Capstone Project #2: Predicting Future Financial Well-Being using CFPB Survey Data

By: Justin Hughes-Coleman

Problem Statement
“Where does all my money go?”

A common issue in modern times is the lack of accessibly sound financial knowledge. If you have a financial advisor or are very financially savvy this isn’t an issue but if you are unsure about credit card minimum payments, this can have a generational impact on your finances.

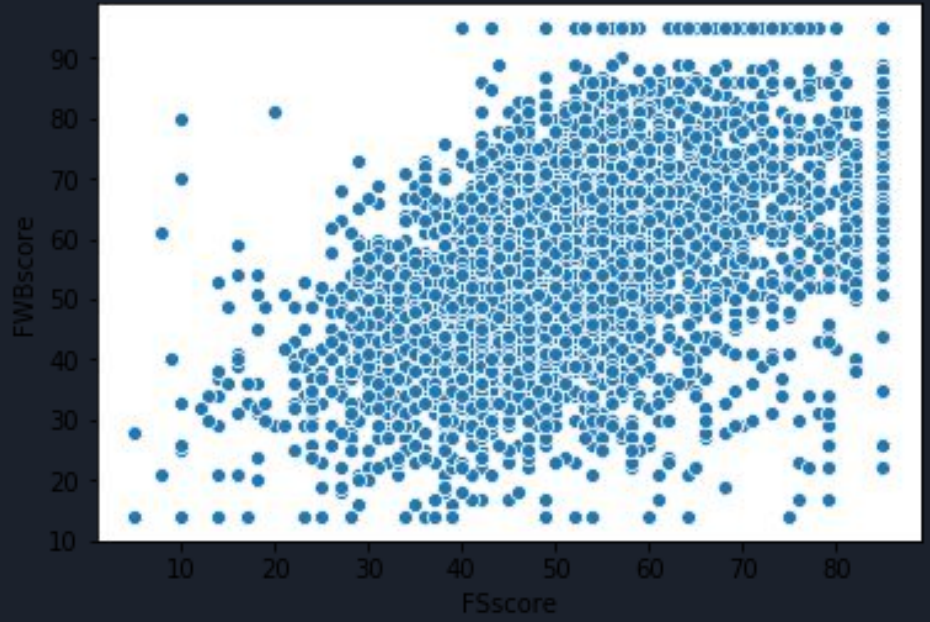
- Using data from the Consumer Financial Protection Bureau, identify the main factors that determine financial well-being
- Develop a model that can later be used by users to map their financial well-being

Data Description

- Dataset was from the Consumer Financial Protection Bureau (CFPB)
- CFPB administered a Financial Well-Being survey that would ask its participants various financial literacy questions in order to gain understanding.
- The dataset also includes exhaustive socio-economic descriptors of the participants
- Dataset also provided percentage statistics for each variable.

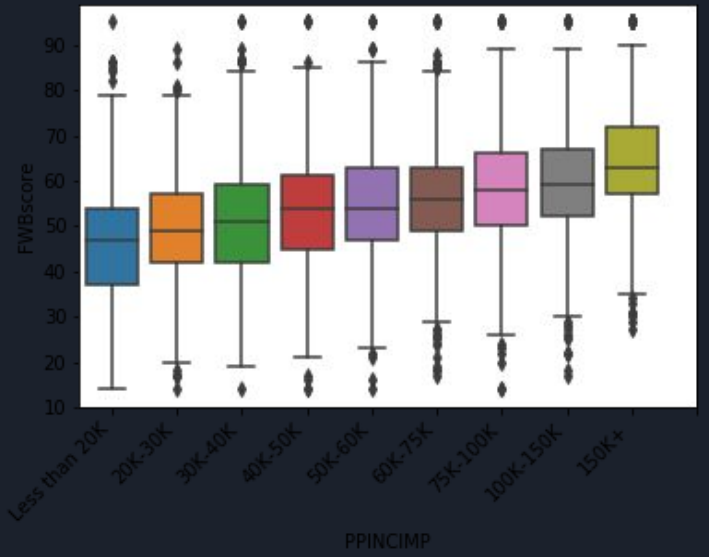
Exploratory Data Analysis

- Dataset has strong positive correlation between the continuous variables
- The financial literacy questions seem to have the most correlation
- Household income showed a surprisingly high amount of outliers that score low on financial well being but were making above average salaries.



Statistical Analysis

- The financial literacy questions regarding self-assessment proved to be the most statistically significant.
- The coefficients on these questions were the highest without raising the p-value
- Income showed a general positive correlation, more than other socio-economic features



Feature Selection

- 31 features out of the original 217 were selected for model training and testing.
- These were the most impactful features based on their p-values < 0.05 (alpha)
- The coefficient of each of these variables greatly outsize the others
- Surprisingly, a lot of the features are not socio-economic but rather financial literacy indicators

Results

- The OLS model performed the best after feature extraction
- Further development into a user interface so participants can get a comprehensive materials to aid in the financial literacy journey

	R-Squared	RSME
OLS Regression(before FS)	0.996	11.03
OLS Regression(after FS)	0.996	11.13
Linear	0.40	11.27
Ridge	0.41	N/A
Neural Network	N/A	10.59

Thank you for reading!