**SHWETANSHU ROHATGI Project.R Code**

# Hotel Room Pricing In Indian Market  
# NAME: Shwetanshu Rohatgi  
# EMAIL: shwetanshu.rohatgi@gmail.com  
# COLLEGE : Maharaja Surajmal Institute of Technology.  
  
##setting the directory and assigning a variabel to the data frame  
setwd("D:/IIML DSA/Final Project")  
  
#Reading the dataset and creating a data frame  
hotel.df<-read.csv(paste("Cities42.csv",sep = ""))  
  
#Viewing the data  
View(hotel.df)  
  
#Removing the repeated date by gsub command  
  
hotel.df$Date<-gsub("18-Dec-16", "Dec 18 2016", hotel.df$Date)  
hotel.df$Date<-gsub("21-Dec-16", "Dec 21 2016", hotel.df$Date)  
hotel.df$Date<-gsub("24-Dec-16", "Dec 24 2016", hotel.df$Date)  
hotel.df$Date<-gsub("25-Dec-16", "Dec 25 2016", hotel.df$Date)  
hotel.df$Date<-gsub("28-Dec-16", "Dec 28 2016", hotel.df$Date)  
hotel.df$Date<-gsub("31-Dec-16", "Dec 31 2016", hotel.df$Date)  
hotel.df$Date<-gsub("4-Jan-17", "Jan 04 2017", hotel.df$Date)  
hotel.df$Date<-gsub("4-Jan-16", "Jan 04 2017", hotel.df$Date)  
hotel.df$Date<-gsub("8-Jan-16", "Jan 08 2017", hotel.df$Date)  
hotel.df$Date<-gsub("8-Jan-17", "Jan 08 2017", hotel.df$Date)  
hotel.df$Date<-gsub("Jan 4 2017", "Jan 04 2017", hotel.df$Date)  
hotel.df$Date<-gsub("Jan 8 2017", "Jan 08 2017", hotel.df$Date)  
  
#Checking the dates  
  
table(hotel.df$Date)

##   
## Dec 18 2016 Dec 21 2016 Dec 24 2016 Dec 25 2016 Dec 28 2016 Dec 31 2016   
## 1652 1655 1655 1655 1655 1655   
## Jan 04 2017 Jan 08 2017   
## 1652 1653

#Changing dates to factors for labelling   
  
hotel.df$Date<-factor(hotel.df$Date)  
is.factor(hotel.df$Date)

## [1] TRUE

#Checking the labelling  
levels(hotel.df$Date)

## [1] "Dec 18 2016" "Dec 21 2016" "Dec 24 2016" "Dec 25 2016" "Dec 28 2016"  
## [6] "Dec 31 2016" "Jan 04 2017" "Jan 08 2017"

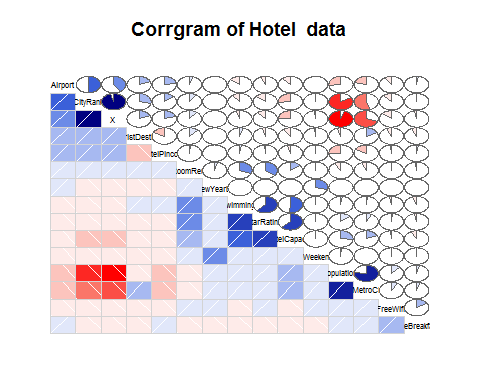
#Analyzing the summary of the data and describing the variables  
  
library(psych)  
describe(hotel.df)

