
Ames House Price Prediction



Group 1

Natasha, Joel Quek, Stephen Zhang

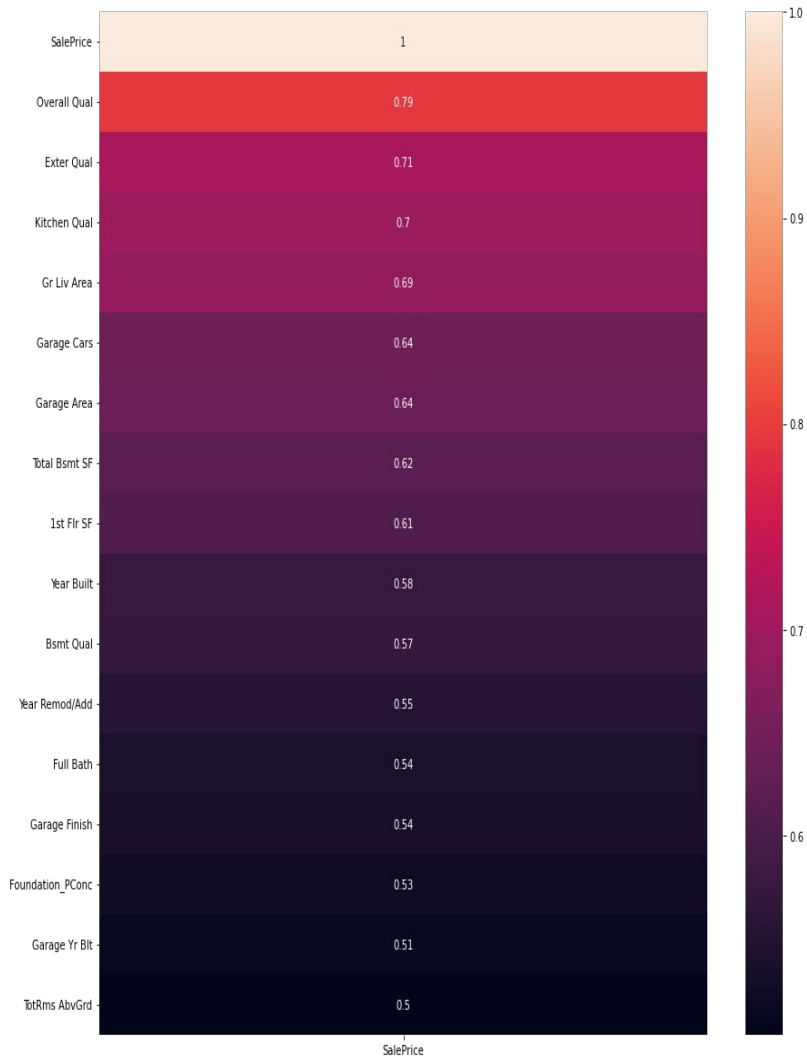
Background and Problem Statement

Objective: Purchasing a house can be a huge commitment . As a real estate agent, we would like our clients

1. to find out which features have greater impact to the house value
2. to find out how much they probably need to spend on purchasing a house with features that they want

Target Audience: Home buyers in Ames

Disclaimer: Data was taken from houses sold in 2006-2010. Probably we will need to retrain our model with a more updated data to ensure better predictions



Data Overview

2051 samples of residential properties sold between 2006 to 2010 in Ames, Iowa.

