***House Price Prediction Using Machine***

***Learning Techniques***

# Acknowledgment

In This Project, I Have Investigated All Factors That Affect The House Price And Buying Decision. I Have Analyzed All The Factors That Affect The House Prices. For This Purpose. I Am Very Thankful To My Friends And Family Who Helped Me Through This Study.

# Introduction

The real estate sector is an important industry with many stakeholders ranging from regulatory bodies to private companies and investors. Among these stakeholders, there is a high demand for a better understanding of the industry operational mechanism and driving factors. Investment is a business activity that most people are interested in this globalization era. There are several objects that are often used for investment, for example, gold, stocks and property. In particular, property investment has increased. In determining the price of home, the developer must calculate carefully and determine the appropriate method because property prices always increase continuously and almost never fall in the long term or short. There are several approaches that can be used to determine the price of the house, one of them is the prediction analysis can be observed by human senses, including the size of the house, the number of bedrooms, the availability of kitchen and garage, the availability of the garden, the area of land and buildings, and the age of the house, while the concept is an idea offered by developers who can attract potential buyers, for example, the concept of a minimalist home, healthy and green environment, and elite environment.

# Problem Statement

The real estate sector Emerged Rapidly from Past Decades and It Has Grown Faster During Past Years. The Competition Has Increased. In The Present the

Market the demands is increased day by day.

* In Such Competitive Market It’s Very Difficult to predict the exact price of the house because The Market Changed By Day To Day.
* Understanding Various Factors That Influence Price Of the house .
* Understanding Customer’s Behaviors Before Buying a house.

# Hardware And Software Requirements And Tools Used

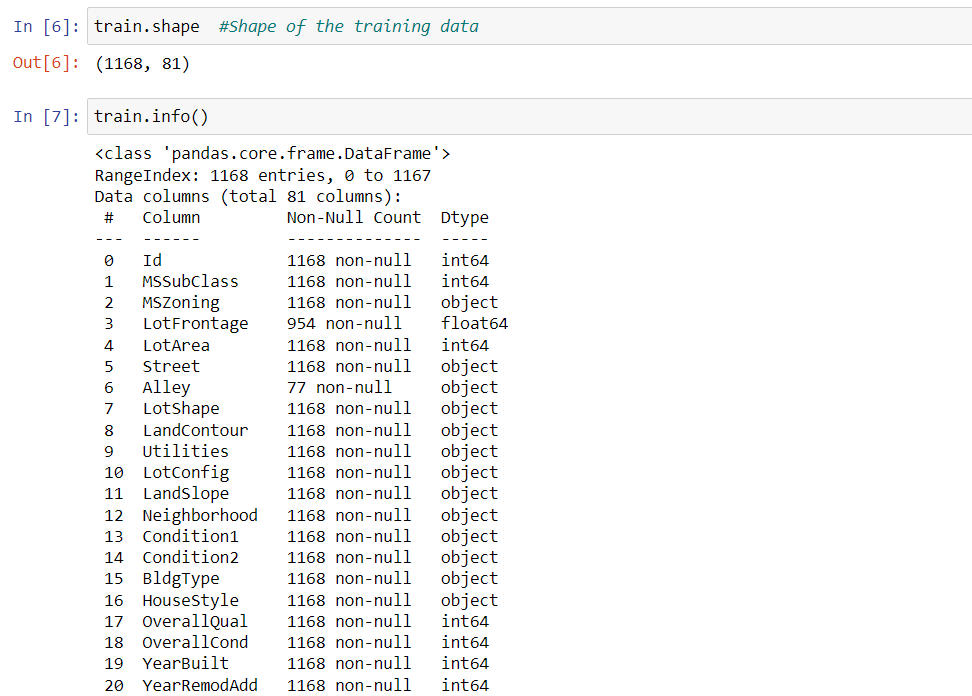
**Hardware Requirements**: We Need A Laptop With Minimum 4 Gb Ram And 500 Gb Of Heard Dish.

