USA House Price Analysis

Bhavya Yasaswini Gadu

Abstract:

The real estate markets show much interest to analyse and predict where property prices are moving towards. Prediction of property prices is becoming increasingly important and beneficial. It hight affects the market's growth and are also a good indicator of both the overall market condition and the economic health of a country. Using the data set available, analysis was made on how different features affects to house prices in USA.

Dataset:

1. The dataset is downloaded from Kaggle website and it contains the house prices in the year 2014 for different cities and streets of USA.
2. Dimensions - 4600 X 17
3. Features - Date, Price, Bedrooms, Bathrooms, Sqft_living, Sqft_lot, Floor, Waterfront, View, Condition, Sqft_above, Sqft_basement, Yr_built, Yr_renovated, Street, City, Statezip, Country

Approach:

- The various python libraries such as Numpy, pandas, Matplotlib are used for the purpose of mathematical calculation, extraction of data and virtualization respectively.
- ***Data Pre-processing-***
    1. The data taken does not have any null values in any of the columns.

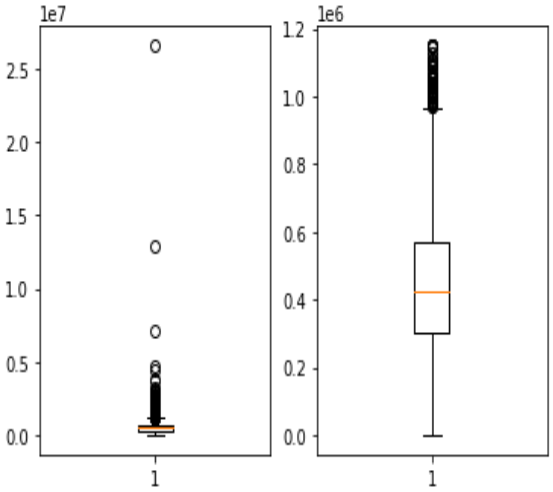
```
date             0
price            0
bedrooms         0
bathrooms        0
sqft_living      0
sqft_lot         0
floors           0
waterfront       0
view             0
condition        0
sqft_above       0
sqft_basement    0
yr_built         0
yr_renovated     0
street           0
city             0
statezip         0
country          0
dtype: int64
```

2. Outliers are the data points that differ from other observations and are generally due to experimental errors which cause deviations while analysing the data. So, the outliers are to be found and handled which can be done using Boxplots.
3. The inter quantile range of the dataset features are-

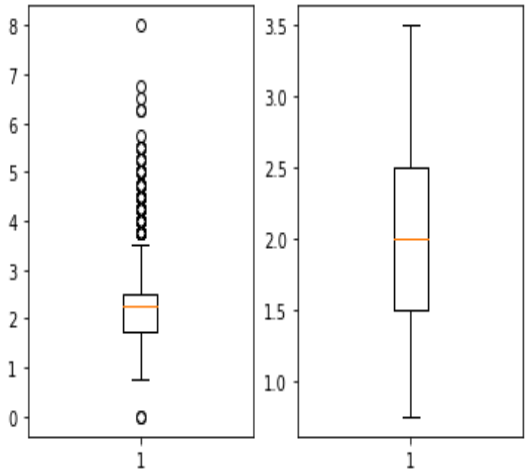
```
price          332087.50
bedrooms            1.00
bathrooms           0.75
sqft_living      1160.00
sqft_lot         6000.50
floors              1.00
waterfront          0.00
view                0.00
condition           1.00
sqft_above       1110.00
sqft_basement     610.00
yr_built           46.00
yr_renovated     1999.00
dtype: float64
```

4. The outliers are removed by eliminating the values outer the IQR. The following graphs show the boxplots before and after removing outliers.

