

***Housing Price Prediction Project***

Submitted by:

P.V.RAGHAVULU

# ACKNOWLEDGMENT

Thanks for giving me the opportunity to work in Flip Robo Technologies as Intern and would like to express my gratitude to Data Trained Institute as well for trained me in Data Science Domain. This helps me to do my projects well and understand the concepts.

Resources Referred – Google, GitHub, Blogs for conceptual referring.

# INTRODUCTION

## Business Problem Framing

A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia.

The company is looking at prospective properties to buy houses to enter the market. You are required to build a model using Machine Learning in order to predict the actual value of the prospective properties and decide whether to invest in them or not.

For this company wants to know:

* + Which variables are important to predict the price of variable?
  + How do these variables describe the price of the house?

## Conceptual Background of the Domain Problem

Houses are one of the necessary needs of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market and there are various companies working in the domain. The company is looking at prospective properties to buy houses to enter the market.

## Motivation for the Problem Undertaken

* Houses are one of the necessary need of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market and there are various companies working in the domain. Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

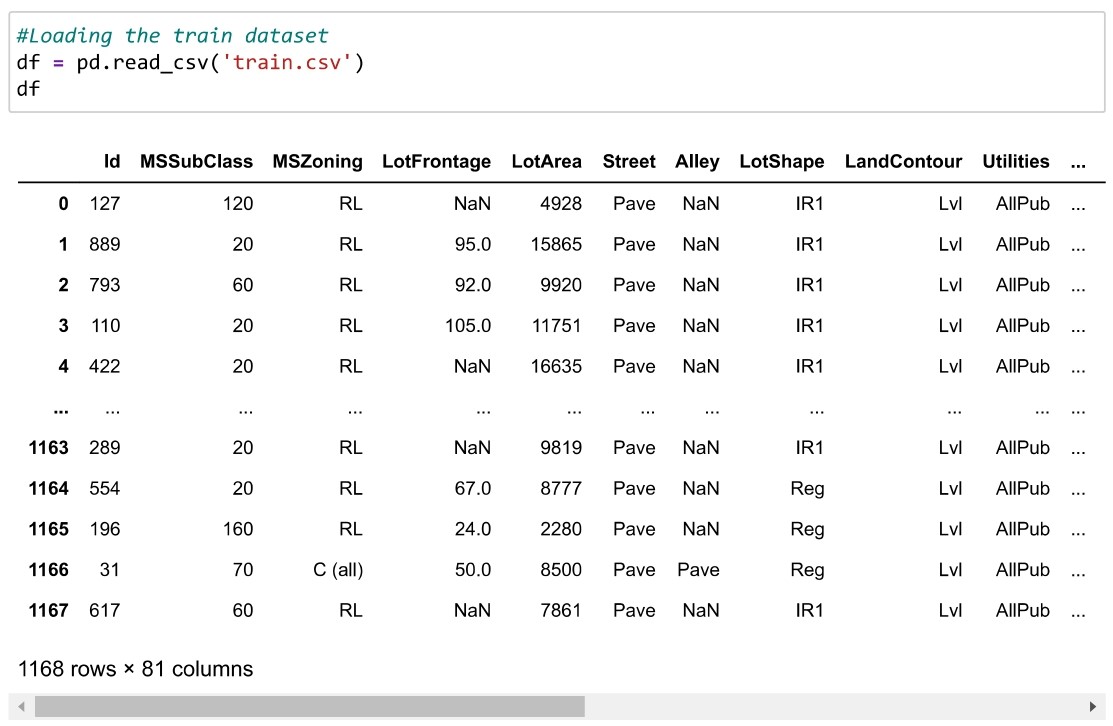
# Analytical Problem Framing

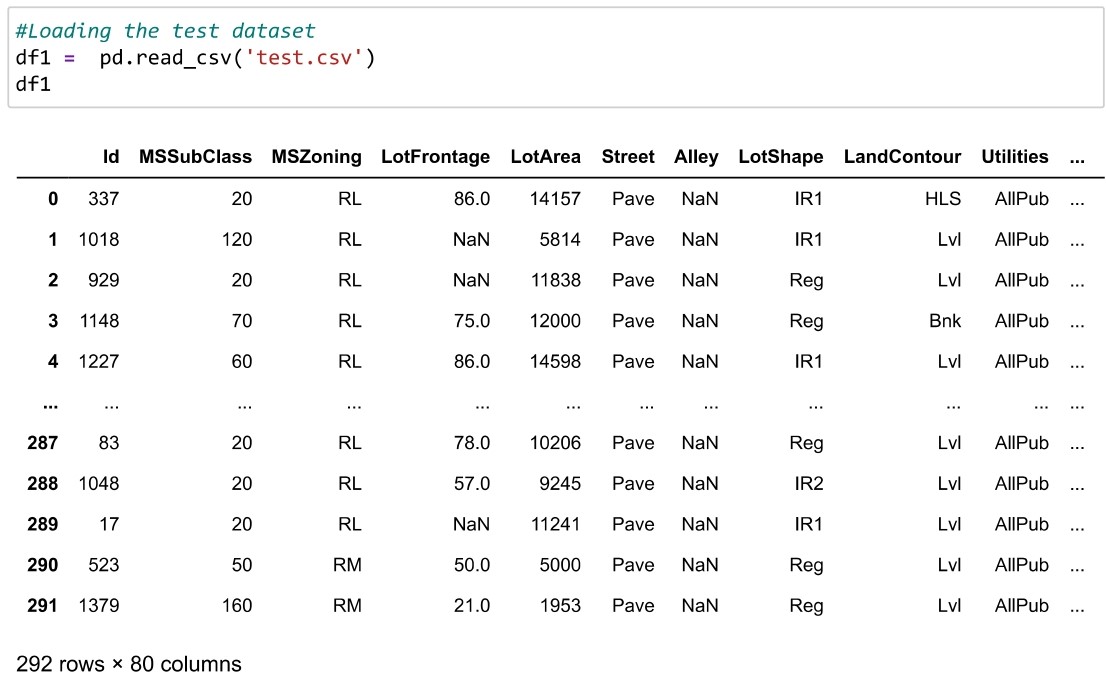
## Mathematical/ Analytical Modelling of the Problem

