A Data Driven Approach to Understanding the Attrition of Women in Software Engineering

by

Madeleine A. Golison

B.S. Computer Science (2016)
Tufts University

Submitted to the System Design and Management Program
in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

at the
Massachusetts Institute of Technology

May 2024

© 2024 Madeleine Golison. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Madeleine A. Golison
MIT System Design and Management Program
May 10, 2024

Certified by: Bruce G. Cameron
Senior Lecturer
Thesis Supervisor

Accepted by: Joan Rubin
Executive Director
MIT System Design and Management Program
A Data Driven Approach to Understanding the Attrition of Women in Software Engineering

by

Madeleine A. Golison

Submitted to the MIT System Design and Management Program on May 10, 2024

in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

Abstract

Data from large tech companies shows that 15% or fewer software engineers are women. While Tech companies blame the university pipeline, studies from McKinsey and Accenture found that Tech company “bro culture” was influencing the pipeline of women out of Tech. However, in the MIT Women in Software Engineering survey, of the 183 respondents, most women reported planning on staying in Tech when leaving SWE roles. This formed the hypothesis that female software engineers were leaving SWE roles for reasons other than “bro culture.” Understanding and improving the attrition of women in the software engineering career path is important because the representation of women in the field is already so small, so any attrition is consequential.

Overall, many factors were found to have influenced the retention of women in software engineering roles. Notably, culture was not the most important reason for women leaving the software engineering career path. The primary reason directly stated in the open-ended survey responses was “burnout,” but this was closely followed by reasons such as finding other opportunities outside of Tech, a desire for better work-life balance, and the lack of diversity. While these explicitly stated reasons were easily noted, predictive models (using logistic regression and tree-based methods) were needed to identify factors that were not obvious to respondents.

The predictive models identified the primary reasons women leave SWE roles by comparing women who planned to remain in the SWE career path and those who did not. The top reasons identified were not enjoying programming, believing that better opportunities existed outside of software engineering, and having their team co-located. The last reason, team co-location, was identified as being related to various other environment factors related to imposter syndrome and was likely a proxy for these other factors. Women in the age range of 25 – 44 seemed to be particularly at risk of leaving the career path, and between the general population and the specific 25 – 34 and 35 – 44 age groups, each had different factors that were most important.

Given these results, several recommendations exist for improving attrition for women in the software engineering career path. The key recommendations include improving manager feedback processes, diversity, work-life balance, and opportunities to work on high-visibility initiatives.

Thesis Supervisor: Bruce G. Cameron
Title: Senior Lecturer
Acknowledgments

First, I would like to thank my thesis advisor, Professor Bruce Cameron. A year and a half ago, I came to Professor Cameron with a thesis topic that I knew aligned directly with his work. When we first spoke, Professor Cameron asked me what other questions I had considered. Upon hearing my initial thesis idea (this one), he encouraged me to investigate more and offered his support for this thesis. Without Professor Cameron’s initial encouragement and constant advice and support, this thesis would not exist! Thank you, Bruce.

Second, I would like to thank Joan Rubin for all the help navigating the SDM program and the thesis these last few years. Joan has been a wonderful guide; her advice and expertise on all things MIT has been crucial for me. I would also like to thank my SDM peers, especially those who spent many weekends working away on their theses with me. Their support and company allowed for constant progress on this thesis.

Next, I would like to thank my family and friends. My parents, sisters, and in-laws always asked how I was progressing and continuously encouraged me. A special thank you is warranted for Professor Andrew Philips, my brother-in-law, who offered his expertise in methods by reviewing Chapter 3: Methods.

Finally, I would like to thank my husband, Michael. He has been my biggest cheerleader since long before I came to MIT, but I certainly needed that support most in the last few years. From taking on more at home to listening to my thesis ideas and real-world problems that related to it, Michael was a big reason I made it here.
# Table of Contents

ABSTRACT ................................................................................................................................. 3  
ACKNOWLEDGMENTS .............................................................................................................. 4  
TABLE OF CONTENTS ............................................................................................................. 5  
LIST OF FIGURES .................................................................................................................... 7  
LIST OF TABLES ....................................................................................................................... 9  
ACRONYMS, TERMS, AND DEFINITIONS .................................................................................. 10  

## CHAPTER 1 INTRODUCTION ................................................................................................. 11  
1.1 PERSONAL MOTIVATION .................................................................................................... 11  
1.2 RESEARCH OVERVIEW .................................................................................................... 12  
1.3 RESEARCH MOTIVATION ................................................................................................ 13  
1.4 THESIS HYPOTHESES AND APPROACH ....................................................................... 14  
1.5 RESEARCH SCOPE AND LIMITATIONS ......................................................................... 16  
1.6 THESIS STRUCTURE ......................................................................................................... 17  

## CHAPTER 2 LITERATURE REVIEW ..................................................................................... 20  
2.1 OVERVIEW OF THE PROBLEM ....................................................................................... 20  
2.2 THE FUNNEL INTO TECH .............................................................................................. 22  
2.3 THE FUNNEL OUT OF TECH ......................................................................................... 26  
2.4 THE CURRENT EXPERIENCE OF WOMEN IN TECH .................................................... 26  
2.5 THE RESEARCH GAP: WHAT WE DON’T KNOW ............................................................ 46  

## CHAPTER 3 RESEARCH METHODS ...................................................................................... 47  
3.1 OVERVIEW ....................................................................................................................... 47  
3.2 UNANSWERED QUESTIONS ............................................................................................ 47  
3.3 MAIN HYPOTHESES ...................................................................................................... 49  
3.4 SURVEY DESIGN ............................................................................................................. 50  
3.5 DATA CLEANING ............................................................................................................ 55  
3.6 QUANTITATIVE METHODS ............................................................................................. 56  
3.7 METHODS FOR TEXT ANALYSIS .................................................................................. 62  
3.8 METHODS LIMITATIONS ............................................................................................... 65  

## CHAPTER 4 RESULTS DEMOGRAPHICS ............................................................................. 68  
4.1 RESULTS OVERVIEW ...................................................................................................... 68  
4.2 CLEANED DATA .............................................................................................................. 68  
4.3 DEMOGRAPHICS OVERVIEW ........................................................................................ 69  
4.4 CURRENT CAREER DEMOGRAPHICS ........................................................................... 69  
4.5 DEPENDENT VARIABLE DEMOGRAPHICS .................................................................... 72  

## CHAPTER 5 RESULTS ANALYSIS ...................................................................................... 85  
5.1 QUANTITATIVE ANALYSIS OVERVIEW ........................................................................ 85  
5.2 RESULTS FOR ALL VARIABLES AND ALL SURVEY RESPONSES ................................. 89  
5.3 AGE GROUP MODELS ..................................................................................................... 103  
5.4 RESULTS FROM OPEN-ENDED SURVEY RESPONSES ................................................ 109  
5.5 HYPOTHESES REVIEW .................................................................................................. 117
CHAPTER 6 CONCLUSIONS .................................................................................................................. 121
  6.1 SUMMARY OF RESULTS .............................................................................................................. 121
  6.2 RECOMMENDATIONS .................................................................................................................. 123
  6.3 AREAS FOR CONTINUED RESEARCH ...................................................................................... 125
REFERENCES ..................................................................................................................................... 128
APPENDIX A: MIT WOMEN IN SOFTWARE ENGINEERING SURVEY ............................................. 133
List of Figures

Figure 1.1: Thesis Structure...........................................................................................................19
Figure 0.1: Gender population at Tech Companies(We Ran the Numbers, and There Really Is a Pipeline
Problem in Eng Hiring., n.d.) ........................................................................................................21
Figure 0.2: “Employed college graduates, by level of highest degree, minor field of highest degree, and
major occupation: 2021”(National Survey of College Graduates (NSCG) 2021 | NSF - National Science
Foundation, n.d.) .........................................................................................................................25
Figure 0.3: Women in Technical Roles(Women in the Workplace 2022, n.d.) ..............................27
Figure 0.4: Women's Experience of Microaggressions(Women in the Workplace 2018, n.d.) ........28
Figure 0.5: Feedback loop for Being an Only .................................................................................29
Figure 0.6: Gender of VC Partners Leading Funding Rounds(Women VCs Invest in Up to 2x More
Female Founders | Journal | Kauffman Fellows, n.d.) .................................................................30
Figure 0.7: Gender and Fundraising Over Time(Snellman & Solal, 2023) ......................................31
Figure 0.8: Women in VC Feedback Loop.....................................................................................32
Figure 0.9: Women's Workplace Experiences(Women in Tech | Accenture and Girls Who Code, n.d.)...35
Figure 0.10: Doubling Women in Tech by 2030(Women in Tech | Accenture and Girls Who Code, n.d.)36
Figure 0.11: Pull Request Acceptance Rate(Terrell et al., 2017) ....................................................39
Figure 0.12: Gender Neutral vs. Gendered PRs(Terrell et al., 2017) ..............................................39
Figure 0.13: Career and Family by Advanced Degree, Harvard and Beyond Fifteen Years after
College(Goldin, 2021) ..................................................................................................................42
Figure 0.14: Coding Outside of Work(Stack Overflow Developer Survey 2022, n.d.) .......................43
Figure 3.1: Dependent Variables for the MIT Women in Software Engineering Survey .................51
Figure 3.2: Reasons for leaving tech roles(Women in Tech | Accenture and Girls Who Code, n.d.) ....52
Figure 3.3: Survey Question Example .........................................................................................55
Figure 3.4: Sigmoid Function(“Advantages and Disadvantages of Logistic Regression,” 2020) ......59
Figure 3.5: Model Explainability with SHAP Values(https://github.com/shap/shap) ....................62
Figure 3.6: BERTopic Model Layers(https://maartengr.github.io/) .............................................64
Figure 4.1: What is Your Current Career? ...................................................................................70
Figure 4.2: What is your Future Career? ....................................................................................71
Figure 4.3: Respondents on if they will remain in SWE roles .......................................................72
Figure 4.4: Ages of Respondents ...............................................................................................73
Figure 4.5: Ages of all Profession Developers(Stack Overflow Developer Survey 2023, n.d.) ..........74
Figure 4.6: Respondents that plan on leaving SWE in the next three years .................................75
Figure 4.7: Dependents of those than were in the remain SWE group .........................................76
Figure 4.8: Dependents of those in the non-remain SWE group ..................................................76
Figure 4.9: Respondents highest level of schooling completed ...................................................77
Figure 4.10: When respondents first learned to program ..............................................................78
Figure 4.11: Company size reported by respondents .................................................................79
Figure 4.12: Company Size for those planning to leave SWE roles ............................................79
Figure 4.13: Company Size for those planning to remain in SWE roles .......................................80
Figure 4.14: Reported current work setting of respondents .........................................................81
Figure 4.15: Preferred work setting of respondents .....................................................................81
Figure 4.16: Respondents' years of programming experience ......................................................82
Figure 4.17: Remain SWE years of programming experience ......................................................82
Figure 4.18: Non-Remain SWE years of programming experience ..............................................83
Figure 5.1: MIT Women in Software Engineering models developed from survey data ................88
Figure 5.2: SHAP values for numerical IVs from Random Forest model ......................................92
Figure 5.3: SHAP values for categorical IVs from Random Forest model ....................................93
Figure 5.4: SHAP values for numerical IVs from the XGBoost model ................................................................. 94
Figure 5.5: SHAP values for categorical IVs from the XGBoost model ................................................................. 95
Figure 5.6: Comparing current and preferred work setting ...................................................................................... 99
Figure 5.7: Comparison of features between co-located and not co-located SWEs .............................................. 100
Figure 5.8: A comparison of work settings for women that remain in or leave SWE roles ................................ 102
Figure 5.9: Ages 25 - 34 Random Forest important features .................................................................................... 104
Figure 5.10: Ages 25-34 XGBoost important features ............................................................................................. 105
Figure 5.11: Ages 35 - 44 Random Forest important features .................................................................................. 106
Figure 5.12: Ages 35 - 44 XGBoost important features ........................................................................................... 107
Figure 5.13: Each BERTopic themes' word scores ................................................................................................. 110
Figure 5.14: The GPT model themes identified from the open-ended survey question 1: Why did you (or will you) leave a software engineering role? ........................................................................ 112
Figure 5.15: Average daily hours worked as a SWE ............................................................................................... 113
Figure 5.16: BERTopic word scores for the responses to "What is/was your favorite thing about software engineering?" .................................................................................................................. 115
Figure 5.17: GPT model outputs for the responses to "What is/was your favorite thing about software engineering?" .................................................................................................................. 115
Figure 5.18: GPT top 5 reasons for staying in SWE ............................................................................................... 116
Figure 5.19: Themes for liking SWE from the thematic analysis of text responses ........................................... 116
Figure 6.1: A visual representation of the top factors influencing women in SWE .............................................. 121
List of Tables

Table 0.1: Involvement in Household Chores by Gender(Cerrato & Cifre, 2018) ........................................45
Table 5.1: MIT Women in Software Engineering independent variables ..............................................87
Table 5.2: Logistic Regression Important Variables .................................................................................90
Table 5.3: Comparing important features across all models .................................................................96
Table 5.4: Model accuracies for age group-based models ...................................................................104
Table 5.5: The most important variables across age groups for the XGBoost models with numerical IVs
..........................................................................................................................................................108
Table 5.6: BERTopic model output ........................................................................................................110
Table 5.7: Explicitly stated reasons for leaving SWE roles from the thematic analysis of the text responses
..........................................................................................................................................................113
Table 5.8: BERTopic themes for answering the question, "What is/was your favorite thing about software
engineering?" .........................................................................................................................................114
Acronyms, Terms, and Definitions

SWE – Shorthand for software engineer

Tech – The industry that represents Tech companies.

Tech company – “A company that focuses primarily on the manufacturing, support, research and development of — most commonly computing, telecommunication and consumer electronics-based — technology-intensive products and services, which include businesses relating to digital electronics, software, optics, new energy and internet-related services such as cloud storage and e-commerce services.”

— Wikipedia article “Technology Company” (“Technology Company,” 2023)
Chapter 1  Introduction

1.1 Personal Motivation

Over the course of my eight years working in Tech, I have held an array of jobs, but I have spent most of my career as a software engineer. When I was working as a software engineer, I enjoyed my role immensely. I loved my team, the company had excellent management and the culture and values aligned well with mine. There was good representation of women in the organization, and I never endured any of the typical problems that women generally face while working in software engineering in Tech. However, I noticed that in software engineering roles specifically (at the various Tech companies I had worked for), few women remained in these roles long enough to make it to the senior or manager levels.

Many young women in software engineering appeared to be joining from university programs, but very few senior software engineering roles seemed to be held by women. I even noticed that some women who had been at my company for over ten years had begun their careers as software engineers and moved into program and product management roles instead of remaining on the SWE career track. Given that my organization did not appear to have all of the typical Tech culture issues that are blamed for women leaving Tech (Women in Tech | Accenture and Girls Who Code, n.d.), I did not understand why women were not staying in the SWE career path. I noticed the same phenomenon from many of my peers who began careers as software engineers at other companies.

I asked my ex-SWE peers why they decided to leave their SWE roles, but remain working in Tech. I did not understand why women were choosing to leave a role that appeared to be not only interesting and fun, but also quite flexible in when and where the work is done. Most of the women I spoke with could not pinpoint what pushed (or drew) them to other roles in Tech, but they planned to stay in Tech unanimously. This (anecdotally) suggested that Tech culture was not driving them out of SWE roles. I found this both intriguing and alarming. If company culture is not the reason...
for women leaving SWE roles, then solving the "poor culture" issue, which is very often cited as "the problem" (Women in Tech | Accenture and Girls Who Code, n.d.), will not help retain women in SWE roles. After making this discovery, I wanted to see if women were genuinely deciding to remain in Tech after leaving SWE roles. If this was the case, then Tech culture could not solely be blamed for women leaving SWE roles.

1.2 Research Overview

It is estimated that between 12% and 17% of software engineers at the largest software companies in the US are women (We Ran the Numbers, and There Really Is a Pipeline Problem in Eng Hiring., n.d.). While there are gender imbalances in several fields, such as firefighting (around 5% female (U.S. Fire Department Profile | NFPA Research, n.d.) and nursing (about 12% male (Data Spotlight, n.d.)), the disparity in software engineering is particularly worrying. Software companies have been an enormous and growing fraction of total GDP, a significant driver of economic productivity, and at an individual level, a driver of wage growth and economic mobility. Further, imbalances relatively early in the field of software engineering may set the future balance as stereotypes and role models become hardened over time. These are just the negative consequences for women in the workplace.

When it comes to women working in STEM roles, and more specifically computer science roles, there is a good deal of research on the university pipeline of women. There are certainly fewer women being trained in computer science, and as such, Tech companies tout this as the reason for the low representation of women in software engineering roles (Varma & Hahn, 2008). While the “pipeline problem” is indeed a problem there are still more women being trained in computer science than there are hold SWE roles (as outlined in the literature review). So, the pipeline cannot be purely to blame.

There are an additional two problems with the current research on women in software engineering. The existing research often groups software engineering with traditional engineering fields when analyzing the work experiences of women in STEM. However, the experiences of women in software engineering can vary quite dramatically from those in traditional engineering fields.
Second, the research also groups together the experiences of all women in Tech (not just SWEs) when analyzing women's experiences working in Tech. The problem with grouping women this way is that women in software engineering specific roles might have vastly different experiences from women in other areas of the business. There is not existing research that tries to address these potential differences.

From sources such as the 2020 Accenture report, the viral blog post about Susan Fowler at Uber, and the Women in the Workplace McKinsey Report it is known that women (in general) leave Tech because of the reported “bad” culture. However, my hypothesis is that female software engineers are leaving software engineering roles for reasons more nuanced than those that have been previously identified, such as culture. The reason we need to understand the attrition of women in SWE roles specifically is because the pipeline is already too small. Fixing the experiences of female software engineers will be critical in fixing the pipeline out of software engineering.

### 1.3 Research Motivation

The university pipeline into computer science is known to have many issues, but those have been well studied. What happens to women after they receive training in computer science and are hired as SWEs in Tech is a mystery. The MIT Women in Software Engineering survey quickly proved the first key hypothesis as being correct: female SWEs plan to (or do) leave software engineering at significant rates. Of the current SWE respondents surveyed, 16% planned to leave software engineering in the next three years (survey result are available in Chapter 5).

For society, more generally, having poor representation of women and under-represented groups in the development of software products can lead to more disastrous consequences. When a group lacks representation in the development of products these products are less likely to be built with them in mind. For example, when the Apple credit card debuted in 2019 it was shown to have bias because women were being offered smaller lines of credit (*The Apple Card Didn’t “See” Gender—and That’s the Problem* | *WIRED*, n.d.).
The representation of women in software engineering is already abysmal. As a result, any attrition of women from the career is significant. If the reasons why women leave software engineering roles can be better understood, then perhaps interventions to combat these reasons might improve the rates of women that remain in software engineering careers. The university pipeline is only one part of the puzzle--another piece is encouraging female software engineers to stay in the career path.

### 1.4 Thesis Hypotheses and Approach

The primary hypothesis generated helped frame the context to the secondary hypotheses: women leave SWE predominantly for other roles in Tech. If this proved to be true then it would suggest hypothesis 1.a, culture is not the most important reason for why women leave software engineering roles. Given that the preliminary analysis found that women were leaving SWE roles for other roles in Tech, the factors for why women were leaving SWE roles applied to SWE roles only and not Tech altogether.

