**Project Output : Hotel Rate affecting factor analysis.**

> hotel.df< -Cities42

> str(hotel.df)

Classes ‘tbl\_df’, ‘tbl’ and 'data.frame': 13232 obs. of 19 variables:

$ CityName : chr "Mumbai" "Mumbai" "Mumbai" "Mumbai" ...

$ Population : int 12442373 12442373 12442373 12442373 12442373 12442373 12442373 12442373 12442373 12442373 ...

$ CityRank : int 0 0 0 0 0 0 0 0 0 0 ...

$ IsMetroCity : int 1 1 1 1 1 1 1 1 1 1 ...

$ IsTouristDestination: int 1 1 1 1 1 1 1 1 1 1 ...

$ IsWeekend : int 1 0 1 1 0 1 0 1 1 0 ...

$ IsNewYearEve : int 0 0 0 0 0 1 0 0 0 0 ...

$ Date : chr "Dec 18 2016" "Dec 21 2016" "Dec 24 2016" "Dec 25 2016" ...

$ HotelName : chr "Vivanta by Taj" "Vivanta by Taj" "Vivanta by Taj" "Vivanta by Taj" ...

$ RoomRent : int 12375 10250 9900 10350 12000 11475 11220 9225 6800 9350 ...

$ StarRating : num 5 5 5 5 5 5 5 5 4 4 ...

$ Airport : num 21 21 21 21 21 21 21 21 20 20 ...

$ HotelAddress : chr "90 Cuffe Parade, Colaba, Mumbai, Maharashtra" "91 Cuffe Parade, Colaba, Mumbai, Maharashtra" "92 Cuffe Parade, Colaba, Mumbai, Maharashtra" "93 Cuffe Parade, Colaba, Mumbai, Maharashtra" ...

$ HotelPincode : int 400005 400006 400007 400008 400009 400010 400011 400012 400039 400040 ...

$ HotelDescription : chr "Luxury hotel with spa, near Gateway of India" "Luxury hotel with spa, near Gateway of India" "Luxury hotel with spa, near Gateway of India" "Luxury hotel with spa, near Gateway of India" ...

$ FreeWifi : int 1 1 1 1 1 1 1 1 1 1 ...

$ FreeBreakfast : int 0 0 0 0 0 0 0 0 1 1 ...

$ HotelCapacity : int 287 287 287 287 287 287 287 287 28 28 ...

$ HasSwimmingPool : int 1 1 1 1 1 1 1 1 0 0 ...

> describe(hotel.df)

vars n mean sd median trimmed mad min max range skew kurtosis

CityName\* 1 13232 NaN NA NA NaN NA Inf -Inf -Inf NA NA

Population 2 13232 4416836.87 4258386.00 3046163.0 4040816.22 3846498.95 8096.0 12442373 12434277.0 0.68 -1.08

CityRank 3 13232 14.83 13.51 9.0 13.30 11.86 0.0 44 44.0 0.69 -0.76

IsMetroCity 4 13232 0.28 0.45 0.0 0.23 0.00 0.0 1 1.0 0.96 -1.08

IsTouristDestination 5 13232 0.70 0.46 1.0 0.75 0.00 0.0 1 1.0 -0.86 -1.26

IsWeekend 6 13232 0.62 0.48 1.0 0.65 0.00 0.0 1 1.0 -0.51 -1.74

IsNewYearEve 7 13232 0.12 0.33 0.0 0.03 0.00 0.0 1 1.0 2.28 3.18

Date\* 8 13232 NaN NA NA NaN NA Inf -Inf -Inf NA NA

HotelName\* 9 13232 NaN NA NA NaN NA Inf -Inf -Inf NA NA

RoomRent 10 13232 5473.99 7333.12 4000.0 4383.33 2653.85 299.0 322500 322201.0 16.75 582.06

StarRating 11 13232 3.46 0.76 3.0 3.40 0.74 0.0 5 5.0 0.48 0.25

Airport 12 13232 21.16 22.76 15.0 16.39 11.12 0.2 124 123.8 2.73 7.89

HotelAddress\* 13 13232 NaN NA NA NaN NA Inf -Inf -Inf NA NA

HotelPincode 14 13232 397430.26 259837.50 395003.0 388540.47 257975.37 100025.0 7000157 6900132.0 9.99 249.76