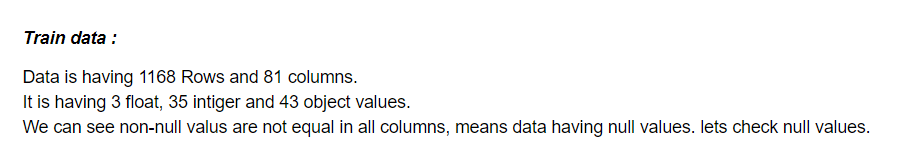
**Software Requirement**: We Need Anaconda Installed In Our Hardware. It Comes With All The Packages Required For Data Analysis And Visualization. Anaconda Having Jupyter Notebook, The Best Environment For Data Analysis. Library Used: We Have Used Mainly Four Libraries For Data Analysis, Mathematical Calculation, And Visualization Of Data. Numpy Is Used For Numerical Calculation And Pandas Is Used For Making Data Frame And Pre- Processing Of The Data. For The Visualization Part, We Have Used Matplotlib And Seaborn Package. Both The Package Provide A Wide Variety Of The Graphs For Data Visualization And Analysis. Coding Language: Python

**Analytical Summary**

# Dataset For This Project

For This Project, We Are Using the US-based housing company named Surprise Housing House Dataset that contain 1460 rows and 81 columns

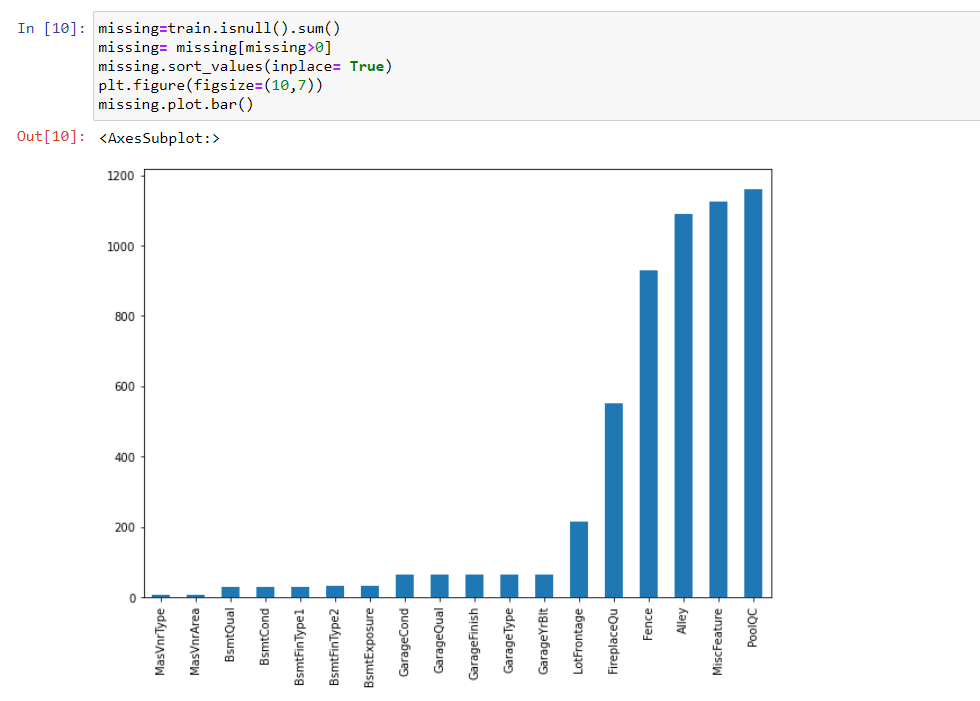




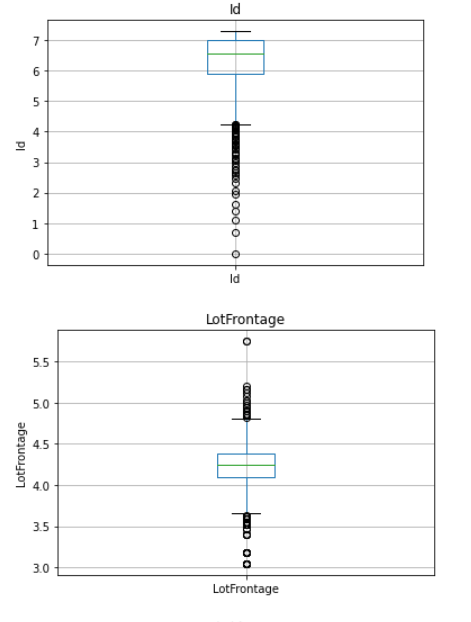
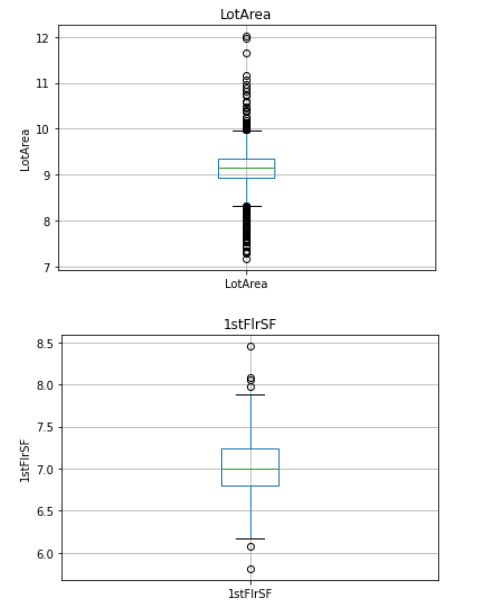
**Data Preprocessing Steps**

