**Analysis of Hotel Room Pricing In the Indian Market**

**Name: Raghu Ram Chinthakrinda**

**E-mail:** [**edm14b010@iiitdm.ac.in**](mailto:edm14b010@iiitdm.ac.in)

**College: IIITDM KANCHEEPURAM**

**Mobile No: 7382998506**

File: Raghu\_Ram\_Chinthakrinda\_Capstone\_Project\_Code.R

RAM

Tue Jul 11 16:56:05 2017

# Hotel Room Pricing In The Indian Market  
# NAME: Raghu Ram Chinthakrinda  
# EMAIL: edm14b010@iiitdm.ac.in  
# COLLEGE : IIITDM KANCHEEPURAM  
  
#set the working directory  
setwd("D:/capstone project")  
  
#Reading the dataset  
hotel.df<-read.csv("Cities42.csv")  
  
#Viewing the data  
View(hotel.df)  
  
#attaching the dataset  
attach(hotel.df)  
  
#first 10 rows of the dataset  
head(hotel.df)

## X CityName Population CityRank IsMetroCity IsTouristDestination  
## 1 1 Mumbai 12442373 0 1 1  
## 2 2 Mumbai 12442373 0 1 1  
## 3 3 Mumbai 12442373 0 1 1  
## 4 4 Mumbai 12442373 0 1 1  
## 5 5 Mumbai 12442373 0 1 1  
## 6 6 Mumbai 12442373 0 1 1  
## IsWeekend IsNewYearEve Date HotelName RoomRent StarRating  
## 1 1 0 Dec 18 2016 Vivanta by Taj 12375 5  
## 2 0 0 Dec 21 2016 Vivanta by Taj 10250 5  
## 3 1 0 Dec 24 2016 Vivanta by Taj 9900 5  
## 4 1 0 Dec 25 2016 Vivanta by Taj 10350 5  
## 5 0 0 Dec 28 2016 Vivanta by Taj 12000 5  
## 6 1 1 Dec 31 2016 Vivanta by Taj 11475 5  
## Airport HotelAddress HotelPincode  
## 1 21 90 Cuffe Parade, Colaba, Mumbai, Maharashtra 400005  
## 2 21 91 Cuffe Parade, Colaba, Mumbai, Maharashtra 400006  
## 3 21 92 Cuffe Parade, Colaba, Mumbai, Maharashtra 400007  
## 4 21 93 Cuffe Parade, Colaba, Mumbai, Maharashtra 400008  
## 5 21 94 Cuffe Parade, Colaba, Mumbai, Maharashtra 400009  
## 6 21 95 Cuffe Parade, Colaba, Mumbai, Maharashtra 400010  
## HotelDescription FreeWifi FreeBreakfast  
## 1 Luxury hotel with spa, near Gateway of India 1 0  
## 2 Luxury hotel with spa, near Gateway of India 1 0  
## 3 Luxury hotel with spa, near Gateway of India 1 0  
## 4 Luxury hotel with spa, near Gateway of India 1 0  
## 5 Luxury hotel with spa, near Gateway of India 1 0  
## 6 Luxury hotel with spa, near Gateway of India 1 0  
## HotelCapacity HasSwimmingPool  
## 1 287 1  
## 2 287 1  
## 3 287 1  
## 4 287 1  
## 5 287 1  
## 6 287 1

#last 10 rows of the dataset  
tail(hotel.df)

## X CityName Population CityRank IsMetroCity IsTouristDestination  
## 13227 13227 Manali 8096 44 0 1  
## 13228 13228 Manali 8096 44 0 1  
## 13229 13229 Manali 8096 44 0 1  
## 13230 13230 Manali 8096 44 0 1  
## 13231 13231 Manali 8096 44 0 1  
## 13232 13232 Manali 8096 44 0 1  
## IsWeekend IsNewYearEve Date  
## 13227 1 0 Dec 24 2016  
## 13228 1 0 Dec 25 2016  
## 13229 0 0 Dec 28 2016  
## 13230 1 1 Dec 31 2016  
## 13231 0 0 Jan 04 2017  
## 13232 1 0 Jan 08 2017  
## HotelName RoomRent StarRating  
## 13227 Manuallaya The Resort Spa in the Himalayas 9668 5  
## 13228 Manuallaya The Resort Spa in the Himalayas 9668 5  
## 13229 Manuallaya The Resort Spa in the Himalayas 9668 5  
## 13230 Manuallaya The Resort Spa in the Himalayas 9668 5  
## 13231 Manuallaya The Resort Spa in the Himalayas 9668 5  
## 13232 Manuallaya The Resort Spa in the Himalayas 7666 5  
## Airport HotelAddress HotelPincode  
## 13227 41 Sunny Side Chadiyari, Manali, Himachal Pradesh 175131  
## 13228 41 Sunny Side Chadiyari, Manali, Himachal Pradesh 175131  
## 13229 41 Sunny Side Chadiyari, Manali, Himachal Pradesh 175131  
## 13230 41 Sunny Side Chadiyari, Manali, Himachal Pradesh 175131  
## 13231 41 Sunny Side Chadiyari, Manali, Himachal Pradesh 175131  
## 13232 41 Sunny Side Chadiyari, Manali, Himachal Pradesh 175131  
## HotelDescription FreeWifi  
## 13227 Manali ski-in/ski-out resort with spa, indoor pool 0  
## 13228 Manali ski-in/ski-out resort with spa, indoor pool 0  
## 13229 Manali ski-in/ski-out resort with spa, indoor pool 0  
## 13230 Manali ski-in/ski-out resort with spa, indoor pool 0  
## 13231 Manali ski-in/ski-out resort with spa, indoor pool 0  
## 13232 Manali ski-in/ski-out resort with spa, indoor pool 0  
## FreeBreakfast HotelCapacity HasSwimmingPool  
## 13227 1 55 0  
## 13228 1 55 0  
## 13229 1 55 0  
## 13230 1 55 0  
## 13231 1 55 0  
## 13232 1 55 0

#omitting the na from the dataframe  
hotel.df=na.omit(hotel.df)  
  
#dimensiions of the data frame   
dim(hotel.df)

## [1] 13224 20

#column names of the data frame  
colnames(hotel.df)

