). The self-belief was measured based on the following three factors: data science self-concept, data science fixed mindset, and data science anxiety. The self-concept refers to one's perception of their behavior, abilities, and distinctive characteristics (Bailey, 2003). Therefore, the data science self-concept refers to students' perception of their behavior, abilities, and expertise in data science. The survey results demonstrated a significant increase in the data science self-concept, meaning that students demonstrated positive behavior, high abilities, and expertise in solving or working on data science problems. Moreover, the students' perceptions regarding the data science fixed mindset and data science anxiety decreased as they progressed through the semester. This suggests that students developed confidence while applying their data science skills and knowledge (Buckley et al., 2019)

The quantitative study results confirmed that students were in the process of identity formation. But it was unclear about the factors that were contributing to the identity formation. The qualitative study explored how the interaction and participation in The Data Mine helped students develop identities. The result of the study revealed that all nineteen of the interviewed students demonstrated the development of a data science identity either in the form of a data science person, data scientist, or both. But the interesting point is that students serving as the TAs for corporate partner projects demonstrated a dual identity. The corporate partner TAs demonstrated a data science identity and an emerging leader identity. Therefore the results of perception of identity formation have two broad themes, one for data science identity and the second one for emerging leader identity, followed by sub-themes for each broad theme.

The data science identity theme revealed that most students, at the start, had limited to no knowledge of data science disciplines, and few demonstrated moderate to high knowledge of data

science. After joining The Data Mine and participating in various events, and collaborating with peers on problem-solving, it helped all nineteen students develop competencies and grow in data science skills and knowledge. Since students were developing confidence and self-efficacy in approaching the data science problems, they planned to pursue a data-oriented career in the future; it is essential to note that identities are dynamic (Godwin & Potvin, 2017; Kim & Sinatra, 2018; Serpe & Stryker, 2011); they keep changing as the individual progresses. Individuals can have multiple identities as they keep acquiring new skills and knowledge. Identity answers the questions: Who am I? How do others see me? How did I become what I am today? How do I see myself in the future? (Avraamidou, 2019, 2020). Responding to these questions is not easy, as multiple factors play an essential role in shaping the perceptions of identity for an individual. Prior studies (Carlone & Johnson, 2007; Godwin, 2016; Hazari et al., 2010) have focused on the following three attributes: recognition, interest and performance, and competence for creating STEM or disciplinary identity. But the results of our study have highlighted that it is not sufficient to define STEM or disciplinary identity based on recognition, interest and performance, and competence (Kim & Sinatra, 2018). Now, if we take a deeper dive into the identity formation of the twelve corporate partners TAs, we find that they did develop data science skills but also identified themselves as leaders, as they were constantly engaged in managing students and student teams. The emerging leader identity resulted from the following developmental factors: faculty/staff recognized the leadership skills; peers/team members acknowledged the leadership skills of TAs; and student TAs found their role meaningful as they contributed to the growth and success of The Data Mine.

It was evident from the student interviews that interaction with peers, faculty, and staff at The Data Mine was an important factor that helped students to learn and grow. Also the results of the quantitative study conducted in phase 1 demonstrated an increase in the perceptions of identity formation, along with an increase in the perceptions of socialization and sense of belonging. Also, a study by Hazari et al. (2020) has revealed that sense of belonging is crucial for identity formation, and sense of belonging can only be experienced if students socialize (Weidman et al., 2014b). Therefore the qualitative study helped us understand the influence of socialization and sense of belonging on identity formation. Therefore, I would like to discuss the qualitative study results from the interactionist lens (Kim & Sinatra, 2018). The interactionist lens has been used in sociology to study the identity formation (Serpe & Stryker, 2011). This lens helps to account for

the various social factors such as the influence of environment/context, peers, engagement in activities, etc., on the identity development of an individual (Serpe & Stryker, 2011). Kim and Sinatra (2018) have argued that we have approached a level where we need to interpret the STEM identity formation process differently. Just perceiving the STEM identity formation from gender or self-perception is not enough. It is essential to account for the social interaction, sense of belonging, and other conducive and challenging factors that play an essential role in shaping the identity.

Since The Data Mine is a learning community, and interaction plays an important role, therefore, our results reveal that influence of social factors such as interaction with peers, faculty, or staff (Weidman et al., 2014b), living in Hillenbrand Hall (Hurtado et al., 2020), working in groups on the group and corporate projects, attending seminar classes, managing teams (Ellingsen et al., 2021; Jaijairam, 2016), etc., had a significant impact on the identity formation of students in The Data Mine. Avraamidou (2020) also viewed identity formation from the lens of communities of practice. The study by Avraamidou (2020) defined identity as a process of becoming. This means that individuals learn from socialization, participate in communities of interest, and gradually, the identity changes as the individual develops new skills and engages in learning (Lave & Wenger, 1991). We see the same results in the case of The Data Mine: students were in the process of developing data science identity but corporate partner TAs were also developing other identities, such as identities as leaders. Students were constantly negotiating to answer who they are and what they want to be in the future (Avraamidou, 2020; Holland et al., 2001). All the students intended to go into a data-oriented field but not necessarily become data scientists. Still, they wanted to apply the skills they developed at The Data Mine to real-world projects. Therefore we can say that identity is a plural construct as it can integrate multiple characteristics that an individual develops through various social interactions (Avraamidou, 2019, 2020; Serpe & Stryker, 2011). Considering the opportunities provided to students in The Data Mine, all students were in the process of developing a data science identity, and corporate partner TAs at the same time were also forming a leader identity. Prior studies by Carroll & Levy (2010) and Priest (2012) have revealed that the perception of leadership is socially constructed. The social environment and experiences play an important role in developing the leader identity. The leader identity can be described as a relational identity (Priest, 2012; Serpe & Stryker, 2011). Student TAs described themselves as leaders because of the interpersonal relationship they developed with their peers,

team members, and mentors (Sluss & Ashforth, 2007). The corporate partner TAs played an essential role in managing the teams, acting as a liaison between students and corporate partners, and interacting with faculty and staff at The Data Mine to keep corporate projects on track. Komives et al. (2005) and Priest (2012) have argued that student leader plays a crucial role in an educational setting as they meaningfully engage with peers, mentors, faculty in the institution to address any complex challenges that student faces. This interaction with various community members allows the student to evolve as a leader. Student leaders—in our case, the corporate partner Tas—were gaining experiences as they progressed from their first year of being corporate TA to the second year. This shift in the position made them competent in handling the issues and moving from a novice to an expert leader (Lord & Hall, 2005).

Therefore, viewing the identity development process from the interactionist lens helped us understand the influence of social interaction, contexts, roles performed, and participation in events on the students' identity formation. The Data Mine engaged the students in multiple social interactions with faculty, staff, and peers as a learning community. It allowed students to grow, evolve as individuals, and continuously see themselves differently.

9.5 Growing into an experienced member of The Data Mine

As posed by the guiding RQ 2, the study's overarching goal was to understand how students grew into experienced members. To develop a better understanding of student experiences and their growth journey, students were grouped into five clusters based on the results of their qualitative coding. Also, narrative analysis was used to restory the student experiences, and therefore some representative students were identified from each cluster to narrate the experiences.

Students in all five clusters demonstrated growth, but every cluster had a different growth journey. Cluster 1 consisted of one student, Finley, an international student, who wanted to join the computer science learning community but was on the waiting list, so he joined The Data Mine. He described joining the Data Mine as a blessing in disguise. Also, he participated in The Data Mine remotely for a semester due to COVID-19. He got multiple opportunities to grow and learn. He also acknowledged that he had a moderate level of coding skills before joining The Data Mine. Still, he learned numerous new data science and statistics skills such as leadership, professional, and research skills through interaction with peers, faculty, and staff at The Data Mine. He also serves as a TA for The Data Mine seminar class.

Cluster 2 consisted of four students. All four students had coding knowledge before joining The Data Mine. Three out of the four, River, Reef, and Tommy, considered themselves moderately good with coding as they knew the logic behind coding. But one student, Arrow, demonstrated a high level of coding skills. All the students in this cluster found The Data Mine a wonderful experience as they could use their existing coding knowledge and learn new concepts. Students in this cohort also developed research skills, mentoring skills, group work skills, etc. Four out of two students were TAs, Arrow was the head TA for a seminar, and Tommy was a corporate partner TA. River was a part of research with the professor-led cohort, and therefore, he conducted research in particle physics. Since he was a chemical engineering graduate, he never imagined working in particle physics. Reef was a computer engineering major. He joined The Data Mine as he was fascinated by the corporate partner cohort; he worked with a reputed giant company and gained extensive data science and group work experience.

Clusters 3, 4, and 5 consisted of students who identified as novices concerning programming and data science skills. They had very limited to no programming knowledge before joining The Data Mine. Cluster 3 consisted of three students, Marion, Jackie, and Jamie. Marion described that The Data Mine helped him grow as a person and the experiences that he had were rare for any undergraduate students to have in their four-year college experience. Starting in Data Mine with no data science skills, today, Marion perceived himself as a front-end developer, scrum master, a leader, and a researcher. Jackie also mentioned that The Data Mine helped him develop data science skills, and he felt happy that he could use his skills to solve real-world problems. He gave an example of his internship experience, where he used his data science skills to develop a dashboard for a production system for a manufacturing plant. Jackie, too, is a corporate partner TA and manages an interdisciplinary team. Jamie also grew into a mature learner. She got various opportunities to participate in The Data Mine. Jamie mentioned that she did not just develop data science skills but also people management skills. She was the vice-president and president of the Data Mine Advisory Board (DMAB). Her responsibility was to conduct outside events by inviting speakers, managing the team, and organizing events. She also served as a liaison between the Data Mine students and faculty as a part of DMAB; she gave feedback to faculty and staff on various challenges students faced. Jamie was also a corporate partner TA and identified herself as a leader. All three students acknowledged that they developed networking skills and described the interaction with faculty, mentors, staff, and peers as an important reason for their growth.

Cluster 4 consisted of Orion, Quinn, Robbie, Morgan, Kendall, and Mickey. They all joined The Data Mine as novice learners and developed multiple data science, research, and teamwork skills through participation. All six students participated in the corporate partner projects. Four out of six students, Orion, Quinn, Robbie, and Morgan, were corporate partner TAs. Mickey managed the social media for the DMAB, and Kendall was a junior year student in The Data Mine. All the students found participation in seminar class and interaction with peers, faculty, and staff valuable. Orion, Kendall, and Quinn shared their specific experiences in participating in outside events. Kendall described how participating in the Outside Events helped her decide her career goal. Mickey was a very closed-off person before joining The Data Mine, but after being in The Data Mine, she developed confidence in communicating with others, collaborating with peers, and also developed presentation skills. Orion, Quinn, Robbie, and Morgan developed team management skills and served as scrum masters for their team. All six students developed data science, research, and mentoring skills by participating in various events. Collaborating with peers helped them learn group work, communication, and presentation skills.

Cluster 5 consisted of Harley, Frankie, Riley, Rio, and Skylar. This cluster consisted of all female students, out of which Harley, Riley, and Skylar were from non-STEM disciplines. All the students in this cluster were novice learners in computational and data science skills. They all joined The Data Mine to develop data science skills and were part of the corporate partner cohort as they found it a perfect opportunity to gain some real-world project experience. Students from non-STEM backgrounds initially felt intimidated, but the inclusive environment of The Data Mine and interaction with the faculty and staff made them feel comfortable. They found their peers very motivating and happy to help. All students participated in the seminars, outside events, and corporate partnerships to develop data science and research skills. In this cluster, Frankie, Riley, Rio, and Skylar were corporate partner TAs who manage their teams. Frankie and Rio were from a STEM background. Still, before joining The Data Mine, they always questioned the reason for including a mandatory statistics course in their undergraduate-level coursework. Being in The Data Mine, they got to apply the concepts of their statistics course to solve real problems, which changed Frankie and Rio's perception. Riley and Skylar never imagined them becoming TAs in The Data Mine, as they were from a non-STEM background. They felt motivated, and they believe that staff, faculty, TAs, mentors of The Data Mine want students to succeed in their life irrespective of their prior backgrounds. Harley, another non-STEM background student, wanted to go to graduate

school to study Information Systems. Being in The Data Mine helped her overcome her fear of mathematics and computing. Harley did excellent in her graduate school interview, and she was accepted into a reputed Public University in an Information Systems program. All students in this cluster believed that for any students interested in learning and developing data science skills, The Data Mine would help them develop the required skills and grow.

