**Hotel Room Pricing In The Indian Market Analysis**

Reading the data into R

cities <- read.csv("Cities42.csv")

DATA SUMMARY

> summary(cities)  
 X CityName Population CityRank IsMetroCity IsTouristDestination IsWeekend   
 Min. : 1 Delhi :2048 Min. : 8096 Min. : 0.00 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.: 3309 Jaipur : 768 1st Qu.: 744983 1st Qu.: 2.00 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median : 6616 Mumbai : 712 Median : 3046163 Median : 9.00 Median :0.0000 Median :1.0000 Median :1.0000   
 Mean : 6616 Bangalore: 656 Mean : 4416837 Mean :14.83 Mean :0.2842 Mean :0.6972 Mean :0.6228   
 3rd Qu.: 9924 Goa : 624 3rd Qu.: 8443675 3rd Qu.:24.00 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
 Max. :13232 Kochi : 608 Max. :12442373 Max. :44.00 Max. :1.0000 Max. :1.0000 Max. :1.0000   
 (Other) :7816   
 IsNewYearEve Date HotelName RoomRent StarRating Airport   
 Min. :0.0000 Dec 21 2016:1611 Vivanta by Taj : 32 Min. : 299 Min. :0.000 Min. : 0.20   
 1st Qu.:0.0000 Dec 24 2016:1611 Goldfinch Hotel : 24 1st Qu.: 2436 1st Qu.:3.000 1st Qu.: 8.40   
 Median :0.0000 Dec 25 2016:1611 OYO Rooms : 24 Median : 4000 Median :3.000 Median : 15.00   
 Mean :0.1244 Dec 28 2016:1611 The Gordon House Hotel: 24 Mean : 5474 Mean :3.459 Mean : 21.16   
 3rd Qu.:0.0000 Dec 31 2016:1611 Apnayt Villa : 16 3rd Qu.: 6299 3rd Qu.:4.000 3rd Qu.: 24.00   
 Max. :1.0000 Dec 18 2016:1608 Bentleys Hotel Colaba : 16 Max. :322500 Max. :5.000 Max. :124.00   
 (Other) :3569 (Other) :13096   
 HotelAddress HotelPincode HotelDescription  
 The Mall, Shimla : 32 Min. : 100025 3 : 120   
 #2-91/14/8, White Fields, Kondapur, Hitech City, Hyderabad, 500084 India: 16 1st Qu.: 221001 Abc : 112   
 121, City Terrace, Walchand Hirachand Marg, Mumbai, Maharashtra : 16 Median : 395003 3-star hotel: 104   
 14-4507/9, Balmatta Road, Near Jyothi Circle, Hampankatta : 16 Mean : 397430 3.5 : 88   
 144/7, Rajiv Gandi Salai (OMR), Kottivakkam, Chennai, Tamil Nadu : 16 3rd Qu.: 570001 4 : 72   
 17, Oliver Road, Colaba, Mumbai, Maharashtra : 16 Max. :7000157 (Other) :12728   
 (Other) :13120 NA's : 8   
 FreeWifi FreeBreakfast HotelCapacity HasSwimmingPool   
 Min. :0.0000 Min. :0.0000 Min. : 0.00 Min. :0.0000   
 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.: 16.00 1st Qu.:0.0000   
 Median :1.0000 Median :1.0000 Median : 34.00 Median :0.0000   
 Mean :0.9259 Mean :0.6491 Mean : 62.51 Mean :0.3558   
 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.: 75.00 3rd Qu.:1.0000   
 Max. :1.0000 Max. :1.0000 Max. :600.00 Max. :1.0000

Linear Regression Model:

model <- lm(RoomRent ~ StarRating + IsWeekend + IsTouristDestination + IsMetroCity + IsNewYearEve + FreeWifi + FreeBreakfast + HasSwimmingPool + Airport,data = cities )

summary(model)

> model <- lm(RoomRent ~ StarRating + IsWeekend + IsTouristDestination + IsMetroCity + IsNewYearEve + FreeWifi + FreeBreakfast + HasSwimmingPool + Airport,data = cities )  
> summary(model)  
  
Call:  
lm(formula = RoomRent ~ StarRating + IsWeekend + IsTouristDestination +   
 IsMetroCity + IsNewYearEve + FreeWifi + FreeBreakfast + HasSwimmingPool +   
 Airport, data = cities)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-10451 -2349 -712 978 310394   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -7726.314 399.765 -19.327 < 2e-16 \*\*\*  
StarRating 3014.799 98.267 30.680 < 2e-16 \*\*\*  
IsWeekend -96.609 124.494 -0.776 0.4378   
IsTouristDestination 2330.496 133.374 17.473 < 2e-16 \*\*\*  
IsMetroCity -1867.441 135.859 -13.745 < 2e-16 \*\*\*  
IsNewYearEve 887.140 182.783 4.854 1.23e-06 \*\*\*  
FreeWifi 562.209 224.987 2.499 0.0125 \*   
FreeBreakfast 309.581 123.255 2.512 0.0120 \*   
HasSwimmingPool 1875.286 156.573 11.977 < 2e-16 \*\*\*  
Airport 11.310 2.721 4.156 3.26e-05 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 6636 on 13222 degrees of freedom  
Multiple R-squared: 0.1817, Adjusted R-squared: 0.1811   
F-statistic: 326.2 on 9 and 13222 DF, p-value: < 2.2e-16

Result: We can see from above regression model that variabe RoomRent significantly depends upon all variables except IsWeekEnd. So let’s remove this variable to make our model even better.

