HOUSING PRICE PREDICTIONS IN AMES, IOWA

Abhay Aanabathula

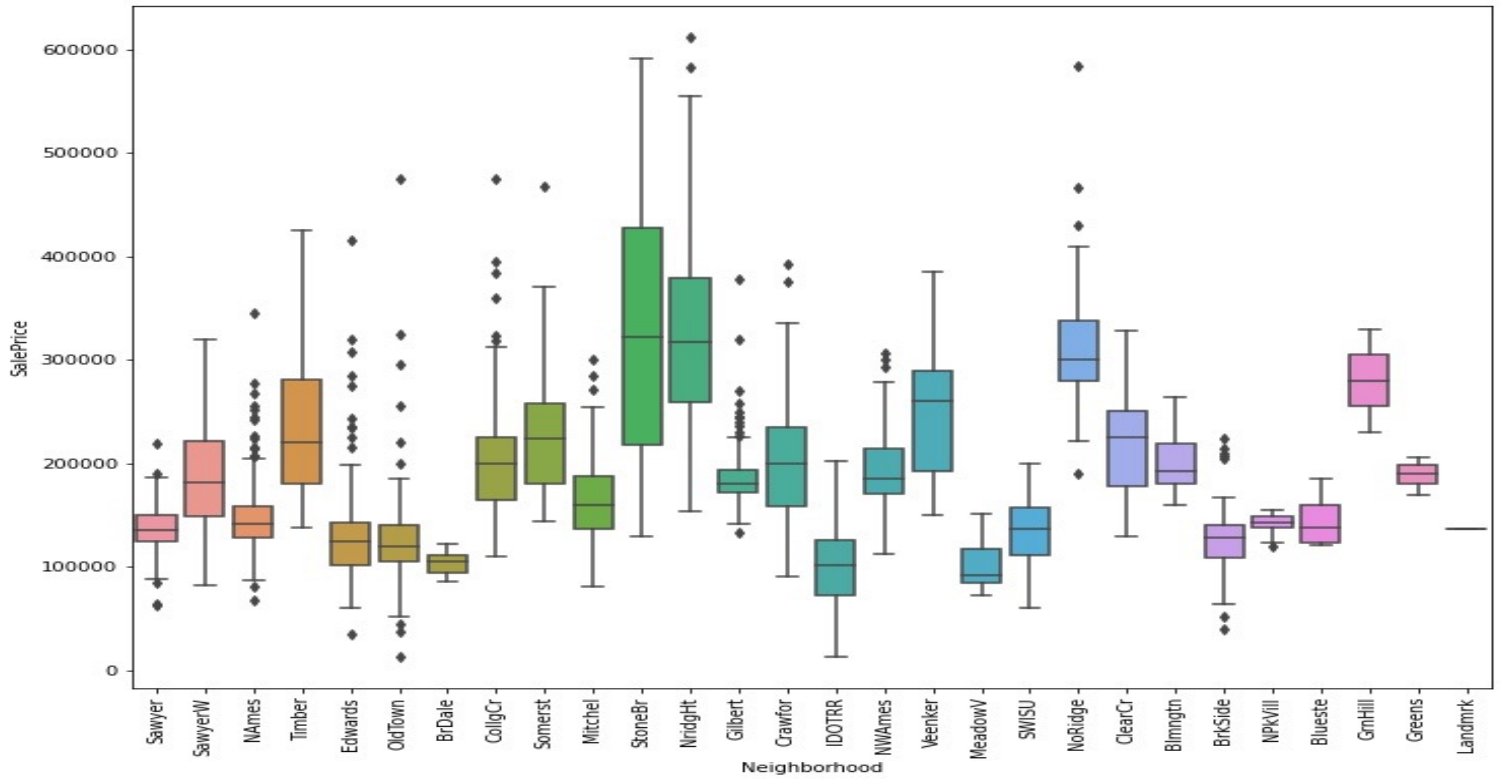
BACKGROUND

- Population of more than 65,000 residents
- Home of Iowa State University with a student population of 35,000
- A growing city!
  - Business relocation and expansion
- No. 8 for "Best Towns for Millennials in America" by Niche Ranking
- Named the healthiest city in America by USA Today

MY ROLE AND GOAL

- My realty firm needs to stay ahead of the competition!
- Home needs to fall within client's budget but also have the features the customer desires
- New homes are coming on the market all the time
- Predicting the prices of these homes will help improve customer service and potentially increase closing rates for the firm