Here our target variable is the “Sale Price” and need to predict the prices of the house. As the data is continuous and our problem is Regression.

## Data Sources and their formats

The Data is provided by Flip Robo Technologies, and it has Train and Test Data Set and need to train our data in Train dataset and need to load the Test dataset to make the predictions.



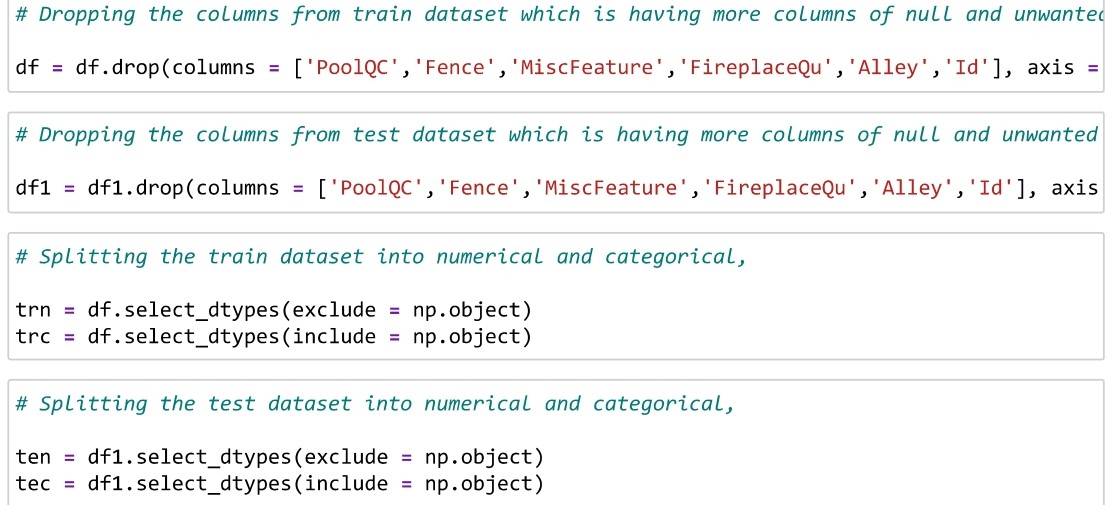


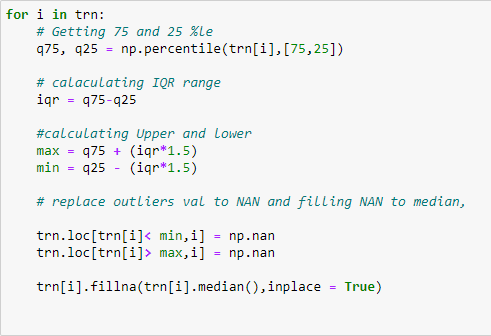
Data is having Null values and We need to treat the missing values and can be discussed in pre-processing.

