University of Moratuwa
MBA in Information Technology
Department of Computer Science & Engineering

Cover Sheet for Assignment

Name with Initials: K.S.H. Sarathchandra

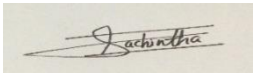
Student Registration No: 189123B

Title of Assignment: Exercise 6.3 – House Price Prediction
Assignment No: 02 **Group** ☐ **Individual** ☑

Subject Code: CS5122
Subject: Descriptive & Predictive Analytics

Lecturer: Dr. Uthayashanker Thayasivam

Student's Statement:
I certify that I have not plagiarized the work of others or participated in unauthorized collusion when preparing this assignment.

Signature:  **Date: 16/03/2019**

Office use only:

On/ before deadline ☐ Extension Given ☐ Late Submission ☐

Signature:……

Marks Given:

Exercise 6.3 – House Price Prediction

Q.01

List 4 question that you may want to explore from the dataset.

- How many distinct records are there in the dataset?
- What columns contain outliers?
- What is the relationship between the area and the market value of a house?
- What is the average square feet of a house?
- What is the average market value of a house?
- What is the average age of a house?

Q.02

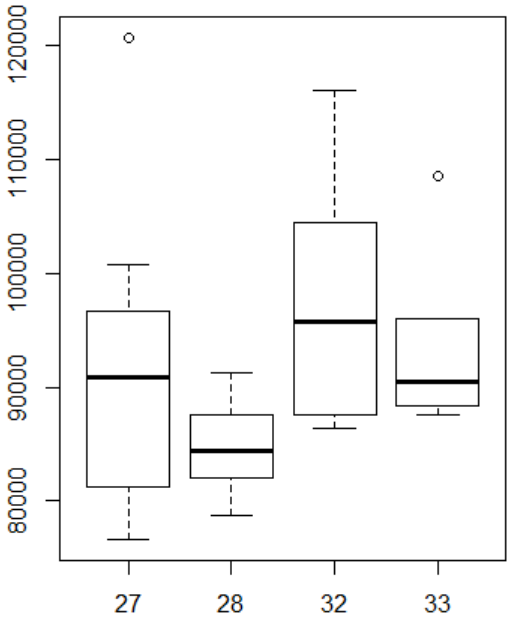
By analyzing statistical properties of data (e.g. mean, std, min, correlation, etc.) and Visualization what you can claim about the data set?

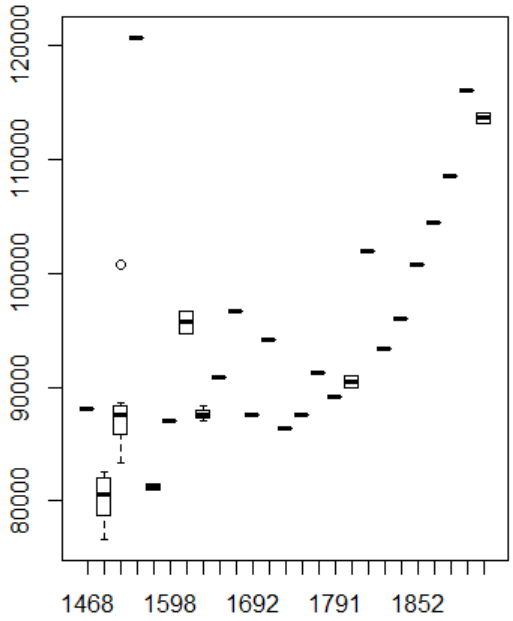
Mean Median and Mode or Original Data Set