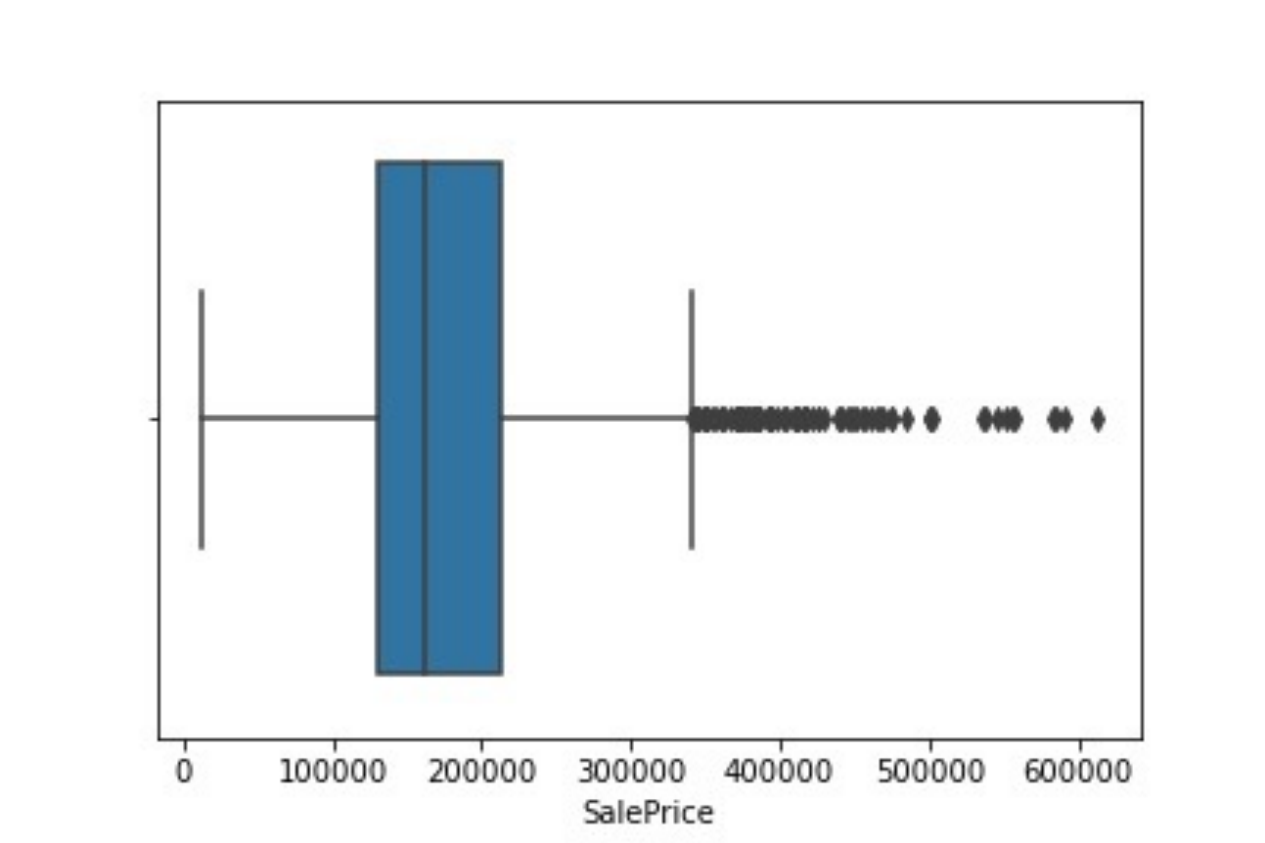
NEIGHBORHOOD VS SALE PRICE



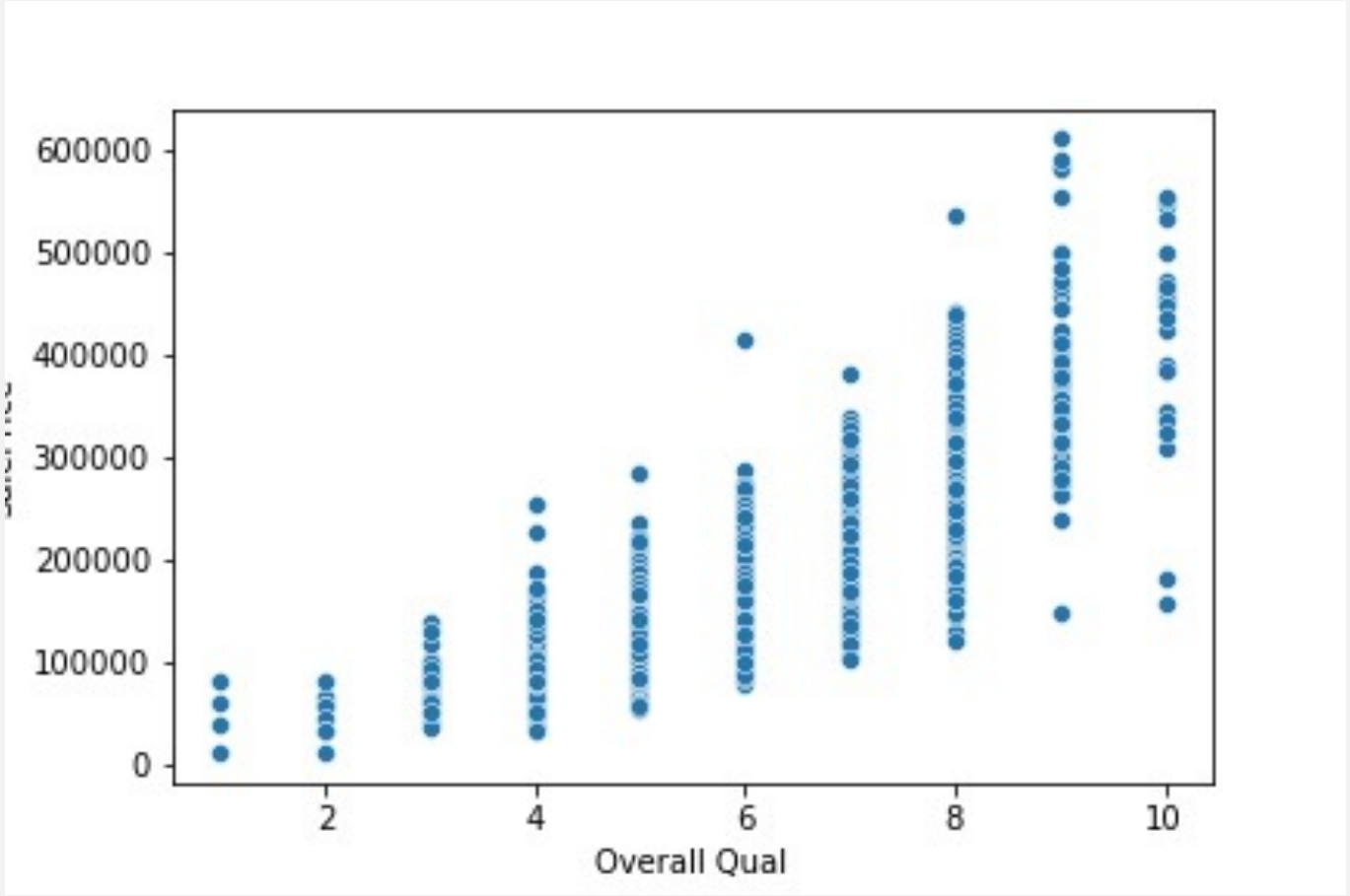
- Most expensive: Stone Brook, Northridge Heights, North Ridge, Veenker, Green Hills