## Data Pre-processing Done

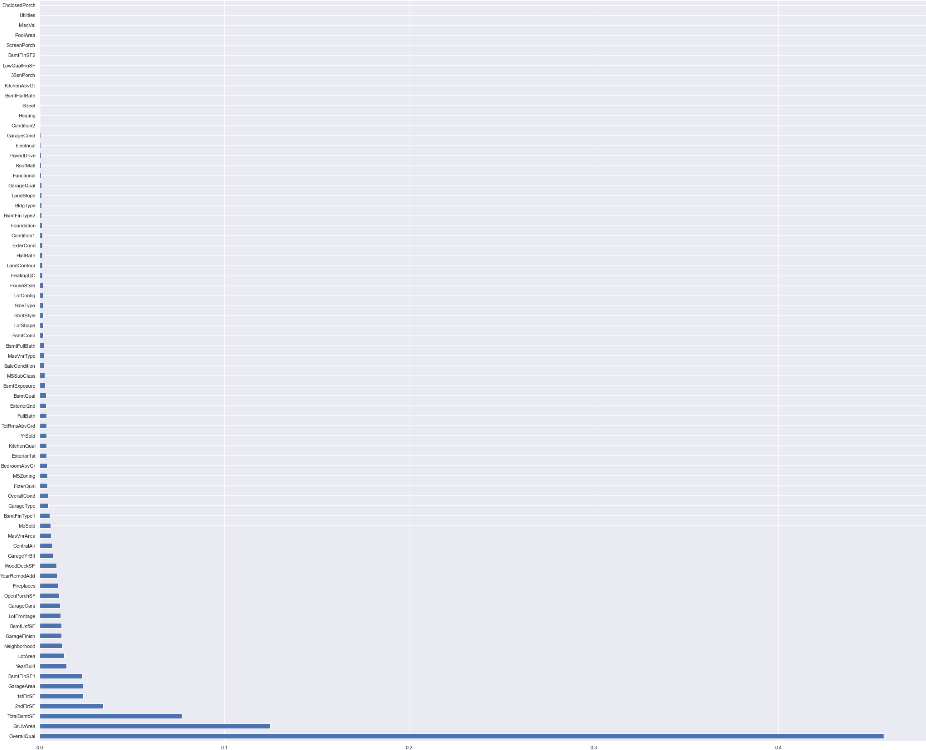
## 

## 









* Data Inputs- Logic- Output Relationships

## The below plot shows that rating of over all conditions of house which is greater than 5 has price of

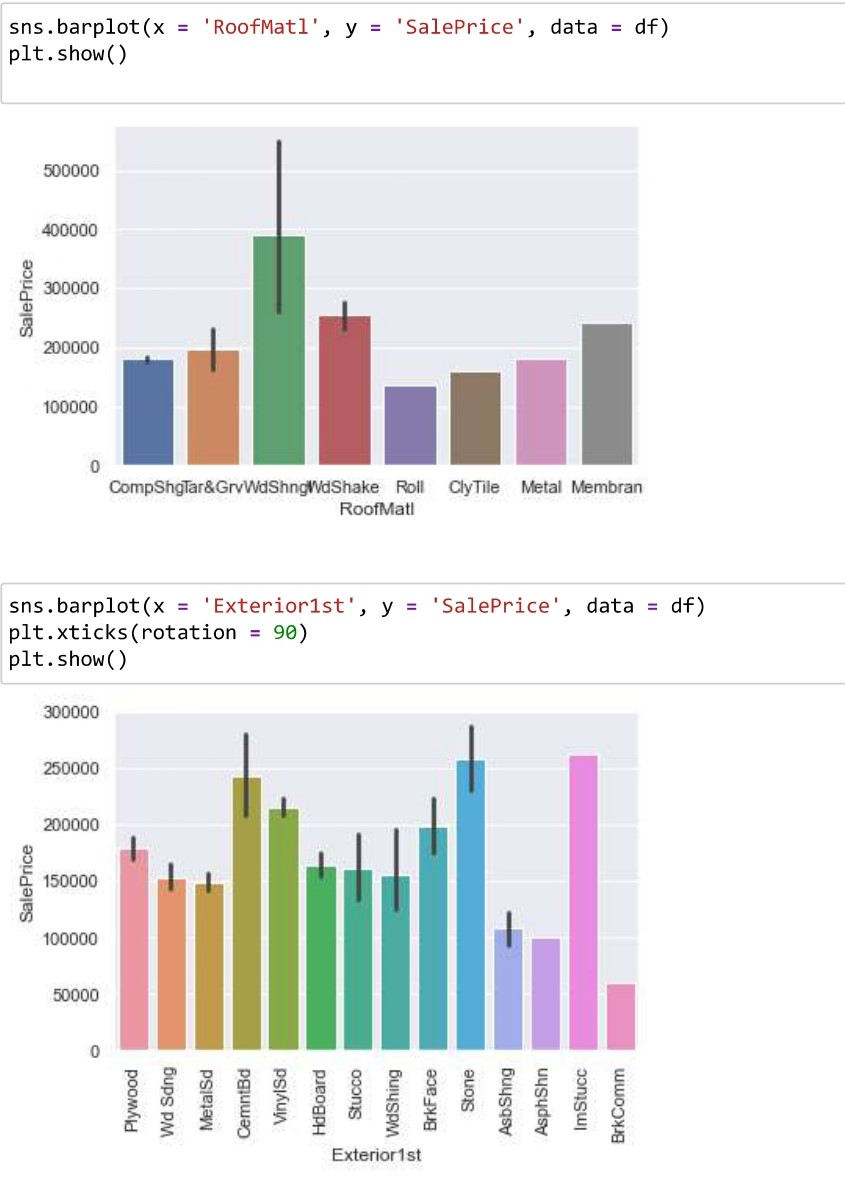
> 1.5L and the house which has rating of >8 of material and finishing of house price is >30L.