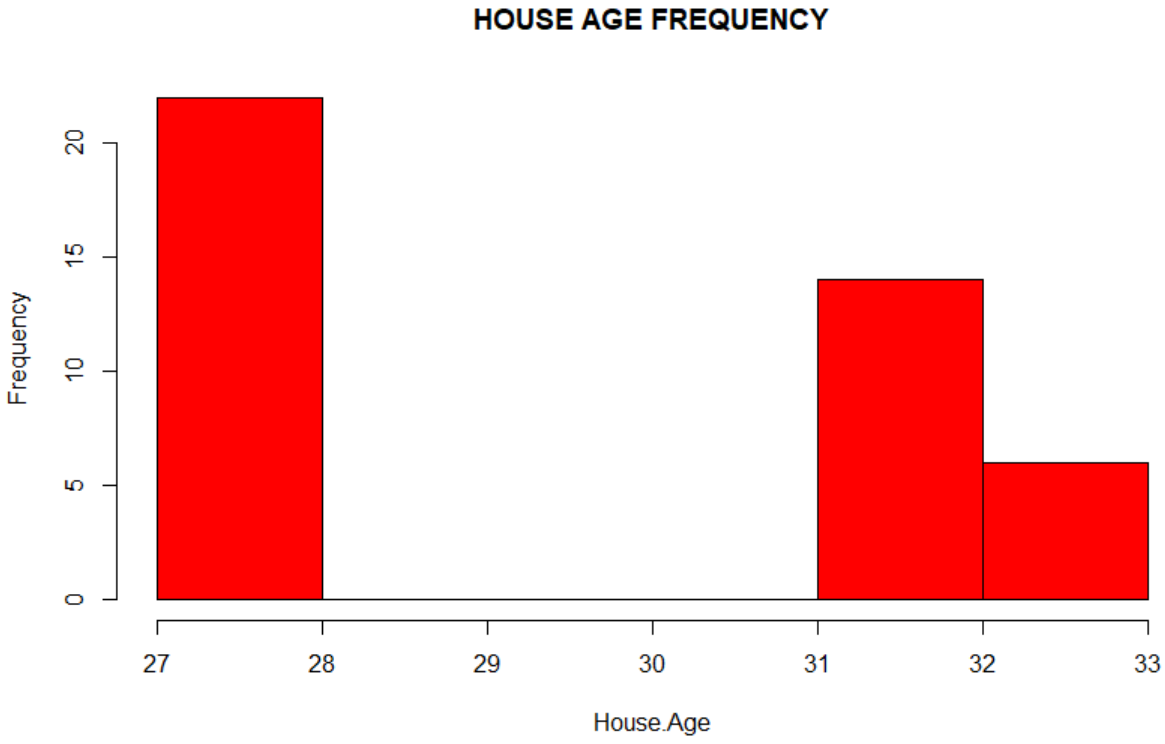
```
   House.Age          Square.Feet       Market.Value
 Min.   :27.00   Min.   :1468   Min.   : 76600
 1st Qu.:28.00   1st Qu.:1520   1st Qu.: 86575
 Median :28.00   Median :1666   Median : 88500
 Mean   :29.83   Mean   :1695   Mean   : 92069
 3rd Qu.:32.00   3rd Qu.:1807   3rd Qu.: 96525
 Max.   :33.00   Max.   :2372   Max.   :120700
```

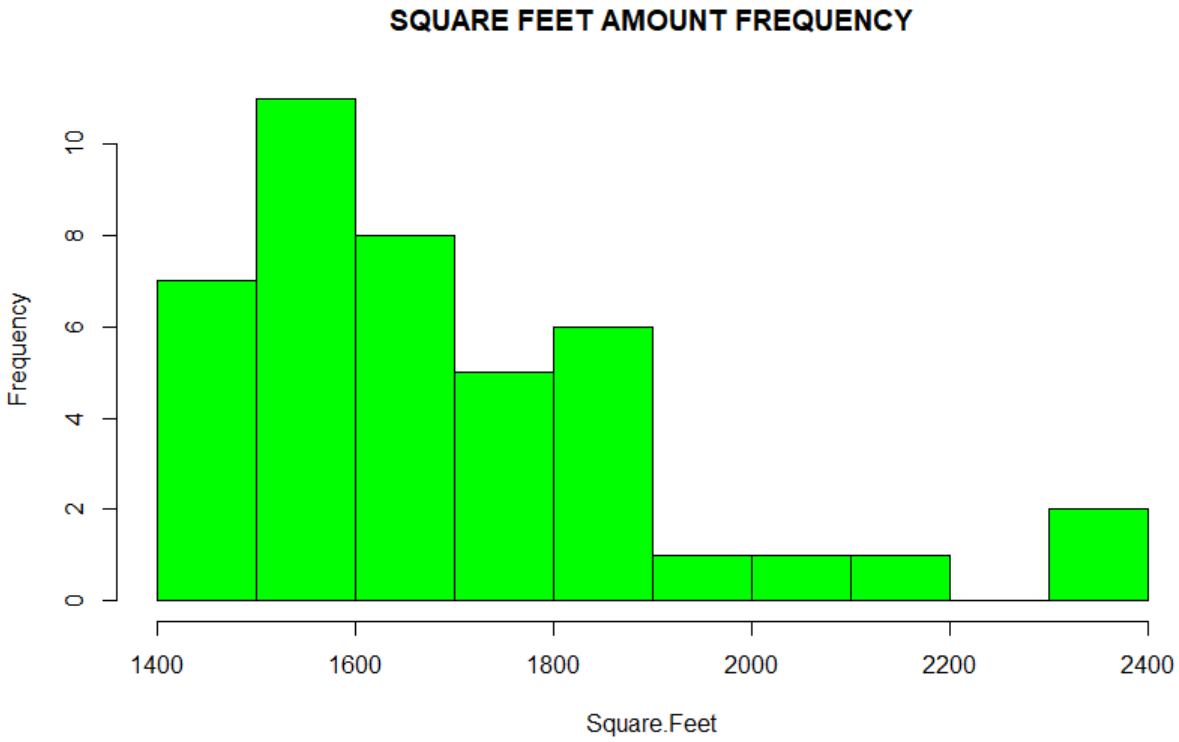
According to the above summary of the data set,