This framing is important when considering the findings from the other hypotheses. For example, if this hypothesis is true: *disliking working with your manager relates to leaving or wanting to leave a SWE role*, since we know that these women are leaving SWE for other roles in Tech, then having a bad manager might be a SWE problem specifically and would not be a reason for women leaving Tech in general.

#### 1.4.1 Primary Hypotheses

- The majority of female software engineers that have left or hope to leave software engineering roles go to or plan to go to non-software engineering roles in Tech, but not leave Tech altogether
- Culture is not the most important reason for why women leave software engineering roles
1.4.2 Secondary Hypotheses

- Disliking working with your manager relates to leaving or wanting to leave a SWE role.
- Female software engineers that work more hours are more likely to have left (or plan to leave) software engineering roles.
- Female software engineers that report less flexibility in their work environment are more likely to have left (or plan to leave) software engineering roles.
- Women leave software engineering roles not because they dislike coding.
- Women that stay in software engineering (or plan to) are mostly remote.
- Women with dependents do not stay in software engineering roles as much as those who do not.
- Women with no other women on their direct teams are more likely to leave software engineering roles.

The approach for investigating these hypotheses was through a survey designed with these hypotheses in mind. The survey had three sets of questions:

1. General background questions about the respondent (mostly on a Likert scale). These questions were hoping to find relationships between leaving SWE roles and a variety of work factors without explicitly asking the respondents. Since many women cannot pinpoint why they want to leave SWE roles, these questions were helpful for identifying those unknowns.
2. Direct questions about work experience (on a Likert scale)
3. Open ended text responses asking respondents why they planned to leave or remain in SWE roles

This work employed several predictive methods, natural language processing methods and qualitative thematic analyses to analyze the survey data. Based on the overall top factors identified across the different analysis methods, several initial recommendations were made for improving rates of women staying in software engineering in Tech.

The population of the MIT Women in Software Engineering Survey and subsequent analysis was women that have worked in Tech as a software engineer, whether currently or in the past. This
population was of interest as data about current software engineers exhibited the experiences of women that planned to stay in the software engineering career and those that planned to leave. Previous female software engineers were also included in the survey population to ensure the representation of women who had left the software engineering career path and were not just "planning" to. All respondents had at least two years of experience in the software engineering role.

1.5 Research Scope and Limitations

The scope of this research was identifying 1. if women that leave software engineering roles predominantly stay in Tech and 2. for the populations of women that left SWE roles, what were their main reasons why. This also included the reasons why women in software engineering decided to stay. The women represented by the research were based in the US.

Since only women were surveyed it is not known whether men have similar attrition rates; however, it is not in scope as we are not trying to prove that women are having better or worse experiences as men in SWE roles. The attrition rates and experiences of men are out of scope as their experiences relative to women do not change the outcome of this study. However, we then cannot assert whether women's experiences are better or worse than men in software engineering roles in Tech.

Other major limitations were related to the survey itself. The distribution of the survey was mostly in "women in Tech" spaces, such as reddit subgroups, LinkedIn groups and other similar forums. The problem with this distribution is that it was difficult to find respondents that had left Tech altogether as they are much less likely to frequent online Tech communities. So, the results cannot say much about women that have left Tech altogether. Also, almost no current SWE respondents said they planned to leave Tech altogether, so the results are very much focused on women that leave SWE for other roles in Tech.

The second limitation with the survey was the number of responses. While there were enough responses to perform the quantitative data analysis, the simpler and more interpretable models would have performed better with 5 - 10x more data points.
1.6 Thesis Structure

This paper is organized into six main chapters that are outlined below in Figure 1.1. The first is Chapter 1, the Introduction, which is the current chapter. Following the first chapter is Chapter 2: Literature Review.

The literature review has four main sections. The first section frames the problem(s) given what is known from existing research. The second two sections outline the pipelines of women coming into and leaving Tech and software engineering roles. The fourth section then focuses on the current experiences of women in Tech.

Chapter 3 describes and evaluates the different research methods. The first section outlines the survey design and question sections. The second section steps through the same hypotheses outlined above. The last two sections outline the quantitative and qualitative methods used to analyze the survey responses.

Chapter 4 is the demographics analysis from the MIT Women in Software Engineering survey. The chapter first analyzes the entire population of the survey and then compares the population groups based on whether they were still in software engineering roles and whether they planned to be in the future.

In Chapter 5 the independent variables that were identified from the survey questions and were analyzed. This included an analysis of variables based on the outputs of predictive methods applied to the survey data. The predictive methods were then compared. The survey text response analysis followed, which contained natural language processing as well as a qualitative thematic analysis of the responses. Finally, a summary of the results of the original hypotheses was done.
The last chapter, Chapter 6, is the conclusion. The conclusion provides a short summary of all the results found in Chapters 4 and 5, and then provides various potential recommendations for combating the most significant factors identified for why women leave SWE roles.
Figure 1.1: Thesis Structure
Chapter 2  Literature Review

Examining the current research around women working in Tech discovered gaps in understanding why women leave software engineering roles, specifically, and where the current research does not generalize well. Before analyzing the existing research, understanding why it matters that women leave software engineering roles is important—hint: it has to do with the fact that there are so few to begin with. Digging into the funnel of women into and out of tech shows where women are "falling off" of the software engineering career path (spoiler: it is not explicitly a university pipeline problem). Finally, a deep dive into the current experiences of women in Tech was done to show the known, general reasons for why women leave Tech. While the existing research defines general reasons for why women leave Tech, a major gap in the research was revealed as there is not much information on why women leave software engineering roles specifically.

2.1 Overview of the Problem

What is considered a Tech Company? Wikipedia defines a Tech Company as:

A company that focuses primarily on the manufacturing, support, research and development of — most commonly computing, telecommunication and consumer electronics-based — technology-intensive products and services, which include businesses relating to digital electronics, software, optics, new energy and internet-related services such as cloud storage and e-commerce services.

— Wikipedia article “Technology Company” (“Technology Company,” 2023)

While this definition is not representative of all views, it is the best available definition for the commonly shared understanding of what Tech is and it is what I will be referring to when the term "Tech" is used in the rest of the thesis. The major reason for the definition and distinction is that the experiences of women in software engineering may be vastly different in other industries such as retail, agriculture, transportation, etc., and we do not want to assume that their experiences are the same.
It is estimated that somewhere between 12% and 17% of software engineers at some of the largest software Tech companies in the US are women (We Ran the Numbers, and There Really Is a Pipeline Problem in Eng Hiring., n.d.).

Aline Lerner from Interviewing.io noted that getting an accurate count of women software engineers is nearly impossible as most Tech companies do not report their numbers. She notes that in the above the number of women in roles such as product/program management and UX design were excluded, rightfully, as these roles do not require any programming. However, at least another 10% needs to be subtracted as more than 10% of women in technical roles are quality assurance engineers—while QA is a more technical role, it does not require programming for product development and is not on the same career ladder as software engineers. To remain conservative with the estimates, women in other less common technical roles will not be removed from this group—these roles might include sales engineering, data science, customer support engineering, etc.

Based on the estimations made above, we assume that 10 - 16% of all software engineers in the
US are women. This range is backed up by data provided by the Accenture *Resetting the Tech Culture* report, as they estimate that "women hold just 16% of engineering roles" (*Women in Tech | Accenture and Girls Who Code*, n.d.). It is important to note that much of the data utilized to generate these percentages comes from large companies that choose to report this data. This range may be an upper bound on the industry demographics, relative to smaller players who do not publish their data.

Given there are so few women in software engineering roles in Tech to begin with, any attrition of women to other roles in Tech, or roles outside of Tech, is a problem. However, before diving into the experience of current and previous software engineers in Tech, it is important to understand whether the low starting point is not necessarily a pipeline problem. Knowing that women who are trained in computer science choose not to be software engineers could provide evidence that there is something wrong with the experience of being a female software engineer in Tech companies.

### 2.2 The Funnel Into Tech

*It is not just a pipeline problem.*

The lack of women studying computer science at the university level is a problem that is frequently studied, when it comes to improving representation of women in STEM roles; this problem has come to be known as the "leaky pipeline" (Camp, 2002). Varma and Hahn (2007) simplify the leaky pipeline metaphor in their paper, "Gender and the pipeline metaphor in computing":

"The science, technology, engineering, and math (STEM) career track from elementary school to initial employment has been depicted as a pipeline. It is generally believed that if a sufficient number of women are encouraged to pursue sciences and mathematics in their elementary and high-school years, exposed to technology early on, and persuaded to enter science and engineering programmes in college/university, the gender disparities now present in STEM would disappear. However, women’s percentages in STEM decrease as they progress through the
pipeline. Men, for the most part, travel smoothly from the beginning to the end of the pipeline and thus dominate STEM. The pipeline is said to be leaky – there is steady attrition of females at every level of STEM, from elementary school into the workplace, in most industrialized countries around the world."

Big Tech companies frequently assert the leaky pipeline/insufficient numbers of women with computer science university degrees for the low numbers of female software engineers in their companies (*Facebook Blames Lack of Available Talent for Diversity Problem* - *WSJ*, n.d.). However, this theory is problematic as the rate of women studying computer science is going up. Earlier in the pipeline this is exhibited--in 2020, 31% of AP Computer Science test takers were women, and this rate has been going up year over year (*AP Computer Science Exam Attendance U.S. 2014-2021*, n.d.). Illustrated below, from 2014 to 2022 the rate of women taking the AP Computer Science exam has grown by over 50%.

![Figure 2.2: AP Computer Science Test Takers by](AP Computer Science Exam Attendance U.S. 2014-2021, n.d.)

The numbers have been increasing at the university level as well. Currently, about 20% of
computer science bachelor degrees go to women (Women, Minorities, and Persons with Disabilities in Science and Engineering: 2021 | NSF - National Science Foundation, n.d.). This has increased over the last ten years, from roughly 17.6% in 2010 to over 20% in 2019.

### Table 325.35

<table>
<thead>
<tr>
<th>Year</th>
<th>Bachelor's degrees</th>
<th>Females as a percent of total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Annual percent change</td>
</tr>
<tr>
<td>2010-11</td>
<td>43,066</td>
<td>8.8</td>
</tr>
<tr>
<td>2011-12</td>
<td>47,406</td>
<td>10.1</td>
</tr>
<tr>
<td>2012-13</td>
<td>50,961</td>
<td>7.5</td>
</tr>
<tr>
<td>2013-14</td>
<td>55,271</td>
<td>8.5</td>
</tr>
<tr>
<td>2014-15</td>
<td>59,586</td>
<td>7.8</td>
</tr>
<tr>
<td>2015-16</td>
<td>64,402</td>
<td>8.1</td>
</tr>
<tr>
<td>2016-17</td>
<td>71,416</td>
<td>10.9</td>
</tr>
<tr>
<td>2017-18</td>
<td>79,597</td>
<td>11.5</td>
</tr>
<tr>
<td>2018-19</td>
<td>88,633</td>
<td>11.4</td>
</tr>
</tbody>
</table>

**Table 325.35.** Degrees in computer and information sciences conferred by postsecondary institutions, by level of degree and sex of student: 1964-65 through 2018-19

Table 2.1: Computer science bachelor’s degrees by gender (Digest of Education Statistics, 2020, n.d.)

Along with this growth of female computer science graduates, software engineers also come from other education pathways besides university computer science programs. While most software engineers do have college degrees (Stack Overflow Developer Survey, 2022), less than half of people working in "computer and information sciences" actually have a degree in computer sciences (National Survey of College Graduates (NSCG) 2021 | NSF - National Science Foundation, n.d.).
In the table above the box annotated "2" is the number of people with degrees in CS that work in CS occupations vs. the box annotated "1" which is the number of people from any degree working in CS occupations. The data above suggests that roughly 45% of people working in CS occupations have a degree in CS. The remaining 55% of software engineers who do not have a degree in computer science are either self-taught programmers or have gone through a coding bootcamp/other certificate program (Stack Overflow Developer Survey 2022, n.d.).

Based on the total number of total CS graduates (3.1 million) and the fact that just over 20% of CS degrees go to women we can estimate that roughly 650,000 of CS degrees were awarded to women in 2021. However, in the same year it was reported that over a million women work in the same field (National Survey of College Graduates (NSCG) 2021 | NSF - National Science Foundation, n.d.). These statistics illustrate that the low rate of female software engineers is not a pipeline-only problem. Given that we have estimated that around 15% of software engineers in Tech are women, women are receiving 20% of computer science degrees, and over half of the people that work in computer sciences have a degree in it means that the education pipeline cannot be blamed for the
low rates of women in software engineering roles in Tech.

Disproving the hypothesis that the university pipeline "problem" as a key reason for low rates of women in software engineering roles in Tech suggests the need to examine other hypotheses. While there are other issues with the "funnel in" beyond in the university pipeline, such as women with computer science degrees choosing other careers right out of university, we will focus on what happens to the women who start careers in software engineering in Tech, but then choose to leave.

2.3 The Funnel Out of Tech

The 2020 Accenture report, *Resetting Tech Culture*, indicates that "women leave tech roles at a 45% higher rate than men" and "50% of women who take a tech role drop it by the age of 35, compared to approximately 20% in other types of jobs"(*Women in Tech | Accenture and Girls Who Code*, n.d.). In this context, "tech roles" does not mean just software engineering, but all roles in Tech that might require some background technical knowledge. In the appendix, Accenture explains "tech roles" include solutions architect, data science, and other roles that require some technical knowledge.

Another study found that women in STEM roles (respondents were mostly engineers) were 807% more likely to leave their jobs than women working in other fields(Glass et al., 2013). The same study was able to control for factors that are usually blamed for the attrition. For example, for female professionals the study could not determine major differences in job characteristics or family factors in STEM roles from those of other professions. The following findings illustrate a baseline understanding of the general experiences of women in Tech and engineering and what might be responsible for attrition.

2.4 The Current Experience of Women in Tech

Over the last ten years, research on the experiences of women working in Tech companies has greatly expanded. Understanding the general issues that women experience in Tech illuminates some of the reasons why female software engineers leave their roles. Based on the information
that is currently available, several themes were identified for why women leave Tech. These themes can be categorized as such: the lack of social connections, bro culture, gender biases, a hostile work environment, lack of female role models, and work and family balance issues.

### 2.4.1 Social Connections

**Being the "Only"
**

In the figure below from Women in the Workplace McKinsey study in 2022, the percentages of men and women in technical departments versus all other roles at Tech companies is shown. In 2018, while 34% of Tech companies were men and 18% were women in technical departments, by 2022 the percentage increased for men and decreased for women. As this trend continued, the likelihood of women being "onlys" on their teams has increased.

![WOMEN’S RELATIVE REPRESENTATION IN TECHNICAL ROLES DECLINED BETWEEN 2018 AND 2022](Figure 0.3: Women in Technical Roles(Women in the Workplace 2022, n.d.)

Being an "only" woman on a team can have a significant negative impact on a woman’s professional experiences(Women in the Workplace 2021, n.d.). Thirty two percent of women in technical roles are "often the only woman in the room at work"(Women in the Workplace 2022, n.d.). While this problem exists for all roles in Tech, the problem is more significant for female software engineers as the rate of female software engineers is lower than the rate of women in
Tech overall. Given that 10-15% of software engineers are women, and development teams are generally composed of anywhere from four to six engineers, the odds are that there will be just one female engineer on any team.

Women who are "onlys" are more likely to have issues being included in their workplace. This is illustrated below; women who are "onlys" experience various microaggressions, including their abilities being challenged as well as generally experiencing disrespectful behavior from others.

Comparing the experiences of women and men who are "onlys" the group that experienced the most microaggressions were the women "onlys". The experience of women non "onlys" is significantly better, but it is interesting to note that they still experience more microaggressions than even male "onlys". In the study, Mckinsey (2018) concludes that the experience of being a woman "only" contributes to them being "1.5 times more likely to think about leaving their job". One area of note is that women who reported being "onlys" on teams in this survey might be more
likely to work at companies with fewer women in general. Fewer women may choose to work at
certain companies because of their reputations of bro-culture, bad work-life balance, etc. So, it
might be poor company culture that has this effect—it might not necessarily be because of being
an "only." This is an area that will be followed up upon in the later data analysis.

In Resetting the Tech Culture Accenture survey, it was noted that women "who leave tech roles in
the workforce, or who are likely to leave in the near future, identify a non-inclusive company
culture as the major driver." Given that women are becoming less represented in technical
departments women are more likely now to be "onlys" in their work environments. With women
"onlys" being more likely to experience issues with microaggressions, we can deduce that being
an only is contributing to the attrition of women in technical departments. This leads back to the
original issue—fewer women working in technical departments.

The feedback loop represented by the data found above might look something like this:

![Figure 0.5: Feedback loop for Being an Only](image)

Madeleine Golison
2.4.2 Women in VC

There is a feedback loop in Silicon Valley that helps perpetuate the lack of senior women in Tech companies, which fuels the lack of female mentors and representation. The initial problem begins with the lack of venture dollars in Silicon Valley going to female-founded startups. Men run 92% of venture funded companies in the United States (A Data-Driven Look at Diversity in Venture Capital and Startups, n.d.). Why does this happen?

If one examines the typical path of how people end up becoming venture capitalists, it becomes clear why the funding goes to male-founded startups. In 2014 only 6% of partners at venture capital firms were women (Balachandra & Davis, n.d.). It has been hypothesized that because there is low representation of women in VCs, fewer VC dollars go to female-founded startups (Empowering Female Founders, 2023). In Figure X below, in every start-up funding round, female partners lead more often for female founding teams than male partners. (Women VCs Invest in Up to 2x More Female Founders | Journal | Kauffman Fellows, n.d.)

![Figure 0.6: Gender of VC Partners Leading Funding Rounds](image-url)
When female founded startups receive funding, it is shown that if the round was led by a female partner that could potentially have negative effects in the future. In a study in 2022, Snellman and Solal found that "firms with female founders who received funding from female rather than male VCs are two times less likely to raise additional financing"(Snellman & Solal, 2023).

They hypothesize this might be because "future investors may discount a female entrepreneur’s competence as the key factor in an early stage investment decision, when the investment comes from a female investor."(Snellman & Solal, 2023) So when women do get funding, if the funding came from a female investor, they are less likely to succeed in future rounds of funding. This is supported by the fact that from 2013 to 2020 only 20 out of the 2000 companies that IPOed were founded by women.(Female Founder IPO Data: Only 20 Women Have Taken US Startups Public, n.d.)

