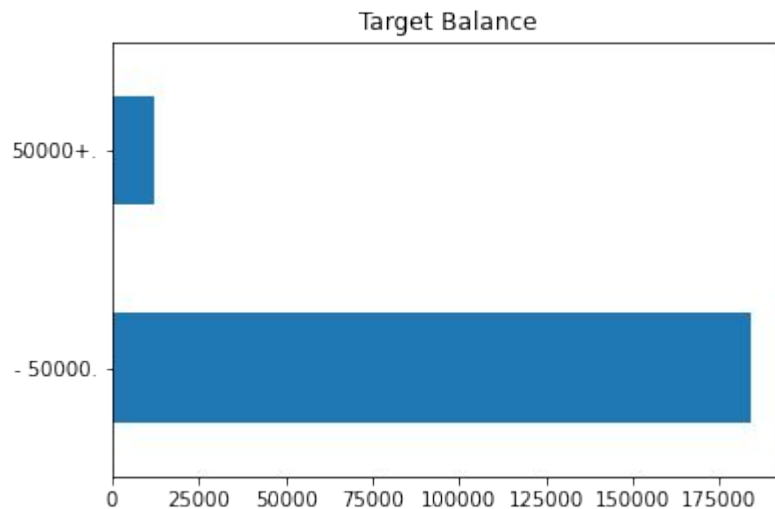


Census Income

Sibongile Toure

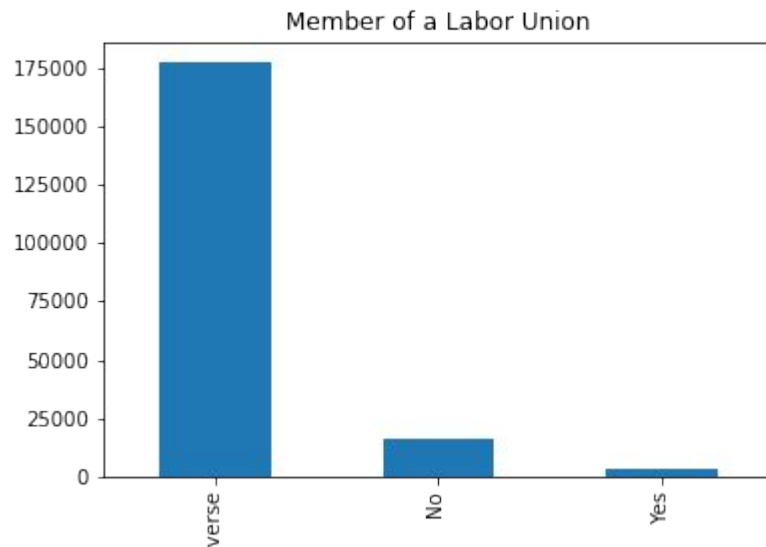
EDA

- Unbalanced Dataset
- Features
 - 33 Nominal
 - Marital Status
 - Class of Worker
 - 7 Continuous
 - Age
 - Number of Hours Worked

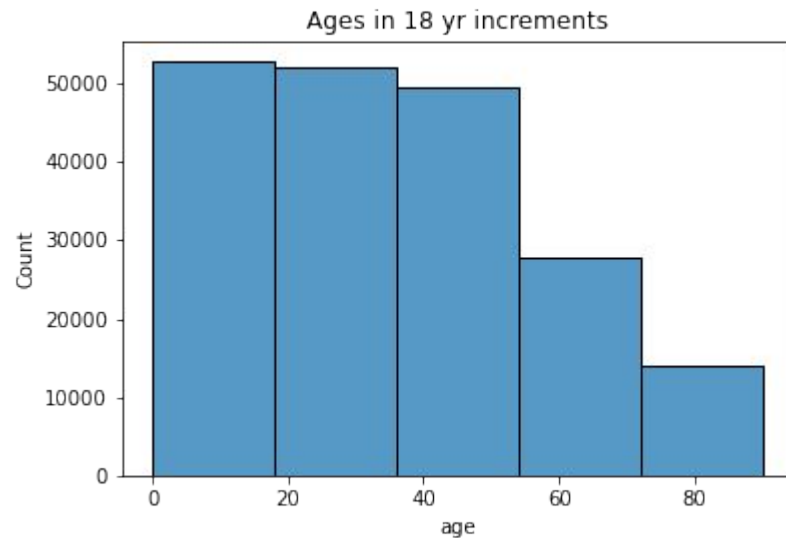
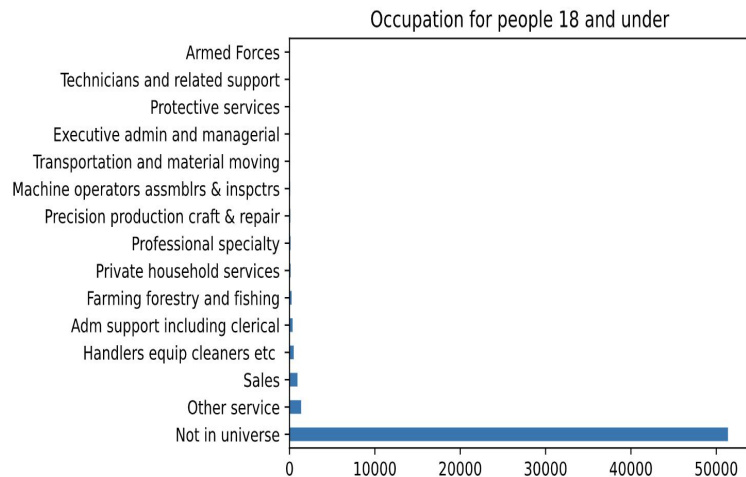


