Exploring the Impact of Demographic and Personal Factors on Job Satisfaction Attitudes

Kirsten Fung, Angela Hong, Britney Hoang

DS 3001 : Machine Learning

Professor Johnson

February 29, 2024

I. Summary

Our central research question is: *What impacts do demographic and personal factors influence attitudes towards job satisfaction?* There are a myriad of factors that contribute to one's personal feelings towards their job. We are interested in exploring demographic factors and personal factors in order to evaluate the degree of impact each factor has on satisfaction. The findings of this research question could aid us in making more informed decisions about our career goals and paths, understand immutable factors, and investigate what factors impact one's happiness and attitude towards their given workplace. The aim of the research question is to understand the extent to which these demographic and personal factors influence an individual's attitudes towards job satisfaction.

For the initial exploration and cleaning stage, we first identified the variables from the GSS dataset that we wanted to test to measure job satisfaction. Demographic factors (i.e. age, sex, and race) and personal factors (i.e. income, work status, and stress levels) were established to be variables that we wanted to focus on as we believed they would be most correlated to a person's overall job satisfaction. We ensured that all these variables included data from 1973 through 2022 to allow graphs of trends over time. Additionally, for each variable, unknown or missing responses were replaced with NaN values and eventually dropped from the dataset to provide more clarity for the users interpreting the results. Further, since we were interpreting data that was spread across five decades we knew that having a large volume of unknown data would have limited our ability to make sound conclusions.

For the visualization process, we primarily wanted to identify whether any of our chosen factors influence a person's satisfaction with their job. First, we looked at job satisfaction over time, showing that there has not been a drastic change over the years. The response of "very satisfied" persisted at around 50% every year, while "very dissatisfied" was only about 5% every year. Heatmaps of job satisfaction against hours worked were also plotted, which were later broken down into our chosen demographic factors as well. The results showed that the majority of people are "very satisfied" with their 40 hour work weeks, however, unexpectedly, the percentage of dissatisfied employees did not grow as the hours increased. Looking at this data separated into sex and race, the heat maps showed that sex did not have a significant difference in satisfaction responses. Yet, race had a noticeable increase in job satisfaction for the individuals categorized as "other". To analyze our personal factors, a kernel density plot of job satisfaction vs. work status displayed that those with full-time and part-time jobs were most satisfied while the unemployed, looking for a job, laid off, and keeping house individuals, were unsurprisingly most unsatisfied with their job. Lastly, stress level and income were found to have no concrete relationship to job satisfaction. These results helped us draw a fundamental conclusion in regards to our research question.

II. Data

Initially, we examined the General Social Survey (GSS) dataset, which contains demographic information and attitudes on various topics for U.S respondents. We identified potential influential variables we thought would best fit our analysis. Our focus was on selecting variables with available data spanning from 1973 to 2022, allowing for a comprehensive scope for our investigation. With this in consideration, our chosen variables related to job satisfaction were grouped into demographic and personal aspects. The variables “year” and “hours worked” were general measures that we could pull analysis from. “Age”, “sex” and “race” were our demographic data pieces, while “occupation”, “highest degree”, “stress level”, “marital status”, “work status”, “number of children”, and “income” were our personal data pieces. Recognizing the significance of the year variable, we ensured its inclusion in all datasets, foreseeing its use for comparative assessments over time.

To import the data in Colab, we forked an existing repository that contained code to begin our process of querying through the GSS dataset. Non-essential columns were omitted from our analysis, with missing values in other columns resolved through exclusion. However, exceptions were made for the "hours_worked" variable, as we perceived its relevance to job satisfaction. We replaced occurrences of zero frequency or missing responses with NaN for specific variables such as "job_satisfaction," "sex," "degree," "stress," "marital," "race," and "income." Variables like "age" and "hours_worked" required minimal intervention as they were already well-maintained. All NaN values were dropped to allow for clarity during the visualization process. Additionally, columns were renamed to enhance readability and clarity. For example, “satjob”, “indus10”, and “hrs2” were renamed to “job_satisfaction”, “occupation”, and “hours_worked”, respectively. Lastly, our cleaned data was converted into a CSV file that was automatically imported into Colab and was uploaded to the GitHub repository.