HotelDescription\* 15 13224 3.46 0.48 3.5 3.41 0.74 3.0 5 2.0 0.93 0.78

FreeWifi 16 13232 0.93 0.26 1.0 1.00 0.00 0.0 1 1.0 -3.25 8.57

FreeBreakfast 17 13232 0.65 0.48 1.0 0.69 0.00 0.0 1 1.0 -0.62 -1.61

HotelCapacity 18 13232 62.51 76.66 34.0 46.03 28.17 0.0 600 600.0 2.95 11.39

HasSwimmingPool 19 13232 0.36 0.48 0.0 0.32 0.00 0.0 1 1.0 0.60 -1.64

se

CityName\* NA

Population 37019.65

CityRank 0.12

IsMetroCity 0.00

IsTouristDestination 0.00

IsWeekend 0.00

IsNewYearEve 0.00

Date\* NA

HotelName\* NA

RoomRent 63.75

StarRating 0.01

Airport 0.20

HotelAddress\* NA

HotelPincode 2258.86

HotelDescription\* 0.00

FreeWifi 0.00

FreeBreakfast 0.00

HotelCapacity 0.67

HasSwimmingPool 0.00

Test for the factors affecting the most to the hotel prices.

> cor(hotel.df$CityRank,hotel.df$RoomRent)

[1] 0.09398553

> cor(hotel.df$CityRank,hotel.df$IsNewYearEve)

[1] -0.0006326444

> cor(hotel.df$CityRank,hotel.df$Airport)

[1] 0.505912

> cor(hotel.df$CityRank,hotel.df$StarRating)

[1] -0.133381

> cor(hotel.df$CityRank,hotel.df$RoomRent)

[1] 0.09398553

> cor(hotel.df$RoomRent,hotel.df$IsMetroCity)

[1] -0.06683977

> cor(hotel.df$RoomRent,hotel.df$IsTouristDestination)

[1] 0.122503

> cor(hotel.df$RoomRent,hotel.df$IsWeekend)

[1] 0.004580134

> cor(hotel.df$RoomRent,hotel.df$HasSwimmingPool)

[1] 0.3116577

> cor(hotel.df$RoomRent,hotel.df$FreeBreakfast)

[1] -0.01000637

> cor(hotel.df$RoomRent,hotel.df$FreeWifi)

[1] 0.003627002

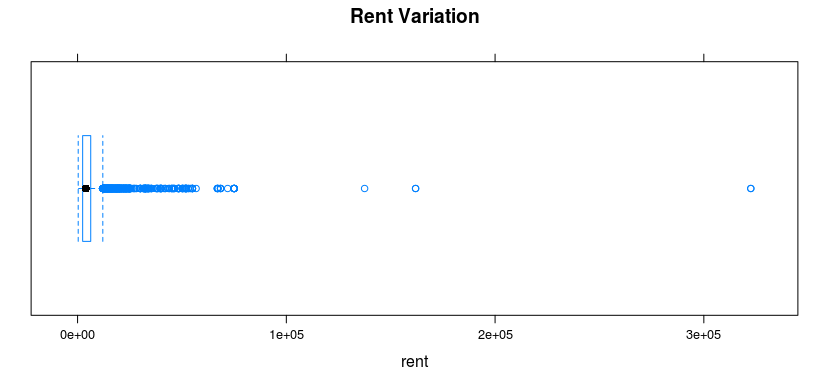
> cor(hotel.df$RoomRent,hotel.df$Population)

[1] -0.08872806

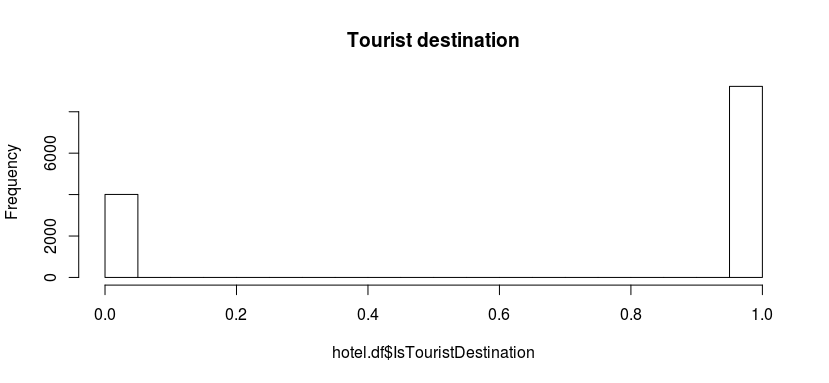
> cor(hotel.df$RoomRent,hotel.df$HotelCapacity)

