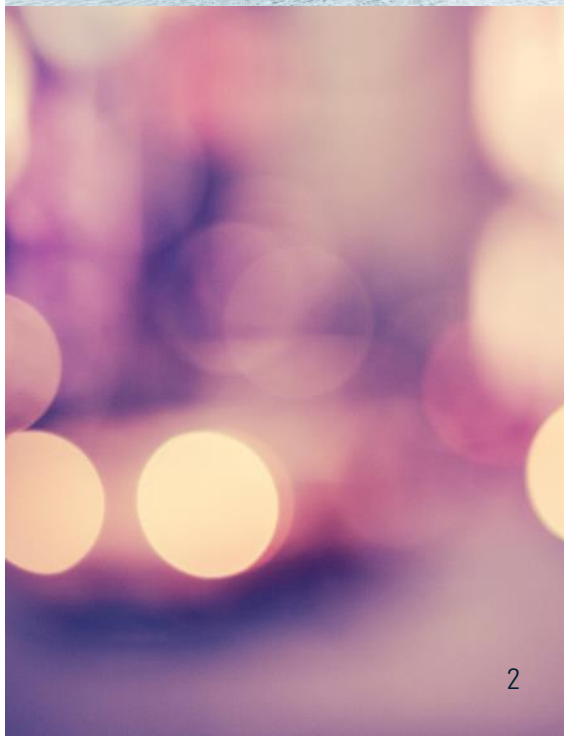
ADULT INCOME MODEL

Jody Jacobs

AGENDA

- Introduction to the Data
- Key Findings
- Strengths and Limitations of the Model
- Final Recommendations



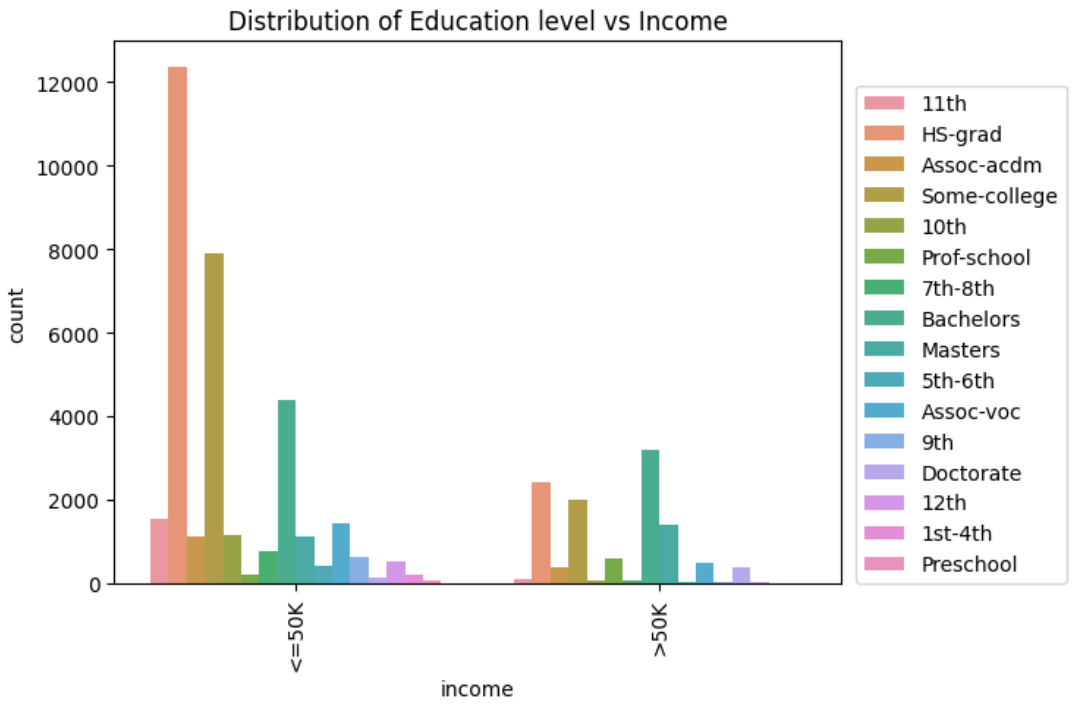
INTRODUCTION

This dataset was taken from the 1994 Census database and can be found on the UCI repository. It consists of 48,842 rows and includes 15 columns.

The goal of the model is to predict whether an adult's income will be less or more than $50000,00, based on various demographic data such as gender, education, age, and more.