A notable challenge arose with the "occupation" variable, prompting us to combine the array of diverse entries into 10 general occupation categories, including education, services, wholesalers, and healthcare. Additionally, the fact that we had to condense our options to only 13 variables was difficult as we were constrained by the availability in the dataset as well as wanting to ensure that our conclusions could be applied to the general population. Many other factors were considered but were not chosen in the end. Lastly, although we intentionally chose variables that spanned across 5 decades, the group discussed and ultimately decided that it would be best to concentrate only on the last decade of responses to ensure relevance of the analysis. These years still contained data for all of the necessary variables related to job satisfaction, allowing us to analyze the most cohesive and applicable information to best answer our research question. The primary focus of our analysis centered on the "job_satisfaction" variable, chosen for its potential to offer valuable insights into individuals' attitudes toward their work, a critical aspect of our research. This variable encapsulated respondents' satisfaction with their day-to-day work from a holistic perspective.

III. Results

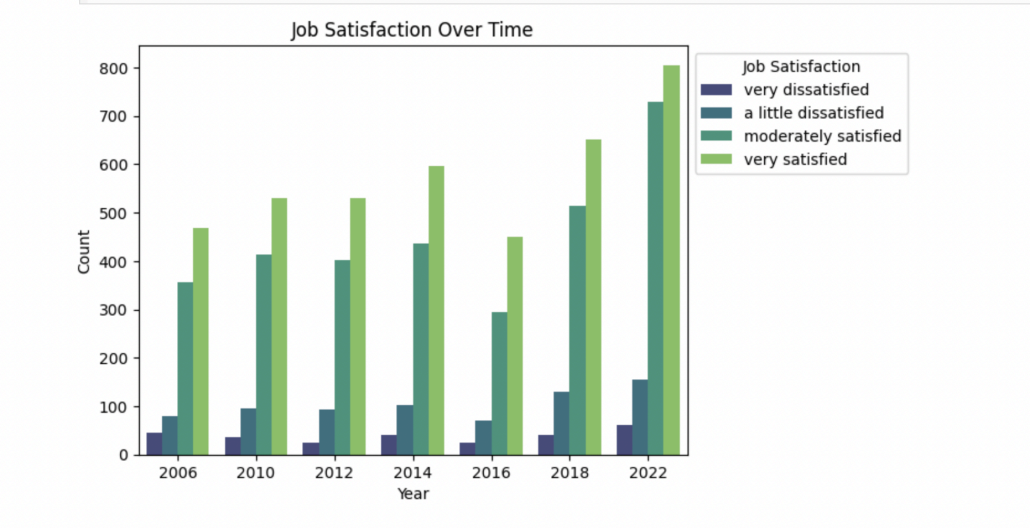

Figure 1: Countplot of Job Satisfaction vs. Year

Figure 1 depicts the relationship between job satisfaction and year. Noticeably, the figure ranges from the years 2006-2022. This decision to focus on this timeframe was deliberate as it allowed our group to concentrate on the most recent decade to further ensure this timeline's relevance in our analysis. Furthermore, a countplot was used for our initial visualization because it was beneficial to see the distribution of our categorical variable "job_satisfaction" to identify any patterns or trends within the frequency of our variable against time. Notably, in this visualization, observe that in 2016, there seems to be a dip in the number of responses followed by the increase in frequency in 2018. Additionally, the year 2022 recorded the highest number of valid respondents. While the data may imply an overall positive trend in job satisfaction with every year, it is important to note that we were merely looking at the number of valid respondents given our categorical data.