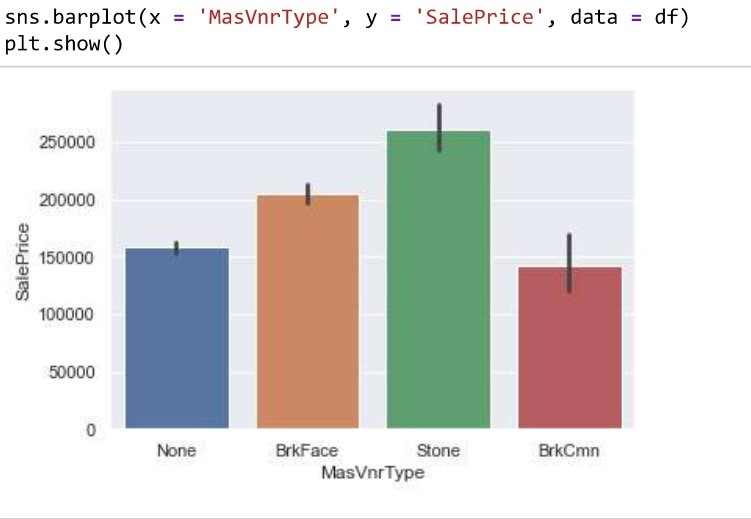
## Also, the plot shows that most of the house roof style is shed / hip / flat.



The below plot explains that most of the roof material in houses are Wood Shingles and the exterior used on houses are Imitation Stucco and Masonry veneer type of most of the houses are Stone and the houses which is satisfying these conditions are selling the house price

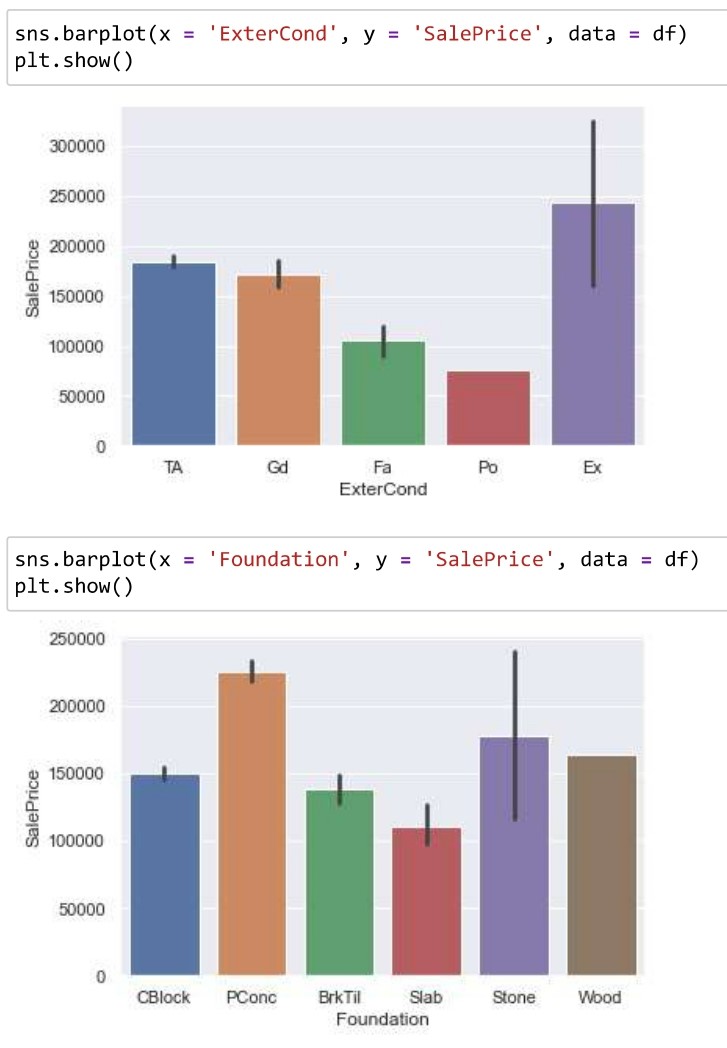
## >25L.

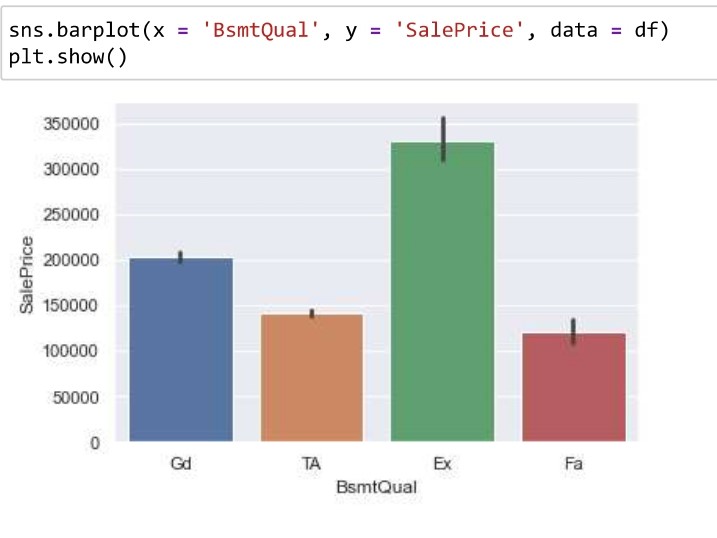




The exterior condition and basement of house which is having rating of Excellent has price of >20L.