To discuss the growth journey of students, we need to consider the results of the three sub-questions related to benefits and challenges of participation, sense of belonging/socialization, and identity formation in The Data Mine. Based on the thematic analysis results, it was documented that students identified numerous benefits in terms of participation opportunities, very few challenges, a high level of socialization and interaction, and demonstrated the development of self-identity in terms of data science identity and relational identity in terms of leader identity. The overarching theme that was identified after analyzing all the five clusters was that all students grew into experienced learning community members. It is understood that the growth story of each participant is different as they came from different disciplinary backgrounds. Still, it is essential to note that they all were newcomers to The Data Mine and all students had the same goal to develop data science skills. Few students who had some prior coding knowledge somewhat knew about what they would gain after participating in The Data Mine; for example, students in clusters one and two had some prior programming knowledge. Therefore, they knew they would develop their data science skills and apply them to real-world problem-solving. But they did not know that they would develop a portfolio of skills, such as research, mentoring, teamwork, leadership, networking, professional skills, etc. Finley, during the interview, mentioned that when all students look back after spending a semester in The Data Mine, they find that they have grown so much as an individual.

However, if we look at clusters three, four, and five, all the students identified themselves as novices concerning programming and data science skills. They had very limited to no programming knowledge. Most of the students in those clusters did not know what they would learn in The Data Mine. But it is important to note that they too grew in many areas. Students with limited to no knowledge of data science became corporate partner TAs, managed teams, and worked on real-world projects. Cluster five had students from non-STEM backgrounds, and they mentioned that they got equal opportunities similar to STEM students to grow, learn, become TAs, and feel connected with the staff, faculty, and peers at The Data Mine. After spending a substantial

amount of time in The Data Mine, all the students in clusters three, four, and five agreed that they learned data science skills, developed leadership skills, teamwork skills, research skills, grew up in the social environment, and learned from one another. At the time of the interview, all the students mentioned that they wanted to pursue a data-oriented career, and they knew how to apply their data science knowledge to their area of interest or discipline.

Regarding the learning trajectory of students in The Data Mine, we find that most of the learning happened through engagement in activities and interaction, and collaboration with peers, mentors, staff, and faculty. These findings suggest that The Data Mine fosters a situated learning environment where learning occurs when students interact with one another and other learning community members. Interaction can occur through participation in various activities such as seminars, research, working on corporate projects, and collaborating with faculty, peers, and mentors. Interaction is key to The Data Mine. Therefore, it brings together interested students, faculty, corporate mentors, and significant others under the same roof and engages them in the data science practices (Wenger, 1998). Prior studies by (Jaiswal, Lyon, Magana, et al., 2021; Jaiswal, Lyon, Perera, et al., 2021) have viewed The Data Mine from the lens of communities of practice. The studies identified that newcomers entering The Data Mine had limited knowledge of data science, statistics, computing, research, or teamwork. In some ways, they were novices. Therefore, when these novice learners started their journey in the Data Mine, they were at the periphery of the community of practice. The Data Mine allowed them to interact with other members who acted like facilitators such as the peers, faculty, and mentors. These facilitators actively engage the newcomers in data science practices by allowing them to participate in seminars, outside events, research, corporate partnerships, etc., in a socially collaborative setting away from the traditional classroom environment (Gundlach & Ward, 2021).

Students also described instances of their interaction with other students in The Data Mine and their interaction with faculties, staff, TAs, corporate mentors, and graduate students. They explained how those interactions helped them grow. Below is a student quote describing some milestones of their growth journey in The Data Mine. Quinn joined The Data Mine in his first year. As a newcomer to The Data Mine, he met Ms. E, Dr. W, and Ms. M. Quinn got an opportunity from Ms. E to be a part of Corporate Partners. Quinn also got an actuarial science scholarship. He did internships in the summer. Quinn saw a natural linear progression in him and all the

opportunities he got at The Data Mine helped him become an experienced member of The Data Mine.

> Yes, certainly so my freshman year. It definitely felt like I was just kind of getting to know everybody. Like I met Ms. E and Dr. W and Ms. M when she came on board too and it was kind of just about getting acclimated and figuring out what kind of environment it was like. Then naturally as it went, then I got asked by Ms. E to be on the corporate partner's team in my freshman year. I got the actuarial science scholarship the end of my freshman year. Here, which I believe the people that I have met in the Data Mine helped me to get that. Uhm, and I also gradually have gotten up to lead a team after that. And then I eventually did my internship this past summer and Dr. W helped me with that. So, it's kind of like I just kind of progressively moved up. I joined the Data Mine kind of got acclimated, and then I slowly was joining projects gained new knowledge from them. Then getting like a real summer job internship, so it has felt like a very natural linear progression in the last two years.

The gradual growth of students from being a newcomer and transforming into experienced members of The Data Mine can be viewed from the lens of legitimate peripheral participation. The newcomers in The Data Mine were provided several opportunities to interact with faculty, mentors, staff, and TAs and participate in events that provided them opportunities to meet like-minded people (Lave & Wenger, 1991). Since the Data Mine allowed the members to engage in a collaborative environment that promotes socio-constructivist learning, the members were provided with ample opportunities to develop skills, create artifacts, experiences, stories, and contribute to the knowledge (Jaiswal, Lyon, Magana, et al., 2021; Jaiswal, Lyon, Perera, et al., 2021). This continuous participation and skill development also allowed the students to believe that they were developing a data science identity and emerging as a data science person, data scientist, or a combination of these. Also, after spending substantial time in The Data Mine, regularly interacting with the members, developing team management skills, and contributing to the learning community's growth, the student TAs developed leader identities. This movement of newcomers who developed into experienced members of the learning community helped the students to make progress on their learning trajectory (Lave & Wenger, 1991) and become leaders, mentors, guides, and role models for incoming newcomers to The Data Mine.

9.6 Implications

Implications for Learning Communities

The results of this study have specific implications for institutions and learning communities. I will describe how the professionals engaged in managing learning communities can benefit from this study in the subsequent paragraphs.

This study focused on the living-learning community and demonstrated its growth from its inception till the Fall of 2021. The study identified the various benefits and challenges students faced in The Data Mine. It was evident from the study that students got multiple opportunities to participate in various academic events, creating channels for constant communication (such as Piazza) that allowed them to interact with faculty and meet like-minded peers. It is essential for learning communities to allow students to interact beyond the classroom environment (Hoffman et al., 2002; Hurtado et al., 2020). Creating online channels for communication can help the students and faculty to interact and share information constantly (Meishar-Tal & Pieterse, 2019). This dissertation study also identified the importance of peer TAs and graduate students. The Data Mine students found it very comfortable talking to peer TAs and graduate students when they had any questions. The learning communities are primarily for undergraduate students (Adams, 2014; Flynn et al., 2016), but allowing graduate students to be part of the learning community will help the undergraduate students learn from the senior members. Having peer-TAs in the learning community is always beneficial as it allows the students to feel comfortable sharing their opinions or seeking advice from a peer (Priest & de Campos Paula, 2016). The Data Mine also has subject matter experts such as data scientists to support students and answer any technical queries. The learning communities focused on computation-related disciplines should also employ experts from faculty to address students' queries.

The Data Mine allowed the students to work on real-world corporate projects. Allowing the students to work on real-world corporate projects, The Data Mine helped them develop and apply their theoretical knowledge to solve real problems. The learning communities must engage students in project-based learning, allowing them to turn their disciplinary knowledge into a practical application (Ellingsen et al., 2021). Engaging in project-based learning will help students develop competence, resulting in higher self-efficacy and confidence (Cortázar et al., 2021; Fini et al., 2018; Jaiswal, Karabiyik, Thomas, et al., 2021). Students in The Data Mine also got opportunities to network and build connections that helped them get internships and jobs. The

learning community must continue to organize events or help students build connections that could help them get internships or jobs in the future (Bridgstock, 2017).

In summary, the role of learning communities or institutions should be to equip the newcomers or novice learners with the skills that make them ready to enter the industry. The curriculum should focus on 21st-century skills required by the industries. Providing networking opportunities and helping the prepare a job portfolio will help solve the issue of the employment gap.

Implications for Data Science Education

The study results also have implications for data science education. The study suggests that undergraduate students must be engaged in data science education from their first year onwards (National Academies of Sciences, Engineering, and Medicine, 2018) as the results of the study demonstrated that it takes time for students to develop and master data science skills. Also, the study demonstrated that the interdisciplinary nature of the data science allowed the students from various disciplines to be the part of The Data Mine, meaning the discipline of data science should not be restricted to computer science or data science majors; rather, students from different disciplines must be motivated to pursue data science education (Gundlach & Ward, 2021; Jaiswal, Lyon, Magana, et al., 2021). Another interesting finding was that involving students in corporate partner projects helped the students to apply their disciplinary knowledge to real world problem solving that improved their self-efficacy and made them feel more competent (National Academies of Sciences, Engineering, and Medicine, 2018). Therefore, the data science educators must find venues that allow the students to apply their disciplinary knowledge to real-world problem solving. The data science education must be imparted in an active project based learning environment that allows students to work alongside of their peers and learn from one another as it helps to master the complex concepts faster (Ellingsen et al., 2021). Lastly, the discipline of data science is multifaceted; therefore, just helping students to develop data science competencies is not enough, the data science educators must encourage students and provide them opportunities to develop research, communication, presentation and teamwork skills (Hardin, 2017; Koby & Orit, 2020).

Implications for Education Research

The result of the study has implications for education research as well. The results demonstrated that learning was implicit and it occurred when students, faculties, mentors, staff interacted with one another (Lave & Wenger, 1991; Priest, 2012). Therefore, it is important for the education researchers to identify the factors that can promote interaction and collaboration among the student and faculty, and design interventions to measure the impact of those factors on student learning. Another important finding of this study was that the identity formation is relational, which means that an individual can have multiple identities. Therefore, it is important for the education researchers to understand the impact of social factors on the identity formation process of STEM students (Kim & Sinatra, 2018).

9.7   Future Work

The study lays the groundwork for future research in the area of learning communities and also characterizes the growth of students and their experiences by participating in a learning community. In the future, the researchers should try to address the following questions:

- How do female students in The Data Mine perceive themselves in terms of self-efficacy and data science identity?
- How does a student's motivation influence the development of data science skills?
- What strategies do students employ to become self-regulated learners?
- How do students mitigate the conflicts while working on a corporate partner project in The Data Mine?
- How does The Data Mine support the underrepresented students to develop their identity?
- How do students benefit from working with culturally diverse students on corporate team projects?

9.8   Conclusion

Learning communities are considered one of the high-impact practices that help to improve student retention. Therefore, understanding the student experiences in the learning communities is crucial. There have been several quantitative studies conducted to understand the student learning

experiences, but there are minimal qualitative studies pertaining to understand the various psychosocial factors that influence student experiences in the learning community. Therefore, through this study, we explored the various participation and collaboration opportunities provided by The Data Mine. This living-learning community helped students socialize, develop a sense of belonging, develop identity, and finally grow into experienced members of the learning community.

The study's findings revealed that, after spending more than one year in The Data Mine, each student demonstrated growth in their data science skills, professional skills, communication skills, and developed identities. The study's findings emphasize that interaction and collaboration with various members in The Data Mine and participating in various events helped the students socialize and develop a sense of belonging and identities. The interaction helped the newcomers to move from the periphery to the center of the community. Another intriguing finding of this study is that students serving as corporate TAs demonstrated a leader identity and the data science identity. From this finding, we can interpret that leadership is socially constructed; therefore, allowing the students to participate in events, interact with peers and mentors, and give them a leadership position will help them develop a leadership identity. The study contributes to the body of literature in the area of learning communities, as the study identified how interaction and collaboration could help students to develop a sense of belonging, data science identity, leader identity and also at the same time help students learn job-oriented skills that make them equipped to meet the current workforce demand. The study also addresses the research needs to conduct studies to identify the psychosocial factors influencing student experiences in the learning community. This study also contributes to understanding identity formation from the interactionist lens. The study also contributes to the body of literature in data science education. The prior studies have revealed that the interdisciplinary nature of data science poses a challenge for educators to impart data science education. Therefore, higher education institutions must put an intentional structure such as learning communities to impart data science education. In the case of Purdue University, the learning community served as an effective mechanism to bring people with different backgrounds but with similar interests together and engage them in data science activities that helped develop data science and psychosocial skills. The study also revealed that engaging students in an active learning project-based environment helped them to master complex data science concepts efficiently and learn the interdisciplinary application of data science. Lastly, the

study also emphasizes that the higher education institution must foster a healthy social environment as that can help students to develop multiple positive identities.

