Adult Income

Presented by Kristina DiBella Silva

Mission Asset Fund (MAF)

Mission Asset Fund is a non-profit organization with a mission to help low income and immigrant families. As well as other causes such as Urban and Community Economic Development and Immigrant's Rights.

WE ARE ON A MISSION TO CREATE A FAIR FINANCIAL MARKETPLACE FOR HARDWORKING FAMILIES.

Information: https://greatnonprofits.org/org/mission-asset-fund-maf

Our Purpose

- MAF wanted to explore the probability of an individual's annual income being higher or lower than $50,000.

- Inquire whether gender and cultural background are determining factors to higher or lower annual earnings.

- This project investigates the skills and background that could potentially impact income.

- Observations like hours worked per week, marital status, years of education, and capital gains and losses play a part in understanding an individual's ability to obtain their income.

- This is a Binary Classification problem for predicting if an adult will make more than 50K or less.

Individual Background

This data originally had a column of 41 Native Countries. In order to consolidate this for machine learning we reframed it to 4 Continents. There were no countries listed for continent of Africa.

➔ North America is dominant.
➔ Regardless of the Native background there is a lesser count for those that make greater than 50K.


Continent by Income
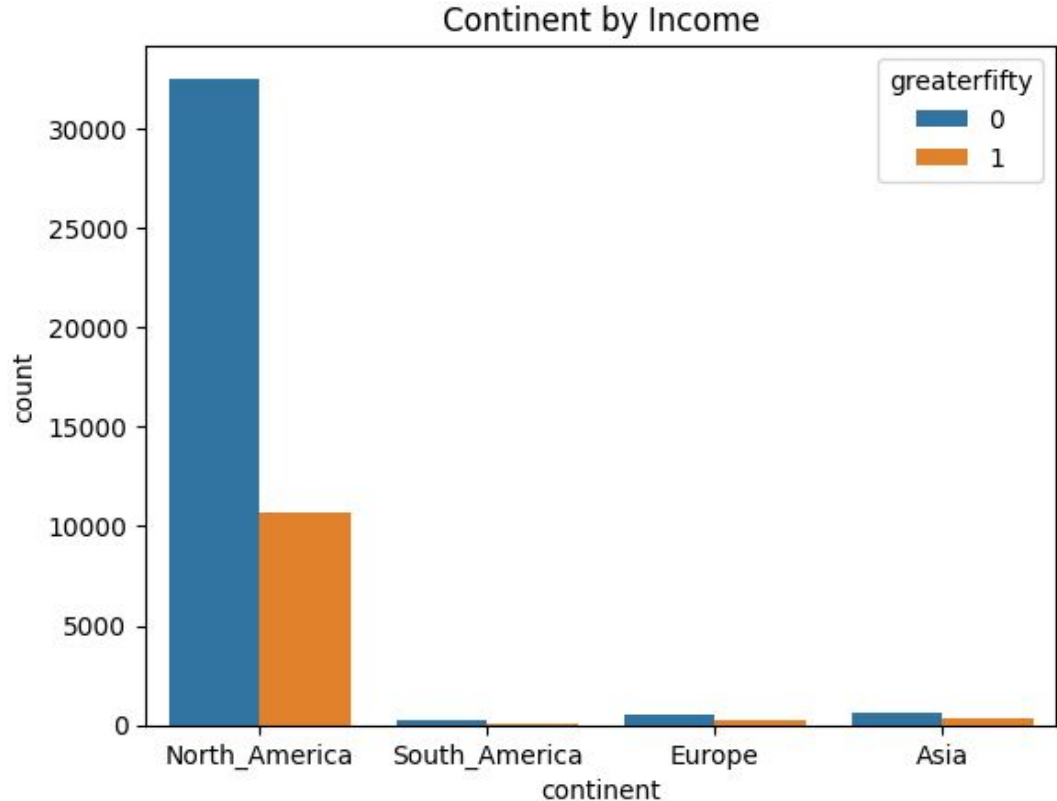
Education

➔ Those with PreK - High School
education have the highest count of 50K
or less.

➔ Bachelors level has the most instances
of individuals earning greater the 50K.

➔ Masters, Prof-School and Doctorate
tend to earn more than 50K.

