Ames Housing Pricing Prediction Model

Predicting Housing Prices for a Real Estate Rental Agency Expanding in Ames, Iowa

Our Goal

Problem Statement

Our company, ACME Housing Inc, is a small housing investment company that earns the majority of its revenue by purchasing and renting units, and they are looking to expand their market share in Iowa.

Ideally, this model would allow us to:

1. Determine the underlying value of houses in the Ames, Iowa region
2. Identify which features contribute the most to housing prices, and are most desirable to renters
3. Determine if specific features are more popular in Ames, in the case that ACME decides to purchase and build a housing development

Accurate & Easy to Use

1. Accuracy - Model has to predict sale price well based on unknown data

2. Less Variables - Many variables are subjective and require domain knowledge and a trained appraiser.

   Less variables = less time spent and less error to subjectivity

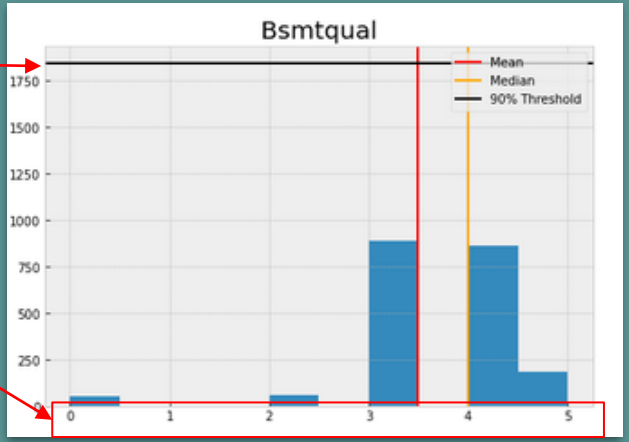
Steps Taken to Analyse the Data

1. If more than 90% of the sample was in a specific category in of the data

2. Ordinal Data was changed numericised, for example:

Na - 0
Po - 1
Fa - 2
TA - 3
Gd - 4
Ex - 5

3. Specific sets of normalised data was collected into similar groups to ensure that there was enough data in larger nomalised groups



1.5.4.1 Neighborhood

Neighborhood will be grouped into smaller groups of similar medians in the following category:

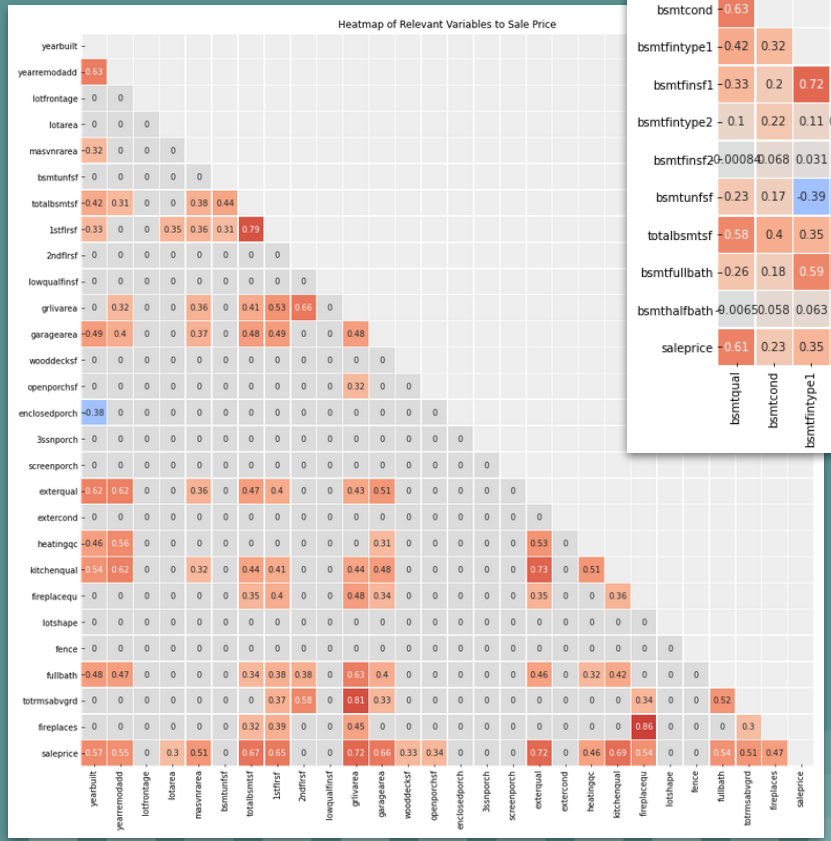
- upperclass - StoneBr, NridgHt, NoRidge, GrnHill
- uppermiddle - Veenker, ClearCr, Somerst, Timber, CollgCr, Crawfor, Blmngtn, Greens, NWAmes, SawyerW, Gilbert, Mitchel
- middleclass - NPkVill, NAmes, Blueste, Landmrk, SWISU, Sawyer
- lowermiddle - BrkSide, Edwards, OldTown
- rural - BrDale, IDOTRR, MeadowV