For these VC funded startups, those with at least one female founder will hire six times more women than those who do not.(Startups With At Least 1 Female Founder Hire 2.5x More Women. | Journal | Kauffman Fellows, n.d.) Considering the statistic mentioned earlier, that 92% of venture
funded companies in the US are run by men, fewer women are hired by these startups and therefore are less likely to sit in leadership positions at these startups. With fewer women having been in higher ranks of technical leadership, they are less likely to be in the pipeline for becoming venture capitalists. In a literature review from the Harvard Kennedy School in 2019, Siri Chilazi explains that there are several hypotheses for why such few women end up in venture capital, and one such is that women have less experience in Tech leadership and entrepreneurship, which are highly desirable backgrounds for VC. (Advancing Gender Equality in Venture Capital, 2023)

![Figure 0.8: Women in VC Feedback Loop](image)

Being in senior leadership in Tech has a multitude of effects beyond being a prerequisite for venture capital. A very important role of senior leadership in Tech is setting the company culture. In 2002 a study of almost 2000 members of the Australian institute of management showed "strong and positive relationships among leadership and organizational culture" (Sarros et al., 2002) and "the more transformational leadership used, the greater the leadership outcomes and the more performance oriented, socially responsible and supportive the organisational culture" (Sarros et al., 2002). Given that most VC funded Tech companies are (young, the average age being 31) male founded and led (Frick, 2014), bro culture becomes a likely possibility in these startups.
2.4.3 Poor Company Culture

Bro Culture

What is "bro culture" and what is its role in Tech companies? Bro culture is, by definition, a non-inclusive one. In Brotopia, Emily Chang describes bro culture as a company culture that was invented and championed by male-dominated startups in the late 1990s and included "insane work hours, drinking...prizing youthful brilliance over experience." Bro culture was also markedly different from nerd culture in that it was not only necessary to have good technical skills, but also be daring and cool in order to work in these companies (Chang, 2018). Chang describes this change as occurring because of the influx of money in Tech, which changed the perception of the industry. In the 2020 survey and analysis, "Resetting the Tech Culture" (Accenture-A4-GWC-Report-Final1.Pdf, n.d.) Accenture gives a thorough overview of how this "bro culture" affects women in Tech.

Accenture Resetting the Tech Culture

The Accenture study was organized in a way that sought responses from both men and women in technical roles. The study included 488 men and 1502 women, with 533 women who had left (or were planning on leaving technical roles within the next two years. Their grouping of technical roles were those where employees utilized "technical skills such as coding, math or engineering on a daily basis—or who have progressed up through the organization by using these skills". These employees were surveyed on a variety of factors that are known to "influence retention and advancement of young women in tech''. Their method for defining inclusive vs non-inclusive company cultures is outlined below:

"How we define and measure workplace culture:

STEP 1: Using a linear regression model, we analyzed the responses to our survey to identify the cultural factors that positively and significantly influence the retention/advancement of women in core programs/roles. These factors were grouped into four buckets.

STEP 2: We built a model to quantify the impact of the cultural factors on the
retention and advancement of women.

STEP 3: We scored every respondent on the incidence and strength of these factors in their workplace.

STEP 4: We segmented respondents to find two core analysis groups: the top 20% of respondents by culture score (“more-inclusive”) and the bottom 20% (“less-inclusive”).” (Women in Tech | Accenture and Girls Who Code, n.d.)

By their method, we can define more inclusive cultures as those in the top 20% of respondents and less inclusive are those in the bottom 20%.

In the study results, the highest-level effect of bro culture on women in Tech is by influencing them to leave Tech altogether. Accenture found that 37% of women cited poor company culture as the reason why they left tech, and that makes it the number one reason women leave Tech overall(Women in Tech | Accenture and Girls Who Code, n.d.). When women were asked how companies could improve the rates of women in technical roles, over 50% noted that access to role models as well as better workplace culture were the most important(Women in Tech | Accenture and Girls Who Code, n.d.). Better workplace culture means a culture that is inclusive.

The same study found that workplaces in Tech that were reported to be more inclusive had much better experiences and outcomes for women employees. In figure x below, 85% of women reported loving their job in inclusive workplaces, whereas just 28% did in less-inclusive workplaces.
Part of the culture problem stems from the fact that many companies do not realize that they have a culture problem in the first place. In the survey, Accenture found that C-suite level executives at Tech companies are generally unaware of the difficulties that women face in Tech. According to the survey, over 75% of HR leadership believes that their company culture "enables women to be successful in technology roles" and 45% said that it is "easy for women to thrive in tech". The problem with this dissonance is then only 38% of HR leaders surveyed believe that building a better company culture is a good way to help retain women in technical roles (Women in Tech | Accenture and Girls Who Code, n.d.).

Interestingly, the report highlights that a "full 91% of the SHROs in our survey told us that attracting women with Tech experience/education is critical for their company’s success"; however, most do not believe that the company culture makes a difference. If the issue of poor company culture cannot be identified, then it certainly will not be addressed. If companies solely focus only on recruiting more women, but not on retaining these women, then the attrition problem...
will continue to exist. Filling more roles with women does not guarantee lower dropout rates.

If less inclusive companies had cultures more aligned with those in the 20% most inclusive companies, Accenture estimated that the number of women in Tech could nearly double by 2030.

![Doubling the number of women in tech by 2030](image)

*Figure 0.10: Doubling Women in Tech by 2030 (Women in Tech | Accenture and Girls Who Code, n.d.)*

### 2.4.4 Hostile Work Environment

A "hostile work environment" can be a catch all for anything that goes beyond poor company culture. Company culture is a macro-level experience that is set from the highest levels of a company. A hostile work environment could be influenced by the company culture, but even companies with good culture can have a hostile work environment at the individual level.

A hostile work environment encompasses anything that makes the workplace an unpleasant place to be, which can include "ill treatment or harassment at work, bad psychological environment..."
represented by work-related stress, administration's autocratic treatment, conflicts, and employees' exposure to threats including all kinds of harassment"(Abbas et al., 2017). A hostile work environment is problematic for more than just the obvious reasons of making women unhappy in their roles. It has been shown that a hostile work environment can create higher rates of poor worker performance and higher employee attrition(Abbas et al., 2017). So beyond contributing to the attrition of women in Tech, a hostile work environment also directly hurts a company's performance.

In *Brotopia*, Emily Chang interviewed hundreds(McGrane, 2018) of female software engineers about hostile experiences they have had while working in Tech. The motivation for learning about more female software engineers' experiences came from Chang's knowledge of the (then) viral, first-hand experience of Susan Fowler at Uber and her hope to highlight others' stories.

**Susan Fowler at Uber**

Problems with Uber's work environment, for women, began to surface in February of 2017 when Susan Fowler, a mid-level site reliability engineer, posted on her blog about her experiences at the company, after she had left(*Reflecting on One Very, Very Strange Year at Uber*, 2017). Fowler described the blatant sexual harassment that came from her manager and the subsequent lack of action from the company when she tried to surface the issues to HR(*Reflecting on One Very, Very Strange Year at Uber*, 2017).

Broad dissemination of the post brought to light many problems at Uber, and how the problems at the company were not just experienced by Susan Fowler. In the following months, other women spoke about their experiences at Uber, highlighting the notion that the company did not take action when employees were reported for behaving inappropriately. Uber was accused of valuing employees that were good at their jobs over all else and did not care about making the work environment inclusive(Isaac, 2017). This is problematic because of what was mentioned earlier: inclusive companies perform better. Uber followed the public allegations with an internal investigation that culminated in the firing of "20 employees over harassment, discrimination and inappropriate behavior"; some of those fired included senior executives(Isaac, 2017).
While the example of Uber is just one anecdote of the hostile experiences of women in Tech, it does illustrate the gaslighting\(^1\) experienced by women when trying to uncover bad behavior of colleagues in the workplace. Emily Chang, through interviews with female software engineers, discovers most women have had experiences that could be defined as hostile. These experiences include one woman being choked by a male colleague at a post-work social event, another woman being told things like “the only thing you’re good for is being taken out to the back parking lot and being raped,” and most women saying they put up with “sexual advances and people hitting on you 24/7”.

### 2.4.5 Gender Biases

While gender bias exists in many occupations, it is helpful to understand how it influences the experiences of women in Tech, and more specifically, software engineering. While much of the research on women in Tech is more general, there is some research about women's specific experiences in software engineering when it comes to gender bias and two such examples are below.

In 2017, (Terrell et al., 2017) in the computer science departments at California Polytechnic State University—San Luis Obispo and North Carolina State University built a study to analyze the open source software contributions of women vs. men on Github, which is the largest host of open source software. In the study, the researchers found that overall women’s code changes to open source software were accepted more often than men's (Terrell et al., 2017).

---

\(^1\) Definition: psychological manipulation of a person usually over an extended period of time that causes the victim to question the validity of their own thoughts, perception of reality, or memories and typically leads to confusion, loss of confidence and self-esteem, uncertainty of one's emotional or mental stability, and a dependency on the perpetrator (Miriam Webster).
While this seems positive, there are some caveats that illustrate bias in the open-source process. When controlling for whether a user’s Github account can be associated with a gender, the results change.

Figure 0.11: Pull Request Acceptance Rate (Terrell et al., 2017)

Figure 0.12: Gender Neutral vs. Gendered PRs (Terrell et al., 2017)
When the gender of the user is known when creating a request for a code change, women's requests are not accepted as often as men's. However, when the gender is unknown, women's requests for changes are accepted far more often (Terrell et al., 2017).

Similar experiences have been reported by female software engineers at Facebook (now Meta). According to a 2017 article published by the Wall Street Journal, female software engineers at the company claimed that their requests for code changes were rejected 35% more often than those of their male colleagues (Facebook’s Female Engineers Claim Gender Bias - WSJ, n.d.). These engineers also found that they waited, on average, 3.9% longer to have their code changes accepted, and got 8.2% more questions on the changes they attempted to make (Facebook’s Female Engineers Claim Gender Bias - WSJ, n.d.).

Facebook ended up doing their own investigation into this claim. In Facebook's analysis, they asserted that the gender gap was smaller in practice by saying that the higher rejection rates were more related to the engineer's level, not gender (Facebook’s Female Engineers Claim Gender Bias - WSJ, n.d.). However, this implies that female software engineers at Facebook do not advance to more senior levels. Either way, bias is involved, whether it is related to what is deemed quality code, or who is allowed to advance at the company.

2.4.6 Work and Family

Claudia Golden, in her book Career and Family, explores the relationship between people’s careers and how they manage having a family. She posits that the "fundamental problem for women trying to attain the balance of a successful career and a joyful family are time conflicts" (Goldin, 2021). Given that most roles in Tech require working far more than 40 hour weeks it is no surprise that these roles conflict with a woman's choice to have a family.

Roles in Tech, like roles in many corporations, are defined as "greedy roles" by Goldin. Goldin explains that greedy roles exist when an "individual puts in overtime, weekend time, evening time" and when they earn much more, "even on an hourly basis" than the person who is not (Goldin, 2021). These greedy roles directly conflict with family life in that the person needs to be available.
and flexible around the clock, which can conflict with family needs. Goldin explains that greedy roles "heavily penalize employees who have even brief career disruptions and those who do not work exceptionally long and grueling hours" (Goldin, 2021). While this is true for most of Corporate America, software engineers anecdotally work many more hours than the traditional 40-60+ hour weeks are not unusual and the hours can be longer if a software engineer is on call (Goldin, 2021). Thus, the issue will be even more pronounced in the software engineering careers.

Goldin shows that women with advanced degrees frequently leave full-time work for part-time work after having children. This phenomenon is most pronounced for those with a BA or MA degree and least for those with an MD. Given that roughly 90% of software engineers work full-time (and some) I hypothesize that female software engineers are leaving the role due to work and family balance issues.
Figure 0.13: Career and Family by Advanced Degree, Harvard and Beyond Fifteen Years after College (Goldin, 2021)

Besides the long working hours, software engineers are expected to improve on their skills outside of work and most do, as illustrated in the Stack Overflow 2022 Developer Survey. It is quite common to be asked in the interview process for software engineering roles what your current side projects are (Should a Lack of Side Projects Raise Flags in an Interview?, 2019). These expectations assume developers have time outside of work for these projects and they illustrate bias against candidates who spend time with their family instead of doing these.
Only 12.41% of software engineers surveyed do not code outside of work (Stack Overflow Developer Survey 2022, n.d.).

For women who have the financial choice to not work in greedy roles, because they are comfortably living off prior, spousal, or family support, many choose not to. In the five years after a woman has a child, if the woman has a spouse that is in the top-earning group, spouses "that earned above the median salary for MBA men," they are 32% less likely to be still working (Goldin, 2021). This finding is consistent with the notion that women do not progress to more senior levels in software engineering. Given that software engineering is a greedy role, women in greedy roles leave full time work more often after having children, and female software engineers do not advance in Tech as often as their male counterparts, we can hypothesize that the greediness of the software engineering career pushes women out of the role.

In a 2017 study from the University of Wisconsin's school of psychology and Texas State University's business school, about why women leave engineering fields, survey results showed...
the largest problem for women was the "inflexible and demanding work environment that made it difficult to balance work and family roles" (Fouad et al., 2017). Study participants explained their intense workload, travel expectations and failures in trying to make their current role more flexible made it challenging to raise a family and continue to work (Fouad et al., 2017). The penchant to work extreme hours is commonplace in Tech. The belief that programmers must work 60-80 hour weeks to make a company successful is still one that is deep rooted in Silicon Valley culture (Chang, 2018).

While Tech companies are not blatantly anti-family, the benefits often skew towards attracting young, single people. For example, Google, Facebook, Apple and other large Tech companies provide free dinner for employees, which incentivizes employees to stay in the office later (Chang, 2018). Most of these companies also provide funding for egg-freezing, which incentivizes employees to delay starting a family; however, of all of the benefits that these companies offer, none of these companies provide onsite or subsidized daycare (Chang, 2018). This is important to note because the brunt of household duties still fall on women, as shown below (Cerrato & Cifre, 2018).
Table 0.1: Involvement in Household Chores by Gender (Cerrato & Cifre, 2018)

The combination of a lack of family benefits, the prevalence of greedy jobs and the hours of work that they require, and women still doing the majority of the household duties make it extremely hard for women to have enough hours in the day to remain working in Tech.

2.4.7 The Influence of COVID and Remote Work

Post-COVID, the rise of remote work made balancing home and work duties seemingly more manageable. Delving into the need for women to have flexibility in their roles will show how work/family balance issues can be somewhat assuaged. In the 2023 Women in the Workplace study by McKinsey they found that "one in 5 women say flexibility has helped them stay at their organization or avoid reducing their hours. A large number of women who work hybrid or remotely point to feeling less fatigued and burned out as a primary benefit" (Women in the Workplace 2022, n.d.).
While some companies have kept the possibility of remote working options, others have completely flipped to a strict Return to Office Guideline. Amazon, Google, Apple and many others have decided that their employees will need to be in office at least three days per week. In the 2023 IWG Women Hybrid Workers Sentiment Survey they found that "The majority of female workers say that, if their current employment did not have a hybrid work option, they would look for another job (72%)." I hypothesize that remote working arrangements for female software engineers increase their likelihood of staying in the role as women have reported the benefits of having more flexible working arrangements.

2.5 The Research Gap: What We Don't Know

While there is significant research to illustrate high-level reasons for women leaving Tech generally, the question that is not fully answered is why attrition of women in software engineering roles is there specifically. Additionally, the literature does not explain why women might leave software engineering roles for other roles in Tech. To better understand why women leave software engineering roles, examining the specific experiences of women in these roles will give a better picture of how relevant the general research of "women in Tech" is to them. Finally, learning why women leave software engineering roles for other roles in Tech will provide information on the specific challenges of software engineers that are not experienced by all women in Tech. It will be helpful to not only know what challenges push women into different occupations, but to also learn what are the most significant factors in why women make a change from the software engineering role specifically.
Chapter 3  Research Methods

3.1 Overview

The following chapter highlights the methods utilized for investigating gaps in knowledge of why women leave software engineering roles. The approach taken was predominantly data driven. A survey was developed and distributed among various channels and several forms of analysis were employed. Quantitative analytical methods were used to compare all potential survey variables together to understand the importance of each variable in their relationships to women remaining in software engineering (or not). For the open-ended survey questions two methods of analysis were used. The first was quantitative and consisted of using natural language processing methods to highlight key themes in the text. The second method used was a qualitative review of the open-ended survey questions using thematic analysis, with the hope of corroborating the results from the quantitative analysis.

3.2 Unanswered Questions

The initial motivation for this research initially came from what was not included in the Accenture Resetting Tech Culture survey. The survey done by Accenture has three major limitations:

3.2.1 Survey Demographics

In the Accenture survey they consider both women currently in technical roles and women that have left technical roles, which helps differentiate the experiences of those that are assumed to be choosing to stay and those that choose to leave technical roles. However, Accenture's definition of technical roles is as follows:

“Women in core technical roles: use technical skills such as coding, math or engineering on a daily basis—or who have progressed up through the organization by using these skills. Example roles include Software Developer, Data Scientist, and Technical Solution Architect.”(Women in Tech | Accenture and Girls Who Code, n.d.)
While this definition is good for understanding the experiences of generally technical women in Tech companies, it is not specific enough to understand the unique experiences of women in software engineering roles. For example, team structures vary significantly between technical roles and thus the experiences in these roles will be different. For example, in a role such as solutions architect the employee is working closely with external customers versus mostly with internal development teams. One hypothesis is that working on a software engineering team comes with a unique cultural experience as the software engineers are mostly only working with other software engineers.

3.2.2 The Focus of the Survey

As mentioned, Accenture, as well as the research evaluated in the literature review, do not focus on learning about women in software engineering-specific roles. Beyond missing information on cultural experiences of software engineers specifically, there has been little to no research done about other aspects of the role that might explain attrition. Some other aspects not yet considered are whether a woman has dependents, work setting (remote, vs hybrid, vs fully in office), years of experience, company size, etc.

3.2.3 A Missing Group: Women Leaving Software Engineering, but Staying in Tech

While the Accenture survey investigated why women stay in or leave technical roles, they do not specify where those that have left technical roles go next. One major hypothesis is that most women in software engineering want to go to other roles in Tech (and not leave Tech altogether). If this hypothesis is true, then the high-level company culture would not necessarily be the reason for leaving the role. If this group exists—women prefer to stay in Tech than leave altogether—then more needs to be understood on why the software engineering role is the reason for their leaving.

These three deficiencies were illuminated broadly in the existing research and influenced the
design for the MIT Women in Software Engineering survey conducted for this research. The largest gap in the existing research is that there is no information about women that leave software engineering, but stay in Tech. Are women switching into other roles in Tech from software engineering? And if so, why?

### 3.3 Main Hypotheses

The following hypotheses were outlined, and they influenced the design of survey questions. The primary hypotheses guided the overall theme of the survey while the secondary hypotheses support the second primary hypothesis by investigating what the most important factors are for why women leave software engineering roles.

#### 3.3.1 Primary Hypotheses

- The majority of female software engineers that have left or hope to leave software engineering roles go to or plan to go to non-software engineering roles in Tech, but not leave Tech altogether
- Culture is not the most important reason for why women leave software engineering roles

#### 3.3.2 Secondary Hypotheses

- Disliking working with your manager relates to leaving or wanting to leave a SWE role
- Female software engineers that work more hours are more likely to have left (or plan to leave) software engineering roles
- Female software engineers that report less flexibility in their work environment are more likely to have left (or plan to leave) software engineering roles
- Women leave software engineering roles not because they dislike coding
- Women that stay in software engineering (or plan to) are mostly remote
- Women with dependents do not stay in software engineering roles as much as those who do not
• Women with no other women on their direct teams are more likely to leave software engineering roles

The goal of these hypotheses is to cover a variety of areas where reasons for attrition can be different for female software engineers than women in Tech in general. The highest level hypothesis is that women in software engineering leave the role for reasons other than the reasons outlined for women in Tech generally, which were outlined in the literature review.