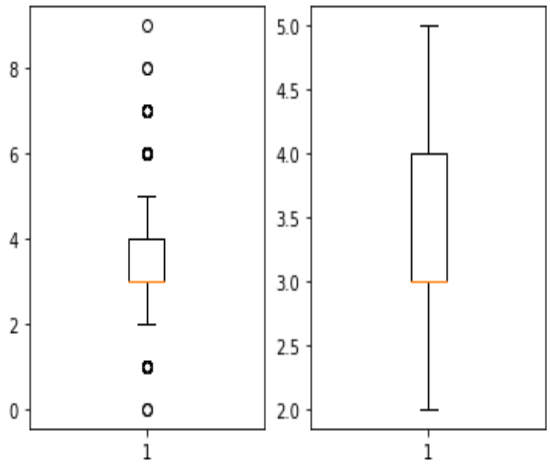
Price-



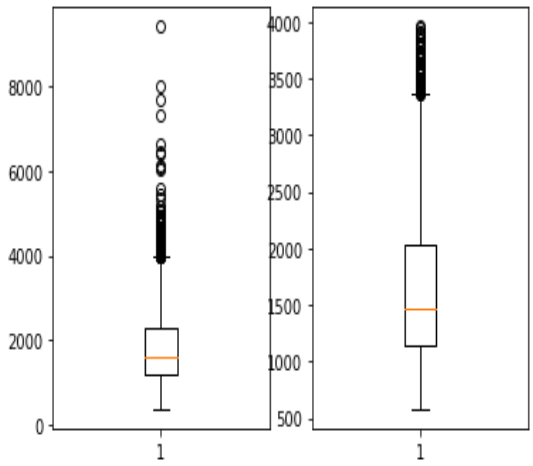
Bathrooms-



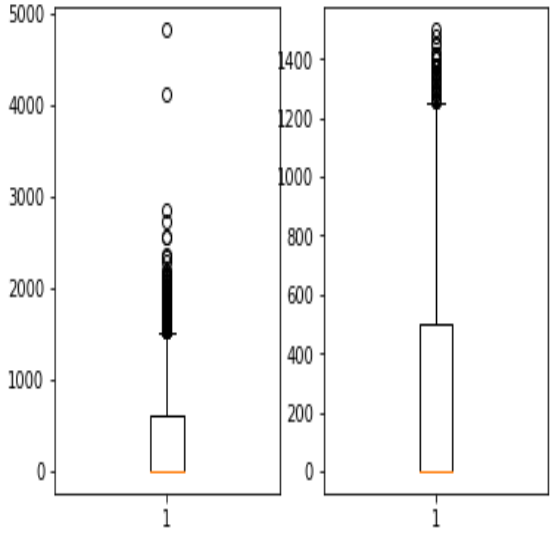
Bedrooms-



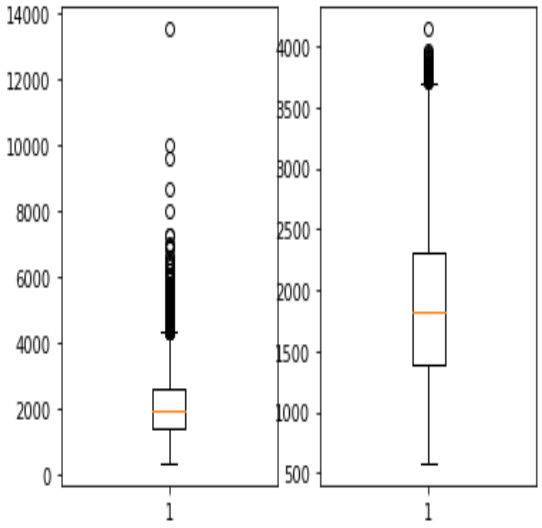
Squareft_above-



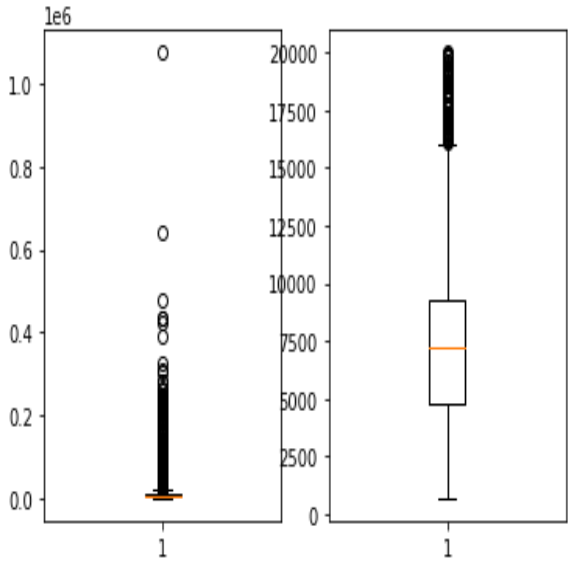
Sqft_basement-



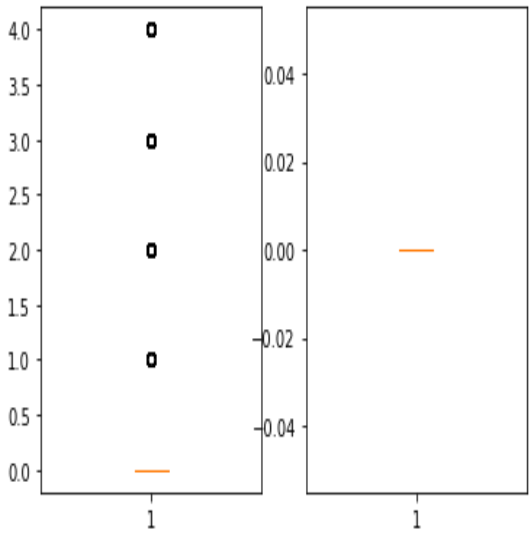
Squareft_living-



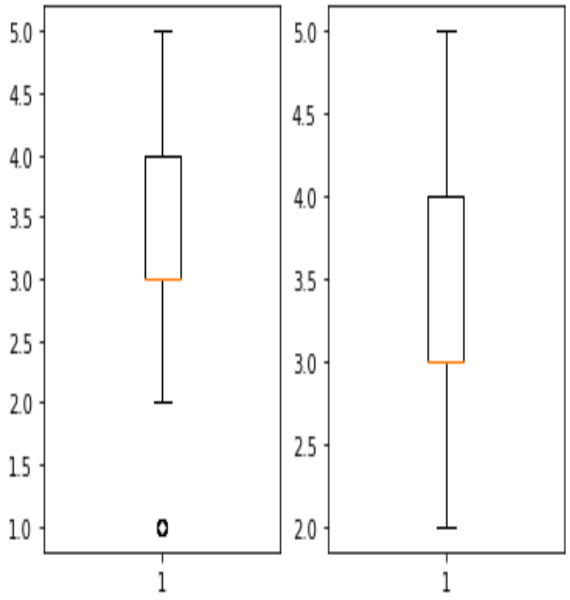
Sqft_lot-



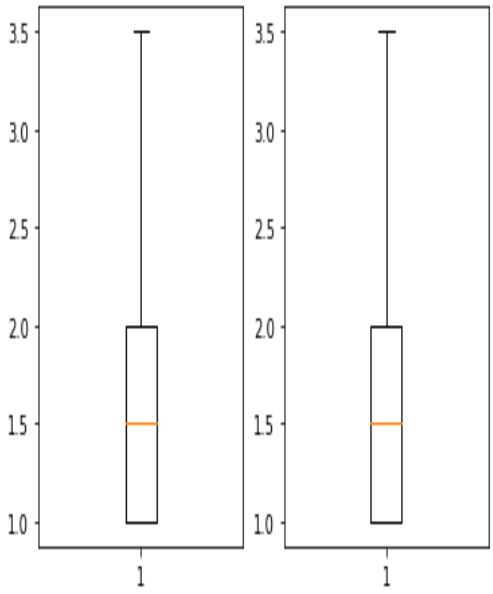
View-



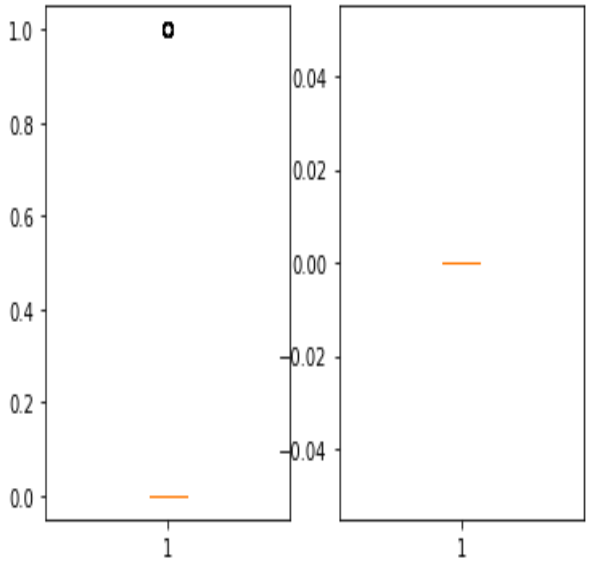
