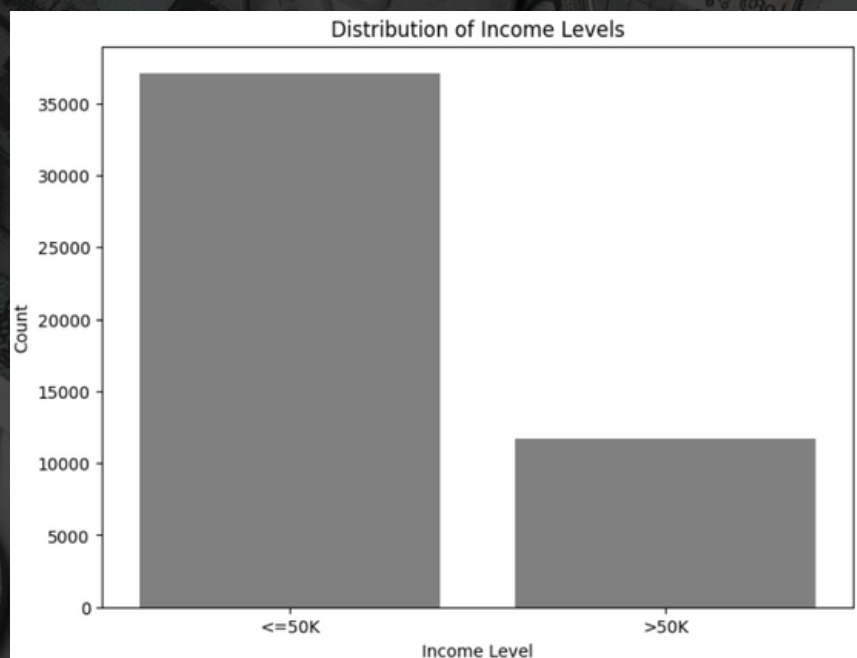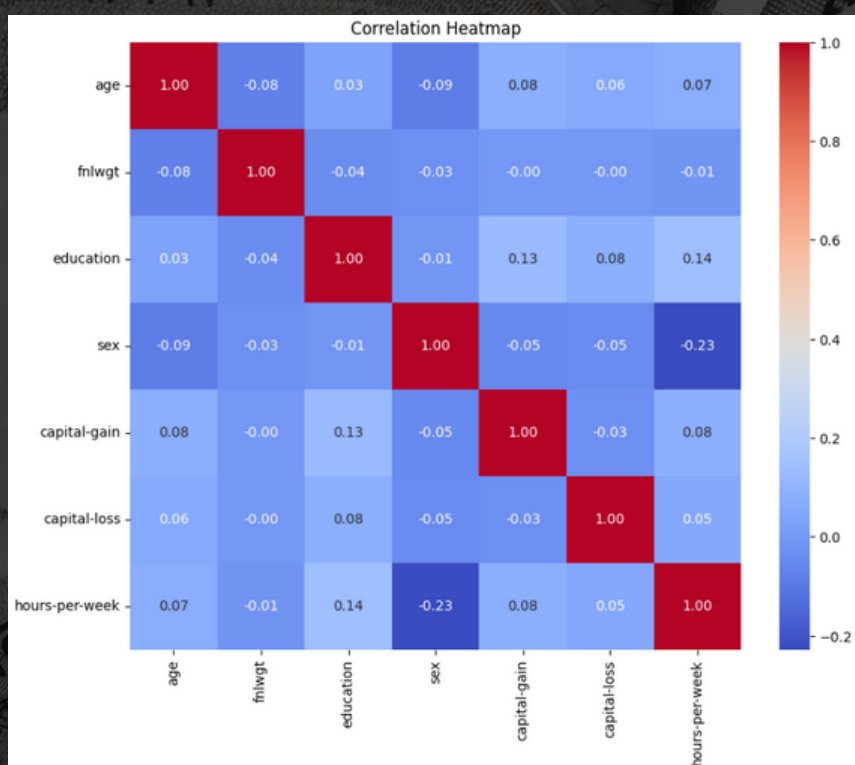