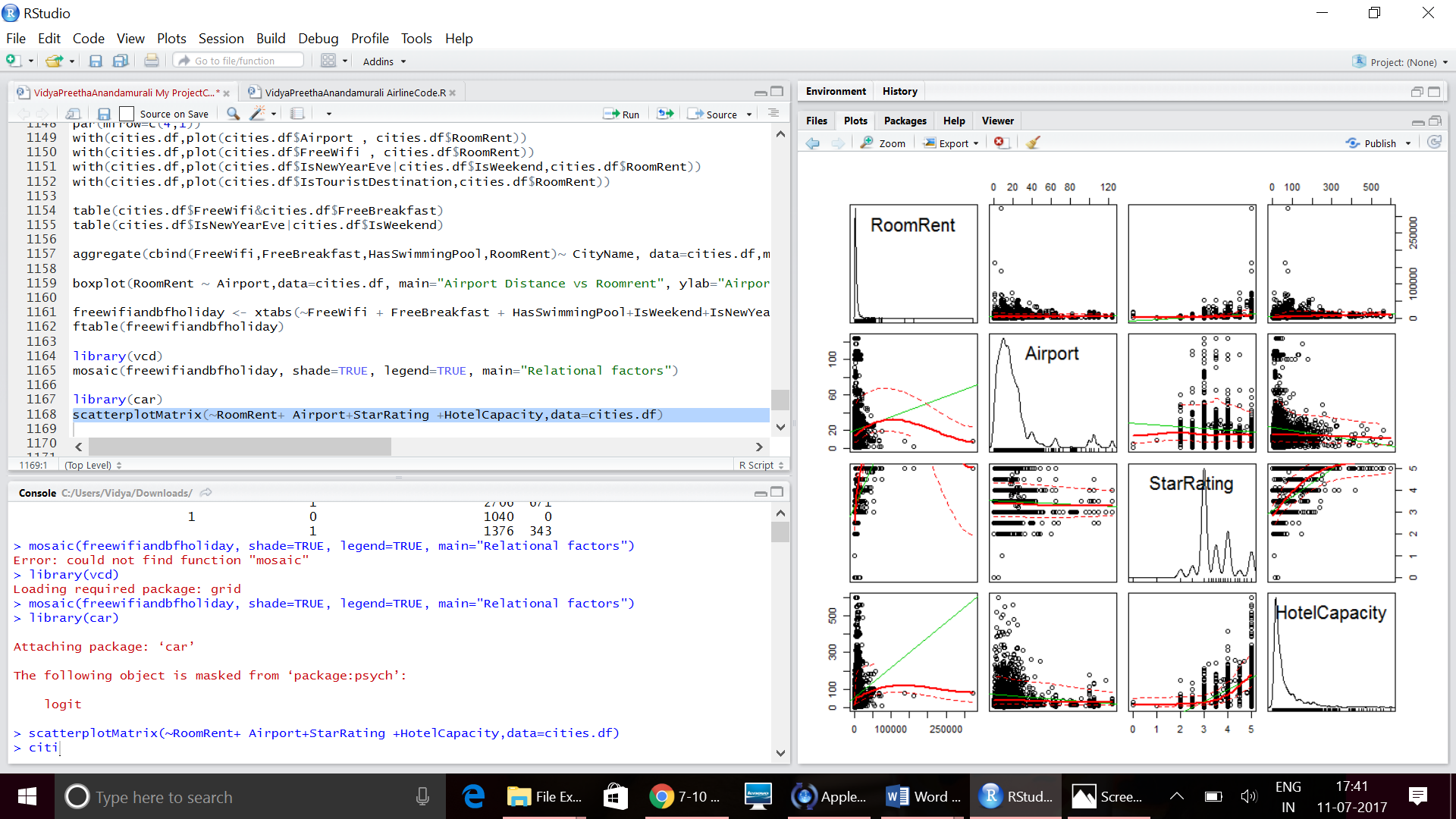
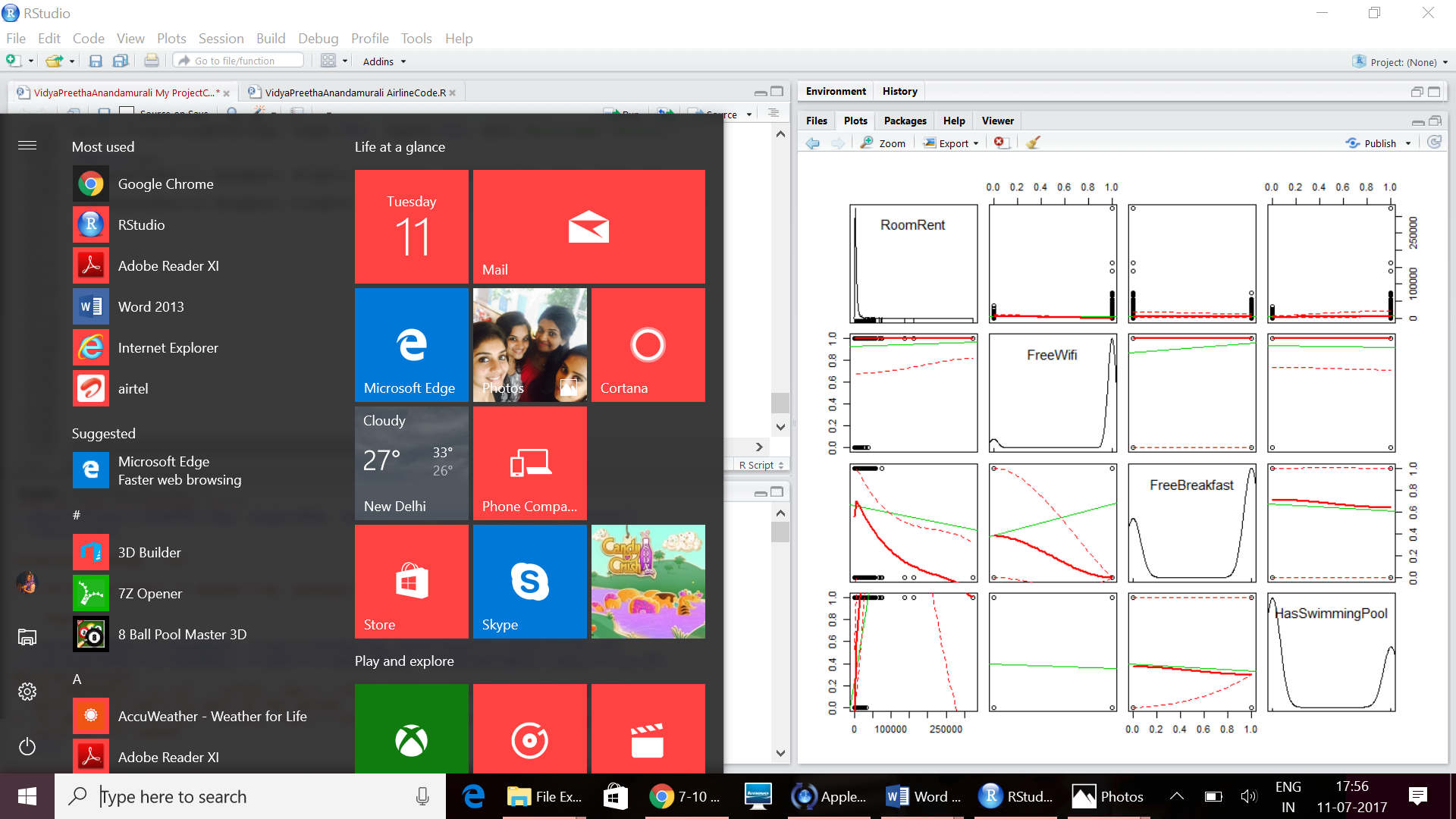
**Output for the combined data set of the 42 cities**

**Summary Statistics**

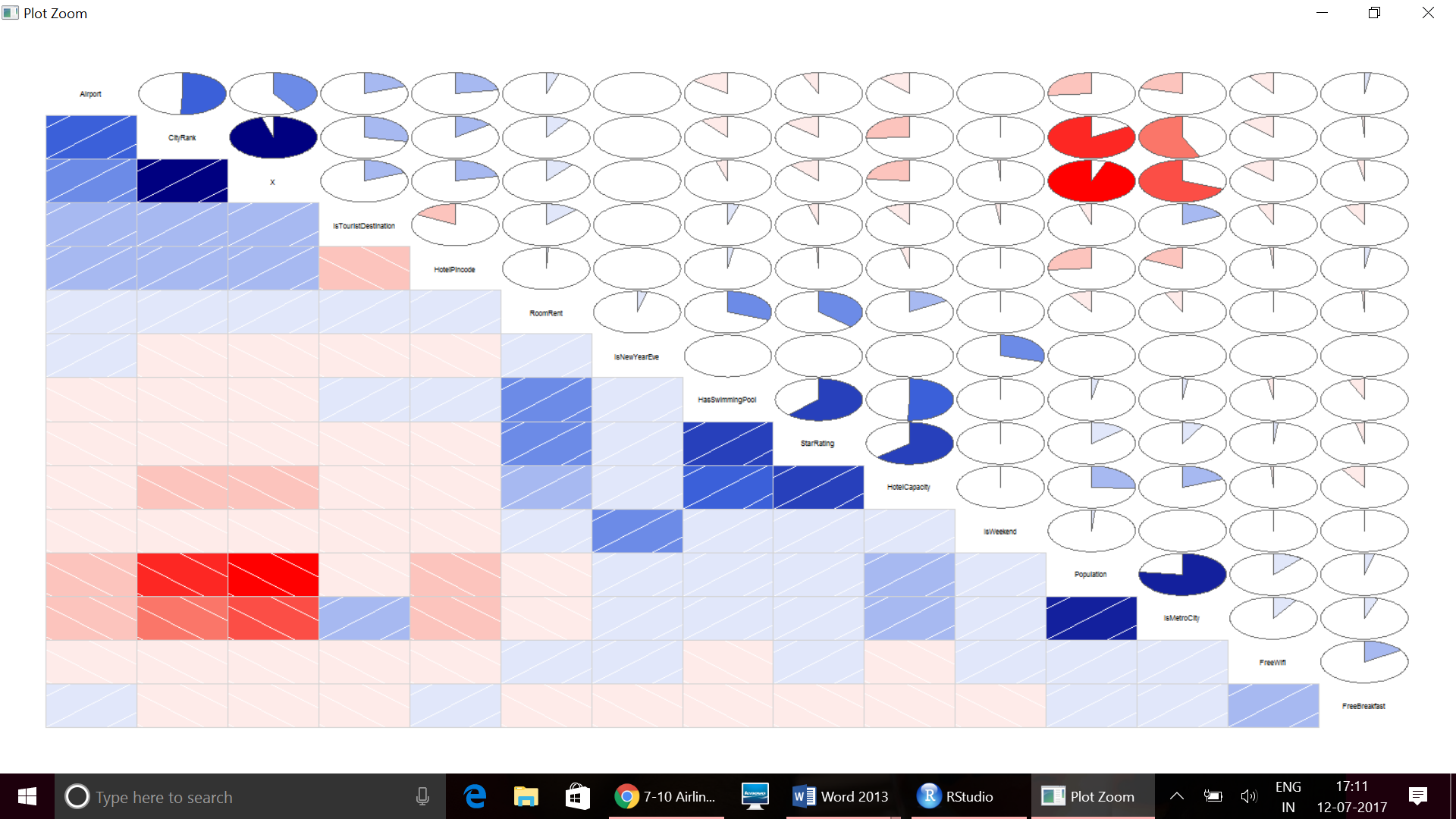
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n mean sd median min max  X 13232 6616.50 3819.89 6616.5 1.0 13232  CityName\* 13232 18.07 11.72 16.0 1.0 42  Population 13232 4416836.87 4258386.00 3046163.0 8096.0 12442373  CityRank 13232 14.83 13.51 9.0 0.0 44  IsMetroCity 13232 0.28 0.45 0.0 0.0 1  IsTouristDestination 13232 0.70 0.46 1.0 0.0 1  IsWeekend 13232 0.62 0.48 1.0 0.0 1  IsNewYearEve 13232 0.12 0.33 0.0 0.0 1  Date\* 13232 14.26 2.82 14.0 1.0 20  HotelName\* 13232 841.19 488.16 827.0 1.0 1670  RoomRent 13232 5473.99 7333.12 4000.0 299.0 322500  StarRating 13232 3.46 0.76 3.0 0.0 5  Airport 13232 21.16 22.76 15.0 0.2 124  HotelAddress\* 13232 1202.53 582.17 1261.0 1.0 2108  HotelPincode 13232 397430.26 259837.50 395003.0 100025.0 7000157  HotelDescription\* 13224 581.34 363.26 567.0 1.0 1226  FreeWifi 13232 0.93 0.26 1.0 0.0 1  FreeBreakfast 13232 0.65 0.48 1.0 0.0 1  HotelCapacity 