Predicting Housing Prices

Agenda

1) Data Overview & Methodology
2) Model 1: Basic Linear Regression Model
3) Model 2: +Feature Engineering
4) Model 3: Lasso CV
5) Conclusions & Recommendations

Introduction

The goal of this project is to utilize the dataset provided to accurately predict the Sale Price of homes sold in Ames, Iowa.

Success will be evaluated using the RMSE which in this instance represents the average difference between a prediction and the actual sale price of a home.

Data Overview & Methodology

Data Summary

Features: 79

Train: 2051

Test: 878

Null Values:

9822 Train

4171 Test

Target: Sale Price

Mean: $181k

Min: $12.8k

Max: $612k

Time Frame: 2006 - 2010

Methodology

EDA & Cleaning

⬆ ⬇

Feature Engineering

⬆ ⬇

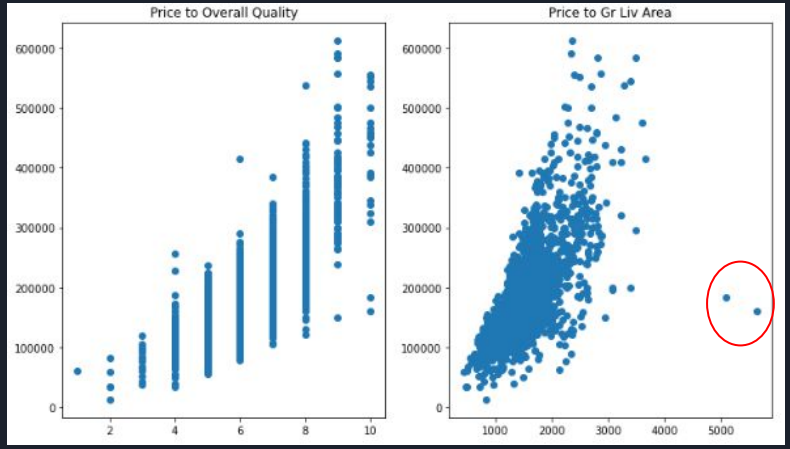
Build, Evaluate & Iterate Model

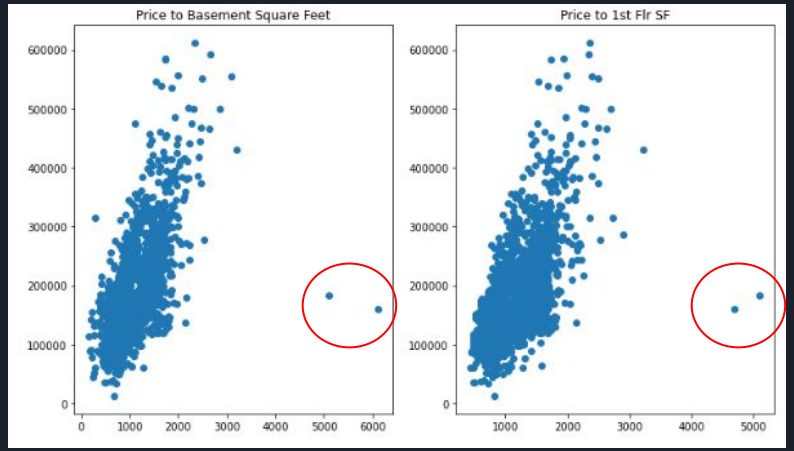
EDA & Cleaning

- Identify the Highest Correlated Features to Sale Price [Overall Quality, Liv Area, Sq Ft, Rooms]

- Two Outliers with low Sale Price relative to size

- Impute or Drop Nulls
  - If large percent of a feature is null then drop the entire column
  - Otherwise impute either the mean or 0





Model 1
Basic Linear Regression

Features: Six

Overall Qual, Gr Liv Area, Garage Area, Lot Area, TotRms AbvGrd, Bedroom AbvGr

Dataset	RMSE	R2
Train	35,243	0.8
Validation	33,741	0.81
Test	36,539	N/A

Model 2
+ Feature Engineering

Features: 204 Features

Engineered Features:

-Dummies of all Categorical Features

-Polynomials (Square Feet, Bedrooms, Bathrooms, Overall Quality)

-New Features (Total Square Feet & Total Bathrooms)

Dataset	RMSE	R2
Train	19,107	.94
Validation	22,715	.92
Test	24,589	N/A

Model 3
LassoCV

Features: 204

Scale Features

Lasso Cross-Validation Estimator

Dataset	RMSE	R2
Train	19,121	.94
Validation	22,577	.91
Test	24,464	N/A

Model Evaluation Summary

LassoCV Wins!

Model	RMSE
Baseline	79,012
Basic Linear Regression	36,539
Linear Regression + FE	24,589
LassoCV	24,464

Conclusion & Recommendations

Consider a wood shingle roof as we saw a $~45k increase in Sale Price for wood shingle roofed homes versus standard composite roofs.

Improve Overall Quality Score- Identify any outdated/unfinished sections of your home and update/remodel to improve your overall quality. Every one unit increase in quality score resulted in $7k increase in home value.

Increase Greater Living Area Square Feet- Adding another story, expanding walls, finishing unfinished or unlivable space is one of the most efficient ways of increasing Sale Price.