MODEL OF AMES IA HOUSING PRICES

By: Nathan Perry

HOW WAS THE DATA HANDLED

- Data in the model is from train.csv
- Null values were filled in using the data dictionary where possible
- Otherwise the values were imputed with the mean of the feature
- Outliers were removed from the sale price column.

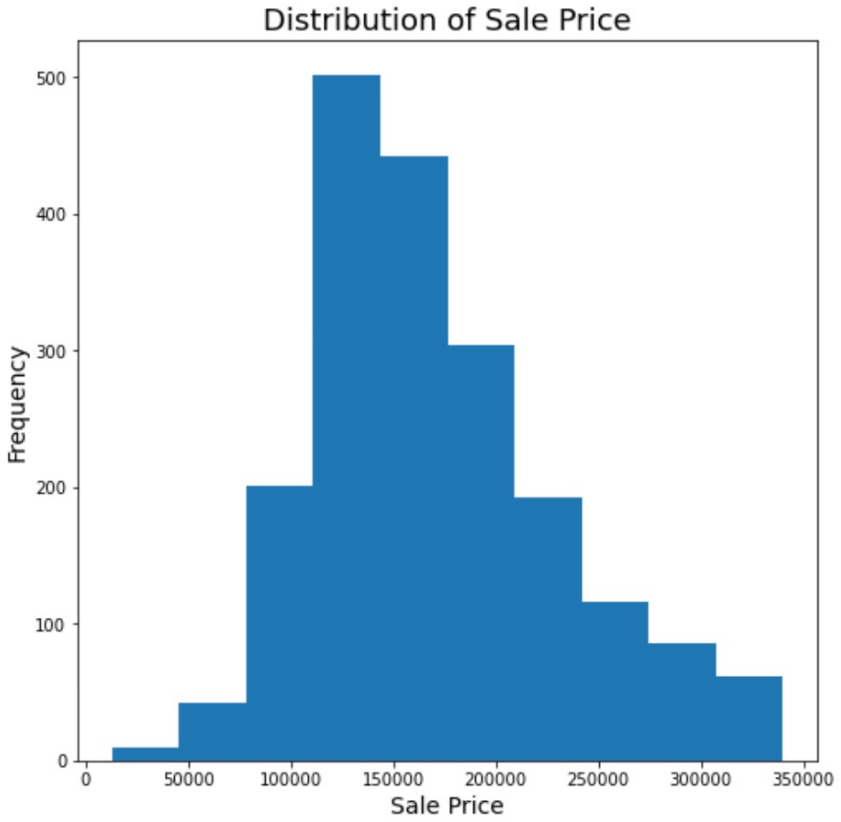
PROBLEM STATEMENT

- Make recommendations on maximizing the value of a home
- To do this I will make a model to predict the price of a home based on a multitude of features
- I am looking for statistically significant features, especially those with a larger effect on the price
- I am also hoping that my model will be able to predict future prices