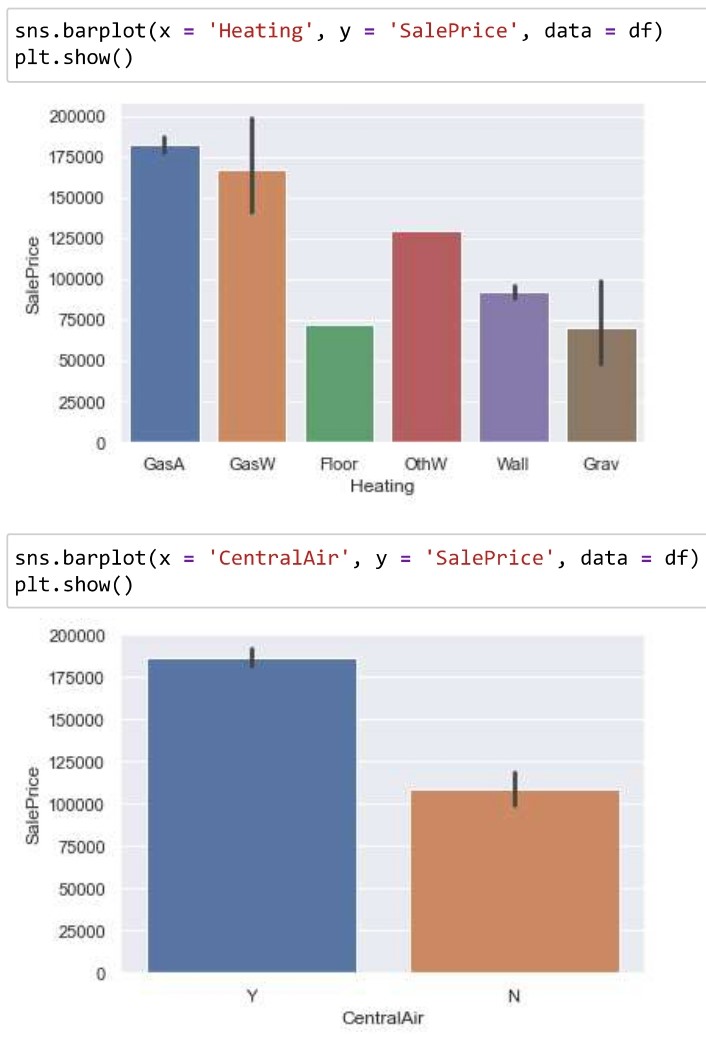
## Most of the house possess the foundation type as Poured Concrete and it has the high price of >20L.

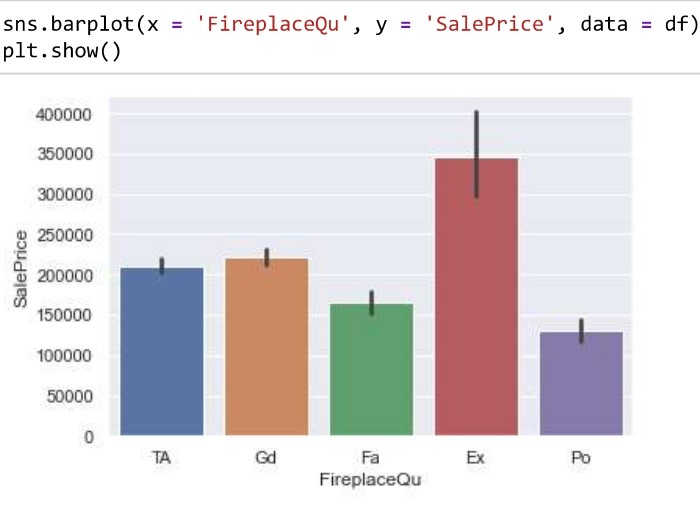




The plot shows that most of the house heating type is Gas forced warm air furnace and central air conditioned and fireplace quality should be excellent has price of

## >1.5L.





* Hardware and Software Requirements and Tools Used

Model training was done on Jupiter Notebook. Kernel Version is Python3.

# Model/s Development and Evaluation

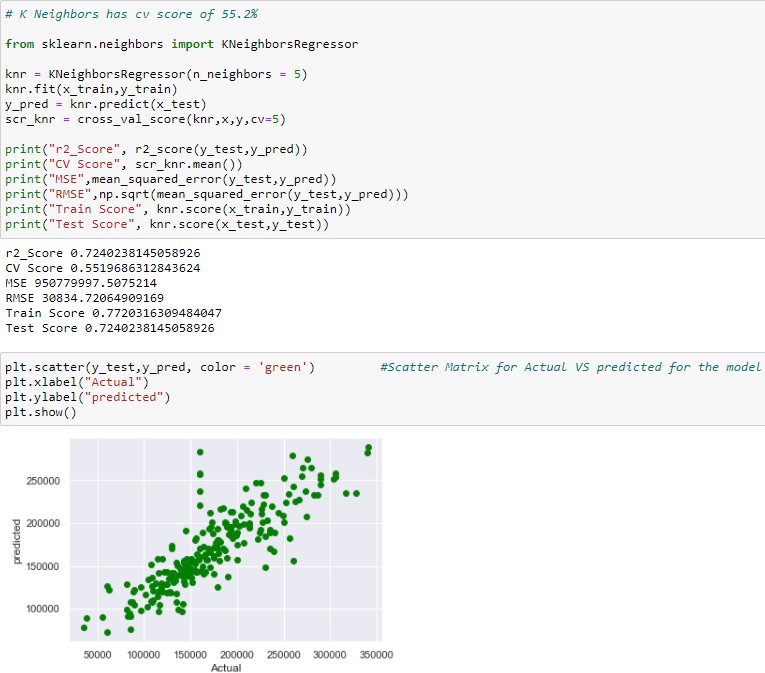
## Testing of Identified Approaches (Algorithms)

