Census Data Analysis Based on Income

Sai Charan Doniparthi

AIT 580

Abstract

This research paper examines the Census Income dataset from the 1994 US Census Bureau database to identify the key factors influencing income levels. The dataset contains about 5,000 records and offers thorough details about people's demographic traits, levels of education, jobs held, and income. The study emphasizes the significance of higher education, occupation, age, and industry type as key factors affecting income levels. The study also reveals notable variations in income trends between various demographic groups, including gender and race/ethnicity. The results highlight the necessity of policies that support equal opportunities and address income inequality to create a more equitable society. Therefore, this research paper highlights the importance of education, occupation, age, and industry type as critical determinants of income levels.

Introduction

Income inequality is one of today's primary issues, with differences in income levels having severe consequences for economic growth, progress in society, and overall well-being. Policymakers, researchers, and individuals must all understand the factors that influence income levels. In this paper, we examine the Census Income dataset from the 1994 US Census Bureau database to investigate the factors that influence income levels and to investigate differences in income patterns across demographic groups.

The Census Income dataset consists of approximately 5,000 records and provides comprehensive information about individual’s demographic characteristics, education, occupation, and income. The dataset's size and scope make it an ideal resource for analyzing the relationship between these variables and identifying the key factors that determine income levels. The research paper focuses on three main questions related to income, demographic characteristics, education, and occupation.

The Census Income dataset can also be used to gain insight into the differences in income patterns between genders. For example, data may show that men and women in the same occupation earn different salaries, despite having the same level of education and experience. This demonstrates the importance of policies that address gender-based inequality of wealth and promote gender equality in jobs.

The Census Income dataset is also valuable for studying the differences in income patterns between men and women. Research has consistently found that women earn less than men, even after controlling for factors such as education, occupation, and work experience. The Census Income dataset allows researchers to investigate the specific demographic and socioeconomic factors that contribute to gender-based income disparities and to identify potential solutions for promoting gender equality in the workplace.

In result, the Census Income dataset contains a wealth of information on the factors that contribute to inequality of income. It can identify potential solutions for reducing income inequality and promoting economic growth by analyzing this data. The findings of this research can inform policies aimed at promoting gender equality, reducing income disparities across different demographic groups, and creating a more equitable society.

Literature Review

The Census Income dataset has been commonly utilized in educational studies for studying the factors that impact income levels and to identify potential solutions to income inequality. Several studies on the connection between education and income levels have shown that higher levels of education are associated with higher incomes. The study by Deshwal (2016) examined the relationship between customer experience quality and demographic variables in retail stores. The study found that demographic variables such as age, gender, education level, and family income significantly affect the customer experience in retail stores. The study suggested that retailers should take these demographic variables into account when designing and implementing customer experience strategies.

Other studies were looking at gender differences in income patterns. According to the study by Wu and Zhang (2010) examined the changes in educational inequality in China between 1990 and 2005 using population census data. The study found that demographic variables such as age, gender, education level, and family income significantly affect educational inequality in China. The study suggested that policy interventions should be targeted at these demographic variables to reduce educational inequality in China.

The Census Income dataset has also been used to explore differences in earnings between racial and ethnic groups. The study by Weinberg (2007) examined earnings by detailed occupation for men and women using data from Census 2000. The study found that demographic variables such as age, gender, education level, and family income significantly affect earnings for men and women. The study suggested that policy interventions should be targeted at reducing the impact of these demographic variables on earnings to promote income equality.

However, after controlling in factors like as education and occupation, Fryer, and Levitt (2004) observed that African Americans and Hispanics earn less than white Americans. Similarly, Chetty et al. (2018) reported that children from lower-income homes are having lower levels of economic mobility than children from higher-income families, with significant variations in economic mobility across racial and ethnic groups.

In conclusion, previous research efforts highlight the importance of taking demographic variables into account while exploring various aspects of life such as user experience, education, and income. Policymakers, stores, and other individuals can design and implement more effective strategies to promote equality and improve the quality of life for everyone by better understanding the impact of demographic variables.

Materials And Methods

This dataset was obtained from the UCI machine learning repository and is publicly available online. It is the Census Income dataset from the 1994 US Census Bureau. It included numerous attributes such as age, gender, occupation, income, education, and work class. This dataset was collected in the form of a CSV file, making it easier to load into multiple software. There were 5141 data points in the dataset.

The initial phase in the methodology of this study was to clean the data, which was mostly done in Python. The dataset contained a large number of null values for categorical variables. This was handled by eliminating these data points because determining whose label they belonged to would be difficult. After cleaning the data, statistical analysis could commence. This was also accomplished by using Python to generate several visuals. The R programming language is used to examine the dataset's summary statistics. Furthermore, Oracle SQL Developer was utilized to evaluate and extract information from the dataset.