3.4 Survey Design

For the purpose of the survey, "software engineers" include not only those on the software engineering career track (associate, mid-level, senior, staff, etc), but also those who directly manage software engineers, as the mid-level engineering managers generally have at least one foot in the technical work that their reports are doing.

3.4.1 The Groups that were Analyzed

Based on outstanding hypotheses, several groups' experiences needed to be analyzed.

• Female software engineers that want to stay in software engineering for at least the next three years. While intention does not guarantee that these women will stay as software engineers indefinitely, since the average time to promote a software engineer is 2 - 3 years (Tang, 2022), these software engineers intend on staying in the role long enough to be on a promotion path indicating that they see a future in the role.

• Female software engineers that want to leave software engineering in the next three years and previous female software engineers. Grouping these together allows us to identify reasons for wanting to no longer be in the role.
One important caveat to mention is that there is no way to know whether these statements will hold true in the future. For example, a woman wanting to leave software engineering may not actually find another suitable role, or they may get a better manager/team, etc. However, the intention still means there are reasons "right now" for wanting to make a change—and knowing these reasons can still clarify how to retain others.

### 3.4.2 The identification of software engineering-specific work factors

Before deciding on what factors could influence the experience of female software engineers, I determined which factors were already evaluated in the Accenture *Resetting Tech Culture* survey. Most of the questions were related to company leadership, organization DEI efforts, employee feelings of psychological safety, and employee flexibility. Some categories, such as company leadership and DEI efforts, were less of a priority for the MIT Women in Software Engineering
survey as the experience of company leadership is likely to be similar across technical roles and different teams within the same company. However, categories related to feelings of psychological safety and employee flexibility were important to include in the survey because these experiences are more likely to be determined at the team level\(^2\). The list below (a snippet from the Accenture methodology section) influenced a subset of questions in this survey; however, not all were included for brevity.

**Reasons for leaving tech roles**

We asked women who had left tech roles - or were likely to leave such roles within two years - why. Respondents were presented with 13 options (including “Other” and “Don’t know”). We then grouped these options into four buckets:

**JOB ROLE FACTORS:**
- I don’t / didn’t like my line manager/supervisor
- The work is / was boring/repetitive
- I wasn’t able to advance at the pace I wanted

**CULTURE FACTORS:**
- I wanted a better work/life balance
- Hard to balance work and family commitments
- I couldn’t thrive because of the company culture
- Because of sexual harassment or discrimination
- Because of racial harassment or discrimination

**DIVERSITY FACTORS:**
- Lack of senior leaders / role models from my racial/ethnic background
- Lack of senior leaders / role models of my gender
- Lack of colleagues from my ethnic / racial background
- Lack of colleagues of my gender

**PULL FACTORS:**
- Another role is/was more attractive

*Figure 3.2: Reasons for leaving tech roles (Women in Tech | Accenture and Girls Who Code, n.d.)*

\(^2\) https://www.glassdoor.com/Reviews/Employee-Review-Amazon-RVW78822254.htm. Amazon, and many other large Tech companies, are known for having drastically different experiences for different organizations and teams
3.4.3 Question Selection

Before selecting questions, potential dependent variables needed to be noted. The target dependent variable selected (and illustrated above) was a Boolean response where a respondent was either currently a software engineer and intends to stay in the role versus currently being a software engineer and hoping to leave (or a software engineer that had already left the role).

The questions that were selected for the survey fell into several categories, which are outlined below. The questions in the survey were designed to be less upfront about the known sentiment of the respondent, but were more aimed at capturing all different factors of a respondent's environment. The goal of framing the questions this way was to see if there were relationships between any given factors and the attrition of female software engineers. The complete survey can be reviewed in the appendix.

3.4.4 Survey Sections

- Demographics of survey respondents
- Career Information - what was the current or previous career of the respondent
- Company Information - what were different properties of the company where the respondent worked as a software engineer
- Working Experiences - the questions that directly asked for the feelings of the respondent

Demographics

Simple questions were included to verify the identities and segment the respondents. Some important questions asked were whether the respondent identifies as a woman, what is their age, what is their education level/degree, when did they learn to program, do they have dependents, etc.

Career and Company Information

The first two questions in this section helped determine the potential dependent variables.

Question 1: What is your current career?
● Software Engineering: a role that spends at least 50% of the time coding, or managing those who do
● Other career in software: any role where you are working with software, but coding 0 to 50% of the time
● Other career not in software

Question 2: In three years, what career do you see yourself in?

● Software Engineering: a role that spends at least 50% of the time coding, or managing those who do
● Other career in software: any role where you are working with software, but coding 0 to 50% of the time
● Other career not in software

The following questions were related to the length of tenure in software engineering roles and what the characteristics of the company and manager were when they were either working as software engineers or where they were working when they used to be software engineers (for the ex-software engineer group).

*Working Experiences*

The Working Experiences section of the survey asked more directed questions about whether the respondent liked programming, what their working hours were, did they enjoy working with their manager, and their thoughts on pair programming. The set of questions in the Working Experiences section explicitly ask the feelings of the respondents while working as software engineers, whether current or past, and are on a *never to always* sliding scale. The hope with this section of questions is to ensure that perceived experiences are included when evaluating the hypothesis that culture is the most important variable in predicting women leaving software engineering roles.

Example from this section:
Open Ended Questions

A small subset of open-ended questions were included in the survey. These were included to expose any potential gaps in the survey questions by allowing respondents to provide their own ideas for why they believe they have left a software engineering role.

3.4.5 Circulation

After publishing the survey, I distributed it across a wide variety of channels. These included the Reddit group r/womenintech, Elpha (a women in tech online forum), LinkedIn, Rands Leadership Slack group, and through my own network of women at Apple, Google, Adobe, Snapchat and Amazon. The survey was in circulation during the summer and fall of 2023.

3.5 Data Cleaning

The survey produced roughly 340 responses; however, after a preliminary review of the data it became evident that many rows were missing responses. The following process was outlined to get the correct data for analyzing from the raw data.

- Removing rows where the dependent variable could not be identified, which was if either of the two key questions from the Career and Company Information section of the survey were not answered.
- Removing rows where a respondent said they did not identify as a woman
- Removing rows where a respondent said they had never been in a software engineering role or had reported never having coded before
- Merged the columns for duplicate questions that were created to be either worded in the present tense or past tense for purposes of the survey
- Removed all remaining rows that had any NAs for categorical and numerical columns.
- Normalized all numerical columns by converting values to 0 - 1 range

The number of usable data points were roughly 180, depending on the analysis conducted. To ensure data consistency, since the number of responses were low, the data could be visually checked. Rows were scanned to make sure a respondent did not select the same response for the entire survey and open-ended responses were read to ensure responses were either empty or legitimate.

### 3.6 Quantitative Methods

The bulk of the survey data analysis used data science and machine learning methods to investigate whether different factors (independent variables) were most related to women deciding to remain in software engineering careers. The overall strategy with trying to identify these most important variables with quantitative methods was to identify the best possible model (via various trials of different methods and tuning parameters) and then utilize this model to explain the most important variables found. The rest of the models developed, depending on their accuracy, were used to verify and support the "best" model's findings. All models were testing all secondary hypotheses, since the primary hypotheses were found in the initial demographic review of the data.

#### 3.6.1 Dependent Variable -- "Remain SWE"

The main dependent variable was named "Remain_SWE" and was a Boolean variable representing the following:

1. True: current software engineer and planned to be in three years
2. False: ex-software engineers or software engineers with plans on leaving a software engineering role within the next 3 years
3.6.2 Independent Variables

The predictive methods mentioned below--logistic regression, random forest and gradient boosting--were all utilized in two distinct ways. The first set of models were generated with numerical independent variables and the second utilized categorical variables. The main reason these two sets of models were created was that setting some of the variables to numeric may have unintended consequences.

For example, when the survey responses for how flexible the location of work is included a range from onsite to hybrid to fully remote, there may not actually be an ordered relationship across the choices (i.e. more remote equals more likely to stay in SWE vs hybrid being the best option for women to stay in SWE) and so we do not want to assume each question's responses are related. Creating the models both ways limited missing important information about ordinal and categorical variables.

Using Numerical IVs

Since most of the multiple-choice survey questions were on a Likert scale, multiple choice responses could be converted to a numerical range in order to reduce the total number of variables and reduce bias in using categorical variables only. For example, the possible responses for the question "How much do you enjoy your software engineering role?" included, 'Not at all', 'A little', 'A moderate amount', 'A lot', 'A great deal'. Since it would be ideal to preserve order information (i.e. "Not at all" is more similar to "A little" than "A great deal"), converting these responses to numbers instead of new dummy variables would reduce the number of variables, which is better for models with less data, and might be more accurate by preserving the order of the variables. Using numerical variables limited the number of IVs to 43 (vs. over 100).

Using Categorical IVs

Since converting the multiple-choice responses to a numerical scale assumes that there is an order to the responses for each question, creating models without this assumption was necessary to see whether the assumption held up. For each multiple-choice question, dummy variables had to be created in order to create models that did not assume a relationship between each question's responses. Doing this took the number of variables in each model from "n" to more than "3n" (as
most multiple-choice questions had 3+ answer options). When creating models with the categorical variables there were roughly 180 complete responses and 123 variables. Ideally, there should be 10n responses for n variables, but as seen here there were almost as many variables as responses, especially considering the data would be split into the test and train subsets. However, it was still important to generate at least one model for each method mentioned below using the categorical variables as converting these variables to numerical implies a continuous relationship for each variable's options.

For example, if age groups were assumed to be continuous then using numerical variables would illustrate a linear relationship between age and remaining in software engineering or nothing at all. However, when using categorical variables, one option for age groups is the 30-40 age group. This group might have a more significant relationship to remaining in software engineering given that this age group is when many women are having families, and if categorical variables were not used this relationship might not be observed.

Working Experiences Questions

For the Working Experiences group of questions, just one set of variables was created because the responses for these questions were already numerical. However, some of the Working Preferences questions needed to be removed from the models as they introduced multicollinearity, which can bias the coefficients and lead us to incorrect feature strengths. For example, "I enjoy programming professionally" was heavily correlated (over .75) with "How much do you enjoy your software engineering role?" so one needed to be removed.

3.6.3 Splitting the Data

The models outlined below were created with various subsets of the data to improve generalization and to ensure variable importance was not random. The traditional split of 70/30 train/test data was not used for each baseline model since the dataset is so small. The baseline for the models was an 80/20 split. Missing 30% of the data during training was less important than improving the model overall, so models with smaller test sets were created, where as little as 10% of the data went to the test set to ensure getting the best possible results.
3.6.4 Baseline Method: Logistic regression

"Logistic regression is a classification algorithm used to find the probability of event success and event failure. It is used when the dependent variable is binary (0/1, True/False, Yes/No) in nature. It supports categorizing data into discrete classes by studying the relationship from a given set of labeled data. It learns a linear relationship from the given dataset and then introduces a non-linearity in the form of the Sigmoid function." (‘Advantages and Disadvantages of Logistic Regression,’ 2020)

\[ f(x) = \frac{1}{1 + e^{-\beta x}} \]

**Figure 3.4: Sigmoid Function** (‘Advantages and Disadvantages of Logistic Regression,’ 2020)

Logistic regression is used as a baseline model for a few reasons. It generally is not as accurate as some of the other methods that predict classification; however, for each variable the model provides coefficients that are very helpful for explainability. The coefficients show the direction of the association (positive or negative) as well as the strength (the size of the coefficient), and this information can help indicate which variables are most important (at least in this model). Since I was not necessarily creating the models to predict the outcome of a woman in software engineering, I was less concerned with the accuracy of the model than the significance of each variable and the strength of the coefficient. However, accuracy was moderately important as showing there was a relationship between the dependent and independent variables was important for credibility of the coefficients. The most important variables in this model were selected by identifying the variables with large coefficients and then ensuring that each of those variables were statistically significant (p < .05), meaning that they were unlikely to be important by chance.
3.6.5 Decision Tree Methods

The second set of methods created were both based on decision trees. Decision tree learning is a supervised machine learning method where classification is predicted by a "learned" decision tree, which is a flow of decision rules based on the factors provided. These methods are useful because they tend to be more accurate than logistic regression and can still provide feature importance. Additionally, the following methods can better handle high-dimensional (and sparse) data. Using a combination of methods is important for robust recommendations and ensuring that results can be reproduced.

Random Forest

"A decision tree method of this kind combines the predictions of numerous decision trees, or forests, to arrive at a final prediction. Each decision tree in a random forest is constructed using a unique bootstrap sample of the data and a unique subset of the predictor variables known as a random subspace. As a result, the predictions of the trees are aggregated either by majority vote for classification issues or by average for regression problems."

(“Logistic Regression Vs Random Forest Classifier,” 2023)

XGBoost

XGBoost is similar to Random Forest, but it usually outperforms it, when comparing model accuracy scores. What is different between the two is that XGBoost uses gradient boosting to minimize the loss of adding new models using gradient descent for optimization. With XGBoost, the hyperparameters differ from random forest and the model can better handle missing values and categorical variables so it was included in the analysis to maximize performance. Given that XGBoost can handle different types of data well, a larger model was built with all variables to see the relative significance of the variables across their groupings.

Grid Search for Cross Validation

For both the random forest and XGBoost models, cross validation was performed with Grid Search. Grid Search is a method that creates all models for a specified subset of hyperparameters in order to find the best hyperparameters for each model type. The ideal hyperparameters varied
based on the variable subsets and data splits, but the hyperparameters searched remained constant. The Grid Search method also handles splitting the data itself.

Random Forest hyperparameters searched:

```python
param_grid = {
    'n_estimators': [25, 50, 100, 150],
    'max_features': ['sqrt', 'log2', None],
    'max_depth': [1, 2, 3, 6, 9],
    'max_leaf_nodes': [2, 3, 6, 9],
}
```

XGBoost hyperparameters search:

```python
parameters = {
    'max_depth': [2, 10, 1],
    'n_estimators': [25, 100, 40],
    'learning_rate': [0.1, 0.01, 0.05]
}
```

### 3.6.6 Model Explainability with Shapley Values

Various methods for model interpretation and feature importance were explored. However, I decided to use Shapley values as they are not affected by multicollinearity like permutation importance is.

"SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions (see papers for details and citations)." (Welcome to the SHAP Documentation — SHAP Latest Documentation, n.d.)
Additionally, this method is model agnostic so it can be applied no matter which ML method is used. Therefore, it was able to be used for all models created.

### 3.7 Methods for Text Analysis

As mentioned, at the end of the survey there were several open-ended questions. The open-ended questions were looking to find answers to this primary hypothesis:

*Culture is not the most important reason for why women leave software engineering roles.*

The close-ended survey questions assume alternative reasons for why women leave SWE roles. The open-ended text responses ensure that the assumptions made did not limit the potential variables. Additionally, the question asking why women stay was critical for making recommendations for improving attrition of female software engineers.

Two of the questions were analyzed quantitatively, with Natural Language Processing methods
and word clouds, to support any primary qualitative analysis done for these responses. The two key survey questions analyzed were:

1. If you left a role in the software engineering career path, or plan to do so in the next few years, what would be the primary reasons as to why?

and

2. What is/was your favorite thing about software engineering?

The goal of analyzing these two main questions was to capture any insights that were not discovered in the closed form survey question.

3.7.1 Natural Language Processing Methods

BERTopic

The baseline text method utilized was BERTopic, which is "a topic modeling technique that leverages transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions." (Grootendorst, 2022) Maarten Grootendorst (2022) developed the method to be well understandable and explainable for general use. Grootendorst illustrates the steps as follows:
Generating "topics" from the open-ended survey questions allowed for a first step in understanding the responses (quantitatively). The less explainable methods--ChatGPT and Qualitative Thematic Analysis, were also used to validate outcomes from the BERTopic models.

*OpenAI’s ChatGPT*

OpenAI’s documentation explains that their GPT models "have been trained to understand natural and formal language. Models like GPT-4 allow text outputs in response to their inputs...[They] can be used across a great variety of tasks including content or code generation, summarization, conversation, creative writing, and more." (*OpenAI Platform, n.d.*)

The OpenAI ChatCompletion API was utilized to extract themes from the responses of the two survey questions mentioned above. The inputs to the API call was one string, which was a concatenation of all text responses for a single survey question, and the output was a list of 10-15 themes.

```
response = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
```

Figure 3.6: BERTopic Model Layers(https://maartengr.github.io/)
As mentioned above, the results of this analysis were not considered conclusive, but more as a starting point for the qualitative thematic analysis, which was the main method for analyzing the text data.

3.7.2 Qualitative Thematic Analysis

"Braun and Clarke's thematic analysis method is an iterative process consisting of six steps: (1) becoming familiar with the data, (2) generating codes, (3) generating themes, (4) reviewing themes, (5) defining and naming themes, and (6) locating exemplars." *(Thematic Analysis - an Overview | ScienceDirect Topics, n.d.)* Interestingly, this process is not too different from the quantitative methods listed above, except for that the codes and themes are human generated.

Utilizing thematic analysis allowed for either validating or invalidating the quantitative methods results, which is important since the data being analyzed was not quantitative in nature. This means that the GPT and BERT models are not always right. Given that the number of responses was small, a thematic analysis could be easily done to either support or invalidate the first two models' results. However, if this survey was able to generate thousands of responses a thematic analysis would take quite a bit longer. Being able to use a combination of both qualitative and quantitative methods on the text responses allows for fast analysis with the ability for a real person to back-up the results.

3.8 Methods Limitations

3.8.1 Survey Distribution
One significant note about the survey distribution is that it was very hard to find ex-SWEs that are no longer in Tech. This could mean one of two things. One could be that there are truly very few women that leave Tech altogether. The other potential reason could be because the women in this group are not likely a part of the same online communities as women SWEs. Since the survey was distributed among women at large Tech companies, Tech leadership and other online women in Tech communities, it is less likely to find women that have left Tech altogether in these forums.

A lack of female respondents that have left Tech altogether is not a problem for the main hypothesis, as understanding why women leave software engineering, but choose to stay in Tech, is the group that is not represented by current research. However, if there are software engineering specific reasons for women leaving Tech altogether, this analysis cannot conclude anything about these (potential) differences.

Another potential issue with the distribution method of the survey is that it was mostly distributed to contacts at large Tech companies. While there is a good amount of representation from smaller companies, most of the respondents were still from the largest companies, so this may have affected the results.

### 3.8.2 Number of Responses

While there were roughly 300 total respondents, there were only about 180 completed surveys. Since there are between 40 and 45 total variables considered in the models, there certainly were not enough responses to provide high confidence (and statistical significance) of more than a few variables. If there were more like 10x responses then the confidence that any variable’s influence was statistically significant would be much higher.

### 3.8.3 Men were not Surveyed

While other surveys have included men in their respondents, this survey did not. In hindsight it would have been helpful to know if men reported the same experiences as women; if their reasons for leaving Tech roles were the same (or different) any recommendations could be generalized. However, because the focus of the study was to understand why women (specifically) leave these
roles, it seemed the experiences of men would not be helpful for this analysis.