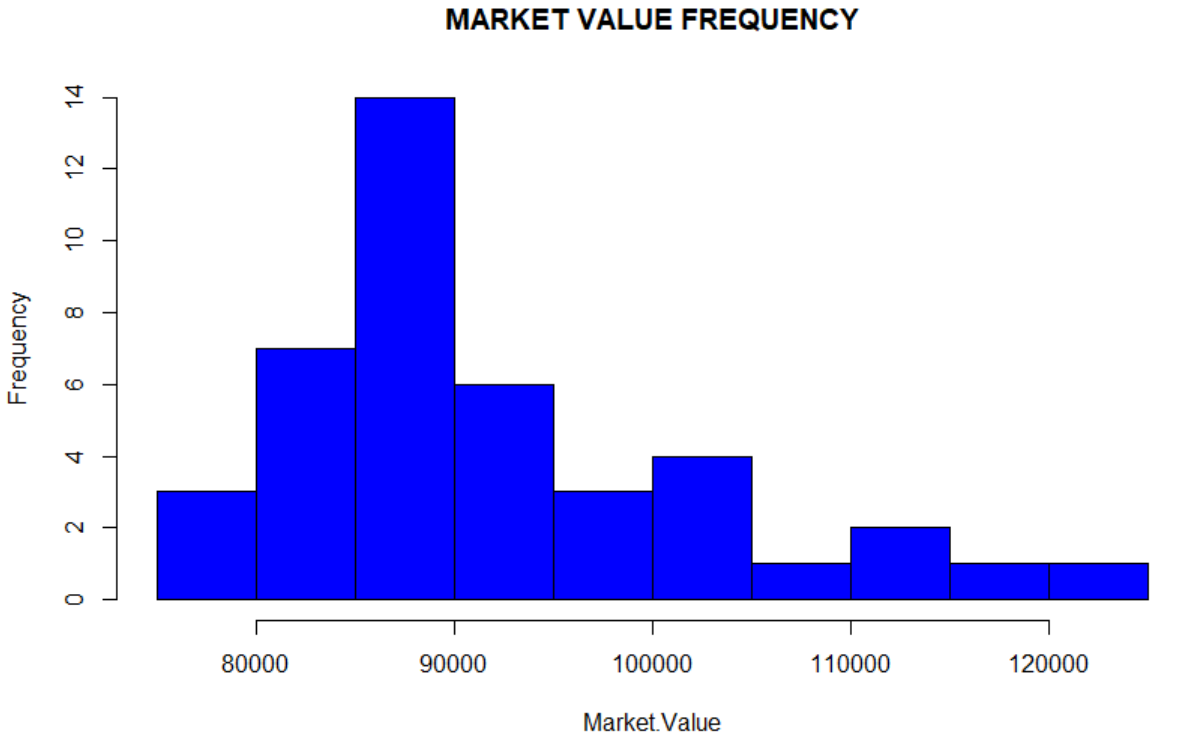
- Minimum age of the house is 27 & maximum age of the house is 33. The average of the age of a house would be 29.83.
- Square feet of a house range from 1468 to 2372. And the average square feet of a house would be 1695.
- Market value of a house range from $76600 to $120700.And the average market value of a house would be $92069.