In preprocessing stage, we analysis our dataset and perform the various type of analysis

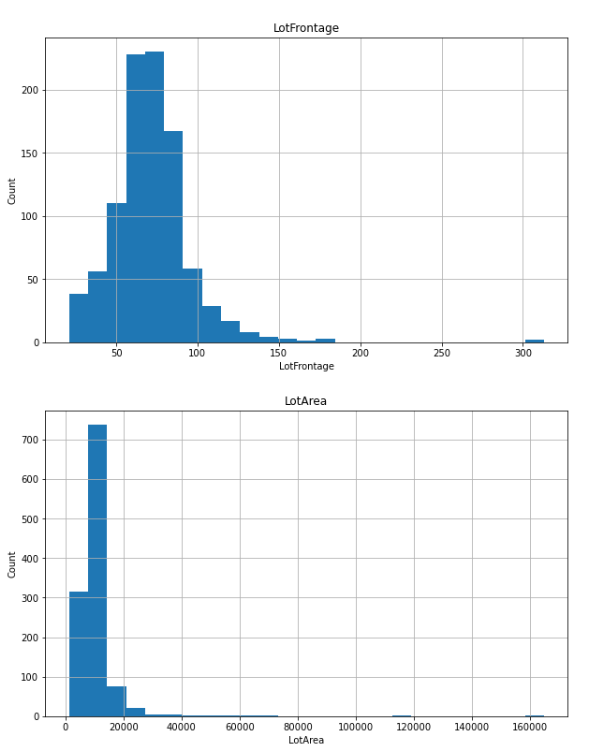
* We check our dataset contain any null value and we found the many columns and rows contain the null value and we treat all values



* we found the outliers in different columns and we remove the outliers using these Z- Score

* we also found the skewness in different columns. We need to treat them.



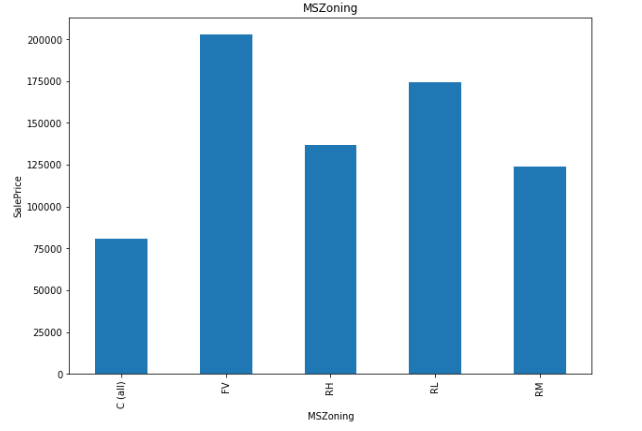
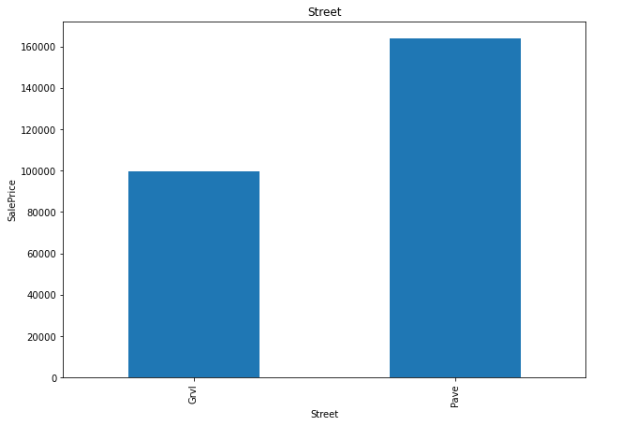
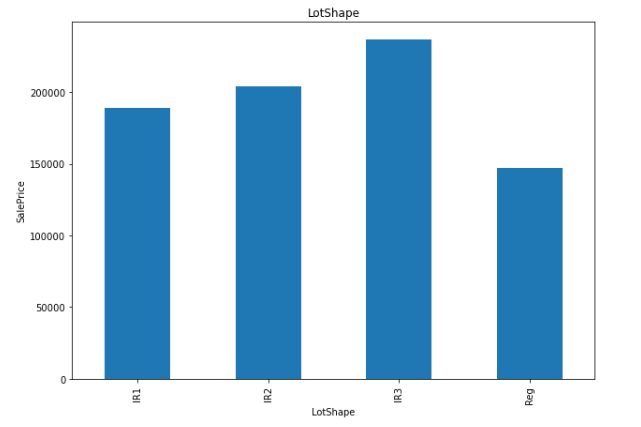
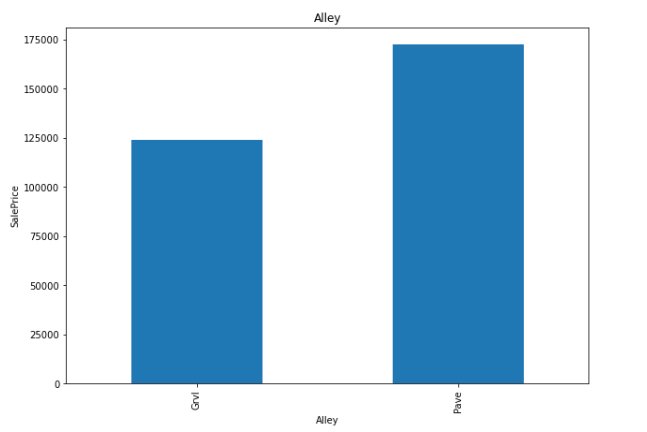
* Also, we plot the various type of plots to find the inside of the data and also check the relation of each column to the price or target columns.
* Our dataset contain many columns object columns and we need to Encode it
* we need to scale the values
* We have 81 features so we check the multicollinearity of the columns.