## [1] "X" "CityName" "Population"   
## [4] "CityRank" "IsMetroCity" "IsTouristDestination"  
## [7] "IsWeekend" "IsNewYearEve" "Date"   
## [10] "HotelName" "RoomRent" "StarRating"   
## [13] "Airport" "HotelAddress" "HotelPincode"   
## [16] "HotelDescription" "FreeWifi" "FreeBreakfast"   
## [19] "HotelCapacity" "HasSwimmingPool"

#Removing the repeated date by gsub command  
Date<-gsub("18-Dec-16", "Dec 18 2016", Date)  
Date<-gsub("21-Dec-16", "Dec 21 2016", Date)  
Date<-gsub("24-Dec-16", "Dec 24 2016", Date)  
Date<-gsub("25-Dec-16", "Dec 25 2016", Date)  
Date<-gsub("28-Dec-16", "Dec 28 2016", Date)  
Date<-gsub("31-Dec-16", "Dec 31 2016", Date)  
Date<-gsub("4-Jan-17", "Jan 04 2017", Date)  
Date<-gsub("4-Jan-16", "Jan 04 2017", Date)  
Date<-gsub("8-Jan-16", "Jan 08 2017", Date)  
Date<-gsub("8-Jan-17", "Jan 08 2017", Date)  
Date<-gsub("Jan 4 2017", "Jan 04 2017", Date)  
Date<-gsub("Jan 8 2017", "Jan 08 2017", Date)  
  
  
#Checking the dates  
table(Date)

## 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   
Date<-factor( Date)  
is.factor( Date)

## [1] TRUE

#Checking the labelling  
levels( 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)  
summary=describe(hotel.df)  
summary[,c(3,4,5,8,9)]

## mean sd median min max  
## X 6620.42 3817.72 6620.5 1.0 13232  
## CityName\* 18.06 11.72 16.0 1.0 42  
## Population 4411981.74 4255094.69 3046163.0 8096.0 12442373  
## CityRank 14.84 13.51 9.0 0.0 44  
## IsMetroCity 0.28 0.45 0.0 0.0 1  
## IsTouristDestination 0.70 0.46 1.0 0.0 1  
## IsWeekend 0.62 0.48 1.0 0.0 1  
## IsNewYearEve 0.12 0.33 0.0 0.0 1  
## Date\* 14.26 2.82 14.0 1.0 20  
## HotelName\* 841.33 488.27 827.5 1.0 1670  
## RoomRent 5475.67 7335.01 4000.0 299.0 322500  
## StarRating 3.46 0.76 3.0 0.0 5  
## Airport 21.16 22.77 15.0 0.2 124  
## HotelAddress\* 1202.98 582.06 1261.0 1.0 2108  
## HotelPincode 397428.71 259916.08 395003.0 100025.0 7000157  
## HotelDescription\* 581.34 363.26 567.0 1.0 1226  
## FreeWifi 0.93 0.26 1.0 0.0 1  
## FreeBreakfast 0.65 0.48 1.0 0.0 1  
## HotelCapacity 62.53 76.68 34.0 0.0 600  
## HasSwimmingPool 0.36 0.48 0.0 0.0 1

#detaching the hotel.df dataset  
detach(hotel.df)  
  
#taking only numeric dataset by eliminating test  
numeric\_data=hotel.df[c(4,3,5,6,7,8,11,12,13,17,18,19,20)]  
attach(numeric\_data)  
  
#checking the important variables  
library(Boruta)

## Warning: package 'Boruta' was built under R version 3.4.1

## Loading required package: ranger

## Warning: package 'ranger' was built under R version 3.4.1

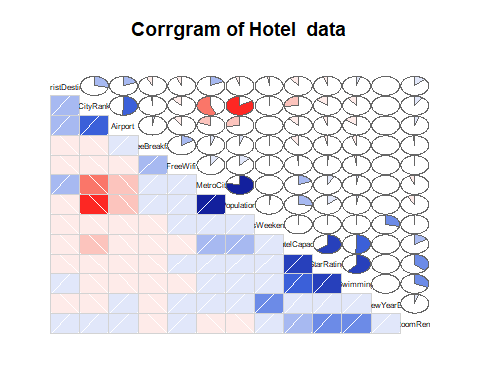
l=Boruta(RoomRent~.,data = numeric\_data)  
print (l)

## Boruta performed 11 iterations in 5.308567 mins.  
## 10 attributes confirmed important: Airport, CityRank,  
## FreeBreakfast, FreeWifi, HasSwimmingPool and 5 more;  
## 2 attributes confirmed unimportant: IsNewYearEve, IsWeekend;

#Taking Y = RoomRent, identifying the most relevent predictor variables by correlation corrgram  
#Corrgram  
  
library(corrgram)

## Warning: package 'corrgram' was built under R version 3.4.1

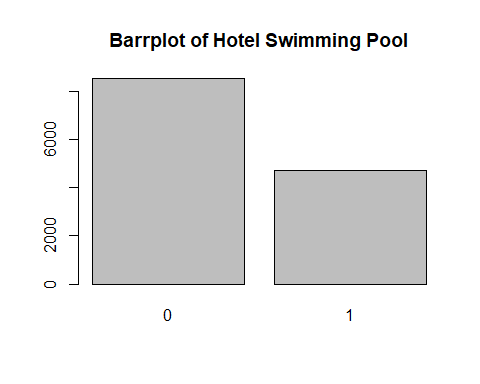
corrgram(numeric\_data, order=TRUE, lower.panel=panel.shade,  
 upper.panel=panel.pie, text.panel=panel.txt,  
 main="Corrgram of Hotel data")



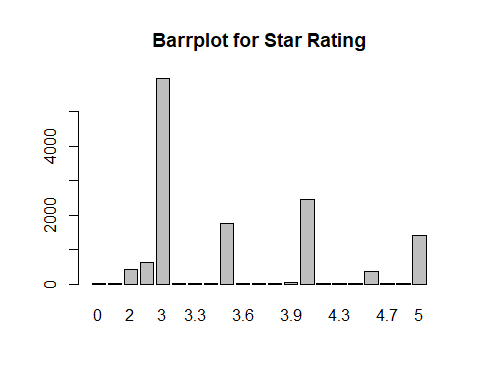
#more correlation is shown in swimming pool, starrating, hotel capacity w.r.t to room rent  
  
  
#corrleation between room rent and remaining variables  
cor(RoomRent,numeric\_data)