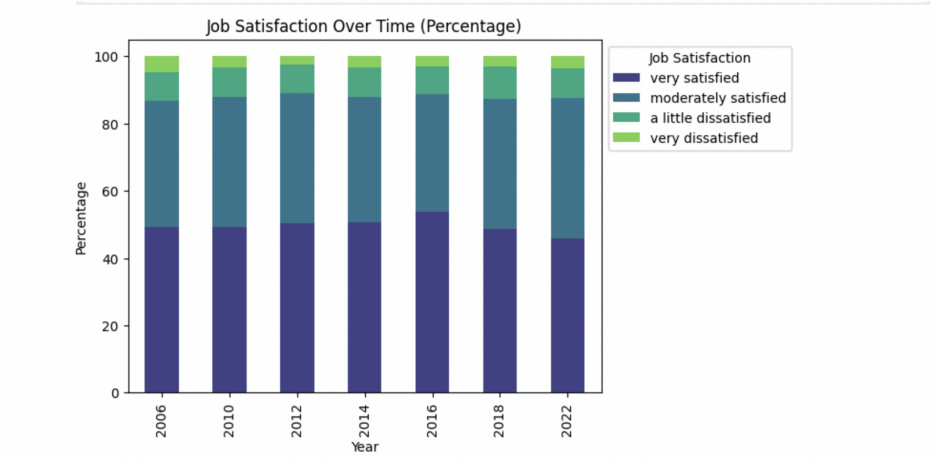

Figure 2: Stacked Bar Graph of Job Satisfaction

In Figure 2, a stacked bar graph based on percentages reveals a stable job satisfaction landscape from 2006 to 2022. Notably, around 50% of respondents consistently report being "very satisfied," with another substantial segment, approximately 40%, expressing "moderate satisfaction." "Very dissatisfied" consistently represents the lowest percentage. Additionally, 2016 exhibits a marginal increase of approximately 5% in "very satisfied" responses compared to other years, but overall, no discernible trend or substantial variation over time is evident. The data instills confidence in the observation that job satisfaction levels remain consistently positive, with minimal fluctuations, and that "very dissatisfied" responses consistently hold the lowest representation.

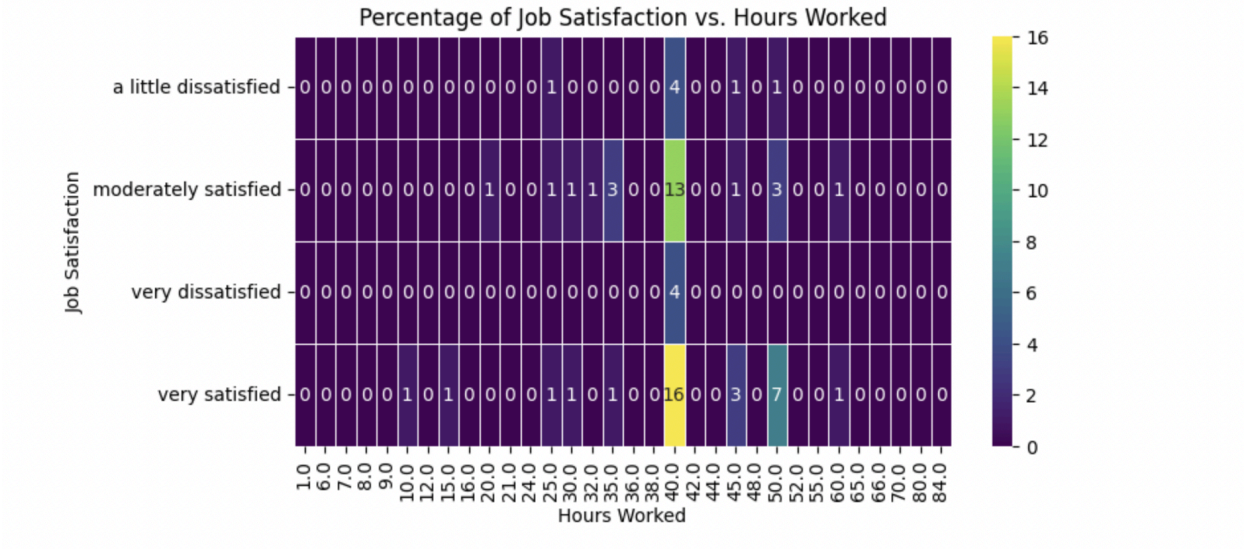

Figure 3: Heatmap of Overall job satisfaction vs. Hours Worked (percentage)

In Figure 3, a heatmap correlating overall job satisfaction with hours worked, presented in percentage terms, reveals an intriguing association. Surprisingly, individuals adhering to the standard 40-hour work week exhibit higher job satisfaction levels. Given that 40 hours is a common standard for most Americans, it is noteworthy that a considerable percentage of respondents fall within this interval. Specifically, 16% of participants express being "very satisfied," while 14% report "moderate satisfaction" with their occupations. This observation underscores the potential impact of standard working hours on job satisfaction levels.