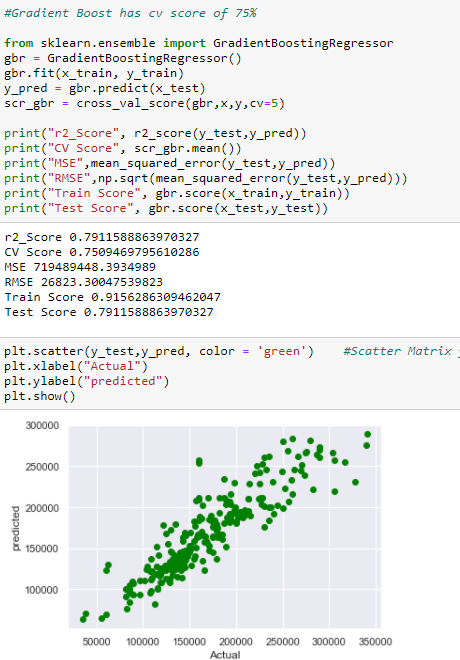
Random Forest K Neighbours Ada Boost Gradient Boost

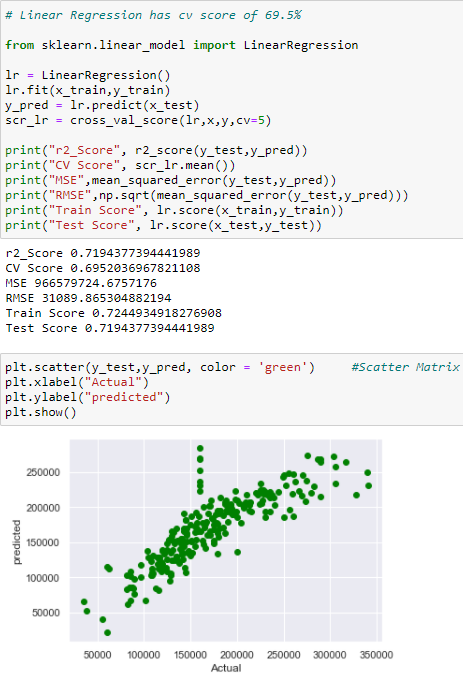
Linear Regression

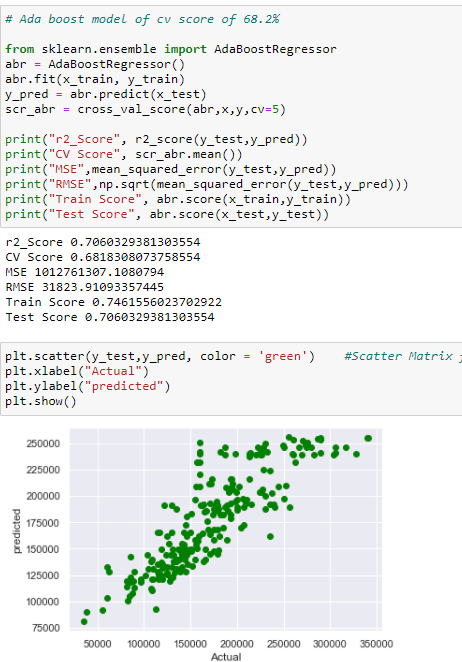
## Text Description automatically generatedRun and evaluate selected models