More better regression model:

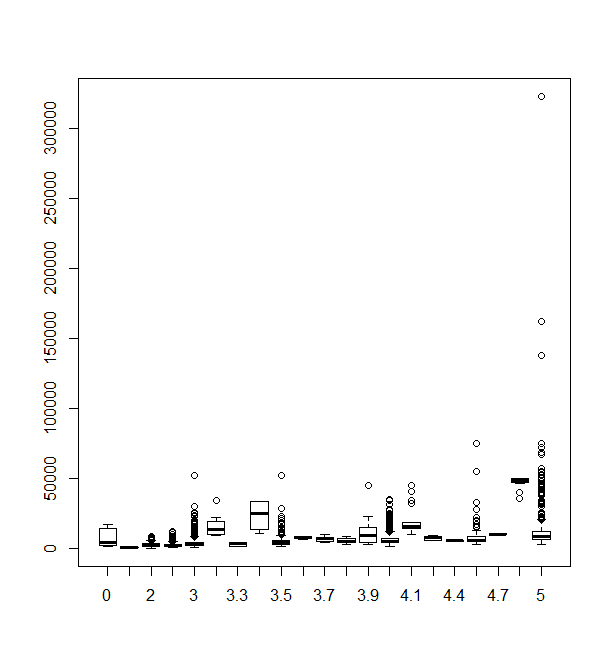
> model1 <- lm(RoomRent ~ StarRating + IsTouristDestination + IsNewYearEve + FreeWifi + FreeBreakfast + HasSwimmingPool + Airport + IsMetroCity,data = cities)  
> summary(model1)  
  
Call:  
lm(formula = RoomRent ~ StarRating + IsTouristDestination + IsNewYearEve +   
 FreeWifi + FreeBreakfast + HasSwimmingPool + Airport + IsMetroCity,   
 data = cities)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-10452 -2361 -711 984 310352   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -7782.035 393.257 -19.789 < 2e-16 \*\*\*  
StarRating 3014.683 98.265 30.679 < 2e-16 \*\*\*  
IsTouristDestination 2332.651 133.343 17.494 < 2e-16 \*\*\*  
IsNewYearEve 845.669 174.794 4.838 1.33e-06 \*\*\*  
FreeWifi 561.683 224.983 2.497 0.0126 \*   
FreeBreakfast 310.458 123.248 2.519 0.0118 \*   
HasSwimmingPool 1874.913 156.569 11.975 < 2e-16 \*\*\*  
Airport 11.303 2.721 4.153 3.30e-05 \*\*\*  
IsMetroCity -1868.073 135.855 -13.751 < 2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 6636 on 13223 degrees of freedom  
Multiple R-squared: 0.1816, Adjusted R-squared: 0.1812   
F-statistic: 366.9 on 8 and 13223 DF, p-value: < 2.2e-16

Result : We can see from above regression model that RoomRent depends upon all variables significantly. Three most significant and important variables are StarRating , IsTouristDestination and HasSwimmingPool as coefficients of these variables are highest as given by the above model.

Let’s see the dependence of dependent variable RoomRent upon these three important variables.

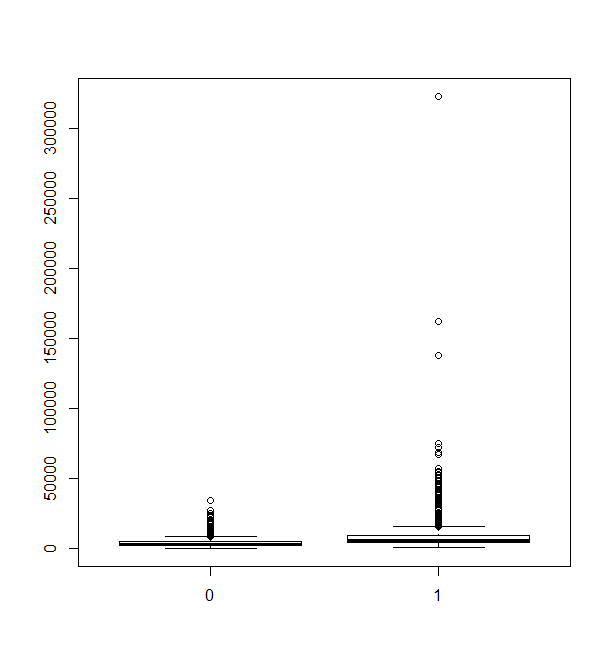
Boxplot of RoomRent vs StarRating

> boxplot(RoomRent ~ StarRating,data = cities)



Boxplot of RoomRent vs HasSwimmingPool

> boxplot(RoomRent ~ HasSwimmingPool,data = cities)



Boxplot of RoomRent vs IsTouristDestination

> boxplot(RoomRent ~ IsTouristDestination,data = cities)

