**Ex.No:4**

AIM:

Explore various variable and row filters in R for cleaning data. Apply various plot features in R on sample data sets and visualize.

Dataset:

This dataset provides comprehensive information for house price prediction, with 13 column names: price, area, bathroom, bedroom, basement, stories, etc.,

PROGRAM:

data <- read.csv('Housing.csv')

data <- read.csv('Housing.csv')

summary(data)

price area bedrooms bathrooms Min. : 1750000 Min. : 1650 Min. :1.000 Min. :1.000 1st Qu.: 3430000 1st Qu.: 3600 1st Qu.:2.000 1st Qu.:1.000 Median : 4340000 Median : 4600 Median :3.000 Median :1.000 Mean : 4766729 Mean : 5151 Mean :2.965 Mean :1.286 3rd Qu.: 5740000 3rd Qu.: 6360 3rd Qu.:3.000 3rd Qu.:2.000 Max. :13300000 Max. :16200 Max. :6.000 Max. :4.000 stories mainroad guestroom basement

Min. :1.000 Length:545 Length:545 Length:545

1st Qu.:1.000 Class :character Class :character Class :character

Median :2.000 Mode :character Mode :character Mode :character

Mean :1.806

3rd Qu.:2.000

Max. :4.000

hotwaterheating airconditioning parking prefarea Length:545 Length:545 Min. :0.0000 Length:545

Class :character Class :character 1st Qu.:0.0000 Class :character

Mode :character Mode :character Median :0.0000 Mode :character

Mean :0.6936 3rd Qu.:1.0000 Max. :3.0000

furnishingstatus

Length:545

Class :character

Mode :character

mean = mean(data$bedroom)

print(mean)

[1] 2.965138

median = median(data$bedrooms)

print(median)

[1] 3

install.packages("dplyr")

Installing package into ‘/usr/local/lib/R/site-library’ (as ‘lib’ is unspecified)

iqr = IQR(data$bedrooms)

print(iqr)

[1] 1

median(data$basement, na.rm = TRUE)

[1] "no"

median(data$bedrooms, na.rm = TRUE)

[1] 3

mean(data$bedroom, na.rm = TRUE)

[1] 2.965138

print(mean(data$bedrooms))

[1] 2.965138

print(var(data$bedrooms))

[1] 0.5447383

print(sd(data$bedrooms))

[1] 0.7380639

mean = mean(data$bedroom)

print(mean)

[1] 2.965138

mode = function(){

return(sort(-table(data$bedrooms))[1]) }

mode()

3

-300

mean(data$parking, na.rm = TRUE)

[1] 0.693578

quartiles = quantile(data$bedrooms)

print(quartiles)

0% 25% 50% 75% 100%

1 2 3 3 6

max=max(data$bedrooms)

min=min(data$bedrooms)

range= max-min

print(range)

[1] 5

r=range(data$bedrooms)

print(r)

[1] 1 6

summary=summary(data)

print(summary)

price area bedrooms bathrooms Min. : 1750000 Min. : 1650 Min. :1.000 Min. :1.000 1st Qu.: 3430000 1st Qu.: 3600 1st Qu.:2.000 1st Qu.:1.000 Median : 4340000 Median : 4600 Median :3.000 Median :1.000 Mean : 4766729 Mean : 5151 Mean :2.965 Mean :1.286 3rd Qu.: 5740000 3rd Qu.: 6360 3rd Qu.:3.000 3rd Qu.:2.000 Max. :13300000 Max. :16200 Max. :6.000 Max. :4.000 stories mainroad guestroom basement

Min. :1.000 Length:545 Length:545 Length:545

1st Qu.:1.000 Class :character Class :character Class :character

Median :2.000 Mode :character Mode :character Mode :character

Mean :1.806

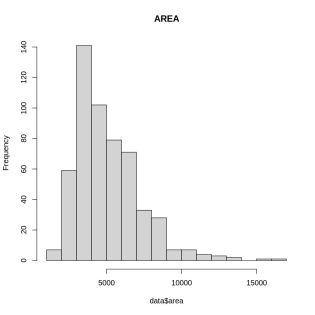
3rd Qu.:2.000

Max. :4.000

hotwaterheating airconditioning parking prefarea Length:545 Length:545 Min. :0.0000 Length:545

