Predicting House Sale Price in Ames, Iowa

Erik Lindberg

General Assembly Project 2
8/14/2020

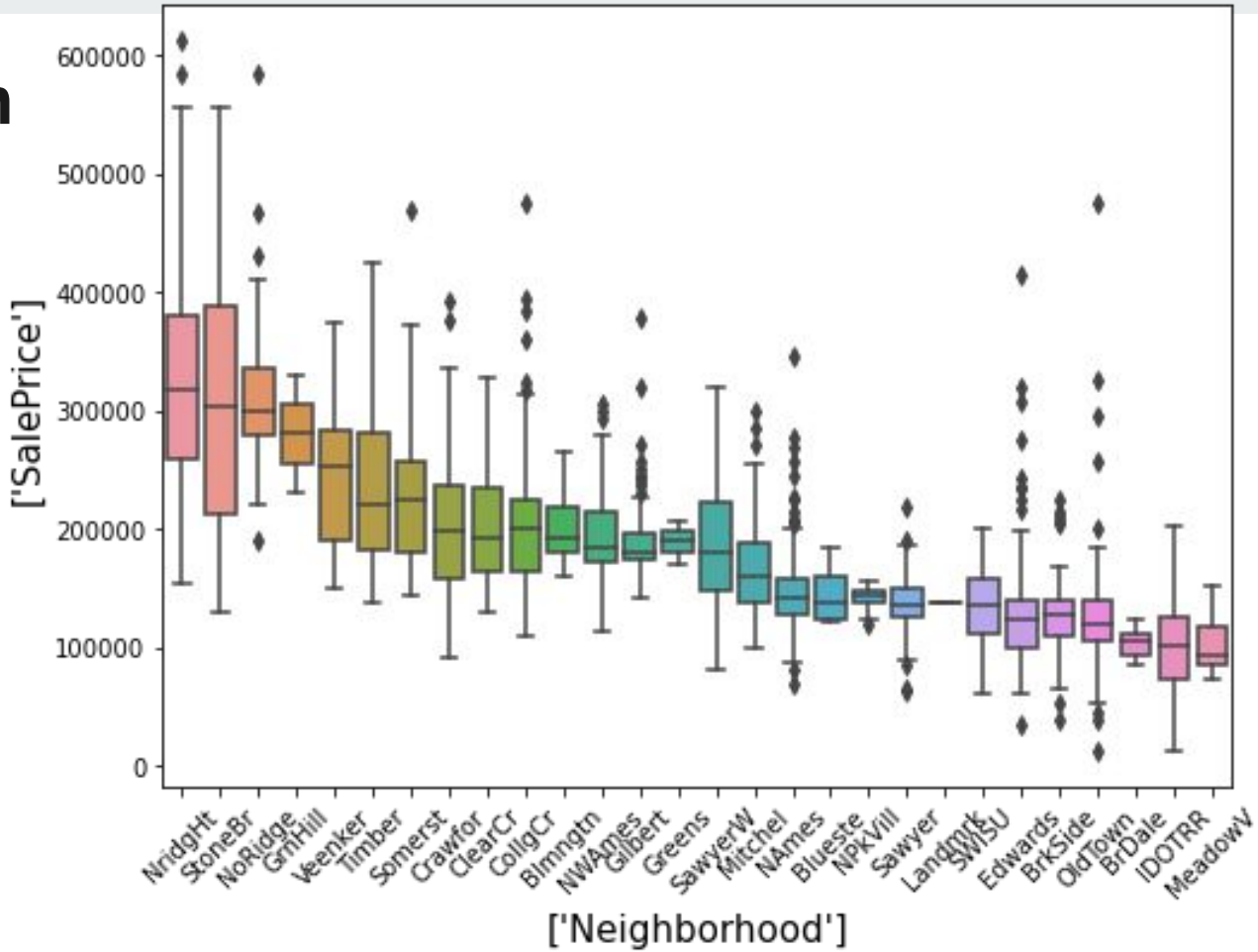
Problem Statement

The Gardner Real Estate Company is looking into increase sales in the Ames, Iowa region. To increase market share the company has decided that an application where customers can input specific house details and receive a sale price would set them apart from the competition.