One example of where it could have been good to have respondents from both men and women is for any question where each group might have reported the inverse of the other. For example, if women reported fully in office work as being their reason for leaving software engineering roles, but men reported that being fully remote as their reason for leaving software engineering roles, then recommending fully remote work to retain women might have negative effects on a company. In future research, understanding if women and men have vastly different experiences would help guide better recommendations for Tech companies.
Chapter 4 Results Demographics

4.1 Results Overview

The results from the MIT Women in Software Engineering survey are organized into three main sections. The first was the demographics section, where information about the survey respondent population was better understood. It was important to analyze the backgrounds of the survey respondents because knowing whether the data set was biased towards a certain group helped provide context for the results.

The second section were the results of the quantitative analysis of the survey. Initially, the results of the various predictive methods were analyzed for the entire set of survey respondents. Then, several follow up analyses were done with certain subsets of the data, based on the demographics. For example, more attrition of female software engineers was observed for women in their 30s, and understanding why that group was different than the entire population helped develop targeted insights. The quantitative analysis also included a preliminary analysis of the text responses with natural language processing methods.

The third analysis of the data was the qualitative analysis of the open ended survey responses. Reading these responses and understanding the themes was important as respondents included unique perspectives and experiences that could not be captured in simple survey questions.

4.2 Cleaned Data

Overall, there were 250 women that responded with their gender identity being female and their years of programming experience being at least two. This was the first method utilized for a quick filter of potential "good" responses. However, once all rows that were not complete were removed from the dataset (for all independent variables that were measured) there were roughly 180-185 completed responses. The demographics and results outlined below were based on only these complete survey responses.
4.3 Demographics Overview

The results of the survey were mostly developed utilizing quantitative methods. However, to provide better context on the respondents' work experiences, understanding the demographics of the respondents was important as knowing the backgrounds of these women showed who the results really applied to. There were several sub-groups that were not well represented by the data, so understanding the shortcomings of the data was important when considering the results. A deep-dive on the demographics of the survey respondents showed which groups were well represented and which groups might require further analysis to better understand their work experiences. The demographic areas analyzed were career preferences, personal information, education experiences, company information and programming career information.

4.4 Current Career Demographics

The current career of respondents was predominantly software engineer, followed by other careers in software, and a few responses from those in non-software careers. These responses were kept as the original question that guided the survey questions and hypotheses was "Why do women leave software engineering roles," and to answer that question it did not matter where the engineers left to, just that they were no longer software engineers.

However, this breakdown should still be noted as the group that left software engineering altogether is not well represented and could require additional investigation if the question needing addressing is "Why do software engineers leave Tech." This question was not investigated for two reasons:

1. It was too difficult to get many responses from women leaving software engineering and Tech
2. This group has been somewhat captured in existing research. Female software engineers fall into the category of women in technical roles that leave Tech and there has been significant research on why women in technical roles leave Tech.
Since the results cannot discern between why women leave software engineering and stay in Tech vs why women leave software engineering \textit{and} Tech, the rest of the analyses better apply to the former group.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure4_1.png}
\caption{What is Your Current Career?}
\end{figure}

When asked the question, "In three years, what career do you see yourself in" the respondents had a slightly different breakdown than the current career breakdown.
Many respondents reported being software engineers and they planned on staying in software engineering. While 142 respondents reported being currently in software engineering roles, only 120 reported planning on staying in SWE roles. Of the respondents who reported being current software engineers, only 6 respondents wanted to leave software altogether. The other 20 planned on staying in the Tech industry, but no longer wanted to be SWEs.

The responses to these initial two questions illustrated, for this specific set of respondents, that women planning to leave software engineering roles overwhelmingly want to stay in careers in Tech. This further fueled the hypothesis that women leave software engineering roles for different reasons than why women leave Tech altogether, since these women are not actually planning on leaving Tech.

It is important to note that this theme was consistent in the respondents' current careers, too. Given that all respondents were either considered current software engineers or previous software engineers, many more respondents reported that they were currently working in Tech. This finding
may have been very specific to this data set, as the distribution channels may have limited responses from those who no longer work in Tech. However, one point was true even with this knowledge, female software engineers mostly want to stay in Tech: of the 22 respondents that planned on leaving software engineering roles, 16 of these women were planning on staying in Tech.

4.5 Dependent Variable Demographics

In the rest of the demographics and results sections, compared groups may be referred to as Remain SWE and Non-Remain SWE. These groups are based on combinations of the groups mentioned above. Various demographics sections make comparisons based on these groups, so defining them before quantitative analysis was necessary.

Remain SWE:
This group contains the women that reported being current software engineers (142 count) and plan on remaining SWEs for at least the next three years (142 - 22 leaving = 120 count).

Non-Remain SWE:
This group contains those that had already left SWE roles (41 count) plus the SWEs that planned on leaving SWE in the next three years (41 + 22 = 63 count).

Figure 4.3: Respondents on if they will remain in SWE roles
4.5.1 Personal

Age
The average age of respondents skewed quite young, which could either have been the result of the survey distribution channels or could be a predictor of SWE attrition.

![Figure 4.4: Ages of Respondents](image)

The average age of the dataset was under the age of 34 and might have been lower if the respondents' specific ages were known. This phenomenon was likely because those in Tech are younger on average than those in other careers. The 2023 Stack Overflow Developer survey found that the majority of software developers belong in the 25-34 age group, which is in line with the age ranges found for women SWEs in this study. (*Stack Overflow Developer Survey 2023*, n.d.) However, this skew does not seem to be related to just the SWE role. At Apple, Amazon and Google over half of all of their employees are between the ages of 20 and 30. (*Top 5 US-Based Tech Companies by Number of Employees in 2024*, n.d.) This is significantly lower than the median age of the American workforce, which in 2022 was 42, and this median age was consistent across men and women. (*Median Age of the Labor Force, by Sex, Race, and Ethnicity*, n.d.)
It is important to note that while the respondents skewed young overall, the respondents that said they were planning on leaving a SWE role in the next three years had the least attrition in just the youngest age group. For example, while there were 59 total respondents between the ages of 35-44, of those 59 responses, 25 women planned to no longer be SWEs. That is over 42% of respondents in this age group. For the older age groups (not including 65-74 as there were not enough responses) this percentage was between 30 and 40%. In the 25-34 age group, 27% of respondents planned on leaving a SWE role. So, it appeared that a spike in [planned] attrition was observed in the 35-44 age group and then tapered off. Given this information, there is not a linear relationship between SWE attrition and age.

The following illustrates the ages of respondents that planned on leaving SWE in the next three years. Those that have already left have not been included as we do not know at what age they left the SWE role.

Figure 4.5: Ages of all Profession Developers (Stack Overflow Developer Survey 2023, n.d.)
The 35-44 age group became an outlier and created additional questions around how age relates to women in SWE. Three conflicting hypotheses were developed:

1. Women in this age range generally have very busy family lives and want more flexibility (and hope to work less) in their careers and they do not believe SWE roles can provide this.
2. Women in this age are likely to be at the experience level where opportunities in management or other Tech related roles become more appealing.
3. Women have accumulated enough experience in Tech that they can take other roles to leave an unappealing SWE role.

Comparing the low attrition group (25-34) to the high attrition group (35-44) showed whether these age groups had specific factors that were more likely to affect their career plans. The analyses on these specific groups can be found below in the predictive model section.
Dependents
In the Remain SWE group, 33% reported having dependents, with most of those dependents being children. In the Non-Remain SWE group, about 42% of respondents reported having dependents (again, mostly children).

Figure 4.7: Dependents of those than were in the remain SWE group

Figure 4.8: Dependents of those in the non-remain SWE group
This is significant for two main reasons. The first is that women with dependents were well represented in both main populations. The second is that the population with a higher ratio of dependents was the Non-Remain SWE group. This finding supports the hypothesis that women with dependents are more likely to not remain in SWE roles. However, it does not mean they are directly related.

A few respondents did directly say, in the open-ended questions, that if they were to leave a software engineering role in the future that it would be because they wanted to spend more time with their children and could not sustain the time needed to dedicate to the SWE role. These reports align with the existing research on why women might leave Tech.

4.5.2 Education

For the survey respondents, nearly all had at least a bachelor's degree and most had a master’s or higher. Between those who were in the Remain SWE group and those that were in the Non-Remain SWE group there were no discernible differences in education level.

The following chart shows the degree levels of all respondents. These are not broken down by dependent variable groups.

![Figure 4.9: Respondents highest level of schooling completed](image)
Most respondents reported having a computer science degree, and this did not significantly differ between the Remain SWE and Non-Remain SWE groups.

There were also no significant differences between the dependent variable groups for when respondents first learned how to program. Interestingly, almost as many respondents learned programming in high school as they did in their undergraduate degrees, but still, this was consistent across the two groups.

![Figure 4.10: When respondents first learned to program](image)

4.5.3 Company

*Company Size*

When considering all respondents, over half came from large companies. This may have biased the results to be more representative of female software engineers in Big Tech versus those in startups or smaller companies.
Was the breakdown of company sizes different for those who planned to stay in or leave software engineering roles?

Figure 4.11: Company size reported by respondents

Figure 4.12: Company Size for those planning to leave SWE roles
The two groups were not largely different when it came to company size; however, it was clear that in the Non-Remain SWE group a higher percentage of respondents came from mid-size companies vs larger companies.

**Location**

Most respondents appeared to work at flexible (location) workplaces. Nearly all reported working either fully remote or hybrid (about 3 days a week in office). Very few responded as working fully onsite. This is important to note because there was very little representation in the data for those that reported working fully onsite.

Another small note is that between the current work settings and preferred work settings of respondents, almost no one wanted to be working fully onsite. The majority preferred being fully remote, but over a third of respondents still preferred being hybrid. This finding aligns with women in the workplace more generally and their needs for more flexible work environments. In a follow up study, it would be helpful to find more respondents from the "fully onsite" group, to see if their working experiences differ from the respondents in the other categories.

Due to this distribution of data, the results of the predictive models will not be able to address the
hypothesis that women in remote SWE roles are more likely to stay in SWE roles.

![Histogram showing current work setting of respondents]

**Figure 4.14: Reported current work setting of respondents**

![Histogram showing preferred work setting of respondents]

**Figure 4.15: Preferred work setting of respondents**

4.5.4 Career

*Years of Programming Experience*

Most survey respondents were classified as mid-career, as most reported having 2-10+ years of *programming* experience. The majority reported having more than 10 years of programming
experience, which illustrated a well experienced population. Because of this, the reports of our respondents were less likely to be due to a brief good or bad experience in their SWE careers.

Figure 4.16: Respondents' years of programming experience

When those who were in the Remain-SWE group vs the Non-Remain SWE group of respondents were analyzed, it seemed that at every level of experience there was some attrition. I expected that those in the 10+ years of programming experience would not have much attrition, but there still was roughly 30% expected for this group. However, the groups with less experience did have much higher attrition levels than those with more experience.

Figure 4.17: Remain SWE years of programming experience
On the left were those that will remain in SWE and on the right are those that planned to leave (or left SWE roles). For those with under 2 years or 6-10 years of experience in programming, they had the worst attrition rates of over 40%. The 2 to 5 years of experience group had something in between.

Similar to looking at the ages and attrition to determine which groups were most at risk for attrition, years of experience was important, too. The new graduates, or those with very little programming experience, were more likely to leave SWE roles. This could be because they discovered that they have other interests early on, or they might have been more affected by a one-off bad experience when determining if they wanted to stay in their SWE career.

For those in the 6-10 years of experience group, we might have seen something similar as the mid-career age group described above. As the data shows (later) in the predictive models, the top two reasons for leaving SWE roles are because respondents do not enjoy software engineering and respondents believe that there are better opportunities elsewhere. There are various reasons this group might be leaving, and it could have to do with them either having enough experience to move into roles that they see as being better opportunities or it could be that they are leaving due to family responsibilities. Based on this, the mid-career group of respondents needed to be
analyzed independently to understand their experiences, as these two hypotheses are very different. The models for the "middle" age group are a subset of all models developed, and these models should help clarify the two possibilities for this spike of attrition in mid-career individuals.

Given what is known about the demographics of the MIT Women in Software Engineering survey respondents, the analysis of the data was influenced. While an initial run of predictive models illustrated the reasons women might stay in or leave a SWE role for the population as a whole, more targeted analyses occurred with specific age groups.
Chapter 5  Results Analysis

5.1  Quantitative Analysis Overview

To establish an understanding of the important variables when considering what might influence women to remain in or leave software engineering roles, the three methods (logistic regression, Random Forest and XGBoost) were utilized to develop predictive models for different slices of data. The Shapley Values method was utilized for each model developed to understand the importance (strength and direction) of each variable on the models.

The initial analysis of the survey respondents was all inclusive of the data and variables. Various additional analyses were done with unique slices of data and variables to learn more about areas of interest. One subset of data that was further analyzed was for the mid-career women. These respondents appeared to have more attrition than all respondents overall, so understanding if their experiences were different from all other respondents was important.

5.1.1  Defining Independent Variables

After converting the survey questions and responses into a data frame that removed heavily (>75%) correlated variables, the models were left with 35 independent variables. The models determined which of these 35 variables contributed to predicting whether a woman would Remain in a SWE role. These variables map directly to the survey questions listed below.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Survey Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERS_Dependents</td>
<td>Do you have dependents?</td>
</tr>
<tr>
<td>PERS_Learn_Prog</td>
<td>When did you learn to program?</td>
</tr>
<tr>
<td>PERS_Yrs_Prog</td>
<td>How many years have you programmed?</td>
</tr>
<tr>
<td>PERS_Is_Manager</td>
<td>Are you a SWE manager?</td>
</tr>
<tr>
<td>COMP_Size</td>
<td>What is your company size?</td>
</tr>
<tr>
<td>Question</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------</td>
</tr>
<tr>
<td>COMP_Advocate_Pros</td>
<td>Does your company have mentorship programs?</td>
</tr>
<tr>
<td>COMP_Curr_Setting</td>
<td>What is your current work setting?</td>
</tr>
<tr>
<td>TEAM_CoLo</td>
<td>Are you colocated with your team?</td>
</tr>
<tr>
<td>TEAM_Women_Team</td>
<td>How many women are on your development team?</td>
</tr>
<tr>
<td>TEAM_Pairing</td>
<td>Does your team practice pair programming?</td>
</tr>
<tr>
<td>MANAGE_Coding</td>
<td>Does your manager code?</td>
</tr>
<tr>
<td>MANAGE_Identity</td>
<td>What is the gender identity of your manager?</td>
</tr>
<tr>
<td>MANAGE_Feedback</td>
<td>How often does your manager provide feedback?</td>
</tr>
<tr>
<td>MANAGE_Liked</td>
<td>Do you enjoy working with your manager?</td>
</tr>
<tr>
<td>PERF_Freq</td>
<td>How often is your performance reviewed?</td>
</tr>
<tr>
<td>PERF_Levels</td>
<td>How well are levels and roles defined?</td>
</tr>
<tr>
<td>PERF_Reviews</td>
<td>Is the performance review process at your company well defined AND consistent across the organization?</td>
</tr>
<tr>
<td>PERF_Fair</td>
<td>Do you feel your performance process is fair?</td>
</tr>
<tr>
<td>SWE_Hours</td>
<td>How many hours do you work a day?</td>
</tr>
<tr>
<td>SWE_Overtime</td>
<td>How often do you work evenings and weekends?</td>
</tr>
<tr>
<td>SWE_Flex_Hrs</td>
<td>Do you have the ability to choose your own working hours?</td>
</tr>
<tr>
<td>SWE_Question_Ability</td>
<td>How often has your ability to do your role been questioned?</td>
</tr>
<tr>
<td>SWE_Fraction_Admin</td>
<td>What percent of your team's administrative work do you do?</td>
</tr>
<tr>
<td>ENV_Values</td>
<td>I am proud of my company culture and values</td>
</tr>
<tr>
<td>ENV_Opps</td>
<td>I feel that there are better career opportunities outside of the software engineering career path</td>
</tr>
<tr>
<td>ENV_Included</td>
<td>I feel socially included on my team</td>
</tr>
<tr>
<td>ENV_Skills</td>
<td>I worry about keeping my technical skills sharp</td>
</tr>
<tr>
<td>ENV_Interviews</td>
<td>I worry about going through technical interview processes</td>
</tr>
<tr>
<td>ENV_Advance</td>
<td>I have opportunities for advancement in my current role</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>ENV_Aligned</td>
<td>My values align with the work that I do</td>
</tr>
<tr>
<td>ENV_Team</td>
<td>I enjoy working with my teammates</td>
</tr>
<tr>
<td>ENV_Confidence</td>
<td>I am confident in my ability to succeed as a software engineer</td>
</tr>
<tr>
<td>ENV_Visible</td>
<td>I have been given as many opportunities to work on high visibility projects as my male counterparts</td>
</tr>
<tr>
<td>ENV_Comp</td>
<td>I feel that I am well compensated</td>
</tr>
<tr>
<td>ENV_Enjoy_Prog</td>
<td>I enjoy programming professionally</td>
</tr>
</tbody>
</table>

*Table 5.1: MIT Women in Software Engineering independent variables*

### 5.1.2 Method Explanation and Development

In the diagram below, all models developed are outlined. Essentially, 6 different models were developed with all responses to remove bias from any one method (using categorical variables vs using numerical variables) and to provide robustness as the different predictive methods can lead to different results.

It is worth noting that the age group models did not employ logistic regression as a method. This is because when segmenting the data, there were not enough data points to get an accurate logistic regression model and the tree-based methods perform better with fewer data points. (Kern et al., 2019)

Ultimately, the synthesis of model results allowed for the identification of the strength and direction of influence of the independent variables outlined above. Refer to the Methods chapter to understand the reasoning for using each of the predictive models.
Figure 5.1: MIT Women in Software Engineering models developed from survey data
5.2 Results for All Variables and All Survey Responses

The first set of models (the 6 listed in the first diagram above) developed were created with all independent variables and all survey responses. The goal of creating these models with both numerical and categorical variables was to understand which variables were the most important across all models developed (in order to reduce the bias of assumptions and increase confidence in these variables being important).

Results for the logistic regression, Random Forest and XGBoost models were compared for the numerical independent variables and then again for the categorical independent variables. The synthesis of the findings of all of these models show for all respondents, which variables were most related to women remaining in software engineering roles. For each model, roughly the top ten variables were mentioned of the 35 total independent variables. Most models reported more than these top 10 as being important, but their importance declined significantly after the first few. Additionally, the Shapley Values method, when applied to most of the models, provided roughly 10-15 important variables. The takeaway here is that many of these 35 variables are related to women remaining in SWE roles; however, the relative importance is not the same across these variables.

5.2.1 Summary of Model Accuracies for All Respondents

<table>
<thead>
<tr>
<th>Accuracies by Model</th>
<th>Logistic Regression</th>
<th>Random Forest</th>
<th>XGBoost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical IV</td>
<td>train: .96</td>
<td>train: .959</td>
<td>train: 1.0000</td>
</tr>
<tr>
<td></td>
<td>test: .70</td>
<td>test: .81</td>
<td>test: .8649</td>
</tr>
<tr>
<td>Numerical IV</td>
<td>train: .84</td>
<td>train: .836</td>
<td>train: 1.0000</td>
</tr>
<tr>
<td></td>
<td>test: .78</td>
<td>test: .784</td>
<td>test: .8378</td>
</tr>
</tbody>
</table>

When examining the important variables in each model, it is critical to note that while the
numerical independent variable models outperformed the categorical (likely because the ratio of responses to variables was higher in the former), the XGBoost model for categorical independent variables performed the best. The important variables for the XGBoost model were the baseline, as it was the best model, and each other model details illustrate how the results can be reproduced with different methods (even if some variables rank slightly different).