Class :character Class :character 1st Qu.:0.0000 Class :character

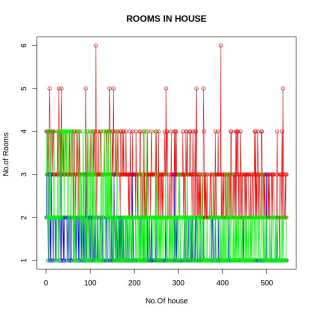
|  |
| --- |
| Mode :character Mode :character Median :0.0000  Mode :character  Mean :0.6936  3rd Qu.:1.0000  Max. :3.0000  furnishingstatus  Length:545  Class :character  Mode :character    T <- aov(data$area~factor(data$price) \*  factor(data$bathrooms))  summary(T)  Df Sum Sq Mean Sq F value  factor(data$price) 218 1.439e+09 6602287 1.942  factor(data$bathrooms) 3 1.481e+07 4938174 1.453  factor(data$price):factor(data$bathrooms) 47 1.696e+08 3608887 1.062  Residuals 276 9.382e+08 3399434  Pr(>F)  factor(data$price) 9.96e-08 \*\*\*  factor(data$bathrooms) 0.228  factor(data$price):factor(data$bathrooms) 0.374  Residuals  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 hist(data$area,main='AREA',) |



OBSERVATION:

Thus the above output shows the visualization of data$area using histogram in R. The data$area is higher in the range of 140 . The data area is lower in two places on 15000 and it values is about 3.

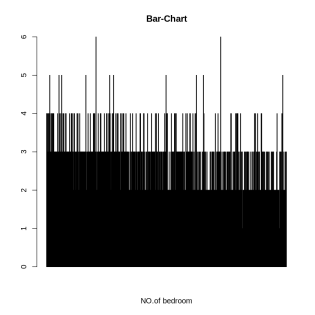
|  |
| --- |
| plot(data$bedrooms, type = "o", col = "red",  xlab = "No.Of house", ylab = "No.of Rooms ",  main = "ROOMS IN HOUSE")  lines(data$bathrooms, type = "o", col = "blue")  lines(data$stories, type = "o", col = "green") |



OBSERVATION:

The above output shows the visualization between Number of house and Number of rooms . The red line shows the number of bedroom in the houses. The green line shows the number stories in the houses and the red line shows the number of bathroom in the houses.There maximum number of bedroom in the houses and the minimum number of bathroom in the houses and the average is stories in the room.

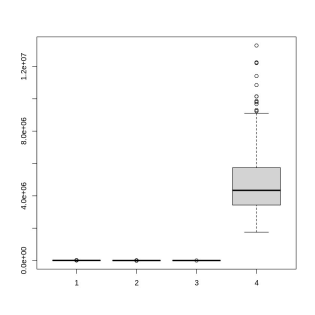
|  |
| --- |
| barplot(data$bedrooms, xlab = "NO.of bedroom", ylab = "", main ="Bar Chart") |



OBSERVATION:

Thus the above output image is the visualization of No.of bedroom using bar-chart in R . The No.of bathroom is higher in 6 present in the dataset. Average bedrooms are in the range of 4 . The minimum number of bedroom in the house is 1. There are many houses which has 2 bedrooms. There is only two houses which has one bedrooms.

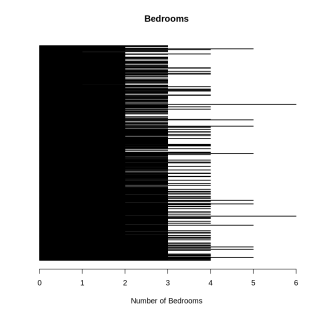
|  |
| --- |
| boxplot(data$area,data$bedrooms,data$bathrooms,data$price) |



OBSERVATION:

The above output is the visualization of area, bedrooms, bathrooms and price. The price range higher than any other value. The bedroom ,bathroom and area which might equal to one another.

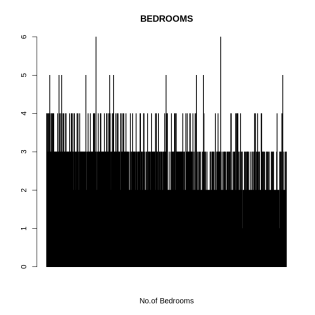
|  |
| --- |
| barplot(data$bedrooms,  main = 'Bedrooms',  xlab = 'Number of Bedrooms', horiz = TRUE) |



OBSERVATION:

Thus the output shows the visualization of bedrooms and bedrooms are higher at the range of 6 and Average bedrooms are in the range of 4 . The minimum number of bedroom in the house is 1. There are many houses which has 2 bedrooms. There is only two houses which has one bedrooms.