Future work will focus on understanding how the skills developed at The Data Mine is helping students to advance in their career. Furthermore, future work will be conducted to understand how The Data Mine is helping underrepresented students grow into experienced members of the learning community.

APPENDIX A: DATA MINE DEMOGRPHICS (2018-2022)

Table 35: Gender Distribution in The Data Mine

	Female	**Male**	**Total**
Fall 2018	29	61	90
Spring 2019	17	53	70
Fall 2019	219	397	616
Spring 2020	103	240	343
Fall 2020	183	436	619
Spring 2021	138	330	468
Fall 2021	246	545	791
Spring 2022	205	497	702

Table 36: Ethnicity, Race and Nationality Distribution in The Data Mine

	2 or more races	**Asian**	**Black or African American**	**Hispanic/ Latino**	**International**	**Native Hawaiian or Other Pacific Islander**	**Unknown**	**White**	**Grand Total**
Fall 2018	8	26	1	3	22	0	0	30	90
Spring 2019	8	17	0	2	17	0	0	26	70
Fall 2019	26	153	18	26	98	1	10	284	616
Spring 2020	14	95	16	16	44	1	3	154	343
Fall 2020	15	195	16	26	101	0	18	248	619
Spring 2021	13	165	8	17	67	0	14	184	468
Fall 2021	22	245	8	27	238	1	10	240	791
Spring 2022	20	216	5	24	222	1	12	202	702

APPENDIX B: SURVEY QUESTIONS

Table 37: Survey Questions for three components of Identity: Recognition, Interest and Performance and Competence

Indicate whether you: (1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree and 5 – Strongly Agree)

Recognition	My parents see me as a data science person
Recognition	Instructors see me as a data science person
Recognition	My peers see me as a data science person
Recognition	I see myself as a data science person
Recognition	Data science is good field for a person like me
Interest	I am interested in learning more about data science
Interest	I enjoy learning data science concepts
Interest	I find fulfillment in applying data science concepts
Performance and Competence	I am confident that I can understand the data science concepts in class
Performance and Competence	I am confident that I can understand the data science concepts used in assignments
Performance and Competence	I can do well on data science assignments
Performance and Competence	I understand concepts I have studied in the data science course
Performance and Competence	Others ask me for help with data science concepts or assignments

Table 38: Survey Questions for Socialization

Please Respond: (Yes or No)

Socialization Staff/Faculty/Mentor	Is there any professor or staff member in the Data Mine or the Data Science Certificate with whom you: - Sometimes engage in social conversation
	Is there any professor or staff member in the Data Mine or the Data Science Certificate with whom you: - Often discuss topics in his/her field
	Is there any professor or staff member in the Data Mine or the Data Science Certificate with whom you: - Often discuss other topics of intellectual interest
	Is there any professor or staff member in the Data Mine or the Data Science Certificate with whom you: - Ever talk about personal matters
Socialization Peer	Is there any student in the Data Mine or the Data Science Certificate with whom you: - Sometimes engage in social conversation
	Is there any student in the Data Mine or the Data Science Certificate with whom you: - Often discuss topics in their field
	Is there any student in the Data Mine or the Data Science Certificate with whom you: - Often discuss other topics of intellectual interest
	Is there any student in the Data Mine or the Data Science Certificate with whom you: - Ever talk about personal matters

Table 39: Survey Questions for Sense of Belonging

Indicate whether you: (1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree and 5 – Strongly Agree)

Sense of Belonging	I feel a sense of belonging to my data science community (Data Mine Learning Community or the Data Science Certificate)
Sense of Belonging	I am a member of my data science community (Data Mine Learning Community or the Data Science Certificate)
Sense of Belonging	I see myself as part of my data science community (Data Mine Learning Community or the Data Science Certificate)

Table 40: Survey Questions for Perceptions of Self-Belief

Indicate whether you: (1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree and 5 – Strongly Agree)

Constructs	Questions
Data Science Self-Concept	I learn data science skills quickly
	I have always believed that I could be good at data science
	In my data science class, I can solve even the most challenging problems or assignments
Data Science Fixed Mindset	I have a fixed level of data science aptitude, and not much can be done to change it
	To be honest, I do not think I can really change my aptitude for acquiring data science skills
Data Science Anxiety	I often worry that it will be difficult for me to complete data science assignments
	I get nervous when trying to solve data science challenges
	I feel helpless when trying to solve data science challenges

Table 41: Survey Questions for Academic/Intellectual Development

Indicate whether you: (1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree and 5 – Strongly Agree)

Academic/Intellectual Development	I am satisfied with the extent of my intellectual development as student in the Data Mine Learning Community or the Data Science Certificate
	My academic experience in the Data Mine Learning Community or Data Science Certificate has had a positive influence on my intellectual growth and interest in ideas
	I am satisfied with my academic experience in the Data Mine Learning Community or Data Science Certificate

APPENDIX C: INTERVIEW PROTOCOL

Thank you for agreeing to participate in this study. Before I start the interview, let me first provide you with some background information about me and the study. My name is Aparajita Jaiswal, and I am a Ph.D. student at Purdue University. I am conducting a study to understand how undergraduate students develop data science skills, form identities, and socialize in The Data Mine. The study is intended to understand how you perceive yourself after completing one year at The Data Mine. I will ask a few questions related to your major, data science skills that you developed, and experiences with your peers, mentors, and faculty at The Data Mine. Please feel free to express your opinions by providing instances and examples. Information that you provide will remain confidential and all data will be deidentified to maintain your privacy. You will be compensated for your time with a $15 Amazon gift certificate for every 30 minutes of your participation. You can also ask me any questions during the interview.

Interview Questions

Introduction

1. What is your major?
2. What are some reasons why you selected that major?
3. Please describe what were your expectations prior to joining The Data Mine?

Possible follow-up questions

   o   What expectations of yours are fulfilled after joining The Data Mine?
   o   In what ways has the Data Mine helped you fulfil those expectations?

Learning through Participation

1. Can you tell me a little bit about how you, specifically, like to participate in The Data Mine?

Possible follow-up question

   o   What type of activities did you participate in?
   o   How your participation helped you achieve the expectations that you identified before joining Data Mine?

2. How has your participation in the activities at The Data Mine changed throughout the year?

Possible follow-up question

   o   [if the answer is positive] In what ways have you become more involved in the activities at The Data Mine? can you give me an example?
   o   [If the answer is negative] Can you describe why the change was limited?

3. How has your participation in activities at the Data Mine influenced the learning of your data science skills? Can you give me some examples of the skills you possessed before joining Data Mine?

Possible follow-up questions

   o   Describe the skills you developed as a result of participation in various activities at the Data Mine?

4. With whom did you tend to interact the most at The Data Mine?

Possible follow-up questions

   o   How has your engagement with other members (peers/seniors/faculty/mentors) at Data Mine impacted development of your data science skills? Please describe the skills that you developed as a result of engagement with other members.

5. Think of one of the most impactful mentors. Please describe what was impactful about that mentoring experience. Can you also provide some examples to explain your response?

6. How has working with senior-year students or graduate students within The Data Mine influenced development of your data science skills?

7. How has your interaction changed with faculty and TAs after completing one year at The Data Mine?

Identity

1. How do you describe yourself in the context of Data Science? Assuming you have spent a substantial amount of time at The Data Mine.

2. As you learn data science skills, do you think there's a line you have crossed (or that you might eventually cross) where you would feel comfortable calling yourself a data scientist or a "Data Science Person?" *["Data Science Person" is an individual who has good understanding of Data Science concepts and enjoys learning data science]*

Possible follow-up question
- o [if the response is positive] Do you think others [mentors, faculty, or peers] recognize you because of your data science skills?
- o [if the response is negative] why do you think so?

3. How and when did you know that you were becoming proficient in applying data science skills?

4. How competent do you feel in using data science skills that you developed at The Data Mine in your Data Science project/assignments? Please share an example.

Possible follow-up question
- o How do projects/assignments in The Data Mine help you develop data science skills?

5. Do you think that acquiring data science skills makes you feel more competent than your peers who do not possess the skills? Can you please elaborate?

Possible follow-up questions
  o If the answer is positive: What specific tools/techniques did you learn and how do you plan to use them in the future?
  o If the answer is negative: Please explain why the projects/assignments were not helpful.

6. Do you think that the knowledge of data science is important to you?

Possible follow-up questions
  o [If yes] Can you tell me about the first time you felt knowledge of data science was important to you?
  o [If No] Why do you think so? Can you elaborate on your response?

7. Can you tell me how you feel about the data science assignments and projects?

Possible follow-up questions
  o [If answer is positive] Please describe what you found interesting and why?
  o [If answer is negative] Please describe what aspects were not interesting and why?

8. What specific concepts in data science do you find interesting? Why do you find them interesting? Please provide some examples.

9. Do you enjoy learning and applying data science skills? If [yes/no] can you share an experience that made you feel so?
  • How important is data science to your major? Can you elaborate your response?
  • In your major, do you use data science skills? If so, can you provide some examples?
    o How competent do you feel in using data science skills that you developed at The Data Mine in your major? Please share an example of a project that you worked on that demonstrates your data science skills.

10. What career or profession do you plan to pursue in the future? Will that involve data science?

11. Which cohort do you belong to within The Data Mine (e.g., industry partnerships, general cohort, visualization cohort, etc.)? Why did you choose that cohort?

Socialization and Sense of Belonging

1. How do you feel as a member of The Data Mine?

Possible follow-up questions
- o [if positive: When did you start to feel that way?]
- o [if negative: Please explain why you don't feel connected to the community]

2. Who are some of the people that made you feel connected or disconnected with The Data Mine? Can you please provide some examples?
3. How do you feel while working on a group-based data science project?

Possible follow-up questions
- o If positive/negative: Please explain an instance when you felt that way.

4. How has working with peers within the Data Mine helped you learn data science?

Possible follow-up question
- o Please give some examples of skills you learnt.

5. How do you describe your social and academic interactions with peers and faculty at The Data Mine?

APPENDIX D. CONSENT FORM FOR SURVEY STUDY

RESEARCH PARTICIPANT CONSENT FORM

Key Information

Please take time to review this information carefully. This is a research study. Your participation in this study is voluntary, which means that you may choose not to participate at any time. You may ask questions to the researchers about the study whenever you would like. If you decide to take part in the study, you will be asked to sign this form, and you will be compensated for your time and effort. Make sure you understand what you will do and any possible risks or benefits. This study will investigate the ways in which students identify them as data scientist and they engage them with other students enrolled in the Data Mine Learning Community at Purdue. Participation in this study will require you to fill out a '20 minutes' survey. The questions on the survey will pertain to development of self-identity and sense of belongingness. You must be at least 18 years of age and currently enrolled in Data Mine learning Community at Purdue to participate in this study.

What is the purpose of the study?

The purpose of this study consists on identifying the educational benefits of participating in any of the data science initiatives at the undergraduate level at Purdue. We would like to enroll up to 200 students per semester for the period of three years. As member or participant of one of the data science initiatives, we would like to hear about your experiences in the Data Mine Learning Community at Purdue and your development as a Data Scientist.

What will I do if I choose to be in the study?

If you choose to participate in the study, you will be required to complete 20 minutes survey. The questions on the survey will pertain to development of self-identity and sense of belongingness. You will be required to include your full name so we can track your progress throughout the duration of your participation in your corresponding data science program. Although you will be required to mention your full name during the survey, after the survey is submitted your name will be de-identified to maintain your privacy throughout the entire analysis.

At this point, all your responses will remain anonymous and no names will be mentioned in any report. Your participation in this survey is anonymous to your instructor and strictly voluntary. Completing this survey has no relation to your grade and you may withdraw at any time. The instructor will not be notified about who is participating in this study. At the end of the semester, when all the grades have been calculated and officially posted, the instructor will be made aware of a summary of the results that contain no identification of the participants.