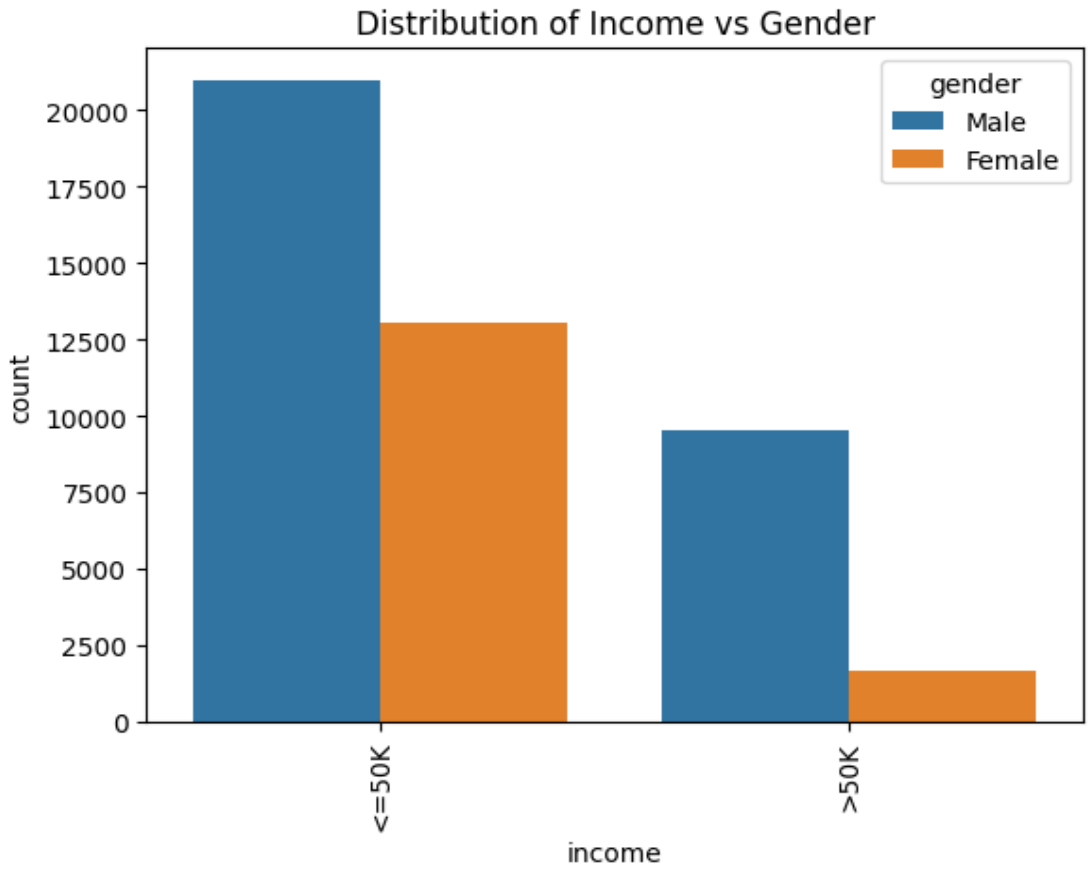
KEY FINDINGS

DISTRIBUTION OF EDUCATION LEVEL VS INCOME



- The graph shows the distribution of education levels and income in the United States. The higher the education level, the higher the income. The distribution of education levels is skewed to the left, meaning that most people have lower levels of education. The distribution of income is also skewed to the left, but to a lesser degree.
- From the graph, we can see that the majority of adults earning less than 50k have a high school diploma or some college education. For adults earning more than 50k, the majority have a bachelor's degree.

DISTRIBUTION OF INCOME VS GENDER



- Based on the graph, it is evident that there is a significant pay gap between men and women in terms of income distribution. The data highlights that there are a higher number of men earning above $50000 as compared to women. Moreover, it is clear that the data collected is mostly male respondents, which may have also influenced the results.

STRENGTHS AND LIMITATIONS OF THE MODEL

Tuned Linear Regression Model

STRENGTHS AND LIMITATIONS OF THE MODEL

STRENGTHS

- It has a high true positive rate, meaning that it correctly predicted Adults earning > $50000 in 88% of the cases
- It has a relatively low false positive meaning that it incorrectly predicted Adults > $50000 in 27% of the cases
- It has a lower true negative rate, meaning that it correctly predicted Adults earning <$50000 in 73% of the cases
- It has lower false negatives meaning that it incorrectly predicted Adults <$50000 in 12% of the cases

LIMITATIONS

- I've observed that most of the adults in the dataset earn less than $50000, which means the model is imbalanced
- The model has low False Positive and False Negatives ratings. However, it's important to note that these ratings are not zero, which means there will still be some incorrect predictions with this model.

RECOMMENDATIONS

- I noticed that the dataset has a larger sample size of males compared to females, which could potentially create bias in the model. It seems that a higher percentage of males are earning over $50000 compared to females. This gender disparity is an important factor that should be taken into account in our society.
- Additionally, it would be beneficial to include location as a demographic feature in order to determine which states have a higher earning potential.




THANK YOU

Presenter Name: Jody Jacobs