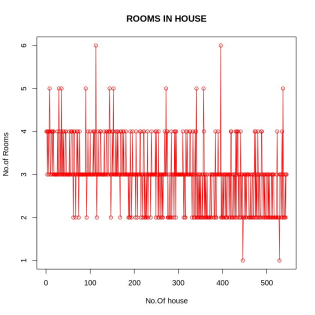
|  |
| --- |
| barplot(data$bedrooms, main = 'BEDROOMS',  xlab = 'No.of Bedrooms', col ='blue', horiz = FALSE) |



OBSERVATION:

The above output picture is the visualization of bedrooms in horizontal way and Major number of bedrooms are in the range 6. The minimum number of bedroom in the house is 1. There are many houses which has 2 bedrooms. There is only two houses which has one bedrooms.

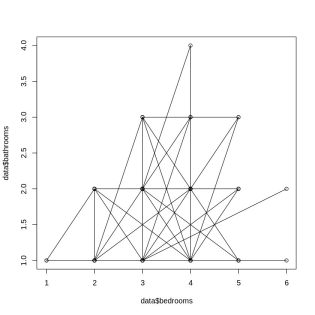
|  |
| --- |
| plot(data$bedrooms, type = "o", col = "red",  xlab = "No.Of house", ylab = "No.of Rooms ",  main = "ROOMS IN HOUSE") |



OBSERVATION:

Thus the above output shows the visualization between Number of rooms and Number of house . The number of rooms are rapidly increased from 2 to 6 .

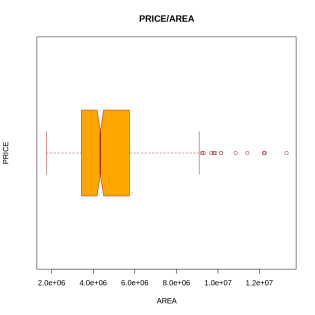
|  |
| --- |
| plot(data$bedrooms,data$bathrooms,type="o") |



OBSERVATION:

Thus the above output picture shows the visualization between data$bedrooms and data$bathrooms in R . The data$bedrooms are higher at 6 when data$bathrooms are at 4.0. The house 2 bathroom in many of the house. The houses has the 2 bedrooms in many of the houses.

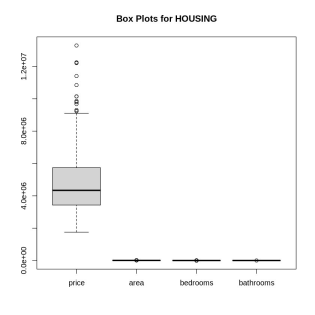
|  |
| --- |
| boxplot(data$price, main = "PRICE/AREA",  xlab = "AREA", ylab = "PRICE",  col = "orange", border = "brown",  horizontal = TRUE, notch = TRUE) |



OBSERVATION:

Thus the output image is the visualization between Area and Price Area is rapidly increased from 4.0e+06 . Area is equal to price of the area between the 4.0e+06to 6.0e=06.

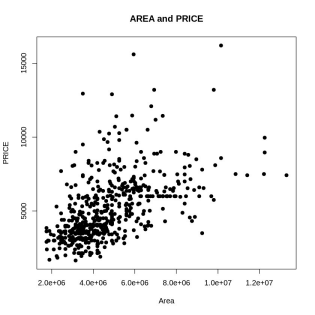
|  |
| --- |
| boxplot(data[, 0:4],  main ='Box Plots for HOUSING') |



OBSERVATION:

Thus the output is the visualization of HOUSING and price is highest at the range of 4.0e+05 . Remaining of these area , bedrooms and bathrooms are least at the range of 0.0e+00. The bedroom ,bathroom and area which might equal to one another.

