RELATIONSHIP BETWEEN PERSONAL CHARACTERISTICS AND INCOME

Made by:
Lucas William
December 19, 2024

TABLE OF CONTENTS

Objectives 3

Methodology 4

Results 4

Conclusion 10

Sources and References 10

Abstract

The following work relates personal characteristics, such as skin color, gender, age, formal professional qualifications, work experience, etc., to the remuneration offered, measuring the influence and impact of each characteristic. The process will analyze the American population of the 1990s, where we will be able to assess, at least to some extent, the lifestyle and cultural characteristics of that people.

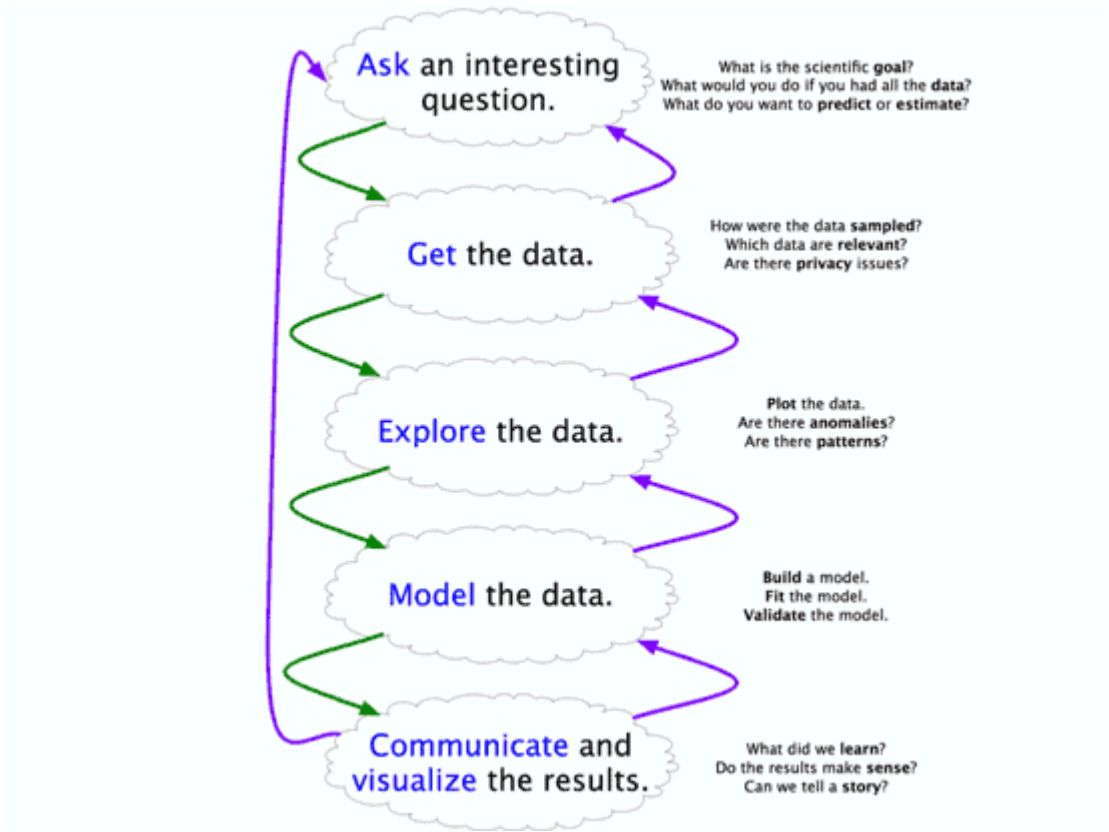
I will carry out the ETL process with a dataset obtained from the Kaggle platform, called Adult Census Income (1994), sourced from the U.S. Census Bureau, the main statistical agency of the U.S. government. The data will be loaded into an Excel spreadsheet, where we will proceed to the cleaning phase. In this phase, we will remove all irrelevant data for the analysis, as well as correct any inconsistencies and handle missing information. Finally, we will use the Business Intelligence tool, Looker Studio, to visualize the data, gaining important insights for decision-making regarding, for example, public policies aimed at closing the wage gap between men and women or other social considerations.

Objectives

Determine the influence of personal characteristics on the salaries of American professionals in the 1990s, seeking patterns and trends that allow conclusions about the nature of American society at the time.

Methodology

We used the workflow outlined by Blitzstein & Pfister, Harvard professors, in the context of data science and data analysis, as illustrated in the diagram below.



1. Question: What is the influence of personal characteristics on salary? Is there any pattern? Any interesting relationships?
2. Data: Adult Census Income (1994), from the U.S. Census Bureau agency;
3. Exploration: Use of Excel tools such as Pivot Tables, Pivot Charts, filters, What-if analysis, and different types of sorting;

4. Data modeling will be done through tables, in the relational sense, where each column represents a characteristic of a person (name, age, gender, etc.), and each row represents a person;

5. We will use Google's Looker Studio tool to build an interactive visualization that tells a story about the data and communicates patterns, trends, and anomalies.

