

# UCI Adult Census (1994)

Zoe Severson, Jasmine Jordan-Lake

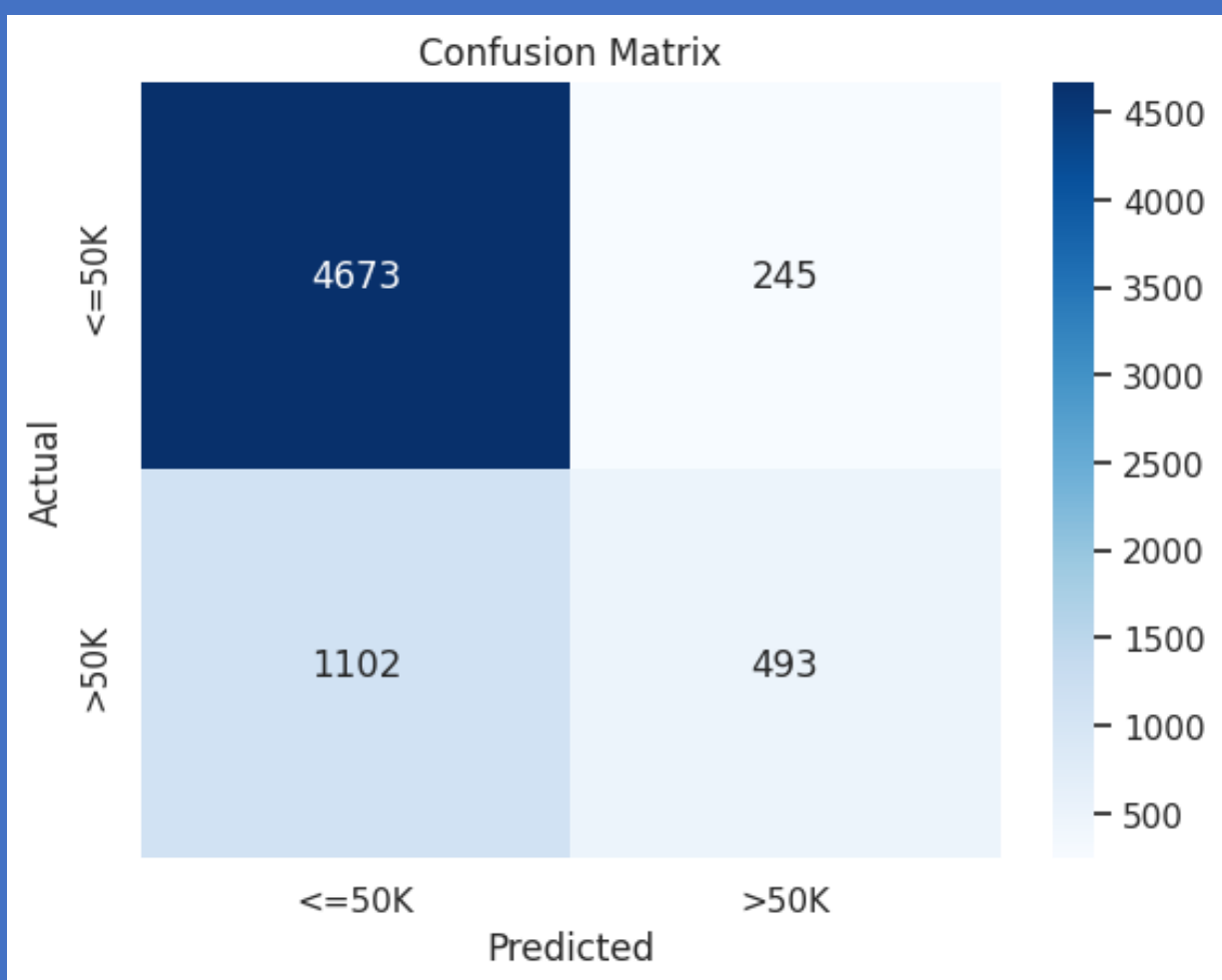
## Introduction

The UCI Adult Census dataset (1994) serves as a valuable resource for understanding the **socioeconomic landscape** and analyzing the **complexities of income levels**. This dataset provides a comprehensive array of demographic attributes, offering insights into the **factors influencing income levels**. Our study utilizes this dataset to delve into possible predictors of income disparity and examine underlying trends.

