**Background**

The Ames housing dataset (<http://jse.amstat.org/v19n3/decock/DataDocumentation.txt>) is a comprehensive list of properties that contains 2930 observations with 82 tabulated variables for homes sold in Ames, Iowa between 2006 and 2010.

**Problem Statement**

Home-buyers like flashy, sexy and noticeable. Home buyers want to live in good neighborhoods. Home buyers like big spaces where they can keep their possessions. Psychologically speaking, home buyers are human just like the rest of us and can be subject to a first impression. These factors were all taken into consideration when analyzing the Ames, IA housing dataset. The goal of this project was to find the factors that were the most influential when predicting the price of a home and making actionable recommendations to homeownder to boost home selling prices in the region surrounding Ames.

**1) EDA, Cleaning and Initial Visualizations**

This dataset contained several columns of mostly null data. The following columns contained over 50% null data and were dropped from the dataset:

1. Alley (Nominal): Type of alley access to property
2. Pool QC (Ordinal): Pool quality
3. Fence (Ordinal): Fence quality
4. Misc Feature (Nominal): Miscellaneous feature not covered in other categories

The strategy used for imputing null values in columns with less than 50% null values consisted of identifying the variable, consulting the data dictionary and replacing a null value with a value determined on a case-by-case basis with influential factors being the variable type (numerical or categorical) and categorical feature options. Features with imputed null values can be found in the following Jupyter Notebook:

01\_EDA\_Cleaning\_Initial\_Visualization.ipynb

Boxplots were created to visualize the interquartile range of relevant numerical data. Outliers were identified and observations can be found in the following Jupyter Notebook:

01\_EDA\_Cleaning\_Initial\_Visualization

A couple of interesting data points were identified within the training data. Two homes (Id 1499, 2181) with ground floor living area sizes ~5000 square feet sold for prices within 15% of the average predicted price in Ames. They both had overall quality ratings of 10/10 and overall condition ratings of 5/10. Without more information regarding the considerations taken into account for both of these ratings I could not come to a conclusion as to why these large spaces sold for the prices that they did.

Correlations were identified via heatmap between the sale price of a home and the other variables in the Ames Housing dataset.

**2) Initial Model Creation**

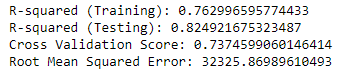
A features matrix (X) and target vector (y) were created based on the following six variables and were chosen as the basis of developing a linear regression model:

1. Overall Qual: Rates the overall material and finish of the house
2. Gr Liv Area: Above grade (ground) living area square feet
3. Garage Area: Size of garage in square feet
4. Garage Cars: Size of garage in car capacity
5. Total Bsmt SF: Total square feet of basement area
6. 1st Flr SF: First Floor square feet

The model was used to predict home prices in Ames and was cross-validated with metric values:



A train/test split was then performed, training data and testing data were scored, training data was cross validated and predictions were generated with metric values:



Predicted and observed home prices in Ames were visualized and observed to have a somewhat linear relationship and high values for R-squared, the correlation coefficient and lowered values for the root mean squared error, a loss function that is used to indicate the accuracy of a model.

**3) Model Tuning**

Categorical variables were chosen based upon the visibility to a potential home-buyer upon seeing the property for the first time. Emotional home-buying decisions are common in the real-estate market and the following property variables were factored into the model due to their potential impact on ones first impression of a home:

1. Neighborhood: Physical locations within Ames city limitsGr Liv Area: Above grade (ground) living area square feet
2. Bldg Type: Type of dwelling
3. House Style: Style of dwelling
4. Exter Qual: Evaluates the quality of the material on the exterior
5. Exter Cond: Evaluates the present condition of the material on the exterior
6. Kitchen Qual: Kitchen quality
7. Garage Qual: Garage quality
8. Garage Cond: Garage quality
9. Paved Drive: Paved driveway

Tuning the model and adding in categorical dummy variables resulted in higher R-squared values and lowered root mean squared error, two indicators of a more accurate model:



**Business Recommendations**

The feature that was found to have the largest positive impact on home sale price was the garage condition. Specifically, the garage had a large positive effect on home sale price if it was in poor shape, fair shape, typical/average shape or good shape, not excellent shape.

The feature that was found to have the largest negative impact on home sale price was the garage quality. Specifically, the garage quality had a negative effect on home sale price if it was in poor shape, fair shape, or typical/average shape; not good shape or excellent shape.

Garage quality and garage condition are both ordinal variables that can take on the following values:

Ex) Excellent

Gd) Good

TA) Typical/Average

Fa) Fair

Po) Poor

NA) No garage

Whether or not the home was located in Old Town also had a negative impact on home sale price.

Ames is a college town directly in the middle of Iowa and it’s top 3 employers are Iowa State University, the City of Ames and Mary Greely Medical Center. The closest comparable cities to Ames are Marshalltown, IA and Des Moines, IA. Marshalltown is a smaller town with a population about one-third the size of Ames. Des Moines is the capital city of Iowa and a hub for the insurance industry. I believe that the model I created is biased to fit the Ames real-estate market specifically and will not generalize well to other cities with differing socioeconomic qualities and characteristics.

In order to increase the value of a home in Ames and improve its sale price I would recommend improving the garage condition as well as improving the overall condition in other areas of the home such as the kitchen, living room, bathrooms, yard, driveway, patio, etc. Do not hesitate to spend money improving these areas of a home, making home investments over the course of time will improve your return on investment and the speed at which the home sells when the time comes.