## CityRank Population IsMetroCity IsTouristDestination IsWeekend  
## [1,] 0.09377257 -0.08839562 -0.06653015 0.1226753 0.004584006  
## IsNewYearEve RoomRent StarRating Airport FreeWifi FreeBreakfast  
## [1,] 0.03849218 1 0.3693641 0.04963384 0.002829088 -0.01032386  
## HotelCapacity HasSwimmingPool  
## [1,] 0.157782 0.3115533

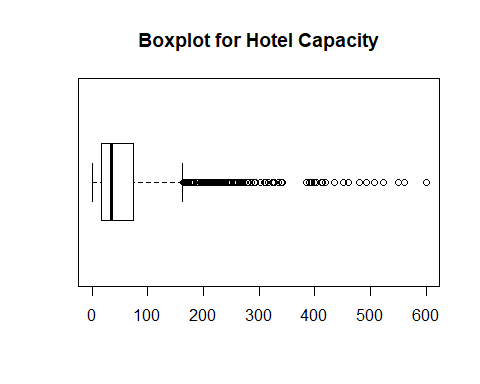
#independent variablea are population, ismetrocity, freebreakfast  
#categorical variables are has swimming pool, starrating, hotel capacity  
  
#Table for HasSwimmingPool  
Swim<-table(HasSwimmingPool)  
barplot(Swim,main="Barrplot of Hotel Swimming Pool")



##hasswimming pool is categorical  
   
#Table for StarRating  
starRating<-table(StarRating)  
barplot(starRating,main = "Barrplot for Star Rating")



##starrating is categorical  
  
#BoxPlot for HotelCapacity  
boxplot( HotelCapacity, main="Boxplot for Hotel Capacity",horizontal = TRUE)

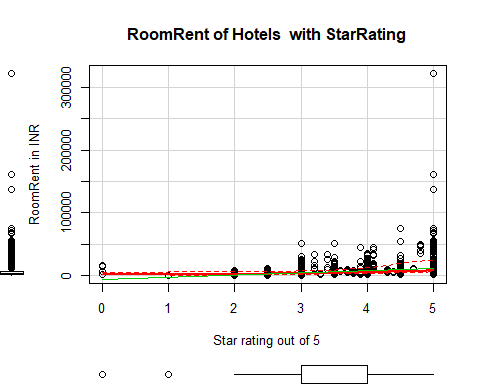


##hotel capacity is continous  
  
  
#Scatterplot pair wise for predictor variable  
library(car)

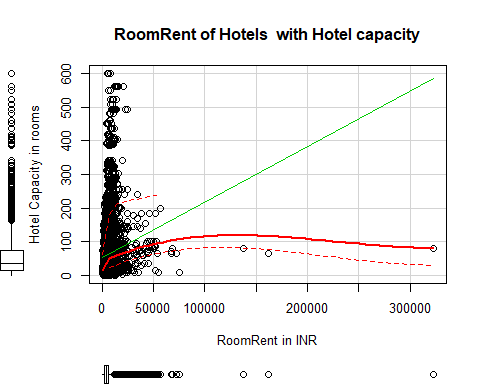
##   
## Attaching package: 'car'

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

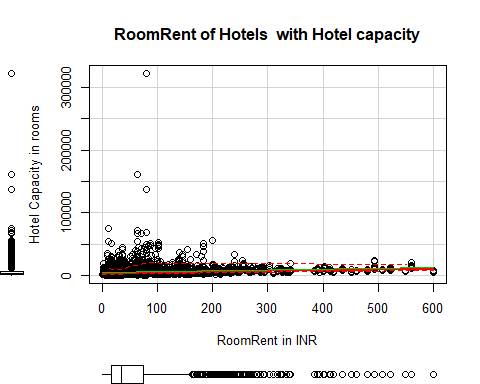
#StarRating Vs RoomRent  
scatterplot( StarRating, RoomRent,main="RoomRent of Hotels with StarRating",ylab = "RoomRent in INR", xlab="Star rating out of 5",cex=1.1)



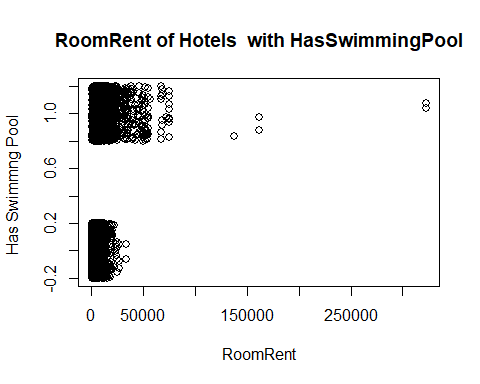
#RoomRent Vs HotelCapacity  
scatterplot( RoomRent, HotelCapacity,main="RoomRent of Hotels with Hotel capacity",ylab = "Hotel Capacity in rooms", xlab="RoomRent in INR",cex=1.1)



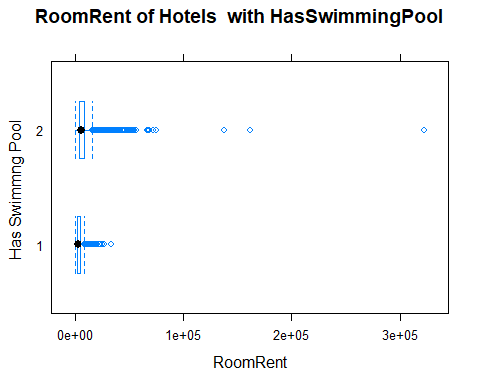
scatterplot( HotelCapacity,RoomRent,main="RoomRent of Hotels with Hotel capacity",ylab = "Hotel Capacity in rooms", xlab="RoomRent in INR",cex=1.1)



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

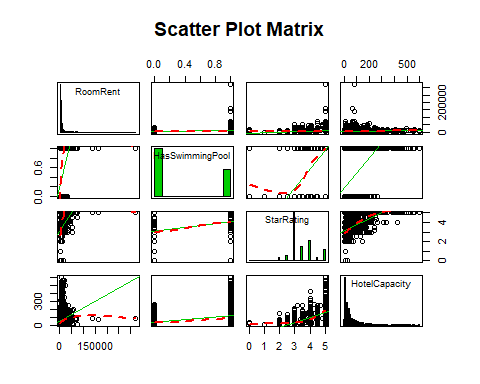


library(lattice)  
bwplot(HasSwimmingPool~RoomRent, data = hotel.df,main="RoomRent of Hotels with HasSwimmingPool",ylab = "Has Swimmng Pool ", xlab="RoomRent" )

