Predicting the price of a house in Ames, Iowa
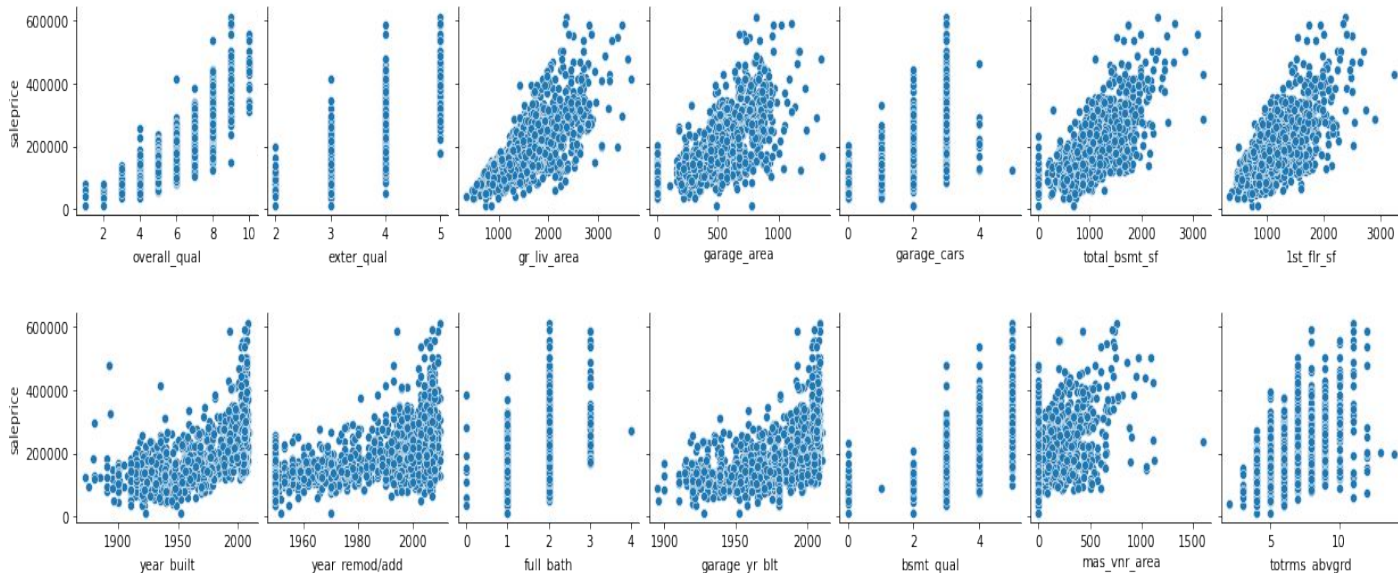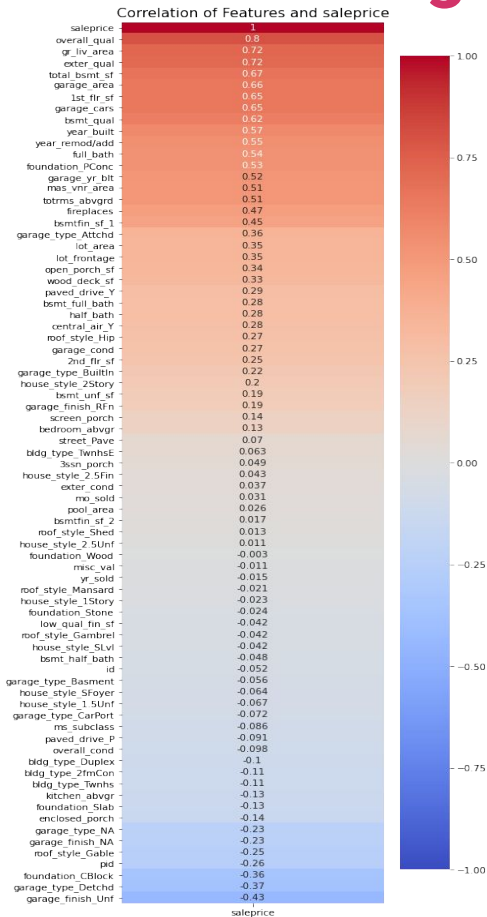
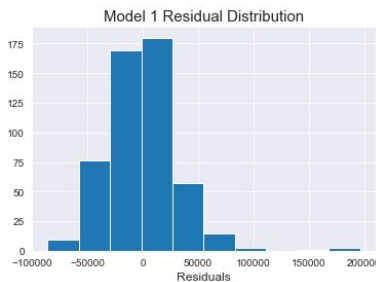Presentation by: Aaron O'Neal