Condition-



Floors-



Waterfront-



5. It can be observed through the plots that the density of the outliers has decreased after pre-processing the data. Now, the data is made ready for making analysis.

Analysis and Visualization:

1. Top and Bottom 10 cities with highest and least average house prices-

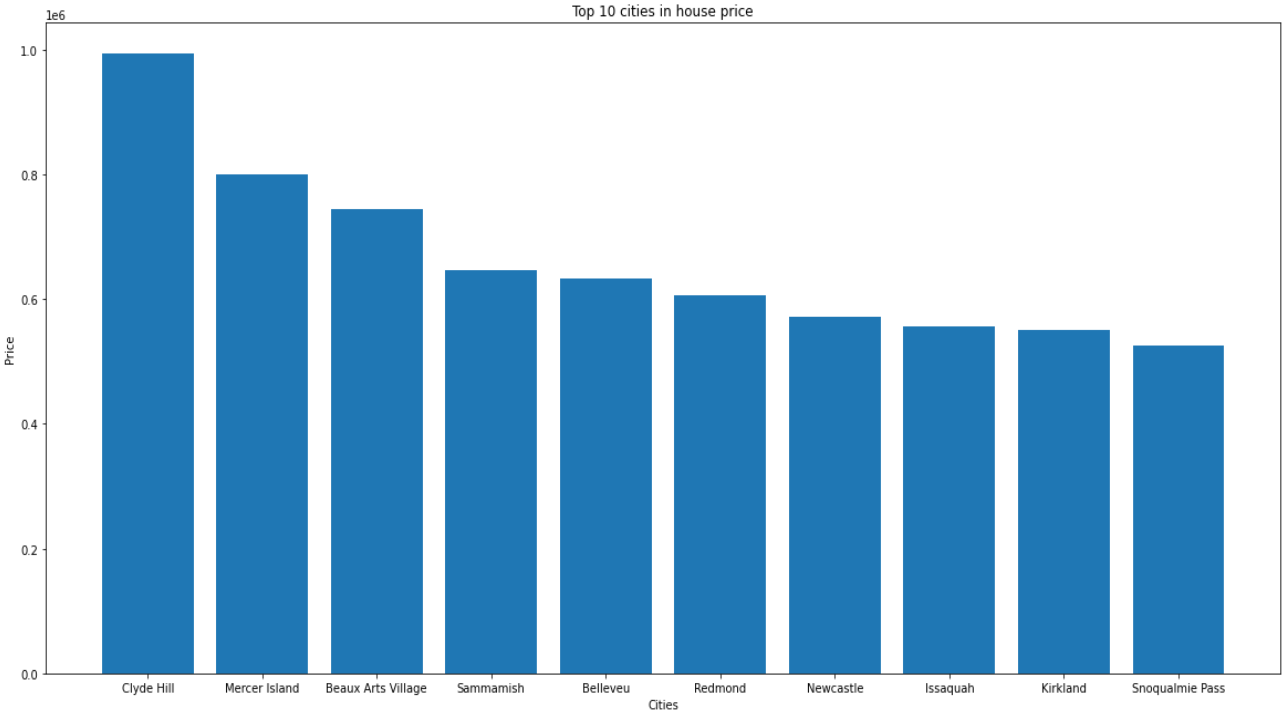
Analysis- To analyse the cost of living in different cities in USA and to know the costliest and cheapest places to live in.