More education, better opportunities.
However, education availability can vary
depending on the area of the world you are
and can be a hardship for lifestyle and income.
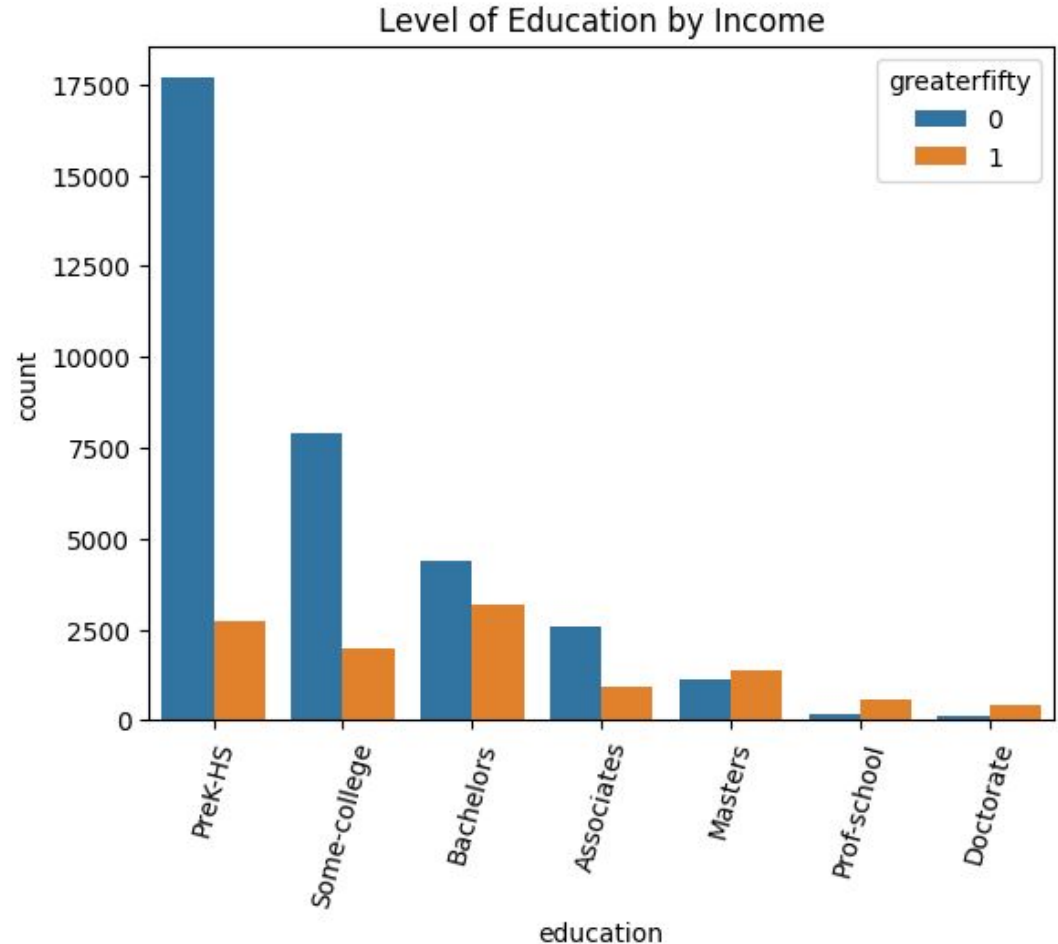
Occupations

- ➔ Males are more common in this dataset than females.
- ➔ Female compensation to be lower in a few fields of work.
- ➔ "Priv-house-serv" is the only occupation where females are more common.
- ➔ Two occupations leading in higher income are "Prof-specialty" and "Exec-managerial".