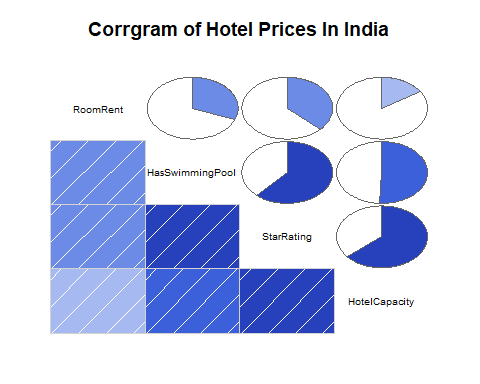


#Scatterplot matrix  
   
scatterplotMatrix( hotel.df[ ,c("RoomRent","HasSwimmingPool","StarRating", "HotelCapacity")],   
 spread=FALSE, smoother.args=list(lty=2),  
 main="Scatter Plot Matrix", diagonal = "histogram")

## 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( RoomRent, HasSwimmingPool, HotelCapacity, 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 0.3115533  
## StarRating 0.3693641  
## HotelCapacity 0.1577820

cov(x,y)

## [,1]  
## HasSwimmingPool 1094.266  
## StarRating 2047.165  
## HotelCapacity 88745.795

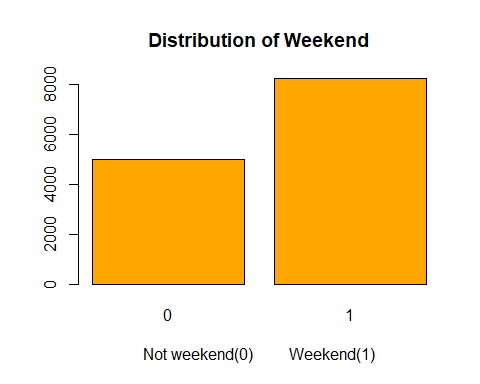
var(x,y)

## [,1]  
## HasSwimmingPool 1094.266  
## StarRating 2047.165  
## HotelCapacity 88745.795

#week4 day6  
#Forming a variable which is having RoomRent less than 1 lakh because the outliers effect the average  
RoomRent1.df <-hotel.df[which( RoomRent<100000),]  
  
#Comparing other factors and their pattern using other trends with roomrent  
   
#Analyzing IsWeekeng effect on RoomRent  
table( IsWeekend)

## IsWeekend  
## 0 1   
## 4988 8236

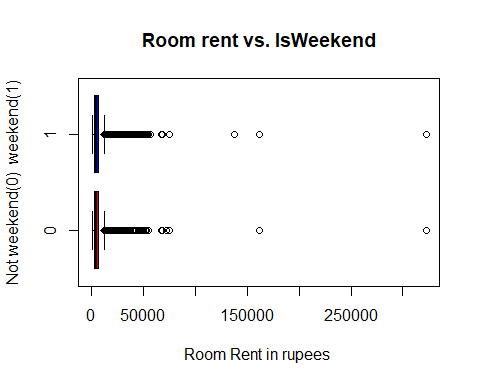
table1<-table( IsWeekend)  
barplot(table1, main="Distribution of Weekend", xlab="Not weekend(0) Weekend(1)", col="orange")



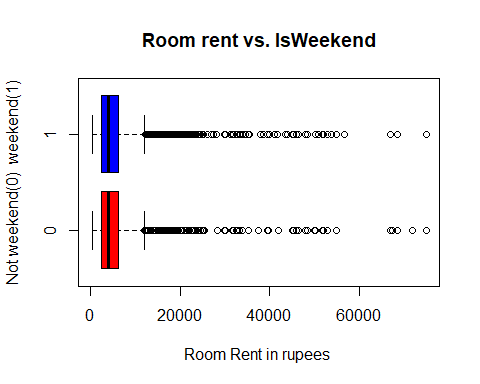
#Effect of Isweekend on RoomRent  
iw= aggregate(RoomRent ~ IsWeekend, data=hotel.df,mean)  
iw

## IsWeekend RoomRent  
## 1 0 5432.466  
## 2 1 5501.836

boxplot(RoomRent~IsWeekend,data=hotel.df, main="Room rent vs. IsWeekend", ylab="Not weekend(0) weekend(1)", xlab="Room Rent in rupees ", col=c("red","blue"),horizontal=TRUE)

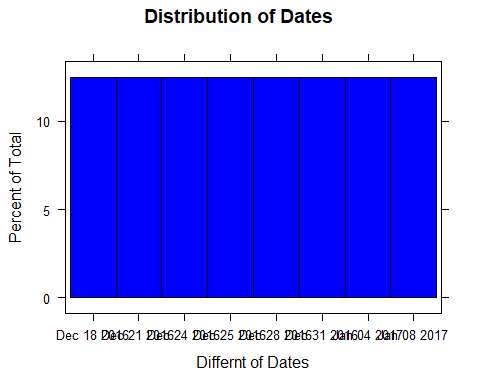


#Without extreme outliers   
boxplot(RoomRent~IsWeekend,data=RoomRent1.df, main="Room rent vs. IsWeekend", ylab="Not weekend(0) weekend(1)", xlab="Room Rent in rupees ", col=c("red","blue"),horizontal=TRUE)



library(lattice)  
histogram(Date, data = hotel.df, main="Distribution of Dates", xlab = "Differnt of Dates", col="Blue")

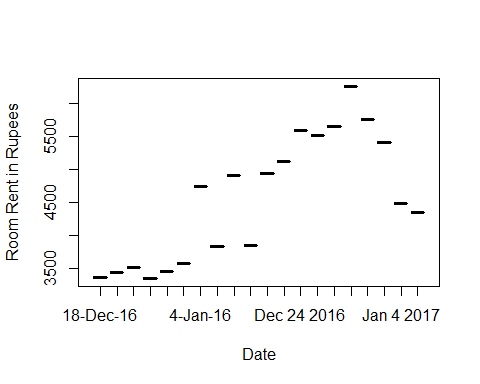
## Warning in histogram.factor(Date, data = hotel.df, main = "Distribution of  
## Dates", : explicit 'data' specification ignored