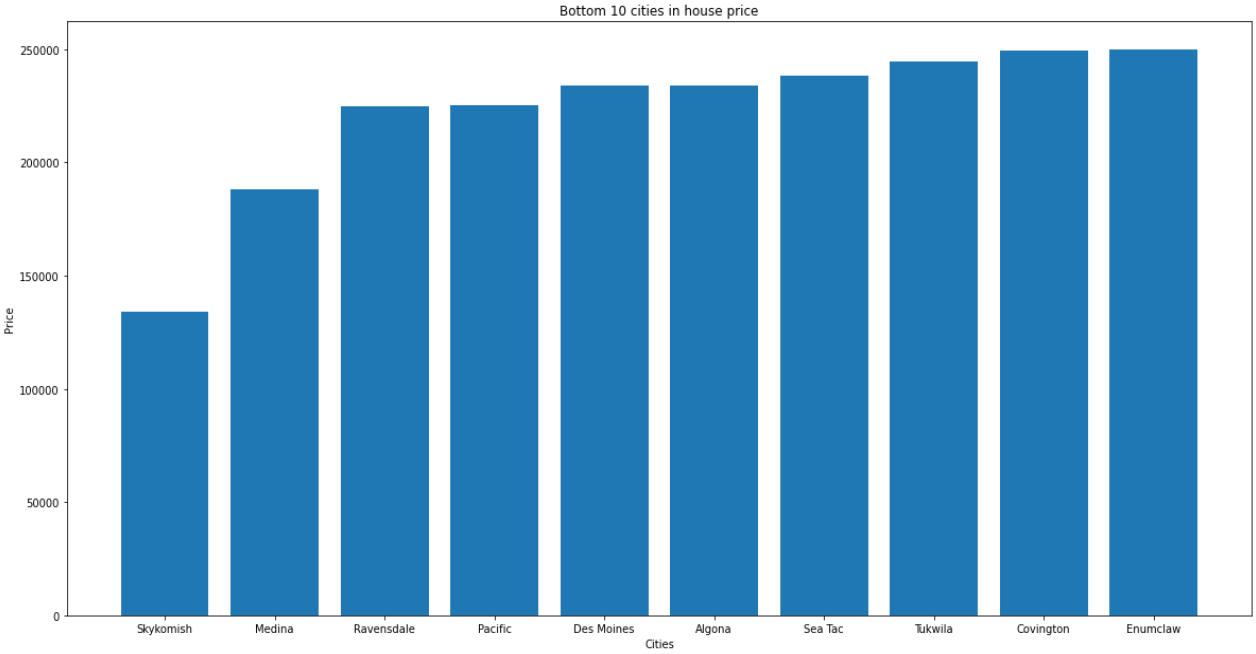
Top

city	price
Clyde Hill	994000.000000
Mercer Island	800674.561404
Beaux Arts Village	745000.000000
Sammamish	647454.671032
Bellevue	633883.650641
Redmond	605649.486708
Newcastle	571520.807692
Issaquah	555559.233165
Kirkland	550991.645012
Snoqualmie Pass	525000.000000

Bottom

city	price
Skykomish	134000.000000
Medina	188000.000000
Ravensdale	225000.000000
Pacific	225233.333333
Des Moines	233892.947368
Algona	234110.000000
SeaTac	238257.739130
Tukwila	244483.833333
Covington	249397.490347
Enumclaw	249839.880952

