#1.libraries used

library(psych)

library(dplyr)

library(corrr)

library(corrplot)

library(RColorBrewer)

library(PerformanceAnalytics)

library(GGally)

#2.Data acquisition

#Reading the data set in csv files

house1<-read.csv(file.choose())

#3.Cleaning the dataset from NA if any

house<-na.omit(house1)

#4. Listing the dataset

house<-as.data.frame(house)

house

#5. cleaning the dataset from NA, if any

house<-na.omit(house)

#6. Data preparation and exploration

str(house)

summary(house)

names(house)

head(house, n=5)

tail(house, n=5)

#7. Splitting the data set into study population and analysis dataset

house\_n<-dplyr::select(house,1:3)

house\_n<-as.data.frame(house\_n)

house\_n

house\_Model<-dplyr::select(house,-1:-3)

house\_Model<-as.data.frame(house\_Model)

house\_Model

#8. Checking the study dataset

str(house\_n)

names(house\_n)

psych::headTail(house\_n)

#9. Visualization of variables

plot(house\_n$Name\_B, col="Green", main="Real estate agencies from where data was collected")

pie(1:4,labels=house\_n$proff,main="Distribution of profession of study population")

#10.converting the variables into the factors

house\_n$Name\_B<-factor(house\_n$Buyer)

house\_n$Proff<-factor(house\_n$Proff, levels=1:4, labels=c("Business","Private-Service","Government-Service","NRI"))

house\_n$Proff

#11.Visualization of the variables

plot(house\_n$Name\_B, col= "blue", main= "The agency from where data is collected")

pie(1:4, labels= house\_n$Proff, main = "DISTRIBUTION OF PROFESSIONS OF STUDY POPULATION")

#12. Summary of study population

table(house\_n$Name\_B)

tables<-ftable(house\_n$Name\_B, house\_n$Proff)

tables

#13. Data exploration of analysis variables datsets

# converting all the variables as in data dictionary

house\_Model$Price<-as.numeric(house\_Model$Price)

house\_Model$Price

house\_Model$Location<-factor(house\_Model$Location, level=1:4, labels= c("GURGAON","DELHI","NIODA","FARIDABAD"))

house\_Model$Location

house\_Model$Prop\_type<-factor(house\_Model$Prop\_type, level = 1:4, labels =c("1BHK","2BHK","3BHK","4BHK"))

house\_Model$Prop\_type

house\_Model$Carpet\_area<as.numeric(house\_Model$Carpet\_area)

house\_Model$Carpet\_area

house\_Model$Year<-factor(house\_Model$Year,levels=1:4,labels=c("2019","2018","2017","2016 and before"))

house\_Model$Year

house\_Model$Lease\_Per<-factor(house\_Model$Lease\_Per,levels=1:4, labels=c("0-5","6-10","11-15","15 and above"))

house\_Model$Lease\_Per

house\_Model$Prop\_purchase<-factor(house\_Model$Prop\_purchase,levels=1:2,labels=c("Loan","Purchased"))

house\_Model$Prop\_purchase

house\_Model$Payoff<-as.numeric(house\_Model$Payoff)

house\_Model$Payoff

house\_Model$Rental\_increment<-as.numeric(house\_Model$Rental\_increment)

house\_Model$Rental\_increment

house\_Model$Main\_Amount<-as.numeric(house\_Model$Main\_Amount)

house\_Model$Main\_Amount

house\_Model$Social\_infra<-as.factor(house\_Model$Social\_infra)

house\_Model$Social\_infra

#14. checking the analysis dataset

str(house\_Model)

names(house\_Model)

psych::headTail(house\_Model)

#15. Descriptive statistics of the variables of the analysis dataset

plot(house\_Model, col="red")

psych::describe(house\_Model)

#16. Descriptive statistics of the continuous variable and plotting the variables

summary(house\_Model$Price)

boxplot(house\_Model$Price, col="red", xlab= "Price", main="BOX PLOT OF THE IDICATIVE SELLING PRICE IN CRORE")

summary(house\_Model$Carpet\_area)

boxplot(house\_Model$Carpet\_area)

summary(house\_Model$Main\_Amount)

boxplot(house\_Model$Main\_Amount)

summary(house\_Model$Rental\_increment)

boxplot(house\_Model$Rental\_increment)

#17. Statistical Model development and evaluation multiple linear regression

Model1<-lm(Price~.,data=house\_Model)

summary(Model1)

step(Model1,direction = "both",trace=1)

anova(Model1)

predict<-predict(Model1)

#18. Diagnostic analysis of plots

plot(predict(Model1),col="blue")

#19. Transfering the output in CSV FILE

write.csv(output, file= "Predicted.csv", row.names=FALSE)

#20. Testing the correlation between the contributing varaibles

corr

# It can also be called using the traditional method

# network\_plot(correlate(mydata), min\_cor=0.5)

mydata<-house\_Model[,c('Price','Carpet\_area')]

mydata

chart.Correlation(mydata,histrogram="TRUE",pch=10)

pairs.panels(mydata, scale=TRUE)

ggpairs(mydata)

ggcorr(mydata, nbreaks=8, palette='RdGy', label=TRUE, label\_size=5, label\_color='red')