ANALYSIS AND PREDICTION OF ADULT INCOMES

U.S CENSUS BUREAU DATASET

Mini Project 2
Yee Rin Lew
15/6/2024

Agenda

Background & Objective

↓

Insights

↓

Summary & Outcome

↓

Suggestion & Limitation

↓

Machine Learning Technique

# Background & Objective

Background & Objective

By looking into a dataset with 14 demographic factors and income level of 32561 individuals, we can

- Understand how various demographic factors influence income levels.
- Gain insights into the economic status of different groups within the population.

With this,

- develop strategies for target marketing with customer segmentation,
- develop pricing strategies that align with the purchasing power of different income segments,
- develop compensation packages that match the financial expectations of potential hires,
- Identify underserved markets and potential areas for expansion by analyzing income distribution.
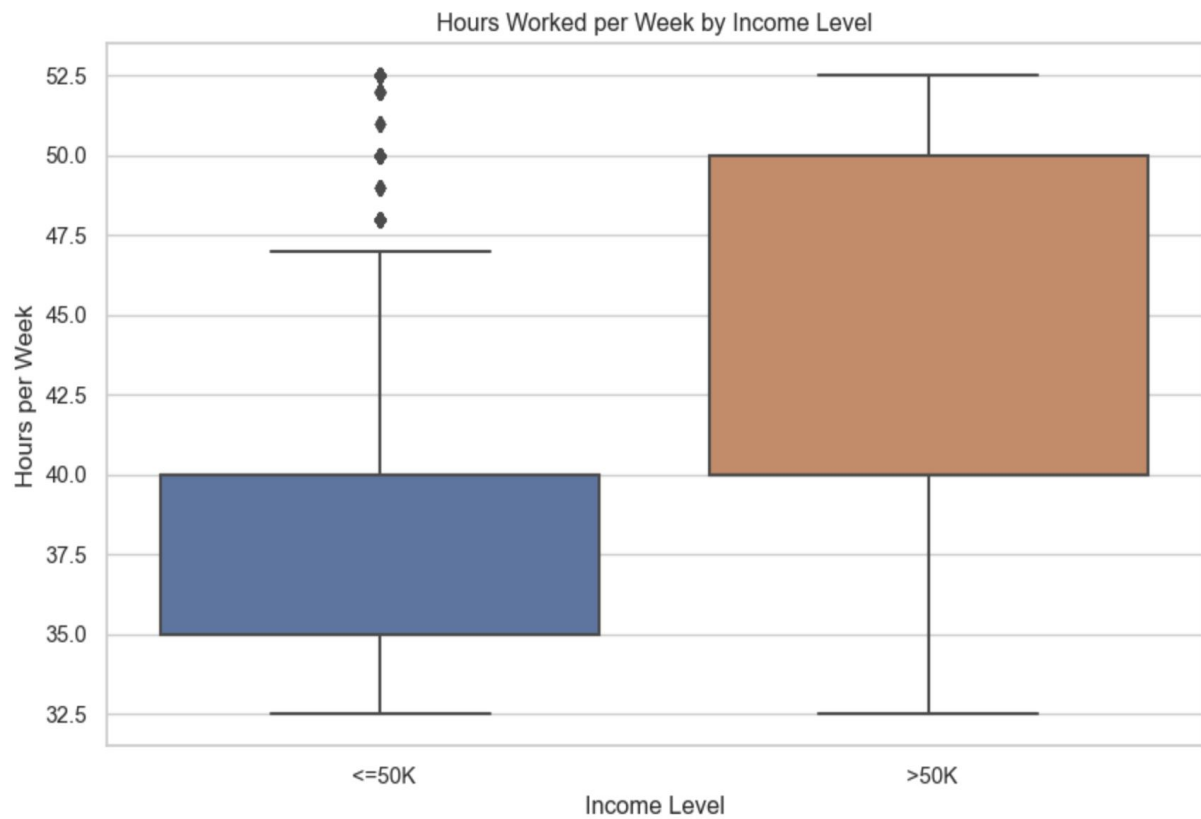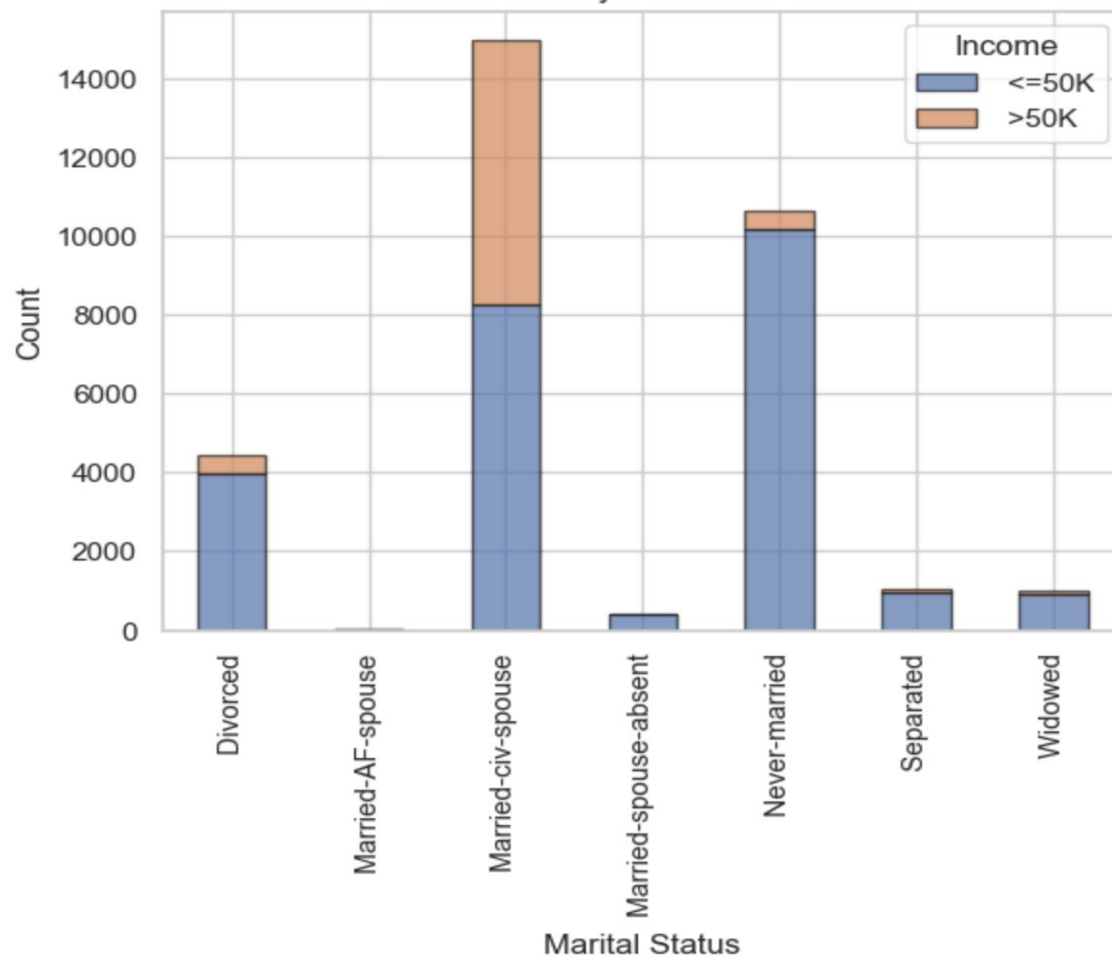
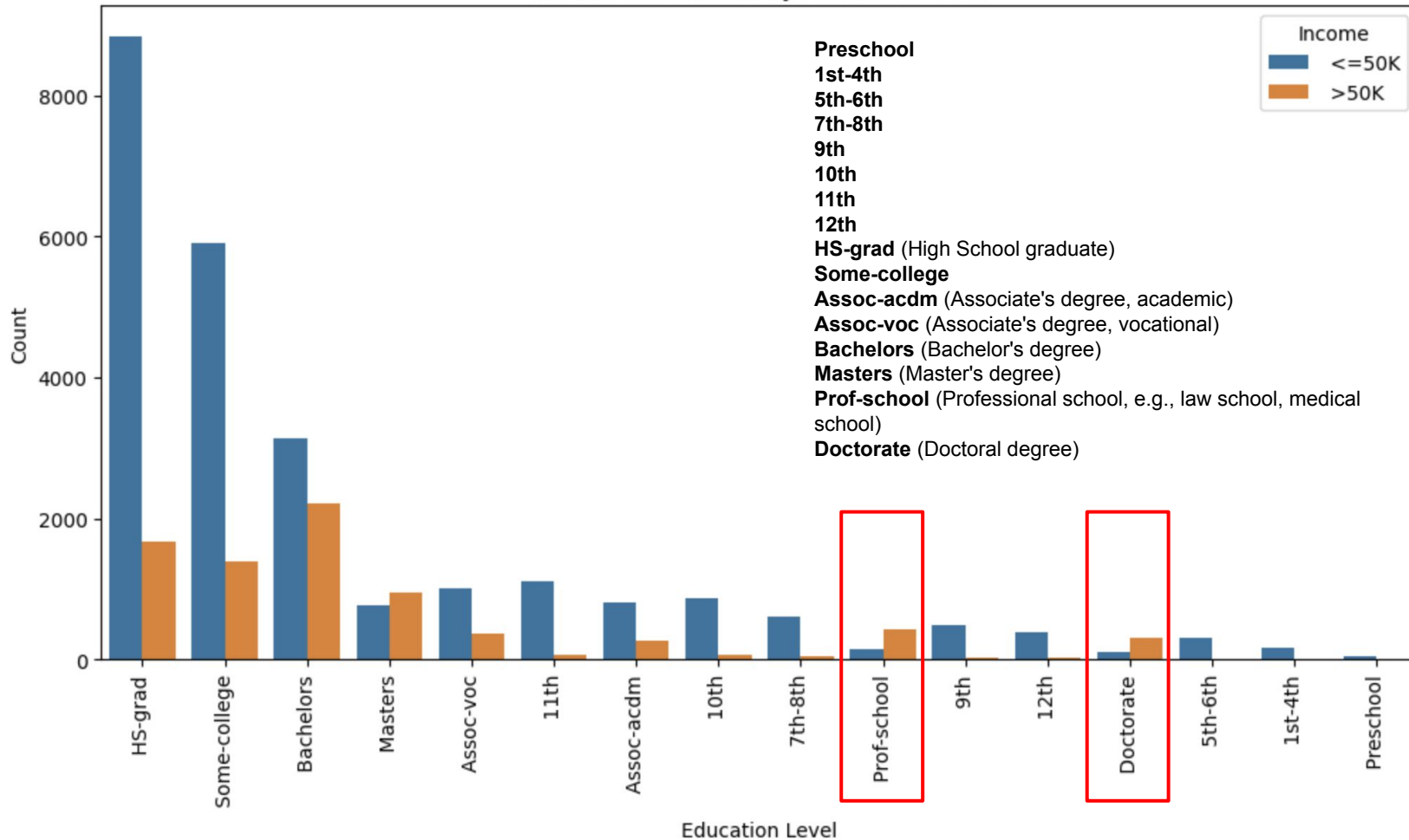## Distribution of Income
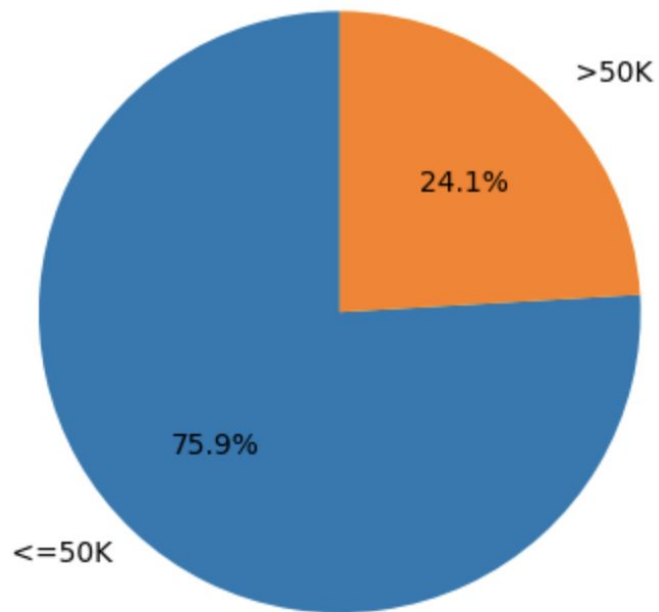
Age Distribution by Income