## vars n mean sd median trimmed  
## X 1 13232 6616.50 3819.89 6616.5 6616.50  
## CityName\* 2 13232 18.07 11.72 16.0 17.29  
## Population 3 13232 4416836.87 4258386.00 3046163.0 4040816.22  
## CityRank 4 13232 14.83 13.51 9.0 13.30  
## IsMetroCity 5 13232 0.28 0.45 0.0 0.23  
## IsTouristDestination 6 13232 0.70 0.46 1.0 0.75  
## IsWeekend 7 13232 0.62 0.48 1.0 0.65  
## IsNewYearEve 8 13232 0.12 0.33 0.0 0.03  
## Date\* 9 13232 4.50 2.29 4.0 4.50  
## HotelName\* 10 13232 841.19 488.16 827.0 841.18  
## RoomRent 11 13232 5473.99 7333.12 4000.0 4383.33  
## StarRating 12 13232 3.46 0.76 3.0 3.40  
## Airport 13 13232 21.16 22.76 15.0 16.39  
## HotelAddress\* 14 13232 1202.53 582.17 1261.0 1233.25  
## HotelPincode 15 13232 397430.26 259837.50 395003.0 388540.47  
## HotelDescription\* 16 13224 581.34 363.26 567.0 575.37  
## FreeWifi 17 13232 0.93 0.26 1.0 1.00  
## FreeBreakfast 18 13232 0.65 0.48 1.0 0.69  
## HotelCapacity 19 13232 62.51 76.66 34.0 46.03  
## HasSwimmingPool 20 13232 0.36 0.48 0.0 0.32  
## mad min max range skew  
## X 4904.44 1.0 13232 13231.0 0.00  
## CityName\* 11.86 1.0 42 41.0 0.48  
## Population 3846498.95 8096.0 12442373 12434277.0 0.68  
## CityRank 11.86 0.0 44 44.0 0.69  
## IsMetroCity 0.00 0.0 1 1.0 0.96  
## IsTouristDestination 0.00 0.0 1 1.0 -0.86  
## IsWeekend 0.00 0.0 1 1.0 -0.51  
## IsNewYearEve 0.00 0.0 1 1.0 2.28  
## Date\* 2.97 1.0 8 7.0 0.00  
## HotelName\* 641.97 1.0 1670 1669.0 0.01  
## RoomRent 2653.85 299.0 322500 322201.0 16.75  
## StarRating 0.74 0.0 5 5.0 0.48  
## Airport 11.12 0.2 124 123.8 2.73  
## HotelAddress\* 668.65 1.0 2108 2107.0 -0.37  
## HotelPincode 257975.37 100025.0 7000157 6900132.0 9.99  
## HotelDescription\* 472.95 1.0 1226 1225.0 0.11  
## FreeWifi 0.00 0.0 1 1.0 -3.25  
## FreeBreakfast 0.00 0.0 1 1.0 -0.62  
## HotelCapacity 28.17 0.0 600 600.0 2.95  
## HasSwimmingPool 0.00 0.0 1 1.0 0.60  
## kurtosis se  
## X -1.20 33.21  
## CityName\* -0.88 0.10  
## Population -1.08 37019.65  
## CityRank -0.76 0.12  
## IsMetroCity -1.08 0.00  
## IsTouristDestination -1.26 0.00  
## IsWeekend -1.74 0.00  
## IsNewYearEve 3.18 0.00  
## Date\* -1.24 0.02  
## HotelName\* -1.25 4.24  
## RoomRent 582.06 63.75  
## StarRating 0.25 0.01  
## Airport 7.89 0.20  
## HotelAddress\* -0.88 5.06  
## HotelPincode 249.76 2258.86  
## HotelDescription\* -1.25 3.16  
## FreeWifi 8.57 0.00  
## FreeBreakfast -1.61 0.00  
## HotelCapacity 11.39 0.67  
## HasSwimmingPool -1.64 0.00

summary(hotel.df)