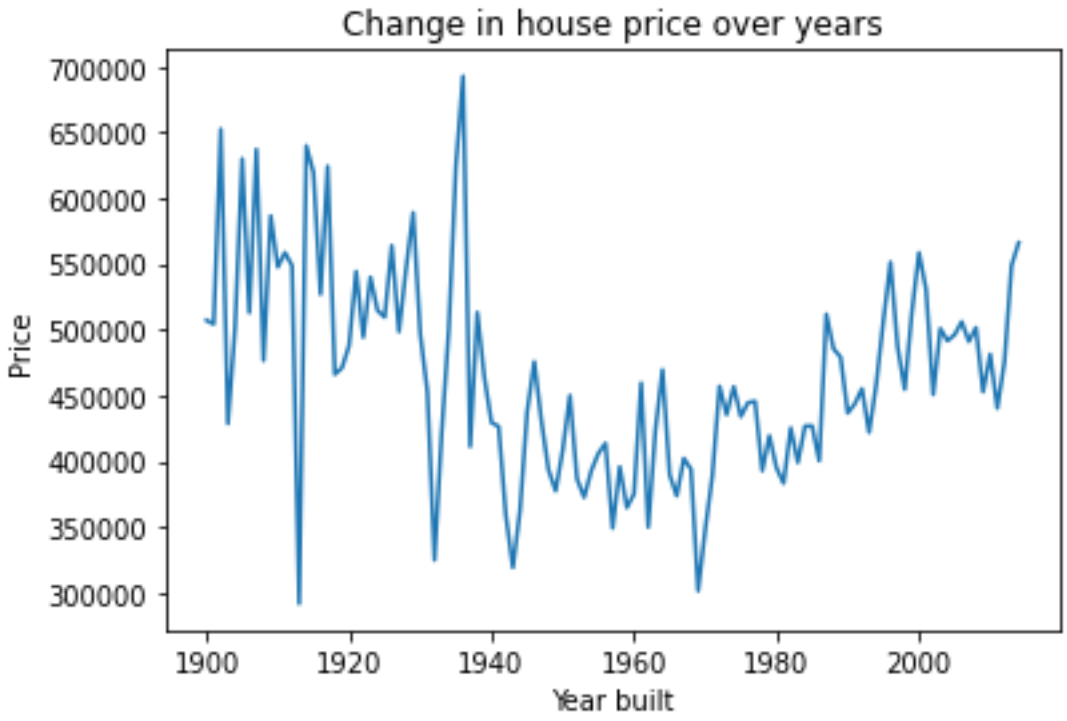




Inference- From the analysis it is observed that Clyde Hill is the costliest city with nearly $10 lakh house price and Skykomish is the cheapest city with less than $1.5 lakh house price in USA.

2. Analysis of house prices over years

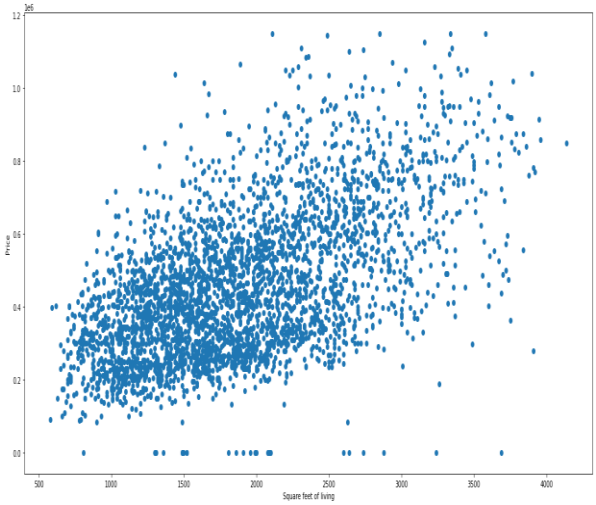
Analysis- To understand the demand of house rates over years and USA.



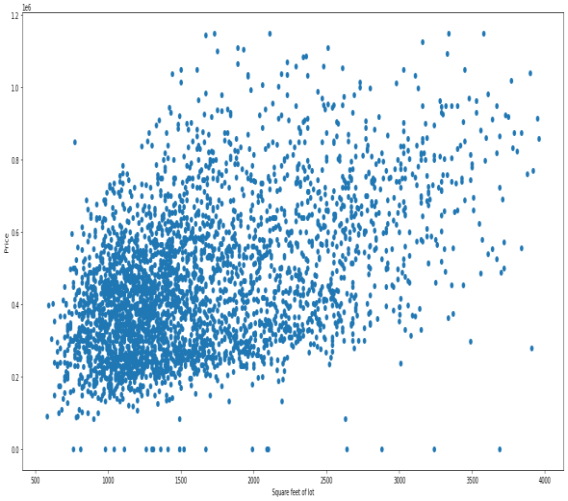
Inference- It can be observed that there is a lot of fluctuations in the demand for houses in the market from 1990 to 2010. The year 1936 has seen a highest demand for houses making the price to hit nearly $70000.