Goal of Project

I have been hired by a real estate agency in Ames, Iowa to help them better understand the housing market in their area. I have been asked to help them develop a model that will accurately predict the price of a house at sale, so they can provide their potential sellers/buyers with more insight.
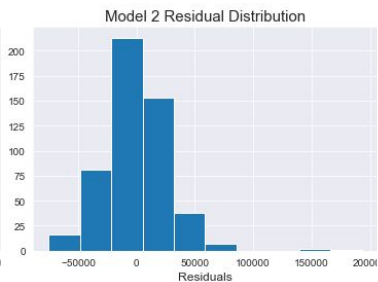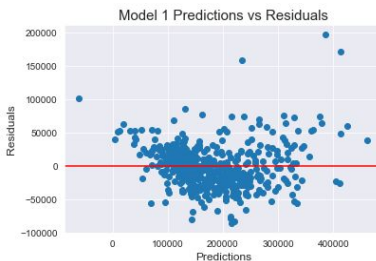
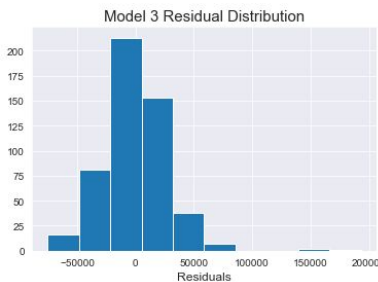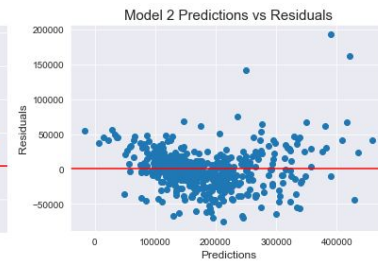# Determining Which Features of a House are Important
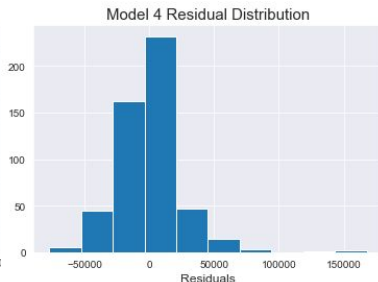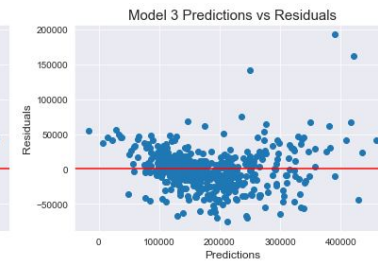
Prediction Models and Their Performance

Model 1 - baseline (>=0.5 corr)
Linear Regression
Cross-val score ~0.8474

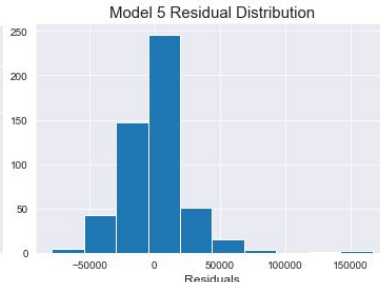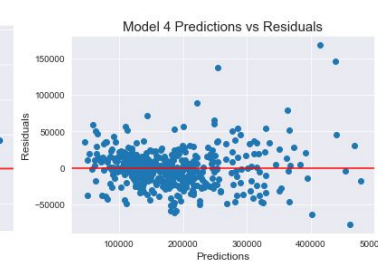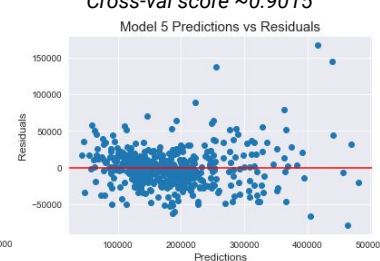Model 2 - all numeric features
Linear Regression
Cross-val score ~0.8797

