Hotel\_pricing In Indian Market

tavva\_prudhvith

January 26, 2018

The purpose of this project is to analyse the pricing strategy of hotels in the Indian hotel industry. Many factors drive hotel room prices. The objective of this project is to identify the factors that matter the most. Think about the following problem: Room Rent = FUNCTION ( Date(s); Hotel Features; External Factors)

## Sumarize the data.

hotel<-read.csv(file.choose(),header=T)  
summary(hotel)

## CityName Population CityRank IsMetroCity   
## Delhi :2048 Min. : 8096 Min. : 0.00 Min. :0.0000   
## Jaipur : 768 1st Qu.: 744983 1st Qu.: 2.00 1st Qu.:0.0000   
## Mumbai : 712 Median : 3046163 Median : 9.00 Median :0.0000   
## Bangalore: 656 Mean : 4416837 Mean :14.83 Mean :0.2842   
## Goa : 624 3rd Qu.: 8443675 3rd Qu.:24.00 3rd Qu.:1.0000   
## Kochi : 608 Max. :12442373 Max. :44.00 Max. :1.0000   
## (Other) :7816   
## IsTouristDestination IsWeekend IsNewYearEve Date   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Dec 21 2016:1611   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 Dec 24 2016:1611   
## Median :1.0000 Median :1.0000 Median :0.0000 Dec 25 2016:1611   
## Mean :0.6972 Mean :0.6228 Mean :0.1244 Dec 28 2016:1611   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 Dec 31 2016:1611   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Dec 18 2016:1608   
## (Other) :3569   
## HotelName RoomRent StarRating   
## Vivanta by Taj : 32 Min. : 299 Min. :0.000   
## Goldfinch Hotel : 24 1st Qu.: 2436 1st Qu.:3.000   
## OYO Rooms : 24 Median : 4000 Median :3.000   
## The Gordon House Hotel: 24 Mean : 5474 Mean :3.459   
## Apnayt Villa : 16 3rd Qu.: 6299 3rd Qu.:4.000   
## Bentleys Hotel Colaba : 16 Max. :322500 Max. :5.000   
## (Other) :13096   
## Airport   
## Min. : 0.20   
## 1st Qu.: 8.40   
## Median : 15.00   
## Mean : 21.16   
## 3rd Qu.: 24.00   
## 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   
##

library(psych)

## Warning: package 'psych' was built under R version 3.4.3

describe(hotel)

## vars n mean sd median trimmed  
## CityName\* 1 13232 18.07 11.72 16 17.29  
## Population 2 13232 4416836.87 4258386.00 3046163 4040816.22  
## CityRank 3 13232 14.83 13.51 9 13.30  
## IsMetroCity 4 13232 0.28 0.45 0 0.23  
## IsTouristDestination 5 13232 0.70 0.46 1 0.75  
## IsWeekend 6 13232 0.62 0.48 1 0.65  
## IsNewYearEve 7 13232 0.12 0.33 0 0.03  
## Date\* 8 13232 14.30 2.69 14 14.39  
## HotelName\* 9 13232 841.19 488.16 827 841.18  
## RoomRent 10 13232 5473.99 7333.12 4000 4383.33  
## StarRating 11 13232 3.46 0.76 3 3.40  
## Airport 12 13232 21.16 22.76 15 16.39  
## HotelAddress\* 13 13232 1202.53 582.17 1261 1233.25  
## HotelPincode 14 13232 397430.26 259837.50 395003 388540.47  
## HotelDescription\* 15 13224 581.34 363.26 567 575.37  
## FreeWifi 16 13232 0.93 0.26 1 1.00  
## FreeBreakfast 17 13232 0.65 0.48 1 0.69  
## HotelCapacity 18 13232 62.51 76.66 34 46.03  
## HasSwimmingPool 19 13232 0.36 0.48 0 0.32  
## mad min max range skew  
## 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 20 19.0 -0.77  
## 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  
## 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\* 1.92 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