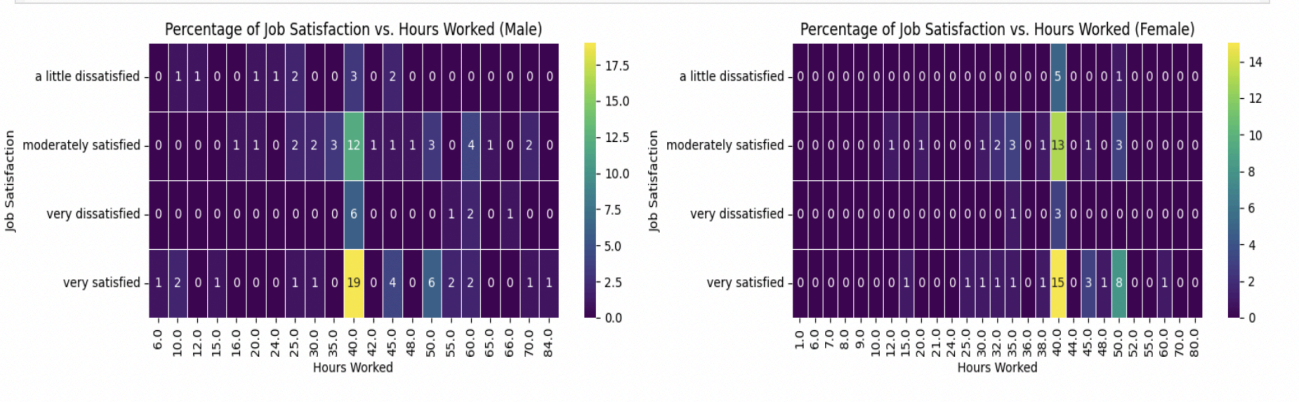

Figure 4: Heatmap of Overall Job Satisfaction vs. Hours Worked (percentage, separated by sex)

In order to further investigate the impact of hours worked on one's overall job satisfaction, we thought it would be helpful to split up the heat map by sex to visualize the results with male and females. When looking at the male heatmap, at a total of 40 hours worked, there were about 19% of men who stated being "very satisfied." On the other hand, there were only 15% of females who responded being "very satisfied." There is a very slight difference here of 4%, but very interesting to observe nonetheless. Next, 12% of males rated "moderately satisfied" at 40 hours a week while females rated 13%—there is not much difference here. Last, at 50 hours a week, 8% of females claimed "very satisfied", while males rated 6%. There are small little differences between the sexes when comparing hours worked with job satisfaction.

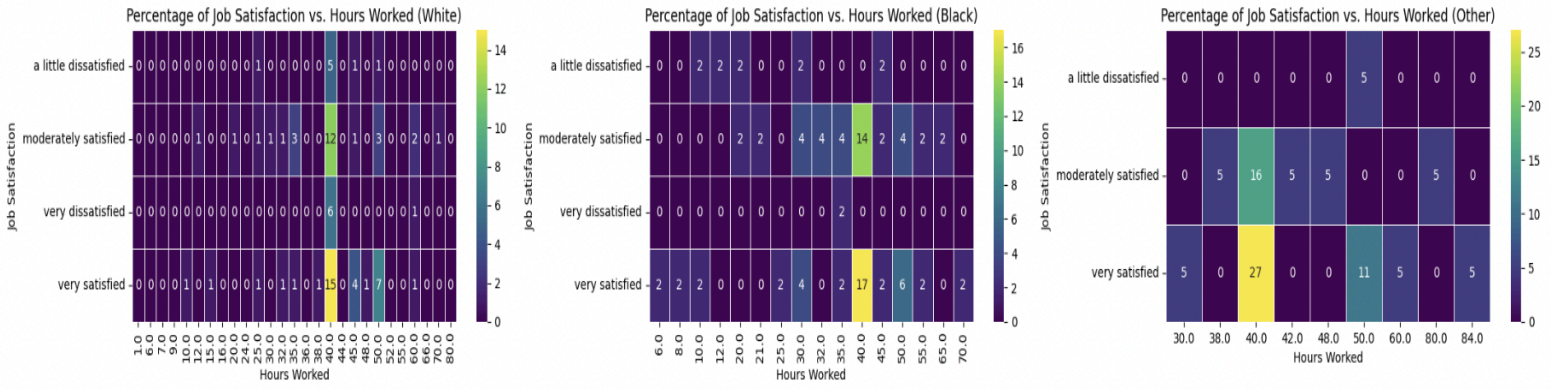

Figure 5: Heatmap of Job Satisfaction vs. Hours Worked (percentage, separated by race)