Feature Engineering

- Dropped features with ~50% NaN values
- Excluded data for people 18 and Under

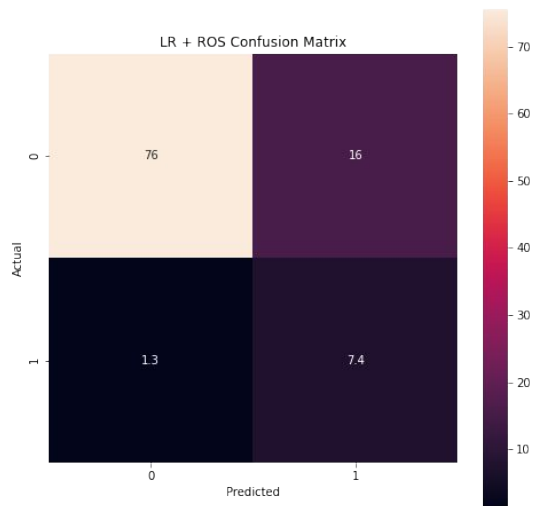
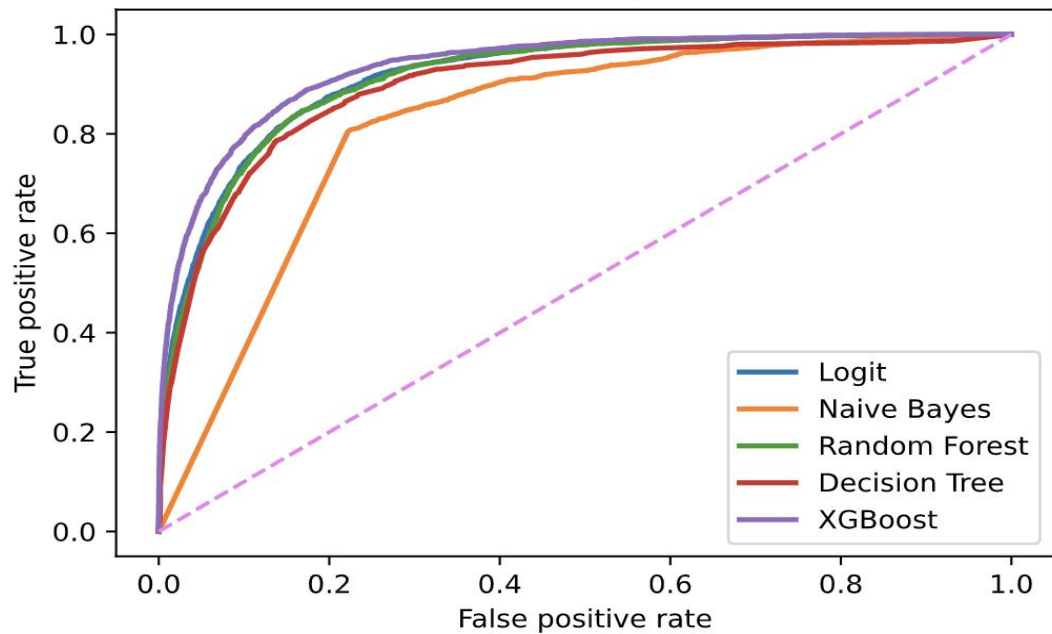


Feature Engineering



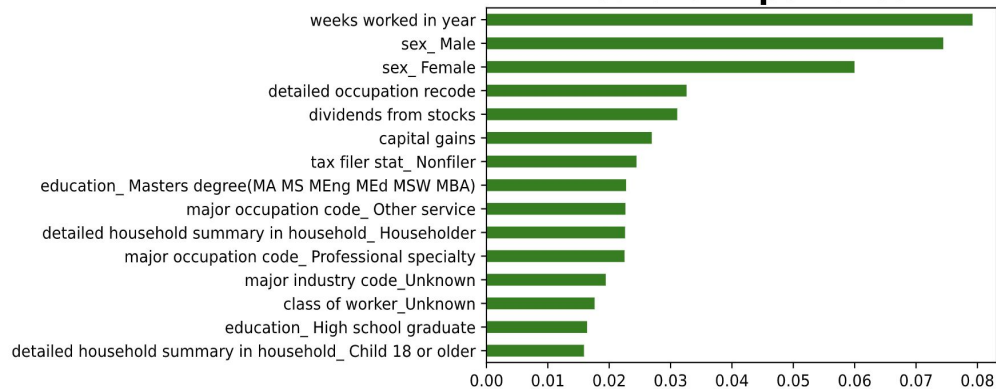
Modeling

ROC Scores

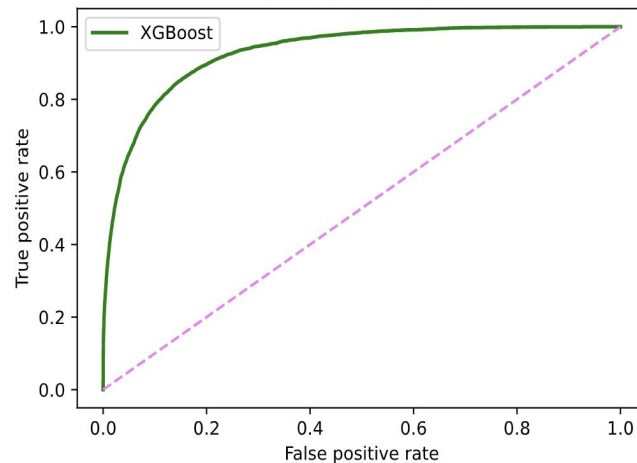


XGBoost

Feature Importance



ROC Score



Takeaways

- By segmenting out people under 18 we are able to get a clearer picture of the workforce
- Certain features such as how many weeks a person worked in a year and their sex has an impact on whether they make above or below \$50000

Future Work

- Include more data from other sources
 - E.g. Neighborhood data can provide insights