Data Visualization

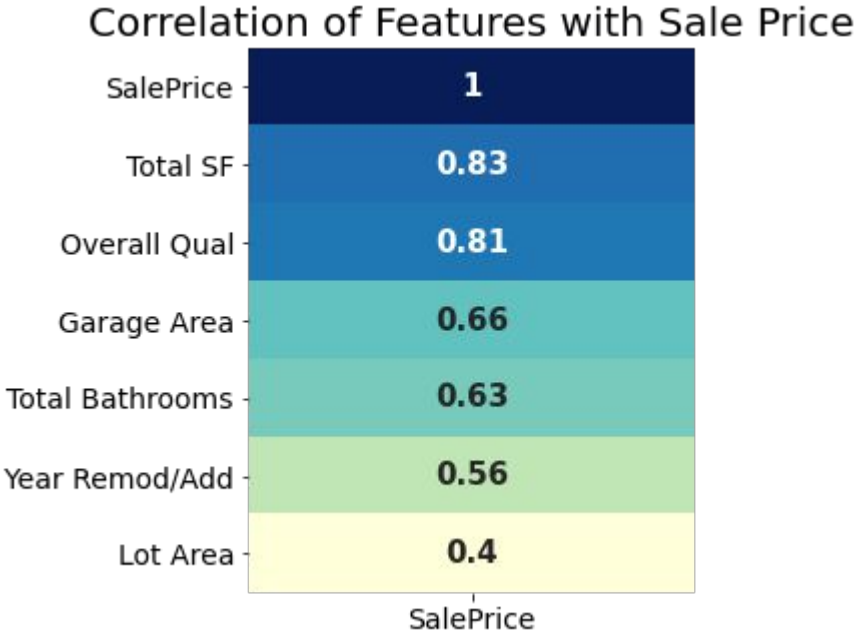
The neighborhood where the house is being sold has a significant impact on the sale price.

Highest priced on avg -
Northridge Heights
Stone Brook
Northridge
Green Hills



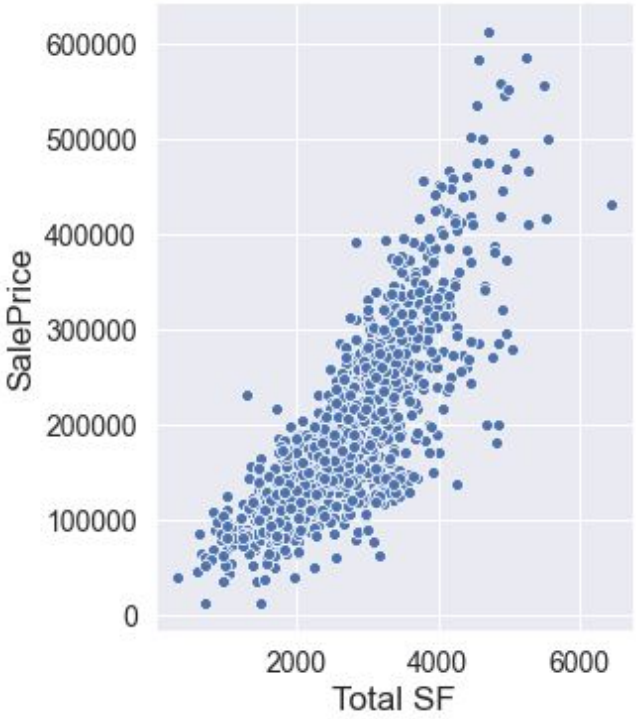
Data Visualization

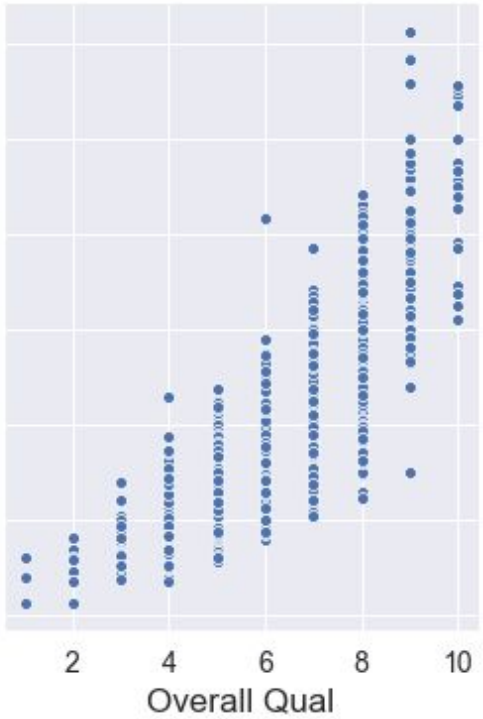
For numeric data, a correlation matrix was created. The features with the highest correlation to sale price were chosen for the model.