In Figure 5, an examination of job satisfaction in relation to hours worked is conducted, with a focus on racial demographics and expressed in percentage terms. The heatmap is structured to compare responses among white individuals, black individuals, and other racial groups. Notable findings include a comparable percentage of 15% for white and 17% for black respondents expressing "very satisfied" for 40 hours worked weekly. However, the "Other" racial group stands out with 27% indicating "very satisfied," showcasing a substantial difference

compared to the other two groups. This suggests that individuals categorized under "Other" (e.g., Asian, Latinx, Indigenous) report higher job satisfaction levels than their black and white counterparts. The racial dimension adds a nuanced layer to our understanding of how job satisfaction varies with hours worked.

Figure 6: Violin Plot of Job satisfaction vs. Hours Worked

For the job satisfaction responses of “moderately satisfied”, “very satisfied”, and “a little dissatisfied”, it almost follows a normal distribution. However, the “very dissatisfied” response looks different than the others. The range of hours worked for “very dissatisfied” is smaller than the rest and starts at around 20 hours a week. Based on this, it seems no one who is unemployed marked “very dissatisfied” as their response. Thus, it was only people who work part-time and/or full time who marked “very dissatisfied” with their job. Additionally, this category is the only one with two peaks (dense in more than one area) showing that a large number of people are more dissatisfied with 60 hour work weeks. Moreover, there also seems to be a wider range of hours worked for the individuals who responded with “a little dissatisfied.”

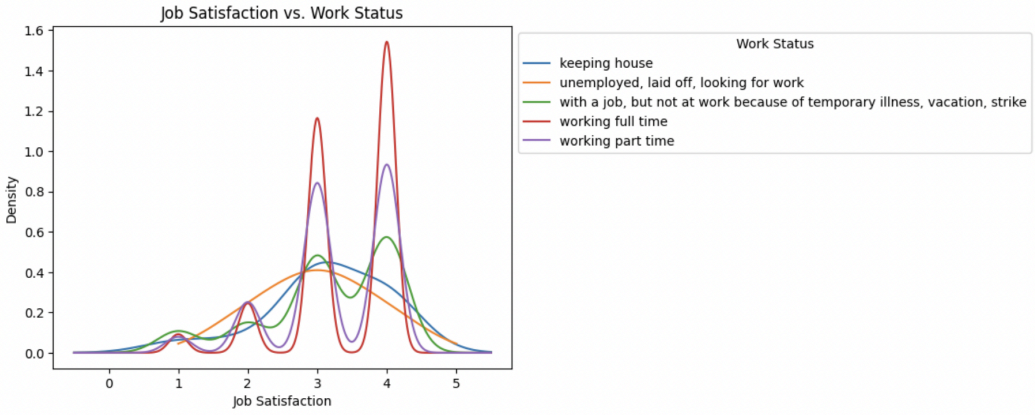

Figure 7: Kernel Density Plot of Job Satisfaction vs. Work Status

When observing the relationship between job satisfaction and work status, the first thing that sticks out are the two notable spikes at "3" and "4" for job satisfaction for people working full-time. Similarly, those working part-time also have two spikes at "3" and "4" for job satisfaction. So, the mean job satisfaction for people working full-time and part-time is expected to be around 3.5 (pretty satisfied). Next, people with a job but are taking a break off work have two, less obvious, spikes at the same places. For work status responses of "unemployed" and "keeping the house", they both vary across the satisfaction range and are more dense at the lower level. Thus, their mean job satisfaction responses are relatively lower than the others.

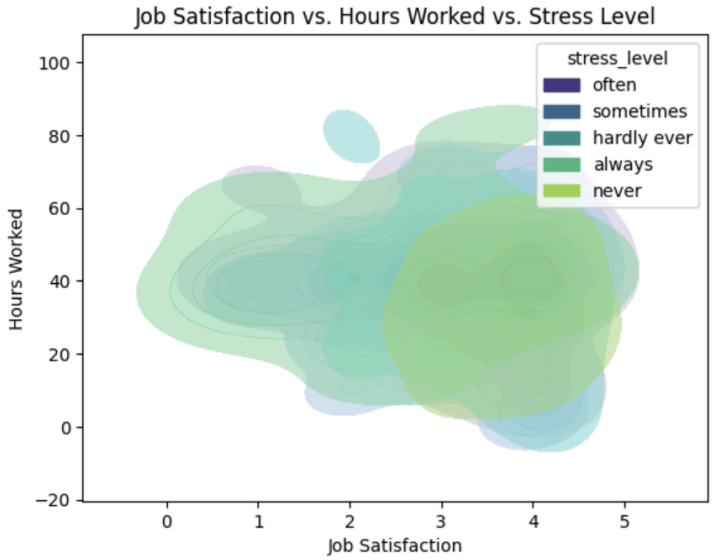

Figure 8: Kernel Density Plot of Job Satisfaction vs. Hours Worked Grouped by Stress Level

