

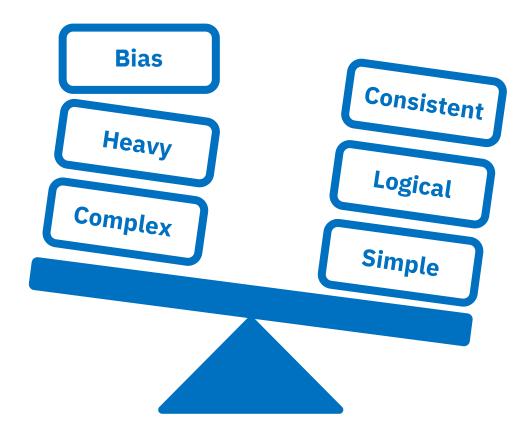
Pricing Model for Ames Iowa Real Estate

October 2019

Modeling is a simple representation of reality.

Project Goal & Design Principle

The goal of a good pricing model for real estate shouldn't be to the get the best statistics metrices, but to build a model simple enough to estimate a fair housing price at a point in history.



One-Pager Overview

Total Score: 23,076

13

Numeric Variables 10

Categorical Variables
Converted to Dummies

Under 100 Total Variables

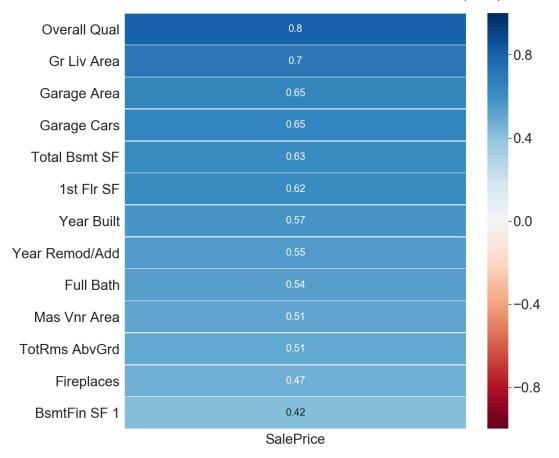
Lasso Regression Power Trans. on Target and Key Variables

Start With Numeric Variables

Started only looking at the numeric variables.

At first go, we chose the 13 numeric variables most correlated to the target.

Correlation Between SalePrice and Numeric Variables (>0.4)



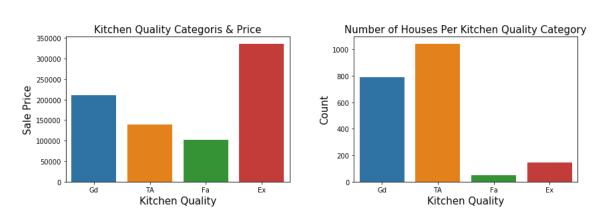
Selected Relevant Dummy Variables

We want to get a very selective group of categorical variables to build a relatively accurate model.

One very important assumption for linear regression models is used here:
Independence among Xi.





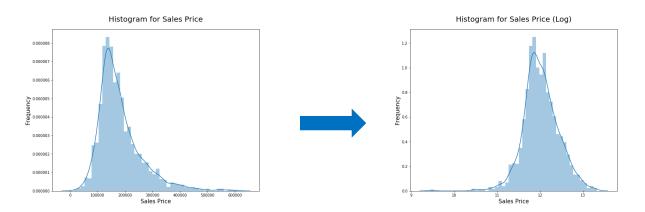


Power Transformation

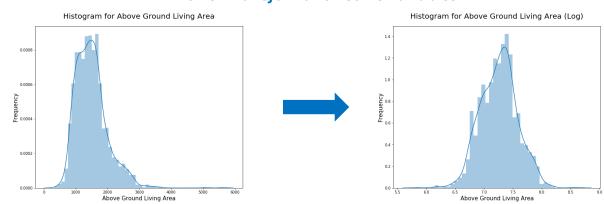
Power transformation on our target gave us a much better distribution.

Following the same logic, taking the log on some of the variables also contributed to a higher score.

Power Transformation on the Target



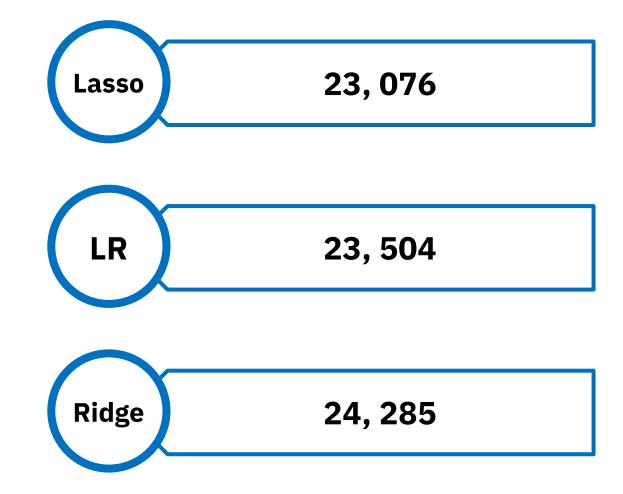
Power Transformation some Variables



Model Selection

After scaling the our features, we compared the results from our 3 models:

- Linear Regression
- Lasso Regression
- Ridge Regression



Further Considerations

The data is subject to historical bias, we need to constantly find more updated housing dataset in a changing housing environment.

To evaluate intrinsic values instead of pricing, it would be useful to collect data such as rental and interest rate.

It would also be interesting to compare real estate pricing in Ames Iowa with the US real estate prices to determine if investing in Ames Iowa is a good idea.

Thank You