## X CityName Population CityRank   
## Min. : 1 Delhi :2048 Min. : 8096 Min. : 0.00   
## 1st Qu.: 3309 Jaipur : 768 1st Qu.: 744983 1st Qu.: 2.00   
## Median : 6616 Mumbai : 712 Median : 3046163 Median : 9.00   
## Mean : 6616 Bangalore: 656 Mean : 4416837 Mean :14.83   
## 3rd Qu.: 9924 Goa : 624 3rd Qu.: 8443675 3rd Qu.:24.00   
## Max. :13232 Kochi : 608 Max. :12442373 Max. :44.00   
## (Other) :7816   
## IsMetroCity IsTouristDestination IsWeekend IsNewYearEve   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :1.0000 Median :1.0000 Median :0.0000   
## Mean :0.2842 Mean :0.6972 Mean :0.6228 Mean :0.1244   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## Date HotelName RoomRent   
## Dec 21 2016:1655 Vivanta by Taj : 32 Min. : 299   
## Dec 24 2016:1655 Goldfinch Hotel : 24 1st Qu.: 2436   
## Dec 25 2016:1655 OYO Rooms : 24 Median : 4000   
## Dec 28 2016:1655 The Gordon House Hotel: 24 Mean : 5474   
## Dec 31 2016:1655 Apnayt Villa : 16 3rd Qu.: 6299   
## Jan 08 2017:1653 Bentleys Hotel Colaba : 16 Max. :322500   
## (Other) :3304 (Other) :13096   
## StarRating Airport   
## Min. :0.000 Min. : 0.20   
## 1st Qu.:3.000 1st Qu.: 8.40   
## Median :3.000 Median : 15.00   
## Mean :3.459 Mean : 21.16   
## 3rd Qu.:4.000 3rd Qu.: 24.00   
## Max. :5.000 Max. :124.00   
##   
## HotelAddress   
## The Mall, Shimla : 32   
## #2-91/14/8, White Fields, Kondapur, Hitech City, Hyderabad, 500084 India: 16   
## 121, City Terrace, Walchand Hirachand Marg, Mumbai, Maharashtra : 16   
## 14-4507/9, Balmatta Road, Near Jyothi Circle, Hampankatta : 16   
## 144/7, Rajiv Gandi Salai (OMR), Kottivakkam, Chennai, Tamil Nadu : 16   
## 17, Oliver Road, Colaba, Mumbai, Maharashtra : 16   
## (Other) :13120   
## HotelPincode HotelDescription FreeWifi FreeBreakfast   
## Min. : 100025 3 : 120 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 221001 Abc : 112 1st Qu.:1.0000 1st Qu.:0.0000   
## Median : 395003 3-star hotel: 104 Median :1.0000 Median :1.0000   
## Mean : 397430 3.5 : 88 Mean :0.9259 Mean :0.6491   
## 3rd Qu.: 570001 4 : 72 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :7000157 (Other) :12728 Max. :1.0000 Max. :1.0000   
## NA's : 8   
## HotelCapacity HasSwimmingPool   
## Min. : 0.00 Min. :0.0000   
## 1st Qu.: 16.00 1st Qu.:0.0000   
## Median : 34.00 Median :0.0000   
## Mean : 62.51 Mean :0.3558   
## 3rd Qu.: 75.00 3rd Qu.:1.0000   
## Max. :600.00 Max. :1.0000   
##

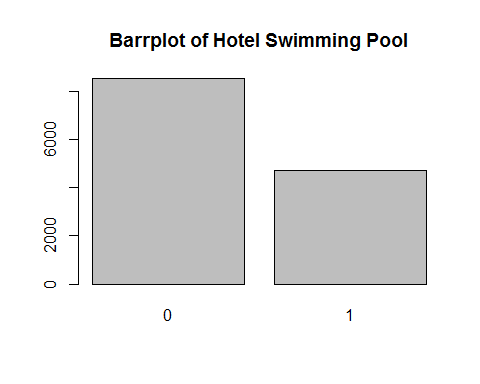
#Taking Y = RoomRent, identifying the most relevent predictor variables by boruta and correlation  
  
#Corrgram  
  
library(corrgram)  
  
corrgram(hotel.df, order=TRUE, lower.panel=panel.shade,  
 upper.panel=panel.pie, text.panel=panel.txt,  
 main="Corrgram of Hotel data")



##through corrgram HasSwimming, StarRating, HotelCapital are very well correlated to RoomRent  
 ##so we can take them as predictors  
  
##Visualizing data for Y as Room rent and X1,X2,X3 as HasSwimmingPool, StarRating and HotelCapacity respectively  
  
 #Table for HasSwimmingPool  
 table(hotel.df$HasSwimmingPool)