Model 3 - num. features scaled
Linear Regression
Cross-val score ~0.8796
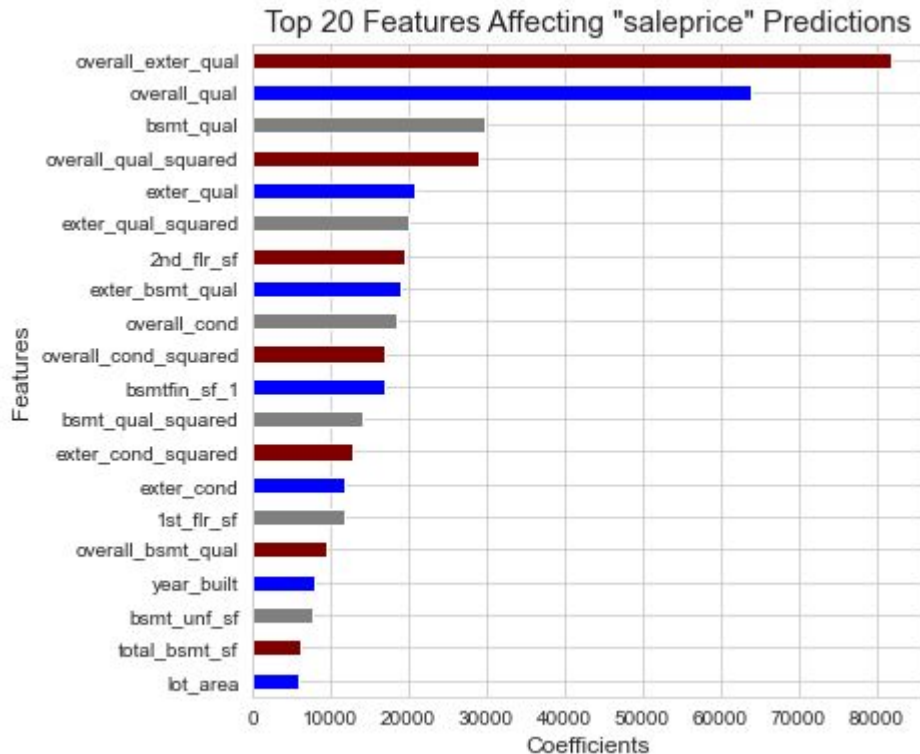
Model 4 - feature interactions
Linear Regression
Cross-val score ~0.9023

Model 5 - all numeric features and
feature interactions
Lasso
Cross-val score ~0.9015
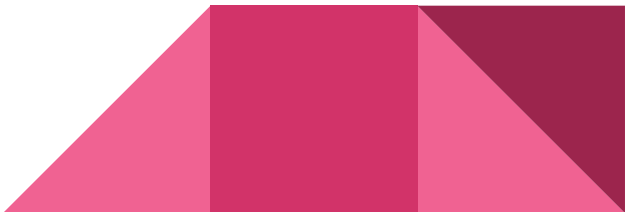
The Lasso Model

Top 20 Features Affecting "saleprice" Predictions

- This model showed the features that best help us predict the price
- It also showed the features that had little effect
- If I were to create another simpler "model", I would reference this Lasso model to remove some of the features that had little effect on the price of a house at sale.

Recommendations

- I would recommend that the real estate agency use my 'Model 4' linear regression model to help them better understand the housing market in Ames, Iowa. They will be able to estimate the price of a house at sale with this model explaining over 90.23% of the variability.
- I would also recommend creating another simpler model based on removing the features with coefficients of 0 or extremely close to 0 in the Lasso model
- Final recommendation is to get more data on higher priced houses to better understand some of the outliers in the data

Questions???