

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

Name with Initials: A.A.P.C.Harischandra

Student Registration No: 189107F

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:

6.3 Exercise

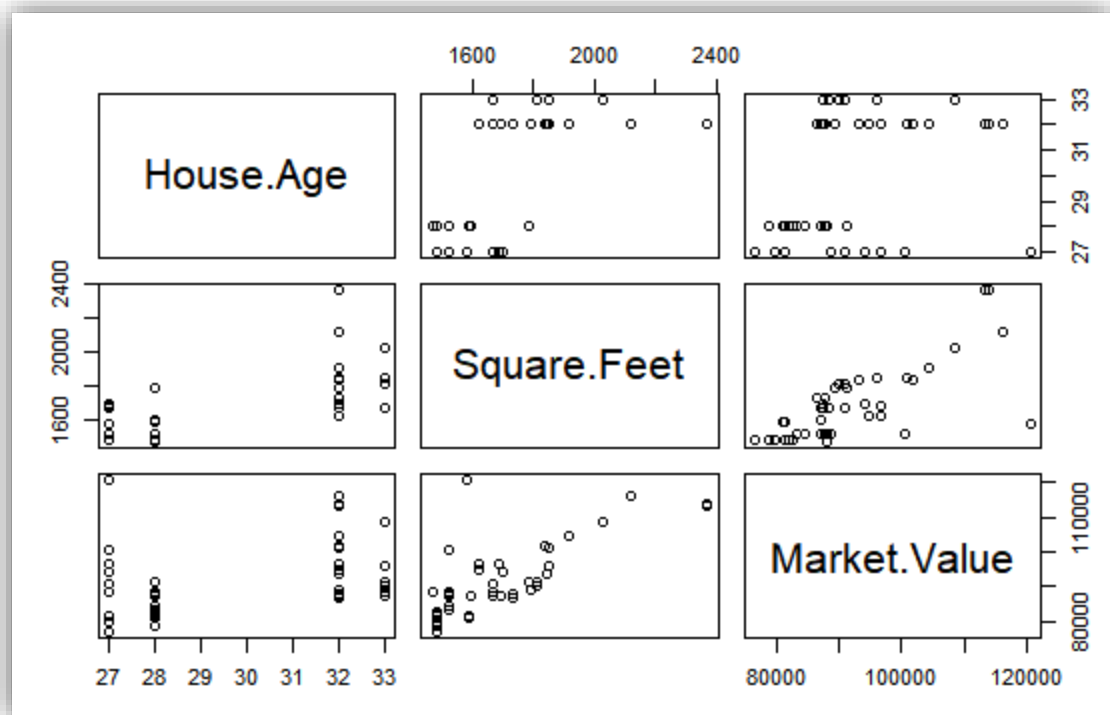
01. List 4 questions that you may want to explore from the dataset.

- What is the relationship between square feet and market value ?
- What is the relationship between age of a house & it's market value?
- Is there any correlation between age and square feet?
- Is there any outliers or independent variables?