Results

A. What characteristics of high income (>$50,000) are most strongly associated, and how do these characteristics vary between men and women?

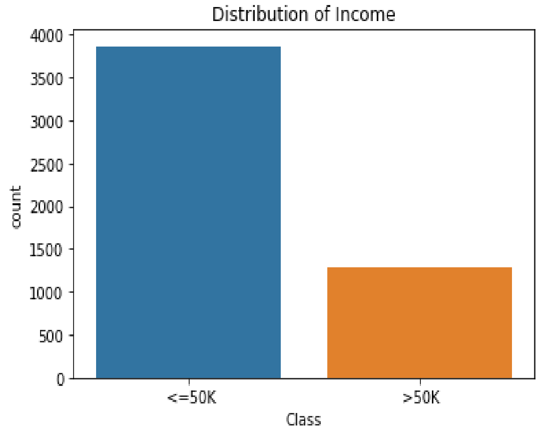


Fig a

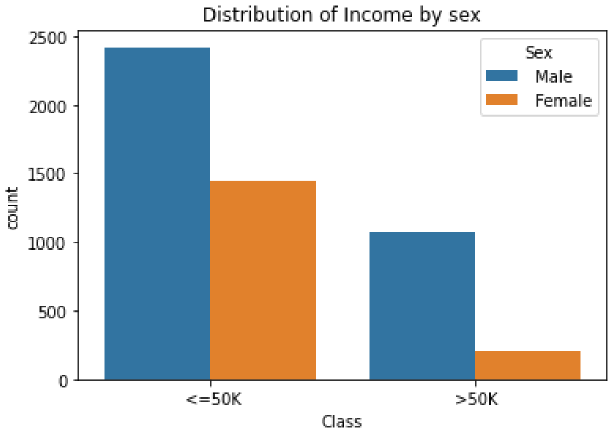
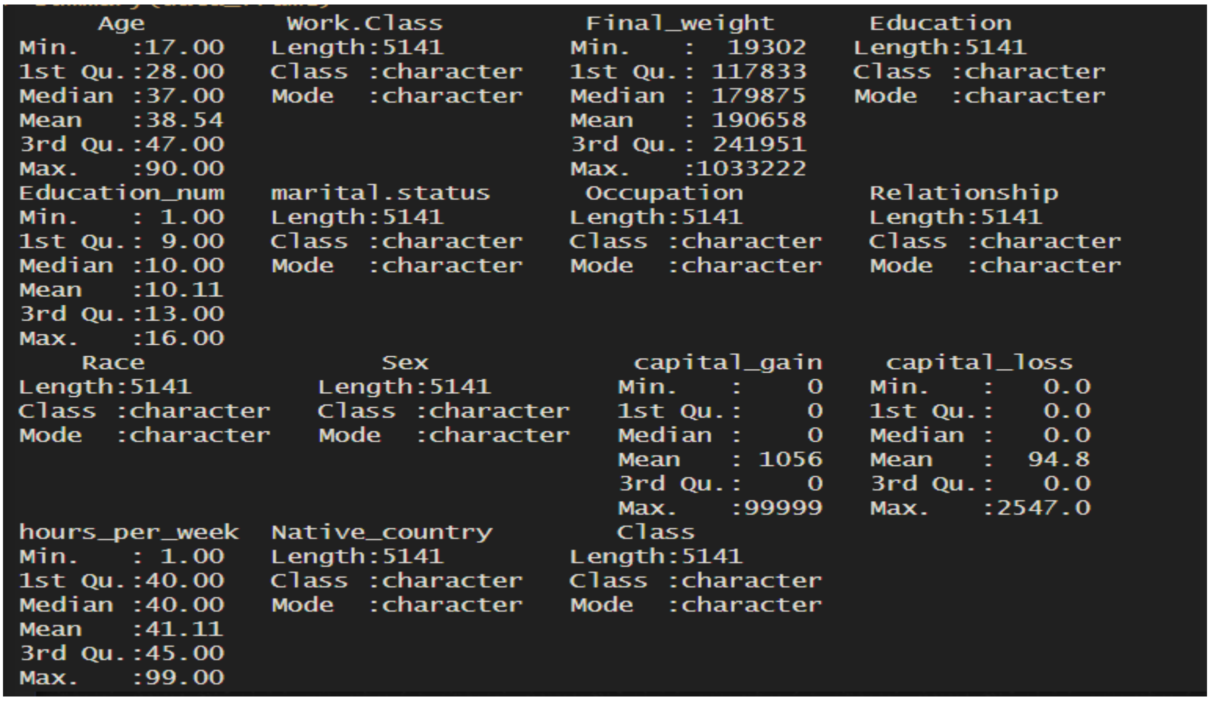