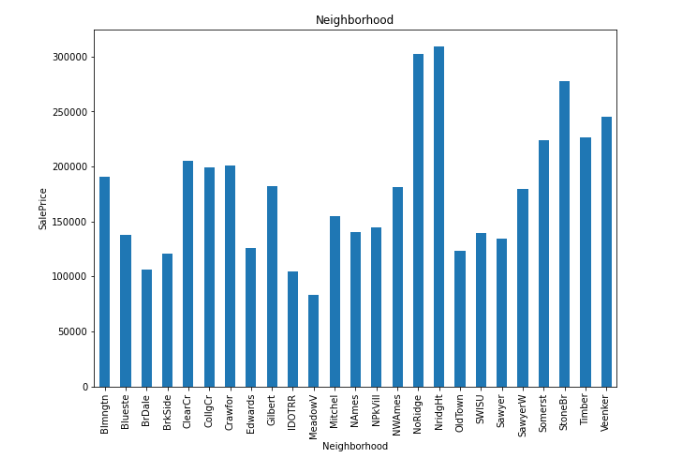
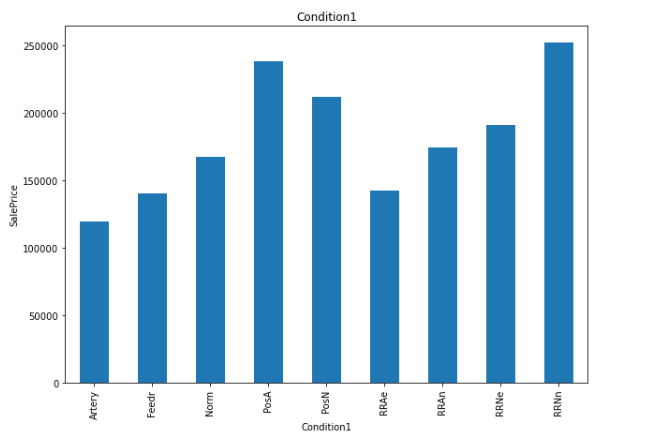
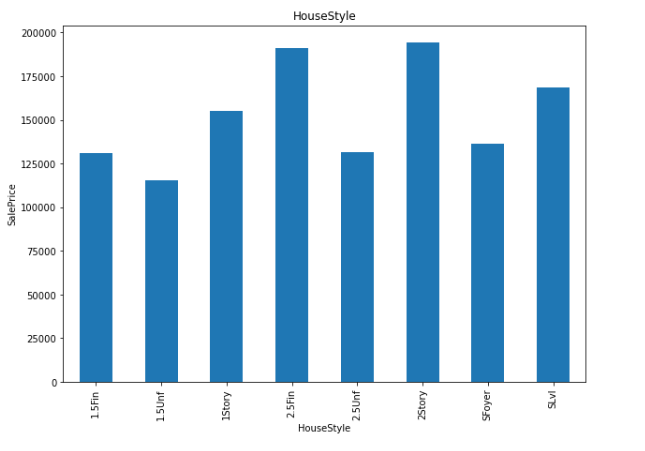
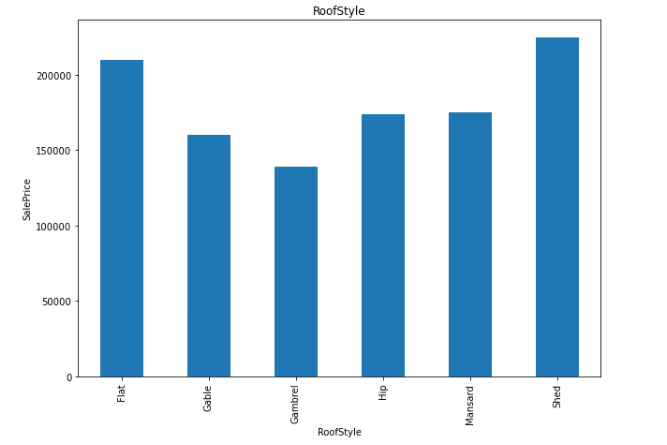
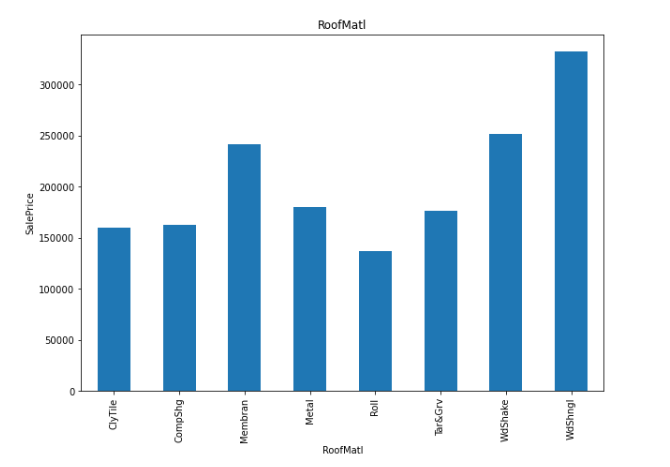
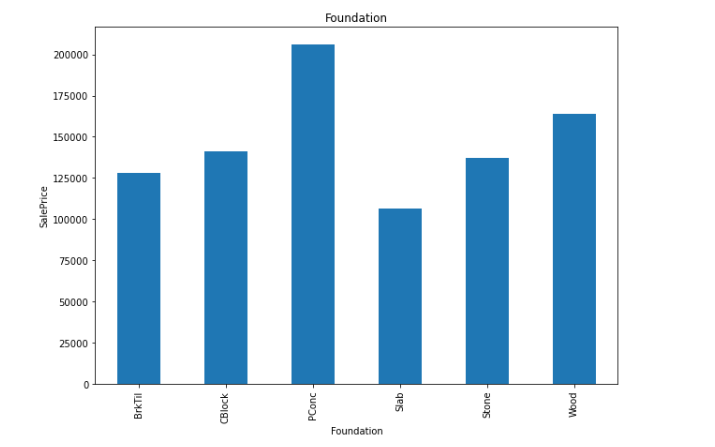
# Conceptual Background Of The Domain Problem