Income by Gender

Hours Worked per Week by Income Level
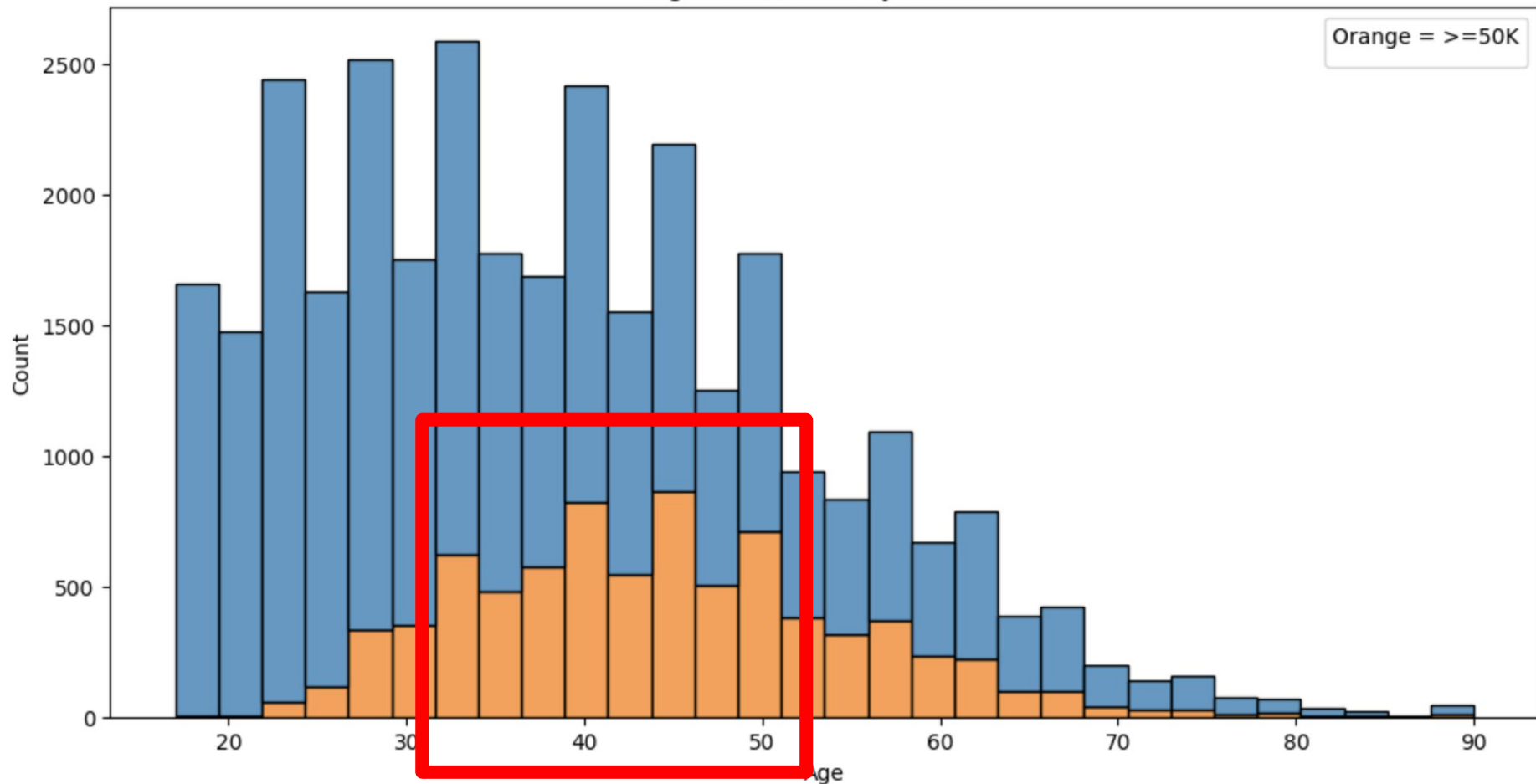