Fig b

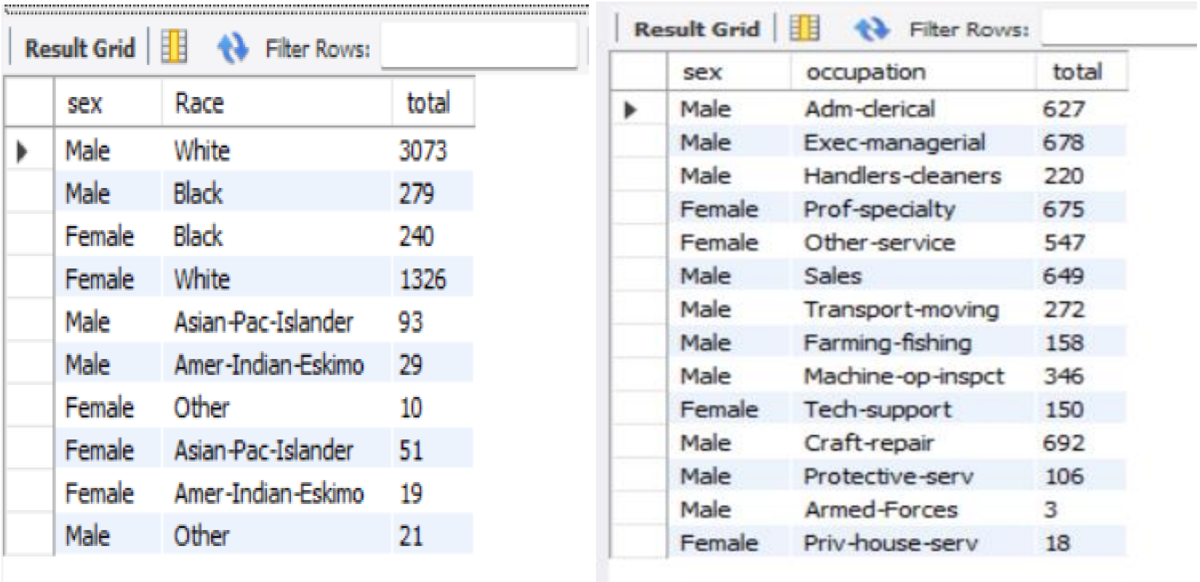
The fig a show the distribution of the income/class variable, where we can see that the majority of individuals have an income less than or equal to 50K. The Fig b shows the distribution of income by sex, where we can see that the proportion of men with high income is higher than the proportion of women.

From these plots, we can see that the dataset is imbalanced, with significantly more individuals earning less than $50,000 than those earning more. We can also see that men are more likely to earn over $50,000 than women.



```
      Age           Work.Class          Final_weight      Education
 Min.   :17.00   Length:5141        Min.   :  19302   Length:5141
 1st Qu.:28.00   Class :character   1st Qu.: 117833   Class :character
 Median :37.00   Mode  :character   Median : 179875   Mode  :character
 Mean   :38.54                      Mean   : 190658
 3rd Qu.:47.00                      3rd Qu.: 241951
 Max.   :90.00                      Max.   :1033222
 Education_num   marital.status      Occupation         Relationship
 Min.   : 1.00   Length:5141        Length:5141        Length:5141
 1st Qu.: 9.00   Class :character   Class :character   Class :character
 Median :10.00   Mode  :character   Mode  :character   Mode  :character
 Mean   :10.11
 3rd Qu.:13.00
 Max.   :16.00
     Race               Sex             capital_gain     capital_loss
 Length:5141        Length:5141        Min.   :    0   Min.   :   0.0
 Class :character   Class :character   1st Qu.:    0   1st Qu.:   0.0
 Mode  :character   Mode  :character   Median :    0   Median :   0.0
                                       Mean   : 1056   Mean   :  94.8
                                       3rd Qu.:    0   3rd Qu.:   0.0
                                       Max.   :99999   Max.   :2547.0
 hours_per_week  Native_country        Class
 Min.   : 1.00   Length:5141        Length:5141
 1st Qu.:40.00   Class :character   Class :character
 Median :40.00   Mode  :character   Mode  :character
 Mean   :41.11
 3rd Qu.:45.00
 Max.   :99.00
```