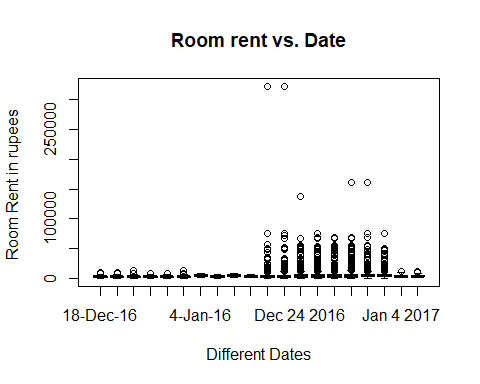
#Effect of different dates on RoomRent  
d = aggregate(RoomRent ~ Date, data = hotel.df,mean)  
d

## Date RoomRent  
## 1 18-Dec-16 3366.795  
## 2 21-Dec-16 3437.545  
## 3 24-Dec-16 3510.795  
## 4 25-Dec-16 3349.591  
## 5 28-Dec-16 3450.045  
## 6 31-Dec-16 3570.318  
## 7 4-Jan-16 4738.548  
## 8 4-Jan-17 3829.615  
## 9 8-Jan-16 4907.419  
## 10 8-Jan-17 3843.077  
## 11 Dec 18 2016 4939.759  
## 12 Dec 21 2016 5131.939  
## 13 Dec 24 2016 5600.656  
## 14 Dec 25 2016 5523.613  
## 15 Dec 28 2016 5654.057  
## 16 Dec 31 2016 6265.332  
## 17 Jan 04 2017 5756.601  
## 18 Jan 08 2017 5408.692  
## 19 Jan 4 2017 4481.400  
## 20 Jan 8 2017 4347.821

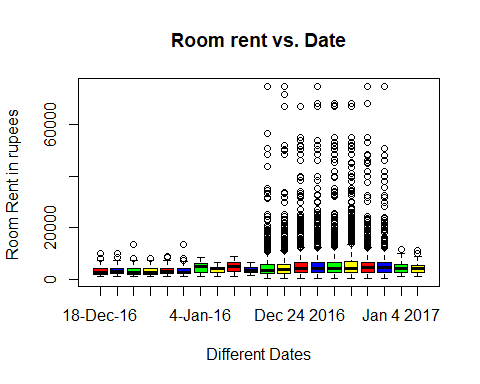
scatterplot(d$Date,d$RoomRent, main="Scatterplot between Date and RoomRent", xlab="Date", ylab = "Room Rent in Rupees")



boxplot(RoomRent~Date,data=hotel.df, main="Room rent vs. Date", xlab="Different Dates", ylab="Room Rent in rupees ", col=c("red","blue","green","yellow"))



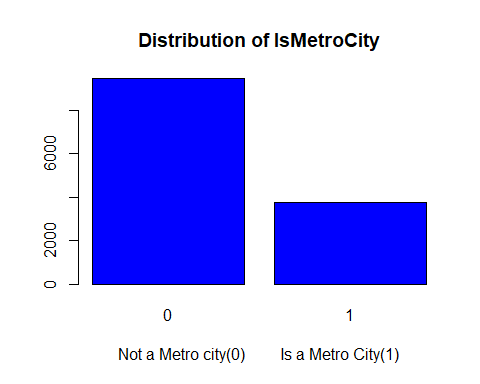
##Without extreme outliers  
boxplot(RoomRent~Date,data=RoomRent1.df, main="Room rent vs. Date", xlab="Different Dates", ylab="Room Rent in rupees ", col=c("red","blue","green","yellow"))



#Analyzing IsMetroCity effect on RoomRent  
table( IsMetroCity)

## IsMetroCity  
## 0 1   
## 9472 3752

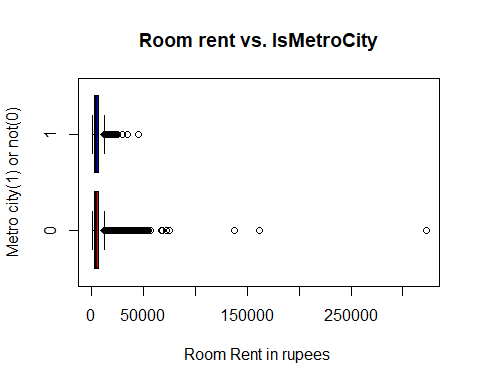
table1<-table( IsMetroCity)  
barplot(table1, main="Distribution of IsMetroCity", xlab="Not a Metro city(0) Is a Metro City(1)", col="blue")



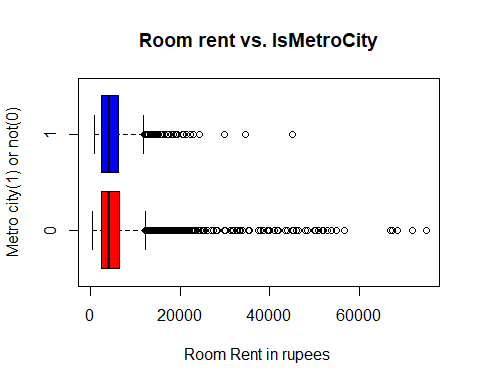
#Effect of IsMetroCity on RoomRent  
imc = aggregate(RoomRent ~ IsMetroCity, data = hotel.df, mean)  
imc

## IsMetroCity RoomRent  
## 1 0 5782.794  
## 2 1 4700.329

boxplot(RoomRent~IsMetroCity,data=hotel.df, main="Room rent vs. IsMetroCity", ylab="Metro city(1) or not(0)", xlab="Room Rent in rupees ", col=c("red","blue","green","yellow"),horizontal=TRUE)



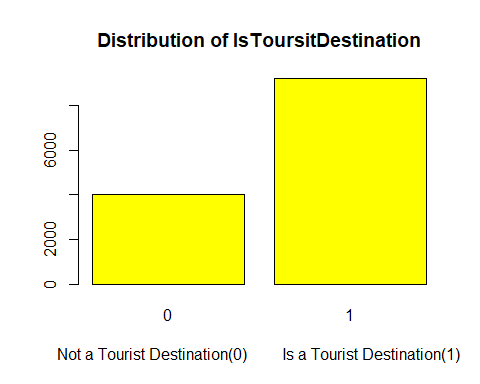
##Without extreme outliers  
boxplot(RoomRent~IsMetroCity,data=RoomRent1.df, main="Room rent vs. IsMetroCity", ylab="Metro city(1) or not(0)", xlab="Room Rent in rupees ", col=c("red","blue","green","yellow"),horizontal=TRUE)