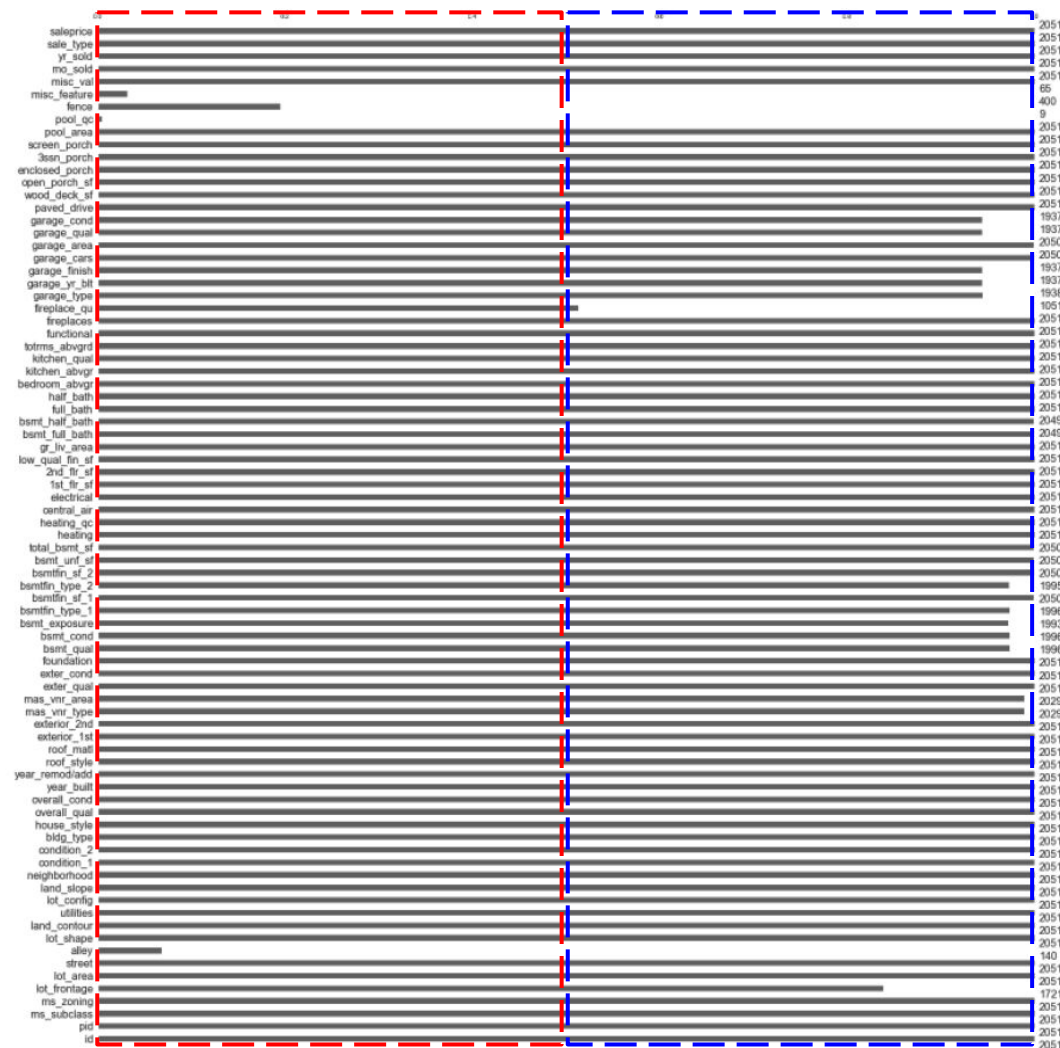
7 Categories of Data:

1. Location and Proximity
2. Square Footage
3. Quality and Condition
4. Parts of the Home
5. Utilities and Intangible Factors
6. Time Factors
7. Home Price [Target]

Pre Processing

1. Our model will fail if we have too many missing values
2. Columns with >50% null values are removed
3. Filling up Missing Values
 - Imputing Median value for numerical features
 - Imputing 'NA' for categorical features

Effect: those imputed values might skew the data as they are treated as actual observations