The kernel density plot observes the relationship between the respondent's stress level and job satisfaction. Notably, respondents who reported being "always" stressed exhibit diverse job satisfaction ratings, showcasing no discernible pattern or consistent trend. The variance in job satisfaction among those consistently stressed suggests individualized responses within this subgroup. Conversely, respondents indicating "never" experiencing stress tend to have higher job satisfaction ratings, averaging around 3.5/4. This aligns with the intuitive expectation that individuals with minimal stress in their occupations are more likely to rate their overall job satisfaction more positively.

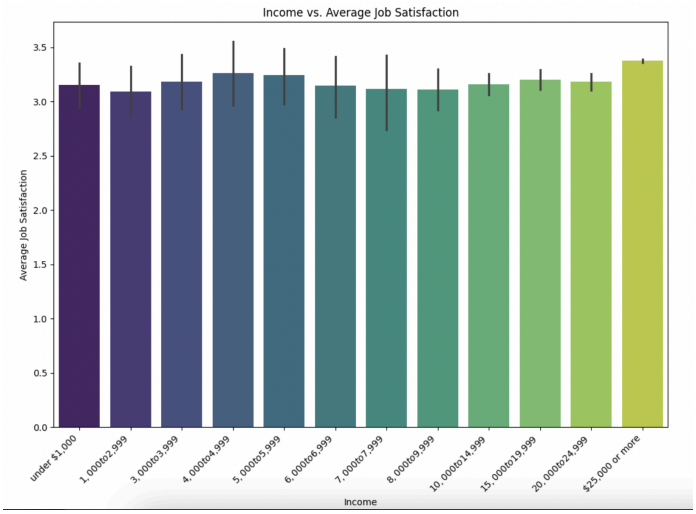

Figure 9: Bar Graph of Income vs. Average Job Satisfaction

Contrary to our initial assumption, all 12 columns present a similarity, lacking distinct variations. Across different income brackets, the average job satisfaction is around 3.4 for an income of $25,000 or more, while income ranges of $1,000 - $2,999 and $8,000 - $9,999 yield slightly lower averages at approximately 3.1. However, these marginal differences are insufficient to draw conclusive relationships. The graph does not support the notion that income serves as a reliable predictor of job satisfaction, indicating that variations in income have minimal impact on respondents' attitudes toward their workplace. Thus, we can conclude that our findings suggest an absence of a distinct relationship between income and job satisfaction in this context.

IV. Conclusion

In summary our analysis explored the relationship between job satisfaction and various factors using the GSS dataset, spanning from the years 1973-2022. The key variables and factors we researched were job satisfaction, demographic factors such as: race, sex, and age and personal factors such as: hours worked, occupation, income, stress level and marital status. After extensive data cleaning and preparing for our data visualizations, we found notable comparisons across groups that helped answer our research question.

Our data exploration began with a countplot highlighting variations in response counts over the years, further, a subsequent stacked bar graph based on percentages revealed consistent levels of "very satisfied" and "moderately satisfied" with no discernible trends. However, we found that the impact of hours worked on job satisfaction was evident in a heatmap, indicating that individuals adhering to the standard 40-hour work weeks tended to express higher job satisfaction, with significant percentages at 16% for "very satisfied" and 14% for "moderately satisfied." Gender differences were also observed in a heatmap separated by sex, with males showing slightly higher satisfaction percentages at specific work hours compared to females.

Racial disparities were evident as well, with the "other" racial category displaying higher satisfaction compared to white and black respondents. The relationship between job satisfaction and stress levels demonstrated no clear pattern for those "always" stressed, while respondents reporting "never" being stressed tended to have higher satisfaction scores. Surprisingly, income did not emerge as a significant predictor of job satisfaction, as indicated by a bar graph grouped by income brackets with consistent average satisfaction ratings across different income levels.