#Analyzing IsTouristDestination effect on RoomRent  
table( IsTouristDestination)

## IsTouristDestination  
## 0 1   
## 4007 9217

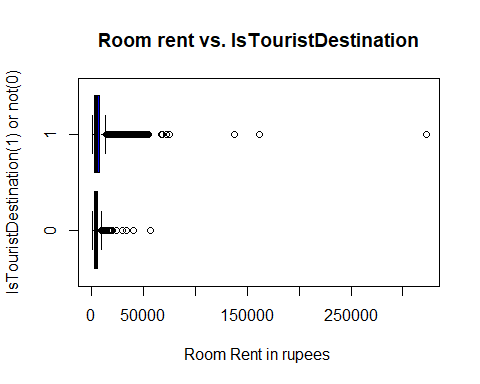
table1<-table( IsTouristDestination)  
barplot(table1, main="Distribution of IsToursitDestination", xlab="Not a Tourist Destination(0) Is a Tourist Destination(1)", col="yellow")



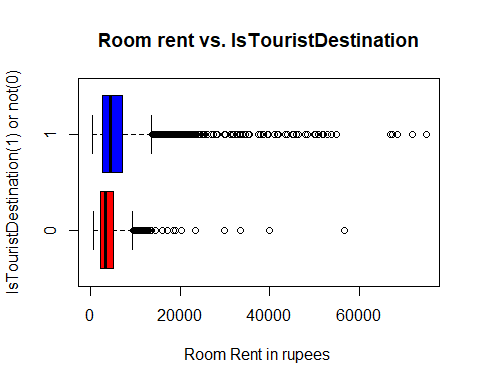
#Effect of IsTouristDestination on RoomRent  
itd = aggregate(RoomRent ~ IsTouristDestination, data = hotel.df, mean)  
itd

## IsTouristDestination RoomRent  
## 1 0 4111.003  
## 2 1 6068.946

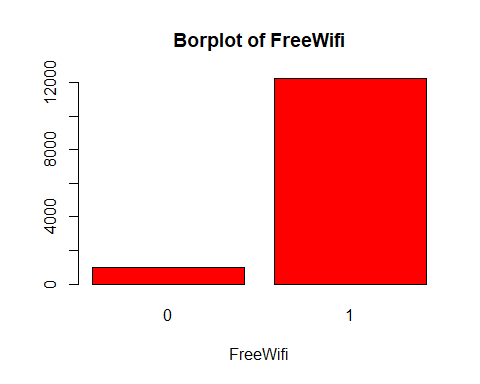
boxplot(RoomRent~IsTouristDestination,data=hotel.df, main="Room rent vs. IsTouristDestination", ylab="IsTouristDestination(1) or not(0)", xlab="Room Rent in rupees ", col=c("red","blue","green","yellow"),horizontal=TRUE)



##Without extreme outliers  
boxplot(RoomRent~IsTouristDestination,data=RoomRent1.df, main="Room rent vs. IsTouristDestination", ylab="IsTouristDestination(1) or not(0)", xlab="Room Rent in rupees ", col=c("red","blue","green","yellow"),horizontal=TRUE)



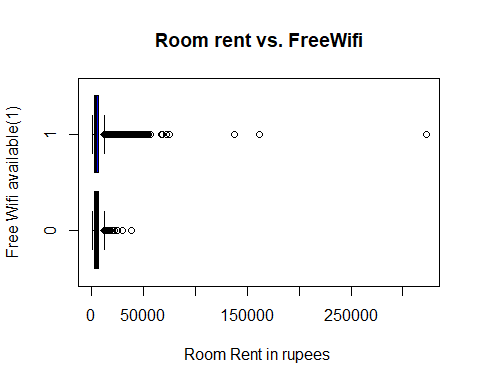
#Analyzing FreeWifi Vs RoomRent  
fw<-table( FreeWifi)  
barplot(fw, main="Borplot of FreeWifi",xlab= "FreeWifi" ,col="red")



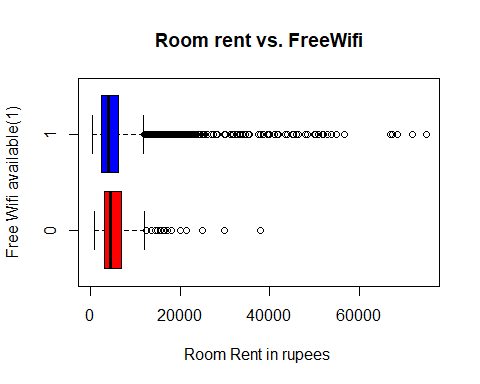
#Effect of FreeWifi on RoomRent  
fw = aggregate(RoomRent ~ FreeWifi, data = hotel.df, mean)  
fw

## FreeWifi RoomRent  
## 1 0 5402.039  
## 2 1 5481.518

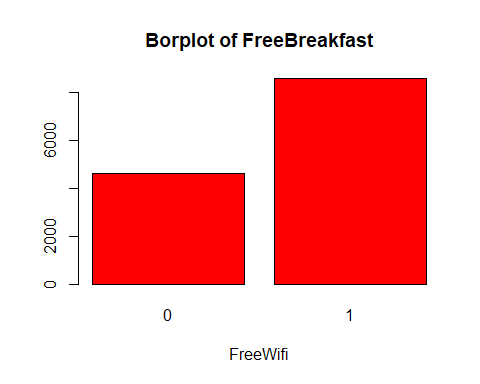
##With extreme outliers of roomrent  
boxplot(RoomRent~FreeWifi,data=hotel.df, main="Room rent vs. FreeWifi", ylab="Free Wifi available(1)", xlab="Room Rent in rupees ", col=c("red","blue","green","yellow"),horizontal=TRUE)



##Without extreme outliers of roomrent  
boxplot(RoomRent~FreeWifi,data=RoomRent1.df, main="Room rent vs. FreeWifi", ylab="Free Wifi available(1)", xlab="Room Rent in rupees ", col=c("red","blue","green","yellow"),horizontal=TRUE)