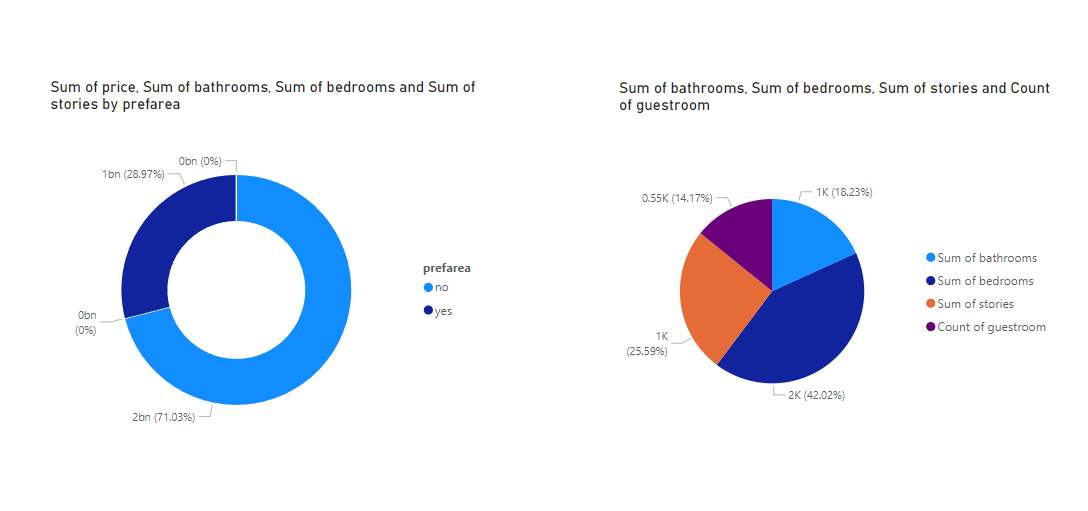
|  |
| --- |
| plot(data$price, data$area,  main ="AREA and PRICE",  xlab ="Area",  ylab =" PRICE ", pch = 19) |

