Ames Housing Project

James sobrino

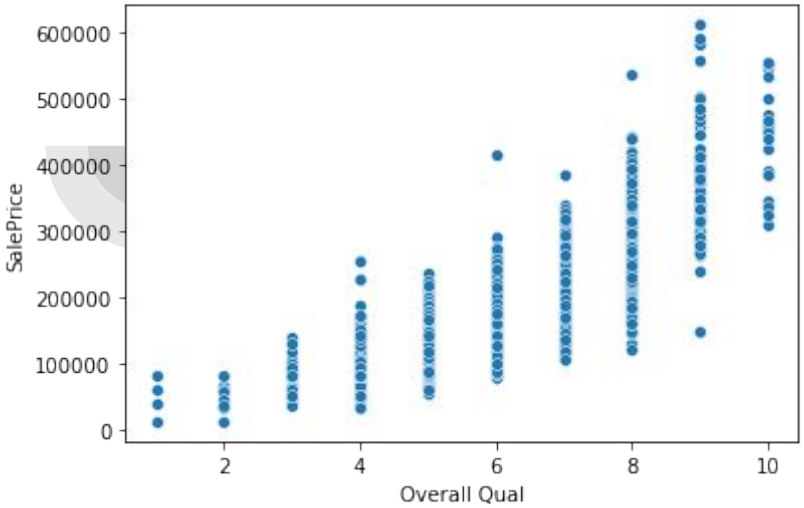
Problem Statement

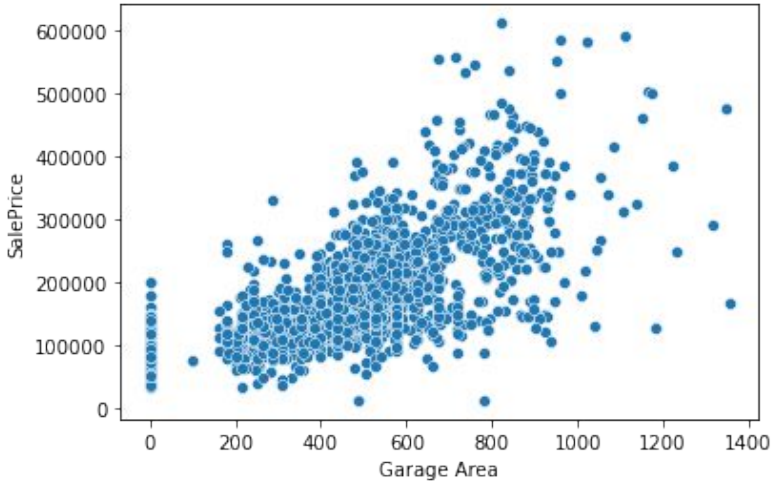
Your company has been hired by local real estate agents to predict housing prices in the city of Ames, Iowa. Using machine learning, create a model that can accurately predict housing prices based on features of the houses in question.

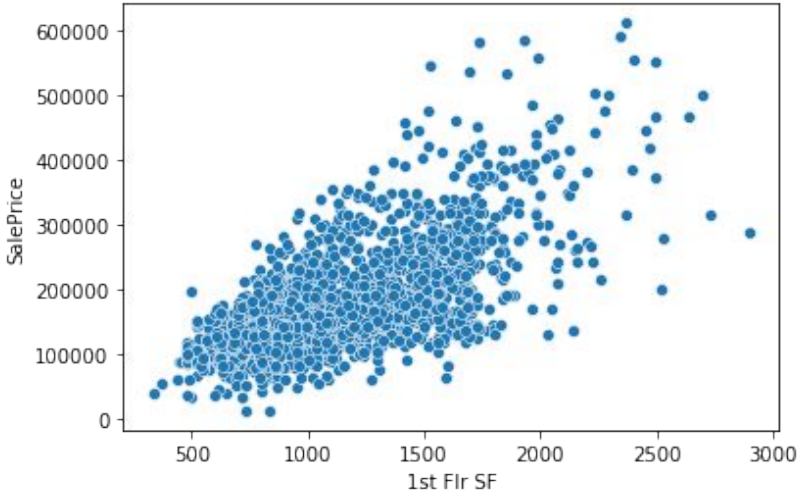
Background

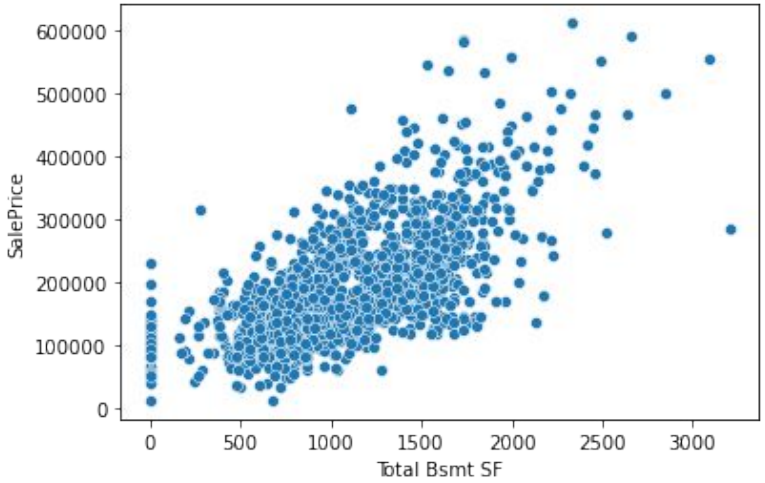
I researched what factors are the best predictors of housing prices. I found a case study of housing prices in Los Angeles which identified the best predictors of housing prices for the area.

Noted Features: Sqft, year built, Land-sqft, pool, access to parking, stories, beds, baths, total rooms, school proximity, park proximity, median household income.


SalePrice
Overall Qual
0
100000
200000
300000
400000
500000
600000
2
4
6
8
10


SalePrice
Garage Area
0
100000
200000
300000
400000
500000
600000
0
200
400
600
800
1000
1200
1400


SalePrice
1st Flr SF
0
100000
200000
300000
400000
500000
600000
500
1000
1500
2000
2500
3000


SalePrice
Total Bsmt SF
0
100000
200000
300000
400000
500000
600000
0
500
1000
1500
2000
2500
3000

Creating the Model

- Cleaned data of null values and some outliers
- Created a heatmap to identify strong correlations for numeric features
- Weeded out weakly correlated features
- Created dummy columns for categorical data identified to be strong factors for housing price from research.

Instantiating a nonoptimal model

- Baseline Model RMSE: about 80000
- Linear Regression Model RMSE: 46000
- Ineffective Model, many of the features were not independent of each other.

Choosing the correct model

- Baseline Model RMSE: about 80000
- Ridge Regression Model RMSE: 25000

Recommendations

- Best neighborhood: Northridge Heights
- Best features: Sq ft for any feature, Garage, total rooms and bath, overall quality, year built/remodeled

Closing Thoughts

There is more work to do with this model, more stringent cleaning needs to be done. There are more features that can be added to the model.

Sadly, there are no more billable hours for this project for the package purchased. Additional purchases will be required to increase the robustness of the model.