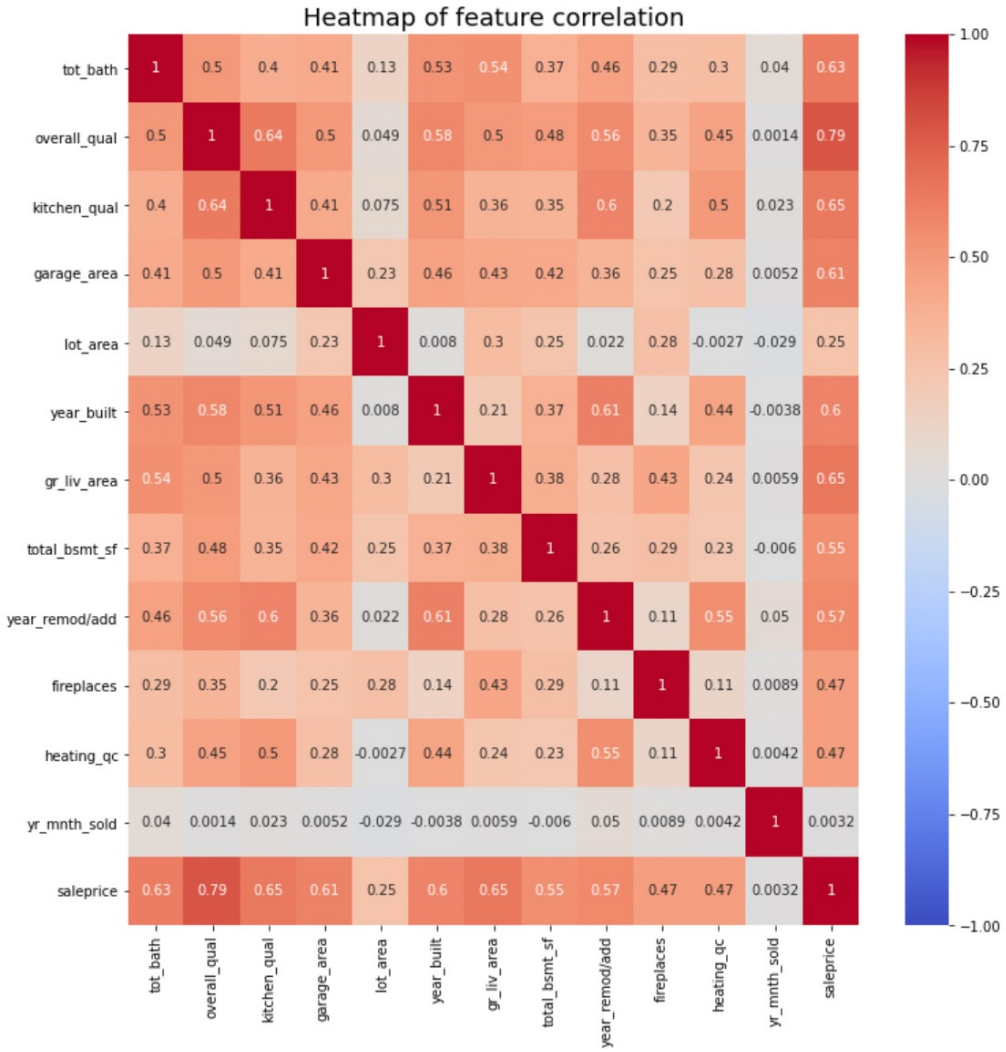
HYPOTHESIS

- Predicted top factors:
  - Time of sale
  - Number of bedrooms
  - Number of bathrooms
  - Square footage of the house
  - Lot size
  - Garage size

WHAT IS THE DISTRIBUTION OF HOME PRICES IN THE AMES AREA



WHICH FEATURES ARE CORRELATED WITH PRICE

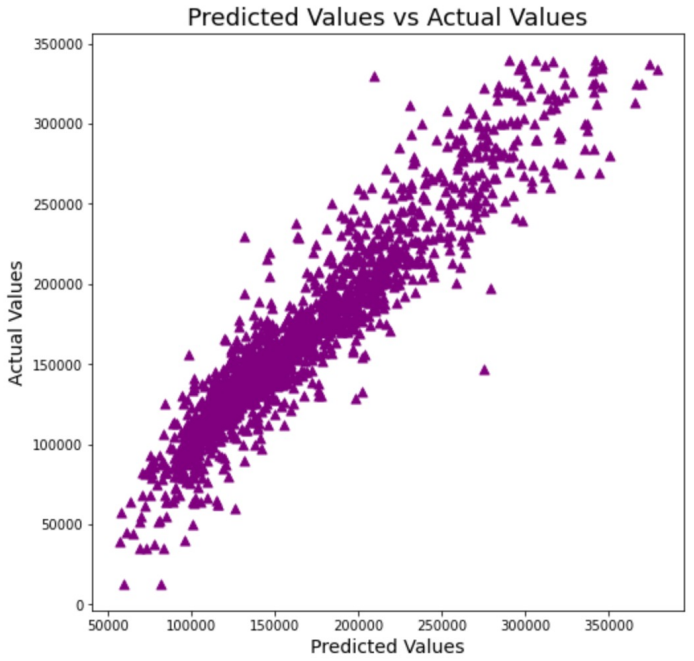


HOW WAS THE MODEL CREATED?