13232 62.51 76.66 34.0 0.0 600  HasSwimmingPool 13232 0.36 0.48 0.0 0.0 1  Visualisations  Scatter Plots for analysis of room rent variation with respect to free wifi, breakfast,distance to airport and  Whether it’s a holiday or a weekend.  C:\Users\Vidya\Pictures\Screenshots\Screenshot (465).png  We can see that the room rent is higher when there is free wifi and when the distance is closer to the airport. Also when it’s a weekend or new years eve prices are showing a higher range than those ,but with a very marginal difference.  **Role of holiday and is weekend on room rent:**  table(cities.df$FreeWifi&cities.df$FreeBreakfast)  FALSE TRUE  5018 8214  table(cities.df$IsNewYearEve|cities.df$IsWeekend)  FALSE TRUE  4989 8243  **Role of each cities**   |  | | --- | | aggregate(cbind(FreeWifi,FreeBreakfast,HasSwimmingPool,RoomRent)~ CityName, d  ata=cities.df,mean)  CityName FreeWifi FreeBreakfast HasSwimmingPool RoomRent  1 Agra 0.9814815 0.4074074 0.42592593 4124.287  2 Ahmedabad 0.8679245 0.8301887 0.26415094 4175.045  3 Amritsar 0.8823529 0.2941176 0.29411765 3444.029  4 Bangalore 0.9634146 