A major takeaway when considering these values is that even with such a small dataset there was very high accuracy in almost every model for both the test and train datasets. Given that the test sets were 20% of the whole set, or 37 responses (equally distributed between remain and non-remaine groups) they generalized well. The high accuracy in these models shows that the important independent variables in these models do predict whether a woman will remain in a SWE role.

### 5.2.2 Summary of Model Results

**Logistic Regression**

Given that this model type generally does not perform well compared to tree-based methods (Kern et al., 2019), the scores were encouraging and they illustrated that there is a relationship between the variables given and whether a woman will remain in a software engineering role. What the results mean is that for the numerical independent variable model, roughly 78% of the time the logistic regression model with numerical variables would predict correctly whether a woman would remain in a SWE role in the next three years.

When examining the p-values for these variables, it was clear which predictors were less likely to be significant due to chance (using $p < .05$)--meaning that we can be confident in these being valid predictors. These variables from the numerical independent variable model (sorted by coefficient size) were:

<table>
<thead>
<tr>
<th>IVs</th>
<th>Coeffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENV_Enjoy_Prog</td>
<td>1.795275</td>
</tr>
<tr>
<td>ENV_Opps</td>
<td>-1.777124</td>
</tr>
<tr>
<td>TEAM_CoLo</td>
<td>-1.343974</td>
</tr>
<tr>
<td>PERS_Is_Manager</td>
<td>-1.106661</td>
</tr>
<tr>
<td>MANAGE_Feedback</td>
<td>0.989733</td>
</tr>
</tbody>
</table>

*Table 5.2: Logistic Regression Important Variables*
Given that these variables were significant as well as had larger absolute value of the coefficients means that these were important predictors in the model and were statistically significant. These variables became the baseline for comparing with other models. Even though the logistic regression models performed worse overall than the tree-based methods, this model type provides statistical significance (p-values), which the others do not. The variables here that overlap with the tree-based methods will be the variables to focus on.

**Categorical IVs**

Interestingly, and notably different from the numerical independent variable logistic regression model was that some discrete options were significant predictors versus an entire question. For example, in the numerical IV models, MANAGE_Feedback (frequency of feedback from manager) was a key important variable. However, in this model "MANAGE_Feedback_At least once a month" was important and some of the other options for MANAGE_Feedback were not. So, in this model, a manager providing feedback at least once a month was strongly related to remaining in software engineering, but the lack thereof was not. Since the model accuracy was not as good for the test set, these variables importances should be considered with caution, but it was helpful to understand which categorical variable options were important compared to the high level question.

**Tree Method: Random Forest**

The important variables in this model were found using the Shapley values method and a summary of these are illustrated below.

The way to interpret the Shapley Value summary is as follows:

- The blue on each dot means that the value of the feature is low. So, for the very first variable, enjoying programming, the blue dot scores mean 0 on the 0 to 1 enjoyment scale. We know in all other models (logistic regression and XGBoost models) this is actually the opposite, so what we are seeing here in both of these is an inversion of the feature values.
- Then the SHAP value, which is the horizontal axis, represents the "coefficient" in the sense that it illustrates the impact on the model.
Based on this information, you would interpret this as not enjoying programming is related to remaining in software engineering. As previously mentioned, this is actually the opposite (Aas et al., 2021).

The following graphics show the important features from the SHAP method applied to the variables from the Random Forest models. The first graphic was created for numerical IVs. The second is the SHAP values from the Random Forest model for categorical IVs.

![SHAP Values](image)

*Figure 5.2: SHAP values for numerical IVs from Random Forest model*
In Figures 5.2 and 5.3 above, in conjunction with the statistically significant variables from the logistic regression models, the most important variables across the models are overlapping. Comparing one more model output, the XGBoost models, will finalize the important variables that were relevant across the board. This exercise brings confidence to the results.

**Gradient Boosted Tree Method: XGBoost**

The most important variables for the XGBoost models, using the Shapley Values for interpretability, were the following:
These are the most important variables from the numerical independent variable model.

![SHAP values for numerical IVs from the XGBoost model](image)

*Figure 5.4: SHAP values for numerical IVs from the XGBoost model*
Figure 5.5: SHAP values for categorical IVs from the XGBoost model

In these plots, SHAP provides "numeric importance values for every feature for every individual prediction" (What Is the SHAP Beeswarm Chart?, n.d.). Given the XGBoost model had the highest accuracy on both the test and train data sets, more weight was given to these important variables when comparing the variables in each other model. Additionally, the XGBoost model for categorical independent variables provided a bit of extra useful information in determining which response options were the most important, too.

5.2.3 Comparing Models' Important Variables for All Survey Data

Each model's top variables are listed in the table below. They are listed in the order of importance for each model, and the variables not listed were not necessarily deemed “unimportant,” but were comparatively less important than those listed here. All variables that were found to be important
for individual models are visible in the SHAP diagrams above.

Table 5.3: Comparing important features across all models

<table>
<thead>
<tr>
<th>Categorical Independent Variables</th>
<th>Numerical Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XGBoost</strong></td>
<td><strong>Random Forest</strong></td>
</tr>
<tr>
<td>ENV_Enjoy_Prog</td>
<td>ENV_Enjoy_Prog</td>
</tr>
<tr>
<td>ENV_Opps</td>
<td>ENV_Ops</td>
</tr>
<tr>
<td>TEAM_CoLo_Yes</td>
<td>TEAM_CoLo_Yes</td>
</tr>
<tr>
<td>ENV_Advance</td>
<td>MANAGEMENT_Feedback</td>
</tr>
<tr>
<td>ENV_Skills</td>
<td>ENV_Advance</td>
</tr>
<tr>
<td>ENV_Aligned</td>
<td>ENV_Aligned</td>
</tr>
<tr>
<td>PERS_Is_Manager_Yes</td>
<td>MANAGEMENT_Feedback At least one month</td>
</tr>
<tr>
<td>MANAGE_Feedback_Never</td>
<td>ENV_Comp</td>
</tr>
<tr>
<td>COMP_Curr_Setting_Fully remote</td>
<td>PERS_Is_Manager_Yes</td>
</tr>
<tr>
<td>ENVIshued_Prog</td>
<td>MANAGE_Coding</td>
</tr>
<tr>
<td>ENV_Included</td>
<td>MANAGEMENT_Feedback</td>
</tr>
</tbody>
</table>

The most important variables for all models are listed highest to lowest based on Shapley Values. Most top variables were consistent across all three models. The Logistic Regression model shows all the top variables; however, only the first five were statistically significant. Given that the XGBoost model was significantly more accurate than the other model types for both sets of independent variables (the first column above) these variables will be the ones analyzed. For the XGBoost categorical independent variable model, none of the variables were outliers, meaning each variable was at least in one other model listed.

5.2.4 Important Variable Interpretation

**ENV_Enjoy_Prog: Enjoying programming**

Every single model developed, including those with different test and train datasets ranked Enjoying Programming as the most important variable, where rating enjoying programming as high meant the respondent would remain in software engineering (positively related). Not only was it always the most important, but this variable consistently had a much stronger relationship to the dependent variable (Remain_SWE) than other variables. While it makes complete sense that enjoying programming more would mean one would be more likely to remain in a software engineering role, there are a few unanswered questions related to this most important variable.
If not enjoying programming is the biggest reason why women leave software engineering roles, then why are there still so many fewer women in software engineering than men? What this finding suggests is that women either do not like programming as much as men or men stay in software engineering roles even if they do not like programming.

In the 2022 Stack Overflow Developer Survey over 72% of respondents said that they programmed in their free time as a hobby. An assumption here is that those that program as a hobby enjoy programming, which feels like a safe assumption. However, in the MIT Women in Software Engineering survey, only 55% of respondents said they enjoyed programming. Since only 5% of the respondents in the Stack Overflow survey were women, even if all of these women were hobbyist programmers, that means at least 67% of the respondents in the survey that said they programmed as a hobby outside of work were men. What the data suggests is that men really do enjoy programming more than women.

It is hard to believe that inherently women enjoy programming less than men. More analysis will be needed to understand why this appears to be the case. A follow up study must define what it means to enjoy programming and what aspects of the software engineering role are bringing developers the most enjoyment. The prevailing theory is that women do not like programming because of the culture, as outlined in Brotopia and the Accenture Women in Tech study, and not because they truly do not like programming. In a follow up study, understanding if enjoying programming is being conflated with other aspects of being in a software engineering role. Some of the findings in the MIT Women in Software Engineering Survey suggests that this could be the case.

**ENV_Opps: Believing that there are better opportunities outside of software engineering**

Across all models, believing that better opportunities existed outside of software engineering was the second most important factor in deciding whether a woman would remain in a software engineering role. Believing there are (or are not) better opportunities out there seems to be a very subjective factor, which makes analyzing it hard. A few ideas come to mind:

Could we objectively say whether the software engineering career path has better opportunities
than others in Tech? Or is it in fact completely subjective? Given that the software engineering career path is generally more financially lucrative than other careers in Tech (up to the highest levels of leadership) it feels like opportunity cannot be tied to compensation. Unless the respondent believes that they will not progress through the software engineering career path at the same rate as they would in other careers in Tech. Then perhaps they believe that they will have better opportunities in other roles, because they do not believe that they can be successful in the software engineering career path.

There is somewhat of a case for women believing that greater compensation is not an important aspect to better career opportunities. Compensation was not one of the top reasons that predicted whether a woman would remain in software engineering, which led to this idea. Ultimately, additional data would be needed to understand whether women believe that other opportunities exist for them because they are being pulled to the new opportunities or they are being stifled in their software engineering role.

Team CoLo: Being co-located with your team

Being co-located with your team was the third most important variable in nearly every model developed, too. While the numerical models showed that the answer "Yes" was related to not remaining in SWE and "No" was related to staying (and "Partially" was in the middle), the categorical models illustrated that answering "Yes" to the question was more important than answering "Partially" or "No" when evaluating a relationship to remaining in SWE.

One theory for why being fully co-located with your team is related to leaving software engineering roles is that this means there is low flexibility in working arrangements. Being co-located was correlated somewhat with current working location. Of the current SWEs that reported that their teams were co-located, 23 out of 28 reported being in office at least three days a week.
When looking at the preferred work location of those co-located with their team it does seem that some of those women would prefer to work remotely.

Another possibility is that being fully co-located requires engineers to be fully immersed in the team and company culture, and if the culture is "bad" then this might be the underlying issue. Interestingly, when looking at the "ENV_" mean values between the co-located vs non-co-located engineers this only partially seemed to be the case.

<table>
<thead>
<tr>
<th>Non-co-located SWEs</th>
<th>Co-located SWEs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ENV_Values</strong></td>
<td><strong>ENV_Values</strong></td>
</tr>
<tr>
<td>68.065421</td>
<td>66.000000</td>
</tr>
<tr>
<td><strong>ENV_Opps</strong></td>
<td><strong>ENV_Opps</strong></td>
</tr>
<tr>
<td>35.970000</td>
<td>42.076923</td>
</tr>
<tr>
<td><strong>ENV_Included</strong></td>
<td><strong>ENV_Included</strong></td>
</tr>
<tr>
<td>64.870370</td>
<td>68.481481</td>
</tr>
<tr>
<td><strong>ENV_Skills</strong></td>
<td><strong>ENV_Skills</strong></td>
</tr>
<tr>
<td>69.572727</td>
<td>79.851852</td>
</tr>
<tr>
<td><strong>ENV_Interviews</strong></td>
<td><strong>ENV_Interviews</strong></td>
</tr>
<tr>
<td>73.576577</td>
<td>81.185185</td>
</tr>
<tr>
<td><strong>ENV_Aligned</strong></td>
<td><strong>ENV_Aligned</strong></td>
</tr>
<tr>
<td>68.148148</td>
<td>68.230769</td>
</tr>
<tr>
<td><strong>ENV_Advance</strong></td>
<td><strong>ENV_Advance</strong></td>
</tr>
<tr>
<td>63.318182</td>
<td>59.291667</td>
</tr>
<tr>
<td><strong>ENV_Team</strong></td>
<td><strong>ENV_Team</strong></td>
</tr>
<tr>
<td>75.522936</td>
<td>69.115385</td>
</tr>
<tr>
<td><strong>ENV_Confidence</strong></td>
<td><strong>ENV_Confidence</strong></td>
</tr>
<tr>
<td>72.891892</td>
<td>60.076923</td>
</tr>
<tr>
<td><strong>ENV_Visible</strong></td>
<td><strong>ENV_Visible</strong></td>
</tr>
<tr>
<td>64.149533</td>
<td>65.807692</td>
</tr>
</tbody>
</table>
The common thread seems to be that imposter syndrome (An Antidote to Impostor Syndrome, n.d.)-related qualities are higher among female software engineers that are co-located. The co-located SWEs had a mean confidence in their "ability to succeed as a software engineer" that was 20% lower than the non-co-located SWE group. Second, the co-located SWEs had 14% higher mean reports of concerns keeping their technical skills sharp. Co-located SWEs experienced being more concerned about the SWE interview processes and reported not liking their teams as much as their non-co-located counterparts. Finally, co-located SWEs were less likely to believe they were being paid fairly.

It makes sense that co-located teammates might increase imposter syndrome in women. It becomes much easier to compare yourself to the people that you work with, when you are in the same place. The last variable, fair compensation, might also be more important for co-located individuals as working closely with peers increases the chances that salary would be talked about and discrepancies between teammates might be found.

**MANAGE_Feedback: Frequency of manager feedback**

It was not surprising that female SWEs that were provided more frequent feedback were more likely to stay in software engineering roles. In fact, in the categorical IV models, two top variables were frequent manager feedback being associated with remaining in SWE and never getting manager feedback being associated with not remaining in SWE roles. Again, this is not entirely surprising; however, what is is that other manager qualities were almost insignificant.

The gender of the manager did not matter as well as how much the respondent liked their manager. It came down to managers doing their job in supporting employees in their career growth. Even the skillset of the manager did not seem to matter. In fact, respondents that had managers who were more technical (still coded at least some of the time) were more likely to leave SWE roles.
This is important to note because many SWE managers find themselves in the role because they have been successful in their SWE careers, not necessarily because of their leadership capabilities. What good software engineering leaders look like would be another area for additional research.

*ENV_Advance, ENV_Skills, ENV_Visible, ENV_Aligned, PERF_Fair*

These independent variables were found to be important across the different models developed. The reason that these are grouped is that they are not particularly surprising.

- **ENV_Advance:** Having opportunities for advancement was related to Remain SWE
- **ENV_Skills:** Having less concern for keeping technical skills sharp was related to Remain SWE
- **ENV_Visible:** Having opportunities to work on high visibility projects was related to Remain SWE
- **ENV_Aligned:** Female SWEs feeling that their values aligned to the work they did was related to Remain SWE
- **PERF_Fair:** Feeling that the performance process was fair was related to Remain SWE

While none of these were surprising, what was interesting is that these variables were more important than most of the remaining variables.

*COMP_Curr_Setting: The respondent’s current work setting*

This survey question and variable became slightly problematic for this analysis. It was not realized until the analysis process that the current company setting was for all respondents currently, and not what their work setting was when they were software engineers. The predictive models showed that being fully remote was actually associated with not remaining in software engineering. However, this could potentially mean that respondents who were in onsite software engineering roles left for remote non-SWE roles, so the results do not necessarily mean that being remote SWEs would be related to more SWE attrition.

To better understand this variable, all current non-SWEs needed to be removed. Then, the Remain and Non-Remain SWE breakdown were able to be examined for those SWEs that reported currently being remote or being hybrid/onsite. These results showed that the current work location of those that plan to remain in SWE roles was not significantly different from those that plan to
leave.

Figure 5.8: A comparison of work settings for women that remain in or leave SWE roles

This follow-up shows that the current company setting of a SWE is not likely an indicator of whether a woman will remain in a SWE role. This (non) finding is significant because a key hypothesis was that female software engineers would be less likely to leave the role if they were granted the ability to work remotely. The evidence here also supports the idea that being on site is not the reason that co-location was associated with higher Non-Remain SWE responses.

PERS_Is_Manager: Whether the respondent reported being a manager

Across all models, being a manager was inversely related to remaining in the SWE career path. This finding produced more questions and hypotheses than answers. What about being a software engineering manager might be the reason women are leaving SWE roles? It could be for reasons related to both push or pull factors.

- **Push**: women could be experiencing more isolation and barriers as they move up in the SWE career path, making them want to try something else (since they are by and large remaining in Tech).
- **Pull**: women gaining management experience as mid-level SWE managers are provided more opportunities to try new roles outside of the SWE career path.

Ultimately, more data is needed to make any determinations on what this phenomenon means. Understanding what roles SWE managers end up in would provide better insights on whether their
leaving is more for push or pull reasons.

The key findings for the entire survey population were that enjoying programming and not feeling like there were better career opportunities outside of the SWE career path were the two most important variables for predicting whether a woman would remain in a SWE role. Unfortunately, there are still many unknowns. What does it mean for someone to enjoy programming? Why do women think there are better career opportunities outside of the SWE career path? Additional research would be required to answer these two key questions.

While these questions cannot be answered now, there is one other area of interest that can be analyzed with the predictive models. When analyzing the demographics, it appeared that the 25-44 age groups had worse attrition than the other age groups. Learning if the two age groups have unique predictor variables will help identify whether the above findings can generalize to all age groups or if women at different stages in life leave software engineering roles for different reasons.

5.3 Age Group Models

Because the data had shown that women in the 25 - 44 age groups had the worst attrition of all respondents, analyzing those groups individually helped outline which factors might have been most significant in these groups specifically. Understanding if the 25 - 44 age groups had unique experiences from the rest of the respondents could influence recommendations for Tech companies trying to lower their female software engineering attrition rates. For both the 25 - 34 and the 35 - 44 age groups, XGBoost and Random Forest models were developed and the important variables in each model were identified. These models were only developed with the numeric independent variables because we had fewer data points for each group.

Model Comparison Results for XGBoost (the best accuracy model)

<table>
<thead>
<tr>
<th>Accuracies by Model</th>
<th>25-34</th>
<th>35-44</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGBoost</td>
<td>train: 1.0000</td>
<td>train: 1.0000</td>
<td>train: 1.0000</td>
</tr>
</tbody>
</table>
Table 5.4: Model accuracies for age group-based models

<table>
<thead>
<tr>
<th></th>
<th>Test</th>
<th>Test</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.89</td>
<td>1.00</td>
<td>.8378</td>
</tr>
</tbody>
</table>

5.3.1 Models for the 25-34 Age Group

Overall, the models built with just one age group each did significantly better than the "All Respondents" models. Below, Figure 5.9, are the most important variables from the Random Forest model (which is inverted, again) and Figure 5.10 are the most important variables from the XGBoost model.

Figure 5.9: Ages 25 - 34 Random Forest important features
In this age group, enjoying programming was still the most important variable. However, feeling like there were better opportunities outside of software engineering fell down the list significantly. It seems this age group remaining in SWE is more likely if the respondent feels included on their team, and, almost "most" importantly, they feel like they have opportunities for advancement. Finally, having flexible work hours and believing their compensation was fair were significant variables in predicting whether respondents in this age group would remain in SWE roles.