#Apply Some Functions  
aggregate(hotel$RoomRent, by=list(freewifi = hotel$FreeWifi, freeBreakfast = hotel$FreeBreakfast, swimmingPool = hotel$HasSwimmingPool,Touristplace=hotel$IsTouristDestination), mean)

## freewifi freeBreakfast swimmingPool Touristplace x  
## 1 0 0 0 0 2961.188  
## 2 1 0 0 0 2788.815  
## 3 0 1 0 0 4108.600  
## 4 1 1 0 0 3286.387  
## 5 0 0 1 0 7185.167  
## 6 1 0 1 0 6675.889  
## 7 0 1 1 0 5079.913  
## 8 1 1 1 0 5517.907  
## 9 0 0 0 1 3677.939  
## 10 1 0 0 1 3257.897  
## 11 0 1 0 1 5909.478  
## 12 1 1 0 1 4395.936  
## 13 0 0 1 1 7446.194  
## 14 1 0 1 1 10818.984  
## 15 0 1 1 1 5240.216  
## 16 1 1 1 1 9215.152

dim(hotel)

## [1] 13232 19

## One-Way Contingency Tables for categorial variables

table1<-with(hotel,table(FreeWifi))  
table1

## FreeWifi  
## 0 1   
## 981 12251

prop.table(table1)\*100 # Percentage For Free-Wifi

## FreeWifi  
## 0 1   
## 7.413845 92.586155

table2<-with(hotel,table(HasSwimmingPool))  
table2

## HasSwimmingPool  
## 0 1   
## 8524 4708

prop.table(table2)\*100 #Percentage for Hotels which have swimming pools.

## HasSwimmingPool  
## 0 1   
## 64.41959 35.58041

table3<-with(hotel,table(FreeBreakfast))  
table3

## FreeBreakfast  
## 0 1   
## 4643 8589

prop.table(table3)\*100 # Percentage For Free Breakfast

## FreeBreakfast  
## 0 1   
## 35.08918 64.91082

table4<-with(hotel,table(IsTouristDestination))  
table4

## IsTouristDestination  
## 0 1   
## 4007 9225

prop.table(table4)\*100 # Percentage For Hotels which is a Tourist Destination.

## IsTouristDestination  
## 0 1   
## 30.28265 69.71735

# Two-Way Contingency tables for categorial variables.

table5<-xtabs(~hotel$IsTouristDestination + hotel$IsWeekend)  
table5

## hotel$IsWeekend  
## hotel$IsTouristDestination 0 1  
## 0 1454 2553  
## 1 3537 5688

table6<-xtabs(~hotel$StarRating + hotel$IsMetroCity)  
table6

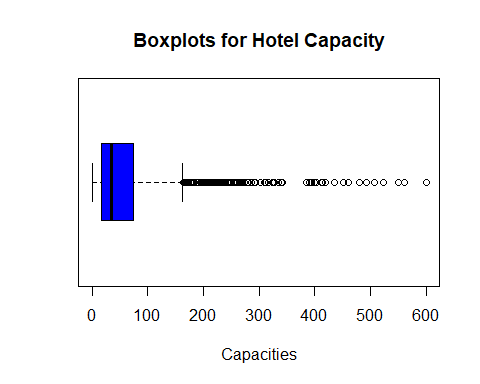
## hotel$IsMetroCity  
## hotel$StarRating 0 1  
## 0 16 0  
## 1 8 0  
## 2 344 96  
## 2.5 456 176  
## 3 4336 1617  
## 3.2 8 0  
## 3.3 16 0  
## 3.4 8 0  
## 3.5 1312 440  
## 3.6 0 8  
## 3.7 24 0  
## 3.8 16 0  
## 3.9 32 0  
## 4 1696 767  
## 4.1 24 0  
## 4.3 16 0  
## 4.4 8 0  
## 4.5 288 88  
## 4.7 8 0  
## 4.8 16 0  
## 5 840 568