0.6219512 0.37804878 4112.803  5 Bhubaneswar 1.0000000 0.7333333 0.20000000 3587.442  6 Chandigarh 0.8095238 0.1666667 0.23809524 4030.940  7 Chennai 0.9615385 0.7692308 0.57692308 4323.647  8 Darjeeling 1.0000000 1.0000000 0.05882353 5458.088  9 Delhi 0.9765625 0.6816406 0.36328125 4318.606  10 Gangtok 0.9375000 0.4375000 0.12500000 4629.648  11 Goa 0.7948718 0.5769231 0.69230769 8170.801  12 Guwahati 1.0000000 0.8333333 0.66666667 5325.812  13 Haridwar 0.8333333 0.5000000 0.16666667 3919.938  14 Hyderabad 1.0000000 0.6250000 0.26865672 3852.175  15 Indore 1.0000000 0.9500000 0.05000000 3414.594  16 Jaipur 0.9296875 0.5182292 0.48567708 7292.022  17 Jaisalmer 0.8787879 0.4128788 0.30303030 5986.072  18 Jodhpur 0.9285714 0.5357143 0.46428571 10661.371  19 Kanpur 1.0000000 1.0000000 0.00000000 3008.562  20 Kochi 0.8684211 0.6447368 0.48684211 6039.609  21 Kolkata 0.9238281 0.5761719 0.35742188 4528.986  22 Lucknow 1.0000000 0.7734375 0.25000000 5879.070  23 Madurai 1.0000000 1.0000000 0.21428571 4768.223  24 Manali 0.8888889 0.7222222 0.05555556 4858.285  25 Mangalore 1.0000000 0.7692308 0.23076923 