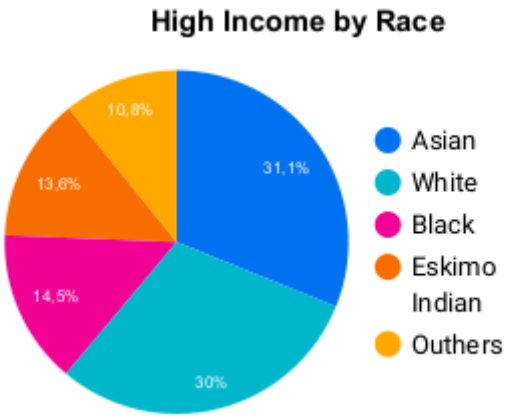
For the sake of simplifying visualizations, we consider an individual with a high income to earn $100,000 or more per year (in 1994 adjusted values).

Results

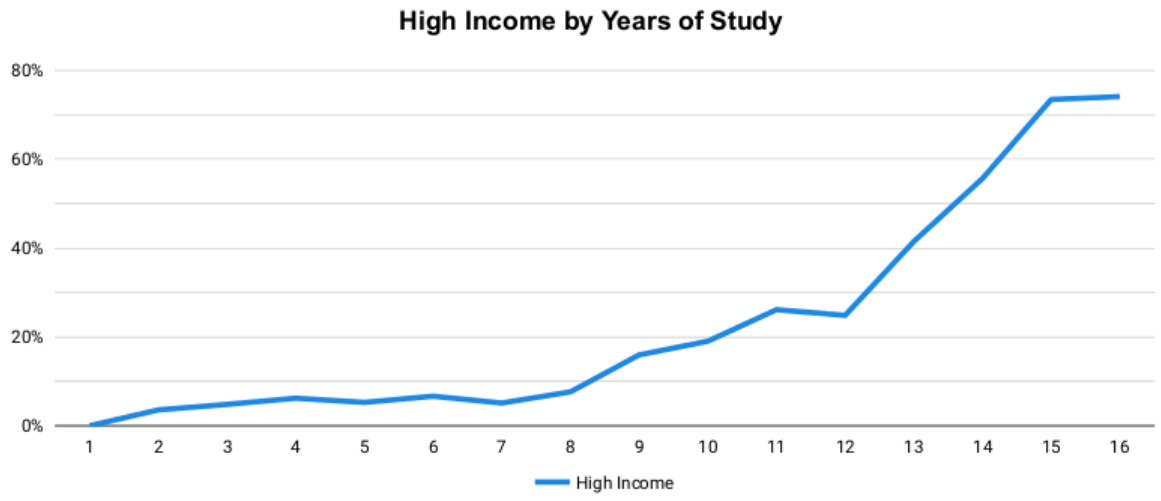
In this section, we will discuss the insights obtained and the conclusions we can draw from the data analyzed.

1. Trends and Anomalies

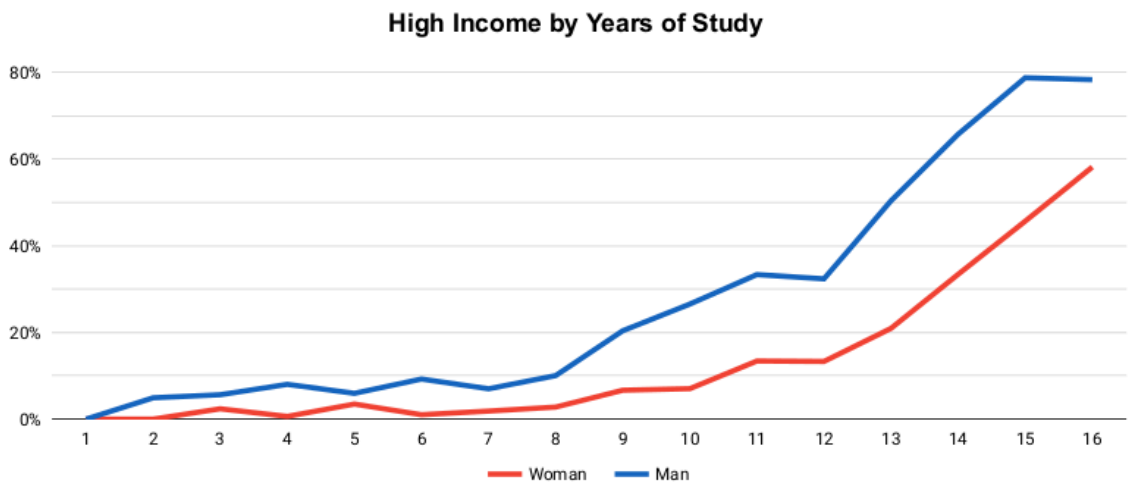
There are five characteristics analyzed in this study: years of education, gender, work class, work sector, and ethnicity. We discovered some interesting trends, such as the fact that Asians tend to fall into high-income groups more frequently than any other ethnicity, including Whites, who rank second, as shown in the following chart:



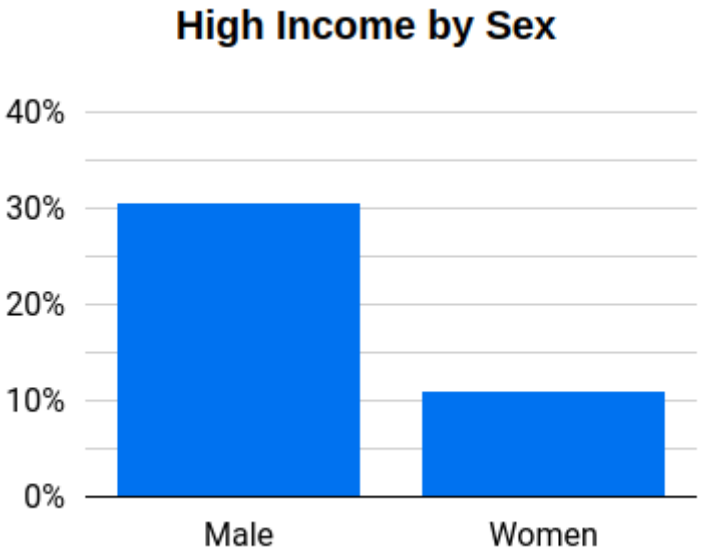
Another strong trend is the relationship between education and high income, which has a correlation coefficient of 0.895, indicating a very significant influence between years of education and high income. This can be observed in the behavior of the following chart, which at certain points follows a pattern extremely close to linear:



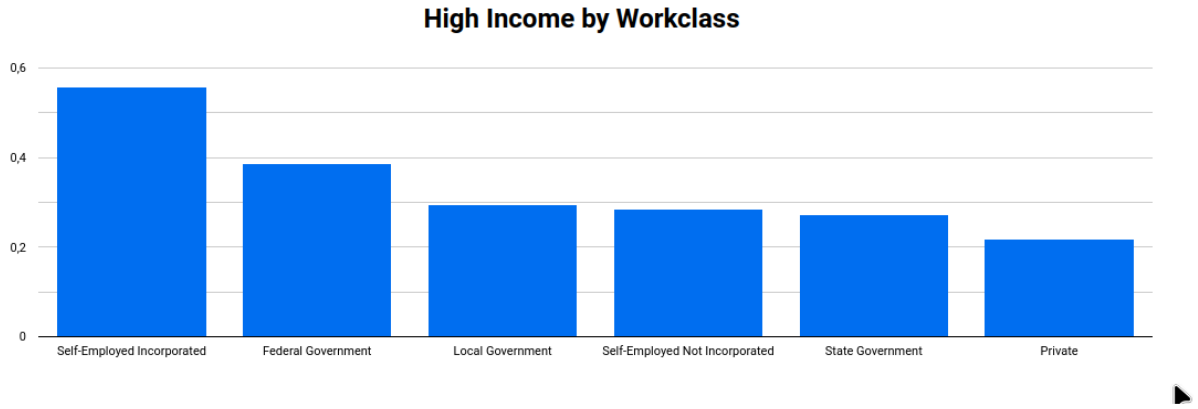
The gender issue is also interesting, as we can see that high-income men are more frequent than high-income women. In fact, the collected sample indicates that there are three times more high-income men than women. Furthermore, when comparing both the gender variable and years of education, we can see that the chance of entering the high-income group increases for both groups. However, male education has a much more significant impact on income than female education, as can be seen in the following chart:



Another chart now shows the relative difference between high-income men and women:



Finally, we would like to highlight the very significant role of work class and position in reaching high income. We find that the most specialized professionals and those in management positions are the highest paid. However, it is interesting to note that the private sector has the lowest number of high-income professionals, while self-employed professionals, i.e., those who offer services through a third-party intermediary (freelancers), have the highest income. Public sector jobs are in the average range, with federal employees being the best paid. The following charts and tables illustrate the situation.



	Occupation	Earn More Than 50 Thousand Dollars
1.	Executive Manager	48,4%
2.	Specialty Worker	44,9%
3.	Protection Services	32,51%
4.	Technical Support	30,5%
5.	Seller	26,93%
6.	Construction and Repair	22,66%
7.	Transport Sector	20,04%
8.	Administrative Work	13,45%
9.	Inspector or operator of Machines	12,49%
10.	Farmer or fisherman	11,57%
11.	Army	11,11%
12.	Trabalhos de Limpeza ou Manuaiss	6,28%
13.	Other Services	4,16%
14.	Private Home Services	0,67%

We can clearly see that the best jobs are those that require a solid education, such as management, technical support, or specialized professionals, while manual labor, known as 'entry-level' jobs, has the lowest average remuneration.

2. Conclusions

We can deduce many interesting results from this data, such as the importance of professional specialization for better financial outcomes in the medium and long term, the importance of discussing gender inequality in the salary scope, the relationship between ethnicity and salary, among other cultural points, such as the relationship between the stereotype of Asians' work ethic and their higher participation in the high-income group. Therefore, it is possible to observe that a statistical survey and data analysis, combined with a digestible presentation of the key points, is essential for achieving good results in decision-making, in any field or era.

3. Sources and References

Kaggle Dataset: https://www.kaggle.com/datasets/uciml/adult-census-income

GitHub: https://github.com/Lucas-Mazzolim/Excel-Projects.git