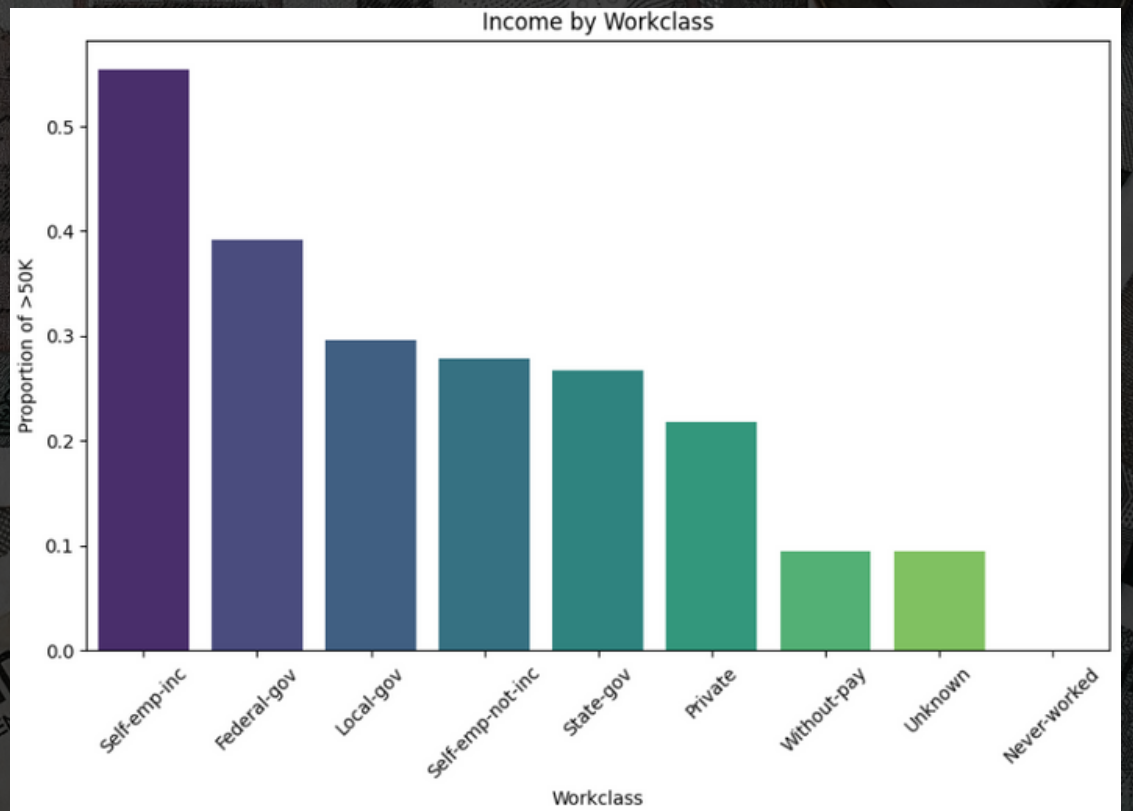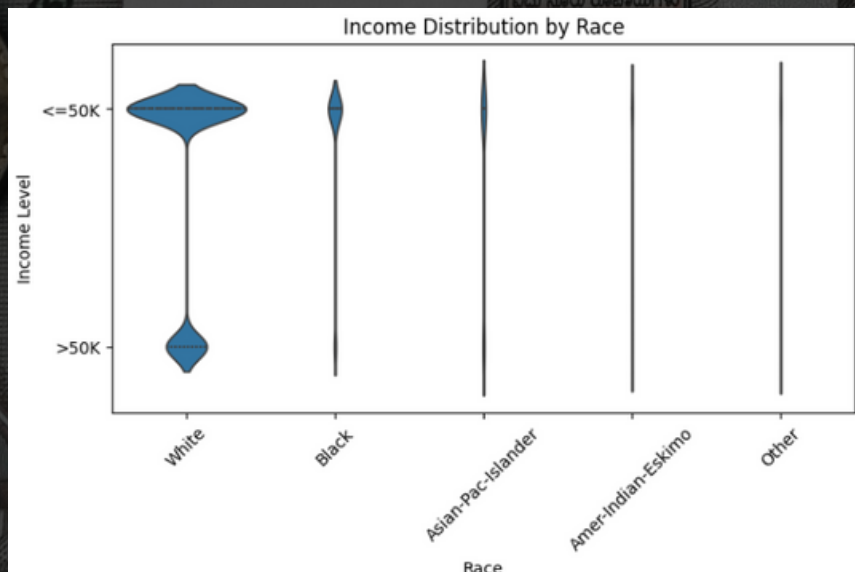