Ames Housing Data & Predictions

Accurately predicting future home sale prices!

Problem Statement/Objective

I am a consultant that just recently gained a client in Ames, Iowa. They have heard of my work and have asked me to help them predict housing prices in the area so that they can have a leg up on the market!
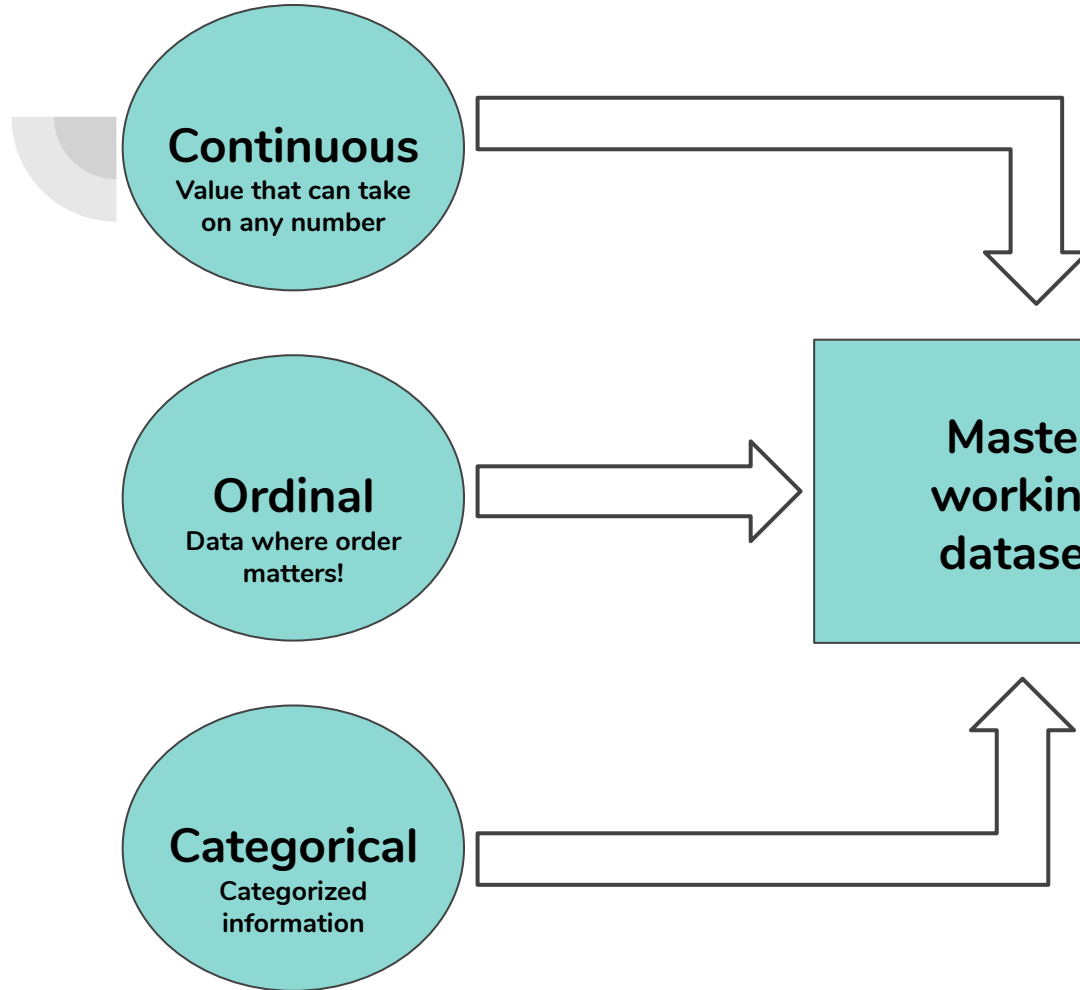
Information about Data Explored

Where: Information collected from the Ames Assessor's Office about residential units sold from 2006 - 2010.

What: I was given a training data set by my client that I can use to create my model for predictions. This data set includes 81 characteristics & 2051 observations (or rows of home sales). Within these data, the observations can be categorized as either: Continuous, Ordinal, or Categorical.