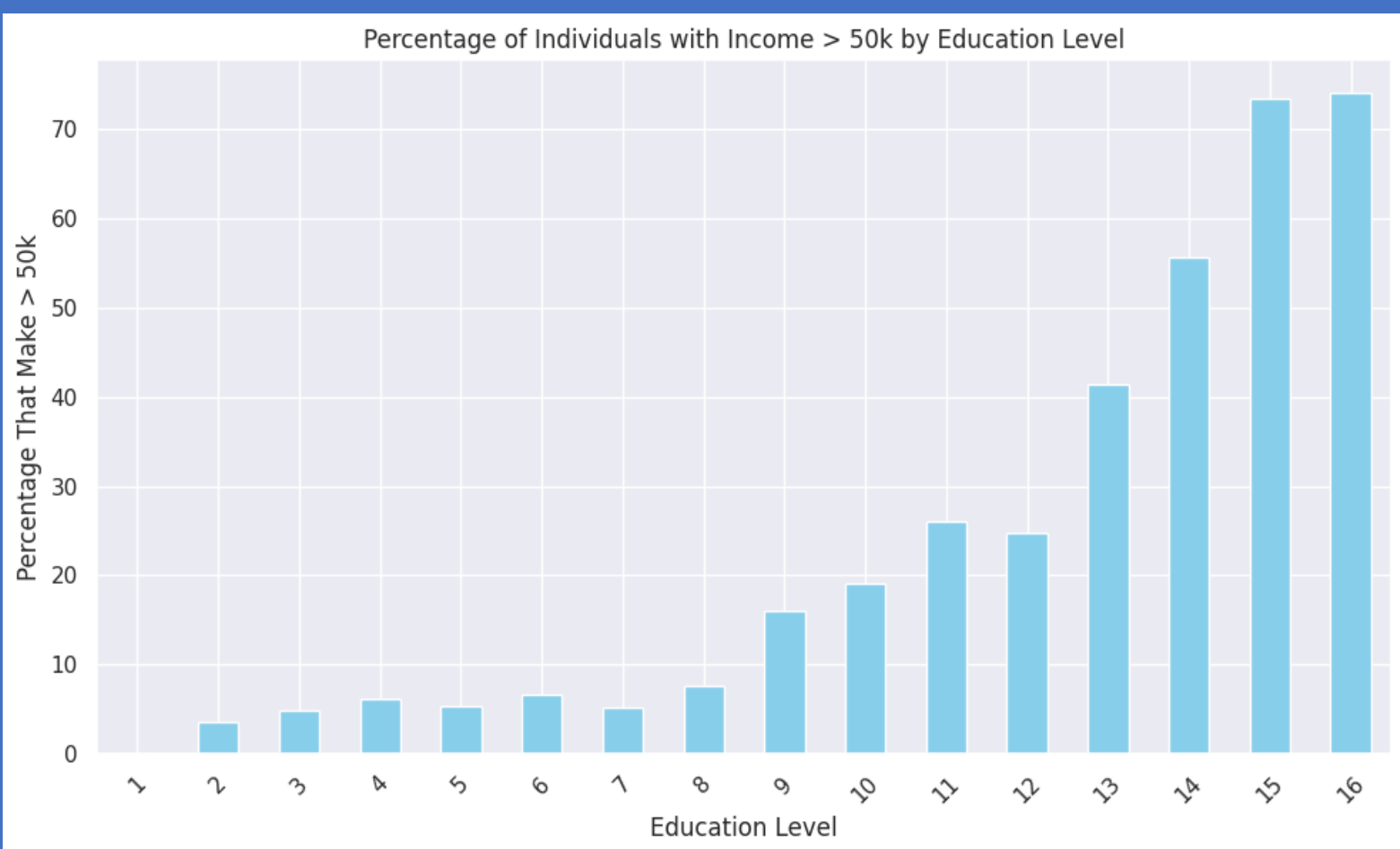
## Objectives

1. Ensure data accuracy and reliability by fixing errors, inconsistencies, and missing values
2. Utilize Naïve Bayes algorithm to predict income levels based on demographic attributes.
3. Uncover correlations and dependencies among demographic factors and income disparity.
4. Offer insights into the socioeconomic determinants influencing income disparities.