- Least expensive: Briardale, Iowa DOT and Railroad, and Meadow Village

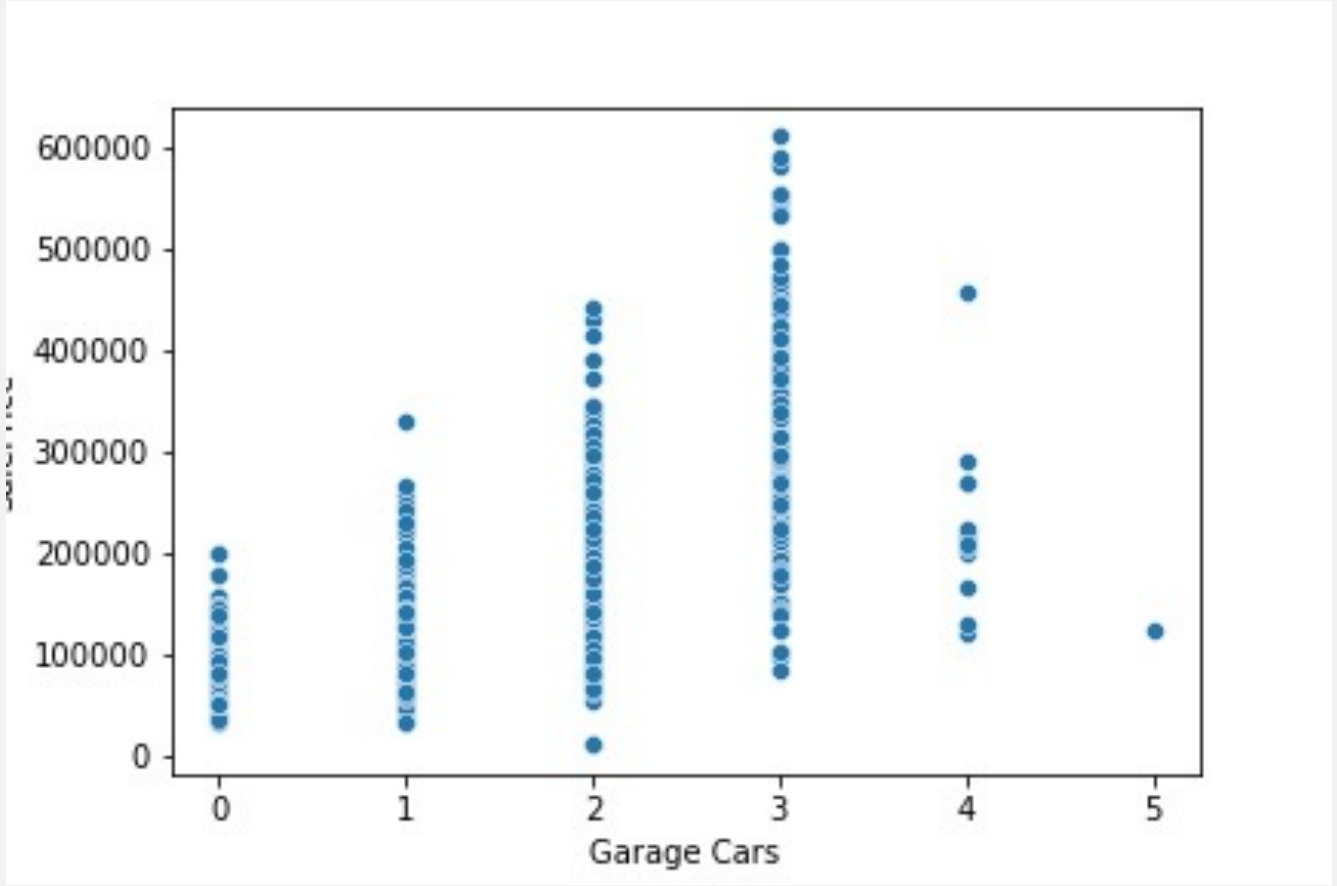
SALE PRICE DISTRIBUTION



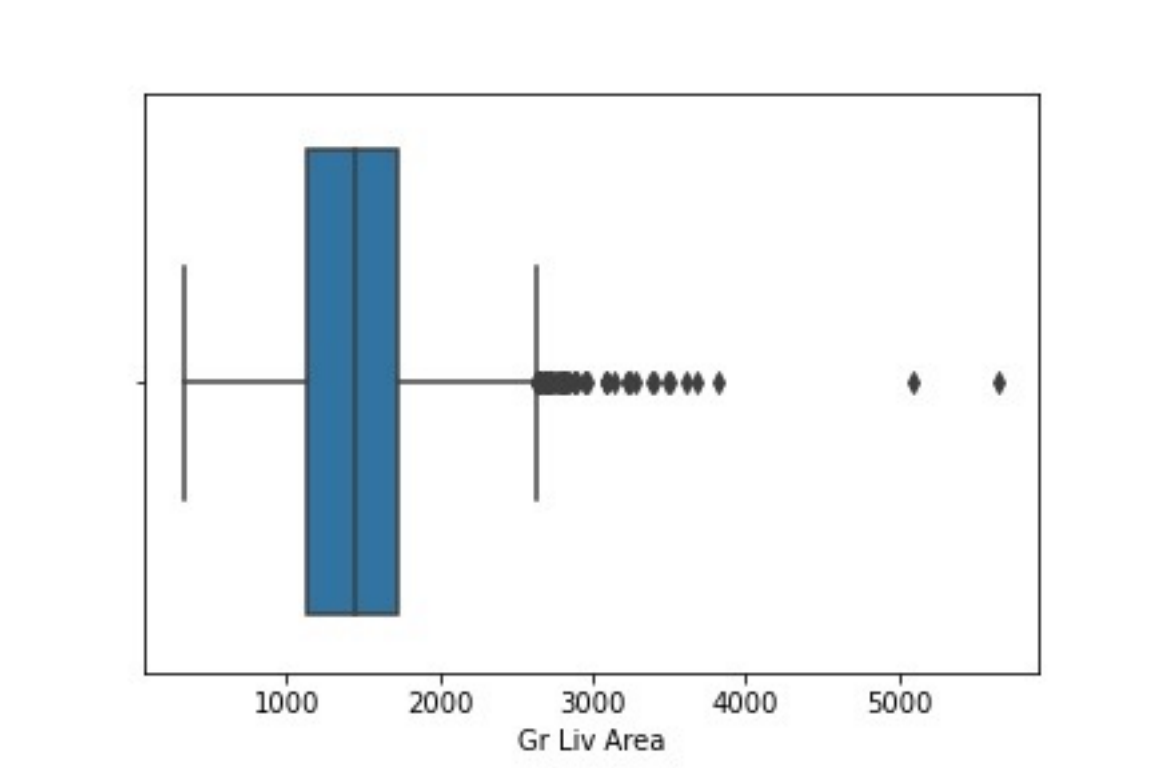
QUALITY VS SALE PRICE



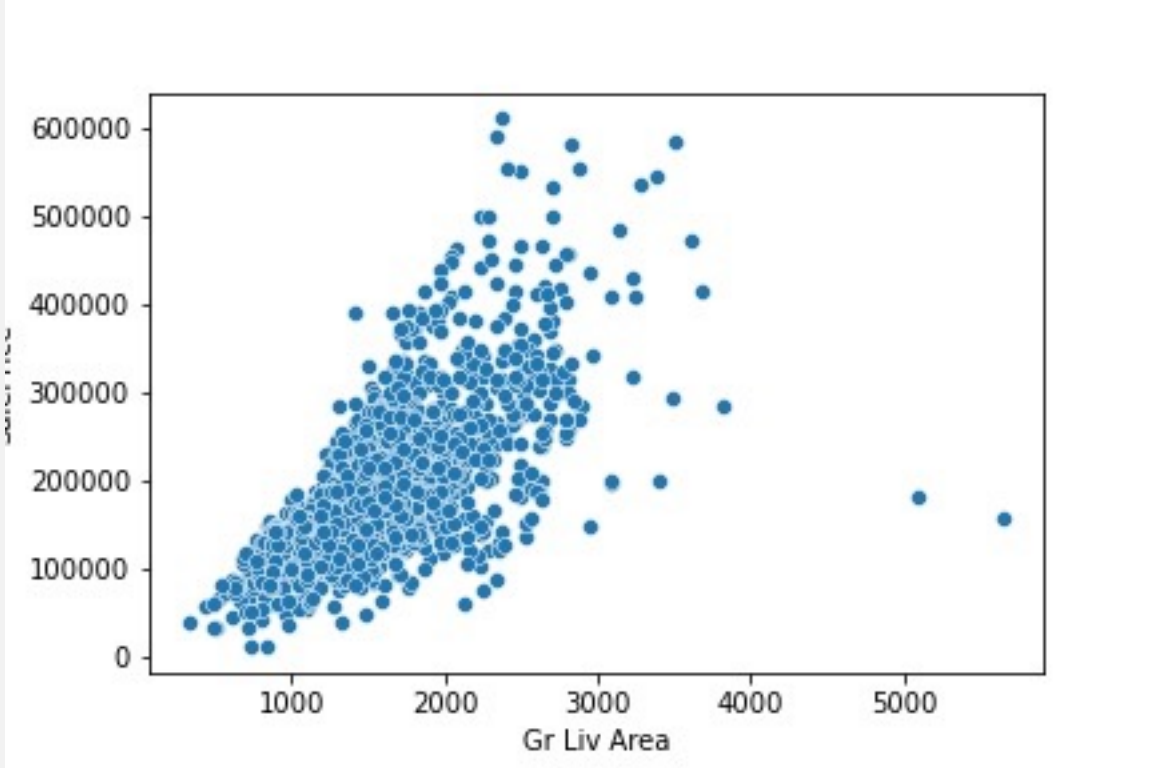