3. Price VS Other Features

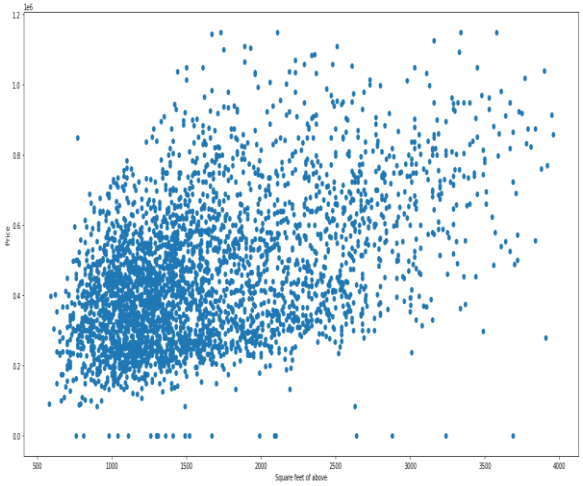
Square feet of living



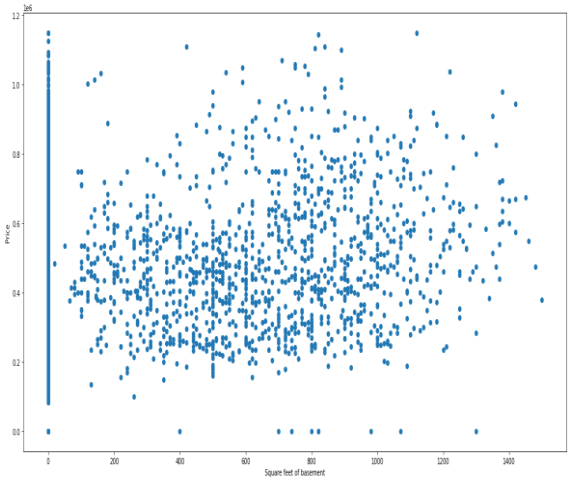
square feet of lot



Square feet of above

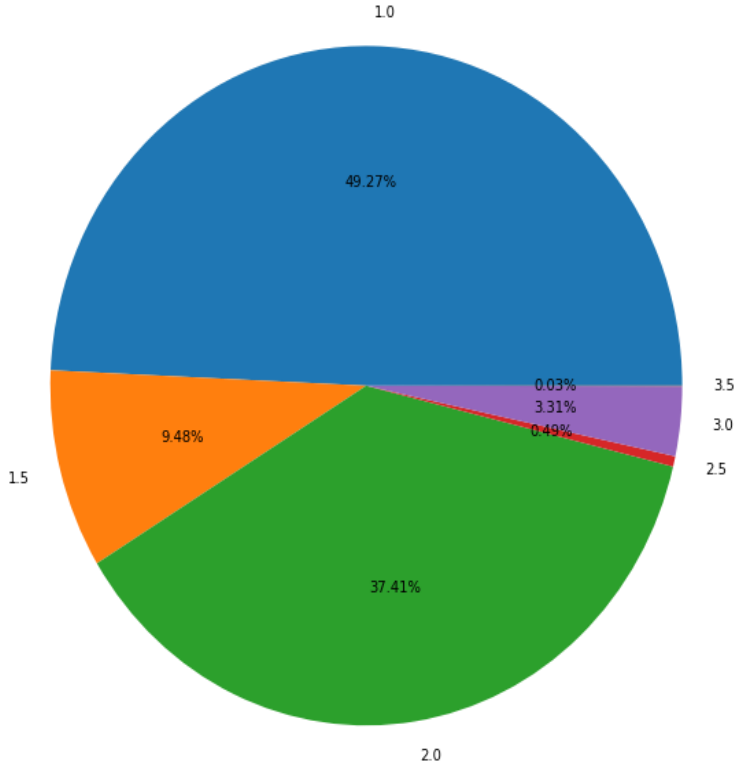


square feet of basement



4. **Analysis on no. of floors in houses**

Analysis- To understand the type of houses that the people are interested in in order to improve the market shares and value.

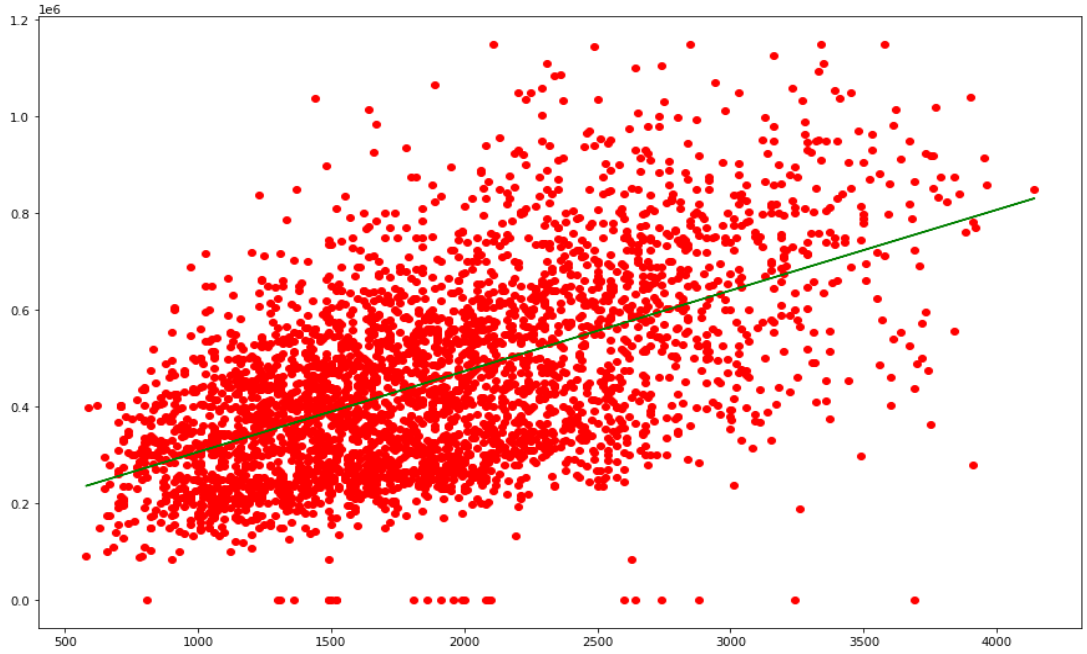


Inference- From the analysis it is understood that maximum population in USA are more interested to have 1 floor house and less than 5% people are with more than 2 floor houses.

