Complete Analysis

By: Andrew Varela

What Problem Am I Solving?

- Original Proposal: Identify market factors that have established effects on homelessness.

- Additional Proposal: Identify market factors that have established effects on homelessness with a scaled version of my Y-variable rate_homeless.

What Data Did I used?

- The project utilizes a data set from Market Predictors of Homelessness

- The data set consists of various housing, economic, safety net, demographic, and climate factors.

What Analysis Methods Did I Use?

Multiple Linear Regression
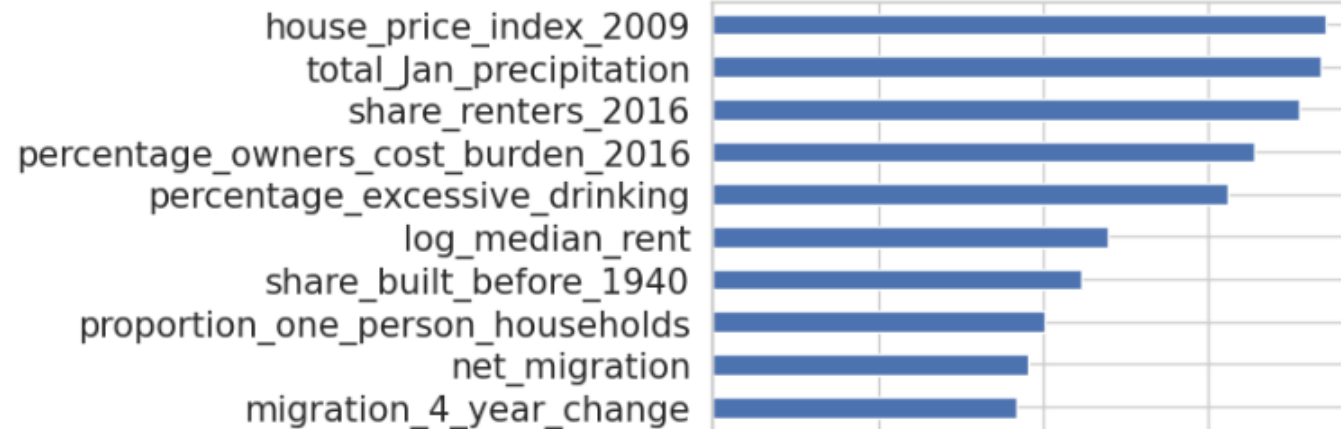
Tree-based methods and XGBoost

Cross Validation

My Analysis

Original Regression Model: R-Squared -> 55.5%
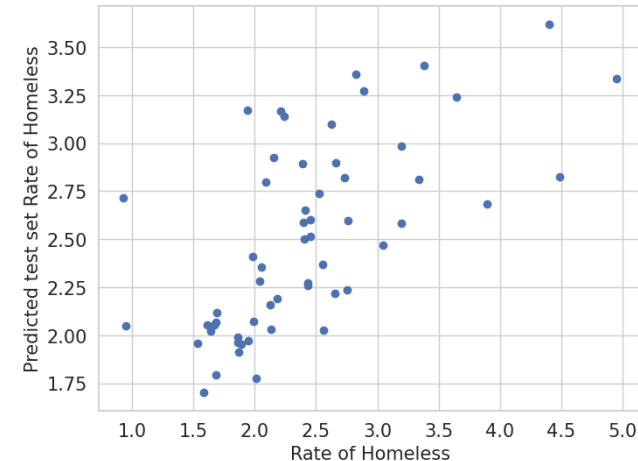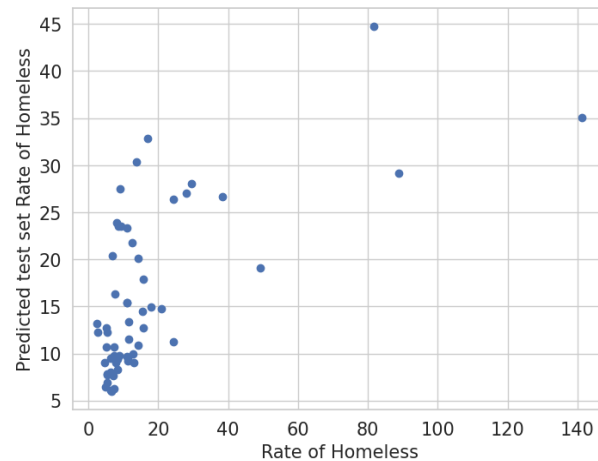
Best regression model: XGBoost

Best Variables by Importance:

Mean Squared Error: 18.482

My Analysis, Additional Proposal

- R-Squared -> 58.7%

- Best Regression Model: Original Model

- Best Variables by Importance: proportion_one_person_household, utility_costs, log_median_rent, share_renters_2016, and total_Jan_precipitation

My Conclusion