5.3.2 Models for the 35 - 44 Age Group

The XGBoost model consistently performed perfectly (accuracy of 1.0) for both the test and train
datasets for this age group only (with different splits of train and test data). While this means the model is overfit to the data, the test and train dataset accuracies were similar enough that it did not matter. What this suggests is that the data in this age group is more similar than all the data in the entire dataset.

In the same method as above, below, Figure 5.11 are the most important variables from the Random Forest model and Figure 5.12 are the most important variables from the XGBoost model.

*Figure 5.11: Ages 35 - 44 Random Forest important features*
In the 35-44 age group, enjoying programming became the second most important variable and feelings of better opportunities outside of software engineering became the most important variable. Given that essentially all of these women plan to stay in Tech, understanding what "better opportunities" are would be helpful. However, the survey did not investigate which roles women were leaving software engineering for and this information would need to be acquired in a future study. One key note is that Remain_SWE does include software engineering managers, so this group does have people-manager representation. Understanding if the women that believe better opportunities are in upper management or in completely different career ladders, such as program management, would be key.
5.3.3 Differences Across Age Groups

Below is a table of the most important variables across all age groups. The “all” group represents all of the data in the data set, while the second two columns are age-group specific. These variables were identified using the XGBoost models with numerical independent variables.

<table>
<thead>
<tr>
<th>All</th>
<th>25 - 34</th>
<th>35 - 44</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENV_Engoy_Prog</td>
<td>ENV_Engoy_Prog</td>
<td>ENV_Opps</td>
</tr>
<tr>
<td>ENV_Opps</td>
<td>ENV_Included</td>
<td>ENV_Engoy_Prog</td>
</tr>
<tr>
<td>TEAM_CoLo</td>
<td>ENV_Adv</td>
<td>TEAM_CoLo</td>
</tr>
<tr>
<td>PERS_Is_Manager</td>
<td>ENV_Skills</td>
<td>ENV_Visible</td>
</tr>
<tr>
<td>ENV_Adv</td>
<td>ENV_Comp</td>
<td>MANAGE_Feedback</td>
</tr>
<tr>
<td>ENV_Values</td>
<td>SWE_Flex_Hours</td>
<td>ENV_Confidence</td>
</tr>
<tr>
<td>PERF_Freq</td>
<td>TEAM_CoLo</td>
<td>ENV_Team</td>
</tr>
<tr>
<td>SWE_Flex_Hours</td>
<td>ENV_Values</td>
<td>PERS_Is_Manager</td>
</tr>
<tr>
<td>MANAGE_Feedback</td>
<td>ENV_Opps</td>
<td>COMP_Curr_Setting</td>
</tr>
<tr>
<td>ENV_Aligned</td>
<td>ENV_Team</td>
<td>ENV_Comp</td>
</tr>
</tbody>
</table>

Table 5.5: The most important variables across age groups for the XGBoost models with numerical IVs

The main takeaway here is that each dataset (all ages and then age group specific) had different important independent variables and we can see that these definitely change over time. While ENV_Opps varied drastically between the two specific age groups, a few commonalities arose that were not represented in the initial analysis with all the data.

1. ENV_Comp: Respondents reporting feeling that they are being compensated fairly were more likely to Remain in SWE roles
2. ENV_Team: Respondents reporting enjoying working with teammates were more likely to Remain in SWE roles

Again, these findings are not surprising, but what is interesting is that these variables were more important to those in the 25-44 age group versus the entire dataset. Clearly, beyond enjoying programming, there was nuance in the other variables that affected a woman's decision to remain in a software engineering role. The text responses in the survey provided additional color on what women believe are the reasons that they would like to leave software engineering. The findings
from the text response analysis complement the initial findings in the predictive models above.

5.4 Results from Open-Ended Survey Responses

At the end of the MIT Women in Software Engineering Survey there were two questions that accepted optional, open-ended text responses. The questions were:

1. If you left a role in the software engineering career path, or plan to do so in the next few years, what would be the primary reasons as to why?
2. What is/was your favorite thing about software engineering?

The goal of adding these questions was to capture information that might not be well represented in the rest of the survey questions. With the first question, the goal was to understand the explicitly stated reasons for why women planned on leaving or had already left SWE roles. The goal of asking the second question was to better understand the reasons why women reported liking software engineering. Given that *enjoying programming* was found to be the biggest reason women choose to stay in SWE roles, these additional open-ended questions can help us better understand why women remain in SWE roles.

These two questions were asked to see if women reported leaving SWE roles for reasons other than the important variables found above and what about software engineering might make women choose to stay. Additionally, the previous survey questions never asked for a respondent's explicitly stated reasons for leaving SWE roles. Noting the differences between stated reasons and deduced reasons will generate a better understanding of women's experiences in SWE roles.

5.4.1 Model Themes for Why Women Leave SWE Roles

Two quantitative methods were utilized to help create an initial summarization of the text responses. BERTopic and OpenAI's ChatGPT were the natural language processing models that quickly pulled the themes from the text in each question mentioned above.
BERTopic Model

The first set of themes were generated from the BERTopic model, which uses clustering to find the text themes and then produces the most important words from each cluster to help explain them. The topic denoted as -1 are the words that were removed as they were overused.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Count</th>
<th>Name</th>
<th>Representation</th>
<th>Representative_Docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>46</td>
<td>-1_management_career_interview_engineer</td>
<td>[management, career, interview, engineer, reit...</td>
<td>I haven’t left any previous, as this is my fi...</td>
</tr>
<tr>
<td>0</td>
<td>31</td>
<td>0_engineer_career_engineers_programming</td>
<td>[engineer, career, engineers, programming, eng...</td>
<td>I left and came back. After a layoff (my thr...</td>
</tr>
<tr>
<td>1</td>
<td>31</td>
<td>1_management_manager_career_opportunities</td>
<td>[management, manager, career, opportunities, p...</td>
<td>I am so tired of all the implicit bias. Climb...</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>2_burnout_leaving_staying_leave</td>
<td>[burnout, leaving, staying, leave, stressburno...</td>
<td>As a manager, I worry a bit about my ability ...</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>3_worklife_volunteer_healthcare_wealthy</td>
<td>[worklife, volunteer, healthcare, wealthy, ret...</td>
<td>Wanting to work more in the fitness industry...</td>
</tr>
</tbody>
</table>

Table 5.6: BERTopic model output

Figure 5.13: Each BERTopic themes’ word scores

The topic 0 is not entirely discernable with the information provided by BERTopic. It could mean that the respondents did not like engineering, but more will need to be discovered in the qualitative thematic analysis.

Topic 1 confirms that women reported leaving SWE for other better opportunities. However, another interesting facet of this topic was shown--it seems like the better opportunities could be in management. We might assume that management does not mean engineering management as engineering management was considered to be within the SWE career path. So, believing other management opportunities were better might mean that the management roles referenced are those in other non-engineering roles, such as program/product management.

Topic 2 is on the other end of the spectrum. Topic 2 is more of a push factor than a pull factor for leaving SWE roles. Burnout appears to be a major reason women reported wanting to leave SWE
roles, when they were directly asked in the survey. This finding aligns with some of the important variables found in the models developed above. Burnout is a newly well-defined term, but it is hard to quantify. In 2016, a journal article from World Psychology succinctly describes burnout as the following:

"Burnout is a psychological syndrome emerging as a prolonged response to chronic interpersonal stressors on the job. The three key dimensions of this response are an overwhelming exhaustion, feelings of cynicism and detachment from the job, and a sense of ineffectiveness and lack of accomplishment. The significance of this three-dimensional model is that it clearly places the individual stress experience within a social context and involves the person's conception of both self and others."(Maslach & Leiter, 2016)

These factors for describing burnout align remarkably well with the key variables that had higher associations with co-located teams. When working physically close to teammates, interpersonal issues compound. Therefore, team co-location being a top 3 variable across all of the quantitative models might really be a red-herring for burnout in a SWE role.

Topic 3 was not surprising, as well. Finding work-life balance and retiring were the last category for why women believed they want/ed to leave SWE. It is not clear which of these is more important to this dataset. Combining this initial analysis with the GPT results and the qualitative thematic analysis was important to back up these results.

**GPT Model**

While the BERTopic model provided the groupings of words that could be considered themes for the open-ended responses, it did not provide as clear of outputs as the GPT model. The prompt provided to the GPT model was:

"Identify the top 10 to 15 themes discussed in the following survey responses. For example: 1) Manager, 2) Diversity, equity, and inclusion."

The output of the model was fifteen different themes, listed in order of frequency.
Figure 5.14: The GPT model themes identified from the open-ended survey question 1: Why did you (or will you) leave a software engineering role?

The output for the GPT-based model supports the existing findings. However, burnout was the top theme here. One concern with this output is that it appears that many respondents must have reported there was not enough diversity and that they had felt discriminated against. These two factors were not explicitly found in the methods utilized above, but are consistent with previous research on women in Tech. However, if women were reporting wanting to leave SWE roles for other roles in Tech for this reason, it would be notable. This would suggest that discrimination and lack of diversity is worse in SWE roles than other roles in Tech. One of the main hypotheses for this study was that poor culture was not the primary reason why women leave SWE roles; however, it appears it still may be a significant reason. The qualitative thematic analysis confirms this.

5.4.2 Thematic Analysis Findings

In Table 5.7, the stated primary reasons for leaving SWE were organized by frequency of occurrence in the open text responses. The counts shows are the number of times the factor was mentioned across all open ended survey responses. What is most notable is that stating leaving SWE because of not liking programming was one of the lower stated reasons. The highest stated reason was stress and burnout, which confirms the previous analysis from the natural language processing models. Both findings would not have been captured by the predictive models.
Table 5.7: Explicitly stated reasons for leaving SWE roles from the thematic analysis of the text responses

Most respondents did not report working too many hours. Given that these same respondents were searching for more work-life balance and less stress in their roles, it seems that the intensity of the work, not just the number of hours, were more indicative of women wanting to leave the SWE roles. However, about a third of respondents did report working 9+ hours a day.

Figure 5.15: Average daily hours worked as a SWE
The themes found with the GPT model were well aligned with those found in the qualitative analysis, especially for the first 6 themes identified by the model. While culture was reported as being a reason for leaving, burnout/stress, a desire for work-life balance, and a desire to work in a more meaningful role were reported more often. Many women actually reported that they would like the option to work part-time in SWE roles, but have not seen that opportunity. A few women mentioned that they took time from the role while they had young children and were able to come back once their kids were in school, while others left and found other part-time work.

While understanding the predicted and reported reasons women leave SWE roles can help the Tech industry identify ways that they can improve the experience for women, learning why women report staying might provide additional information for helping draw women towards the SWE career path.

5.4.3 Model Themes found for Why Women Like Software Engineering

The natural language processing models did a good initial review of finding the key "best thing" about software engineering. "Problem solving", in both models, came up as the most common theme in answering the question: What is/was your favorite thing about software engineering?

BERTopic Outputs:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Count</th>
<th>Name</th>
<th>Representation</th>
<th>Representative_Docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>115</td>
<td>0_engineering_work_challenging_working</td>
<td>[engineering, work, challenging, working, sati...</td>
<td>I enjoy the problem solving. The work itself ...</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>1_solving_problems_problem_logic</td>
<td>[solving, problems, problem, logic, code, custi...</td>
<td>Problem solving, problem solving, Problem sol...</td>
</tr>
</tbody>
</table>

Table 5.8: BERTopic themes for answering the question, "What is/was your favorite thing about software engineering?"
The BERTopic model, however, did not do well in terms of determining a spectrum of reasons. This was likely due to the infrequency of the other major reasons mentioned. When developing the GPT-based model, additional themes were outlined.

**GPT Model Outputs:**

1) Problem solving  
2) Coding  
3) Flexibility in work schedule  
4) Working with and helping people  
5) Challenging projects  
6) Learning new technologies  
7) Building and creating  
8) Collaboration with teammates  
9) Solving complex problems  
10) Impact on overall user experience  
11) Mental challenge and stimulation  
12) Elegant and optimized solutions  
13) Valued skill set  
14) Transforming data into solutions  
15) Building robust systems and applications

As seen in Figure 5.18, the GPT model produced many different "themes," and many of them were not actually unique. Creativity, building, collaborating, creating impact, and problem solving were the general key ideas of the above.

When narrowing the output of the model to be just five themes, those were:

\[\text{Figure 5.16: BERTopic word scores for the responses to } \text{"What is/was your favorite thing about software engineering?"}\]

\[\text{Figure 5.17: GPT model outputs for the responses to } \text{"What is/was your favorite thing about software engineering?"}\]
1. Problem Solving
2. Creativity
3. Fulfillment
4. Flexibility
5. Learning

*Figure 5.18: GPT top 5 reasons for staying in SWE*

While these five themes seem to be representative of the output of the first model, the qualitative thematic analysis allowed for a better understanding of the most important reasons that women like software engineering.

### 5.4.4 Thematic Analysis Findings

There were clearly three very important themes when women described why they enjoy software engineering.

<table>
<thead>
<tr>
<th>Stated Reasons for Liking SWE</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem solving and always learning</td>
<td>65</td>
</tr>
<tr>
<td>Being creative and creating something new</td>
<td>34</td>
</tr>
<tr>
<td>Making an impact</td>
<td>30</td>
</tr>
<tr>
<td>Working with people or other teams</td>
<td>13</td>
</tr>
<tr>
<td>Flexibility</td>
<td>10</td>
</tr>
<tr>
<td>Pay</td>
<td>7</td>
</tr>
<tr>
<td>Fun</td>
<td>5</td>
</tr>
</tbody>
</table>

*Figure 5.19: Themes for liking SWE from the thematic analysis of text responses*

Most women that responded to this question reported that problem solving was one of the reasons they liked software engineering. Interestingly, in the reasons for leaving, a few women reported being bored. Having opportunities to learn and continue problem solving appears to be a major reason why women enjoy SWE roles.

The second is interesting because creativity is not often associated with technical roles and careers. Many women reported that they enjoyed being creative in their SWE roles and liked seeing their work produce something completely new.
The third theme in this text response was that women enjoyed making an impact in their SWE roles. It appeared that the respondents wanted to feel like the work that they do matters to more than just the business, but to the people that the business serves. Helping improve the experiences of others (not only customers, but other teammates and their employers) seemed to be a major consideration for the respondents.

5.5 Hypotheses Review

When evaluating the initial hypotheses, the results were mixed. While we discovered that women leave SWE roles for reasons other than why women seem to leave Tech, they were not necessarily the same reasons previously hypothesized.

Most female software engineers that have left or hope to leave software engineering roles go to or plan to go to non-software engineering roles in Tech, but not leave Tech altogether.

This hypothesis was found to be true. Based on data in the demographics section of the results it was found that only 6 respondents intended to leave Tech altogether.

Culture is not the most important reason for why women leave software engineering roles.

There is more information on the results of testing this hypothesis above. The simple and big idea is that culture is not the most important or only reason why women leave software engineering roles. This might be an outdated belief that is based on historical data. Bro-culture seems to be less relevant to the women that responded to the survey; however, since "stress/burnout" were the most reported reasons for women leaving, it seems that there might be a link to company culture that influences stress and burnout. For example, having a fast-paced culture is not necessarily bro-y or hostile to women, but it could be considered an environment that causes stress and burnout.

Disliking working with your manager relates to leaving or wanting to leave a SWE role

To my surprise, the results for this hypothesis were mixed. While liking your manager was not an important variable when considering all respondents together, it seemed marginally important
for the women in the 25-34 category. It appeared that other manager qualities were more important, like providing frequent feedback and making sure the performance review process is fair.

**Most female software engineers that have left or intend to leave software engineering roles go to or plan to go to non-software engineering roles in Tech, but not leave Tech altogether**

For this specific dataset, this was well supported. Of all responses, only 6 of the 183 full responses were from women that were planning to leave SWE roles and Tech altogether. However, as previously mentioned, the dataset makeup is likely biased based on the survey distribution channels. So, while we cannot know if previous software engineers that left Tech altogether have been well represented, we do know that current software engineers do mostly plan on staying in Tech, even if they want to leave SWE roles.

**Female software engineers that work more hours are more likely to have left (or plan to leave) software engineering roles**

This hypothesis did not seem to be proven as true. What mattered more was whether a woman experienced stress/burnout or had flexibility in what hours they worked.

**Female software engineers that report less flexibility in their work environment are more likely to have left (or plan to leave) software engineering roles**

As mentioned above, this seems well supported in this analysis. While having flexibility was not the most important reason women reported remaining in SWE roles, it was considered important not only in the predictive models, but also in the qualitative analysis of the text responses.

**Women leave software engineering roles not because they dislike coding**

This hypothesis is partially true. It would better be phrased as "Women leave software engineering for more reasons than just disliking coding." Reporting enjoying programming (or not) was the most important factor for remaining in software engineering roles in almost every predictive model developed. However, after analyzing the open text responses, this was not one of the reported most important factors. What I have inferred is that women that decide to remain in SWE roles decided that they like SWE and those that left decided that they did not like SWE
(as much), but it was more correlation than causation.

**Women that stay in software engineering (or plan to) are mostly remote**
This is partially true. Most people that reported wanting to remain in SWE roles were either remote or hybrid. However, it did not appear that being remote was associated with women remaining in SWE roles.

**Culture is not the biggest reason for why women leave software engineering roles**
Culture was not the "biggest" reason for why women leave SWE roles. However, it was reported as being a significant reason by women in the open text responses and appeared to be significant in the predictive model factors, too.

**Women with dependents do not stay in software engineering roles as much as those who do not**
Another interesting finding was that respondents that reported having children were not more likely to leave SWE roles. While this was not proven in the predictive models, many women did report wanting more work-life balance, to spend more time with their families. The outstanding question would be, does everyone feel this way, or just women? Do the women that left for work-life balance reasons have better financial support from their spouses? Are their new roles creating better work-life balance? What are these roles that women have left SWE roles for? There are many areas for future research with this hypothesis.

**Women with no other women on their direct teams are more likely to leave software engineering roles**
The literature review found significant research on the harm of being an "only" in a workplace or on a team. While it may be true that being an "only" is related to having worse experience in Tech, it does not seem to be a clear top reason for leaving SWE roles. In the survey research, it was found that having more women on the same team was not necessarily related to women leaving SWE roles. It appears that the results from this survey might refute existing research--these findings would suggest additional research on the "only" phenomenon should be done. In future studies, it would be ideal to understand if company diversity is as important, if not more
important, than just having another woman on your team. This study did not explore company diversity beyond asking respondents whether they have (or had) other women on their teams.
Chapter 6  Conclusions

6.1 Summary of Results

Overall, many variables were found to have influenced the retention of women in SWE roles. Beginning with the stated reasons women decide to leave SWE roles helps put the key variables found in the predictive models into context. For example, women that reported disliking programming were less likely to remain in SWE roles; however, this factor was not explicitly referenced by women when they were directly asked why they would leave SWE roles. So, these results needed to be analyzed together. A summary of the top five stated and model-identified reasons for women leaving SWE roles can be seen below. The size of the circle represents the importance of the variable; however, the importance of the stated factors cannot be directly compared to the model-identified factors as they were found by different methods.

