Income Class Model

William Chen
Evgeny Didenko
Matthew Hill
August Van Hout

Outline

Problem Statement

Data Wrangling and Cleaning

Exploratory Data Analysis and Visualization

Modeling and Scores

Conclusions

Problem Statement

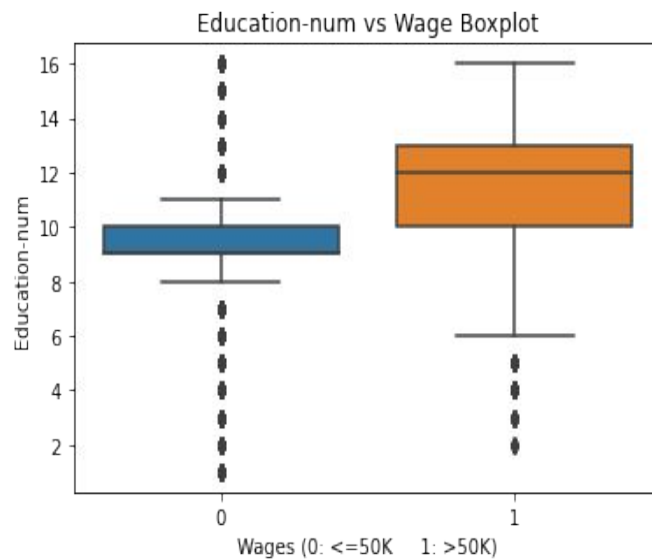- Predicting if a person's income is in excess of $50,000 given certain profile information.
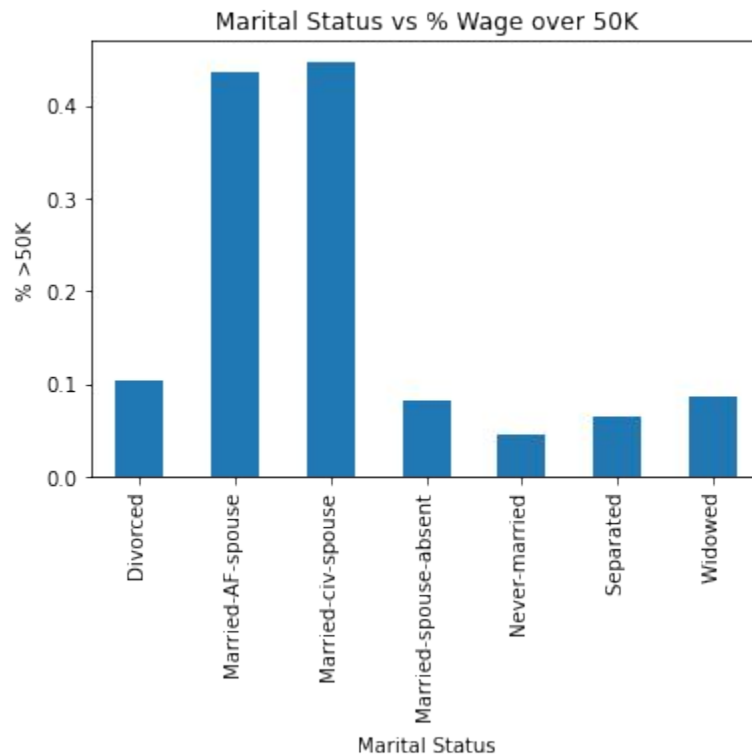
Wrangling and Cleaning