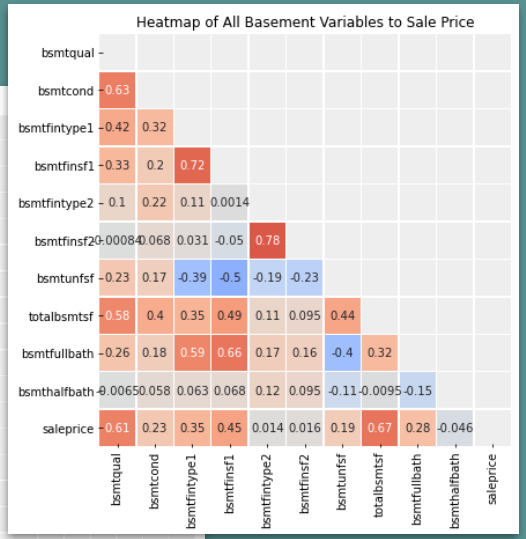
Dropping Variables based on Correlations

Correlations were considered between similar groups to pare down the number of collinear, namely:

- Basement variables
- Garage variables

Correlations of variables with sale price of above 0.3 were also considered to carry forward into more analysis.





Age, Fireplace quality

Engineered Feature	Method
Age	Younger of years from construction or remodel to sale
Fireplacesxqu	Multiplication of Fireplaces & Fireplace Quality

Models Tested

Model	Model Proposal	No. of Variables	RSME
Null Model	Use the Mean	0	79276
Simple Model	Vs. Lot Area	1	54664
Model 1	Base Model	47	25381
Model 2	Drop Variables Based on Lasso	46	25381
Model 3	Most Impt 20 Variables	20	28932
Model 4	Most Impt 40 Variables	40	28975
Model 5	Dropping Variables to 1 per group	37	25682

Final Model

Ridge Regression with Alpha of 1.0

Accuracy

Root Mean Squared Error

- Train Set = 26126
- Test Set (Kaggle) = 29692

Easy to Use

- 37 variables, with as few overlapping variables as possible
- Outside of neighborhood, the model does not require any variables that are directly unrelated to the house, which allows for a faster assessment

Best Discounts

Living Space

- Above ground living space and basement size were the largest contributors to sale price
- The biggest detractor was unfinished basement space
- Buying houses with unfinished spaces and finishing them might be a good way to get a decent price if the cost of finishing it is less than the discount

Neighborhood

- Tier 1 Neighborhoods saw a big rise in price, the difference between the others were much more marginal

Garages

- The larger the garage, the pricier the house
- Students may not need or have as many cars, this may be an easy way to purchase houses at a discount

Data, Data, Data

1. Longer Time Frame - More Data

2. Rental Prices of the area, will give us a good indication of what the best houses to target are

3. Distance from Iowa State University - With a large presence, student housing may be one of the best sources of renters