Continuous
Value that can take on any number

Ordinal
Data where order matters!

Categorical
Categorized information
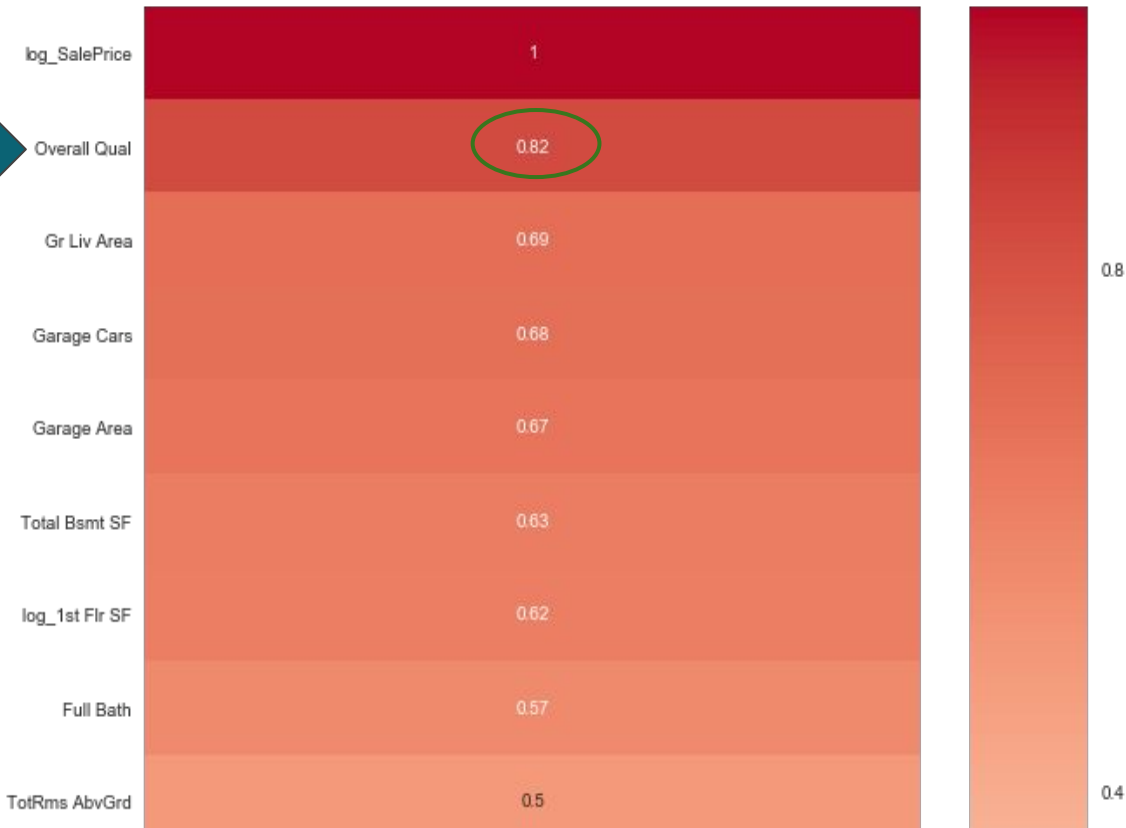
Master working dataset

- Particular columns of interest were all selected and transformed/manipulated to better be useful to create a productive model.
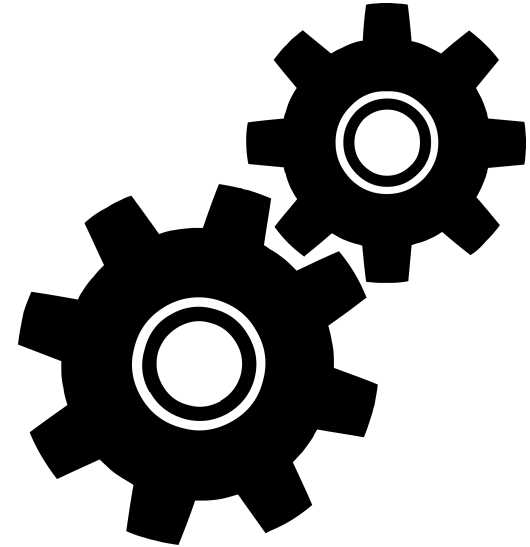
Variable Selection

