

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light greenish-blue. They are positioned diagonally, with the blue one partially covering the green one.

# Housing Data Predictions

By Hammad Afzal



# The problem

What features affect a houses price the most and what model best uses these features to predict price?



# EDA Process

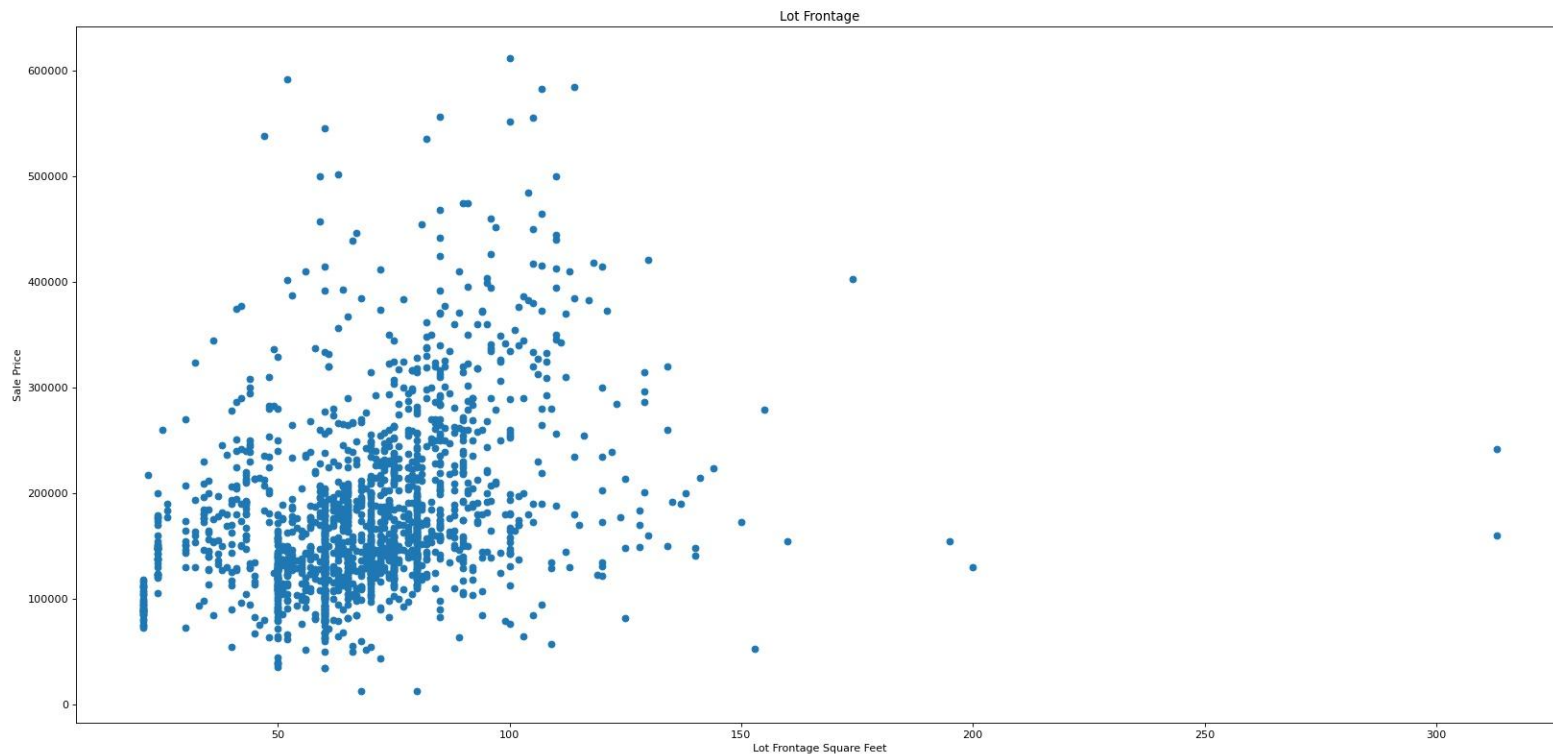
Cleaning decisions

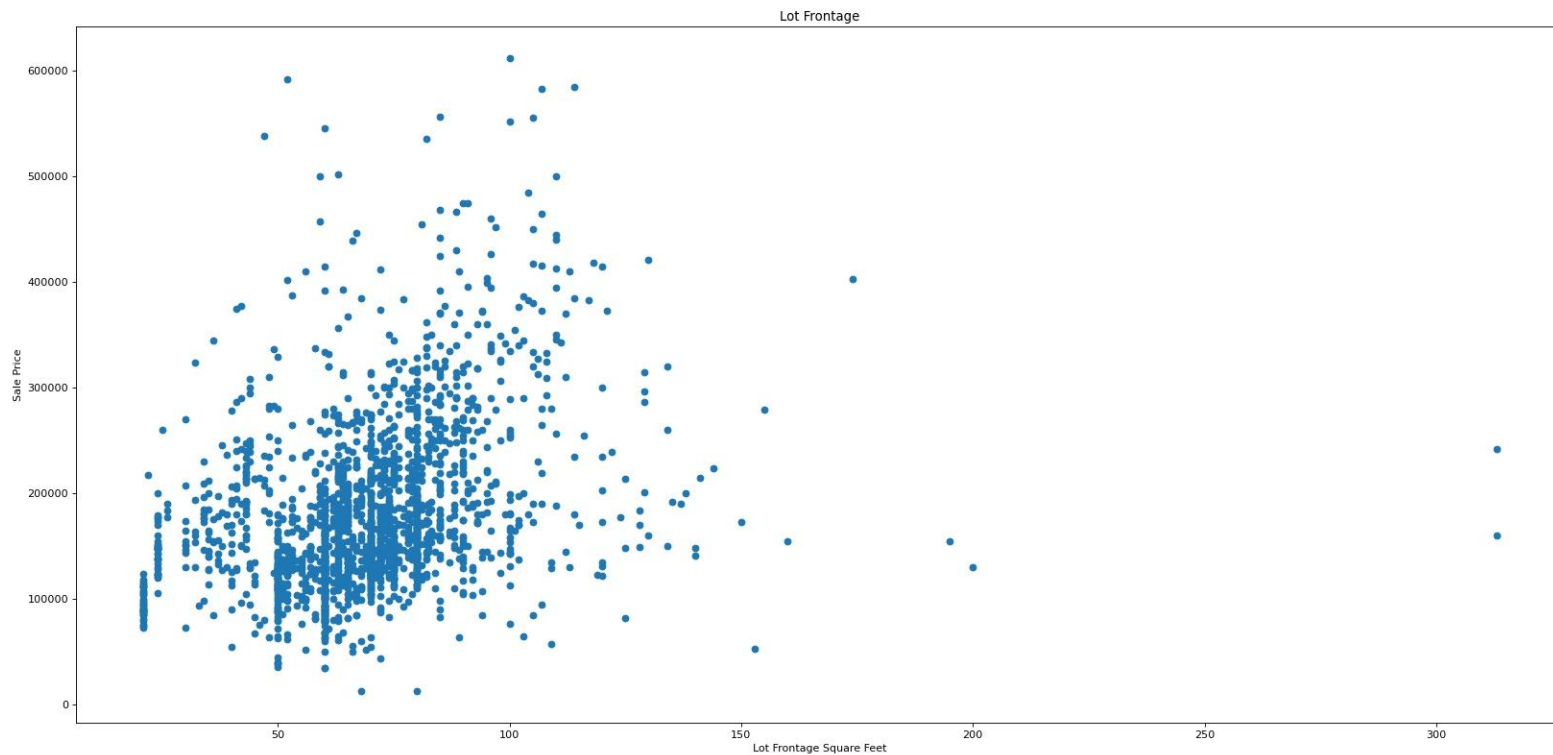
Imputing decisions