[1] 0.1578733

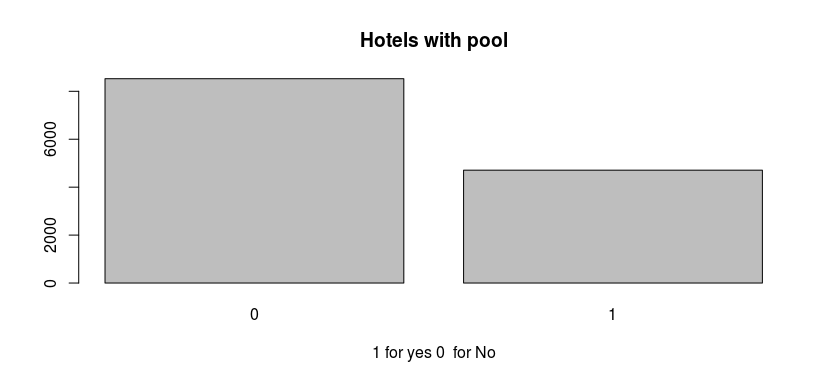
> bwplot(hotel.df$RoomRent,main="Rent Variation",xlab="rent")



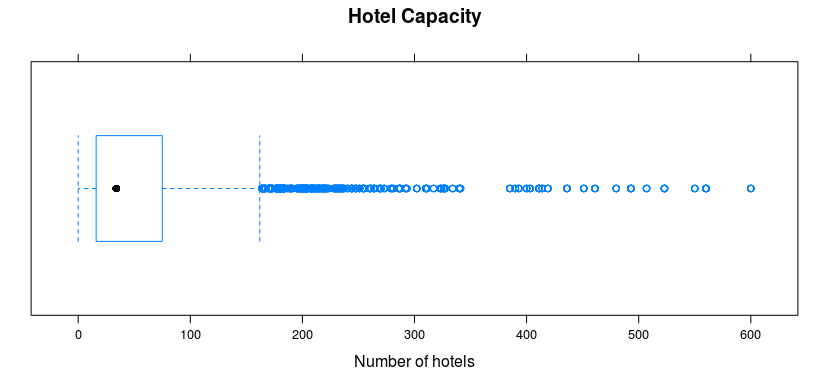
> hist(hotel.df$IsTouristDestination,main="Tourist destination")



> barplot(table(hotel.df$HasSwimmingPool),main="Hotels with pool",xlab = "1 for yes 0 for No")



> bwplot(hotel.df$HotelCapacity,main="Hotel Capacity",xlab="Number of hotels")

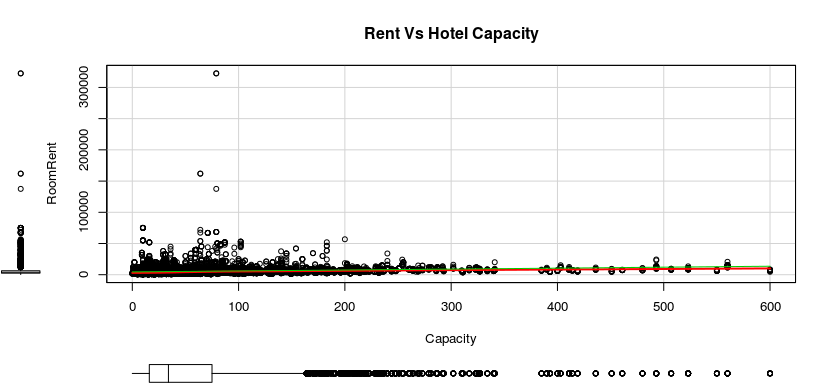


Pairwise varaition of components

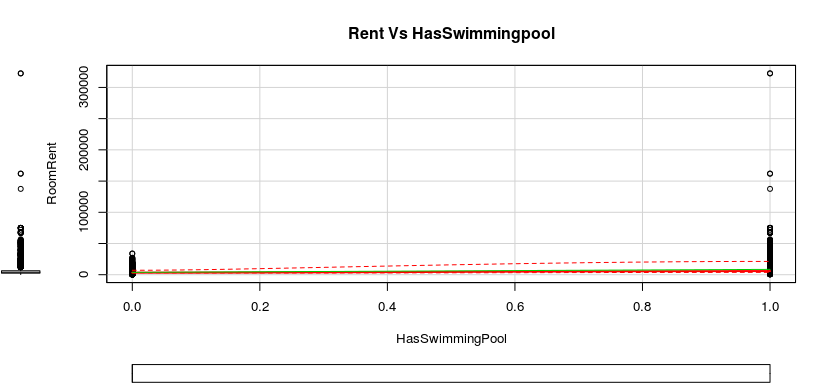
> scatterplot(RoomRent~IsTouristDestination, data=hotel.df,horizontal=TRUE,main="Rent Vs Is tourist destination")