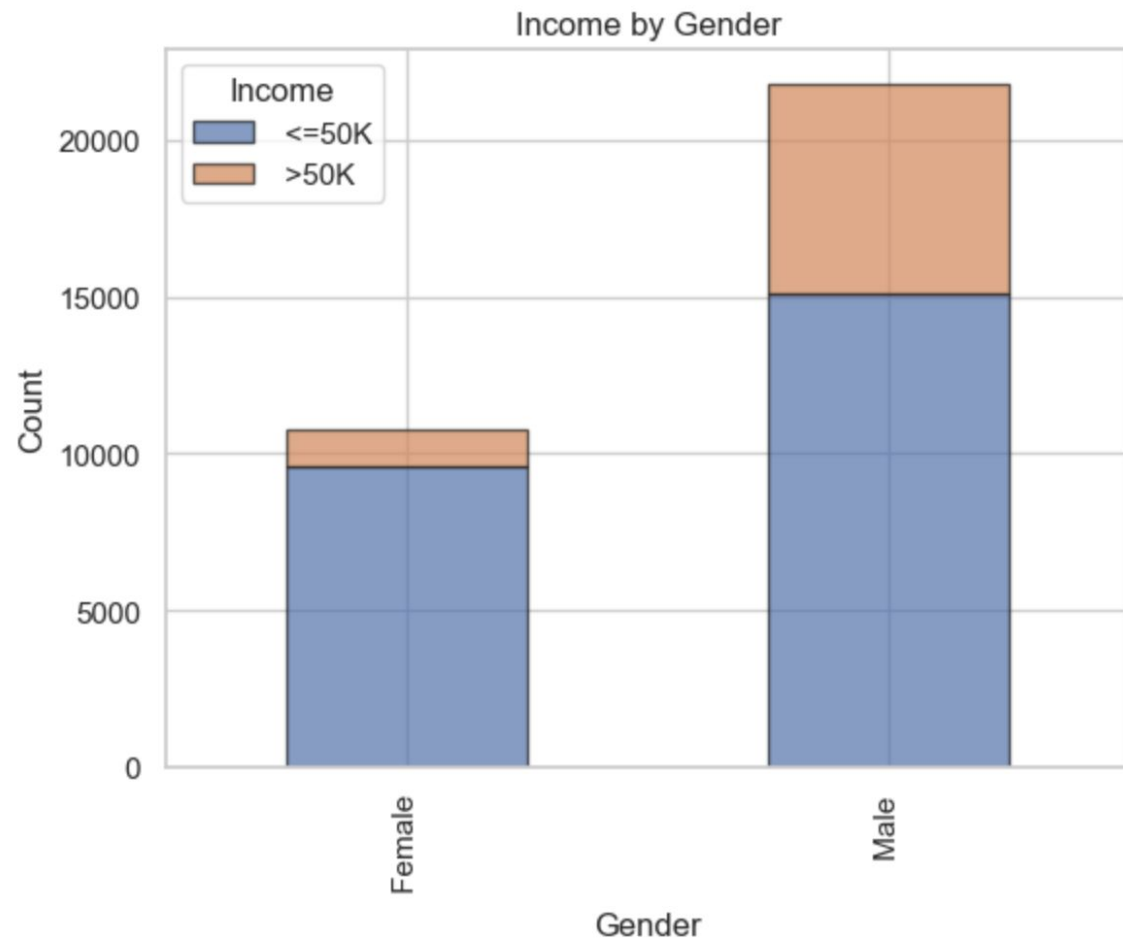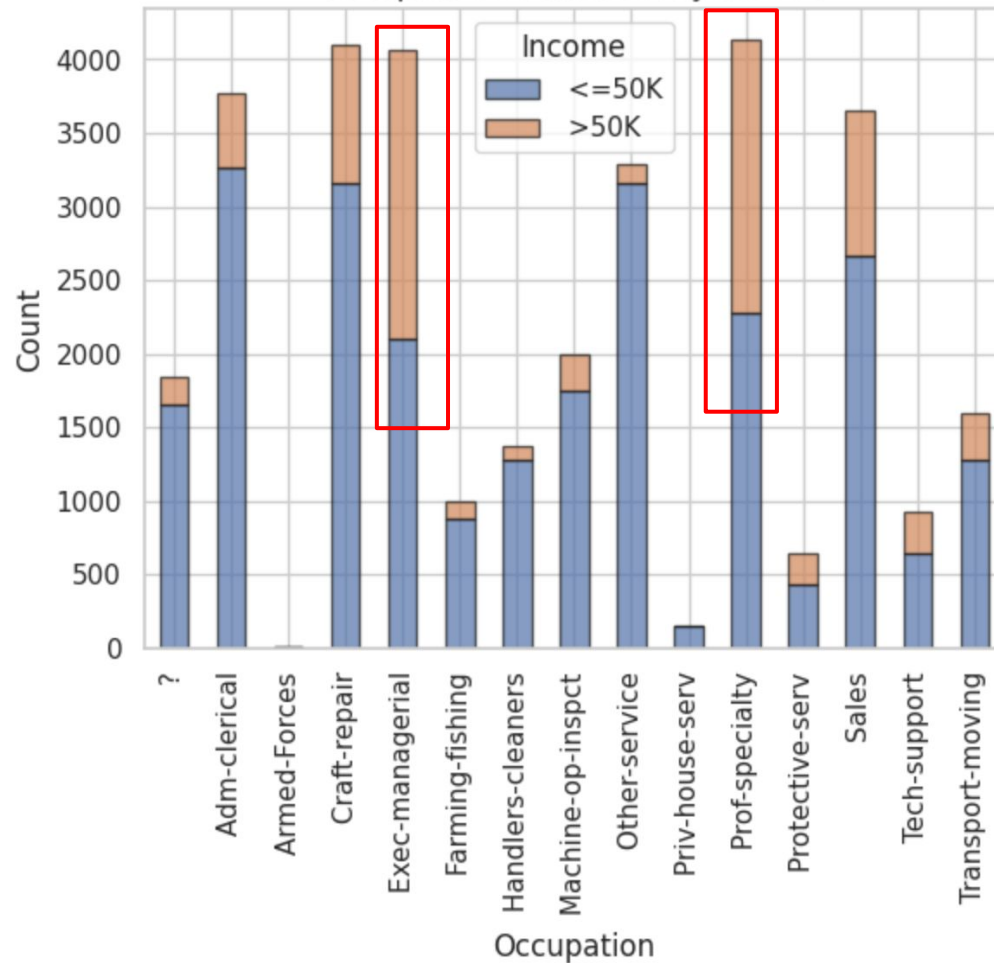
Income by Marital Status