Number of Occupations VS Income of 50K
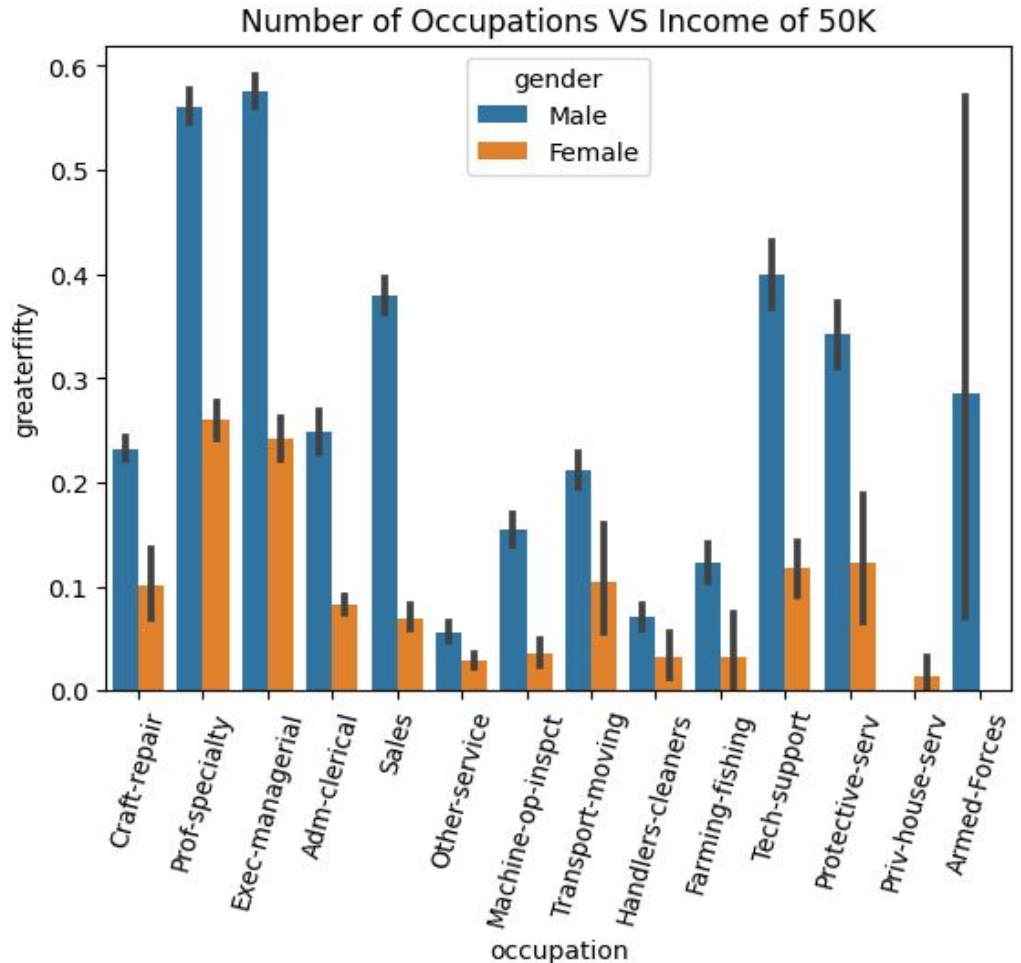
Our Chosen ML Model

➔ The model chosen to predict the class of income is the **Tuned Decision Tree Model**.

➔ This model has an accuracy score of 85.7% and more true positives (income greater than 50K)

➔ All tested models struggled with True predictions.

➔ Chosen model also had False positive which is ideally not wanted.

➔ With an off balance dataset this can be counter balancing since there are not enough instances to accurately predict the minority class.

Test Confusion Matrix
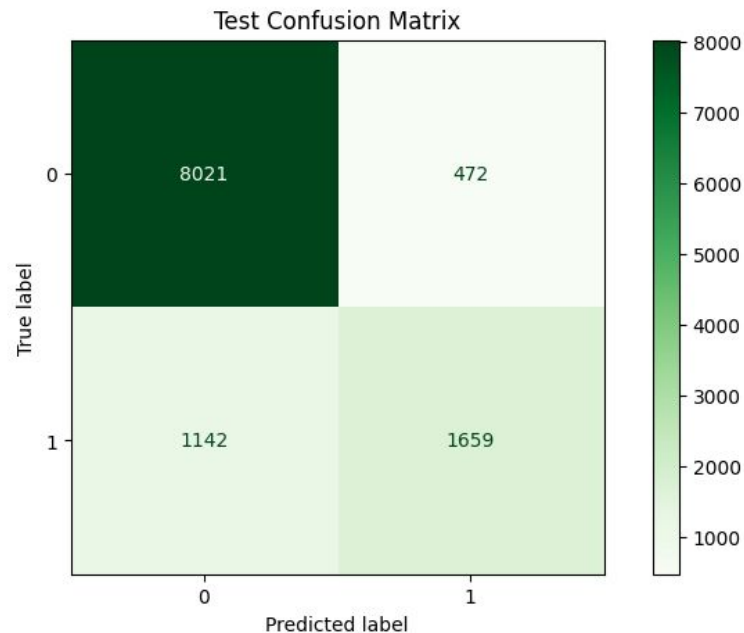
Recommendations

➔ As this dataset is imbalance, we could balance the data with under sampling.

➔ Undersampling would make the 50K or less match the greater than 50K observations

➔ This would be the prefered method, instead of duplicating or randomized additional data to minority class(greater than 50K) to match the majority as that would be great amount of "false".

➔ Evaluate more collected data to improve the model further.