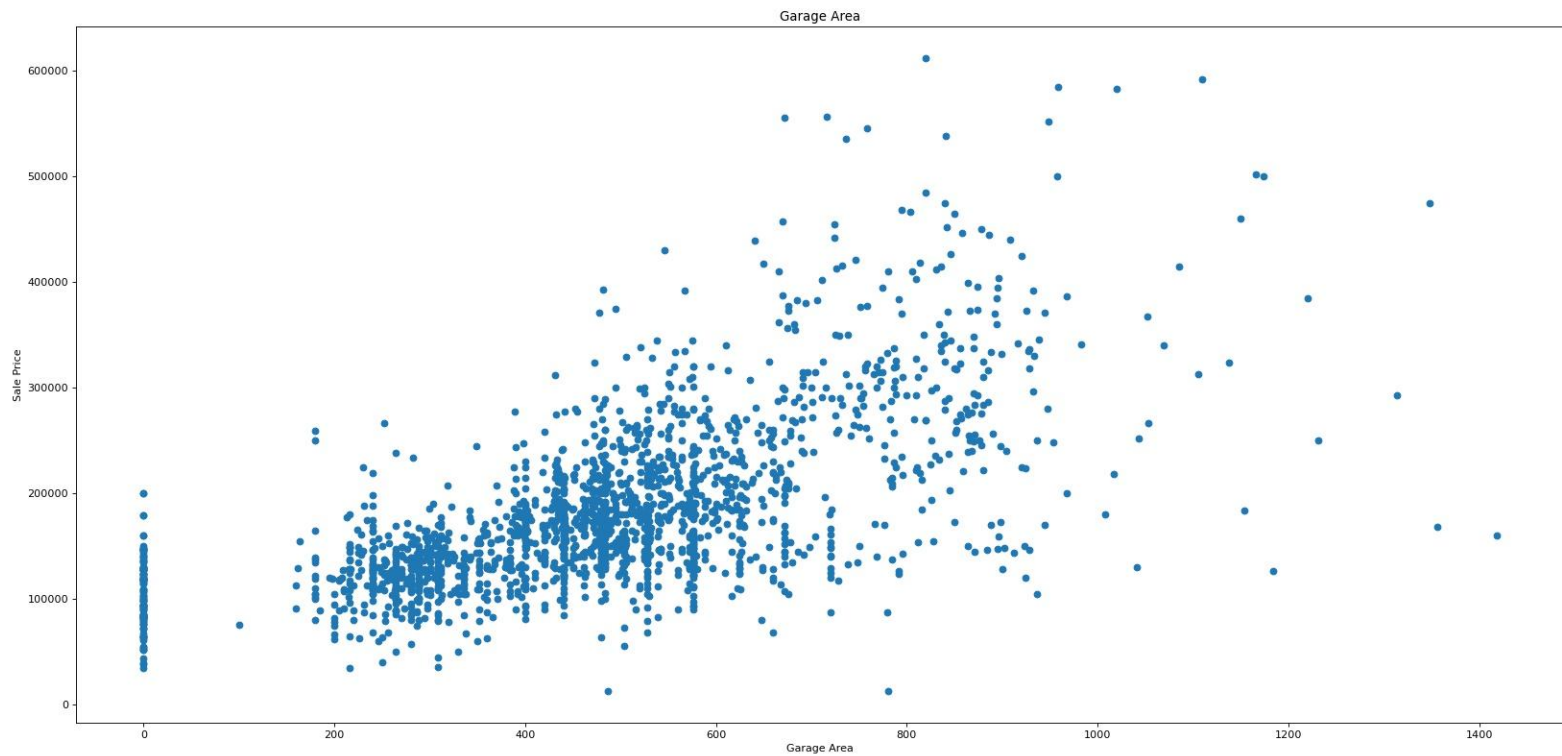
Most valuable features in predictions

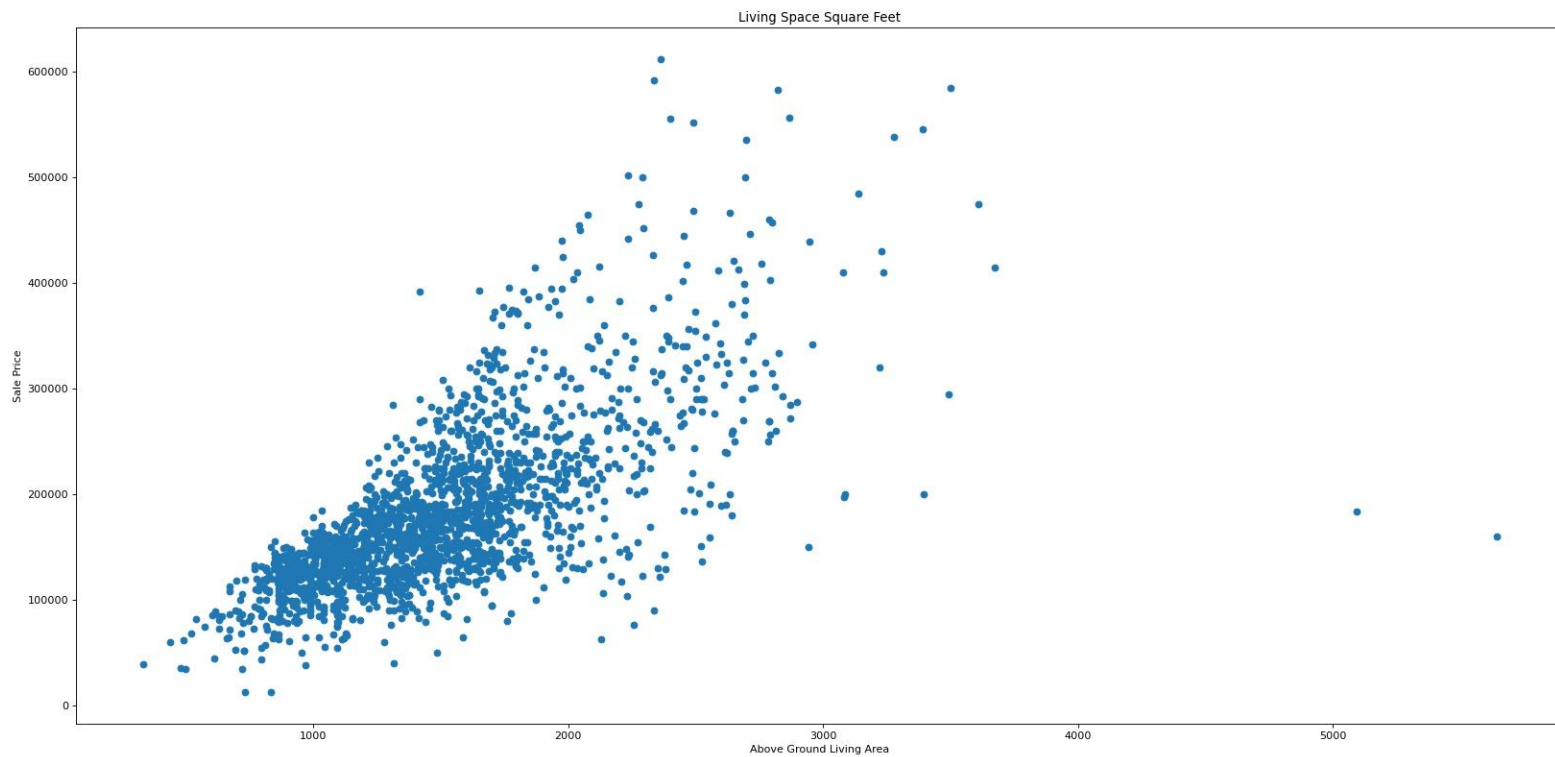


NA does not mean missing data!