- Replacing Non-Values

- Merge "Capital Gain" and "Capital Loss" to "Capital Change"
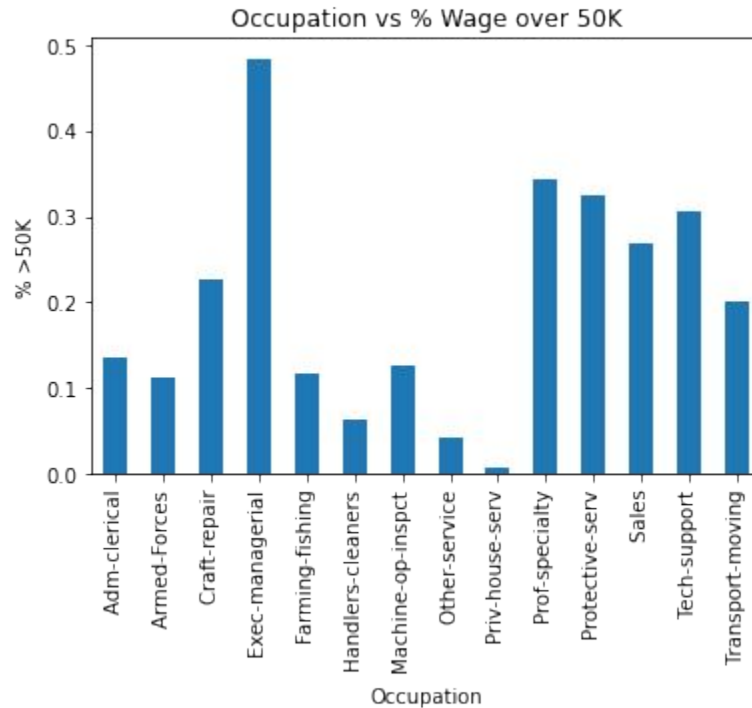
- Changed Native Country to US / Non-US
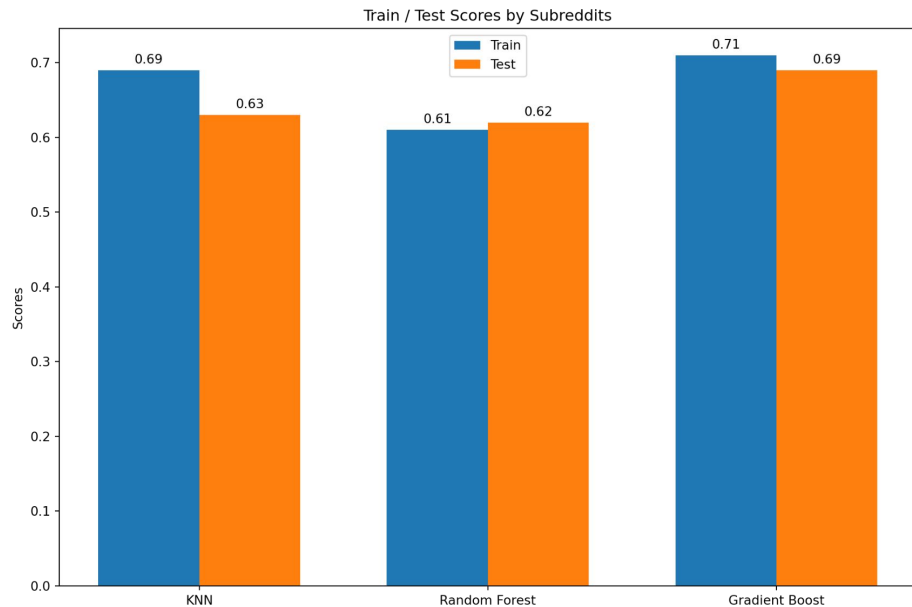
EDA and Visualization

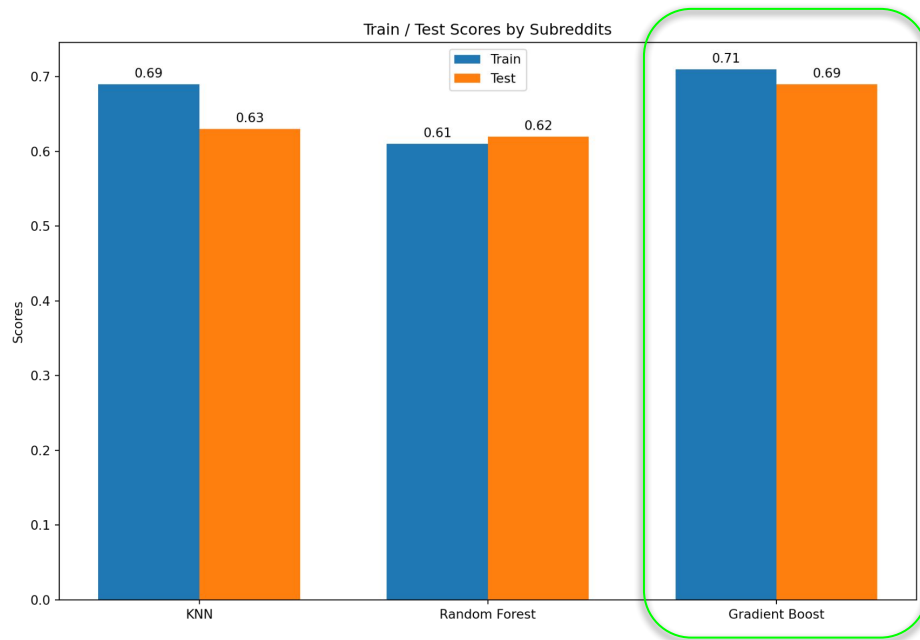Education-num vs Wage Boxplot

EDA and Visualization

Marital Status vs % Wage over 50K

EDA and Visualization

Occupation vs % Wage over 50K

Modeling and Scores - F1 Scores

Train / Test Scores by Subreddits

Modeling and Scores - F1 Scores

Modeling and Scores - Features

Features vs Gradient Boost F1 Score