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

Cover Sheet for Assignment

Name with Initials: T.M.S Vithanage

Student Registration No: 189129A

Title of Assignment: Analysis Home Market Value dataset

Assignment No: 3 **Group** ☐ **Individual** ■

Subject Code: CS5122
Subject: Descriptive and Predictive Analytics

Lecturer: DR. Uthayasanker Thayasivam

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

Date: 16th March 2019

Office use only:
On/ before deadline ☐ Extension Given ☐ Late Submission ☐

Signature: …….

Marks Given:

Home Market Dataset Analysis

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

- What is the relationship between the Square feet of house & its Market Value?
- What is the co-relation between age of House, square feet and its market value?
- For a given Age of house and square feet, can we predict the house value?
- What are the average square feet of a house?
- What is the average market value of a house?
- How market value of house varies with square feet?
- How market value of house varies with age of House?

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

```
summary(homeMarketValue)
  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 summary of data set, we can derive that the minimum age of the house is 27Y and maximum age is 33Y. The average age of the house is 29Y. When considering Square feet of the house, we can say these data only consider the square feet range 1468 to 2372. Not only that, also we can derive that market value of the houses is in the 76,600$ - 120,700$ range. Average market value of the house is 92,069$.

```
#standard deviation
sd(homeMarketValue$House.Age)
1] 2.428657
sd(homeMarketValue$Square.Feet)
1] 220.2567
sd(homeMarketValue$Market.Value)
1] 10553.08
```

When consider Standard Deviation (SD) of the data set, Age of the houses has less SD. That means most of the Age groups are very close to the average. Square Feet (SF) and Market Value (MV) has high SDs. That means both of these values are spread out from the average.

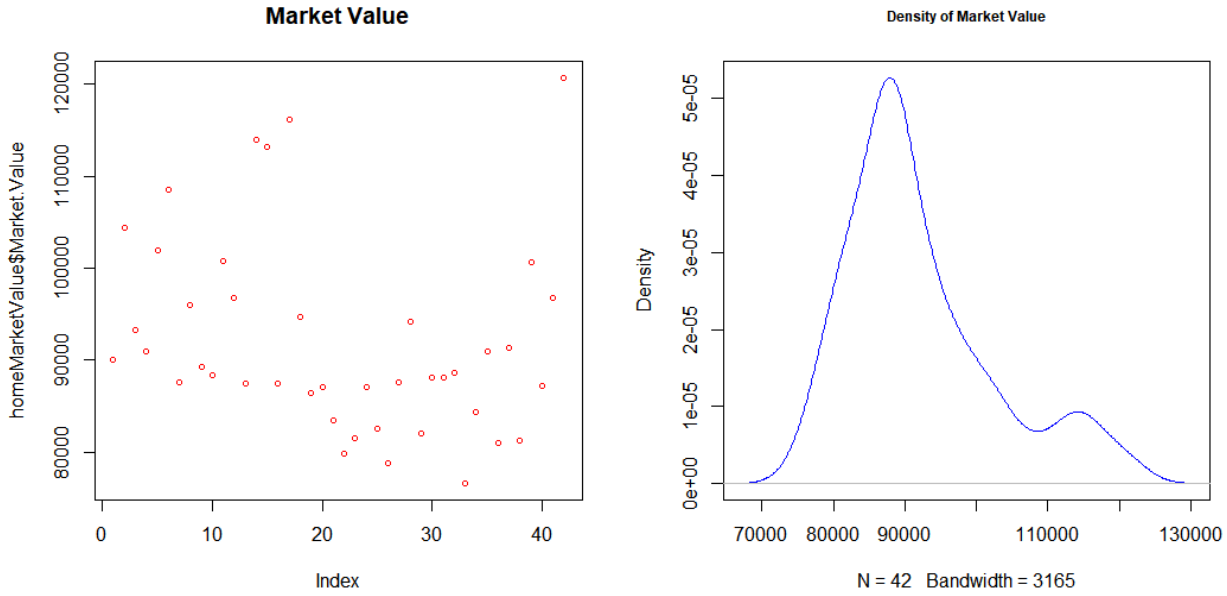
```
#Mode
mode(homeMarketValue$House.Age);
1] "numeric"
mymode <- function(x){
  uniqx <- unique(x)
  uniqx[which.max(tabulate(match(x,uniqx)))]
}
mymode(homeMarketValue$House.Age)
1] 32
mymode(homeMarketValue$Square.Feet)
1] 1520
mymode(homeMarketValue$Market.Value)
1] 87600
```

These are the mode values of the properties of this data set.

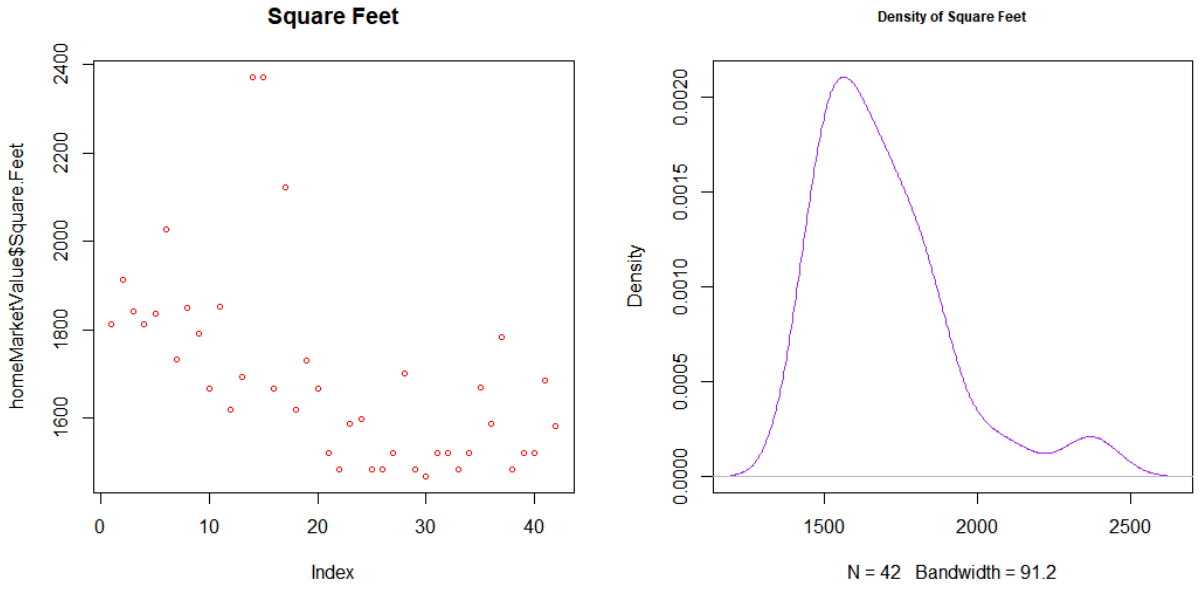
```
ages/ageu
L] 0.08140749
marketValueS/marketValueU
L] 0.1146214
```

Coefficient of Variation (CoV) of the age of house and the market value of the house.

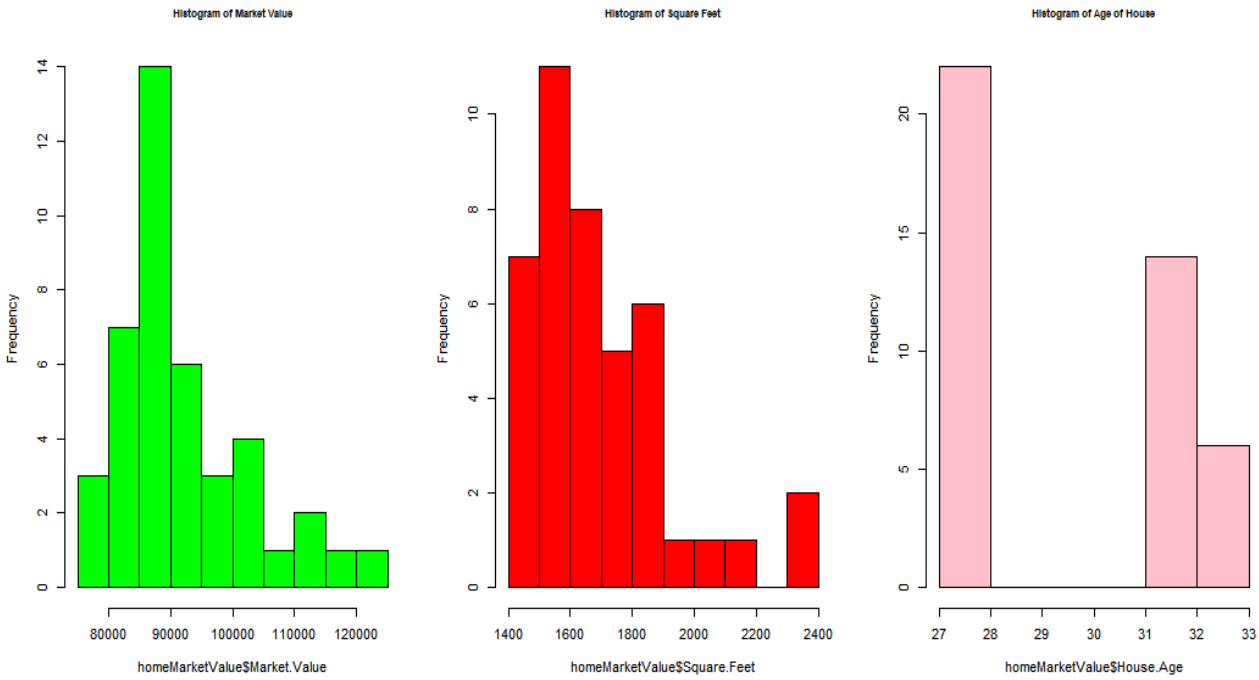
Above insight shows the spread of all three variables.



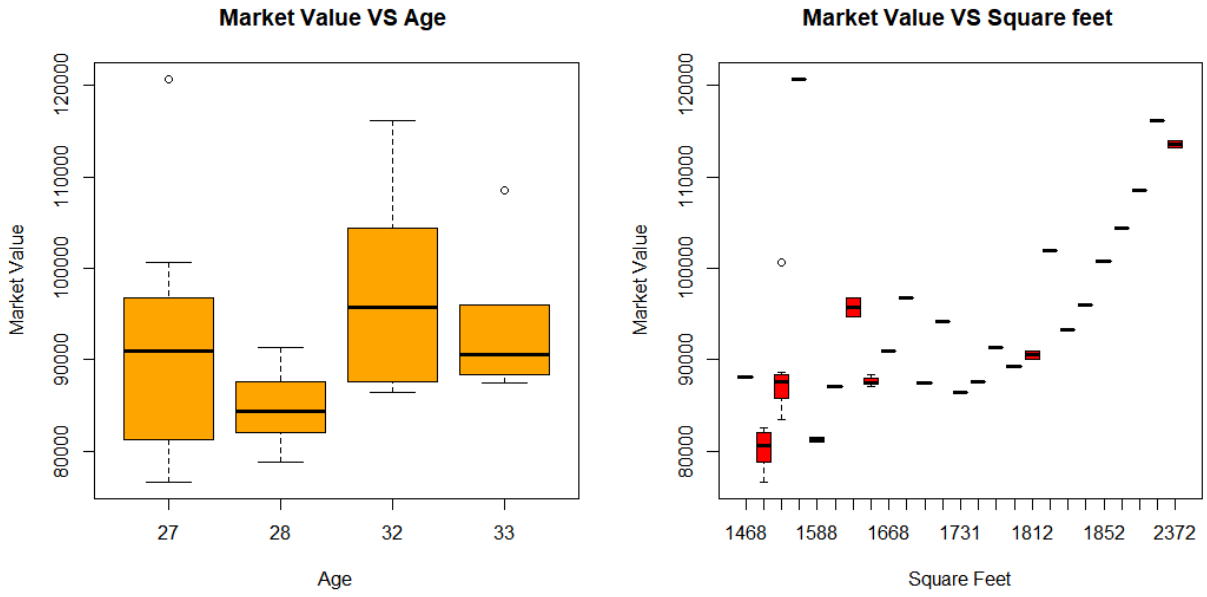