GARAGE SPACE (IN CARS) VS. SALE PRICE



ABOVE GROUND GRADE LIVING AREA

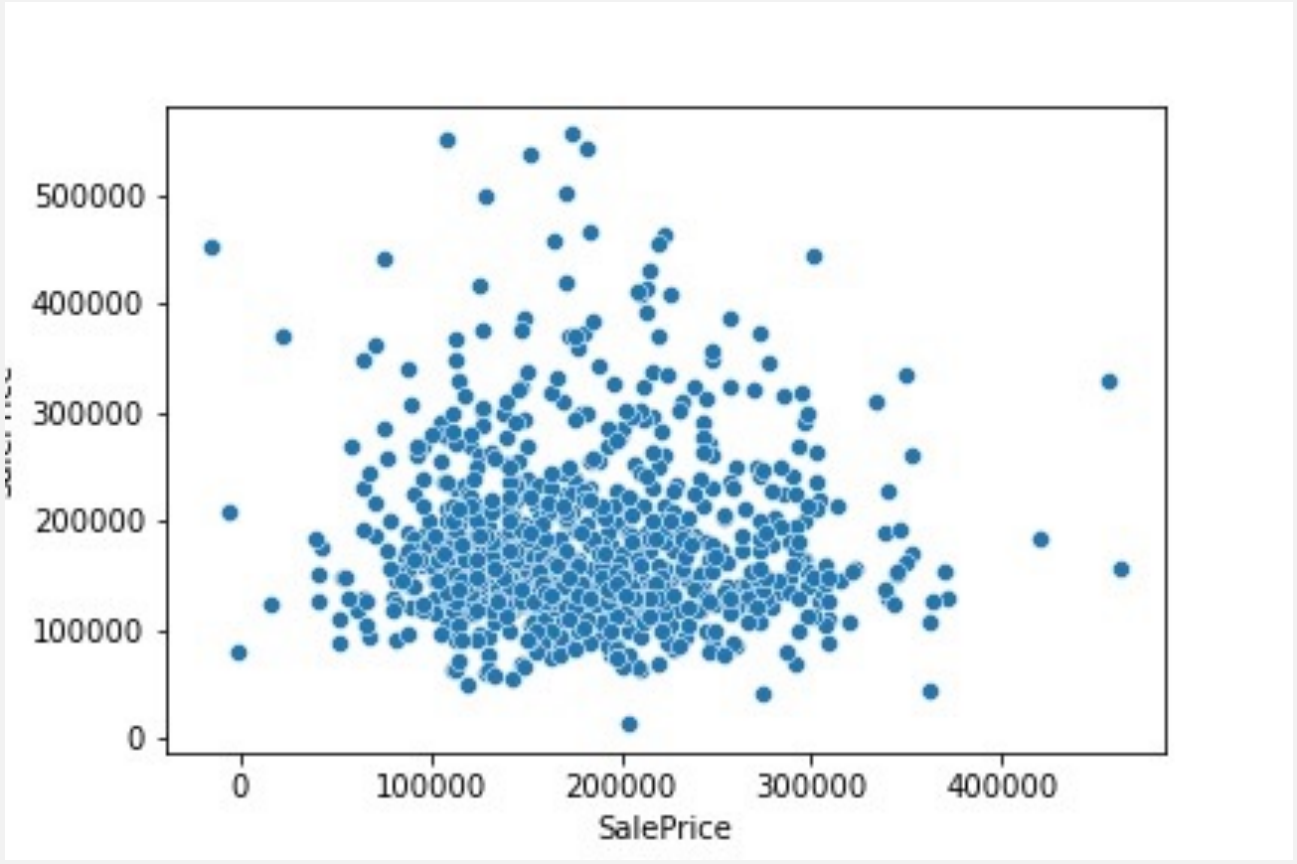


Boxplot



Living area vs. sale price

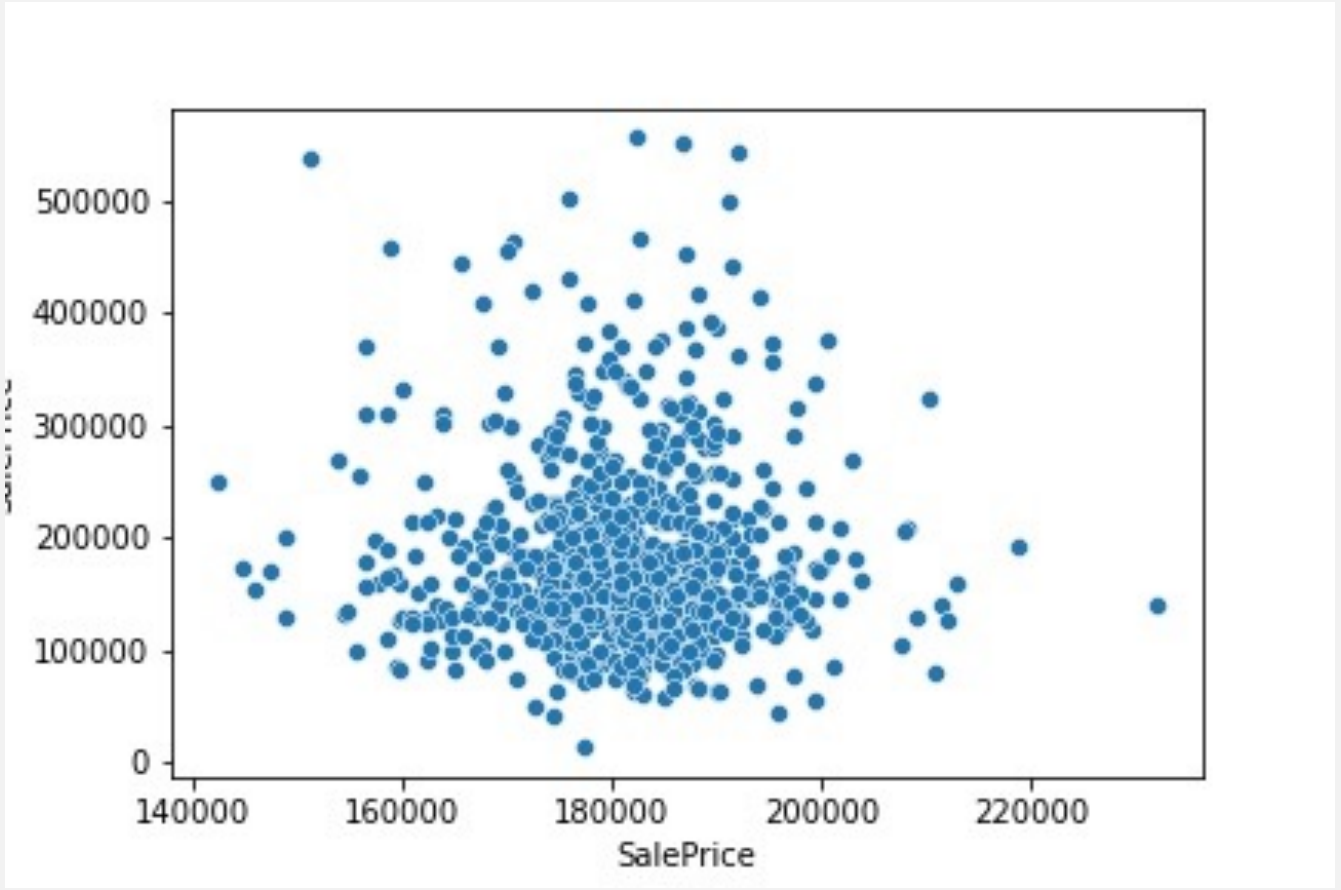
MLR RESULTS (ACTUAL VS PREDICTED)



MLR RESULTS

- A one unit increase in overall quality suggests an increase of 227875.35 in the sale price, while holding all else constant
- A one unit increase in square feet for the above grade ground living area suggests an increase of 48.88 in the sale price, while holding all else constant
- A one unit increase in square feet for the garage area suggests an increase of 76.78 in the sale price, while holding all else constant