table7<-xtabs(~hotel$StarRating + hotel$IsTouristDestination)  
table7

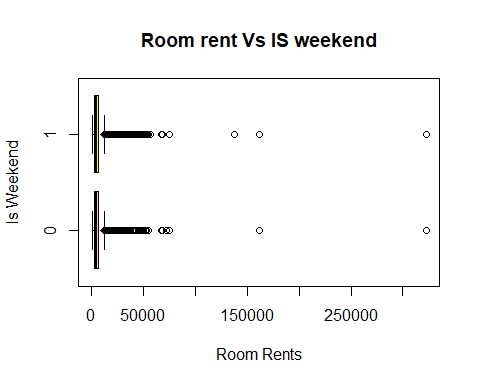
## hotel$IsTouristDestination  
## hotel$StarRating 0 1  
## 0 0 16  
## 1 0 8  
## 2 64 376  
## 2.5 152 480  
## 3 1888 4065  
## 3.2 0 8  
## 3.3 0 16  
## 3.4 0 8  
## 3.5 448 1304  
## 3.6 0 8  
## 3.7 8 16  
## 3.8 16 0  
## 3.9 16 16  
## 4 839 1624  
## 4.1 0 24  
## 4.3 16 0  
## 4.4 8 0  
## 4.5 128 248  
## 4.7 8 0  
## 4.8 0 16  
## 5 416 992

## Boxplots For the variables present in

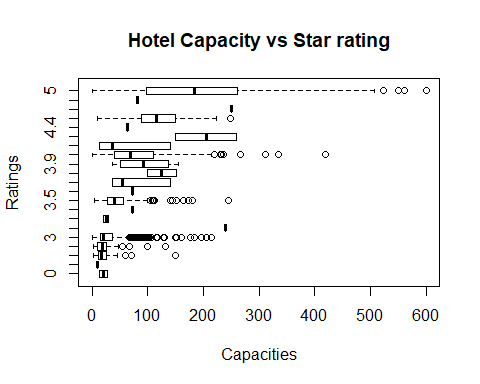
# Boxplot For Capacity.  
boxplot(hotel$HotelCapacity,  
 main="Boxplots for Hotel Capacity",  
 col="blue",  
 horizontal = T,  
 xlab="Capacities")



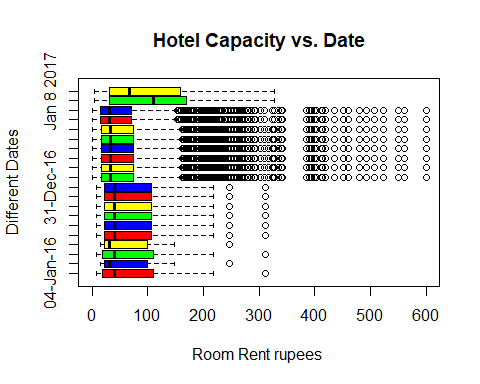
#Boxplot for Room rent and Weekends.  
boxplot(hotel$RoomRent ~ hotel$IsWeekend,  
 horizontal=T,  
 main="Room rent Vs IS weekend",  
 xlab="Room Rents",  
 ylab="Is Weekend",  
col=c("pink","yellow"))



#Boxplots For Capacity and rating OF th Hotels.  
boxplot(hotel$HotelCapacity ~ hotel$StarRating,  
 horizontal=T,  
 main="Hotel Capacity vs Star rating",  
 xlab="Capacities",  
 ylab="Ratings")

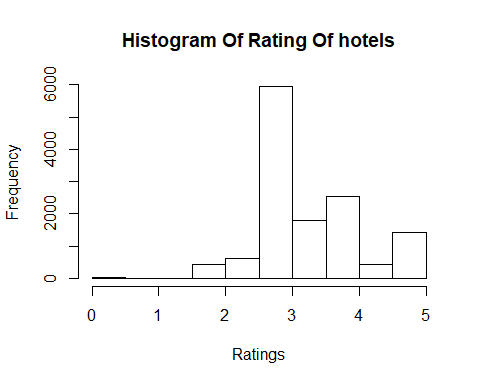


#boxplots FOr Hotel Capacity and Date.  
boxplot(hotel$HotelCapacity~hotel$Date,  
 main="Hotel Capacity vs. Date",  
 ylab="Different Dates",   
 xlab="Room Rent rupees ",  
 col=c("red","blue","green","yellow"),  
 horizontal=TRUE)