Regression Analysis

1. **Square feet of living vs Price**

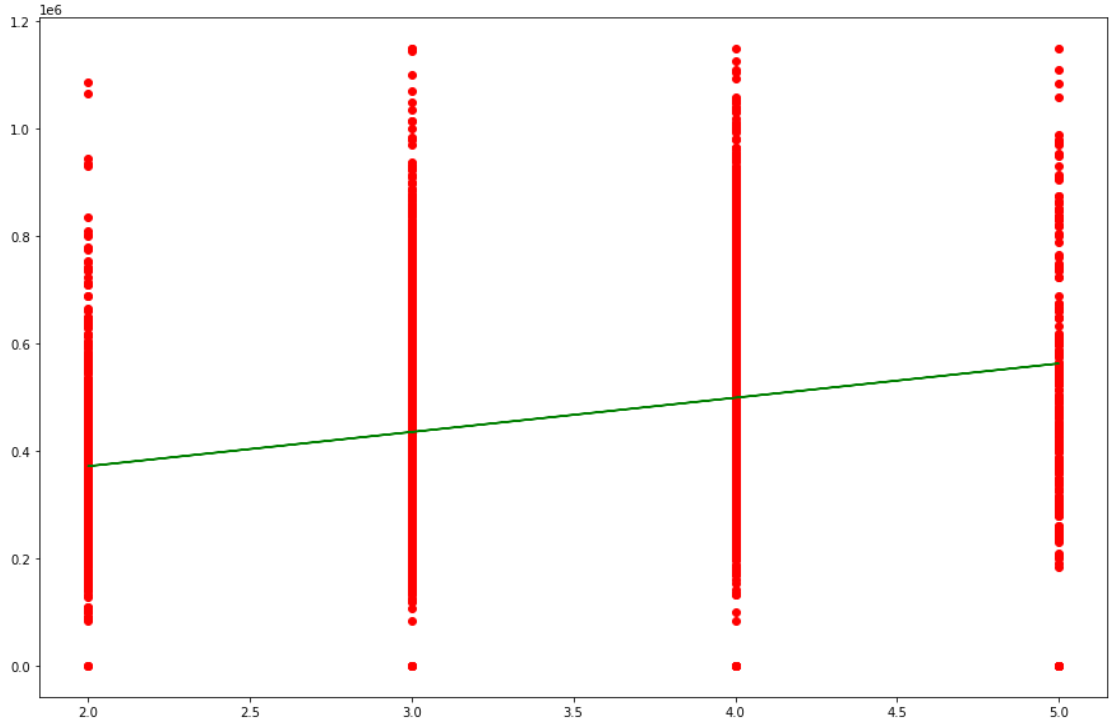
Analysis- To understand how the price value depends on the square feet living of the house.



Inference- From the analysis, it is observed that the square feet of the house is correlated to the price of the house. With the r2 value as 0.76 it can be said that the house price increases when the area of the place increases.

2. No. of bedrooms VS Price

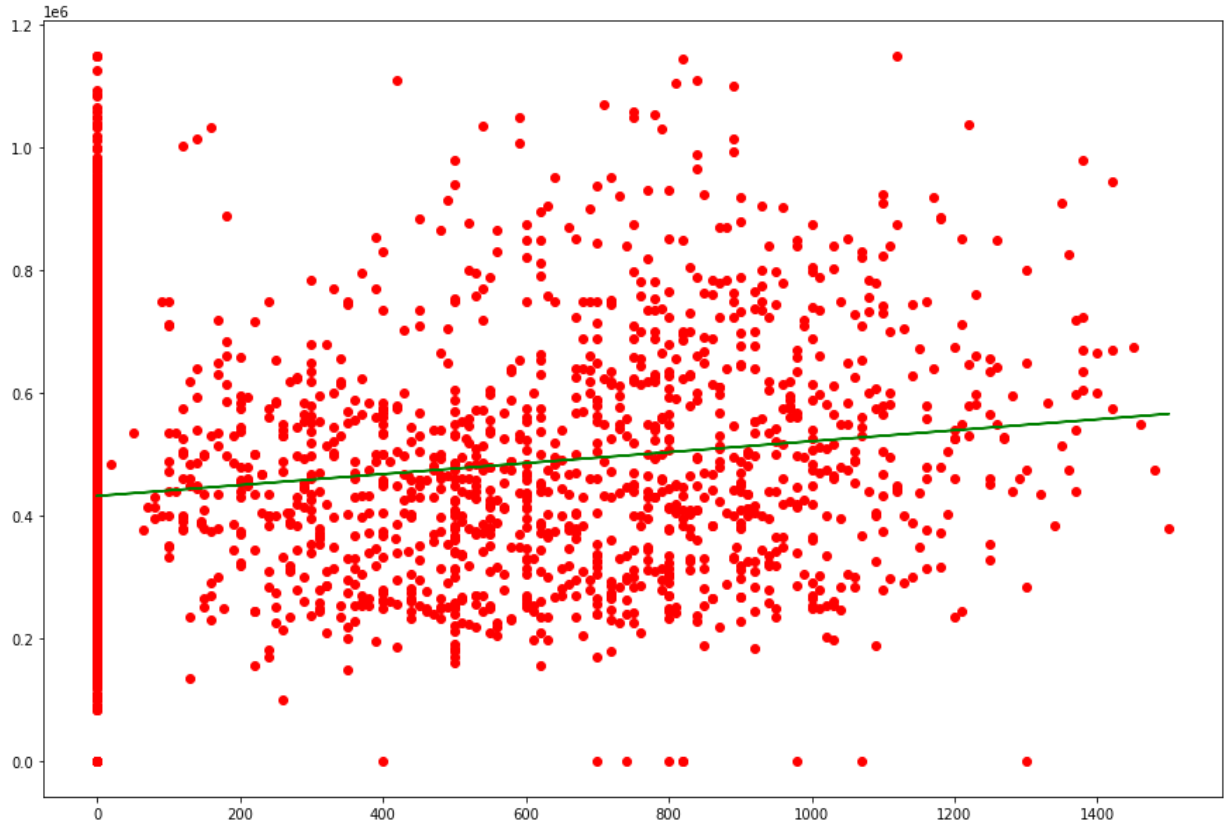
Analysis- To analyse the affect of bedrooms on the house price.