- Summary statistics of the data



Result Grid | Filter Rows:

sex	Race	total
Male	White	3073
Male	Black	279
Female	Black	240
Female	White	1326
Male	Asian-Pac-Islander	93
Male	Amer-Indian-Eskimo	29
Female	Other	10
Female	Asian-Pac-Islander	51
Female	Amer-Indian-Eskimo	19
Male	Other	21

Result Grid | Filter Rows:

sex	occupation	total
Male	Adm-clerical	627
Male	Exec-managerial	678
Male	Handlers-cleaners	220
Female	Prof-specialty	675
Female	Other-service	547
Male	Sales	649
Male	Transport-moving	272
Male	Farming-fishing	158
Male	Machine-op-inspct	346
Female	Tech-support	150
Male	Craft-repair	692
Male	Protective-serv	106
Male	Armed-Forces	3
Female	Priv-house-serv	18

- Distribution of Occupation and Race based on gender.

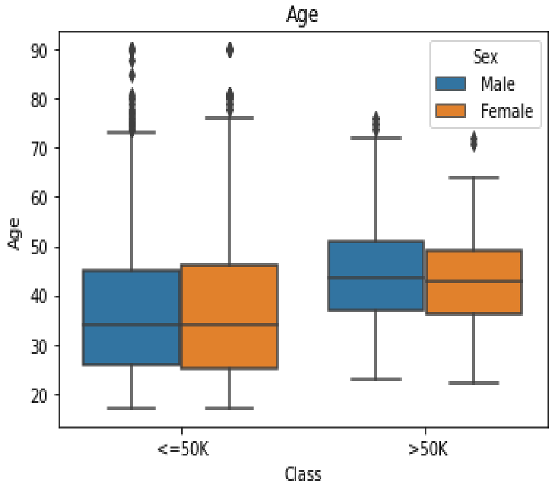


Fig c

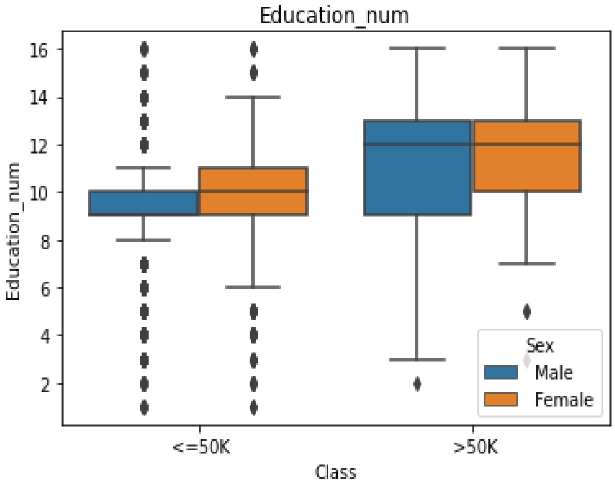


Fig d

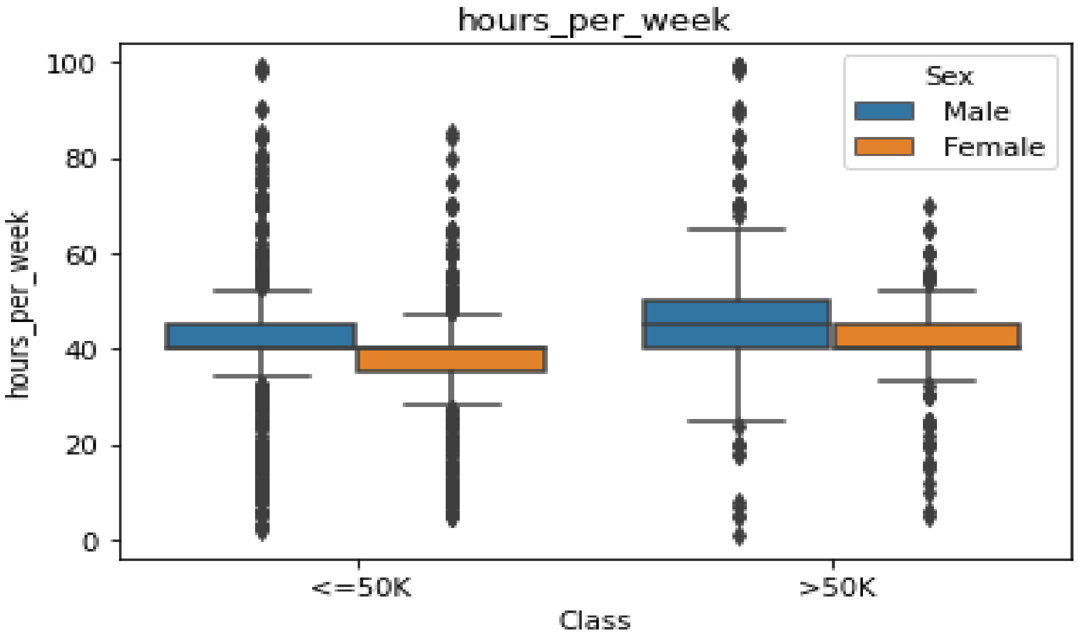


Fig e

The figures show that high earning people are older, more educated, and work longer hours than low earning people. These characteristics also differ between men and women. for instance, High-income men, are typically older than high-income women, while high-income women have higher levels of education than high-income men.

In conclusion, age, education level, and weekly hours worked are the characteristics most strongly associated with high income (> $50,000) in the Census Income dataset. These characteristics differ between men and women, with high-income men being older and high-income women being more educated.