Income Distribution by Education Level

Occupation Distribution by Income

Summary & Suggestion

Summary

Key Findings

1. **Income Distribution:**
   - A significant majority of individuals earn <=50K, with a smaller portion earning >50K.
2. **Demographic Influences on Income:**
   - **Age:** Middle-aged individuals (30-50 years) tend to have higher incomes.
   - **Gender**: Male tends to have a higher income in view of career interruption of female etc.
   - **Education:** Higher education levels (Bachelor's degree and above) are strongly correlated with higher incomes.
   - **Occupation:** Professional and managerial occupations have higher incomes compared to manual labor and entry-level jobs.
   - **Work Hours:** Full-time work (35-50 hours per week) is associated with higher incomes.
   - **Married-civ-spouse (Married, civilian spouse):** Tend to have higher incomes compared to other marital statuses. This group includes many dual-income households, which can contribute to higher overall income.

Suggestion

Targeted Marketing and Sales Strategies

- Use income predictions to create personalized marketing strategies. Tailor product recommendations and promotions based on the predicted income levels of different demographic groups to increase engagement and satisfaction.
- Prioritize high-income segments for premium product offerings and exclusive services to maximize revenue and customer loyalty.

Pricing Strategies

- Develop tiered product lines that cater to various income levels, ensuring broad market appeal. Offer basic, mid-range, and premium versions of products.

Suggestion

Human Resources and Recruitment

- Ensure competitive and equitable compensation packages by regularly reviewing and adjusting salaries based on demographic data and market benchmarks.
- Create targeted career development and mentorship programs to support the advancement of underrepresented groups and improve employee satisfaction and retention.

Market Expansion

- Analyze income and demographic data to identify underserved markets and regions with high-income potential for expansion.

