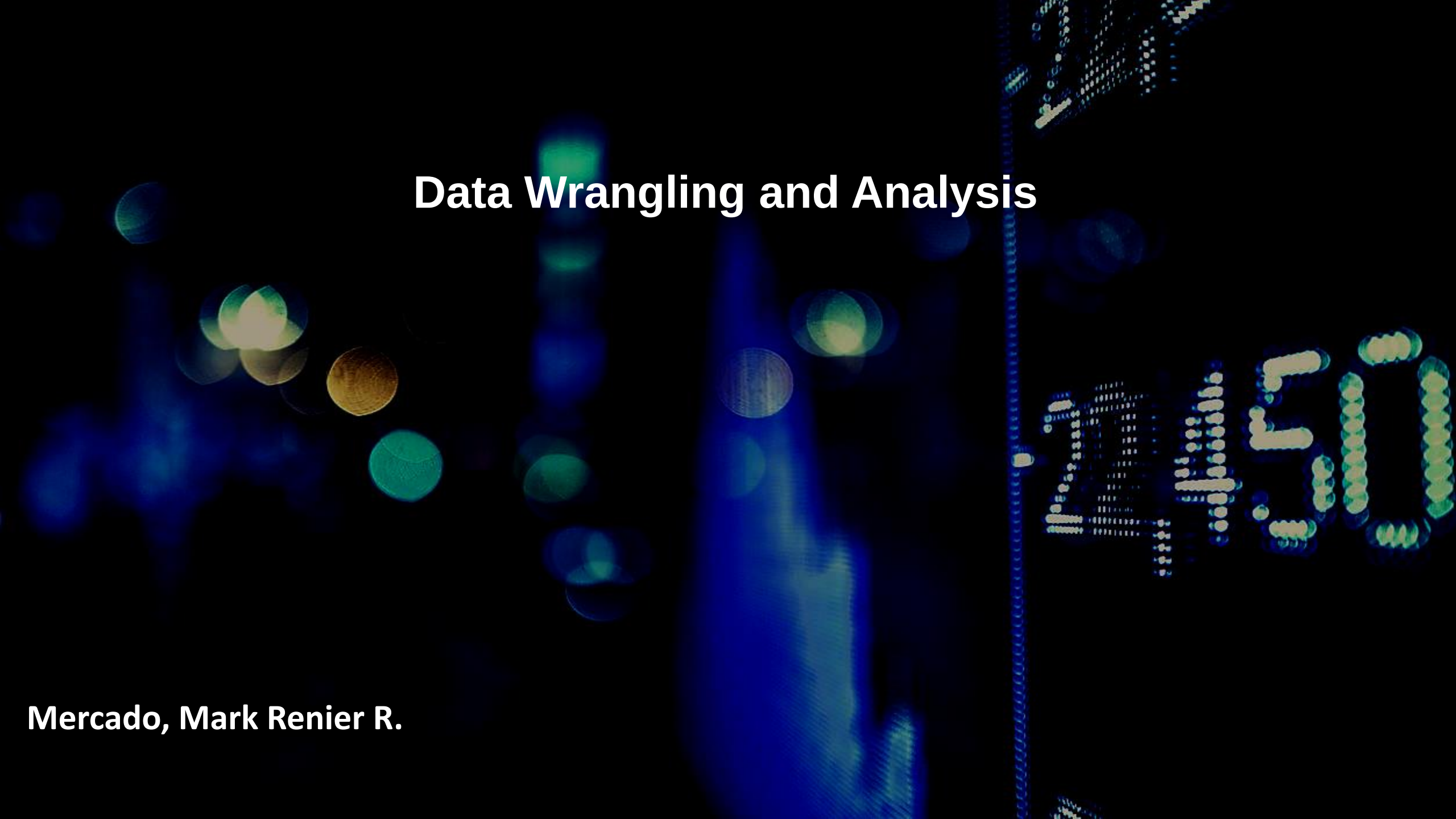


Data Wrangling and Analysis

Mercado, Mark Renier R.