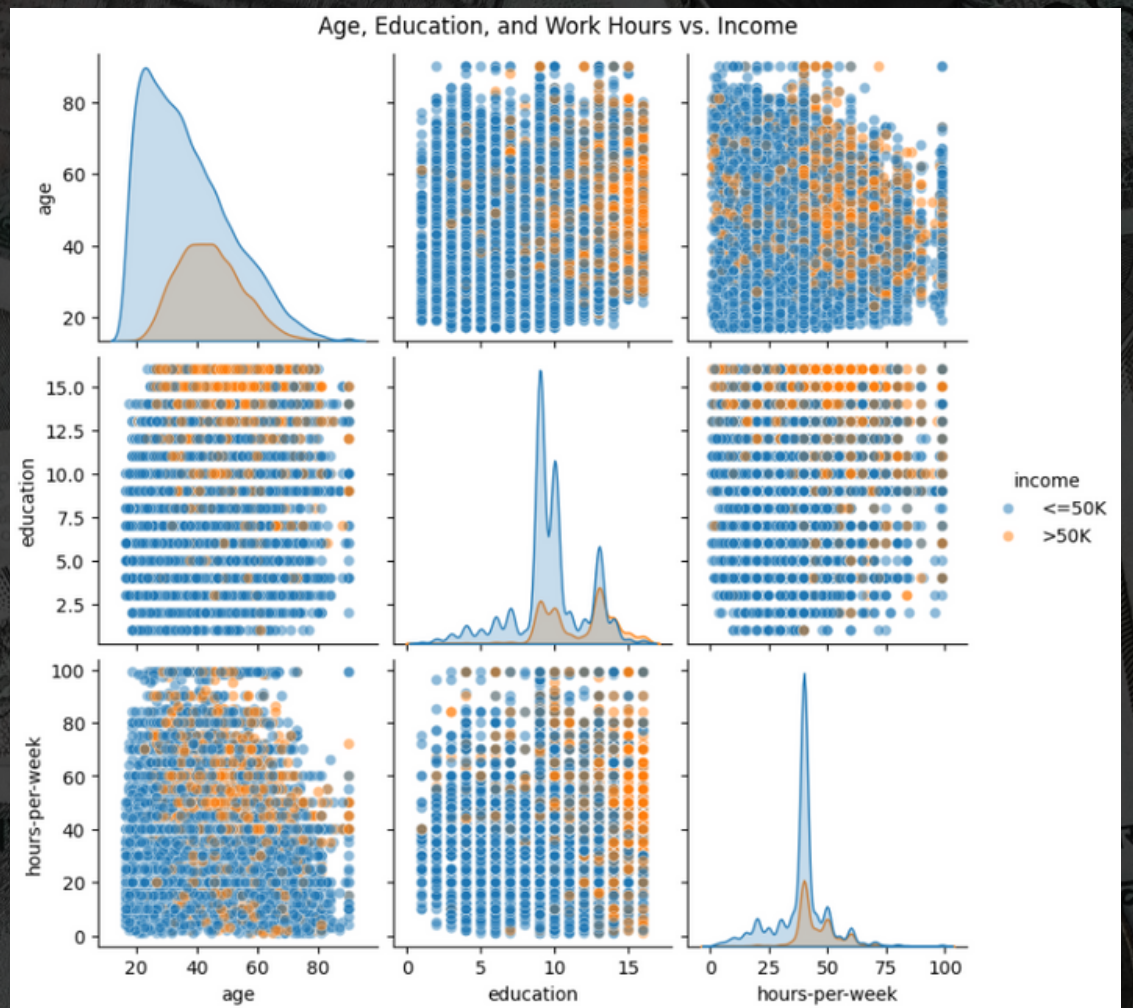# CENSUS INCOME