02. By analyzing statistical 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.

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

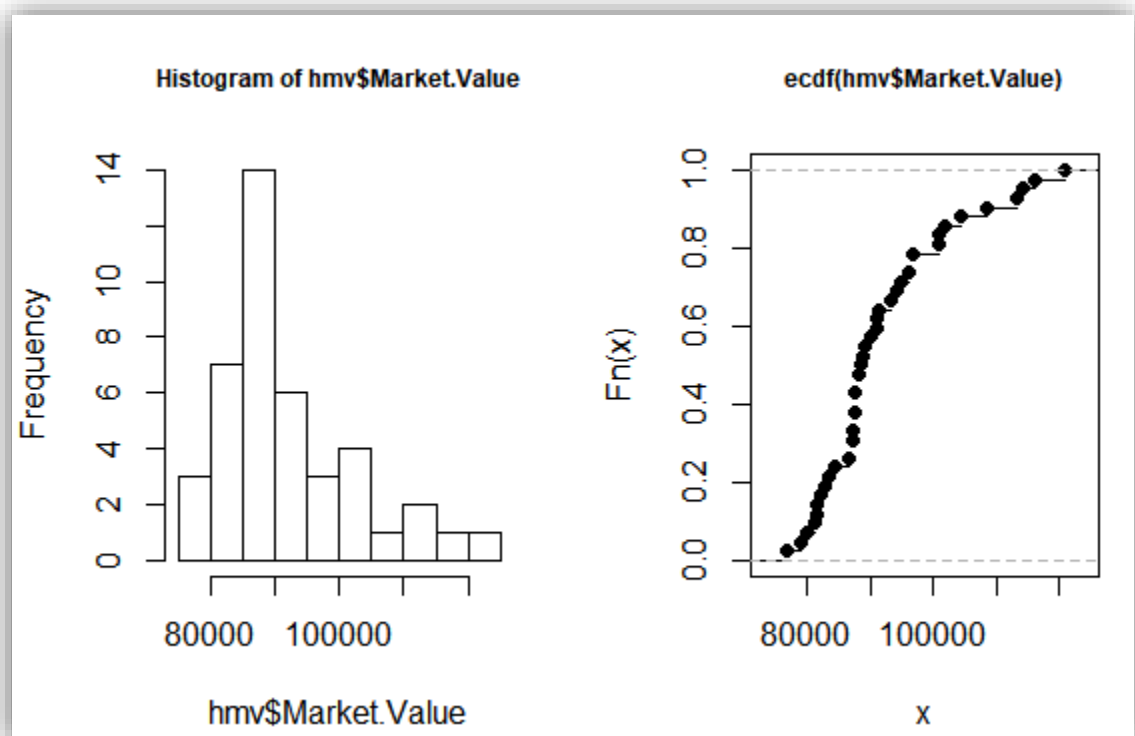
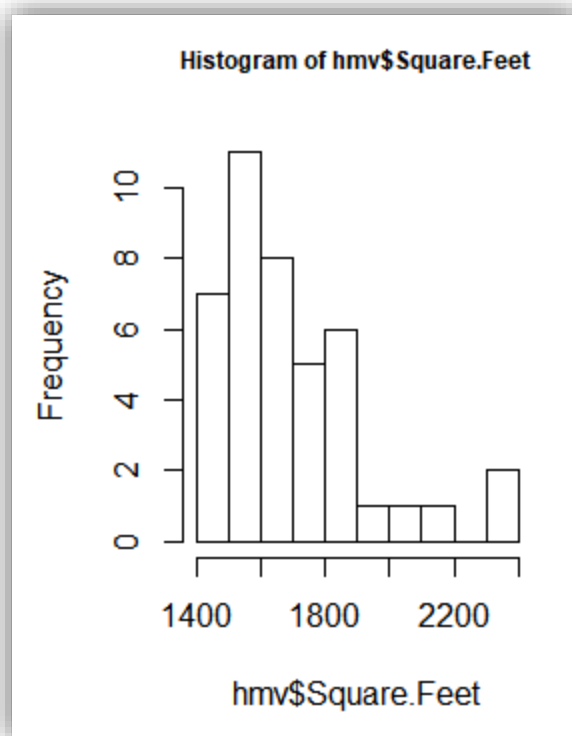
Based on summary of the data set, age of the house vary from 27 to 33. That means minimum age is 27 & maximum age of 33. The average of the age of a house would be 29.83. Square feet of a house range from 1468 to 2372. And the average square feet of a house is 1695. By analyzing the Market value of a houses which vary from \$76600 to \$120700. And it denoted that average market value of a house would be \$92069. Further to based on plot function generates a set of graphs where House age / square feet & Market value is plotted each other.



We can see the given data are continuous data therefore we can use the histogram, where the bins represent ranges of data



By analyzing this graph we can clearly say that highest frequency were happen for 27 house age. And there are no any value between 27 to 31.



We can see the how the frequency were happen for market value and commutative distribution function as well.

	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

Based on core relation function, there are higher positive correlation between square feet of a house and the market value.

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

	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

By analyzing the correlation matrix, we can see a higher positive correlation between square feet of a house and the market value, if so square feet increase the market value increases. Further to analyze the data, age of a house & the square feet of a house can be considered as possible predictors and can build a multiple linear regression model to predict the market value of a house based on age & square feet of house.

04. Predict the market value

	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	