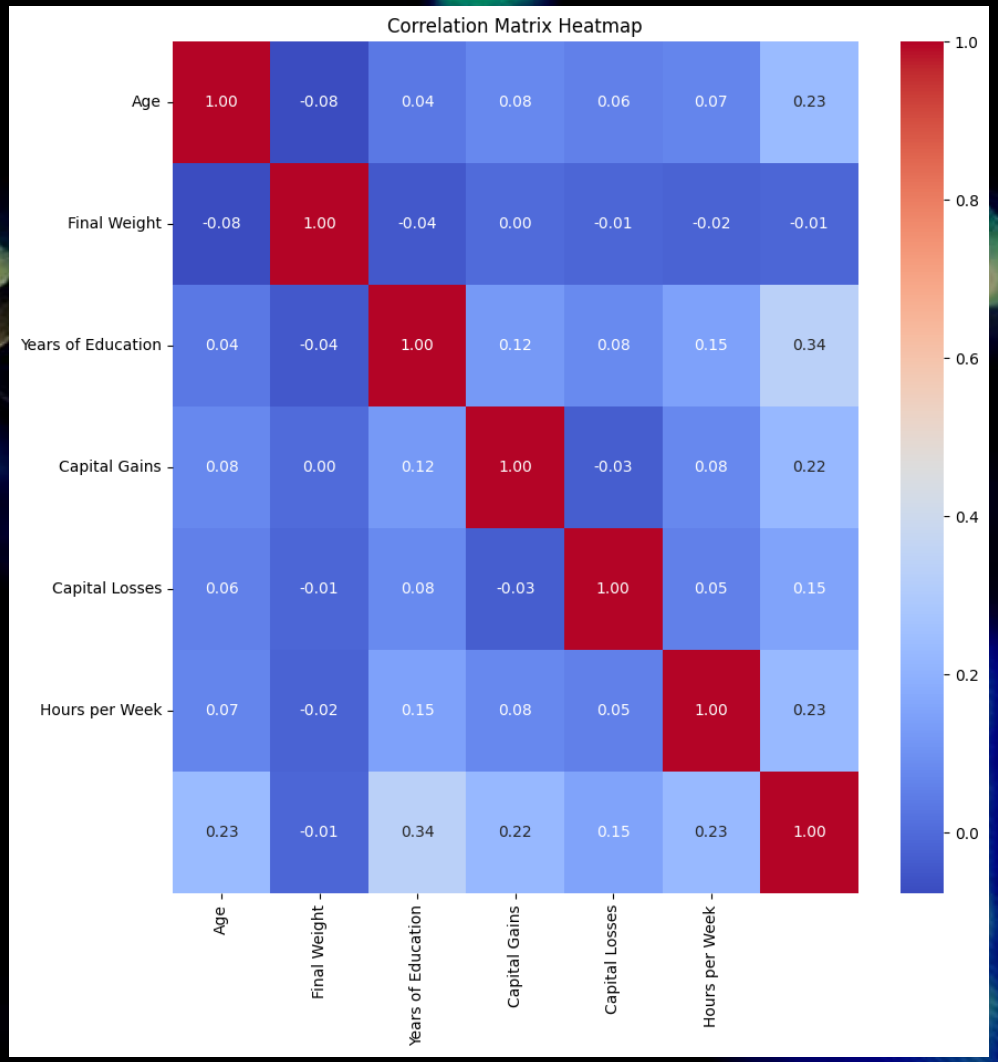
Description of Study

The exploratory data analysis conducted on the provided Dataset are shown in this report. This comprises data on a person's demographic characteristics, such as age, education, occupation, and income. Understanding the data and spotting any patterns or trends that might be of interest are the goals of this study.

