Sahil Wadhwa

Nancy Jiang

Project Report Checkpoint

Title: Auditing race and gender inequality in data science related job market

Abstract

In the twentieth century, one of sociology's findings is that race and gender matter in the job market. Jobs were segregated by race and gender with whites earning more than other people of color and men earning more than women. Race inequality in the job market has been a long-standing interest of scholars. Notably, some research indicates that racial gaps are more significant for women than for men. Women face the unique challenges of lower wages and lower rewards in the global workforce. However, as women's relative share in occupations grows nowadays, the gender inequality gap narrows in most job markets(Stier et al.,2014). Besides, a finding shows that race-based discrimination is weaker in high-paid jobs. Back in 2012, the Harvard Business Review acclaimed data science as "the sexiest job of the21st century". Some may pose the question of whether the statement still holds today. According to the U.S. BUREAU of Labor Statistics, employment in data science is projected to grow 36% from 2021 to 2031, much faster than the average for all occupations, which means employers will create more than 13,500 new data science related job opportunities each year on average, over the decade(Bureau of Labor Statistics 2022). Thus, we are curious about gender-based and race-based inequality in data science.

Introduction

The inequality that we are going to study relates to our quarter 1 project as Sahil studied racial discrimination in employment throughout that project and we are going to study racial and gender discrimination in a different field for the quarter 2 project. There has been previous work related to these problems (gender and racial bias in general employment) but not in a specific field job market. Our quarter 2 project is trying to audit race and gender inequality in data science related job employment, although Nancy's previous work for the Q1 project does not answer these employment bias problems, it focuses on auditing race and gender discrimination in the online housing market, and some methods still work. For example, the profile training with Seleniumis useful for our Q2 project, since we want to study race and gender inequality, we have to control variables. Besides, the ANOVA test method could still be performed within our quarter 2 project. Our quarter 2 project's problem is interesting to study because it will enlighten people on the discrimination prevalent in one of the up-and-coming fields in American employment. Our investigation addresses a massive deficiency in quarter 1 project which is the fact that we were focused on the general question in Q1 project whereas we will be studying a specific field in our Q2 project which should give us more useful information. The primary output of our Q2 project will be a report with the supplement output being a website. Our first step will be collecting as much as possible data and thencleaning it. We will then start to analyze the data through a descriptive analysis where we will mathematically describe and summarize the data that we found. Next, we will undergo diagnostic analysis where we will study the different rates of employment in

the data science field for different genders and races. Finally, we will use visual analysis to communicate our findings.

Methods

We started out our project by compiling all the data that we needed for our tests. We gathered this data by using a job recommendation website where we gathered data for 12 male subjects and 12 female subjects. For each gender, we gathered data within 4 different races (white, black, hispanic, asian). Some of the attributes that we gathered included the location of the recommended job, whether the job was remote, whether the job was full-time, how many employees were at this specific job, what type of field this job was in and what position was being offered. Some secondary attributes that we collected included what the main responsibilities of the job were for the employee and what the positional requirements were for the job.

The three attributes that we decided to focus on were what position was being offered (level of the job), how many employees were at the offered job and the type of job was being offered. The first two attributes would help us decipher what level of job was being offered to each race and each gender. Meanwhile, the type of job attribute would help us analyze the different jobs that are being offered to the different types of races and genders we are looking at.

Our main method that we used was an ANOVA test. We summed up the counts for each of the three attributes which were level, employees and type. We used a bar plot to represent the employee counts for each range of employees and pie plots to represent the level and type attributes.

Results

To be completed

Conclusion

To be completed

Appendix

To be completed

Contributions (so far)

Sahil: Proposal, Obtaining Data, Data Cleaning, Data Modeling

Nancy: Proposal, Obtaining Data, Data Interpretation, Report

References

Bureau of Labor Statistics, U.S. Department of Labor, *Occupational Outlook Handbook*, Data Scientists,

at https://www.bls.gov/ooh/math/data-scientists.htm (visited *November 28, 2022*).

Stier, H., & Yaish, M. (2014). Occupational segregation and gender inequality in job quality: a multi-level approach. Work, Employment and Society, 28(2), 225–246. https://doi.org/10.1177/0950017013510758