Team 2 Executive Report

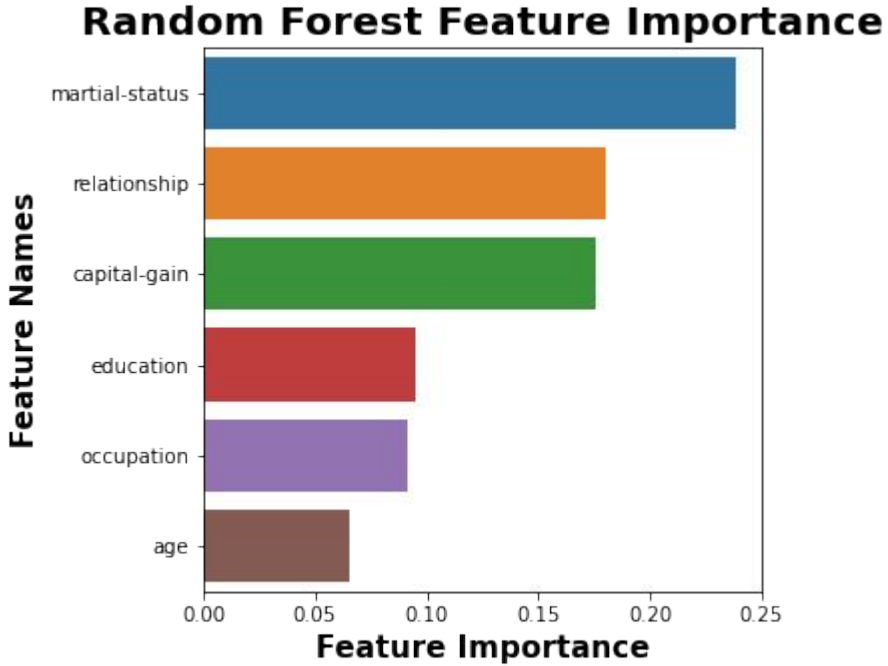
Rosty Hnatyshyn, Kyle Gonzalez, Nicholas Seah, Sang-Hun Sim, Arvind Hasti, Chandana Tarur Veerabhadraiah

Problem Statement

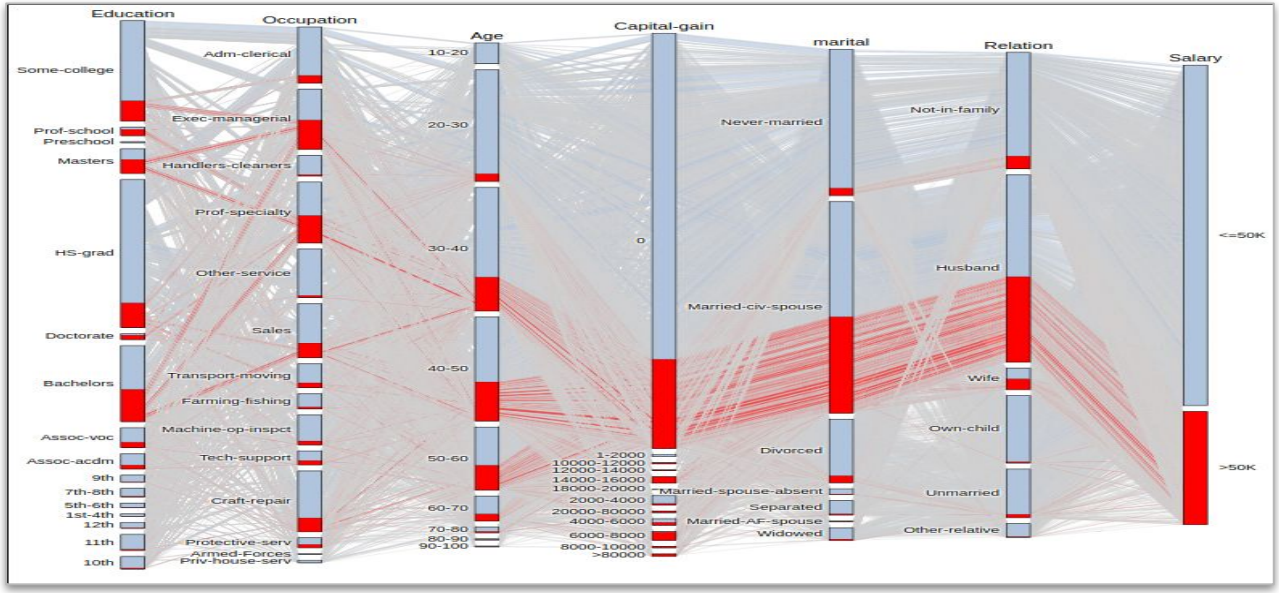
- The purpose of this project is to identify factors that determine an individual's income.
- The data used in this project comes from the United States Census Bureau
- A Random Forest model was used to predict the income of an individual based upon attributes that were provided in the dataset.