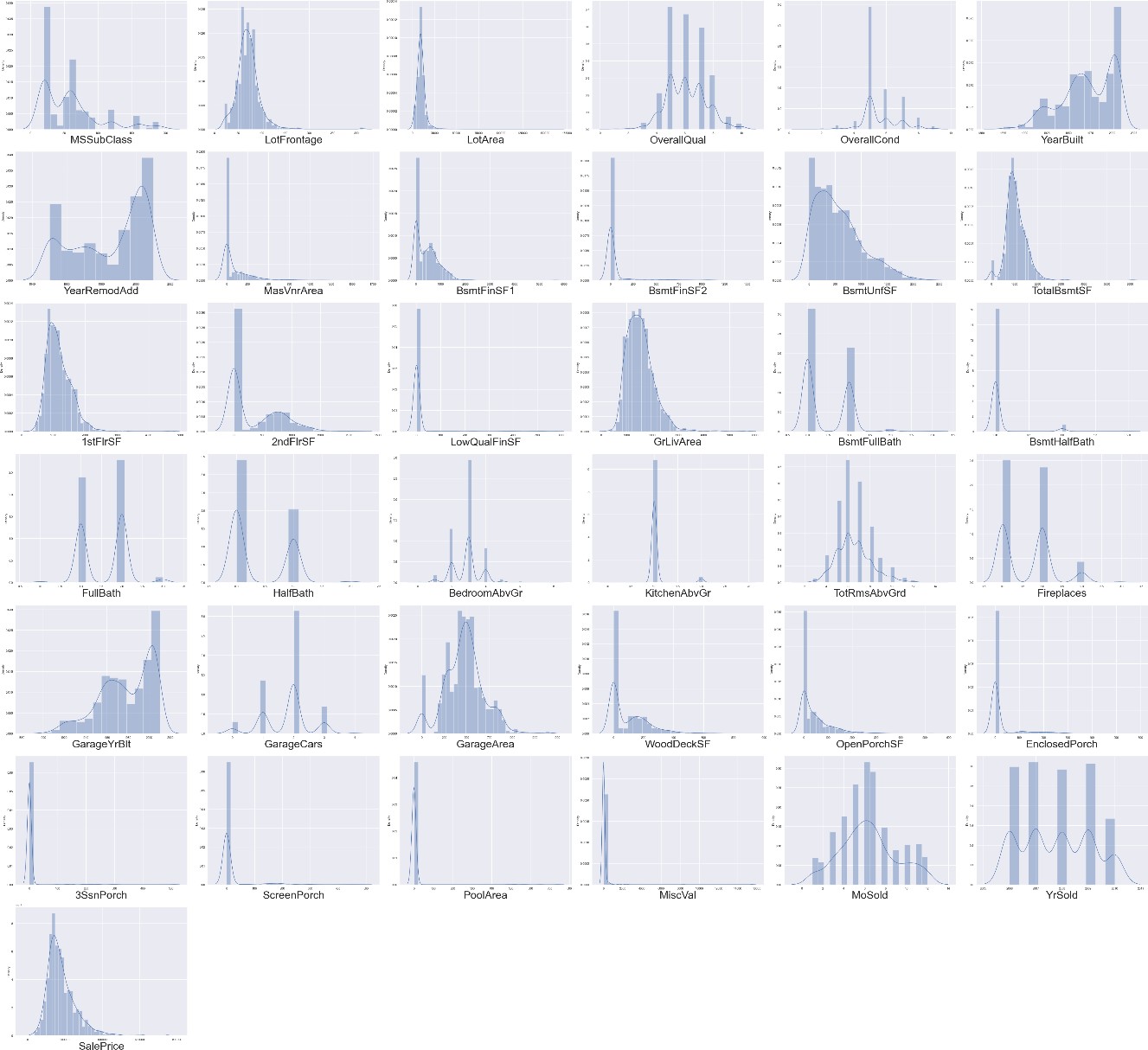


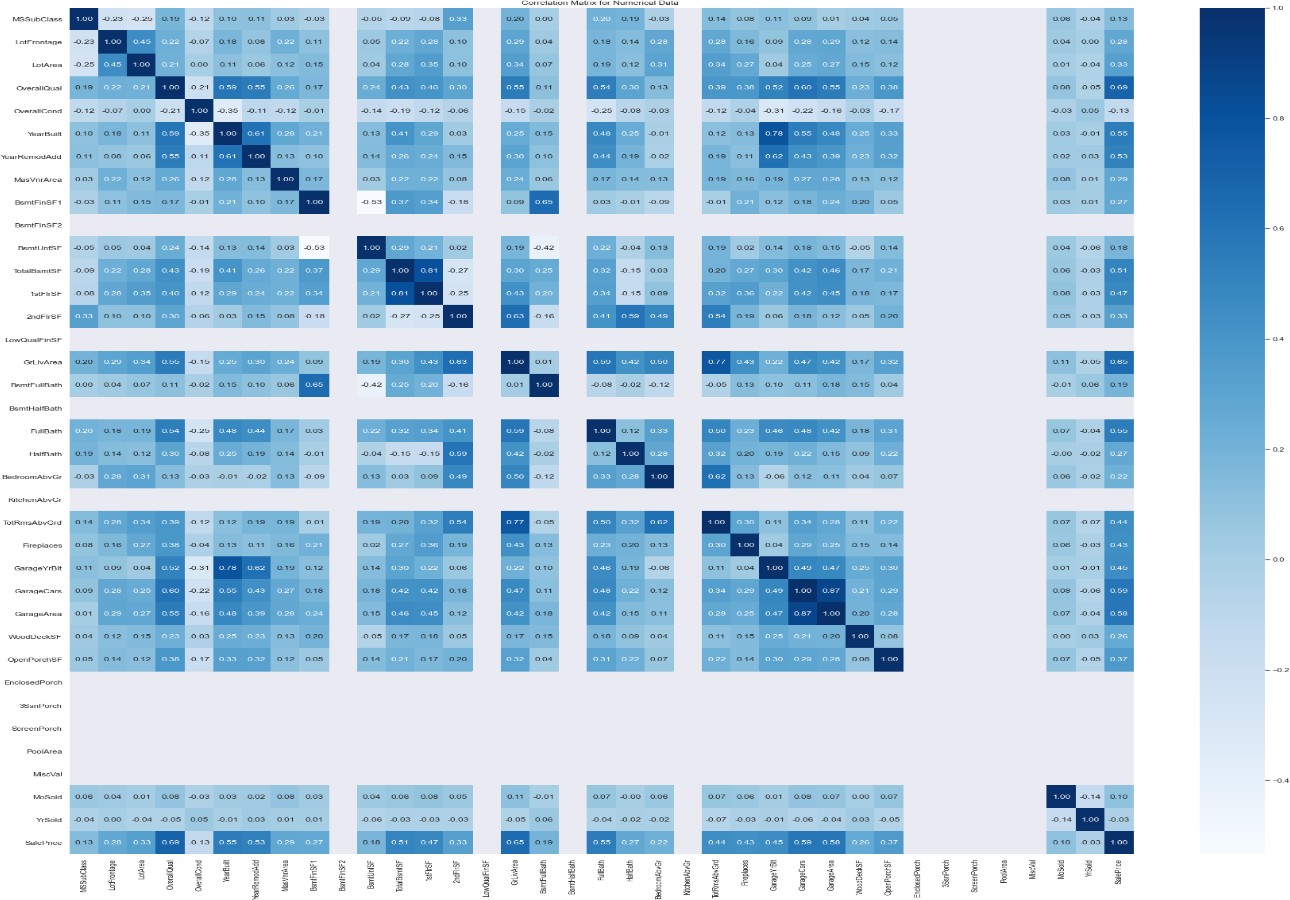
* Visualizations

## Box plot and Dist. plot,

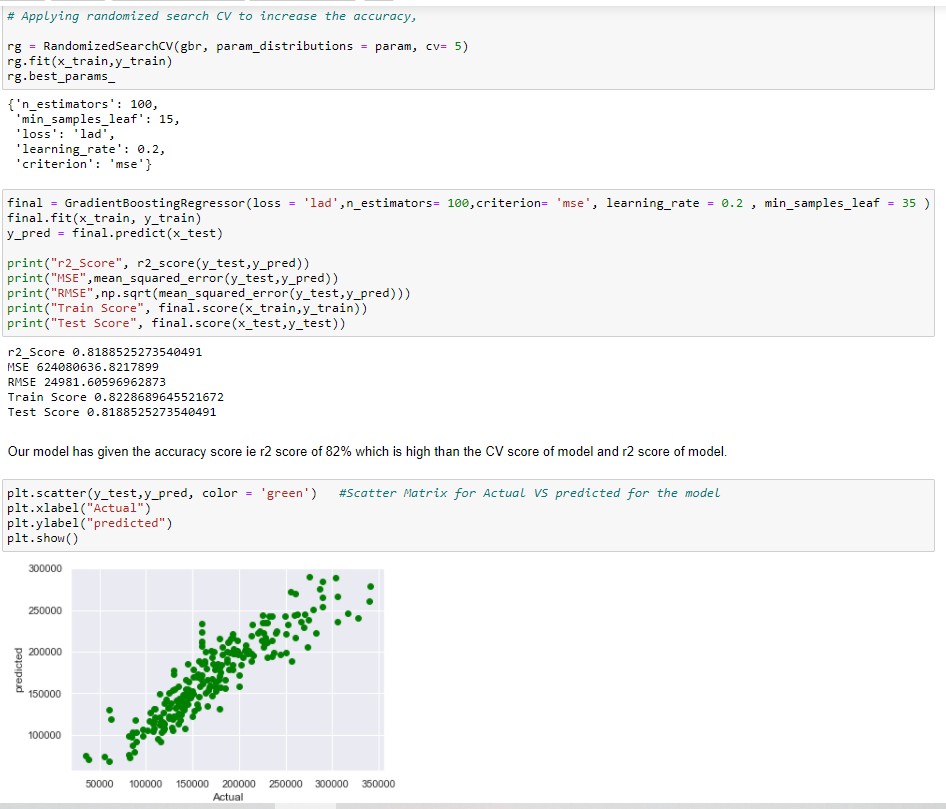








* Interpretation of the Results



# CONCLUSION

## Key Findings and Conclusions of the Study

As this project is about predicting the price of house in US market, The problem is Regression and I have used 5 algorithms to build the model and among all Gradient Boosting algorithm is the best model.

I have used hyper parameter tuning to increase the accuracy score as well and price which predicted was slightly varying from the actual price.

We have 2 datasets for training and testing the data and it has more columns but less rows and removed the columns which is less and no importance using feature importance feature engineering.

So, this will helps the clients to understand the variables which are important to predict the price of house and helps them to get success in their business.

## Learning Outcomes of the Study in respect of Data Science

Random forest and Boosting algorithms work best and Randomized search CV is faster than Grid search CV.

As this dataset has 2 data files and we need to do all pre- processing, EDA to both datasets and train dataset has target variable and test dataset will not have target variable which we need to predict it.

This is kind of different and time taking process but helps me to learn more and most important is that I am getting hands-on experience more on Data Science Concepts.