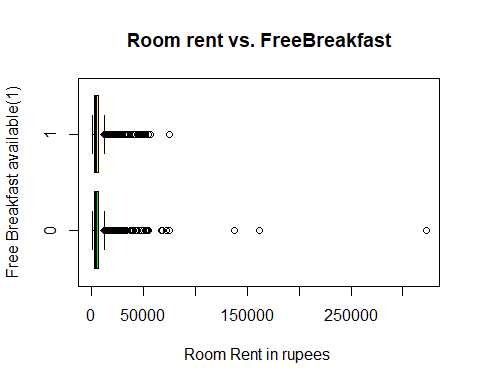
#Analyzing FreeBreakfast Vs RoomRent  
fw<-table( FreeBreakfast)  
barplot(fw, main="Borplot of FreeBreakfast",xlab= "FreeWifi" ,col="red")



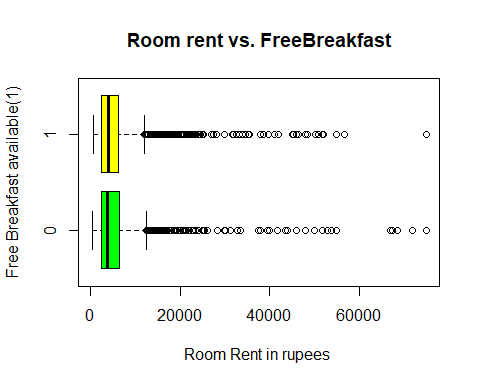
#Effect of FreeBreakfast on RoomRent  
fb= aggregate(RoomRent ~ FreeBreakfast, data =RoomRent1.df, mean)  
fb

## FreeBreakfast RoomRent  
## 1 0 5345.824  
## 2 1 5420.044

##With extreme outliers of roomrent  
boxplot(RoomRent~FreeBreakfast,data=hotel.df, main="Room rent vs. FreeBreakfast", ylab="Free Breakfast available(1)", xlab="Room Rent in rupees ", col=c("green","yellow"),horizontal=TRUE)



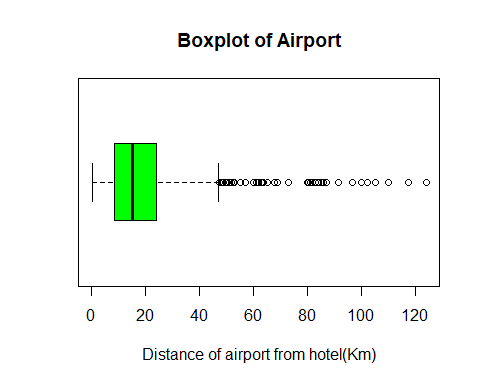
##Without extreme outliers of roomrent  
boxplot(RoomRent~FreeBreakfast,data=RoomRent1.df, main="Room rent vs. FreeBreakfast", ylab="Free Breakfast available(1)", xlab="Room Rent in rupees ", col=c("green","yellow"),horizontal=TRUE)



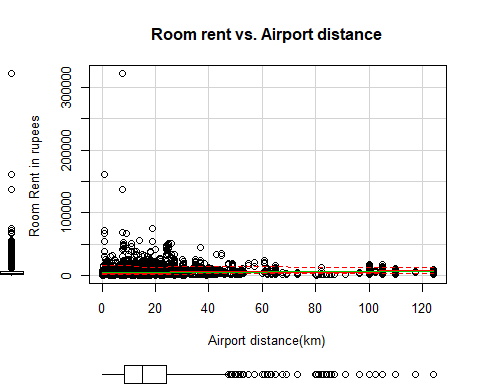
#Analyzing Airport distance from hotel effects in what way on RoomRent  
summary(Airport)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.20 8.40 15.00 21.16 24.00 124.00

boxplot( Airport, main="Boxplot of Airport",xlab= "Distance of airport from hotel(Km)" ,col="green",horizontal = TRUE)



#Effect of Airport distance on RoomRent  
scatterplot( Airport, RoomRent, main="Room rent vs. Airport distance", xlab="Airport distance(km)", ylab="Room Rent in rupees ",cex=1.1)



##Hypothesis  
 #1.Average RoomRent in hotels having swimming pool is more than that which don't have.  
 t.test(RoomRent~HasSwimmingPool, alternative="less")

##   
## Welch Two Sample t-test  
##   
## data: RoomRent by HasSwimmingPool  
## t = -29.007, df = 5011.8, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -4501.796  
## sample estimates:  
## mean in group 0 mean in group 1   
## 3776.576 8549.052

#2.Average RoomRent in hotels with high star rating is high as compared to one which has less star rating.  
 t.test( RoomRent, StarRating)

##   
## Welch Two Sample t-test  
##   
## data: RoomRent and StarRating  
## t = 85.791, df = 13223, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 5347.182 5597.238  
## sample estimates:  
## mean of x mean of y   
## 5475.669994 3.459815

#3.Average RoomRent in hotels providing Free Breakfast is more than that which don't provide.  
 t.test(RoomRent~FreeBreakfast, alternative="less")

##   
## Welch Two Sample t-test  
##   
## data: RoomRent by FreeBreakfast  
## t = 1.0112, df = 6196.5, p-value = 0.844  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf 416.9104  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5578.750 5420.044

#4.Average RoomRent in metro cities hotels is more than that of non metro cities.  
 t.test(RoomRent~IsMetroCity, alternative="less")

##   
## Welch Two Sample t-test  
##   
## data: RoomRent by IsMetroCity  
## t = 10.674, df = 13218, p-value = 1  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf 1249.286  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5782.794 4700.329

#5.Average RoomRent in hotels having more hotel capacity is more compared to one with less capacity.  
 t.test( RoomRent, HotelCapacity)

##   
## Welch Two Sample t-test  
##   
## data: RoomRent and HotelCapacity  
## t = 84.861, df = 13226, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 5288.101 5538.171  
## sample estimates:  
## mean of x mean of y   
## 5475.66999 62.53373

#chi sq test  
 chisq.test(RoomRent,HotelCapacity)

## Warning in chisq.test(RoomRent, HotelCapacity): Chi-squared approximation  
## may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: RoomRent and HotelCapacity  
## X-squared = 1478600, df = 545220, p-value < 2.2e-16

chisq.test(RoomRent,StarRating)

## Warning in chisq.test(RoomRent, StarRating): Chi-squared approximation may  
## be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: RoomRent and StarRating  
## X-squared = 132220, df = 43100, p-value < 2.2e-16

chisq.test(RoomRent,HasSwimmingPool)