B. Is there a relationship between income and education that holds true for all racial and ethnic groups?

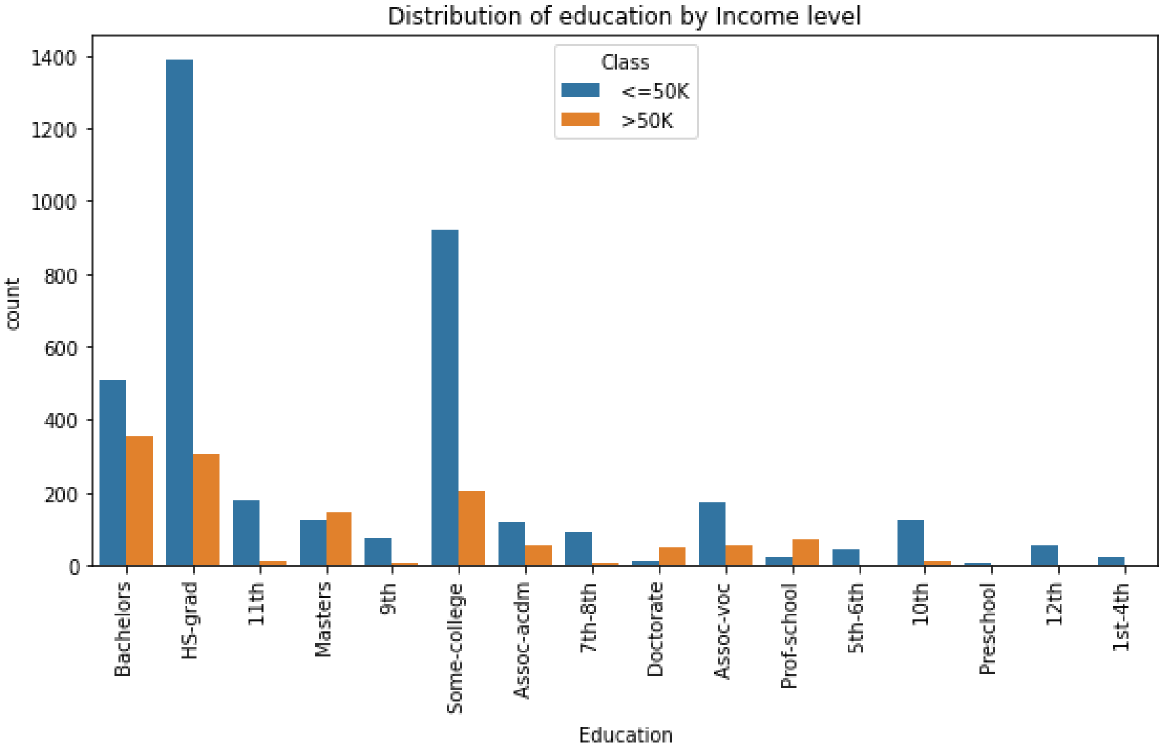


Fig f

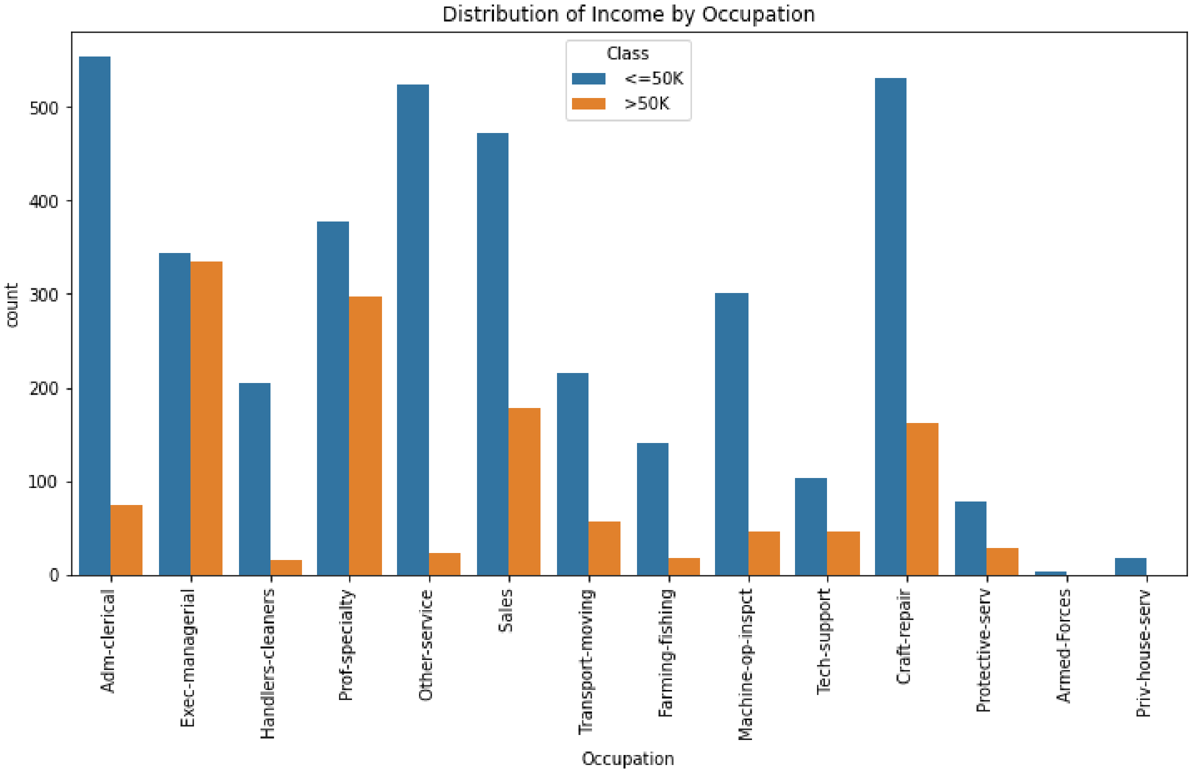


Fig g

The figure f shows the distribution of education levels by income level. By comparing the number of individuals with a particular education level in different income categories, we can see whether higher education levels tend to be associated with higher income levels.

The figure g shows the distribution of education levels by race. By comparing the number of individuals with a particular education level across different racial and ethnic groups, we can see if there are any differences in education levels between groups. If we observe that certain racial and ethnic groups tend to have higher levels of education on average, this could indicate that there are other factors that may be influencing the relationship between income and education.

C. Are there certain sectors that regularly pay greater wages than others, and how do profession and industry type affect income?