> scatterplot(RoomRent~HotelCapacity, data=hotel.df,horizontal=TRUE,main="Rent Vs Hotel Capacity",xlab="Capacity",spread=FALSE)



> scatterplot(RoomRent~HasSwimmingPool, data=hotel.df,horizontal=TRUE,main="Rent Vs HasSwimmingpool")



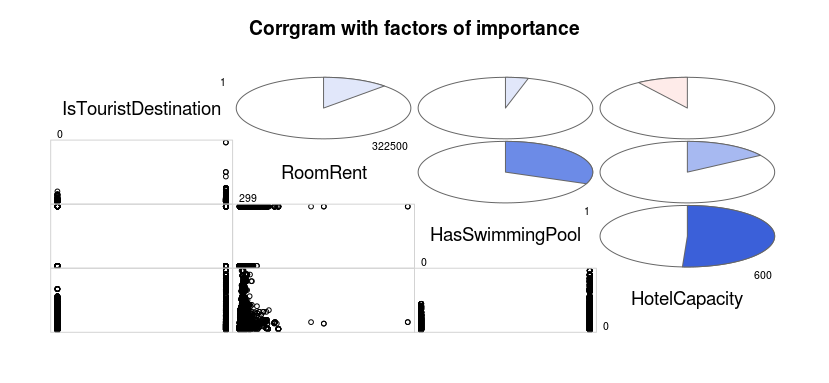
> finalfactor1 <- c("RoomRent", "IsTouristDestination","HasSwimmingPool","HotelCapacity")

> corrgram(hotel.df[,finalfactor1], order=TRUE,

+ main="Corrgram with factors of importance",

+ lower.panel=panel.pts, upper.panel=panel.pie,

+ diag.panel=panel.minmax, text.panel=panel.txt)



Varianvce covariance matrix xreation using vcov

> k<- lm(RoomRent ~IsTouristDestination+HasSwimmingPool+HotelCapacity,data=hotel.df)

> vcov(k)

(Intercept) IsTouristDestination HasSwimmingPool HotelCapacity

(Intercept) 15478.37608 -12495.95953 -1889.07364 -39.5735190

IsTouristDestination -12495.95953 17497.04741 -2044.19506 16.3939779

HasSwimmingPool -1889.07364 -2044.19506 21559.73039 -69.6959561

HotelCapacity -39.57352 16.39398 -69.69596 0.8469172

Hypothesis 1

The rent of hotel depends on the factors as TouristDestination, SwimmingPool, HotelCapacity , metrocity, Isweekend,Starrating , freewifi, cityrank,airport distance,

Isnewyear.

> hyp1<- lm(RoomRent~ IsTouristDestination+ HasSwimmingPool+ HotelCapacity +IsMetroCity+ IsWeekend+IsNewYearEve+ Airport+CityRank+FreeWifi+FreeBreakfast,data=hotel.df)

> summary(hyp1)

Call:

lm(formula = RoomRent ~ IsTouristDestination + HasSwimmingPool +

HotelCapacity + IsMetroCity + IsWeekend + IsNewYearEve +

Airport + CityRank + FreeWifi + FreeBreakfast, data = hotel.df)

Residuals:

Min 1Q Median 3Q Max

-9750 -2376 -844 923 313804

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 945.9588 286.5052 3.302 0.000964 \*\*\*

IsTouristDestination 1707.9465 152.2021 11.222 < 2e-16 \*\*\*

HasSwimmingPool 4532.3518 148.1034 30.603 < 2e-16 \*\*\*

HotelCapacity 4.7858 0.9473 5.052 4.43e-07 \*\*\*

IsMetroCity -1080.5273 180.3110 -5.993 2.12e-09 \*\*\*

IsWeekend -91.8868 128.6561 -0.714 0.475113

IsNewYearEve 893.5627 188.8938 4.731 2.26e-06 \*\*\*

Airport 12.1547 3.0852 3.940 8.20e-05 \*\*\*

CityRank 29.5880 6.7739 4.368 1.26e-05 \*\*\*

FreeWifi 863.1765 232.6489 3.710 0.000208 \*\*\*

FreeBreakfast 282.4669 128.0402 2.206 0.027396 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6858 on 13221 degrees of freedom