##   
## 0 1   
## 8524 4708

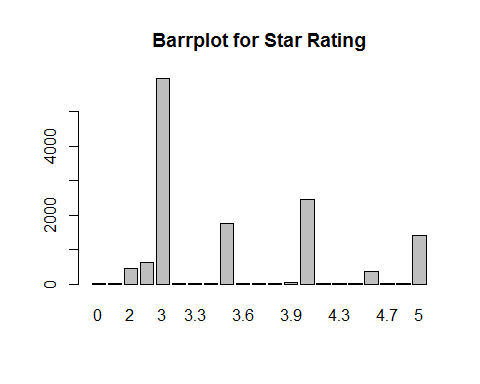
Swim<-table(hotel.df$HasSwimmingPool)  
 barplot(Swim,main="Barrplot of Hotel Swimming Pool")



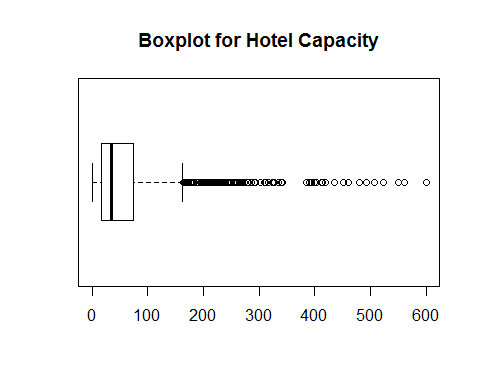
#Table for StarRating  
 table(hotel.df$StarRating)

##   
## 0 1 2 2.5 3 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4 4.1   
## 16 8 440 632 5953 8 16 8 1752 8 24 16 32 2463 24   
## 4.3 4.4 4.5 4.7 4.8 5   
## 16 8 376 8 16 1408

starRating<-table(hotel.df$StarRating)  
 barplot(starRating,main = "Barrplot for Star Rating")



#BoxPlot for HotelCapacity  
 boxplot(hotel.df$HotelCapacity, main="Boxplot for Hotel Capacity",horizontal = TRUE)

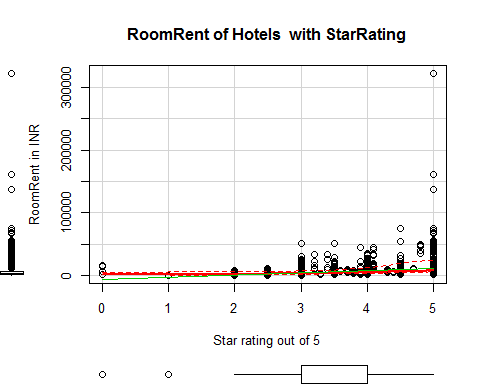


#Scatterplot pair wise for predictor variable  
   
 library(car)

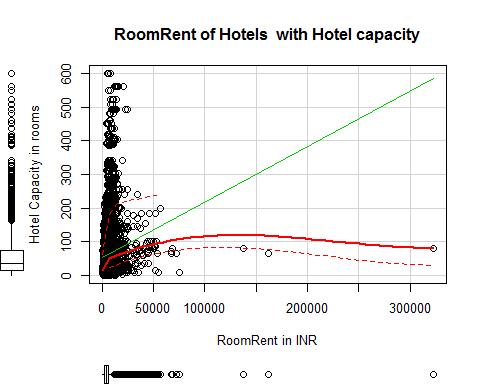
##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

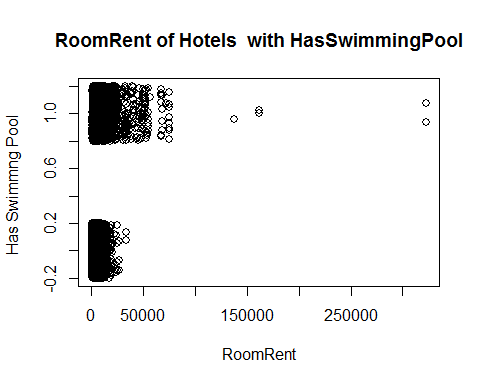
#StarRating Vs RoomRent  
   
 scatterplot(hotel.df$StarRating,hotel.df$RoomRent,main="RoomRent of Hotels with StarRating",ylab = "RoomRent in INR", xlab="Star rating out of 5",cex=1.1)