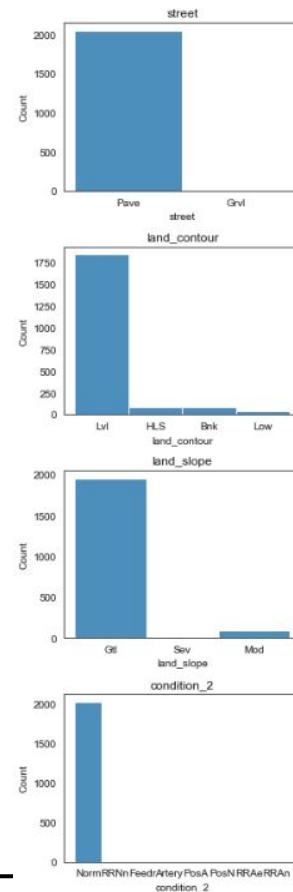
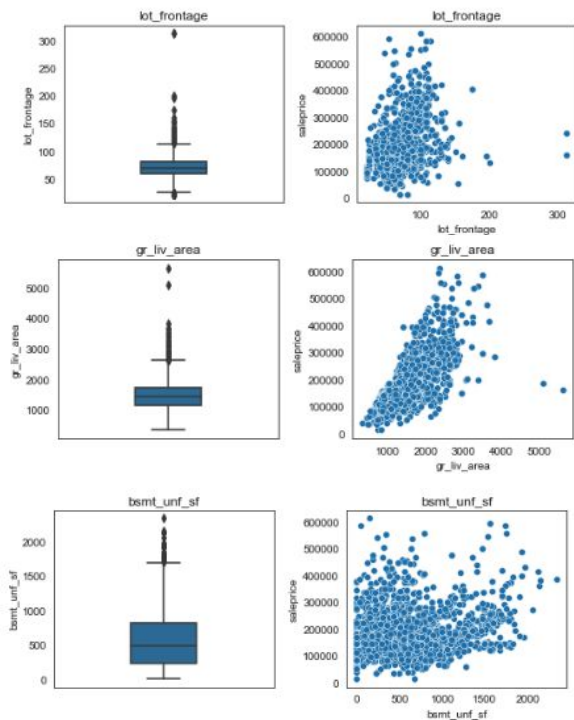


Feature Engineering

Feature Engineering:

1. Some of the data have outliers. Therefore, we are going to scale using the Robust Scaler for numerical values.
2. One Hot Encoding: preprocessing step to convert categorical features to 0 and 1 that can be processed by machine.
3. House age: year sold- year built

Limitation: When most of our data resides on one of the categorical feature, this feature can't be used as a good predictor because there is not much differences for our machine to learn



Best Model: Linear Regression

Reason

We want to predict the value of a variable based on the values of other variables, so we decided to use linear regression.

The Score

R^2 : R-squared is the percentage of the dependent variable variation that a linear model explains.