Heat Correlation Map

- **The heat map was used to help understand which variables have the biggest impact to Sales Prices.**

- **This ultimately help sort out what characteristics can impactfully predict home prices in Ames, IA.**

- **As seen by this heat map, the Overall Quality of the house has a HIGH impact on the price of the house, with a score of 0.82 (out of 1).**
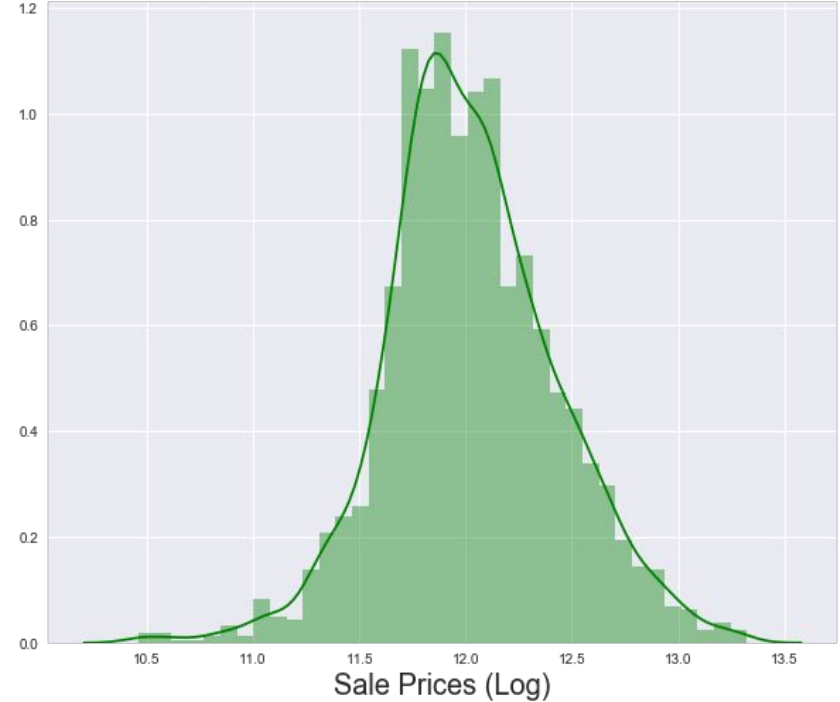

log_SalePrice	1
Overall Qual	0.82
Gr Liv Area	0.69
Garage Cars	0.68
Garage Area	0.67
Total Bsmt SF	0.63
log_1st Flr SF	0.62
Full Bath	0.57
TotRms AbvGrd	0.5

0.8

0.4

Important Variables:Our Target - SALE PRICE!

Histogram Distribution of Sale Prices

Distribution of Log Transformed Sale Prices (Without Sale Outliers)

- Skewed to the right: Expensive home prices.

- More normal distribution. A better target!

Important Variables: Quality of Home

Relationship between Quality of House & Log of Sale Price

- There is a distinct positive pattern between the quality of the house and the price.

- As the Quality increases the Sale Price increases.

- This translates to this characteristic being a strong predictor for our model.

Model Selection

- The model type used is a Linear Regression Model.

- The goal is for our model to predict the price of a home with specific characteristics with little error.

Evaluation Metrics

- In order to know the performance of our model, it was evaluated against 2 different metrics. R-squared & RMSE (Root Mean Squared Error)

- R-squared is on a scale of 0 to 1, with 1 being the best score.

- RMSE measures how for our predictions are from the correct value, on average. The closer we get to 0 (No difference in value) the better.

Model 1: R-squared - 0.81; RMSE - $34,072

- Does not generalize well to unseen data.

Model 2: R-squared - 0.91; RMSE - $21,166

- Generalizes well to unseen data.

Conclusion:

- I have a model that can generate excellent performance for your endeavors.

- More Data! The data set I have based this model off of, is slightly outdated. Obtaining up to date data can potentially improve the model greatly.