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.

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
   I. Whether there is a relationship between the square feet of the house and its market value?
   II. Whether there is a relationship between the square feet of the house and its age?
   III. Does age of house impact on the market value?
   IV. Does square feet of house impact on its market value?
2. 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

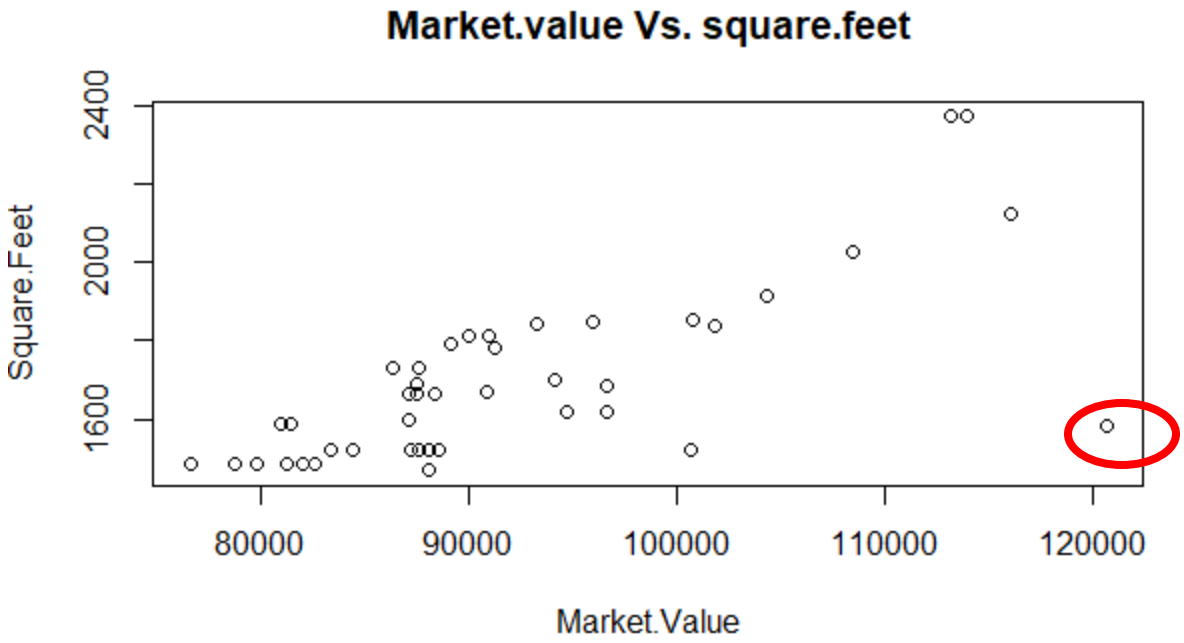
```
> summary(hmv)
   House.Age      Square.Feet    Market.Value
 Min.   :27.00   Min.   :1468   Min.   : 76600
 1st Qu.:28.00   1st Qu.:1520   1st Qu.: 86575
 Median :28.00   Median :1666   Median : 88500
 Mean   :29.83   Mean   :1695   Mean   : 92069
 3rd Qu.:32.00   3rd Qu.:1807   3rd Qu.: 96525
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

   Correlations of the data