The Root Mean Square Error

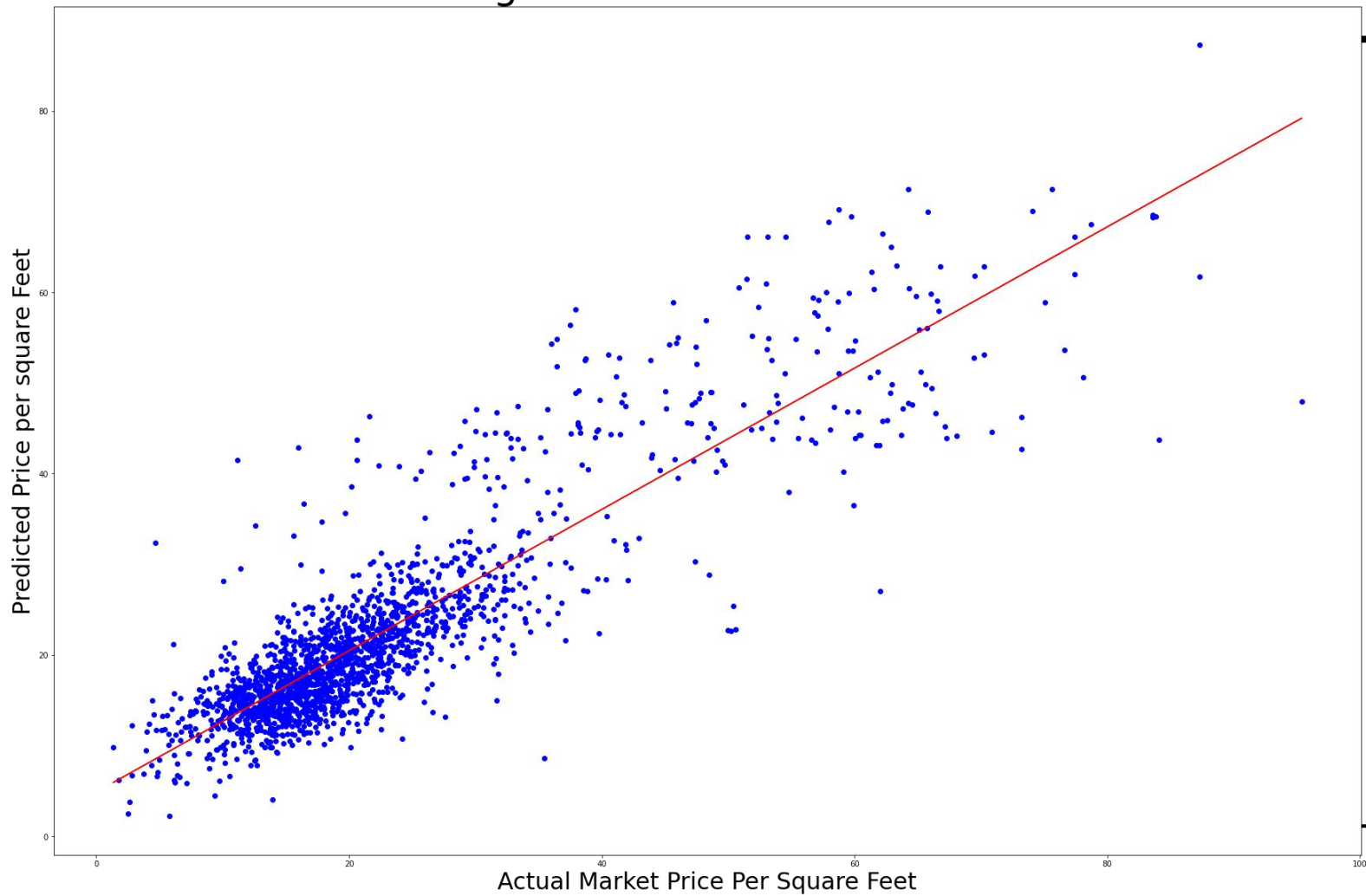
Measure of the differences between values (sample or population values) predicted by a model and the values observed

Criteria for Good Model

How near our predicted value to true value in our Training Data

Takeaway Finding Undervalued Homes based on the Model

Housing Price - Predicted Vs. Actual



Takeaway Factors Affecting Price

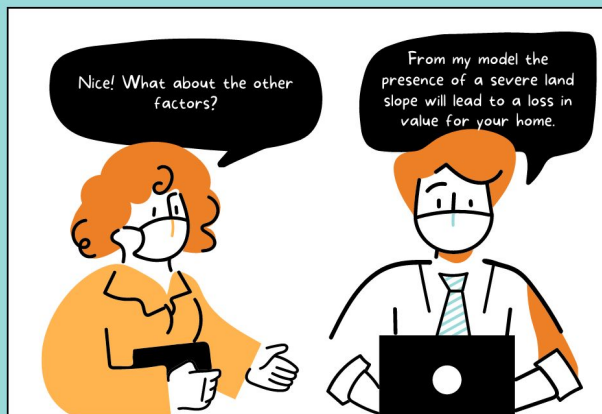
Factors with Positive Effect on Price Per Square Feet

Factor	Effect on Price per Square Feet
Wood Shakes Roof Material	22.19
2-Storey PUD 1946 and newer	21.63
Floating Village Residential	6.129
Near Positive Off-Site Feature - parks, greenbelt etc	4.039
Availability of all Utilities	3.14
Paved Street	2.13

Factors with Negative Effect on Price Per Square Feet

Factor	Effect on Price per Square Feet
Irregular lot shape	-6.89
Frontage on 3 sides of property	-6.02
Gravity Furnace	-4.94
Adjacent to East West Railroad	-4.64
Land contour low depression	-4.31
Severe Land Slope	-2.81

CHOICES



Thank You 😊