120000
110000
100000
90000
80000
27
28
32
33

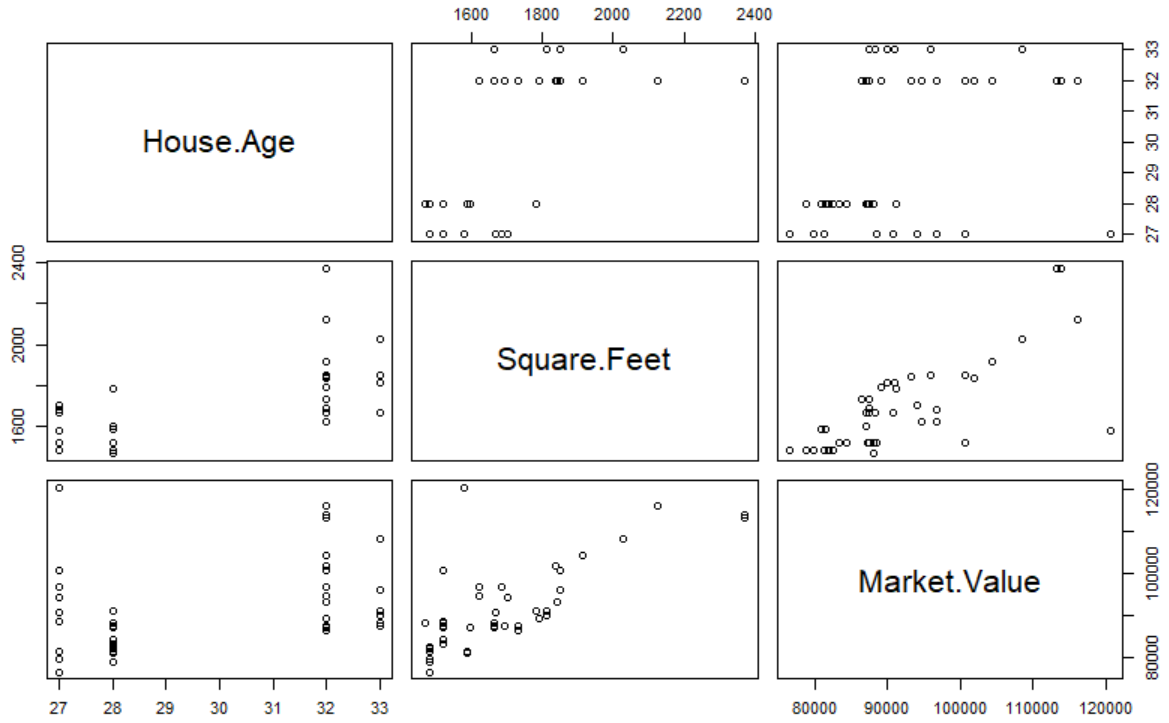

120000
110000
100000
90000
80000
1468
1598
1692
1791
1852


HOUSE AGE FREQUENCY
Frequency
0
5
10
15
20
27
28
29
30
31
32
33
House.Age


SQUARE FEET AMOUNT FREQUENCY
Frequency
0
2
4
6
8
10
1400
1600
1800
2000
2200
2400
Square.Feet


MARKET VALUE FREQUENCY
Frequency
0
2
4
6
8
10
12
14
80000
90000
100000
110000
120000
Market.Value

Below is the plotting of the data set.



- According to the above plot, can see strong correlation between house area and the market value of the house.

	House.Age	Square.Feet	Market.Value
House.Age	1.0000000	0.6456685	0.3614153
Square.Feet	0.6456685	1.0000000	0.7312552
Market.Value	0.3614153	0.7312552	1.0000000

- According to the covariance of the given data set, there is higher positive correlation between square feet of a house and the market value. And also, can be seen a positive correlation between the age of a house & the market value of the house.

Q.03

What regression analysis technique, that is suitable to predict the market value, given the age of a house and square feet? Justify.

As the regression analysis technique to predict the market value the Linear Regression technique can be used. Both Simple and Multiple Linear Regression is suitable for predicting the market value of the given houses.

-House market values can be predicted based on either area or the age of the house. In such cases simple linear regression can be used.

-But in the case of predicting the house market value based on both area and the age of the house, then the multiple linear regressions can be used. Basically, due to the *availability of two attributes and also cause of having a high correlation between two independent variables*, this **multiple linear regression model** is used to predict the market value of a house based on age & square feet.

Q.04

Predict the market value of the following 5 houses.

Predicted Value

```
     fit       lwr       upr
1  93380.45  88492.92  98267.99
2  85593.47  82520.73  88666.21
3  97041.63  93878.98 100204.28
4 112580.90 105506.93 119654.86
5 119937.95 110961.62 128914.27
```