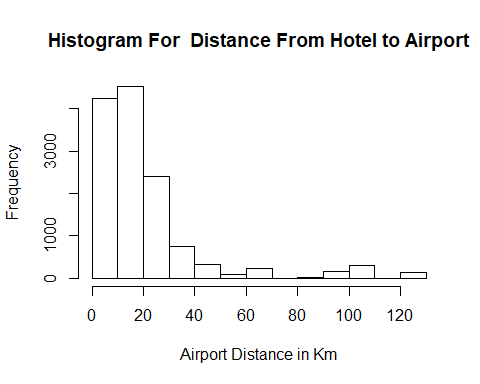


## Histograms Of Suitable Data

# Histogram Of ratings For Hotels  
hist(hotel$StarRating,  
 main="Histogram Of Rating Of hotels",  
 xlab="Ratings",  
 breaks=10)



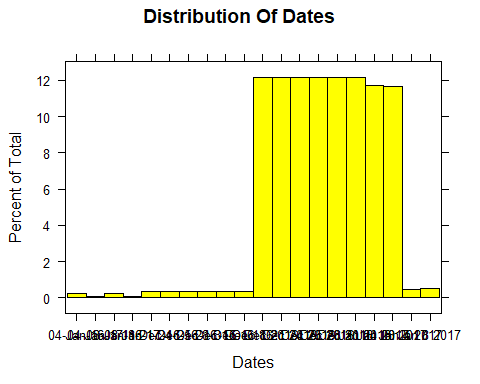
# Histograms for Distance Between Hotel to Airport.  
hist(hotel$Airport,  
 main="Histogram For Distance From Hotel to Airport",  
 xlab="Airport Distance in Km")



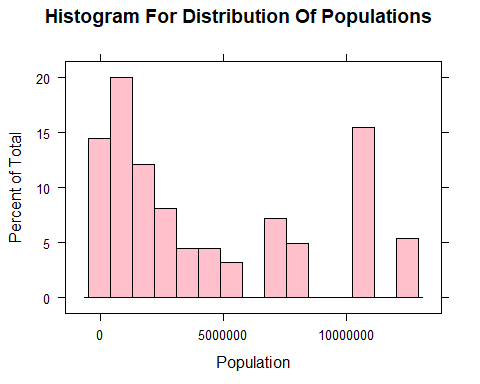
# Using Histogram command using lattice package to find the distributions of Date variable.  
table(hotel$Date)

##   
## 04-Jan-16 04-Jan-17 08-Jan-16 08-Jan-17 18-Dec-16 21-Dec-16   
## 31 13 31 13 44 44   
## 24-Dec-16 25-Dec-16 28-Dec-16 31-Dec-16 Dec 18 2016 Dec 21 2016   
## 44 44 44 44 1608 1611   
## Dec 24 2016 Dec 25 2016 Dec 28 2016 Dec 31 2016 Jan 04 2017 Jan 08 2017   
## 1611 1611 1611 1611 1548 1542   
## Jan 4 2017 Jan 8 2017   
## 60 67

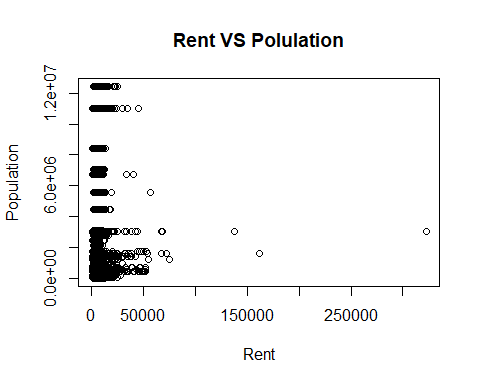
library(lattice)  
histogram(~Date,data=hotel,  
 main="Distribution Of Dates",  
 xlab="Dates",  
 col="yellow")