LASSO REGRESSION WITH POLYNOMIAL FEATURES



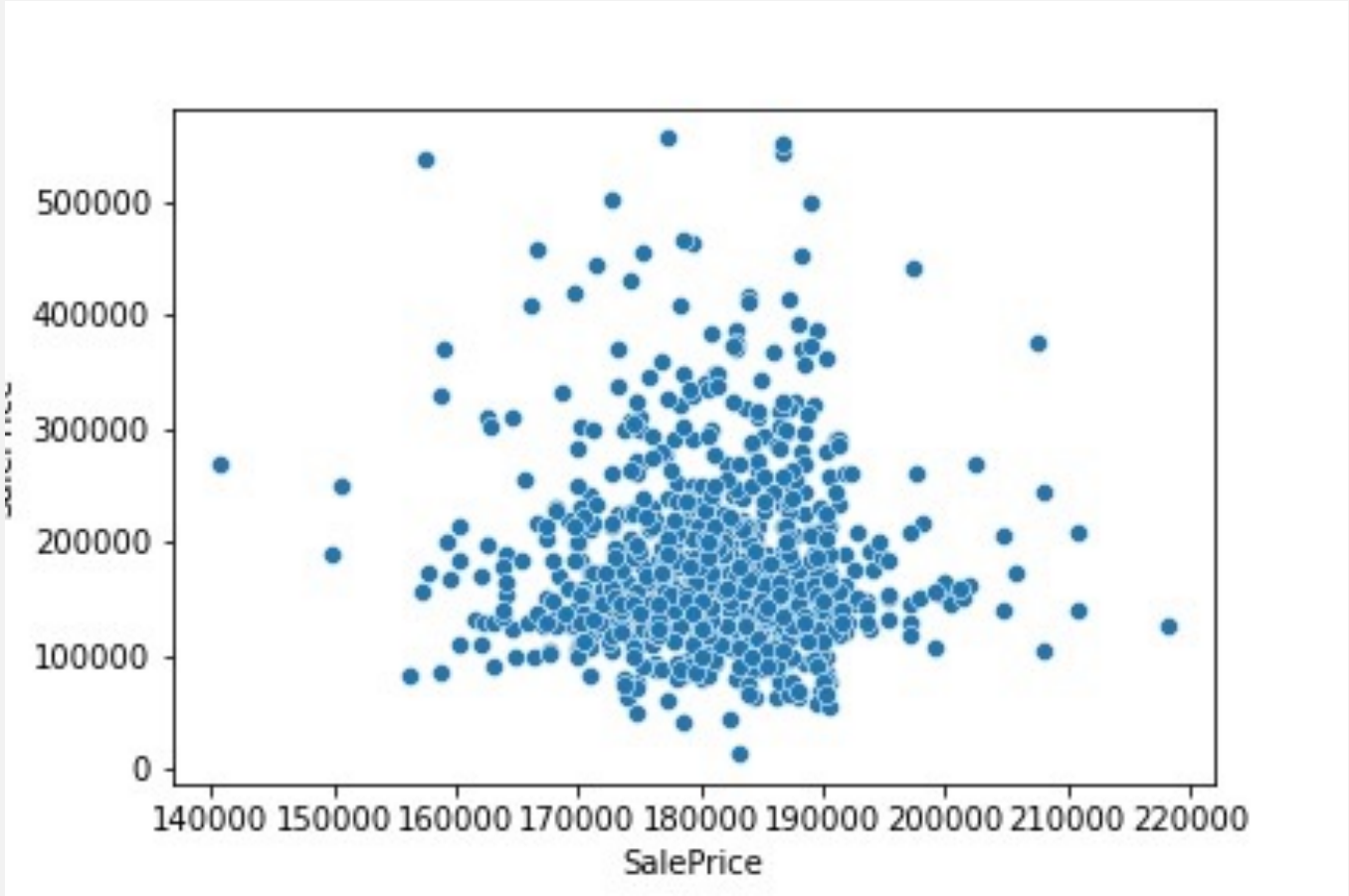
LASSO RESULTS

- Overall quality, above grade ground living area, and the garage area were all found to not have an affect on sale price
- A one unit increase in the year built suggests an increase of 5807.68 in the sale price, while holding all else constant
- A one unit increase in the total rooms above grade suggests a decrease in the sale price by 5753.64, while holding all else constant

LASSO POLYNOMIAL FEATURES

- A one unit increase in the interaction between overall quality and square feet of the 1st floor suggests an increase 5308.18 in the sale price, while holding all else constant.
- A one unit increase in the interaction between the garage area and the number of full bathrooms in the home suggests a decrease of 5478.84 in the sale price, while holding all else constant.
- A one unit increase in the interaction between the garage area measured in cars and the square feet of the basement suggests a decrease of 1766.02 in the sale price, while holding all else constant.
- A one unit increase in the interaction between the surface area of the 1st floor and the masonry veneer area suggests an increase of 267.25 in the sale price, while holding all else constant.

LASSO REGRESSION RESULTS



LASSO RESULTS

- A one unit increase in overall quality suggests an increase of 1874.65 in the sale price, while holding all else constant
- A one unit increase in square feet for the above grade ground living area suggests a decrease of 6097.32 in the sale price, while holding all else constant
- A one unit increase in square feet for the garage area suggests a decrease of 2750.73 in the sale price, while holding all else constant
- A one unit increase in the year built suggests an increase of 5680.67 in the sale price, while holding all else constant
- A one unit increase in the square footage of the 1st floor suggests an increase in the sale price by 9561.98, while holding all else constant
- A one unit increase in the total square footage of the basement suggests a decrease in the sale price by 6690.27, while holding all else constant

CONCLUSIONS

- Definitely would not use these models to make predictions
- Look to restructure my models and run further analysis
- Issues in splitting or maybe the refitting to the test data
- Low bias and high variance
- Did not perform well on the test data
  - Models focused too much on the training data and did not generalize well