```
> cor(hmv)
              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
```

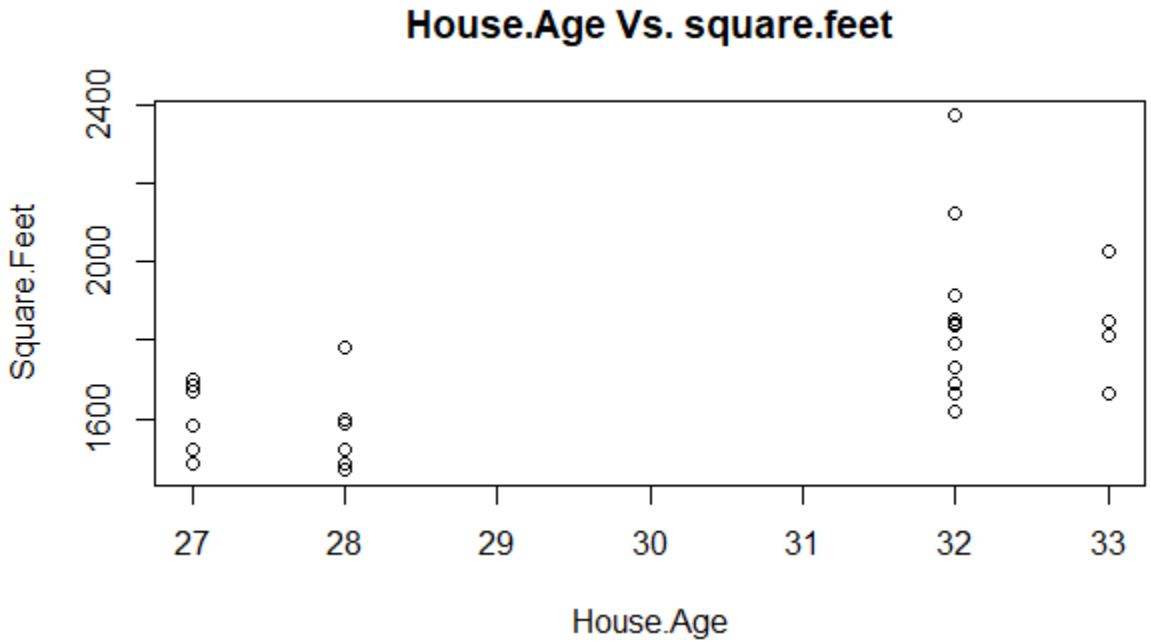
   According to the correlation calculated above, there is a high positive correlation between the square feet of the house and market value such as **0.7312552**.then it says the market value of the house will increase when the square feet is increase positively.
   It can be shown in a graph as below:

```
> plot(hmv$Market.Value, hmv$Square.Feet, xlab = 'Market.Value', ylab = 'Square.Feet',
main = 'Market.value Vs. square.feet')
```

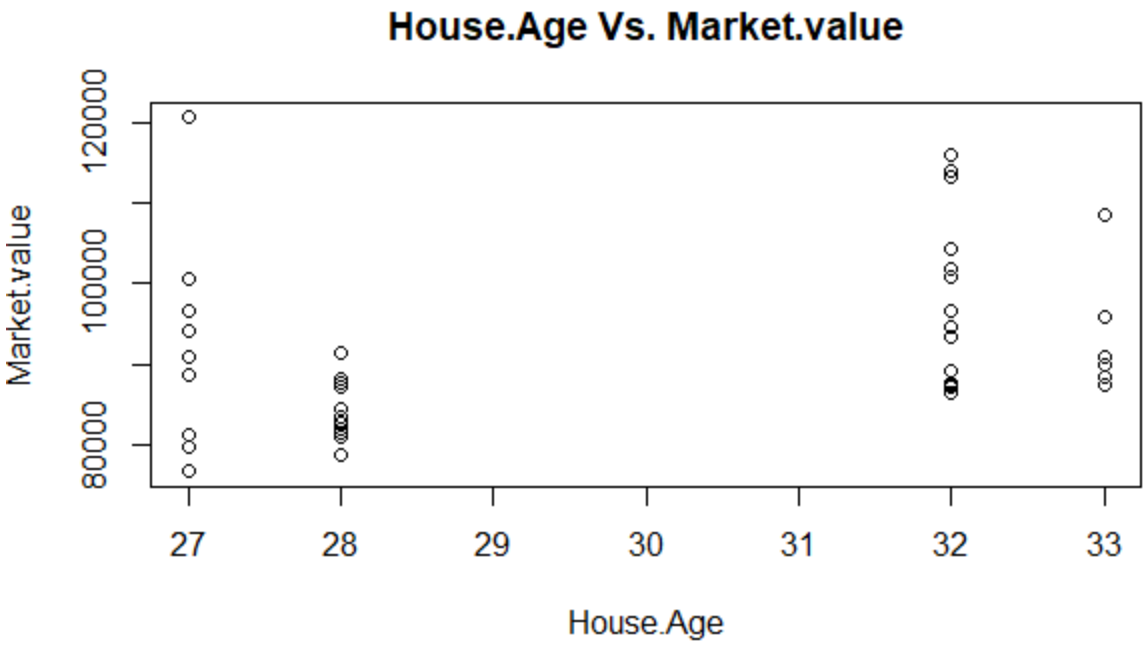


In the graph shows some outliers of the data set with the red color circles. Which means there are some houses which are having less square feet with a higher market value.