![Diagram showing top factors influencing women in SWE roles]

Figure 6.1: A visual representation of the top factors influencing women in SWE

As previously mentioned, burnout was frequently reported as the most important reason women planned to leave SWE roles (or already had). Many of the predictive model factors aligned with this finding, too. Enjoying programming (or not) and being co-located with teammates likely
represented other factors such as the features related to imposter syndrome and burnout. The features more likely associated with burnout were the following:

- Worrying about keeping technical skills over time
- Worrying about going through the technical interview processes in the future
- How much a woman enjoys working with her direct team
- How much confidence a woman has in believing she will be successful in her SWE role
- Women believing they are being paid fairly

While each of these individually are important factors, it is not certain that all of these are related to enjoying programming and burnout. More research is needed to determine if these factors are being conflated with enjoying programming.

The second most important variable in predicting if a woman would remain in a SWE role was if the woman believed (or did not believe) that there were sufficient opportunities in the SWE career path. This variable was prevalent in the stated reasons from the open-ended text responses as well. It appeared there was general concern around opportunities for more visibility and advancement in the SWE career path. This is interesting because there are well defined career paths for SWEs to move up in companies, whether it is in the staff software engineer route or in the engineering management route. However, we do not know if this belief is due to women feeling like they are being pushed into other roles by being stifled in their SWE roles, or if these women genuinely believe that other careers in Tech have more opportunities for them to succeed.

While these were the most important variables from the predictive models, several other variables were stated in the text responses. Women reported wanting to work in another field that is more interesting to them, needing better work-life balance to spend more time with their children, feeling othered for being the only woman, and being tired of the culture.

One final important note is that while these variables were consistently important across models and methods, their general importance varied across age groups. Women in the 35 – 44 age range had unique experiences from those in other age groups. For example, enjoying programming was
not considered the most important variable for this group, and instead, opportunities were. My hypothesis is that women with young families have different needs and desires in work than their counterparts of other ages.

6.2 Recommendations

Based on the important factors from the quantitative and qualitative analyses, several initial recommendations can be made, with the goal of decreasing the attrition of women in SWE roles in Tech. These factors are grouped into push and pull factors. Here, push factors are reasons that drive women to leave SWE roles and pull factors are reasons that make women want to stay.

6.2.1 Limit Push Factors

Some of the clear top push factors could begin to be limited. While there could be recommendations for each important variable, a subset was identified here.

Need for work-life balance
Create a company culture that encourages employees to spend more time with families or reduces the burden of having to “do it all”. For example, companies could provide increased family leave and vacation time, or they might consider subsidizing childcare. Some traditional engineering companies (like those in defense) have every other Friday off. Limiting “after hours” work as being the norm would be critical.

Forced Team Co-Location
Encouraging team leaders to enforce flexibility in work arrangements would improve outcomes for women.

Unfair performance
At the company-level there should be well defined ladders and levels for SWE roles. Tech companies should push managers to provide frequent feedback to their employees and must verify that the managers are following the set performance process in a fair way. Having well-defined roles, levels and performance processes would also illustrate how women can advance in their
SWE careers and have agency in how they progress.

**Not Enough Visibility**
Tech companies should provide feedback loops between employees and their managers to better understand whether managers are providing their employees with equal opportunity for working on higher visibility projects. Managers and senior leadership should provide equal opportunities for SWEs to present their work upward.

**Inadequate Feedback**
As mentioned above, managers need to provide quality feedback to their employees frequently. Tech companies could establish monthly check-ins between managers and their reports in a way that they can track participation in the process.

**Not Aligning with Company Culture**
This issue is harder to solve. If women do not align with company culture it could be for a variety of reasons. However, ensuring that women feel respected and not “othered” in their SWE roles could be critical.

**Lack of diversity**
Tech needs to feel responsible for building diverse companies. Once women are in the door as SWEs in a Tech company, though, there are other ways to improve perceived diversity. Once while I was working at a large Tech company, there was the goal to have a woman on each engineering team. This was completely the wrong way to improve the feeling of “othering”. Tech companies should try and put female SWEs on the same team to reduce the feeling of “othering” for those SWEs.

### 6.2.2 Increase Pull Factors

Pull factors for women in SWE could be increased. The pull factors were discovered when women reported the reasons why they enjoyed software engineering. The top reasons were enjoying problem solving, being creative/creating something new, making an impact, working with other people and teams, and having flexibility in when and where they work.
The biggest impact on women in keeping them in SWE roles with “pull” factors would be making sure that they are working on interesting projects. Managers need to understand what kinds of work make their female employees feel like their work is creative, makes an impact, is challenging enough to require problem solving and allows them to work with others. Another key factor that Tech companies can influence is creating more flexible working arrangement policies. In the past few years, many Tech companies have pivoted away from their COVID, remote-friendly working environments. Most female SWEs would like to be remote, many women reported wanting a hybrid work setting and just a few women reported wanting to be in office full-time. Allowing employees to choose when and where they work, if they are performing well, would be monumental in providing that desired flexibility.

6.3 Areas for Continued Research

When analyzing the survey results, many areas of opportunities for future research became evident. The key areas where this research could be expanded have to do with the diversity of data sources. Additionally, improving the granularity of the data would improve findings and recommendations for Tech companies when it comes to retaining women (and men) in the SWE career path.

The first key limitation when the results were analyzed was that nearly no respondents were planning on leaving SWE and Tech simultaneously. If the data included women that wanted to leave SWE and Tech, we would better understand whether women leave Tech and SWE for different reasons. For example, current research exists for why women leave Tech in general, this study addresses why women in SWE leave SWE roles for other roles in Tech, but the group of women that leave SWE and Tech altogether were not analyzed. Additionally, because the data sources were relatively biased, we cannot conclude whether women in SWE almost always plan to stay in Tech, even if they plan to leave SWE roles. This dataset would suggest this; however, more data and more diverse data sources would better confirm this.

The second limitation of this study was that men were not included in the survey. While the focus of this study was to understand specifically why women leave SWE, understanding if these reasons
are the same (or different) for men would help define recommendations for retaining both groups in SWE. Also, including men in the survey would show whether women truly have worse experiences in SWE than men. While this knowledge is not necessarily needed to improve the experiences of women in SWE roles, it would help companies understand whether their policies and structure favor one gender over another.

The third area for further research would be areas in the current study that would benefit from more detailed data. One area that would be helpful to investigate is when women in SWE leave the SWE career path, where exactly are they going? Are they leaving Tech? When they leave Tech what roles do they go into? When they leave SWE roles, but stay in Tech, are they going into product management, sales, marketing, etc.? Why do they believe they should pursue new opportunities (push vs. pull reasons)? Learning about where women want to go would show what is missing in their current roles.

Several other survey questions have areas for further research that would require additional data to be captured. In the predictive models from this study, “Enjoying Programming” was the most important factor in determining whether a woman would remain in SWE. However, it is not very clear what it means to enjoy programming. In the open text responses, women report enjoying the problem-solving aspects of programming; however, their reported reasons for leaving were more around stress and burnout. Are stress and burnout inherently part of programming? Not more than a few respondents reported leaving software engineering because they did not like programming, so it was not clear if they were conflating enjoying programming and enjoying their experience of working in a programming role at a Tech company.

Another question that might better be solved by interviewing female SWEs is what it means to experience burnout and what exactly about the SWE role contributes to the burnout and stress reported. Is the company culture the reason for burnout? Depending on the sources of burnout, recommendations for limiting burnout will vary dramatically. Given that this was the number one reported reason for women leaving SWE roles for predominantly other roles in Tech, learning more about what is causing the burnout would be the most important area for a follow-up study. We can make assumptions based on some of the top factors from this study that seem to be related
to burnout, but we cannot definitively say what burnout means for women in SWE roles. Future research on women in software engineering will be a critical step in reproducing the results in this thesis and answering the questions that became clear in this initial research step.
References


Madeleine Golison 128 MITsdm Thesis
**Gap in Venture Capital.**


https://dev.to/alvaromontoro/should-a-lack-of-side-projects-raise-flags-in-an-interview-4ifl


Technology company. (2023). In Wikipedia.


Top 5 US-based Tech Companies by Number of Employees in 2024. (n.d.). Retrieved March 17, 2024,
We ran the numbers, and there really is a pipeline problem in eng hiring. (n.d.). Retrieved October 11, 2023, from https://interviewing.io/blog/we-ran-the-numbers-and-there-really-is-a-pipeline-problem-in-eng-hiring#user-content-fnref-4


Women in the Workplace 2022. (n.d.).


Appendix A: MIT Women in Software Engineering Survey

Demographics

Women in Software Engineering
Thank you for taking the time to submit your information about working as a software engineer. We hope to learn more about why some women leave software engineering roles and why women decide to stay, so information about your experience will be very meaningful to this research.

Demographics

Do you identify as a woman?

☐ Yes
☐ No

What is your age?

☐ Under 18

https://mit.e01.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=5V_z55k096Wg2DvE&ContextLibraryID=UR_7R20jQbF7...
18 - 24
25 - 34
35 - 44
45 - 54
55 - 64
65 - 74
75 - 84
85 or older

Do you have children or other dependents?

- Children
- Other dependents
- No/none

When did you first learn how to program?

- Elementary/Middle School
- High School
- Undergraduate degree
- After completing an undergraduate degree
- Other
What is the highest degree or level of school you have completed?

- Less than a high school diploma
- High school degree or equivalent (e.g. GED)
- Some college, no degree
- Associate degree (e.g. AA, AS)
- Bachelor’s degree (e.g. BA, BS)
- Master’s degree (e.g. MA, MS, MEd)
- Doctorate or professional degree (e.g. MD, DDS, PhD)

If you have a university degree, what area is it in? (if no degree select "none")

- Computer Science or Computer Engineering
- All other engineering
- Other STEM - non-engineering
- Social sciences
- Liberal arts/humanities
- Other
- None

What is your employment setting?
What would be your preferred employment setting?

- Fully remote
- Hybrid (2 days at home, 3 days in office)
- Fully onsite

Career Information

Current Career Information (when referring to software engineering roles, these would be any roles where you have programmed > 50% of the time)

What is your current career?

- Software Engineering: a role that spends at least 50% of the time coding, or managing those who do
Other career in software: any role where you are working with software, but coding 0 to 50% of the time

Other career not in software

In three years, what career do you see yourself in?

Software Engineering: a role that spends at least 50% of the time coding, or managing those who do

Other career in software: any role where you are working with software, but coding 0 to 50% of the time

Other career not in software

How many years have you programmed professionally? (where you have spent at least 50% of your time programming)

Under 2 years

2 - 5

6 - 10

10+ years

I have never have programmed professionally

Do you directly manage software engineers?
Programming Company Info - Eng

Company Information

How large is the company that you work for?

- 1 - 10 people
- 10 - 100 people
- 100 - 1000 people
- Over 1000 people

Is your company public or privately held?

- Public
- Private

Of all of your company’s CEO’s direct reports, how many are
women?

○ 0
○ 1
○ 2
○ 3 or more

How large is your direct development team?

○ 1 - 4 people
○ 5 - 9 people
○ 10+ people

Are you colocated with your team?

○ Yes
○ Partially
○ No

Does your team use a 2 - 4 week development cycle (agile processes)?

○ Yes
○ No
Does your development team have daily check-ins/stand ups?

- Yes
- No

How many other female software engineers are on your development team? (not including yourself)

- 0
- 1
- 2+

Does your development team practice pair programming?

- Always
- Most of the time
- About half the time
- Sometimes
- Never
Does your management encourage more pair programming?

- No
- Maybe
- Yes

Does everyone on your development team report to the same people manager?

- Yes
- No

What is your manager's level of coding experience?

- Currently writes code
- Previously wrote code over 50% of the time
- Has never coded, or coded less than 50% of the time in a previous role

How does your manager identify?

- Male
- Female
How often does your manager provide actionable feedback?

- Never
- Once a year
- Once a quarter
- At least once a month

Does your company have any formal mentorship or coaching programs?

- Yes, and I have participated
- Yes, but I have not participated
- No

Does someone other than direct manager help you plan career next steps?

- Yes
- No
How often is your performance reviewed?

- Once a year
- Twice a year
- At least quarterly
- No regular reviews

How well are roles and levels defined at your company?

- Not defined
- Defined vaguely
- Defined for levels, but not role specific
- Defined for both the separate levels and roles

Is the performance review process at your company well defined AND consistent across the organization?

- Not at all
- Somewhat
- Mostly
- Definitely
Do you feel that your performance review process is fair?

- Definitely no
- Mostly no
- Mostly yes
- Definitely yes

Does your organization participate in 360 reviews? (A 360 review is the process of getting feedback on an individual's performance and/or potential from their manager, peers and others who interact with them regularly)

- No
- Somewhat
- Yes

**Working Preferences - Eng**

**Working Preferences as a Software Engineer**

What is your current level designation at your company? (if
your level is not available, pick the closest)

- Junior/associate level
- Mid-level
- Senior-level
- Manager or Principal level

How much do you enjoy your software engineering role?

- Not at all
- A little
- A moderate amount
- A lot
- A great deal

How much do you enjoy working with your manager?

- Not at all
- A little
- A moderate amount
- A lot
- A great deal
How many hours do you work a day?

- Less than 7
- 7 - 8
- 9 - 10
- more than 10

How often do you work on weekends or overnight?

- Never
- Once a year
- A few times a year
- Once a month
- Every weekend

Can you choose your own working hours?

- Yes
- No

Do you enjoy pair programming?

- Yes
- Sometimes
For the following categories, answer these questions based on your experience as a software engineer (whether past or present). Select a value from 0 to 100, 0 being never applicable, 50 being applicable about half of the time and 100 being always applicable.

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Sometimes</th>
<th>About half the time</th>
<th>Most of the time</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am proud of my company culture and values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoy programming professionally</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I feel there are better career opportunities outside of the software engineering career path</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I feel socially included on my team</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I worry about keeping technical skills sharp

I worry about going through technical interview processes

My values align with the work that I am doing

I have opportunities to work with other female software engineers

I have opportunities for advancement in my current role

I enjoy working with my teammates
I am confident in my ability to succeed as a software engineer

I feel like I can be myself at work

I have been given as many opportunities to work on high visibility projects as my male counterparts

I feel that I am well compensated

How often has your ability to do your job been questioned (per year)?

- None
- Once
- Two or more times

What fraction of administrative tasks do you manage for your team?

- None
- 25%
- 50%
- More than 50%

Programming Company Info – Non Eng

Company Information as a Software Engineer (please answer these as of your most recent role where you were programming > 50% of the time)

How large was the company that you worked for as a software engineer?

- 1 - 10 people
- 10 - 100 people
- 100 - 1000 people
- Over 1000 people
Was your company public or privately held?

○ Public
○ Private

Of all of your company’s CEO’s direct reports, how many are women?

○ 0
○ 1
○ 2
○ 3 or more

How large was your development team?

○ 1 - 4 people
○ 5 - 9 people
○ 10+ people

Did your team use a 2 - 4 week development cycle (agile
processes)?

- Yes
- No

Did your development team have daily check-ins/stand ups?

- Yes
- No

How many other female software engineers were on your development team? (not including yourself)

- 0
- 1
- 2+

Did your development team practice pair programming?

- Always
- Most of the time
- About half the time
- Sometimes
Did your management encourage more pair programming?

- No
- Maybe
- Yes

Were you colocated with your team?

- Yes
- Partially
- No

Did everyone on your development team report to the same people manager?

- Yes
- No
What was your manager's level of coding experience?

- Currently writes code
- Previously wrote code over 50% of the time
- Has never coded, or coded less than 50% of the time in a previous role

How did your manager identify?

- Male
- Female
- Non-binary / third gender

How often did your manager provide actionable feedback?

- Never
- Once a year
- Once a quarter
- At least once a month

Did your company have any formal mentorship or coaching programs?

- Yes, and I have participated
Did someone other than direct manager help you plan career next steps?

- Yes
- No

How often was your performance reviewed?

- Once a year
- Twice a year
- At least quarterly
- No regular reviews

How well were roles and levels defined at your company?

- Not defined
- Defined vaguely
- Defined for levels, but not role specific
- Defined specifically for levels and roles
Was the performance review process at your company well defined AND consistent across the organization?

- Not at all
- Somewhat
- Mostly
- Definitely

Did you feel that the performance review process was fair?

- Definitely no
- Mostly no
- Mostly yes
- Definitely yes

Did your organization participate in 360 reviews? (A 360 review is the process of getting feedback on an individual’s performance and/or potential from their manager, peers and others who interact with them regularly)

- No
- Somewhat
- Yes
Working Preferences - Non Eng

Working Preferences as a Software Engineer (please answer these as of your most recent role where you were programming > 50% of the time)

What was your most recent level designation when you were working as a software engineer?

- Junior/associate level
- Mid-level
- Senior-level
- Manager or Principal

How much did you enjoy your software engineering role?

- Not at all
- A little
- A moderate amount
- A lot
- A great deal
How much did you enjoy working with your manager?

- Not at all
- A little
- A moderate amount
- A lot
- A great deal

Roughly how many hours did you work a day (round to closest whole number)?

- Less than 7
- 7 - 8
- 9 - 10
- more than 10

How often did you work on weekends or overnight?

- Never
- Once a year
- A few times a year
- Once a month
- Every weekend
Could you choose your own working hours?

- Yes
- No

Do you enjoy pair programming?

- Yes
- Sometimes
- No

For the following categories, answer these questions based on your experience as a software engineer. Select a value from 0 to 100, 0 being never applicable, 50 being applicable about half of the time and 100 being always applicable.

<table>
<thead>
<tr>
<th>I was proud of my company culture and values</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>About half of the time</th>
<th>Most of the time</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>60</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Never</td>
<td>Sometimes</td>
<td></td>
</tr>
</tbody>
</table>

https://mit.co1.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=5V_s8j5o9086d24eEd&ContextLibraryID=UR_78R205a7B4...
I enjoyed programming professionally

I felt there were better opportunities outside of the software engineering career path

I felt socially included on my team

I worried about keeping technical skills sharp

I worried about going through technical interview processes

My values aligned with the work that I did
<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Sometimes</th>
<th>Most of the time</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>I had opportunities to work with other female software engineers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had opportunities for advancement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoyed working with my teammates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was confident in my ability to succeed as a software engineer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I felt like I could be myself at work</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had been given as many opportunities to work on high visibility projects as my male counterparts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I felt that I was well compensated

How often had your ability to do your job been questioned (per year)?

- None
- Once
- Two or more times

What fraction of administrative tasks did you manage for your team?

- None
- 25%
- 50%
- More than 50%

Manage Questions

As an engineering manager, please rate the following based on your experience.

<table>
<thead>
<tr>
<th></th>
<th>Slightly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Mostly Agree</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel confident in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>managing my direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoy managing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>better than being an</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>individual contributor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>software engineer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoy working</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with my peers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoy working</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with my direct reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My direct reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>respect me</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I feel that my</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ideas are valued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Slightly Disagree Slightly Agree Mostly Disagree Mostly Agree Agree
0 20 40 60 80 100

I feel like I can improve the experience of female software engineers in my organization

Open Ended Questions

If you left a role in the software engineering career path, or plan to do so in the next few years, what would be the primary reasons as to why?

What is/was your favorite thing about software engineering?
If you have left a software engineering role, what do you enjoy most about your new role? Do you miss anything about software engineering?

Is there anything else you’d like to share?

If you’d be open to a short interview about your experiences please provide your contact information:

Name

Email