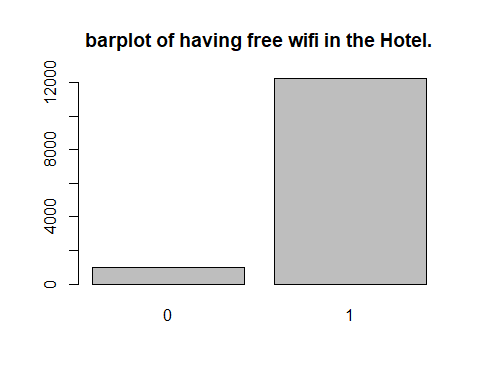
# Using Histogram command using lattice package to find the distributions of population variable.  
histogram(~Population,  
 data=hotel,  
 main="Histogram For Distribution Of Populations",  
 col="pink")



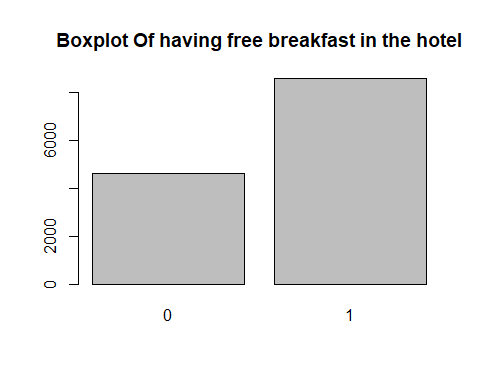
plot(hotel$RoomRent,  
 hotel$Population,  
 main="Rent VS Polulation",  
 xlab="Rent",  
 ylab="Population")



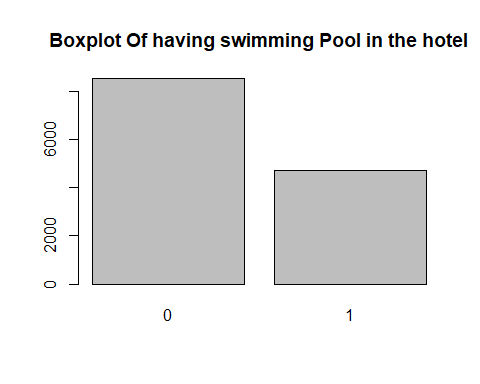
# Including the ones which have free wifi in hotels.  
freewifi<-table(hotel$FreeWifi)  
barplot(freewifi,  
 main="barplot of having free wifi in the Hotel.")



#Including the ones which have free breakfast in the hotels.  
breakfast<-table(hotel$FreeBreakfast)  
barplot(breakfast,  
 main="Boxplot Of having free breakfast in the hotel")



#Including the ones which have swimming pool in the hotels.  
pool<-table(hotel$HasSwimmingPool)  
barplot(pool,  
 main="Boxplot Of having swimming Pool in the hotel")

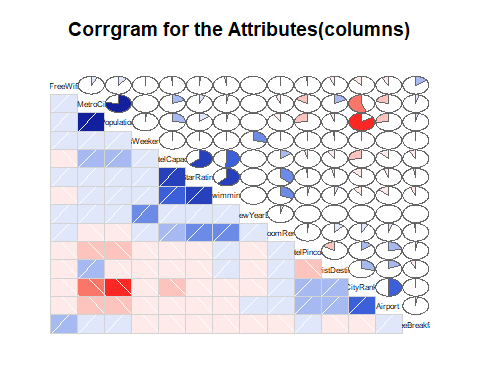


Conclusion For the Above Plots and Barplots: 1.freewifi-According to the graph,there are more hotels having free wifi than the ones who doesn’t have wifi. 2.breakfast-According to the graph,there are more hotels who give free breakfast than the ones who doesn’t give breakfast. 3.pool-According to the graph,there are less number of hotels who have a swimming pool in their hotel.

library(corrgram)

## Warning: package 'corrgram' was built under R version 3.4.3

corrgram(hotel,main="Corrgram for the Attributes(columns)",  
 lower.panel = panel.shade,  
 upper.panel=panel.pie,  
 order=T,  
 text.panel = panel.txt)