How long will I be in the study?

You will participate in a 20-minute survey assessment during the semester. The survey will be conducted at the beginning and end of each semester.

What are the possible risks or discomforts?

Your participation in this study will not involve any potential risks to you over and above those risks that would be encountered in everyday life. The research team will make all possible effort to maintain your confidentiality and although breach to confidentiality is a possible risk related to this study, safeguards are in place as listed in the confidentiality section

Are there any potential benefits?

There are no direct benefits to you for participating in this study. However, this study may allow you to offer your perspective in a confidential way, on the educational aspects of data mine learning community. These perspectives are invaluable, since they point to existing strengths and weaknesses of data-mine activities, as well as your perceptions of both. Your insights also provide us with guidelines for how to best implement such learning resources in the future.

Will I receive payment or other incentive?

To compensate you for your time and effort participating in this study. If you are a first semester student in Data Mine Learning Community, you will be eligible to receive up to $45.00 in the form of an Amazon gift certificate as follows:

- You will have to respond to our initial survey at the beginning of the <u>first semester</u> you will receive $15.00 Amazon gift certificate.

- Then if you plan to complete the second survey study at the end of the first semester you will receive $15.00 Amazon gift certificate.
- Also, if you complete the third survey study at the end of your second semester you will receive $15.00 Amazon gift certificate.

If you are a <u>second semester</u> student in the Data Mine Learning Community, you will be eligible to receive up to $30.00 in the form of an Amazon gift certificate as follows:

- You will have to respond to our initial survey at the beginning of the semester you will receive $15.00 Amazon gift certificate.
- Then if you plan to complete the second survey study at the end of the semester you will receive $15.00 Amazon gift certificate.

Amazon gift certificates will be provided a few days after participants complete each survey. Student will receive an email containing the Amazon gift certificates. Please note that students who complete the survey will only receive the Amazon gift certificates. Incomplete survey responses will be excluded from the study.

Are there costs to me for participation?

There are no costs for you for participating in this study other than the time you spend in completing the survey.

Will information about me and my participation be kept confidential?

Your confidentiality will be maintained, and your identity will not be disclosed. All steps will be taken to maintain your confidentiality by using a pseudonym to identify your data. The survey responses will be kept by Dr. Alejandra Magana, the lead researcher on this project, in a secure and locked location at Purdue University West Lafayette campus. The documents and electronic files will be kept indefinitely for future research purposes. Also, Federal regulations require that consent forms must be kept for a minimum of three years after closure of a study. The only persons who will have access to this data will be Dr. Alejandra Magana, and her PhD student Aparajita Jaiswal.

In addition, the project's research records may be reviewed by departments at Purdue University responsible for regulatory and research oversight.

What are my rights if I take part in the study?

Your participation in this study is voluntary. You may choose not to participate or, if you agree to participate, you can withdraw your participation at any time.

Who can I contact if I have questions about the study?

If you have questions, comments or concerns about this research project, you can talk to one of the researchers. Please contact Prof. Alejandra J. Magana (765) 494-

3994 admagana@purdue.edu.

If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu) or write to:

Human Research Protection Program - Purdue University

Ernest C. Young Hall, Room 1032

155 S. Grant St.,

West Lafayette, IN 47907-2114

Also, to report anonymously to Purdue's Hotline see www.purdue.edu/hotline.

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. I am prepared to participate in the research study described above.

I consent to participate in the study

I do not consent to participate in the study

APPENDIX E. CONSENT FORM FOR INTERVIEW STUDY

Key Information

Please take time to review this information carefully. This is a research study. The IRB Protocol number for the study is: IRB-2021-432, Interviews: Characterizing the Learning, Sociology and Identity Effects of the Data Mine Learning Community. The Principal Investigator (PI) for the study is Prof. Alejandra J. Magana. Your participation in this study is voluntary, which means that you may choose not to participate at any time. You may ask questions to the researchers about the study whenever you would like. If you decide to take part in the study, you will be asked to sign this form, and you will be compensated for your time and effort. Make sure you understand what you will do and any possible risks or benefits. This study will investigate the ways in which students identify them as data scientist or data science person and they engage them with other students enrolled in the Data Mine Learning Community at Purdue. Participation in this study will require you to participate in 60 to 90 minutes long interview. The questions in the interview will pertain to development of identity, socialization and sense of belongingness within the Data Mine Learning Community. You must be at least 18 years of age and currently enrolled in Data Mine learning Community at Purdue to participate in this study.

What is the purpose of this study?

The purpose of this study consists of identifying the educational benefits of participating in any of the data science initiatives at the undergraduate level at Purdue. As member or participant of one of the data science initiatives, we would like to hear about your experiences in the Data Mine Learning Community at Purdue and your development as a Data Scientist or Data Science Person.

What will I do if I choose to be in this study?

If you choose to participate in the study, you will be requested to participate in a 60 to 90 minutes long interview. You can exit the interview at any point of time. The questions of the interview will pertain to development of disciplinary identity, socialization and sense of belongingness. You will be required to fill out a short survey and the consent form prior to the interview. Your participation in this interview is anonymous to your instructor and strictly voluntary. Participating in the interview has no relation to your grade and you may withdraw at any time. The instructor will not

be notified about who is participating in this study. At the end of the semester, when all the grades have been calculated and officially posted, the instructor will be made aware of a summary of the results that contain no identification of the participants.

How long will I be in the study?

You will participate in a 60 to 90-minutes long interview during the semester. However, you can choose to exit the interview session at any point of time.

What are the possible risks or discomforts?

Your participation in this study will not involve any potential risks to you over and above those risks that would be encountered in everyday life. The research team will make all possible effort to maintain your confidentiality and although breach to confidentiality is a possible risk related to this study, safeguards are in place as listed in the confidentiality section.

Are there any potential benefits?

There are no direct benefits to you for participating in this study. However, this study may allow you to offer your perspective in a confidential way, on the educational aspects of Data Mine Learning Community. These perspectives are invaluable, since they point to existing strengths and weaknesses of data-mine activities, as well as your perceptions of both. Your insights also provide us with guidelines for how to best implement such learning resources in the future.

Will I receive payment or other incentive?

To compensate you for your time and effort participating in this study. You will be compensated with a $15 amazon gift card for every 30 minutes of your participation. For example, if your participation lasts 90 minutes, you will be compensated with $45 worth of amazon gift certificate. Amazon gift certificates will be provided a few days after participants complete each interview. Student will receive an email containing the Amazon gift certificates. Please note to receive the Amazon gift certificates you need to participate for at least 30 minutes. Any participation less than 30 minutes will not be compensated, and the interview responses will be discarded from the study.

Are there costs to me for participation?

There are no costs for you for participating in this study other than the time you spend during the interview session.

Will information about me and my participation be kept confidential?

Your confidentiality will be maintained, and your identity will not be disclosed. All steps will be taken to maintain your confidentiality by using a code to identify your data. You will be recognized by your assigned code. The interview data will be kept in a secure and locked location at Purdue University West Lafayette campus. The interview files will be kept indefinitely for future research purposes.

What are my rights if I take part in this study?

Your participation in this study is voluntary. You may choose not to participate or, if you agree to participate, you can withdraw your participation at any time.

Who can I contact if I have questions about the study?

If you have questions, comments or concerns about this research project, you can talk to one of the researchers. Please contact the PI Prof. Alejandra J. Magana (765) 494-3994 admagana@purdue.edu. If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu) or write to: Human Research Protection Program - Purdue University Ernest C. Young Hall, Room 1032 155 S. Grant St., West Lafayette, IN 47907-2114 Also, to report anonymously to Purdue's Hotline see www.purdue.edu/hotline.

 Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. Please choose one of the following options:

I Agree to participate in the research study mentioned above

I Disagree to participate in the research study mentioned above

VITA

Aparajita Jaiswal

Address: 110 Heavilon Hall, Purdue University, West Lafayette, Indiana - 47907
Phone: +1 (765) 476-6761
Email: jaiswal2@purdue.edu
Google Scholar: Aparajita Jaiswal

EDUCATION

August 2018 -May 2022
Doctor of Philosophy (Ph.D.), Technology
Department of Computer & Information Technology,
Purdue Polytechnic, West Lafayette, IN, USA.
Dissertation: "Characterizing the learning, sociology, and identity
effects of participating in the Data Mine."

Committee: Dr. Alejandra J. Magana (Chair), Dr. Mark D. Ward,
Dr. Austin Toombs, Dr. Ida Ngambeki.

2011-2013
Master of Business Administration
Motilal Nehru Institute of Research & Business Administration
University of Allahabad, Uttar Pradesh, India

2008-2011
Bachelor of Business Administration
Birla Institute of Technology, Mesra, Ranchi, India

PUBLICATIONS

Published & In Press Journal Articles:

1. Magana, A.J., Karabiyik, T., Thomas, P.J., **Jaiswal A.,** Perera, V., & Dworkin, J. (2022). Teamwork Facilitation and Conflict Resolution Training in Online and Hybrid Courses During the COVID-19 Pandemic. *Journal of Engineering Education*.

2. Magana, A.J., **Jaiswal, A.,** Madamanchi, A., Parker, L.C., Gundlach, E., & Ward, M.D. (2021) Characterizing the psychosocial effects of participating in a year-long residential research-oriented learning community. *Current Psychology.*

https://doi.org/10.1007/s12144-021-01612-y

3. **Jaiswal A.,** Lyon, J. A., Zhang, Y., & Magana, A.J. (2021) Supporting student reflective practices through modelling-based learning assignments. *European Journal of Engineering Education.* https://doi.org/10.1080/03043797.2021.1952164

4. **Jaiswal A**., Karabiyik, T., Thomas, P.J., & Magana, A.J. (2021). Characterizing Team Orientations and Academic Performance in Cooperative Project-Based Learning Environments. *Education Sciences*, 11(9),520, https://doi.org/10.3390/educsci11090520 **(Selected as one of the *Featured Papers* in the journal)**

5. Karabiyik,T., **Jaiswal, A.,** Thomas, P.J., & Magana, A.J. (2020). Understanding the Interactions between the Scrum Master and the Development Team: A Game-Theoretic Approach, *Mathematics*, 8(9), 1553, https://doi.org/10.3390/math8091553.

6. **Jaiswal, A**., Magana, A.J., Lyon, J.A., Gundlach, E, & Ward, M.D. (2021). Student Experiences within a Data Science Learning Community: A Communities of Practice Perspective. *Learning Community Research and Practice Journal,* 9(1), Article 2. https://washingtoncenter.evergreen.edu/lcrpjournal/vol9/iss1/2

Submitted Journal Articles:

7. Fennell, H. W., Lyon, J. A**., Jaiswal, A**., Li, T. & Magana, A.J. (under review). Investigating learning trajectories of engineering students through hybrid physics labs. *Journal of Science Education and Technology.*

8. Perera, V., **Jaiswal, A.,** Lyon, J. A., Fennell, H. W. & Magana, A.J. (under review). Using epistemic games to characterize students' thinking as they set up a computational physics problem. *Journal of Science Education and Technology.*

9. Lyon, J.A., **Jaiswal A.,** & Magana, A.J. (under review). Student experiences and performance within a designed modeling- based learning intervention. *Advances in Engineering Education.*

Peer-Reviewed Conference Proceedings

10. **Jaiswal, A.,** Patel, K., Patel, D., & Magana, A.J. (2022, *Abstract Accepted, Paper Submitted*). Perceived Scrum Values, Conflict Resolution Ability, and Cultural Self-Awareness among System Design and Analysis Students. *In Proceedings of the 129th ASEE Annual Conference and Exposition.*

11. **Jaiswal, A.,** Patel, D., Zhu, Y., Lee, J., & Magana, A.J. (2022, *Abstract Accepted, Paper Submitted*). A Reflection on Action Approach to Teamwork Facilitation. *In Proceedings of the 129th ASEE Annual Conference and Exposition.*

12. **Jaiswal, A.,** Lyon, A.J., Perera, V., Magana, A.J., Gundlach, E, & Ward, M.D. (2021). Work in Progress: Evaluating Student Experiences in a Residential Learning Community: A Situated Learning Perspective. *In Proceedings of the 128th ASEE Annual Conference and Exposition, Virtual Meeting. July 26-29, 2021.* https://peer.asee.org/38152

13. **Jaiswal, A.,** P. Thomas, T. Karabiyik., V. Perera, & Magana, A.J., (2021). Assessing the Impact of Transition from Face-to-Face to Online Instruction on Team Cooperation. *In Proceedings of the 129th ASEE Annual Conference and Exposition.* 10.18260/1-2—36718

14. Lyon, J.A., **Jaiswal, A.,** Magana, A.J., Gundlach, E, & Ward, M.D. (2021). Student challenges, strategies, and learning within a Data Mine Learning Community. *Frontiers in Education* (FIE).