## Warning in chisq.test(RoomRent, HasSwimmingPool): Chi-squared approximation  
## may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: RoomRent and HasSwimmingPool  
## X-squared = 8387.9, df = 2155, p-value < 2.2e-16

#Generating a multiple linear regression model for RoomRent  
 #1.  
 fit1<-lm(RoomRent~StarRating+HasSwimmingPool+HotelCapacity-1, data = hotel.df)  
 summary(fit1)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity -   
## 1, data = hotel.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8039 -2448 -1251 461 312401   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## StarRating 1396.8552 26.1437 53.430 < 2e-16 \*\*\*  
## HasSwimmingPool 3719.7733 148.8394 24.992 < 2e-16 \*\*\*  
## HotelCapacity -7.6598 0.9417 -8.134 4.53e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6815 on 13221 degrees of freedom  
## Multiple R-squared: 0.4457, Adjusted R-squared: 0.4456   
## F-statistic: 3544 on 3 and 13221 DF, p-value: < 2.2e-16

confint(fit1)

## 2.5 % 97.5 %  
## StarRating 1345.60986 1448.100537  
## HasSwimmingPool 3428.02683 4011.519835  
## HotelCapacity -9.50575 -5.813876

#Coefficents of the model  
 fit1$coefficients

## StarRating HasSwimmingPool HotelCapacity   
## 1396.855201 3719.773331 -7.659813

fit2<-lm(RoomRent~StarRating+HasSwimmingPool+HotelCapacity+IsWeekend+IsTouristDestination-1, data = hotel.df)  
 summary(fit2)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity +   
## IsWeekend + IsTouristDestination - 1, data = hotel.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8326 -2519 -1213 466 312479   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## StarRating 1258.8978 44.5202 28.277 < 2e-16 \*\*\*  
## HasSwimmingPool 3669.9530 148.8953 24.648 < 2e-16 \*\*\*  
## HotelCapacity -6.1759 0.9661 -6.393 1.68e-10 \*\*\*  
## IsWeekend -509.7103 119.2393 -4.275 1.93e-05 \*\*\*  
## IsTouristDestination 1053.6672 124.8271 8.441 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6794 on 13219 degrees of freedom  
## Multiple R-squared: 0.4493, Adjusted R-squared: 0.4491   
## F-statistic: 2157 on 5 and 13219 DF, p-value: < 2.2e-16

confint(fit2)

## 2.5 % 97.5 %  
## StarRating 1171.631760 1346.163815  
## HasSwimmingPool 3378.096887 3961.809159  
## HotelCapacity -8.069513 -4.282225  
## IsWeekend -743.436340 -275.984256  
## IsTouristDestination 808.988161 1298.346319

#Coefficents of the model  
 fit2$coefficients

## StarRating HasSwimmingPool HotelCapacity   
## 1258.897787 3669.953023 -6.175869   
## IsWeekend IsTouristDestination   
## -509.710298 1053.667240

#3.  
 fit3<-lm(RoomRent~StarRating+HasSwimmingPool+HotelCapacity+Airport-1, data = hotel.df)  
 summary(fit3)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity +   
## Airport - 1, data = hotel.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8240 -2382 -1225 384 312742   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## StarRating 1248.4242 33.2338 37.565 < 2e-16 \*\*\*  
## HasSwimmingPool 3903.7508 150.7277 25.899 < 2e-16 \*\*\*  
## HotelCapacity -6.7434 0.9485 -7.110 1.22e-12 \*\*\*  
## Airport 18.8696 2.6165 7.212 5.83e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6802 on 13220 degrees of freedom  
## Multiple R-squared: 0.4479, Adjusted R-squared: 0.4477   
## F-statistic: 2681 on 4 and 13220 DF, p-value: < 2.2e-16

confint(fit3)

## 2.5 % 97.5 %  
## StarRating 1183.281183 1313.56715  
## HasSwimmingPool 3608.302992 4199.19868  
## HotelCapacity -8.602508 -4.88421  
## Airport 13.740834 23.99840

#Coefficents of the model  
 fit3$coefficients

## StarRating HasSwimmingPool HotelCapacity Airport   
## 1248.424165 3903.750835 -6.743359 18.869619

#4.  
 fit4<-lm(RoomRent~StarRating+HasSwimmingPool+HotelCapacity+FreeBreakfast+Airport+Population+FreeWifi-1, data = hotel.df)  
 summary(fit4)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity +   
## FreeBreakfast + Airport + Population + FreeWifi - 1, data = hotel.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8980 -2443 -1022 689 311696   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## StarRating 2.006e+03 6.958e+01 28.825 < 2e-16 \*\*\*  
## HasSwimmingPool 3.171e+03 1.562e+02 20.294 < 2e-16 \*\*\*  
## HotelCapacity -6.179e+00 9.938e-01 -6.218 5.18e-10 \*\*\*  
## FreeBreakfast -1.925e+02 1.245e+02 -1.546 0.122038   
## Airport 1.014e+01 2.670e+00 3.796 0.000147 \*\*\*  
## Population -1.932e-04 1.490e-05 -12.966 < 2e-16 \*\*\*  
## FreeWifi -1.396e+03 2.030e+02 -6.875 6.47e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6735 on 13217 degrees of freedom  
## Multiple R-squared: 0.4589, Adjusted R-squared: 0.4586   
## F-statistic: 1601 on 7 and 13217 DF, p-value: < 2.2e-16

confint(fit4)

## 2.5 % 97.5 %  
## StarRating 1.869276e+03 2.142053e+03  
## HasSwimmingPool 2.864295e+03 3.476753e+03  
## HotelCapacity -8.127409e+00 -4.231542e+00  
## FreeBreakfast -4.365039e+02 5.150750e+01  
## Airport 4.903133e+00 1.537075e+01  
## Population -2.224027e-04 -1.639883e-04  
## FreeWifi -1.793928e+03 -9.979496e+02

#Coefficents of the model  
 fit4$coefficients

## StarRating HasSwimmingPool HotelCapacity FreeBreakfast   
## 2.005665e+03 3.170524e+03 -6.179475e+00 -1.924982e+02   
## Airport Population FreeWifi   
## 1.013694e+01 -1.931955e-04 -1.395939e+03

#detaching the numeric\_data set  
 detach(numeric\_data)