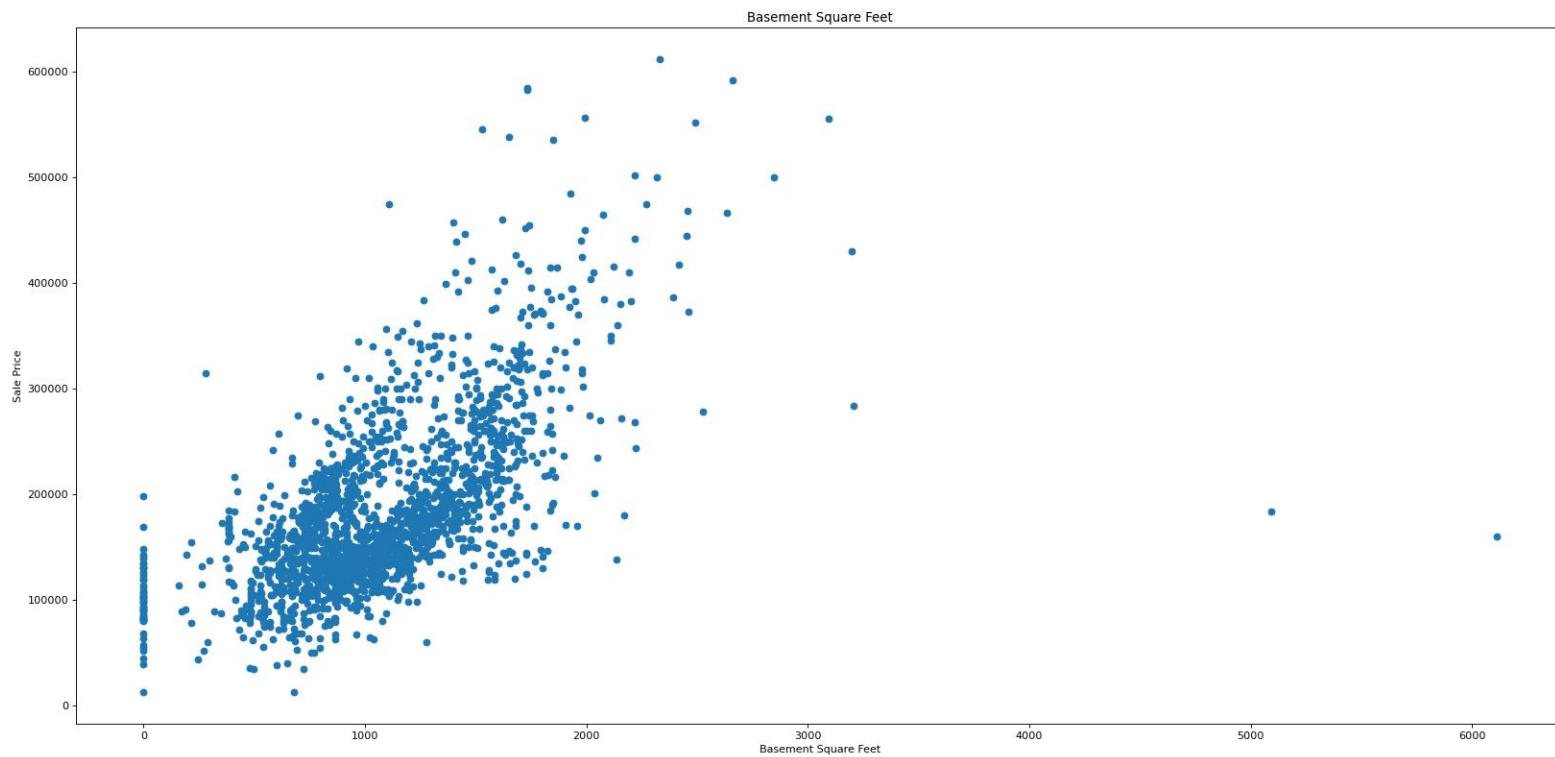


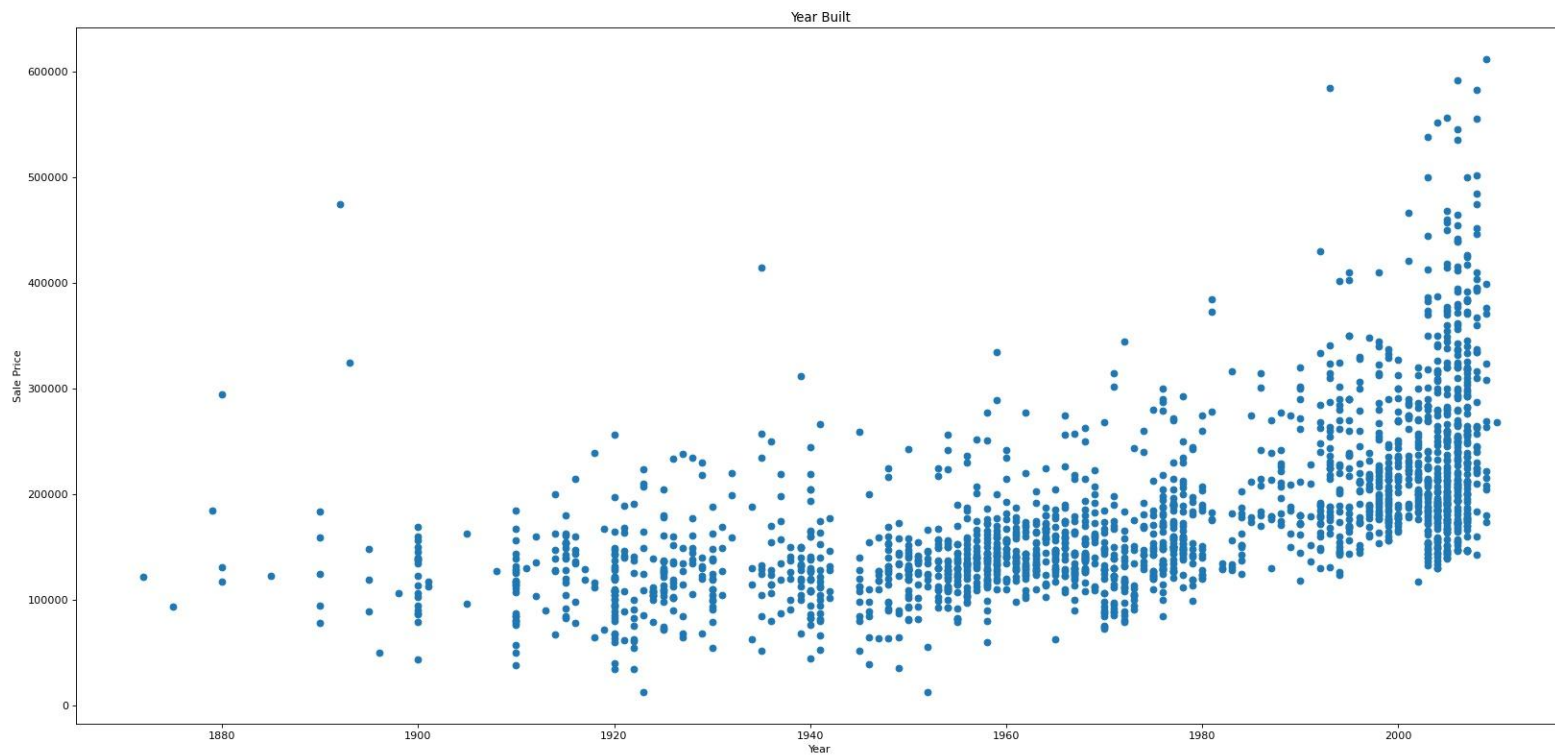


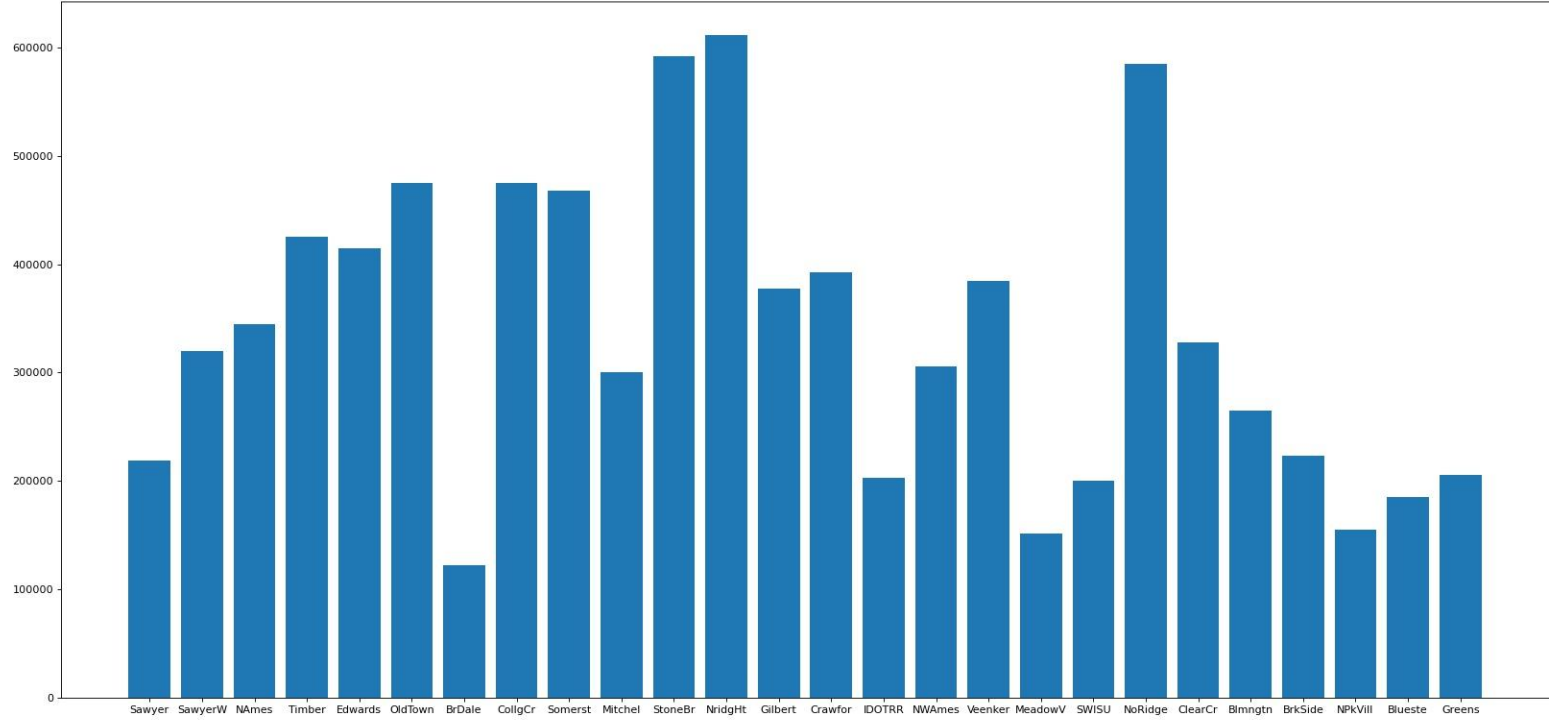














# Homeowner improvement

Establish good heating

Raise quality of kitchen

Remodel home



# Model Fitting

The models were Linear Regression, Lasso, Ridge, and Elastic Net



# Elastic net was the best

Lowest cross val score was higher than the others lowest score



Still a lot to do!