JAY-ANN ALORRO
CPE22S3

This Exploratory Data Analysis hopes to find the correlation between different factors and whether or not they affect the income levels of different random people from the Census data from 1994.

Distribution of Income Levels

Most people included earns less than 50 thousand a year.

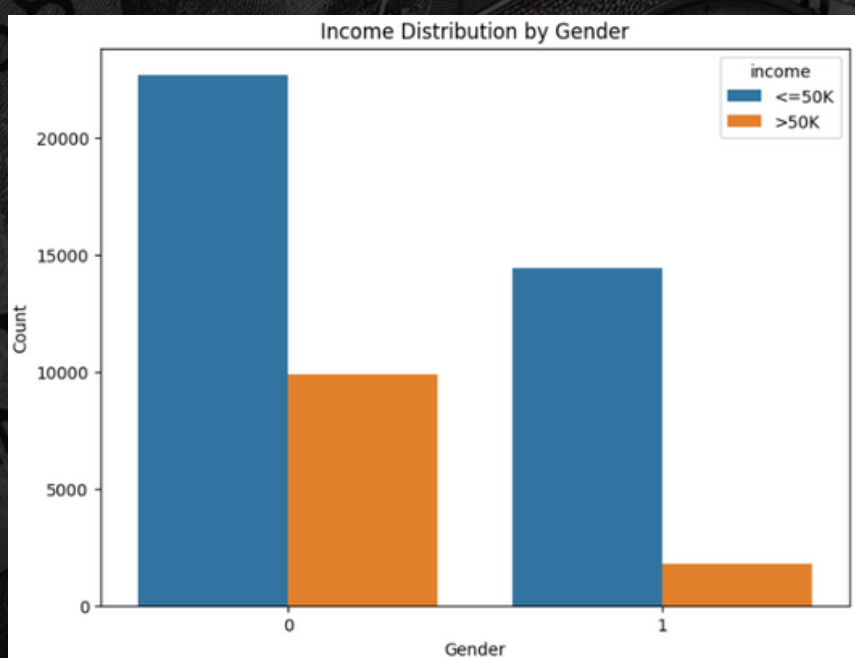Age, Education, and Work Hours vs. Income

Some categories do earn above 50K and it favors males and white people. Because of the data being from 1994, it is understandable that a higher income level prefers the said categories.

Each variable have some kind of relation with the income level which is shown in some of the graphs provided. However, there are little to no correlation to each variable that excludes 'income'.

Income Distribution by Race

Income by Workclass

The individuals in the dataset attained at least a high school education while the smaller group belongs to the ones that have a doctorate. I have found that the higher the education level attained, the higher their chances to be included to the group that earns >50K. Another thing is that when comparing those who have a doctorate and those that wen to a professional school, their ratio belonging in the individuals in the >50K and <=50K income is almost the same.

Most individuals belong to being self-employed which is limitless when it comes to earning. The federal government is the upper class in the public sector which gives them a higher chance to earn more than 50K.

Correlation Heatmap

In conclusion, some variables/features that are provided in the original dataset needed to be removed because they don't have any relevance in analyzing the data and some of them have a lot of bad data. The variables left in the dataset have their own relation with the income level an individual can belong in which I were able to analyze.

Income Distribution by Gender