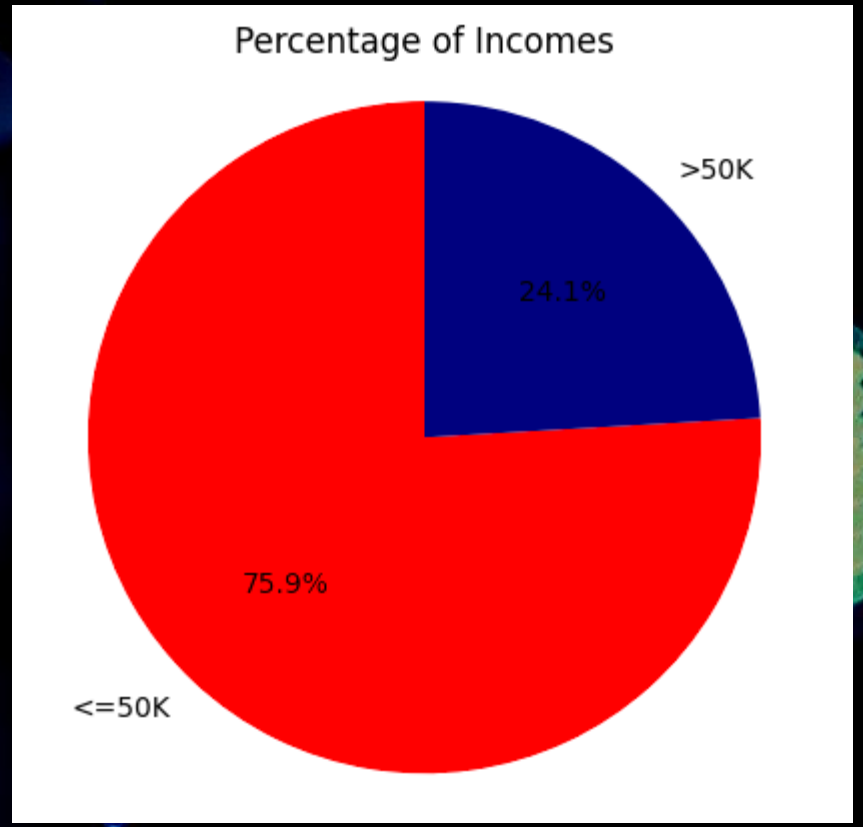
BASIC STATISTICS

Correlations:

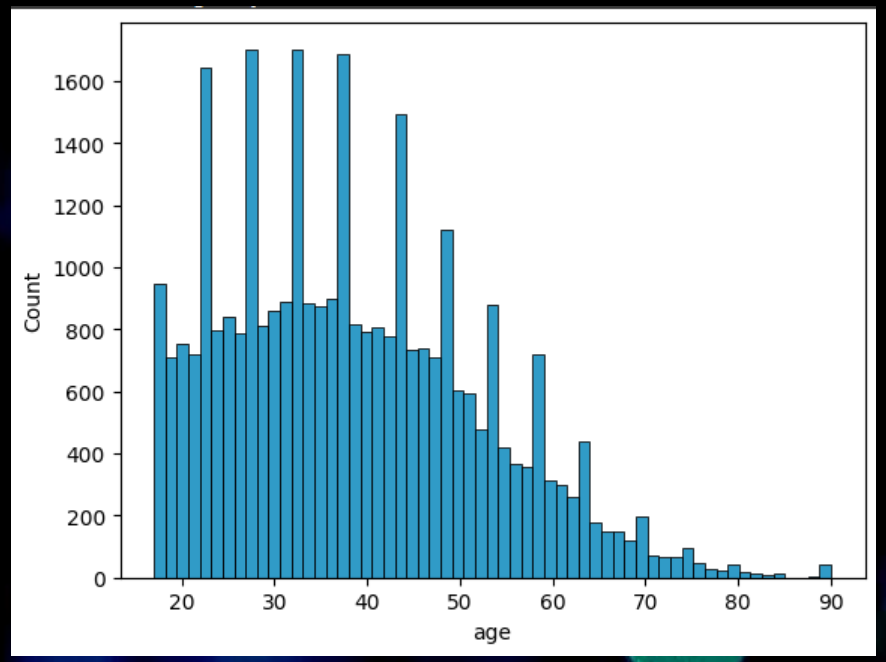
Red boxes indicates the correlations with 1.00



Percentage of income:



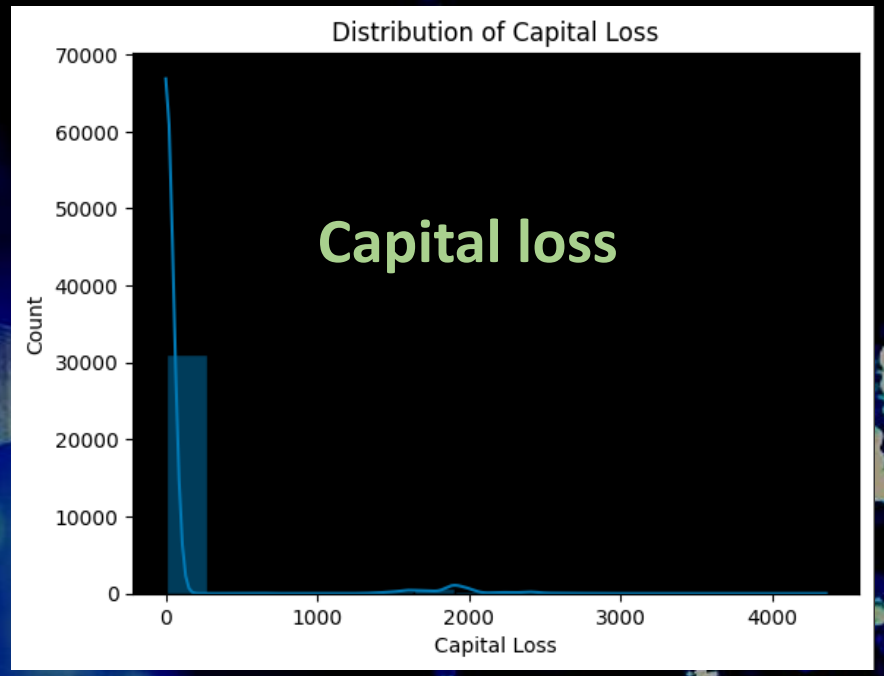
Distribution of age across the population



The line depicts the distribution's structure and reveals that the capital loss has a peak around 0 and is substantially skewed towards lower values. The bars in the graph show how frequently each amount of capital loss actually occurs. The number of people who suffered a specific level of capital loss is shown by the height of each bar.