Stakeholder

- Chief Marketing Officer
- Sales Director
- Customer Service Team
- Human Resource Officer
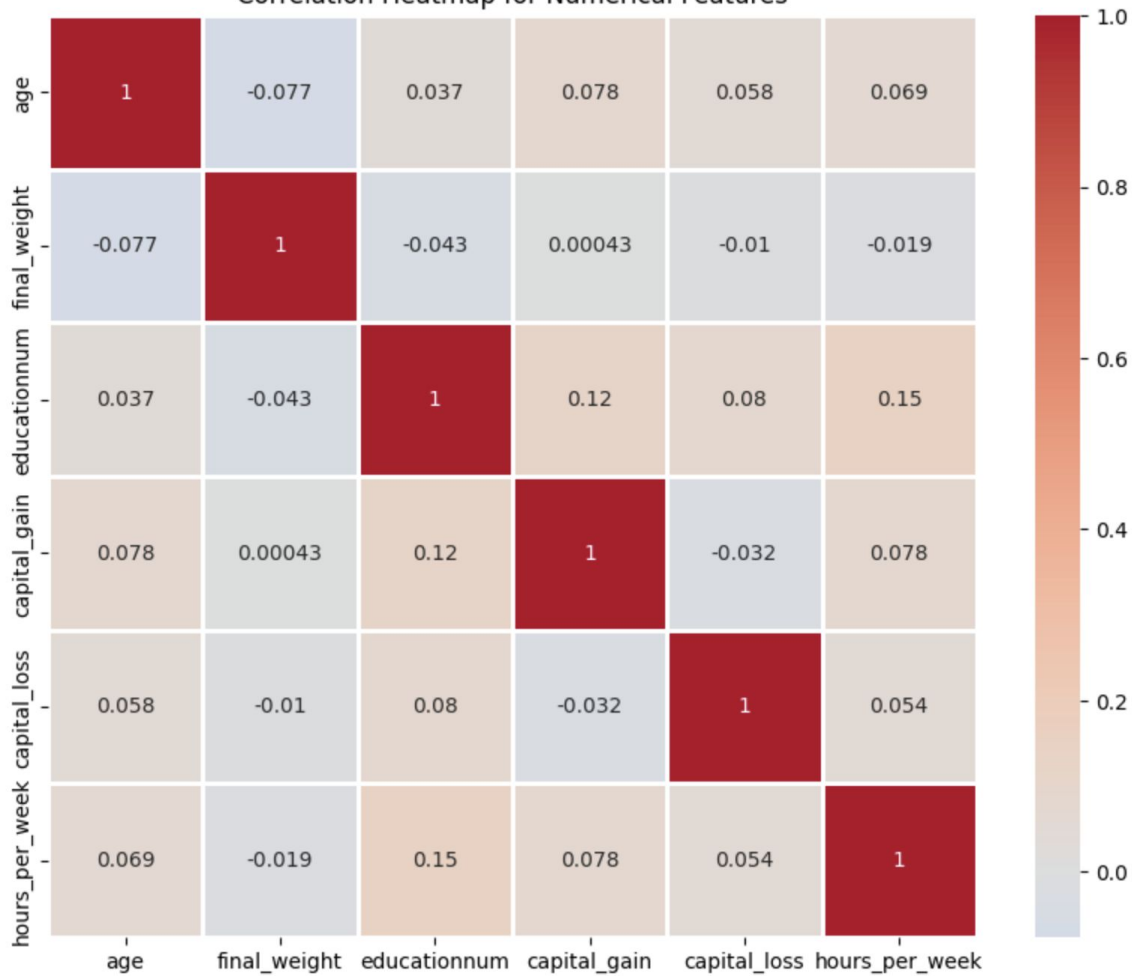- Learning and Development Manager
- CEO

Limitation

Limitation

- primarily covers the United States, limiting its applicability to other regions with different economic and social contexts.