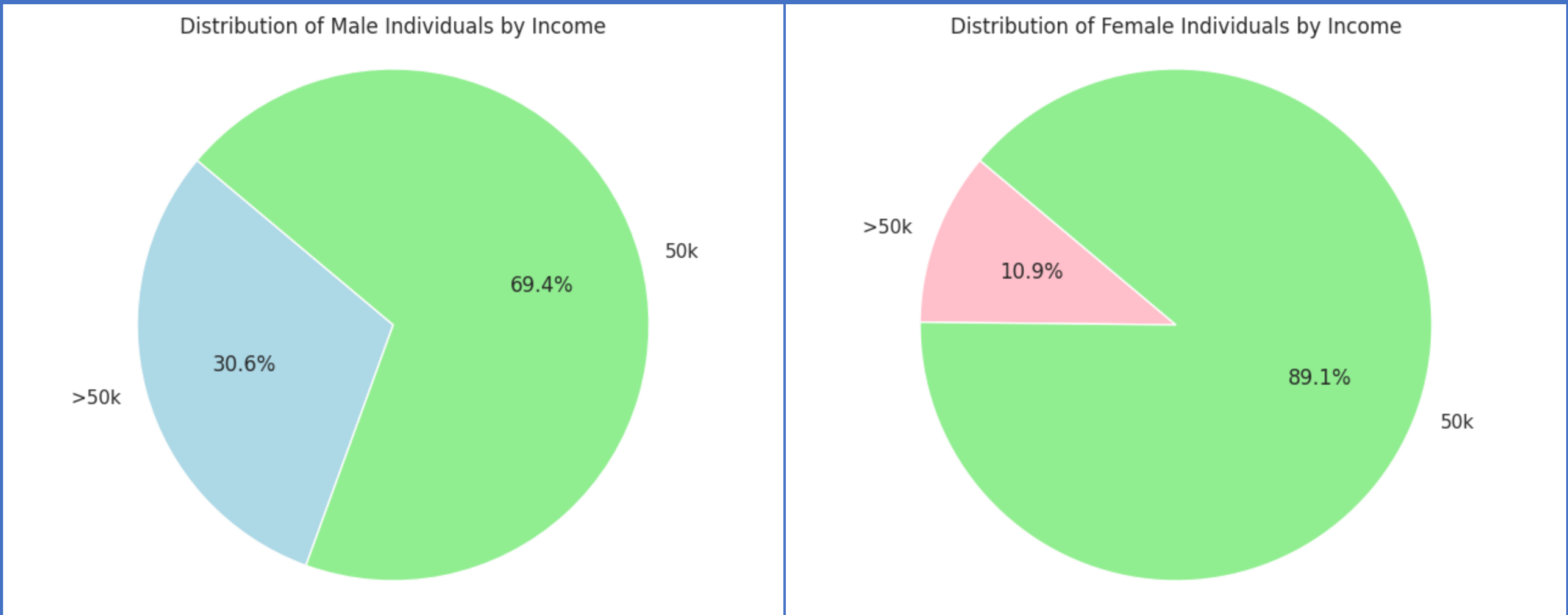
## Results



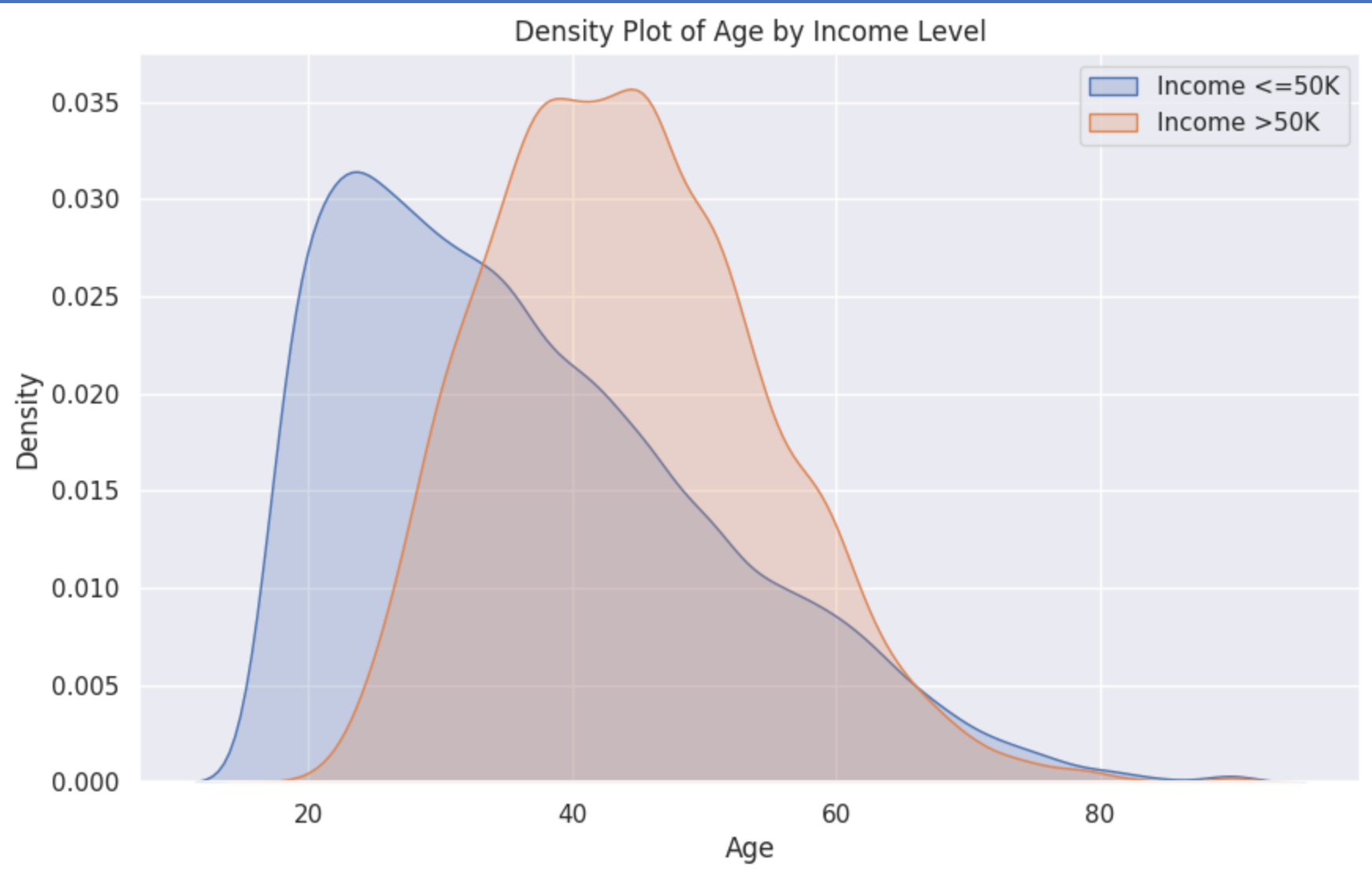
### Education Level



### Sex



### Age



## Methods

We used the Naïve Bayes algorithm to predict income levels (>50k or <=50k) using demographics. The data was split into training and testing sets, with 20% allocated for testing. A Gaussian Naïve Bayes classifier was then trained on the training data to learn patterns in income levels. Predictions were made on the testing set, and model performance was evaluated using accuracy score and a classification report.

## Conclusions

- Analysis of the confusion matrix showcased a robust precision of 81% in classifying individuals with incomes over \$50k, while also pinpointing room for enhancement with a precision of 67% for incomes below \$50k.
- The data suggests a moderately positive correlation between education level and income classification, showing that those who had more schooling typically made more.
- In this dataset, 30.6% of men earn over \$50k, compared to only 10.9% of women, highlighting possible inequalities.
- The majority of individuals earning over \$50k fall within the age range of approximately 35 to 50, whereas the majority of those earning \$50k or less range from approximately 20 to 30 years old.
- Used to gain valuable understanding of the socioeconomic landscape and explore targeted interventions to help promote a greater sense of equity.