Inference- From the graph it is clearly evident that the bedrooms do not correlate with the price of the house independently. So, the house price is not much dependent on the no. of bedrooms it has.

3. Square feet of basement VS Price

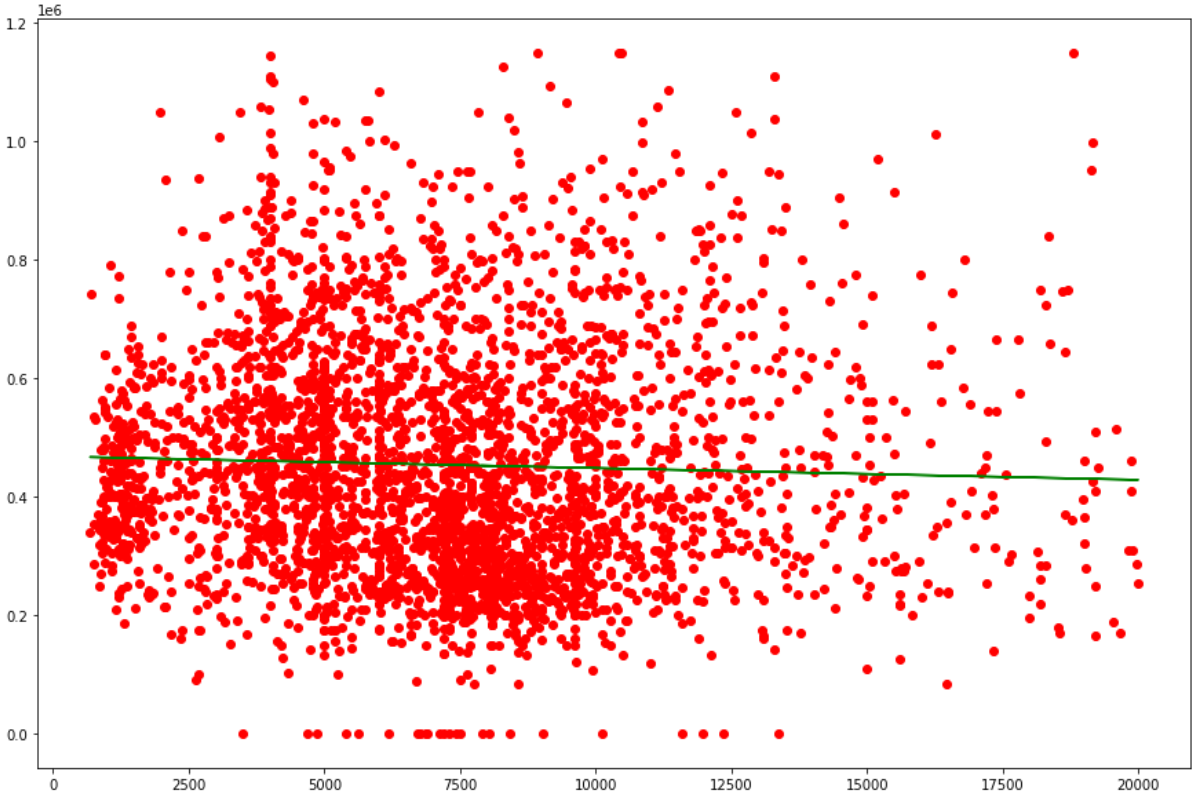
Analysis- To analyse the effect of basement on the cost of price in USA



Inference- The basement area of a house may or may not have effect on the price and it can not be clearly predictable as the r2 score (0.52) lies in the midway of the edges.

4. Square feet of lot VS Price

Analysis- To analyse the house prices in USA with respect to the square feet of house lot.



Inference- The area of lot is also correlated with the price but in an opposite way. It can be observed that increase in lot areas result in slight decrease in house prices which corresponds to the negative correlation.

Conclusion:

- From the analysis, it is observed that Clyde Hill in USA is the place that give retailers profit in house marketing with highest prices.
- The house prices in USA are also unpredictable as there is no pattern followed in the prices in the past years.
- The area of basement is less influencing the house prices than the area of living and area of lot in USA.