Multiple R-squared: 0.1261, Adjusted R-squared: 0.1255

F-statistic: 190.8 on 10 and 13221 DF, p-value: < 2.2e-16

> hyp1$coefficients

(Intercept) IsTouristDestination HasSwimmingPool HotelCapacity IsMetroCity IsWeekend

945.958777 1707.946550 4532.351750 4.785836 -1080.527308 -91.886817

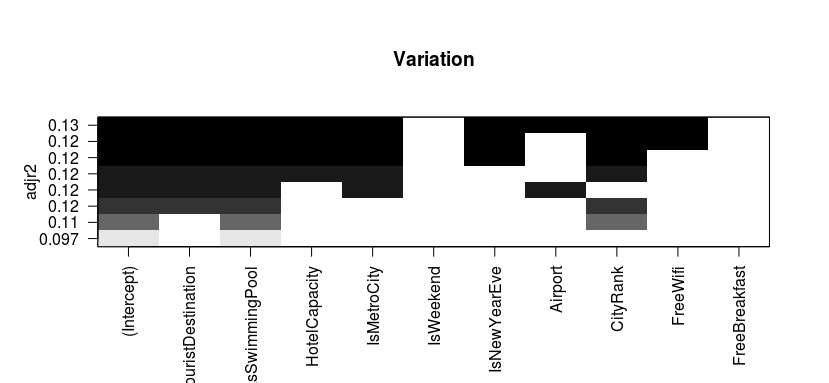
IsNewYearEve Airport CityRank FreeWifi FreeBreakfast

893.562690 12.154724 29.588022 863.176499 282.466875

> install.packages("leaps")

> obs <- regsubsets(mode, data = hotel.df, nbest=1)

> plot(obs,scale="adjr2")



#hypothesis2

#The rent of hotel depends on the factors as TouristDestination, SwimmingPool, HotelCapacity , metrocity, Isweekend,Starrating , freewifi, cityrank,airport distance,

> mode2 <- RoomRent~ IsTouristDestination+ HasSwimmingPool+ HotelCapacity+IsMetroCity+IsNewYearEve+ Airport+CityRank+FreeWifi

> hyp2 <- lm(mode2,data=hotel.df)

> summary(hyp2)

Call:

lm(formula = mode2, data = hotel.df)

Residuals:

Min 1Q Median 3Q Max

-9667 -2395 -868 950 313605

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1013.6273 271.2284 3.737 0.000187 \*\*\*

IsTouristDestination 1674.0802 151.3178 11.063 < 2e-16 \*\*\*

HasSwimmingPool 4534.8862 148.1147 30.617 < 2e-16 \*\*\*

HotelCapacity 4.6024 0.9438 4.877 1.09e-06 \*\*\*

IsMetroCity -1043.4271 179.5160 -5.812 6.30e-09 \*\*\*

IsNewYearEve 852.9641 180.6624 4.721 2.37e-06 \*\*\*

Airport 12.4081 3.0833 4.024 5.75e-05 \*\*\*

CityRank 30.2564 6.7678 4.471 7.86e-06 \*\*\*

FreeWifi 940.7937 229.9815 4.091 4.33e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6859 on 13223 degrees of freedom

Multiple R-squared: 0.1258, Adjusted R-squared: 0.1252

F-statistic: 237.8 on 8 and 13223 DF, p-value: < 2.2e-16

> hyp2$coefficients

(Intercept) IsTouristDestination HasSwimmingPool HotelCapacity IsMetroCity IsNewYearEve

1013.627269 1674.080189 4534.886241 4.602403 -1043.427068 852.964132

Airport CityRank FreeWifi

12.408094 30.256428 940.793703

> summary(hyp1)$adj.r.squared

[1] 0.1254598

> summary(hyp2)$adj.r.squared

[1] 0.1252345

> AIC(hyp1)

[1] 271323.7

> AIC(hyp2)

[1] 271325.1

> finalfactor1 <- c("RoomRent", "IsTouristDestination","HasSwimmingPool","HotelCapacity")

> matrix<-rcorr(as.matrix(hotel.df[,finalfactor1]))

> matrix

RoomRent IsTouristDestination HasSwimmingPool HotelCapacity

RoomRent 1.00 0.12 0.31 0.16

IsTouristDestination 0.12 1.00 0.04 -0.09

HasSwimmingPool 0.31 0.04 1.00 0.51

HotelCapacity 0.16 -0.09 0.51 1.00

n= 13232

P

RoomRent IsTouristDestination HasSwimmingPool HotelCapacity

RoomRent 0 0 0

IsTouristDestination 0 0 0

HasSwimmingPool 0 0 0

HotelCapacity 0 0 0