15. Lyon, J. A., **Jaiswal, A.,** & Magana, A.J. (2020). The use of MATLAB Live as a technology-enabled learning environment for computational modeling activities within a capstone engineering course. *In Proceedings of the 127th ASEE Annual Conference and Exposition.* 10.18260/1-2—35380

16. Mohandas, L., Mentzer, N., **Jaiswal, A.,** & Farrington, S. (2020). Effectiveness of Undergraduate Teaching Assistants in a First-Year Design Course. *In Proceedings of the 129th ASEE Annual Conference and Exposition.* 10.18260/1-2--34503

CONFERENCE PRESENTATION

Jaiswal, A., Lyon, A.J., Perera, V., Magana, A.J., Gundlach, E, & Ward, M.D. (2021). Evaluating Student Experiences in a Residential Learning Community: A Situated Learning Perspective. *ASEE Virtual Annual Conference*.

Jaiswal, A., P. Thomas, T. Karabiyik., V. Perera, & Magana, A.J., (2021). Assessing the Impact of Transition from Face-to-Face to Online Instruction on Team Cooperation. ASEE Virtual Annual Conference.

Jaiswal, A., & Laux, C., (2019). Integration of Six Sigma and Time-Variant Kano Model in Higher-Education Institutions. Fifth International Conference on Lean Six Sigma in Higher Education. Edinburgh, Scotland.

MENTORING

*Mentored a total of **nine undergraduate students***

Fall 2020   Yiqun Zhang*, Computer and Information Technology

Fall 2020   Mahita Mudunuri, Computer and Information Technology

Fall 2021   Yi Zhu*, Computer and Information Technology

Fall 2021   Jin Su Lee*, Computer and Information Technology

Fall 2021   Yuzhe Zhou*, Computer and Information Technology

Fall 2021   Kang Kichang*, Computer and Information Technology

Fall 2021   Junhee Lee*, Computer and Information Technology

Fall 2021   Alex Choi*, Computer and Information Technology

Fall 2021   Jainam Piyush Patel*, Computer and Information Technology

Mentored Student Publication

1. **Jaiswal A**., Lyon, J. A., **Zhang, Y***., & Magana, A.J. (2021) Supporting student reflective practices through modelling-based learning assignments. *European Journal of Engineering Education.*

2. **Jaiswal, A.,** Patel, D., **Zhu, Y.*, Lee, J**.*, & Magana, A.J. (2022, *Accepted*). A Reflection on Action Approach to Teamwork Facilitation. *In Proceedings of the 129th ASEE Annual Conference and Exposition.*

Mentored Student Working Papers

1. **Lee, J.*, Kichang, K.*, Jaiswal, A.,** Patel, D & Magana, A.J. *Exploring cultural values in the context of teamwork*
2. **Patel, J.*, Jaiswal, A.,** Patel, D & Magana, A.J., *How do students' prior teamwork experiences influence their perception of teamwork on a semi-capstone project.*
3. **Zhou, Y.*, Choi, A.*, Jaiswal, A.,** Patel, D & Magana, A.J. *Assessing he role of cultural background on student team communication*

EXPERIENCE

2021   **Developed a graduate-level course with my advisor on Intercultural Competence and Mentoring**

The course that I developed on intercultural competence is based on the four principles of intercultural development:1) increasing personal and cultural self-awareness, 2) increasing awareness of others within their own cultural and personal context, 3) learning to manage emotions, 4) learning to bridge cultural gaps by shifting frames and adapting behaviors. We designed activities, projects and assignments focusing on these principles, and the intent of the class is to help students become reflective practitioners.

2021   **Co-taught an undergraduate research method course.**

The course's objective was to help students identify a research problem, train them on conducting literature reviews and research methods, and guide them on manuscript writing. The course aims to help students develop research skills and present their work at conferences.

2020 – Present   **Graduate Research Assistant, Rocketed Lab, Purdue University**

As a Graduate Research Assistant, my primary responsibility is to conduct research in STEM education, publish our findings in the reputed peer-reviewed journal, and present at conferences. My responsibility is also to develop a research plan for undergraduate students and mentor them. Additionally, our lab collaborates with faculty and students across many departments, and I have had opportunities to work with people from diverse backgrounds.

2019 – 2020   **Graduate Teaching Assistant of a senior-level course, Purdue Polytechnic.**

The course focused on teaching intercultural competence to senior year technology students As a Graduate Teaching Assistant for the senior level intercultural course was to design in-class activities, assignments for the students that could help them understand their own culture and appreciate other cultures. I also had the responsibility to debrief students on their Intercultural Development Inventory (IDI) results and conduct office hours.

2019 – 2020   **Graduate Assistant for Office of Globalization, Purdue Polytechnic.**

As a Graduate Assistant for the office of Globalization, my responsibilities were to help the Globalization Team in their study abroad initiatives. Conduct study abroad fairs and other social events. Conducted the Student Transitions and Registrations (STAR) Program for incoming Freshmen students. Participate in workshop sessions related to intercultural learning, such as Intercultural Development Inventory (IDI) and Beliefs, Events and Values Inventory (BEVI).

2018   **Instructor for a Freshmen level course, Purdue Polytechnic.**

Taught a course on supply chain systems to two sections, each comprising of 60 undergraduate students. The objective of the

course was to help students understand the basics of supply chain management. The course used a project-based cooperative learning approach to teach students the concepts of operations and logistics management.

2014-2017	**Senior Manager Operation Excellence at Uber9 Business Process Services Private Limited (vakilsearch.com)**

Managed a team of 45 people. Followed a lean methodology to get the work done smoothly. Worked with the technology team to develop new tech solutions. Help the team to trouble shoot any work execution issues.

2013-2014	**Worked as Manager Partnerships and Affiliations at Uber9 Business Process Services Private Limited (vakilsearch.com)**

Explored the opportunities of affiliations and partnerships for the company. Participated in various events to create brand awareness and improve brand value. Improved the sales revenue through partnerships and affiliations.

AWARDS AND HONORS

- Dean's Travel Grant by the Department of Computer and Information Technology, 2019.

- Travel Grant by Office of Globalization, 2019.

- Awarded Daniel and Martina Lewis scholarship, Purdue University, 2018.

CERTIFICATION

- Foundations of College Teaching, Purdue University   2021
- Black Belt Six Sigma, Purdue University   2020
- Intercultural Development Inventory (IDI) Administrator, Purdue University   2019
- Green Belt Six Sigma, KPMG, Chennai, India   2015

PROFESSIONAL DEVELOPMENT

Teaching

2021   Foundations of College Teaching
    Attended a month-long training program on teaching large lectures with the use of technologies and ways to promote diversity and equity in a classroom

Diversity

2020   Beliefs, Events and Values Inventory (BEVI) Qualifying Workshop
Attended a three-day workshop on BEVI

2019   Intercultural Development Inventory (IDI) Qualifying Seminar
Attended a three-day workshop on intercultural competence.

Statistics and Data Analysis

2019   Telling story with data
Attended a 2-hour training on analysis quantitative data using SPSS.

2019   Text Analysis Workshop
Attended a 2-hour training on how to analyze text data using packages in R.

2019   Visualizing Social Media network with Gephi
Attended a 2-hour training on data visualization and learned to create network diagrams using Gephi

ACADEMIC SERVICES

2021-2022   Reviewer for American Society of Engineering Education (ASEE) conference

2019-2020   Served as Intercultural Development Inventory (IDI) Administrator

2019-2020   Co-ordinated and conducted Student Transition, Advising, and Registration (STAR) program for incoming Freshmen students at Purdue Polytechnic Institute, Purdue University.

SKILLS

Research Skills: Narrative Inquiry, Thematic Analysis, Semi-Structured interviews, Focus groups Survey designing, Factor Analysis (exploratory & confirmatory), Multivariate Statistics, Hierarchical Clustering, Sentiment analysis,

Software: R, SPSS, Microsoft Office, Minitab, NVIVO, Atlas.ti

REFERENCES

Adams, S. (2014). *The perceived benefits of participation in living and learning communities: The student perspective.* [Masters Theses. 1273]. https://thekeep.eiu.edu/theses/1273

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). *Situated Learning and Education. Educational Researcher, 25*(4), 5–11. https://doi.org/10.3102/0013189X025004005

Arensdorf, J., & Naylor-Tincknell, J. (2016). *Beyond the Traditional Retention Data: A Qualitative Study of the Social Benefits of Living Learning Communities. Learning Communities: Research & Practice, 4*(1), 4.

Asonitou, S. (2013). Improving teamwork and communication skills through an action research project. *Proceedings of WSEAS Conference, Special Session: New Trends and Technologies for Marketing, Management, Economics and Information Processing II, 14–16.*

Avraamidou, L. (2019). Stories we live, identities we build: How are elementary teachers' science identities shaped by their lived experiences? *Cultural Studies of Science Education*, 14(1), 33–59. https://doi.org/10.1007/s11422-017-9855-8

Avraamidou, L. (2020). Science identity as a landscape of becoming: Rethinking recognition and emotions through an intersectionality lens. *Cultural Studies of Science Education, 15*(2), 323–345. https://doi.org/10.1007/s11422-019-09954-7

Bailey 2nd, J. A. (2003). Self-image, self-concept, and self-identity revisited. *Journal of the National Medical Association*, *95*(5), 383.

Balcan, M.-F., Liang, Y., & Gupta, P. (2014). Robust Hierarchical Clustering. *Journal of Machine Learning Research, 15*(118), 4011–4051.

Banks, C. M. (2018). *A Qualitative Study on Teachers' Perceptions of Co-Teaching in Inclusion Classrooms*. [PhD Thesis] Seton Hall University.

Baron, S. I., Brown, P., Cumming, T., & Mengeling, M. (2020). The Impact of Undergraduate Research and Student Characteristics on Student Success Metrics at an Urban, Minority Serving, Commuter, Public Institution. *Journal of the Scholarship of Teaching and Learning, 20*(1). Retrieved

from https://scholarworks.iu.edu/journals/index.php/josotl/article/view/25423

Barshay, J. (2017). Federal Data Show 3.9 Million Students Dropped Out of College With Debt in 2015 and 2016 | Data Mine | US News. https://www.usnews.com/news/data-mine/articles/2017-11-07/federal-data-show-39-million-students-dropped-out-of-college-with-debt-in-2015-and-2016

Bauer, R. N., & Kiger, S. (2017). Nursing living-learning communities and student retention: A qualitative study. *Nursing Education Perspectives, 38*(2), 75–79.

Baum, S., Ma, J., Earnings, K. P., & Notes, T. (2010). *Educational Pays 2010.* https://files.eric.ed.gov/fulltext/ED526357.pdf

Beard, F., & Morton, L. (1998). Effects of internship predictors on successful field experience. *Journalism & Mass Communication Educator, 53*(4), 42–53.

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*(3), 329–345.

Binder, J. F., Baguley, T., Crook, C., & Miller, F. (2015). The academic value of internships: Benefits across disciplines and student backgrounds. *Contemporary Educational Psychology, 41*, 73–82.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991).

    Motivating project-based learning: Sustaining the doing, supporting the learning.

    Educational Psychologist, 26(3–4), 369–398.

Bobilya, A. J., & Akey, L. D. (2002). An Evaluation of Adventure Education Components in a

    Residential Learning Community. *Journal of Experiential Education, 25*(2), 296–304.

    https://doi.org/10.1177/105382590202500208

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in

    work teams: A meta-analysis. *Small Group Research, 31*(3), 305–327.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in*

    Psychology, 3(2), 77–101.

Braun, V., & Clarke, V. (2012). Thematic analysis.   In H. Cooper, P. M. Camic, D. L. Long, A.