Feature Importance

- The model provides variables that are important in predicting a person's income, and we used this list to identify what variables deserved a visualization.

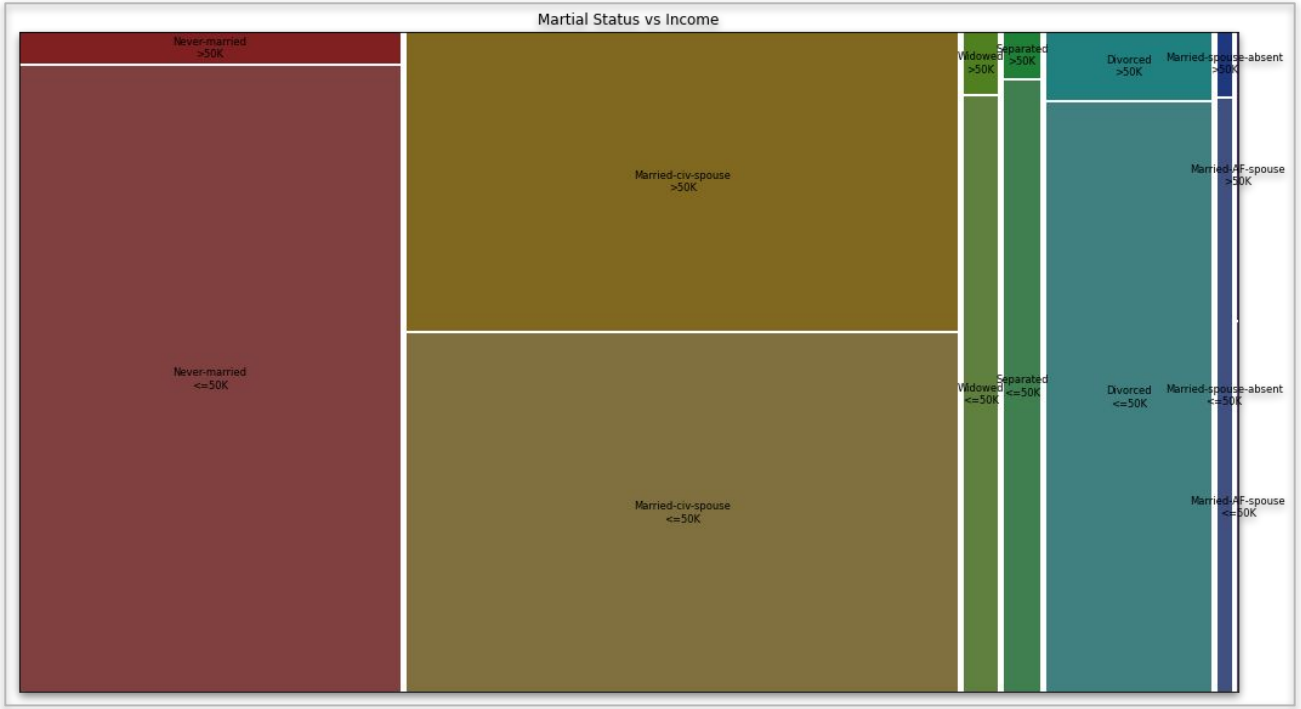


Dataset Analysis: Parallel Set



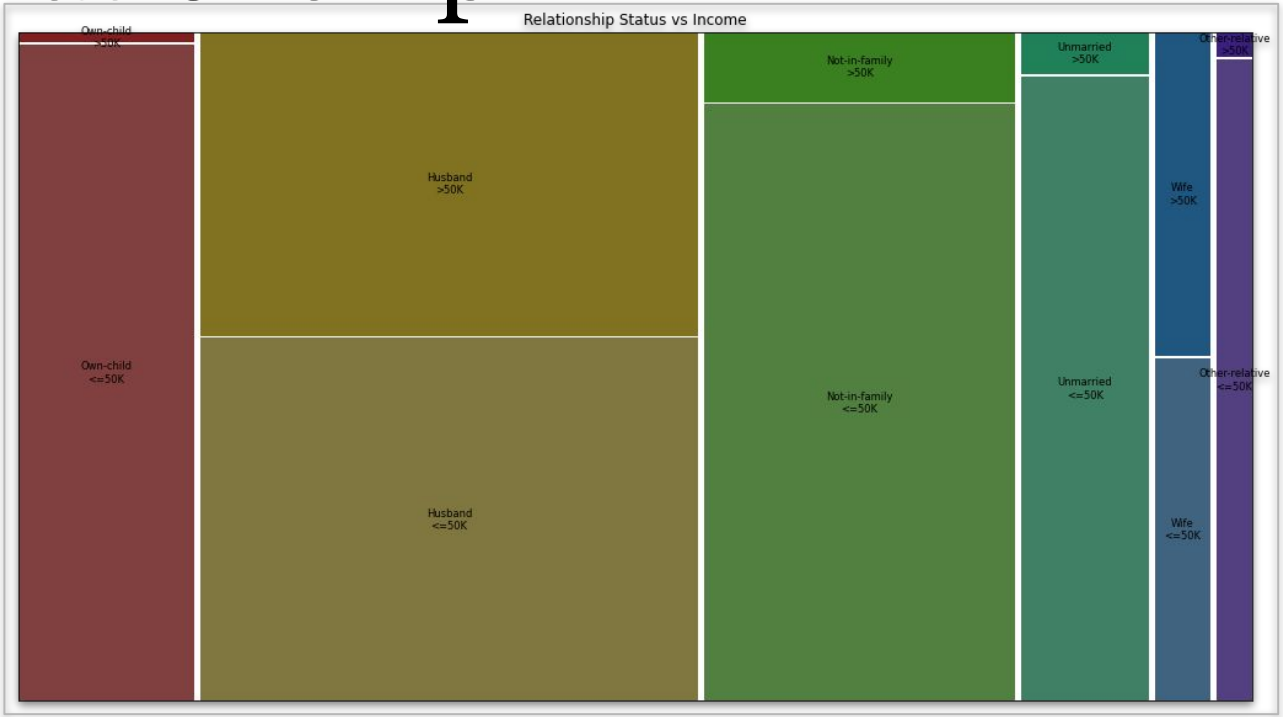
We used the parallel set chart to further confirm what we found in the feature importance chart - this graph clearly shows the groups of people who tend to make over 50K a year. This immediately cuts out a good number of marketing targets for the analysts