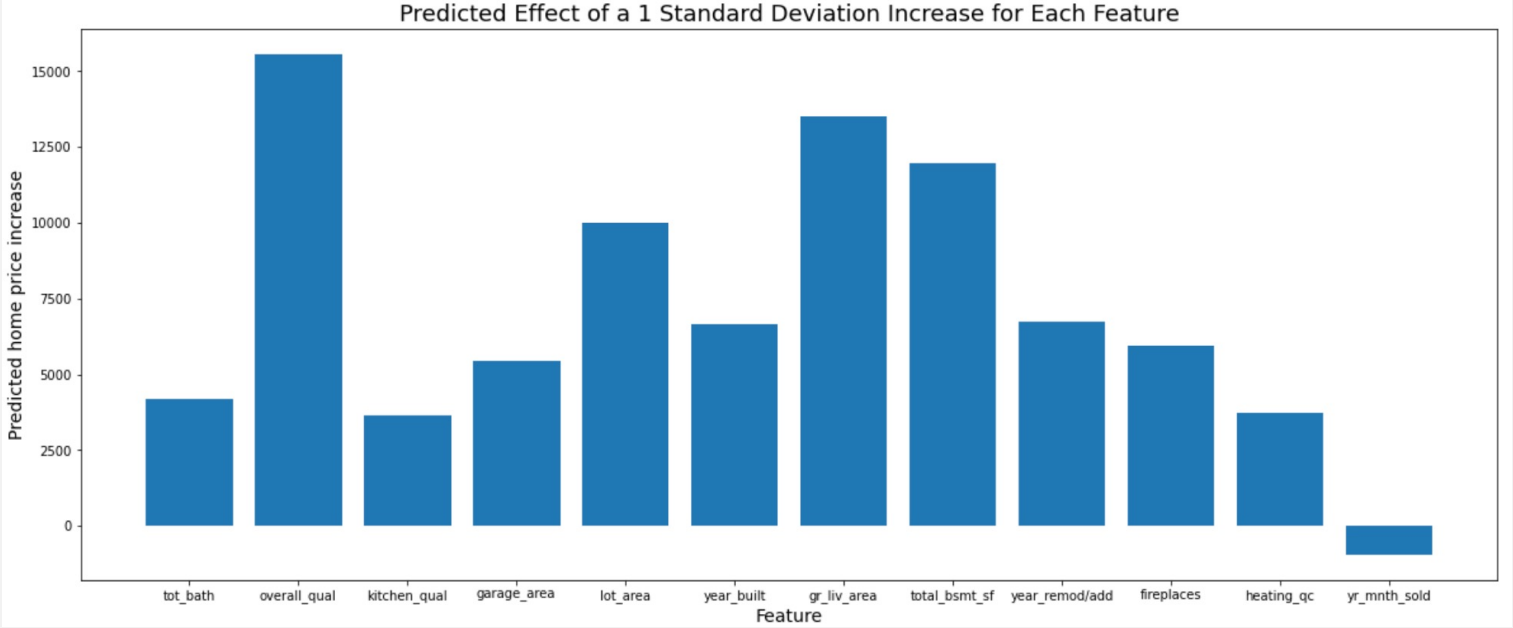
- The selected features were put through polynomial transformation and standardized
- Lasso regression was used to remove insignificant factors
- The natural log of sale price was used when fitting the model

HOW DOES THE MODEL PERFORM

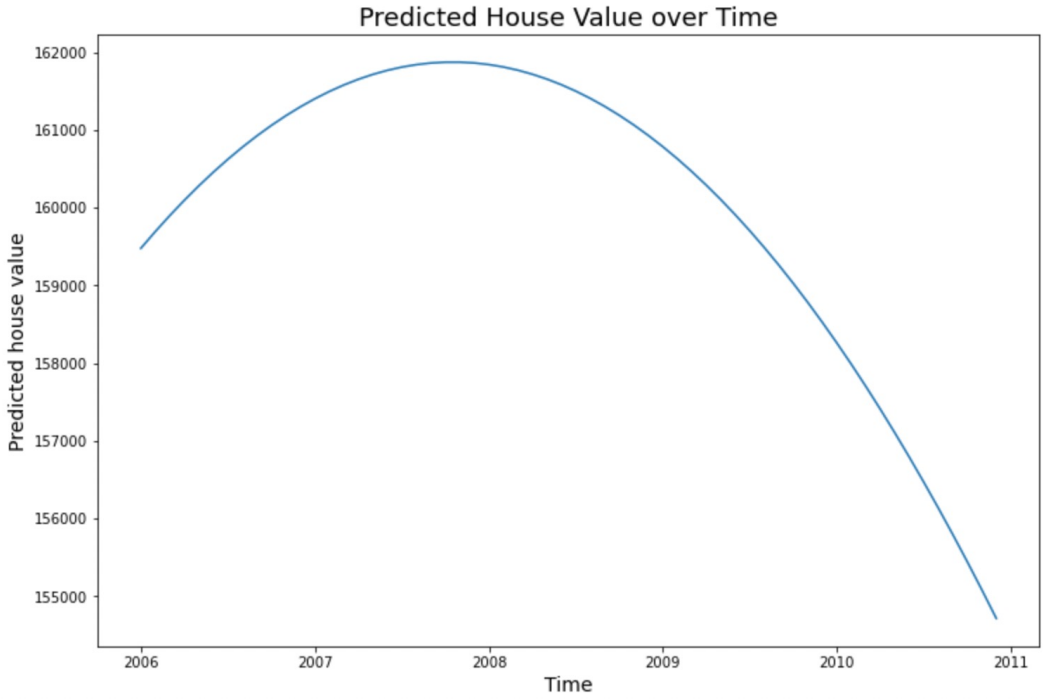
- R^2 on train data: .844
- R^2 on test data: .824
- RMSE on train data: 20695
- RMSE on test data: 25005



WHICH FEATURES HAVE THE GREATEST EFFECT?



THERE IS A STEEP DROP IN PREDICTED HOME PRICE AROUND 2008



CONCLUSIONS

- This model cannot be used to predict future prices
- A well built and maintained home will have significantly higher value than average
- Home and lot size are also main drivers of home value
- Older homes are less valuable, but remodeling can make up much of the lost value