Above diagram shows the density of the Market value. According to this density of market value is higher around 85000.



Above diagram shows the density of the Square Feet. According to this density of Square Feet is higher around 1600.



Above diagram shows the histograms of the all three valuables in this data set. According to this we can claim that most of the houses has market value in range 80,000$ - 10,500$. And also, most of houses are build in the land which has Square Feet in range 1400 – 1900. When we consider age of the houses, many houses in this data set are in between the age limit 27Y- 28Y. There are no houses in between 28Y- 31Y age limit.



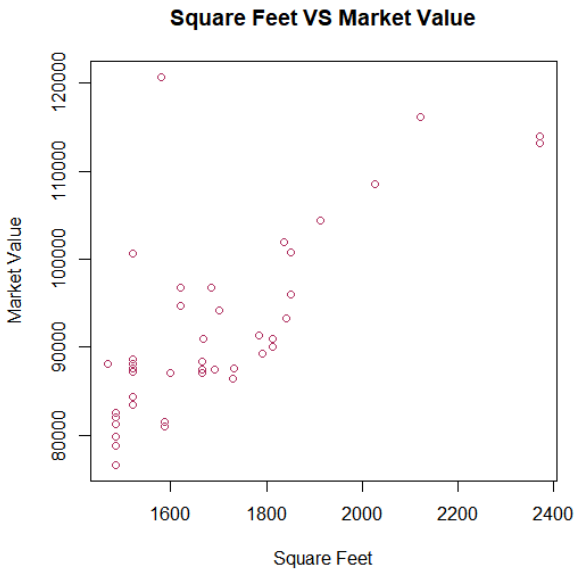
Above diagram shows how market value vary with Age of House and the Square Feet. We can claim that house can reach highest market value in 32Y age limit. And also, lowest market value gain b the 27 age range houses.
Not only that market value is depending on the square feet also. Market Value is increase with the Square Feet of the house. So, Market Value and Square Feet has well enough relationship.

Below image shows how mean of Market value and standard deviation of Market Value are varies according to their Age.

House.Age	mean	sd
27	92166.67	13413.240
28	84853.85	3640.196
32	98064.29	10606.573
33	93566.67	7895.991

This is how correlate Market Value with Age and square feet of the house.