    T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in*

    psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and

    biological (pp. 57–71). American Psychological

    Association. https://doi.org/10.1037/13620-004

Braun, V., & Clarke, V. (2021). Conceptual and design thinking for thematic analysis.

Qualitative Psychology, 9(1), 3–26. https://doi.org/10.1037/qup0000196.

Bridgstock, R. (2017). *The university and the knowledge network: A new educational model for*

    twenty-first century learning and employability. In Graduate employability in context (pp.

    339–358). Springer.

Brower, A. M., & Inkelas, K. K. (2010). Living-learning programs: One high-impact educational

    practice we now know a lot about. *Liberal Education, 96*(2), 36–43.

Brubacher, J. S., & Rudy, W. (1997). *Higher education in transition: A history of American colleges and universities.* Transaction Publishers.

Buckley, J., Grajeda, S. B., Trauth, A., & Roberts, D. (2019). *A Framework for Quantifying Student Self-Confidence and Task Choice in Engineering Design-related Activities.* ASEE Annual Conference & Exposition 2019.

Callanan, G., & Benzing, C. (2004). Assessing the role of internships in the career-oriented employment of graduating college students. *Education & Training, 46*(2), 82–89. https://doi.org/10.1108/00400910410525261

Canez, D. (2018). *21st century learning practices: A case study of student response and motivation in the classroom*. [Masters Thesis]. http://hdl.handle.net/10211.3/201176

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. Journal of Research in Science Teaching: *The Official Journal of the National Association for Research in Science Teaching, 44*(8), 1187–1218.

Carlone, H. B., Johnson, A., & Scott, C. M. (2015). Agency amidst formidable structures: How girls perform gender in science class. J*ournal of Research in Science Teaching, 52*(4), 474–488.

Carrino, S. S., & Gerace, W. J. (2016). Why STEM Learning Communities Work: The Development of Psychosocial Learning Factors through Social Interaction. *Learning Communities: Research & Practice, 4*(1), 3.

Carroll, B., & Levy, L. (2010). Leadership development as identity construction. *Management Communication Quarterly, 24*(2), 211–231.

Cheng, W., & Shwe, T. (2019). Clustering Analysis of Student Learning Outcomes Based on Education Data. *2019 IEEE Frontiers in Education Conference (FIE), 1–7.*

Chin, W. W., Salisbury, W. D., Pearson, A. W., & Stollak, M. J. (1999). Perceived cohesion in small groups: Adapting and testing the perceived cohesion scale in a small-group setting. *Small Group Research, 30*(6), 751–766.

Choi, J.-I., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology Research and Development, 43*(2), 53–69. https://doi.org/10.1007/bf02300472

Ciechalski, J. C., Pinkney, J. W., & Weaver, F. S. (2002). A Method for Assessing Change in Attitude: The McNemar Test. *Poster presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA* (April 1-5, 2002).

Cobo, G., García-Solórzano, D., Morán, J. A., Santamaría, E., Monzo, C., & Melenchón, J. (2012). Using agglomerative hierarchical clustering to model learner participation profiles in online discussion forums. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 248–251.*

Cole, D. G., Newman, C. B., & Wheaton, M. M. (2017). Learning Communities and Students' Sense of Belonging to Their University: First Results from a Longitudinal Study. *AERA Online Paper Repository.*

Columbus, L. (2017). *IBM predicts demand for data scientists will soar 28% by 2020.* https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/?sh=1a56e1ee7e3b

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher, 19*(5), 2–14.

Connelly, L. M. (2016). Trustworthiness in qualitative research. *Medsurg Nursing: Official Journal of The Academy of Medical Surgical Nurses, 25*(6), 435.

Contreras Aguirre, H. C. (2019). *Exploring Latina STEM College Student Success at Southern Hispanic Serving Higher Education Institutions*. ProQuest LLC.

Cook, S. J., Parker, R. S., & Pettijohn, C. E. (2004). The perceptions of interns: A longitudinal case study. J*ournal of Education for Business, 79*(3), 179.

Cortázar, C., Nussbaum, M., Harcha, J., Alvares, D., López, F., Goñi, J., & Cabezas, V. (2021). Promoting critical thinking in an online, project-based course. *Computers in Human Behavior, 119*, 106705.

Cox, M. D., & Richlin, L. (Eds.). (2004). *Building faculty learning communities: new directions for teaching and learning*, 97 (Vol. 15). John Wiley & Sons

Creswell, J. W. (2007). *Five qualitative approaches to inquiry.* Qualitative Inquiry and Research Design: Choosing among Five Approaches, 2, 53–80.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Crisp, G., Baker, V. L., Griffin, K. A., Lunsford, L. G., & Pifer, M. J. (2017). *Mentoring Undergraduate Students: ASHE Higher Education Report,* Volume 43, Number 1. John Wiley & Sons.

Cross, K. P. (1998). Why Learning Communities? Why Now? *About Campus, 3*(3), 4–11. https://doi.org/10.1177/108648229800300303

Crotty, S. K., & Brett, J. M. (2012). Fusing creativity: Cultural metacognition and teamwork in multicultural teams. *Negotiation and Conflict Management Research, 5*(2), 210–234.

Dagley, M., Georgiopoulos, M., Reece, A., & Young, C. (2016). Increasing retention and

    graduation rates through a STEM learning community. *Journal of College Student*

    Retention: Research, Theory & Practice, 18(2), 167–182.

Dagley-Falls, M., Georgiopoulos, M., & Young, C. (2010). Influencing Sense Of Community In

    A Stem Living Learning Community: An NSF Step Funded Project. *ASEE Annual*

    Conference & Exposition, 15–734.

Davis, J. (2006). The importance of the community of practice in identity development. Internet

    Journal of Allied Health Sciences and Practice, 4(3), 5.

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant,

    L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A.,

    Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., Tiruviluamala, N., Uhlig,

    P.X., Washington,T. M., Wesley, C.L., White, D., & Ye, P. (2017). Curriculum

    Guidelines for Undergraduate Programs in Data Science. *Annual Review of Statistics and*

    Its Application, 4(1), 15–30. https://doi.org/10.1146/annurev-statistics-060116-053930.

Deckard, N. M. D. (2017). *Statistics Education for Undergraduate Sociology Majors: Survey*

    Findings across Institutions. Numeracy: Advancing Education in Quantitative Literacy,

    10(2).

Degn, L., Franssen, T., Sørensen, M. P., & de Rijcke, S. (2018). Research groups as communities

    of practice—A case study of four high-performing research groups. *Higher Education,*

    76(2), 231–246.

Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE Handbook of Qualitative Research.* Thousand

    Oaks, CA: Sage.

Dewey, J. (1986). Experience and education. *The Educational Forum, 50*(3), 241–252.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics, 26*(4), 745–766.

Doody, O., & Noonan, M. (2013). Preparing and conducting interviews to collect data. *Nurse Researcher, 20*(5).

Drury, M., Conboy, K., & Power, K. (2012). Obstacles to decision making in Agile software development teams. *Journal of Systems and Software, 85*(6), 1239–1254.

Dunn, M. S., & Dean, L. A. (2013). Together we can live and learn: Living-learning communities as integrated curricular experiences. SCHOLE: *A Journal of Leisure Studies and Recreation Education, 28*(1), 11–23.

Eby, L. T., & Allen, T. D. (2002). Further investigation of protégés' negative mentoring experiences: Patterns and outcomes. *Group & Organization Management, 27*(4), 456–479.

Ekornes, S. (2021). The impact of perceived psychosocial environment and academic emotions on higher education students' intentions to drop out. *Higher Education Research & Development,* 1–16.

Ellingsen, P., Tonholm, T., Johansen, F. R., & Andersson, G. (2021). Learning from Problem-Based Projects in Cross-Disciplinary Student Teams. *Education Sciences, 11*(6), 259.

Elliott, J. (2005). *Using narrative in social research: Qualitative and quantitative approaches*. Sage Publications,

Emden, C. (1998). Conducting a narrative analysis. *Collegian, 5*(3), 34–39. https://doi.org/10.1016/S1322-7696(08)60299-1

English, P., de Villiers Scheepers, M. J., Fleischman, D., Burgess, J., & Crimmins, G. (2021). Developing professional networks: The missing link to graduate employability. *Education + Training, Vol. 63* No. 4, pp. 647-661. https://doi.org/10.1108/ET-10-2020-0309

Eshchar-Netz, L., & Vedder-Weiss, D. (2021). Teacher Learning in Communities of Practice: The Affordances of Co-Planning for Novice and Veteran Teachers' Learning. *Journal of Research in Science Teaching, 58*(3), 366–391.

Amstutz, K et al. (2021). The Examples Book. https://the-examples-book.com/.

Fearon, C., McLaughlin, H., & Eng, T. Y. (2012). Using student group work in higher education to emulate professional communities of practice.*Education + Training*, Vol. 54 No. 2/3, pp. 114-125. https://doi.org/10.1108/00400911211210233

Figgou, L., & Pavlopoulos, V. (2015). Social psychology: Research methods. *International Encyclopedia of the Social & Behavioral Sciences, 22*, 544–552.

Fini, E. H., Awadallah, F., Parast, M. M., & Abu-Lebdeh, T. (2018). The impact of project-based learning on improving student learning outcomes of sustainability concepts in transportation engineering courses. *European Journal of Engineering Education, 43*(3), 473–488.

Flynn, M. A., Everett, J. W., & Whittinghill, D. (2016). The impact of a living learning community on first-year engineering students. *European Journal of Engineering Education, 41*(3), 331–341.

Gee, J. P. (2001). Education identity as an analytic lens for research. *Review of Research in Education, 25,* 99–125. https://doi.org/10.3102/0091732X025001099

Gergorić, T. (2020). Teachers' assessment of active learning in teaching Nature and Society. *Economic Research-Ekonomska Istraživanja, 33*(1), 1265–1279.

Giles Jr, D. E., & Eyler, J. (1994). The theoretical roots of service-learning in John Dewey: Toward a theory of service-learning. *Michigan Journal of Community Service Learning, 1(*1), 7.

Godwin, A. (2016). The development of a measure of engineering identity. *ASEE Annual Conference & Exposition.*

Godwin, A., & Potvin, G. (2017). Pushing and pulling Sara: A case study of the contrasting influences of high school and university experiences on engineering agency, identity, and participation. *Journal of Research in Science Teaching, 54*(4), 439–462.

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. J*ournal of Engineering Education, 105*(2), 312–340.

Goodenew, C. (1993). Classroom belonging among early adolescent students. *Journal of Early Adolescence, 13*, 21–43.

Grata, L. (2019). *Females in STEM: Self-Efficacy Within Women Who Pursue STEM Majors.* (Doctoral dissertation, Duquesne University). Retrieved from https://dsc.duq.edu/etd/1865

Green, W., Ruutz, A., Houghton, L., & Hibbins, R. (2017). *Enabling stories: Narrative, leadership, learning and identity in a faculty-based teaching community of practice.* In Implementing communities of practice in higher education (pp. 159–181). Springer.

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5–17.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries.

    Educational Communication and Technology, 29(2), 75–91.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Sage Publications.

Gundlach, E., & Ward, M. D. (2021). The data mine: Enabling data science across the

    curriculum. *Journal of Statistics and Data Science Education, 29*(1), 74–82.

    https://doi.org/10.1080/10691898.2020.1848484

Gutierrez, L. M. (2012). From the editor: Recognizing and valuing our roles as mentors. *Journal*

    of Social Work Education, 48(1), 1–4.

Hajdarevic, D. (2018). The Role of Product Owners: An Empirical Investigation of the Role of

    Product Owners and Challenges in Agile Scrum Projects.Uppsala Universitet.

Hammar Chiriac, E. (2014). Group work as an incentive for learning—Students' experiences of

    group work. *Frontiers in Psychology, 5*, 558. https://doi.org/10.3389/fpsyg.2014.00558

Handley, K., Sturdy, A., Fincham, R., & Clark, T. (2006). Within and beyond communities of

    practice: Making sense of learning through participation, identity and practice. *Journal of*

    Management Studies, 43(3), 641–653.

Hardin, J. (2017, September 1). *Expectations and Skills for Undergraduate Students Doing*

    Research in Statistics and Data Science | Amstat News.

    https://magazine.amstat.org/blog/2017/09/01/undergraduateexpectations/

Hausmann, L. R., Ye, F., Schofield, J. W., & Woods, R. L. (2009). Sense of belonging and

    persistence in White and African American first-year students. *Research in Higher*

    Education, 50(7), 649–669.