These visualizations provided valuable insights into job satisfaction dynamics. The countplot of job satisfaction over the years (Figure 1) highlighted trends in response counts, prompting the creation of a more insightful stacked bar graph based on percentages (Figure 2). Heatmaps further explored relationships, revealing surprising correlations, such as individuals working a standard 40-hour week expressing higher job satisfaction (Figure 3). The breakdowns by sex and race added depth to the analysis, uncovering differences in satisfaction levels. The exploration of income's impact on job satisfaction challenged preconceptions, demonstrating that income levels did not show a significant influence on job satisfaction (Figure 8). These nuanced findings contribute to a comprehensive understanding of the complex dynamics which we believe influences job satisfaction over time.

In this exploration of job satisfaction dynamics, our project has provided valuable insights, yet there are still limitations and additional outside work that could be done to the scope of this project. First and foremost, an analysis extending beyond our current scope, holds the potential to reveal trends in job satisfaction evolution over more extended periods. While our focus is primarily on quantitative data, the integration of qualitative insights through surveys or interviews could offer a more profound understanding, particularly in the respondents subjective experiences related to work environment, job security, opportunities, and what they deem to be their “work-life balance”. Moreover, discussing job-specific factors such as role responsibilities and organizational culture, could significantly enhance the granularity of our understanding. Although we carefully selected variables like work hours, demographic factors, and stress levels from the GSS dataset, it is also essential to acknowledge the inherent subjectivity in our choices. Variables we deemed influential, such as work status and race, might be open to interpretation and may not universally impact job satisfaction. While our selection aimed for generalization, there is room for acknowledging possible misconceptions and varied perspectives on these variables. Future research could benefit from delving into the nature of these chosen variables, considering the potential for divergent perceptions that might shape the overall landscape of the respondents job satisfaction.

V. Appendix

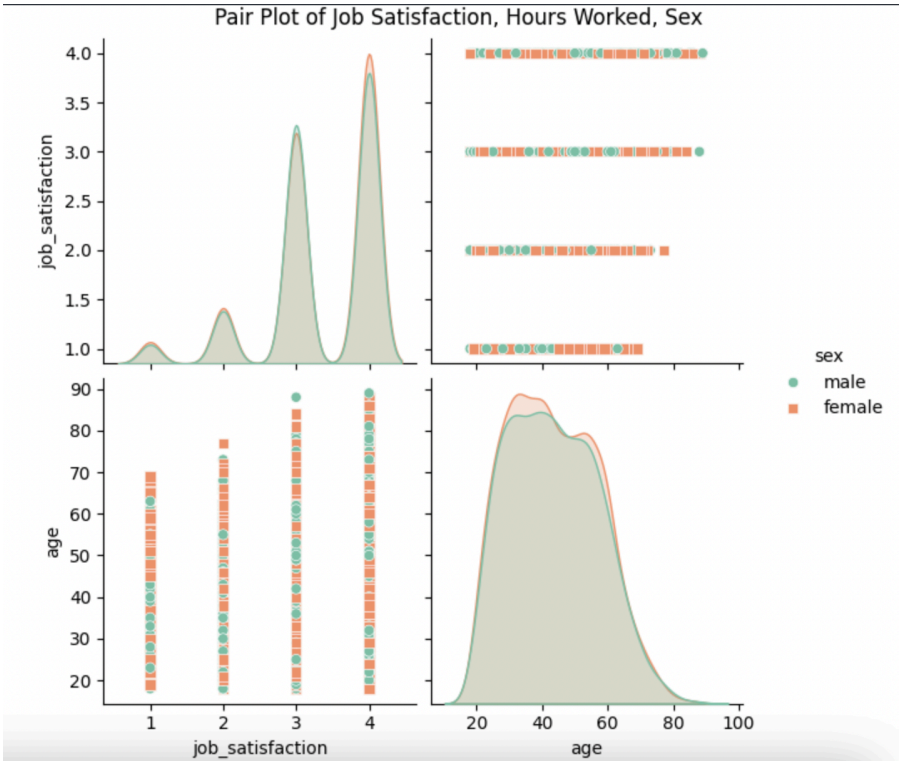

Figure A1: Pairplot of Job Satisfaction vs. Hours Worked (grouped by sex)