#Let us find the Correlation Matrix For Rent,Pincode,airport and rating.  
  
x<-hotel[,c("RoomRent","StarRating","Airport","FreeWifi")]  
   
y<-hotel[,c("RoomRent","StarRating","Airport","FreeWifi")]  
 cor(x,y)

## RoomRent StarRating Airport FreeWifi  
## RoomRent 1.000000000 0.36937343 0.04965324 0.003627002  
## StarRating 0.369373425 1.00000000 -0.06091918 0.018009594  
## Airport 0.049653244 -0.06091918 1.00000000 -0.094523677  
## FreeWifi 0.003627002 0.01800959 -0.09452368 1.000000000

cov(x,y)

## RoomRent StarRating Airport FreeWifi  
## RoomRent 5.377460e+07 2.048375e+03 8287.1785840 6.968633743  
## StarRating 2.048375e+03 5.718875e-01 -1.0485276 0.003568375  
## Airport 8.287179e+03 -1.048528e+00 518.0133284 -0.563666554  
## FreeWifi 6.968634e+00 3.568375e-03 -0.5636666 0.068647130

var(x,y)

## RoomRent StarRating Airport FreeWifi  
## RoomRent 5.377460e+07 2.048375e+03 8287.1785840 6.968633743  
## StarRating 2.048375e+03 5.718875e-01 -1.0485276 0.003568375  
## Airport 8.287179e+03 -1.048528e+00 518.0133284 -0.563666554  
## FreeWifi 6.968634e+00 3.568375e-03 -0.5636666 0.068647130

## Effect Of Star-Rating On Room rent.

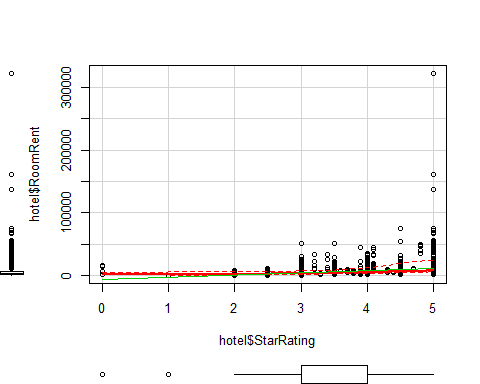
library(car)

## Warning: package 'car' was built under R version 3.4.3

##   
## Attaching package: 'car'

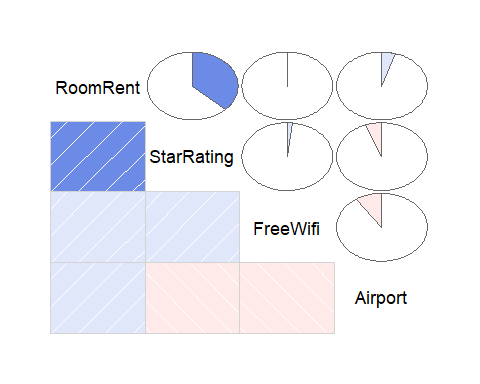
## The following object is masked from 'package:psych':  
##   
## logit

scatterplot(hotel$StarRating,hotel$RoomRent)



## Now,let us visualize the above correlation matrix with corrgram.

library(corrgram)  
 corrgram(hotel[,c("RoomRent","StarRating","Airport","FreeWifi")],  
 order=T,  
 lower.panel=panel.shade,  
 upper.panel=panel.pie,  
 text.panel=panel.txt)



Test1:- Hypothesis:There is no significant difference between the Room Rent of Hotels providing free wifi and those which do not

t.test(hotel$RoomRent ~ hotel$FreeWifi)

##   
## Welch Two Sample t-test  
##   
## data: hotel$RoomRent by hotel$FreeWifi  
## t = -0.76847, df = 1804.7, p-value = 0.4423  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -360.5977 157.5701  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5380.004 5481.518

Result:As p-value is greater than 0.05 it suggests that there is no significant difference between the means of our Room rent and Free -Wifi.So,we could not reject null hypothesis.