1. Marital Status



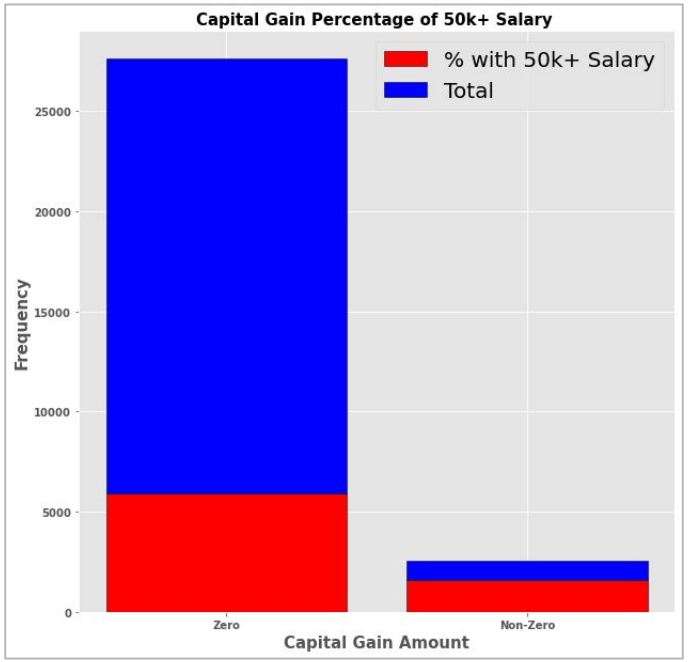
This plot demonstrates some very pure splits in this attribute group: people who are not currently married (never married, divorced, separated, or widowed) tend to make less than $50k a year.

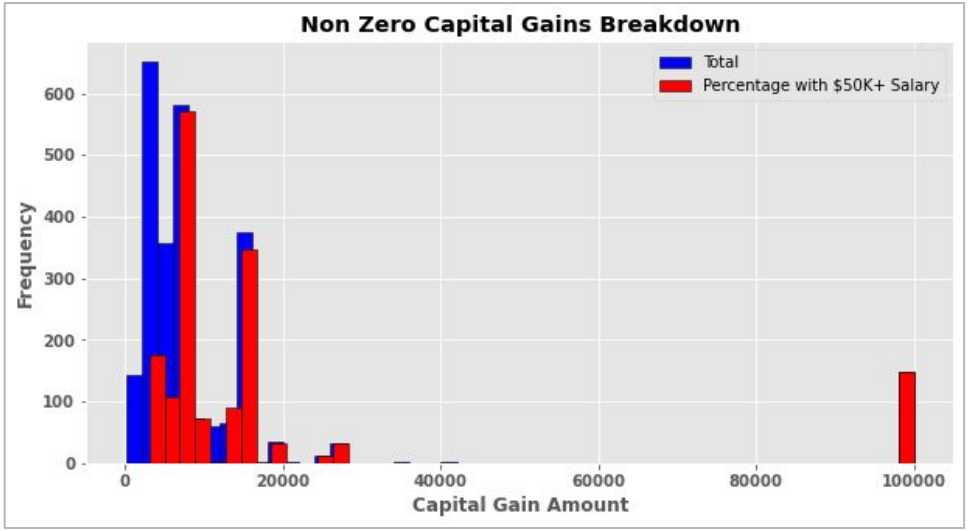
2. Relationship



This plot demonstrates the splits in income for the relationship attribute - unmarried people or people without families should be targeted because they overwhelmingly tend to make under $50k a year.