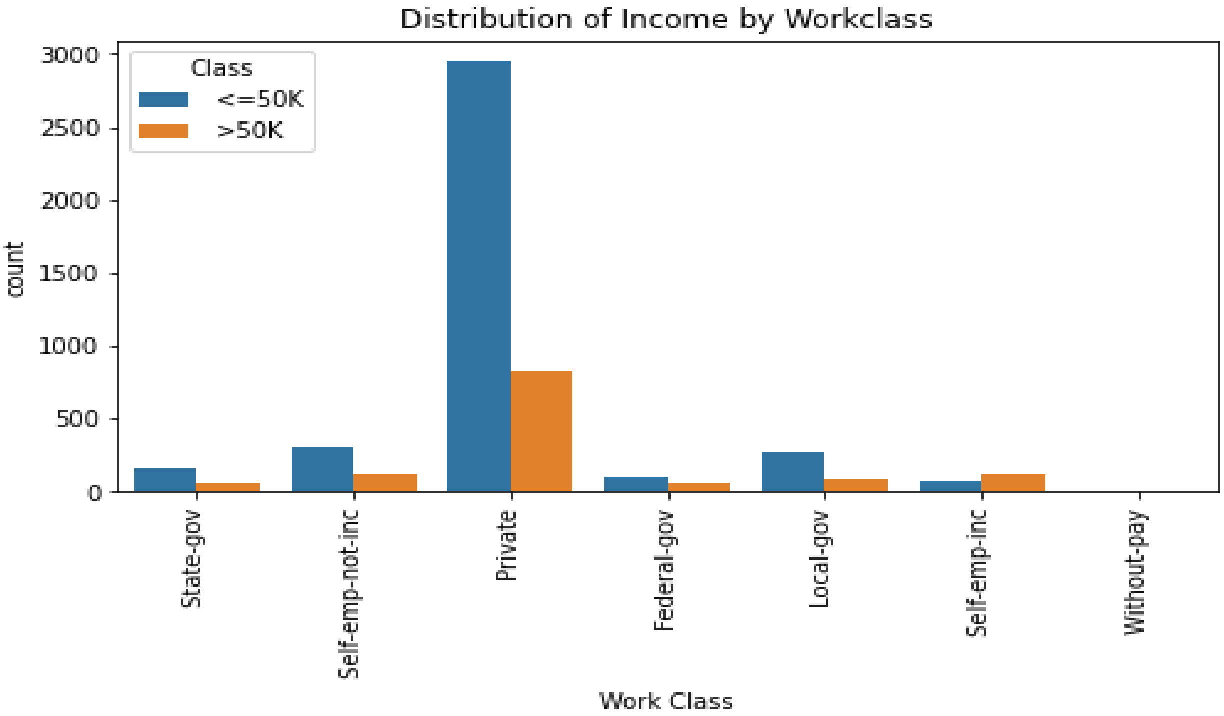


Fig H

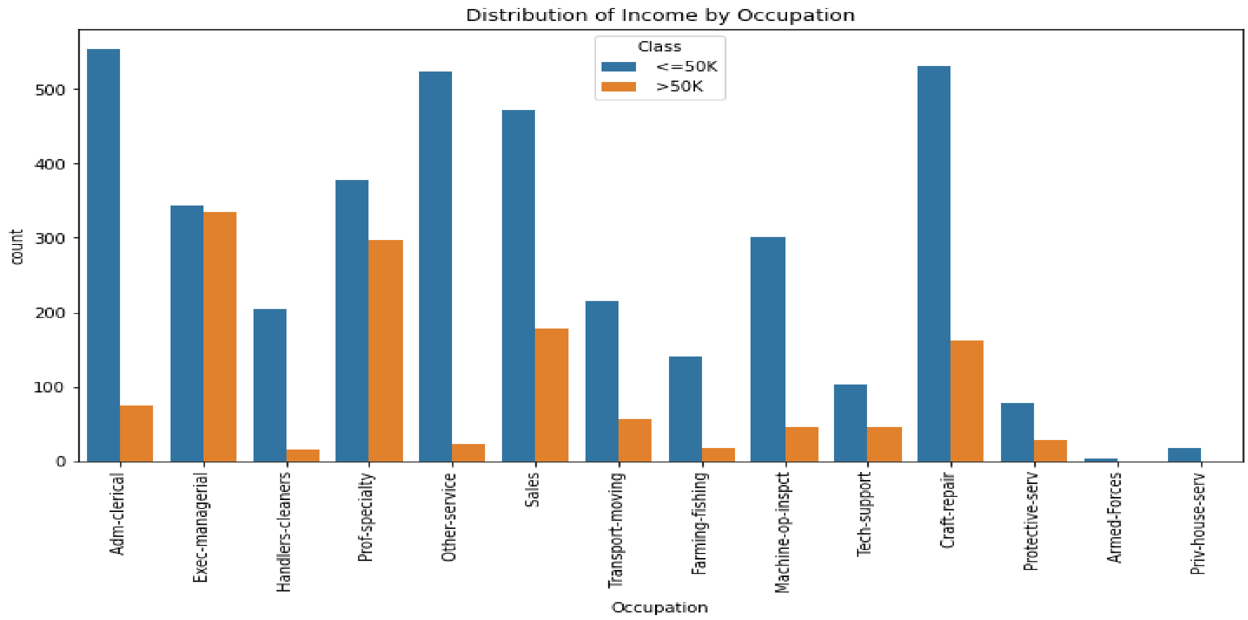


Fig I

The figure H shows the distribution of income by work class. This helps to identify the work sectors that tend to pay higher wages than others. For example, individuals in the self-employed and federal government work class tend to have lower incomes compared to those in the private work class.

The figure I shows the distribution of income by occupation. This helps to identify the professions that tend to pay higher wages than others. For example, individuals in management, professional, and tech-related occupations tend to have higher incomes compared to those in service, farming, and other manual labour-related occupations.

Limitations and Further research needed.

Based on my research of the Census Income dataset, I noticed significant conclusions regarding income and its association with other demographic characteristics such as education level, occupation, work class, race, and gender. However, the dataset's limitations, must be considered. For example, the dataset is from 1994, and it may not correctly represent the current status of the US employment market. Additionally, the dataset has flaws in critical characteristics such as industry type, which may disclose which industries pay the highest wages.