Test2:- Hypothesis:average Room rent hotels coming from tourist is more than people coming from cities.

t.test(hotel$RoomRent,hotel$IsTouristDestination)

##   
## Welch Two Sample t-test  
##   
## data: hotel$RoomRent and hotel$IsTouristDestination  
## t = 85.856, df = 13231, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 5348.337 5598.253  
## sample estimates:  
## mean of x mean of y   
## 5473.9918380 0.6971735

Result:As p-value is 2.2e-16 which is less than 0.05 suggests a significant difference between Room rent and Istouristdestination.So,We can reject Null Hypothesis.

Test3:- Hypothesis:Average Room rent hotels having free breakfast is more than the hotels which doen’t have any breakfast.

t.test(hotel$RoomRent~hotel$FreeBreakfast)

##   
## Welch Two Sample t-test  
##   
## data: hotel$RoomRent by hotel$FreeBreakfast  
## t = 0.98095, df = 6212.3, p-value = 0.3267  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -153.5017 460.9935  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5573.790 5420.044

Result:As p-value is greater than 0.05 it suggests that there is no significant difference between the means of our Room rent and free brerakfast.So,we could not reject null hypothesis.

Test4:- Hypothesis:The Room Rents of Hotels on normal Eve are cheaper than that on New Year’s Eve.

t.test(RoomRent~IsNewYearEve,data=hotel)

##   
## Welch Two Sample t-test  
##   
## data: RoomRent by IsNewYearEve  
## t = -4.1793, df = 2065, p-value = 3.046e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1256.5297 -453.9099  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5367.606 6222.826

result: As p-value is <0.05 ,it suggests that there is a significance differenc ebetween room rent hotels on normal eve and newyear eve.

Test5:- Hypothesis:Room rent of hotels in non-metro cities are more expensive than hotels in metro cities.

t.test(RoomRent~IsMetroCity,data=hotel)

##   
## Welch Two Sample t-test  
##   
## data: RoomRent by IsMetroCity  
## t = 10.721, df = 13224, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 888.0308 1285.4102  
## sample estimates:  
## mean in group 0 mean in group 1   
## 5782.794 4696.073

result: As p-value is <0.05 ,it suggests that there is a significance differenc ebetween room rent hotels in metro cities and non-metro cities.

Test6:- Hypothesis:Room rent of hotels having swimming pools are more expensive than hotels not having it.

t.test(RoomRent~HasSwimmingPool,data=hotel)

##   
## Welch Two Sample t-test  
##   
## data: RoomRent by HasSwimmingPool  
## t = -29.013, df = 5011.3, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5096.030 -4450.942  
## sample estimates:  
## mean in group 0 mean in group 1   
## 3775.566 8549.052

Result:As p-value is <0.05 ,it suggests that there is a significance differenc ebetween room rent hotels having swimming pool and hotels that doesn’t have swimming pool.

## Regression Models

## Regression Analysis(1)

reg3<-lm(RoomRent~ StarRating +IsMetroCity +FreeBreakfast,data=hotel)  
summary(reg3)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + IsMetroCity + FreeBreakfast,   
## data = hotel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9075 -2438 -968 887 311018   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6802.00 290.16 -23.442 <2e-16 \*\*\*  
## StarRating 3656.73 78.22 46.750 <2e-16 \*\*\*  
## IsMetroCity -1568.66 131.25 -11.952 <2e-16 \*\*\*  
## FreeBreakfast 113.01 123.73 0.913 0.361   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6779 on 13228 degrees of freedom  
## Multiple R-squared: 0.1457, Adjusted R-squared: 0.1455   
## F-statistic: 751.8 on 3 and 13228 DF, p-value: < 2.2e-16

reg3$coefficients

## (Intercept) StarRating IsMetroCity FreeBreakfast   
## -6801.997 3656.730 -1568.661 113.012

As Y=F(x) y=Room Rent x1,x2,x3,… are exploratory variables dependnt upon the room rent i.e the factors. Room rent =b0 +b1 *Starrating + b2*ismetrocity +b3*freebreakfast From the fitted coefficients,then b0=-1 (let us assume) b1=3656.730 b2=-1568.661 b3=113.012 The Conclusion will be Room rent =(-1) +3517.46* Starrating + (-1568.661)*ismetrocity +113.012*freebrreakfast