Hazari, Z., Chari, D., Potvin, G., & Brewe, E. (2020). The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates. *Journal of Research in Science Teaching, 57*(10), 1583–1607.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M.-C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching, 47*(8), 978–1003.

Hess, A. J. (2019, June 19). *Graduating in 4 years or less helps keep college costs down—But just 41% of students do*. CNBC. https://www.cnbc.com/2019/06/19/just-41percent-of-college-students-graduate-in-four-years.html

Hessenauer, S. L., & Law, K. (2017). Mentoring: A Natural Role for Learning Community Faculty. *Learning Communities: Research & Practice, 5*(2), 3.

Hoffman, M., Richmond, J., Morrow, J., & Salomone, K. (2002). Investigating "sense of belonging" in first-year college students. *Journal of College Student Retention: Research, Theory & Practice, 4*(3), 227–256.

Holland, D., Lachicotte Jr, W., Skinner, D., & Cain, C. (2001). *Identity and agency in cultural worlds*. Harvard University Press.

Hora, M. T., Wolfgram, M., & Thompson, S. (2017). *What do we know about the impact of internships on student outcomes?* Results from a preliminary review of the scholarly and practitioner literatures. Center for Research on College-Workforce Transitions Research Brief, 2, 1–20.

Hurtado, S. S., Gonyea, R. M., Graham, P. A., & Fosnacht, K. (2020). The Relationship Between Residential Learning Communities and Student Engagement. *Learning Communities Research and Practice, 8*(1), 1–18.

Inkelas, K. (2008). *National Study of Living-Learning Programs*: 2007 report of findings. Retrieved from National Study of Living-Learning Programs website: http://www.livelearnstudy.net/images/2007_NSLLP_Final_Report.pdf.

Inkelas, K. K., Daver, Z. E., Vogt, K. E., & Leonard, J. B. (2007). Living–learning programs and first-generation college students' academic and social transition to college. *Research in Higher Education, 48*(4), 403–434.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3–20.

Jaijairam, P. (2016). First-Year Seminar (FYS)–The Advantages That This Course Offers. *Journal of Education and Learning, 5*(2), 15–23.

Jaiswal, A., Karabiyik, T., Thomas, P., & Magana, A. J. (2021). Characterizing Team Orientations and Academic Performance in Cooperative Project-Based Learning Environments. *Education Sciences, 11*(9), 520. https://doi.org/10.3390/educsci11090520

Jaiswal, A., Lyon, J. A., Magana, A. J., Ellen, G., & Ward, M. D. (2021). Student Experiences within a Data Science Learning Community: A Communities of Practice Perspective. *Learning Communities Research and Practice, 9*(1), 1–38

Jaiswal, A., Lyon, J. A., Perera, V., Magana, A. J., Gundlach, E., & Ward, M. D. (2021). Work in Progress: Evaluating Student Experiences in a Residential Learning Community: A Situated Learning Perspective. *2021 ASEE Virtual Annual Conference Content Access*.

Jaiswal, A., Lyon, J. A., Zhang, Y., & Magana, A. J. (2021). Supporting student reflective

    practices through modelling-based learning assignments. *European Journal of*

    Engineering Education, 1–20.

Jaiswal, A., Thomas, P. J., Karabiyik, T., Perera, V., & Magana, A. J. (2021). Assessing the

    Impact of Transition from Face-to-Face to Online Instruction on Team Cooperation. *2021*

    ASEE Virtual Annual Conference Content Access.

Jessup-Anger, J. E. (2015). Theoretical Foundations of Learning Communities. *New Directions*

    for Student Services, 2015(149), 17–27. https://doi.org/10.1002/ss.20114

Joffe, H. (2012). *Thematic analysis. Qualitative Research Methods in Mental Health and*

    Psychotherapy: A Guide for Students and Practitioners, 1, 210–223.

Jones, R. (1995). Why do qualitative research? British Medical Journal Publishing Group.

    https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2550075/pdf/bmj00599-0006.pdf

Jurado-Navas, A., & Munoz-Luna, R. (2017). Scrum Methodology in Higher Education:

    Innovation in Teaching, Learning and Assessment. *International Journal of Higher*

    Education, 6(6), 1–18.

Karabiyik, T., Jaiswal, A., Thomas, P., & J Magana, A. (2020). Understanding the Interactions

    between the Scrum Master and the Development Team: A Game-Theoretic Approach.

    Mathematics, 8(9), 1553.

Kelly, K., Terre Blance, M., & Durrheim, K. (2006). *Why qualitative research?* Research in

    Practice: Applied Methods for the Social Sciences, 2, 271–284.

Kelly, T., & Howie, L. (2007). Working with stories in nursing research: Procedures used in

    narrative analysis. I*nternational Journal of Mental Health Nursing, 16*(2), 136–144.

Kim, A. Y., & Sinatra, G. M. (2018). Science identity development: An interactionist approach. I*nternational Journal of STEM Education, 5*(1), 1–6.

Kim, Y. S., & Merriam, S. B. (2010). Situated learning and identity development in a Korean older adults' computer classroom. *Adult Education Quarterly, 60*(5), 438–455.

Kirk, D., & Macdonald, D. (1998). Situated Learning in Physical Education. *Journal of Teaching in Physical Education, 17*(3), 376–387. https://doi.org/10.1123/jtpe.17.3.376

Koby, M., & Orit, H. (2020). *Ten Challenges of Data Science Education.* https://cacm.acm.org/blogs/blog-cacm/246219-ten-challenges-of-data-science-education/fulltext

Kohler, D. M. (2010). An application of Vygotsky's social learning theory on calculator self-efficacy and calculator achievement by gender. *Dissertations, Theses and Capstone Projects*. 413. https://digitalcommons.kennesaw.edu/etd/413

Komives, S. R., Owen, J. E., Longerbeam, S. D., Mainella, F. C., & Osteen, L. (2005). Developing a leadership identity: A grounded theory. *Journal of College Student Development, 46*(6), 593–611.

Komolthiti, M. (2016). *Leadership journeys: A narrative research study exploring women school superintendent's meaning-making of leadership development experiences*. Northeastern University.

Kuh, G. D. (2001). Assessing what really matters to student learning inside the national survey of student engagement. Change: *The Magazine of Higher Learning, 33*(3), 10–17.

Kuh, G. D. (2008). *Excerpt from high-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges and Universities, 14(3), 28–29.

Lave, J. (1991). *Situating learning in communities of practice.* In L. B. Resnick, J. M. Levine, &
S. D. Teasley (Eds.), Perspectives on socially shared cognition. (pp. 63–82). American
Psychological Association. https://doi.org/10.1037/10096-003

Lave, J. (1996). Teaching, as learning, in practice. *Mind, Culture, and Activity, 3*(3), 149–164.

Lave, J. (2004). *Situating learning in communities of practice* (pp. 63–82). American
Psychological Association. https://doi.org/10.1037/10096-003

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.*
Cambridge University Press.

Lawthom, R. (2011). Developing learning communities: Using communities of practice within
community psychology. I*nternational Journal of Inclusive Education, 15*(1), 153–164.

Leidenfrost, B., Strassnig, B., Schütz, M., Carbon, C.-C., & Schabmann, A. (2014). The Impact
of Peer Mentoring on Mentee Academic Performance: Is Any Mentoring Style Better
than No Mentoring at All?. *International Journal of Teaching and Learning in Higher
Education, 26*(1), 102–111.

Lenning, O. T., & Ebbers, L. H. (1999). *The Powerful Potential of Learning Communities:
Improving Education for the Future*. ASHE-ERIC Higher Education Report, Vol. 26, No.
6. ERIC.

Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009).
Evolution of Wenger's concept of community of practice. *Implementation Science : IS,
4*(1), 11–11. https://doi.org/10.1186/1748-5908-4-11

Li, L., Worch, E., Zhou, Y., & Aguiton, R. (2015). How and why digital generation teachers use
technology in the classroom: An explanatory sequential mixed methods study.
International Journal for the Scholarship of Teaching and Learning, 9(2), n2.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalist inquiry*. Beverly Hills, CA: Sage.

Lindsay, L. N. (2001). *Transformation of learners in a community of practice occupational therapy fieldwork environment.* [PhD Thesis]. University of Georgia

Liu, R. (2010). *Alienation and First-Year Student Retention.* Professional File. Number 116, Spring. In Association for Institutional Research (NJ1). Association for Institutional Research. https://eric.ed.gov/?id=ED511456

Lord, R. G., & Hall, R. J. (2005). Identity, deep structure and the development of leadership skill. *The Leadership Quarterly, 16*(4), 591–615.

Lyons, J. B. (2012). *A Study of Undergraduate Mentoring and the Factors Affecting Students' Disengagement from Mentoring.* [PhD Thesis] St. John's Fisher College. https://fisherpub.sjfc.edu/education_etd/121

Magana, A. J., Jaiswal, A., Madamanchi, A., Parker, L. C., Gundlach, E., & Ward, M. D. (2021). Characterizing the psychosocial effects of participating in a year-long residential research-oriented learning community. *Current Psychology,* 1–18.

Maltby, J. L., Brooks, C., Horton, M., & Morgan, H. (2016). Long Term Benefits for Women in a Science, Technology, Engineering, and Mathematics Living-Learning Community. *Learning Communities: Research & Practice, 4*(1), 2.

Mann, S. J. (2001). Alternative perspectives on the student experience: Alienation and engagement. *Studies in Higher Education, 26*(1), 7–19.

McDonald, J., & Cater-Steel, A. (2016). *Communities of practice: Facilitating social learning in higher education.* Springer.

Meishar-Tal, H., & Pieterse, E. (2019). Faculty-Students Online Communication Channels. International Association for Development of the Information Society. *Paper presented at the International Association for Development of the Information Society (IADIS) 6th International Conference on Educational Technologies* (Hong Kong, Feb 8-10, 2019)

Merriam, S. B. (2002). *Introduction to qualitative research*. Qualitative Research in Practice: Examples for Discussion and Analysis, 1(1), 1–17.

Morrow, J., & Ackermann, M. (2012). Intention to persist and retention of first-year students: The importance of motivation and sense of belonging. *College Student Journal, 46*(3), 483–491.

Moulton, D., Corcoran, P., James, P., & Rathore, P. (2017). A community of practice for blended science and engineering learning and teaching at UniSA. In *Implementing Communities of Practice in Higher Education* (pp. 253–280). Springer.

Mueller, R. A. (2019). Episodic narrative interview: Capturing stories of experience with a methods fusion. *International Journal of Qualitative Methods, 18,* 1609406919866044.

Mullen, C. A. (2007). Naturally occurring student-faculty mentoring relationships: A literature review. *The Blackwell Handbook of Mentoring: A Multiple Perspectives Approach,* 119–138.

National Academies of Sciences, Engineering, and Medicine. (2018). D*ata science for undergraduates: Opportunities and options*. National Academies Press.

National Center for Education Statistics. (n.d.). *Number and percentage distribution of science, technology, engineering, and mathematics (STEM) degrees/certificates conferred by postsecondary institutions, by race/ethnicity, level of degree/certificate, and sex of student*: 2010-11 through 2019-20. https://nces.ed.gov/programs/digest/d21/tables/dt21_318.45.asp

O'Donnell, V. L., & Tobbell, J. (2007). The transition of adult students to higher education: Legitimate peripheral participation in a community of practice? *Adult Education Quarterly, 57*(4), 312–328.

O'Keeffe, P. (2013). A Sense of Belonging: Improving Student Retention. *College Student Journal, 47*(4), 605–613.

Orsmond, P., Merry, S., & Callaghan, A. (2013). Communities of practice and ways to learning: Charting the progress of biology undergraduates. *Studies in Higher Education, 38*(6), 890–906.

Ortega-Alvarez, J. D. (2019). *Conceptions of teaching among Colombian engineering faculty: An exploratory study* [PhD Thesis]. Purdue University Graduate School.

Osterman, K. F. (2010). Teacher practice and students' sense of belonging. *In International research handbook on values education and student wellbeing* (pp. 239–260). Springer.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193.

Pargetter, R. (2000). Transition: From a school perspective. *Journal of Institutional Research, 9*(1), 14–21.

Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence and voluntary

    dropout decisions from a theoretical model. *The Journal of Higher Education, 51*(1), 60–

    75.

Pattison, S., Gontan, I., Ramos-Montañez, S., Shagott, T., Francisco, M., & Dierking, L. (2020).

    The identity-frame model: A framework to describe situated identity negotiation for

    adolescent youth participating in an informal engineering education program. J*ournal of*

    the Learning Sciences, 1–48.

Petty, J. (2017). Creating stories for learning about the neonatal care experience through the eyes

    of student nurses: An interpretive, narrative study. *Nurse Education Today, 48,* 25–32.

Petty, J., Jarvis, J., & Thomas, R. (2018). Core story creation: Analysing narratives to construct

    stories for learning. *Nurse Researcher, 25*(4), 47–51.

Phoenix, C., Smith, B., & Sparkes, A. C. (2010). Narrative analysis in aging studies: A typology

    for consideration. J*ournal of Aging Studies, 24*(1), 1–11.

Pike, G., Kuh, G., & McCormick, A. (2008). Learning community participation and educational

    outcomes: Direct, indirect and contingent relationships. *Association for the Study of*

    Higher Education Conference in Jacksonville, FL.

Pike, G. R. (1999). The effects of residential learning communities and traditional residential

    living arrangements on educational gains during the first year of college. *Journal of*

    College Student Development, 40(3), 269.

Pistilli, M. D. (2009). *How female learning community students from the colleges of engineering,*

    science, and technology experience Purdue university: A qualitative dissertation [PhD

    Thesis]. Purdue University.

PITAC, U. S. P. I. T. A. (2005). *Computational Science: Ensuring America's Competitiveness. National Coordination Office for Information Technology Research & Development.*

Priest, K. L. (2012). *Constructing Leadership Identities through Participation in a Leadership Living-Learning Community* [PhD Thesis]. Virginia Polytechnic Institute and State University.

Priest, K. L., & de Campos Paula, A. L. (2016). Peer-led learning communities: Exploring integrative high-impact educational practices for leadership education. *Journal of Leadership Education, 15*(1), 86–95.

Priest, K. L., Saucier, D. A., & Eiselein, G. (2016). Exploring Students' Experiences in First-Year Learning Communities From a Situated Learning Perspective. I*nternational Journal of Teaching and Learning in Higher Education, 28*(3), 361–371.

Psaromiligkos, Y., Orfanidou, M., Kytagias, C., & Zafiri, E. (2009). Mining log data for the analysis of learners' behaviour in web-based learning management systems. *Operational Research, 11*(2), 187–200. https://doi.org/10.1007/s12351-008-0032-4

Purdue University (n.d.). *Diversity, Inclusion and Belonging: Campus Population Overview.* https://www.purdue.edu/diversity-inclusion/about-us/stats.html

Rehak, A. M. (2018). *Becoming and Being an Inspire Student: Learning, Participation, and Engagement in a Living-Learning Community* [PhD Thesis]. Indiana University.

Ridwan, M. R., & Retnawati, H. (2021). Application of Cluster Analysis Using Agglomerative Method. Numerical: *Jurnal Matematika Dan Pendidikan Matematika,* 33–48.

Riessman, C. K. (2008). *Narrative methods for the human sciences*. Sage Publications.Inc

Rocconi, L. M. (2011). The impact of learning communities on first year students' growth and development in college. *Research in Higher Education, 52*(2), 178–193.

Roth, W.-M., & McGinn, M. K. (1997). Graphing: Cognitive ability or practice? *Science Education, 81*(1), 91–106. https://doi.org/10.1002/(SICI)1098-237X(199701)81:1<91::AID-SCE5>3.0.CO;2-X

Scanlon, L., Rowling, L., & Weber, Z. (2007). 'You don't have like an identity… you are just lost in a crowd': Forming a student identity in the first-year transition to university. *Journal of Youth Studies, 10*(2), 223–241.

Schneider, M. (2010). *Finishing the First Lap: The Cost of First Year Student Attrition in America's Four Year Colleges and Universities*. In American Institutes for Research. American Institutes for Research. https://eric.ed.gov/?id=ED512253

Schneiderman, J. (2020). *Improving Sense of Belonging Among Freshmen at a Large, Urban High School Through Service Learning*. [PhD Thesis]. John Hopkins University. http://jhir.library.jhu.edu/handle/1774.2/62599

Scott, M. J., & Ghinea, G. (2014). Measuring enrichment: The assembly and validation of an instrument to assess student self-beliefs in CS1. *Proceedings of the Tenth Annual Conference on International Computing Education Research, 123–130*.

Serpe, R. T., & Stryker, S. (2011). The symbolic interactionist perspective and identity theory. *In Handbook of identity theory and research* (pp. 225–248). Springer.

Seymour, E., Hunter, A.-B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education, 88*(4), 493–534.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*(2), 63–75.

Simina, V. K. (2012). Socio-Constructivist Models of Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3128–3131). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_882

Singh, A. (2018). *Investigating the sense of belonging of international students through a predictive model* (p. 12331544) [Doctor of Philosophy, Iowa State University, Digital Repository]. https://doi.org/10.31274/etd-180810-6097

Sluss, D. M., & Ashforth, B. E. (2007). Relational identity and identification: Defining ourselves through work relationships. *Academy of Management Review, 32*(1), 9–32.

Smith, B. L. (2001). The Challenge of Learning Communities as a Growing National Movement. *AAC&U Peer Review 4* (1).

Smith, S. U., Hayes, S., & Shea, P. (2017). *A Critical Review of the Use of Wenger's Community of Practice (CoP) Theoretical Framework in Online and Blended Learning Research,* 2000-2014. Online Learning (Newburyport, Mass.), 21(1), 209-. https://doi.org/10.24059/olj.v21i1.963

Solanki, S., McPartlan, P., Xu, D., & Sato, B. K. (2019). Success with EASE: Who benefits from a STEM learning community? *PloS One, 14*(3), e0213827.

Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative Inquiry, 4*(2), 280–292.

Spanierman, L. B., Soble, J. R., Mayfield, J. B., Neville, H. A., Aber, M., Khuri, L., & De La Rosa, B. (2013). Living learning communities and students' sense of community and belonging. *Journal of Student Affairs Research and Practice, 50*(3), 308–325.

Stahl, G. K., & Maznevski, M. L. (2021). Unraveling the effects of cultural diversity in teams: A retrospective of research on multicultural work groups and an agenda for future research. *Journal of International Business Studies, 52*(1), 4–22.

Stassen, M. L. (2003). Student outcomes: The impact of varying living-learning community models. *Research in Higher Education, 44*(5), 581–613.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*(6), 859.

Taylor, T., & Stamatoplos, T. (1999). First-year learning communities: Redefining the educational roles of academic librarians. *ACRL National Conference Proceedings*. https://via.library.depaul.edu/lib_pubs/42

Teeuwsen, P., Ratković, S., & Tilley, S. A. (2014). Becoming academics: Experiencing legitimate peripheral participation in part-time doctoral studies. *Studies in Higher Education, 39*(4), 680–694.

Thelin, J. R. (2011). *A history of American higher education*. JHU Press.

Thelin, J. R., & Gasman, M. (2003). Historical overview of American higher education. *Student Services: A Handbook for the Profession*, 4, 3–22.

Tight, M. (2020). Student retention and engagement in higher education. *Journal of Further and Higher Education, 44*(5), 689–704.

Tinnell, T. L., Ralston, P. A., Tretter, T. R., & Mills, M. E. (2019). Sustaining pedagogical change via faculty learning community. *International Journal of STEM Education, 6*(1), 1–16.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. ERIC.

Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *The Journal of Higher Education, 68*(6), 599–623.

Tinto, V. (2006). Research and Practice of Student Retention: What Next?*. *Journal of College Student Retention: Research, Theory and Practice, 8*(1), 1–19. https://doi.org/10.2190/4YNU-4TMB-22DJ-AN4W

Tinto, V., Goodsell-Love, A., & Russo, P. (1994). *Building learning communities for new college students.* PSU: National Center on Postsecondary Teaching, Learning, and Assessment.

Tomkin, J. H., Beilstein, S. O., Morphew, J. W., & Herman, G. L. (2019). Evidence that communities of practice are associated with active learning in large STEM lectures. *International Journal of STEM Education, 6*(1), 1–15.

Trostel, P. (2015). It's Not Just the Money. *Government & Civic Life*. 4. https://digitalcommons.library.umaine.edu/mcspc_gov_civic/4

Ujj, C. (2020). *Impact of Living-Learning Communities on Second-Year Students* [PhD Thesis]. University of Georgia.

Virtue, E. E., Maddox, G., & Pfaff, K. (2019). The Lasting Effects of Learning Communities. *Learning Communities Research and Practice, 7*(2), 6.

Vygotsky, L. S. (1978). Mind in society. *In The development of higher psychological processes.* Harvard University Press.

Vygotsky, L. S. (1980). Mind in society: *The development of higher psychological processes*. Harvard University Press.

Walker, C. A. (2018). *Impostor Phenomenon, Academic Self-Efficacy, and Persistence Among African-American Female Undergraduate STEM Majors* [PhD Thesis]. Northeastern University.

Wang, M.-T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119–140.

Weidman, J. (1989). Undergraduate socialization: A conceptual approach. *Higher Education: Handbook of Theory and Research*, 5(2), 289–322.

Weidman, J. C., DeAngelo, L., & Bethea, K. A. (2014a). The Weidman model of undergraduate socialization: Continuity and change. *Annual Meeting of the American Educational Research Association, Philadelpha, PA*.

Weidman, J. C., DeAngelo, L., & Bethea, K. A. (2014b). Understanding student identity from a socialization perspective. *New Directions for Higher Education, 2014*(166), 43–51.

Weidman, J. C., & Stein, E. L. (2003). Socialization of doctoral students to academic norms. *Research in Higher Education, 44*(6), 641–656.

Wellmon, R., Gilin, B., Knauss, L., & Linn, M. (2009). The benefits of an interdisciplinary collaborative learning experience: The student perspective on outcomes. International *Journal of Interdisciplinary Social Sciences, 4*(8), 15–27.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems Thinker, 9(*5), 2–3.

Wenger, E. (2004). Knowledge management as a doughnut: Shaping your knowledge strategy through communities of practice. *Ivey Business Journal, 68*(3), n-a.

Wenger, E. (2011). *Communities of practice: A brief introduction.*
http://hdl.handle.net/1794/11736

Wenger, E., Trayner, B., & Laat, M. de. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework.* Senior Deputy Vice-Chancellor and Deputy Vice-Chancellor (Education) - Papers. https://ro.uow.edu.au/asdpapers/596

Wenger, E., & Wenger, B. (2015). *Communities of Practice. A Brief Introduction.* http://wenger-trayner.com/wp-content/uploads/2015/04/07-Brief-introduction-to-communities-of-practice.pdf

Whannell, R., & Whannell, P. (2015). Identity theory as a theoretical framework to understand attrition for university students in transition. *Student Success, 6*(2), 43–53.

Whipple, W. R. (1987). *Collaborative Learning: Recognizing It When We See It.* In AAHE Bulletin.

Williford, A. M., & Wadley, J. Y. (2008). *How Institutional Research Can Create and Synthesize Retention and Attrition Information.* Professional File Number 108, Fall 2008. In Association for Institutional Research (NJ1). Association for Institutional Research. https://eric.ed.gov/?id=ED504412

Xie, K., Lu, L., Cheng, S.-L., & Izmirli, S. (2017). The interactions between facilitator identity, conflictual presence, and social presence in peer-moderated online collaborative learning. *Distance Education, 38*(2), 230–244.

Yin, Y. (2018). *Behavioral targeting: Early student profile identification for efficient student recruitment* [Master's Thesis]. University of Twente.

Zammit, K., Vickers, M., Hibbert, E., & Power, C. (2017). Equity buddies: Building

    communities of practice to support the transition and retention of students through their

    first year at university. In *Implementing Communities of Practice in Higher Education*

    (pp. 373–394). Springer.

Zheng, R. (2010). Effects of Situated Learning on Students' Knowledge Acquisition: An

    Individual Differences Perspective. *Journal of Educational Computing Research, 43*(4),

    467–487. https://doi.org/10.2190/EC.43.4.c