University of Moratuwa

MBA in Information Technology

Department of Computer Science & Engineering

**Name with Initials: P.M.Upeksha**

**Student Registration No: 189128U**

**Title of Assignment: 6.3 Exercise (Home Market Value Dataset analyze)**

**Assignment No: Group Individual**

**Subject Code: CS5122**

**Subject: Descriptive and Predictive Analytics**

**Lecturer: Dr. Uthayasanker 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:** Pubudu **Date:** 17th April 2018

**Office use only:**

On/ before deadline Extension Given Late Submission

**Signature:**………………………….

**Marks Given:**

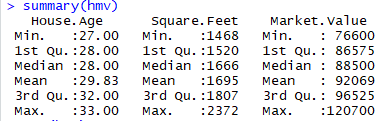
1. List 4 questions that you may want to find out from the dataset. For example, “what is the relationship between age of a house and its market value?”

4 Questions

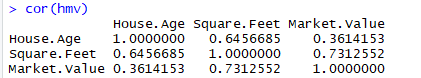
* 1. Whether there is a relationship between the square feet of the house and its market value?
  2. Whether there is a relationship between the square feet of the house and its age?
  3. Does age of house impact on the market value?
  4. Does square feet of house impact on its market value?

1. By analyzing statistics properties of data (e.g., mean, std, min, max, correlation, etc.) and visualization what can you claim about the dataset? Justify each of your claims

*Summary of the data*



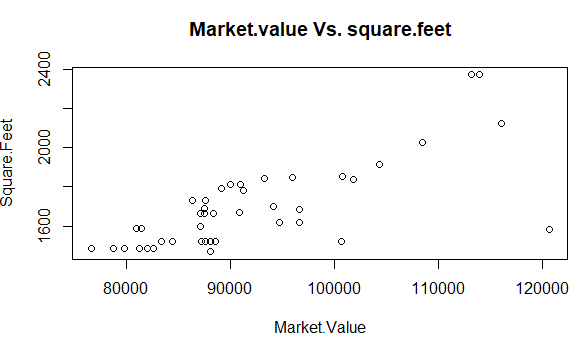
*Correlations of the data*



According to the correlation calculated above, there is a high positive correlation between the square feet of the house and market value such as **0.7312552**.then it says the market value of the house will increase when the square feet is increase positively.

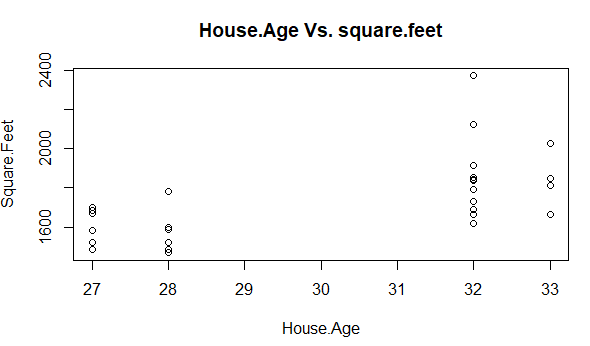
It can be shown in a graph as below:



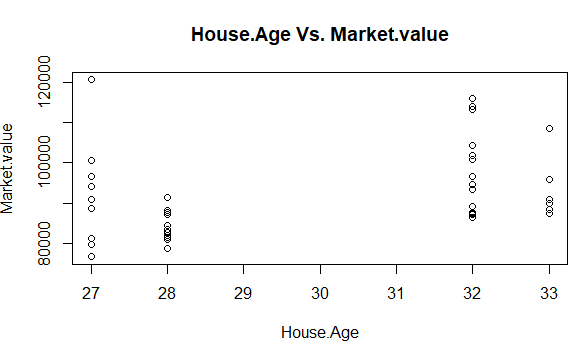


In the graph shows some outliers of the data set with the red color circles. Which means there are some houses which are having less square feet with a higher market value.



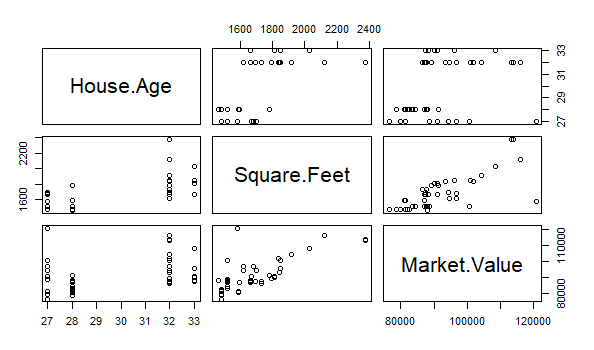


The correlation calculated between age and square feet there is 0.6456685.By looking at the above, we can understand most of the older houses has a larger amount of square feet.



The above graph shows, there is no very significant difference between the market value and age. But the highest market values are for the houses built recently. The correlation we calculate between the age and the market value of a house is around 0.3614153 which implicates a low correlation value.

1. What Regression Analysis technique is suitable to predict the market value, given the age of a house and square feet? Justify.

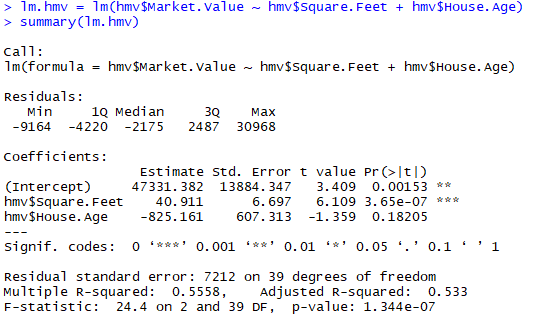


By looking at the plot for all 3 variables and the correlation values, it shows all three variables has positive correlation between each other. Therefore, this can be a multi-attribute data set we need to use the linear regression with the type formula Y = mx + c, and recommend a linear regression technique.

1. Predict the market value of the following 5 houses.

|  |  |
| --- | --- |
| Age | Square Feet |
| 26 | 1650 |
| 28 | 1500 |
| 29 | 1800 |
| 30 | 2200 |
| 31 | 2400 |

*Linear Model*



***Predicted Market Values for the given parameters:***