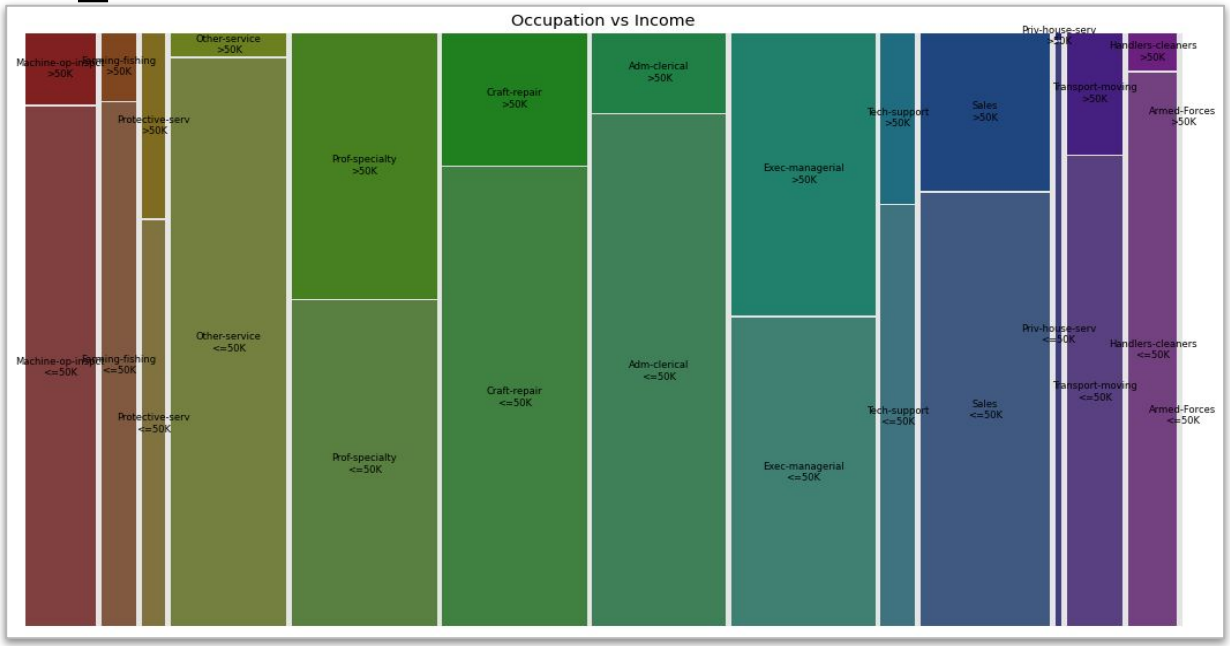
3. Capital Gain





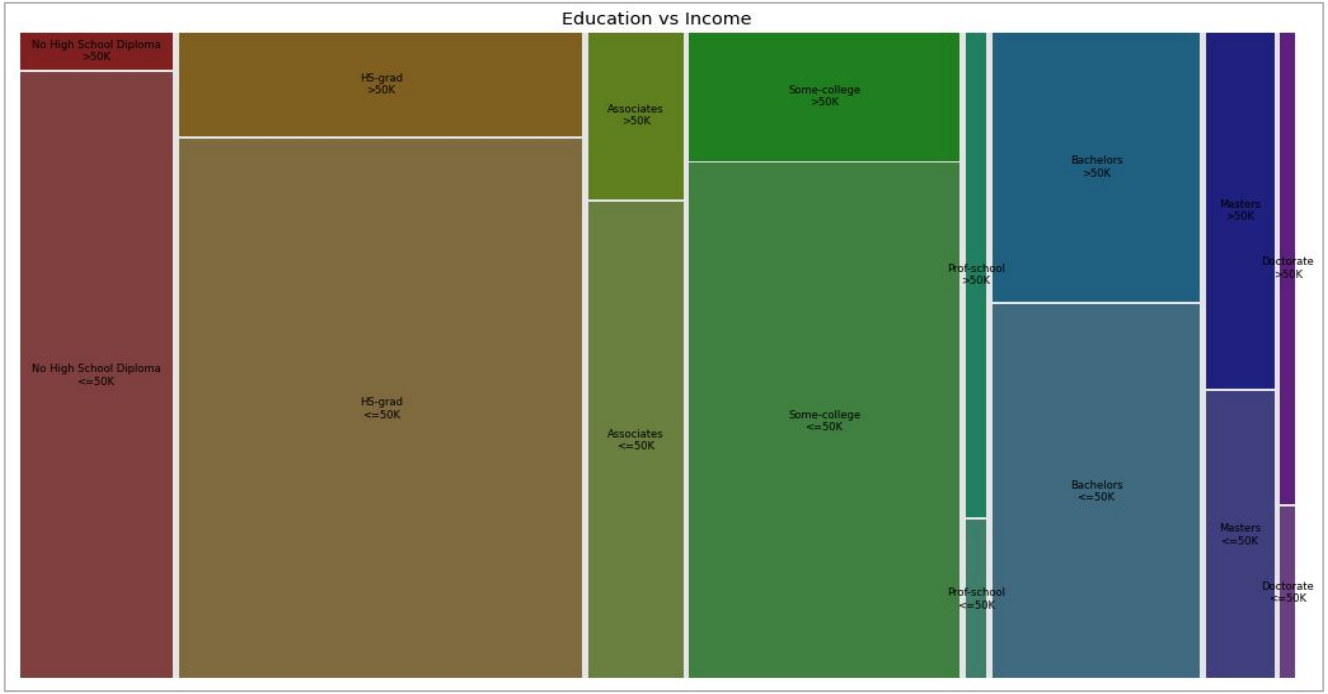
These charts explore the relationship between income and capital gains. It is clear that having some form of capital gains correlates to having a 50k+ salary.