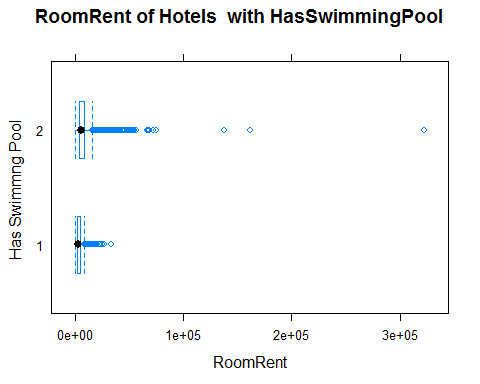
#RoomRent Vs HotelCapacity  
   
 scatterplot(hotel.df$RoomRent,hotel.df$HotelCapacity,main="RoomRent of Hotels with Hotel capacity",ylab = "Hotel Capacity in rooms", xlab="RoomRent in INR",cex=1.1)



#RoomRent Vs HasSwimmingPool  
   
 plot(jitter(hotel.df$RoomRent),jitter(hotel.df$HasSwimmingPool),main="RoomRent of Hotels with HasSwimmingPool",ylab = "Has Swimmng Pool ", xlab="RoomRent",cex=1.1)

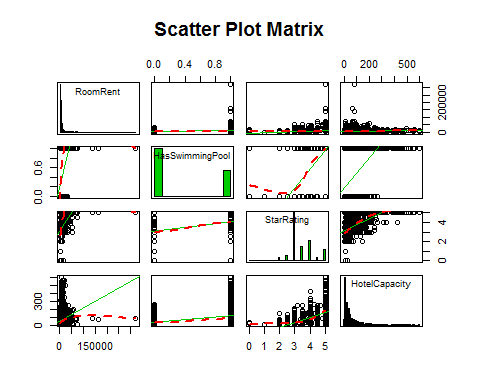


library(lattice)  
 bwplot(HasSwimmingPool~RoomRent, data = hotel.df,main="RoomRent of Hotels with HasSwimmingPool",ylab = "Has Swimmng Pool ", xlab="RoomRent" )

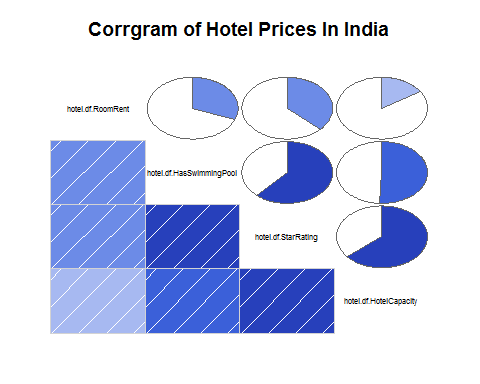


#Scatterplot matrix  
   
 scatterplotMatrix(  
 hotel.df[  
 ,c("RoomRent","HasSwimmingPool","StarRating", "HotelCapacity")],   
 spread=FALSE, smoother.args=list(lty=2),  
 main="Scatter Plot Matrix", diagonal = "histogram")

## Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,  
## spread = spread, : could not fit smooth



#Corrgram of Y, x1, x2, x3  
   
 library(corrgram)  
   
 xyz<-data.frame(hotel.df$RoomRent, hotel.df$HasSwimmingPool, hotel.df$HotelCapacity, hotel.df$StarRating)  
 corrgram(xyz, order=TRUE, lower.panel=panel.shade,  
 upper.panel=panel.pie, text.panel=panel.txt,  
 main="Corrgram of Hotel Prices In India")



#Variance-Covariance Matrix for Y, x1, x2, x3  
  
 x<-hotel.df[,c("HasSwimmingPool","StarRating", "HotelCapacity")]  
 y<-hotel.df[,c("RoomRent")]  
 cor(x,y)

## [,1]  
## HasSwimmingPool 0.3116577  
## StarRating 0.3693734  
## HotelCapacity 0.1578733

cov(x,y)

## [,1]  
## HasSwimmingPool 1094.202  
## StarRating 2048.375  
## HotelCapacity 88753.413

var(x,y)

## [,1]  
## HasSwimmingPool 1094.202  
## StarRating 2048.375  
## HotelCapacity 88753.413