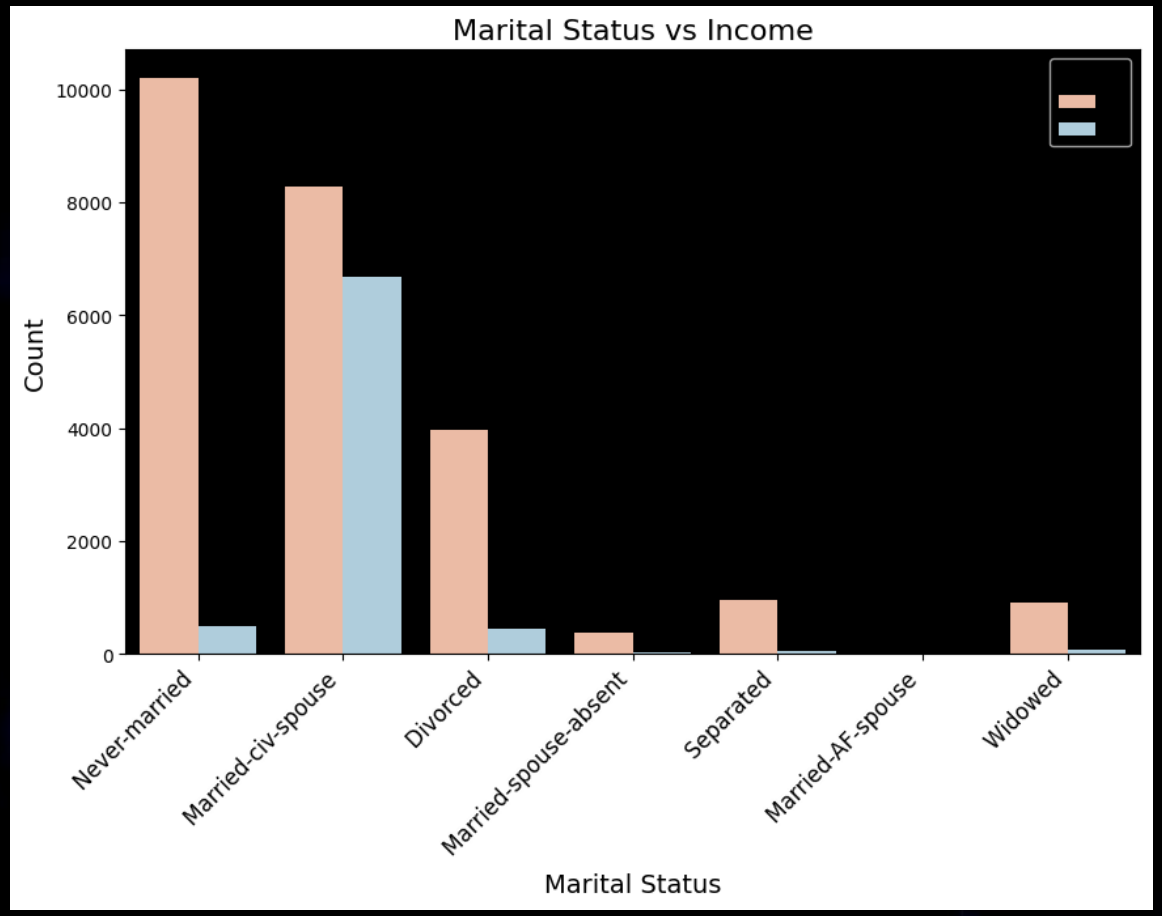
Number of Individuals with capital gain

>0: 2712

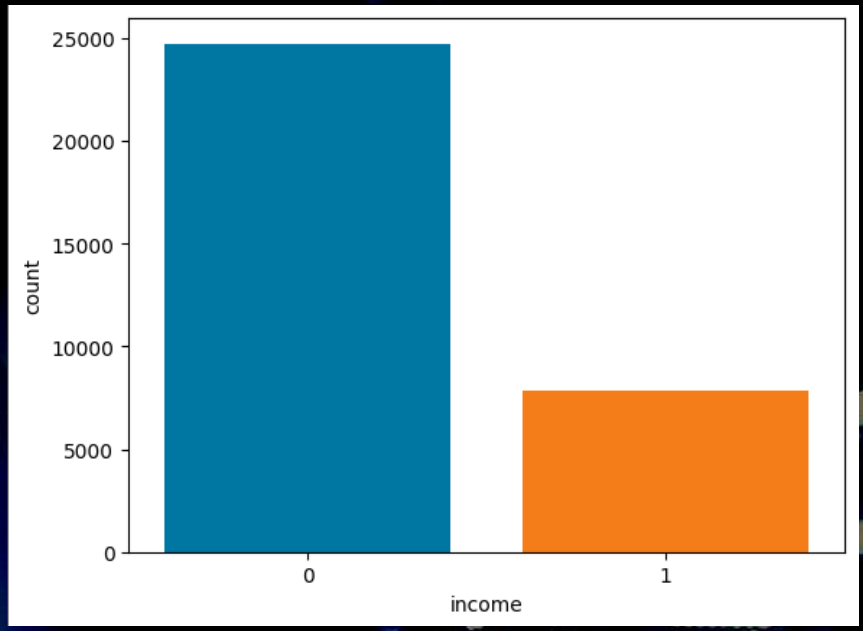


According to the population's age distribution chart, the vast majority of people are between the ages of 20 and 50. Because of the distribution's small rightward skew, more people fall into the older age categories. A generational trend or a big event that occurred around that time may be indicated by the increase in people who are 35 years old. The population appears to be broadly distributed across age groups, with a somewhat higher percentage of older people, according to the figure.

MARITAL STATUS VS INCOME



Relationship between Income and Count



Majority of the population in the dataset has an income less than or equal to 50,000. There are a relatively small number of individuals with incomes greater than 50,000.

Distribution of income levels for different marital status categories. It is evident from the plot that married individuals have a higher income level than non-married individuals.

ANALYSIS:

The distribution of capital losses is strongly skewed to the right, which suggests that the majority of people do not experience substantial losses. Since capital gain and income are positively correlated, those with higher earnings typically have bigger capital gains. All backgrounds and genders exhibit the same trend, with married people often earning more money than single ones.
For these analysis, it provides insight into the different elements that affect a person's level of income. We may better understand the income distribution across various groups and the variables that cause income disparity by evaluating the statistics.

CONLUSION

In conclusion, the ability to understand and get insights from data requires the ability to analyze data using tools like numPy and pandas for data wrangling, visualize data using Pandas and Seaborn, and execute exploratory data analysis on a complicated dataset. As a data science student, these abilities help me manage huge and complicated datasets successfully, extract valuable information from them, and effectively convey the results. I/we can create data-driven decisions that are advantageous to enterprises, organizations, and society at large by putting these talents to use. Therefore, everyone interested in working with data must become proficient in these abilities.

Data:
https://colab.research.google.com/drive/1j8WxMe0BSQ1i-tlo-0x08t84fjmK-nHL?usp=sharing