Our Main Problem Statement Is To Investigate The Factors Affecting and Impact The House Price. Due To Competitive Market, It Becomes More difficult to predict the exact house price base the price current market because this field is changed day by day but we crate a model and tried to predict the price based on the given features in the dataset. To Select The Fair Price. During Investigation We Will Study Factors Like No. of rooms , Agriculture , Industrial , shape of property , Near Flat/Level , public Utilities , Rates the overall material and finish of the house , Rates the overall condition of the house, Original construction date , Type of roof , Roof material , Exterior covering on house , quality of the material on the exterior , Type of foundation , Kitchen quality , Total rooms above grade , Fireplace quality , Garage location , Pool quality , Type of sale and many more we use these features to crate the model and predict the price .

**Output Relationships**

We perforce the various type of plots and fine the relation between the feature and price of the hours.

# CONCLUSION

## Key Findings and Conclusions of the Study

The main goal of this project is to determine the prediction for prices which we have successfully done using different machine learning algorithms like a Random forest, multiple regression, Support vector machine, gradient boosted trees. And research provides us to find the attributes contribution in prediction. So, I would believe this research will be helpful for the peoples.

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

It was very interesting to study and visualize the data using tools available in python. We have visualized the data very deeply and come to the conclusion of what customers need and what factors cause to select the house prices.

## Limitations of this work and Scope for Future Work

The data set was very small and all the conclusions are based on this small dataset. For better and more clarity, we can perform the same steps on the big dataset to make some clear and more accurate decisions.