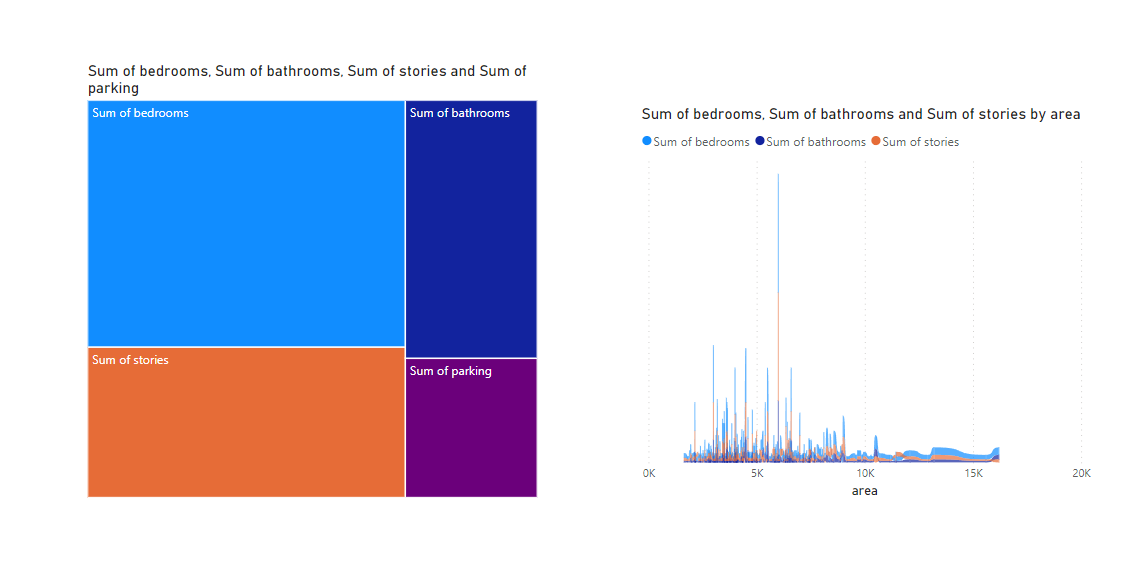


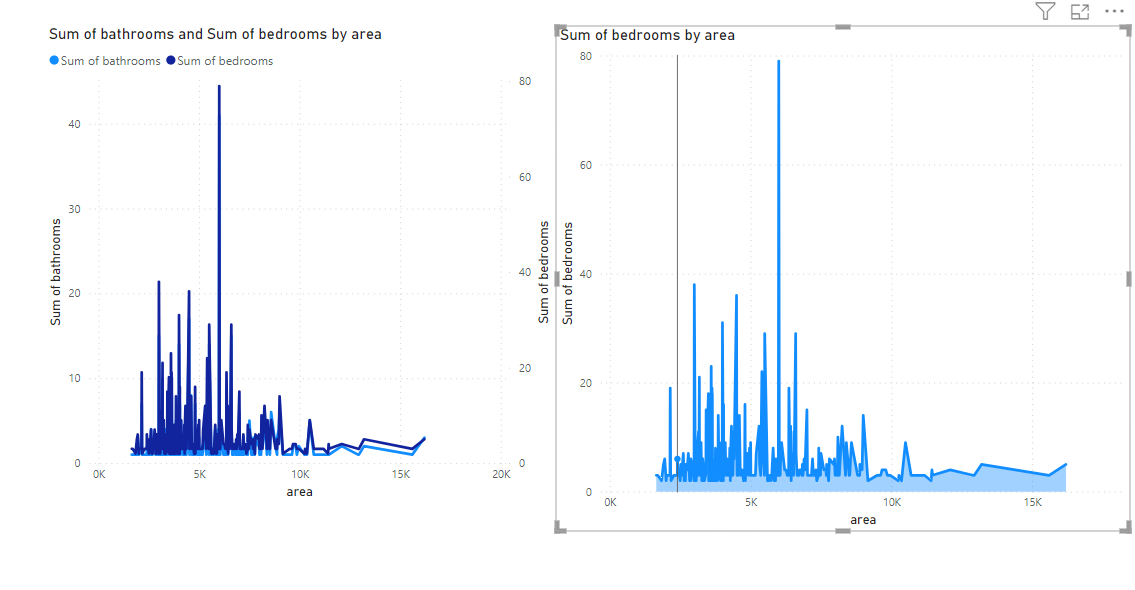
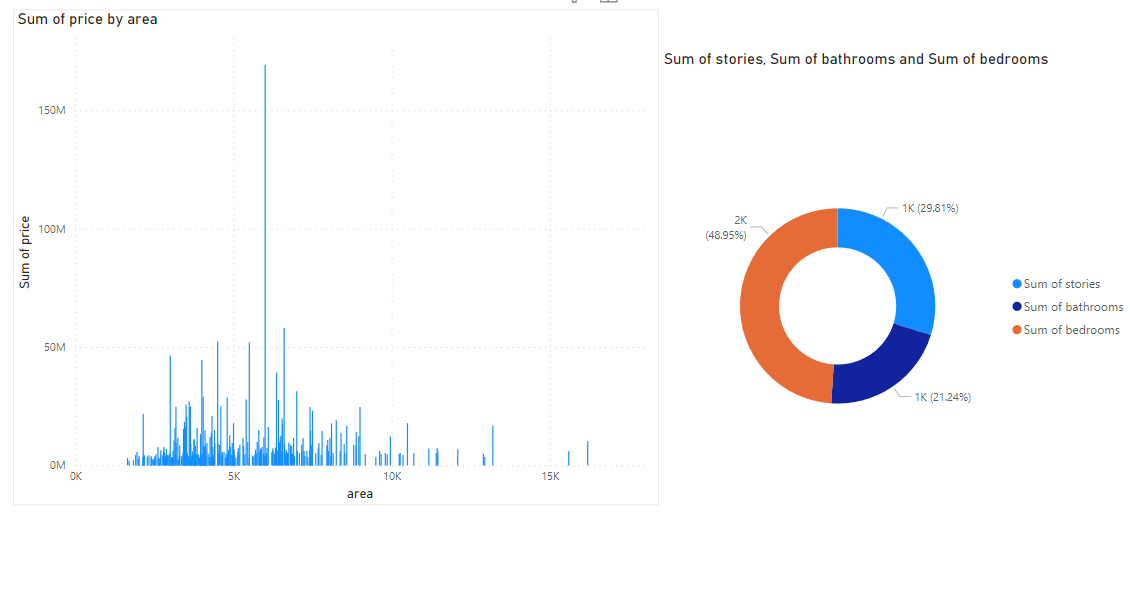
OBSERVATION:

The above output shows that visualization between area and price and area in 1.0e+07 is higher at the range of 15000 in price .The maximum number of areas price where between 2000 to 5000.The highest values occur two times.The lowest value in between 500 to 1000.

Power Bi:





RESULT:

Thus the above code was successfully executed and completed.