```
> plot(hmv$House.Age, hmv$Square.Feet, xlab = 'House.Age', ylab = 'Square.Feet', main =
  'House.Age Vs. square.feet')
```

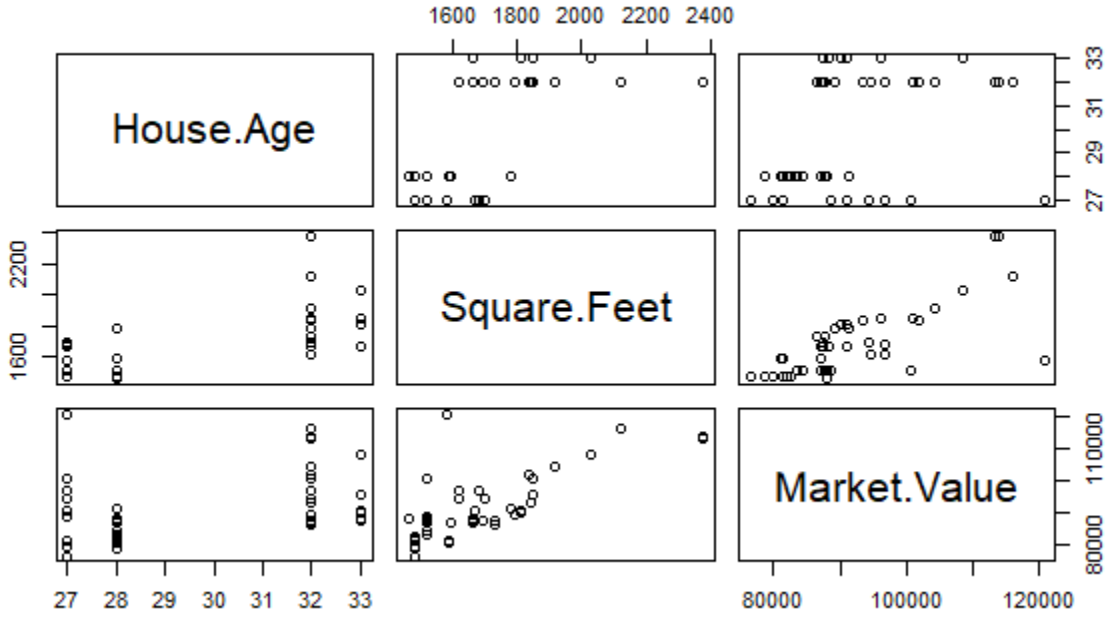


The correlation calculated between age and square feet there is 0.6456685.By looking at the above, we can understand most of the older houses has a larger amount of square feet.



The above graph shows, there is no very significant difference in the market value by its 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.

3. 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.

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

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

Linear Model

```
> lm.hmv = lm(hmv$Market.Value ~ hmv$Square.Feet + hmv$House.Age)
> summary(lm.hmv)

Call:
lm(formula = hmv$Market.Value ~ hmv$Square.Feet + hmv$House.Age)

Residuals:
   Min     1Q Median     3Q    Max
 -9164  -4220  -2175   2487  30968

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     47331.382  13884.347   3.409  0.00153 **
hmv$Square.Feet    40.911      6.697   6.109 3.65e-07 ***
hmv$House.Age    -825.161    607.313  -1.359  0.18205
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7212 on 39 degrees of freedom
Multiple R-squared:  0.5558,    Adjusted R-squared:  0.533
F-statistic:  24.4 on 2 and 39 DF,  p-value: 1.344e-07
```

Predicted Market Values for the given parameters:

```
> newSqure = c(1650,1500,1800,2200,2400)
> newAge = c(26,28,29,30,31)
> Z =hmv$Market.Value
> X= hmv$Square.Feet
> Y = hmv$House.Age
> lm.hmv = lm(Z~X+Y)
> newdata = data.frame(X=newSqure, Y=newAge)
> predictedData = predict(lm.hmv,newdata,level = 0.95,interval = "confidence")
> predictedData
        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
```