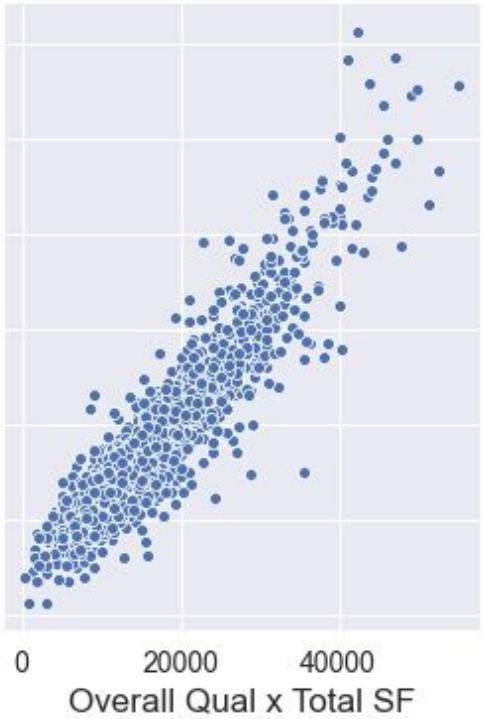


Data Visualization

Interaction terms were explored however the reduced interpretability was the reason they were eventually removed



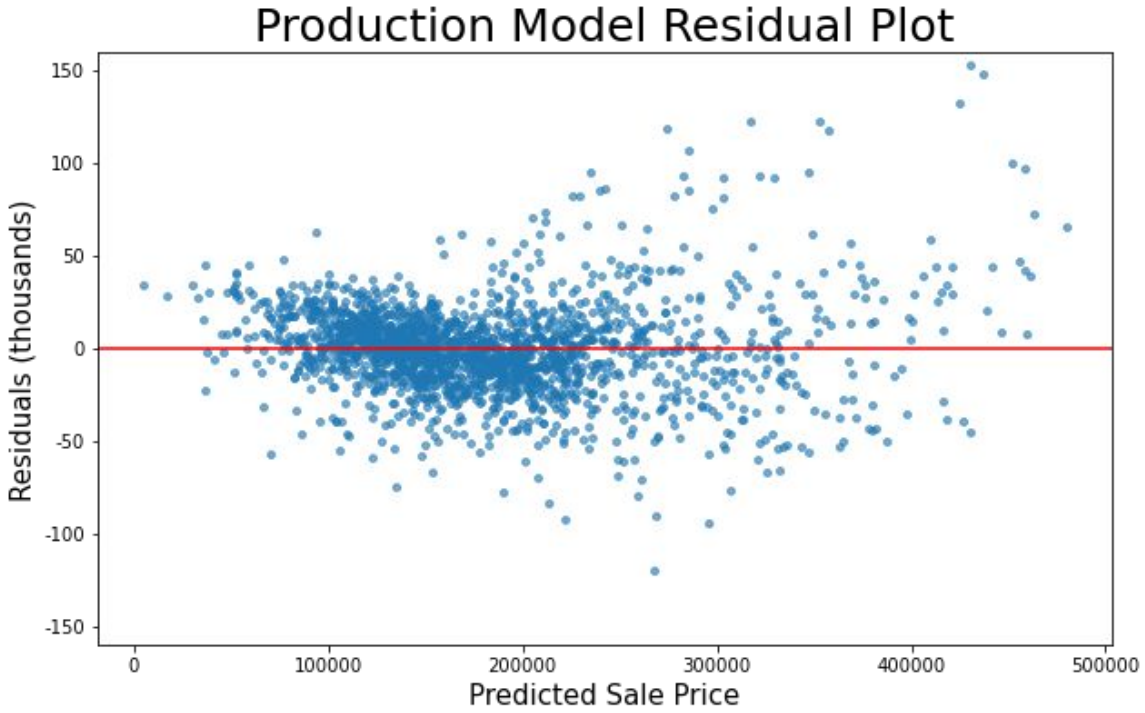




Production Model

Linear Regression

Model Features	---	---
Numeric	---	---
Lot Area	Overall Qual	Garage Area
Total Bathrooms	Total SF	Year Remod/Add
---	---	---
Catagorical	---	---
MS SubClass	Neighborhood	Condition 1
Exter Qual	Kitchen Qual	---



	Lot Area	Overall Qual	Total SF	Garage Area	Year Remod/Add	Total Bathrooms
Dollars increase per	2.0051	10660.0	32.3825	23.8789	248.5279	10930.0

Training R^2 = 0.904

Testing R^2 = 0.873

Conclusions

- Created a linear regression to predict house sale prices in Ames, IA
- While a better predicting model could be found, the interpretability of the above model allows anyone to interpret the results
- Total SF, Overall quality, and Garage area are the 3 features which should be maximized when trying to selling a house in Ames

Future Plans

- Create a subsequent model with more predictive power
- Could be used in conjunction with the basic linear regression to provide more accurate house sale prices for customers