# Regression Analysis (2)

reg1<-lm(RoomRent~StarRating + Airport + HasSwimmingPool + HotelCapacity + FreeWifi,data=hotel)  
summary(reg1)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + Airport + HasSwimmingPool +   
## HotelCapacity + FreeWifi, data = hotel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10799 -2266 -877 967 310428   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7473.711 394.284 -18.955 <2e-16 \*\*\*  
## StarRating 3517.457 111.686 31.494 <2e-16 \*\*\*  
## Airport 25.595 2.603 9.831 <2e-16 \*\*\*  
## HasSwimmingPool 2716.242 158.616 17.125 <2e-16 \*\*\*  
## HotelCapacity -14.751 1.006 -14.665 <2e-16 \*\*\*  
## FreeWifi 210.761 223.335 0.944 0.345   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6687 on 13226 degrees of freedom  
## Multiple R-squared: 0.1689, Adjusted R-squared: 0.1686   
## F-statistic: 537.5 on 5 and 13226 DF, p-value: < 2.2e-16

reg1$coefficients

## (Intercept) StarRating Airport HasSwimmingPool   
## -7473.71059 3517.45652 25.59515 2716.24231   
## HotelCapacity FreeWifi   
## -14.75071 210.76122

As Y=F(x) y=Room Rent x1,x2,x3,… are exploratory variables dependnt upon the room rent i.e the factors. Room rent =b0 +b1 *Starrating + b2*airport +b3*hasswimmingpool +b4*hotelcapacity From the fitted coefficients,then b0=-1 (let us assume) b1=3517.45652 b2=25.59515 b3=2716.24231 b4=-14.75071 The Conclusion will be Room rent =(-1) +3517.46 *Starrating + 25.6*airport +2716.24*hasswimmingpool +(-14.75)*hotelcapacity

## Regression Analysis (3)

reg2<-lm(RoomRent ~ StarRating + Airport + HotelCapacity + IsWeekend + HasSwimmingPool,data=hotel)  
summary(reg2)

##   
## Call:  
## lm(formula = RoomRent ~ StarRating + Airport + HotelCapacity +   
## IsWeekend + HasSwimmingPool, data = hotel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10800 -2267 -880 980 310462   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7312.691 349.532 -20.921 <2e-16 \*\*\*  
## StarRating 3522.885 111.535 31.585 <2e-16 \*\*\*  
## Airport 25.346 2.590 9.786 <2e-16 \*\*\*  
## HotelCapacity -14.777 1.006 -14.695 <2e-16 \*\*\*  
## IsWeekend 40.177 119.938 0.335 0.738   
## HasSwimmingPool 2708.404 158.403 17.098 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6687 on 13226 degrees of freedom  
## Multiple R-squared: 0.1688, Adjusted R-squared: 0.1685   
## F-statistic: 537.3 on 5 and 13226 DF, p-value: < 2.2e-16

reg2$coefficients

## (Intercept) StarRating Airport HotelCapacity   
## -7312.69132 3522.88515 25.34555 -14.77651   
## IsWeekend HasSwimmingPool   
## 40.17697 2708.40430

As Y=F(x) y=Room Rent x1,x2,x3,… are exploratory variables dependnt upon the room rent i.e the factors. Room rent =b0 +b1 *Starrating + b2*airport +b3*hotelcapacity +b4*isweekend +b5*hasswimmingpool From the fitted coefficients,then b0=-1 (let us assume) b1=3522.88515 b2=25.34555 b3=-14.77651 b4=40.17697 b5=2708.40430 The Conclusion will be Room rent =(-1) +3522.89* Starrating + 25.34*airport + (-14.8)*hotelcapacity + 40.17*isweekend + 2708.4*hasswimmingpool.