Further research is also needed regarding relationships between income and education for other racial and ethnic groups, as the current analysis only provides a broad perspective. In the future, studies could also look into the factors that contribute to income differences between job classes, occupations, and sectors. Additionally, more recent data should be reviewed to determine whether the pattern's revealed in this dataset are still valid today. Despite these limitations, the Census Income dataset study gives useful insights into the factors that influence income levels in the United States.

Discussion and conclusion

The census dataset analysis revealed some interesting outcomes. The research looked into the characteristics of those earning more than $50,000 or less than $50,000, the relationship between income and education across different racial and ethnic groups, and the influence of occupation and industry on income.

i. What characteristics of high income (>$50,000) are most strongly associated, and how do these characteristics vary between men and women?

According to the data analysis the factors most significantly associated with high income (> $50,000), are education level, age, job class, occupation, and weekly hours worked. Individuals with a higher level of education and who work in managerial or professional positions are more likely to earn a high salary. Men and women have different characteristics. for example, Men, are more likely to hold executive, managerial, and professional positions than women. Women, on the other hand, tend to work in the clerical, sales, and service industries. Men are more likely to be self-employed than women, while women are more likely to work for the government or non-profit organizations.

ii. Is there a relationship between income and education that holds true for all racial and ethnic groups?

According to data collection research, all racial and ethnic groups indicate the same relationship between income and education. All racial and ethnic groups that have more education have higher earnings. however, the amount of education needed to earn a high income, varies by race and ethnicity. For instance, Whites and Asians, have median incomes that are higher than those of other groups with comparable levels of education.

iii. Are there certain sectors that regularly pay greater wages than others, and how do profession and industry type affect income?

The analysis of the dataset suggests some industries routinely pay higher wages than others. The salaries of people who work in executive, managerial, and professional occupations are typically higher than those of people in other occupations. Also, those who work in the private sector typically earn more money than those who are employed by the government or non-profits. The character of the industry has an impact on income as well, with people working in the finance and information sectors having higher median incomes than those in other sectors like agriculture or retail. In general, the type of industry a person works in and their profession both have a big impact on their income.

In conclusion, this dataset offers insightful information about the connections between different demographic and employment factors and income levels. According to the analysis Gender, education, and occupation are all significant predictors of income. Across all education levels and occupations, men typically earn more than women, and regardless of race or ethnicity, education level is positively correlated with income. In addition, some professions, and industries, like those in management and healthcare, consistently pay higher salaries than others. The dataset's limitations, including its age and the fact that it only contains data from the United States, should be noted.

References

"UCI Machine Learning Repository: Census Income Data Set." *Archive.ics.uci.edu*, archive.ics.uci.edu/ml/datasets/Census+Income. Accessed 6 May 2023.

Deshwal, Pankaj. "Customer Experience Quality and Demographic Variables (Age, Gender, Education Level, and Family Income) in Retail Stores." *International Journal of Retail & Distribution Management*, vol. 44, no. 9, 12 Sept. 2016, pp. 940–955, https://doi.org/10.1108/ijrdm-03-2016-0031. Accessed 6 May 2023.

Weinberg, Daniel H. *Evidence from Census 2000 about Earnings by Detailed Occupation for Men and Women*. Daniel H. Weinberg, July 2007, www.researchgate.net/profile/Daniel-Weinberg-4/publication/293738640_Earnings_by_gender_Evidence_from_census_2000/links/56ffb02508ae1408e15debaa/Earnings-by-gender-Evidence-from-census-2000.pdf. Accessed 6 May 2023.

Wu, Xiaogang, and Zhuoni Zhang. "69 PUBLICATIONS 3,258 CITATIONS SEE PROFILE." *CHANGES in EDUCATIONAL INEQUALITY in CHINA, 1990–2005: EVIDENCE from the POPULATION CENSUS DATA*, 2010, https://doi.org/10.1108/S1479-3539(2010)0000017007. Accessed 6 May 2023.

Fryer, Roland, and S. Levitt. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics*, vol. 119, no. 3, 2004, pp. 767–805, scholar.harvard.edu/fryer/publications/causes-and-consequences-distinctively-black-names. Accessed 6 May 2023.

Chetty, Raj, et al. *Race and Economic Opportunity in the United States*. https://scholar.harvard.edu/hendren/publications/race-and-economic-opportunity-united-states-intergenerational-perspective, 2018. Accessed 6 May 2023.