- the income variable is divided into two broad categories (<=50K and >50K), which may not capture the full range of income levels and nuances within those groups.
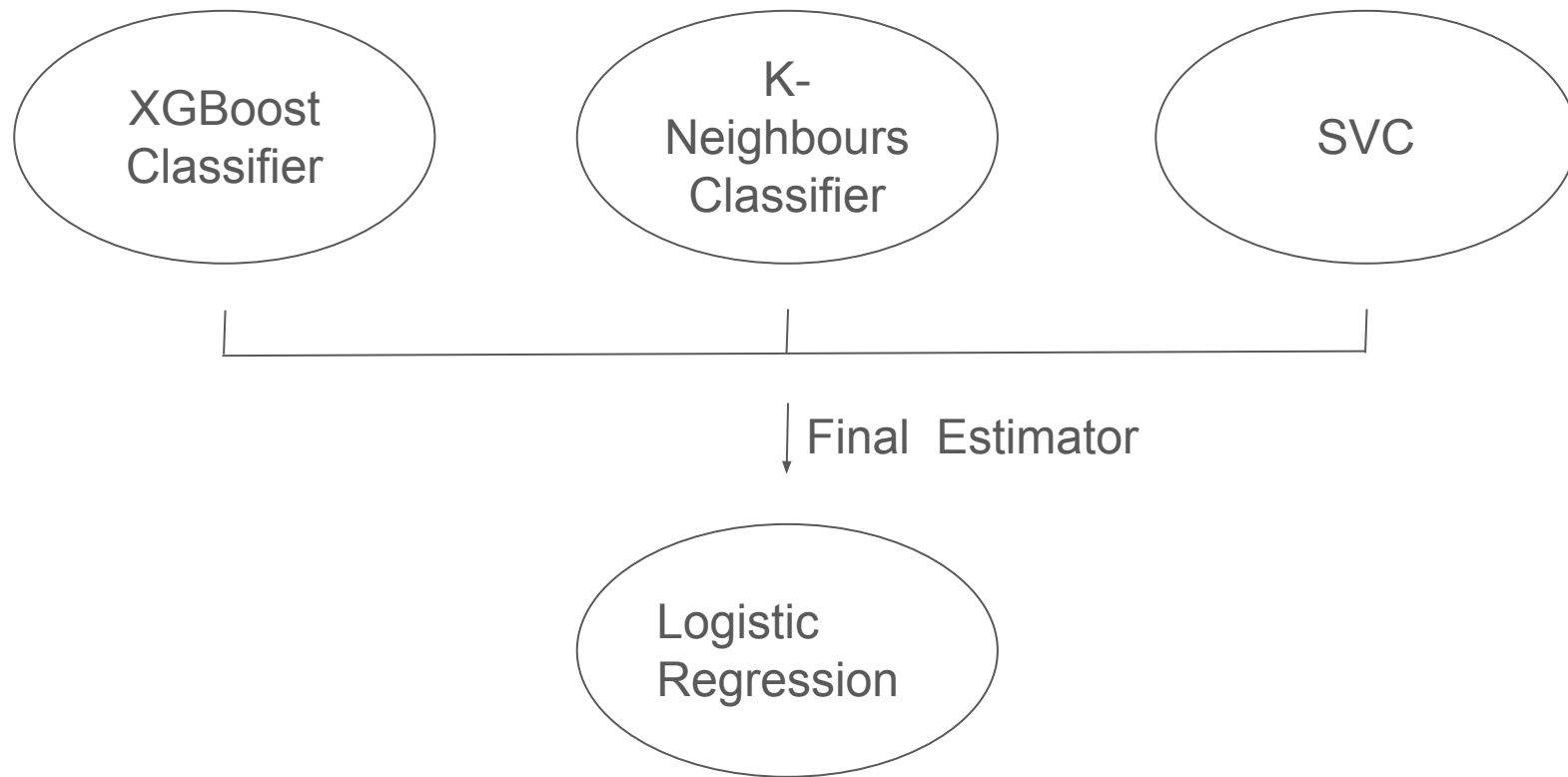
Machine Learning

Dataset

Data Source	United States® **Census** Bureau (Ronny Kohavi and Barry Becker) Data Shape: (32561 x 15)
Data Cleaning	Python · NumPy · pandas
EDA	matplotlib · seaborn · scipy · pandas
Machine Learning & Evaluation	scikit-learn · MLxtend

Correlation Heatmap for Numerical Features

Stacking Classifier

XGBoost Classifier

K-Neighbours Classifier

SVC

Final Estimator

Logistic Regression

Comparison Between Models

ROC AUC Curve

Thank you!