4. Occupation



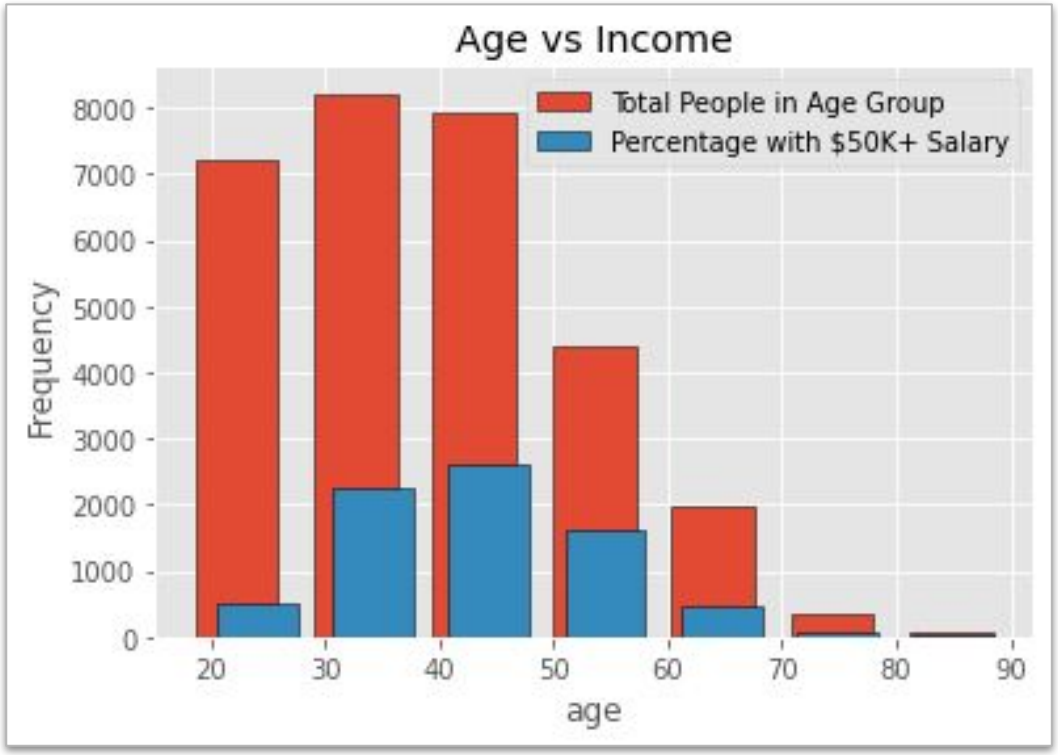
This plot demonstrates the splits in income among different occupations - people who work white collar jobs (management, specialized work) tend to make more than those who work blue collar (factory work, farming, transportation) jobs.

5. Education



We used a mosaic plot to show the relationship between education and income. From this plot we can see that individuals who have doctorate degrees are more likely to make $50k+ per year than individuals without them.

6. Age



This split bar chart shows the distribution of people's income based on their age group. We see that people in their early 20s have the lowest number of people making 50k or more, and people in their 40s have the highest number of people making 50k or more.