```
> cor(homeMarketValue$House.Age, homeMarketValue$Market.Value)
[1] 0.3614153
> cor(homeMarketValue$Square.Feet, homeMarketValue$Market.Value)
[1] 0.7312552
> cor(homeMarketValue[,1:3])
              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
```



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

```
> cor(homeMarketValue)
             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 this diagram there is no negative correlations in the dataset. And also, there is a higher correlation between Market Value & Square Feet and moderate correlation between Market Value & House Age. Not only that there is no higher correlation between Square Feet and House Age. When consider the relationships between these three components, we can't see a relationship between House Age and Square Feet. But Market Value has relationship with other two variables. If square feet increase, then obviously market value is increased. So, we can define Market Value as dependent variable and Square Feet & House Age as independent variables. So, there are multiple independent variables and one dependent variable in this data set. So **multiple linear regression** is the most suitable one for predict market value with given square feet and the age of a house.

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

Age	Square Feet
26	1,650
28	1,500
29	1,800
30	2,200
31	2,400

```
       fit       lwr       upr
 93380.45  88492.92  98267.99
 85593.47  82520.73  88666.21
 97041.63  93878.98 100204.28
112580.90 105506.93 119654.86
119937.95 110961.62 128914.27
```

```
> summary(lm.homeMarketValue)

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

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 **
homeMarketValue$House.Age   -825.161    607.313  -1.359  0.18205
homeMarketValue$Square.Feet   40.911      6.697   6.109 3.65e-07 ***
---
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
```