4110.337  26 Mumbai 0.9325843 0.6629213 0.33707865 6343.730  27 Munnar 0.8780488 0.8536585 0.02439024 7543.500  28 Mysore 1.0000000 0.9000000 0.25000000 3320.869  29 Nainital 0.7777778 0.5555556 0.16666667 6409.833  30 Ooty 0.7058824 0.4117647 0.11764706 6144.257  31 Panchkula 1.0000000 0.6250000 0.25000000 2813.500  32 Pune 0.9600000 0.7466667 0.38666667 3897.652  33 Puri 0.5714286 0.4285714 0.57142857 5708.429  34 Rajkot 1.0000000 1.0000000 0.25000000 4107.078  35 Rishikesh 0.9090909 0.7840909 0.27272727 4943.670  36 Shimla 0.8000000 0.7714286 0.14285714 5780.604  37 Srinagar 0.8000000 0.6000000 0.40000000 10572.025  38 Surat 0.9000000 0.9000000 0.20000000 3660.850  39 Thiruvanthipuram 0.9183673 0.5918367 0.67346939 6726.796  40 Thrissur 1.0000000 0.7500000 0.00000000 3387.844  41 Udaipur 0.9298246 0.7192982 0.45614035 10145.252  42 Varanasi 1.0000000 0.6969697 0.18181818 8675.042  Effect of price and distance from airport using box plots:  C:\Users\Vidya\Pictures\Screenshots\Screenshot (467).png  Result : Hotels that are located closer to the airport cost much more than those that don’t. Outliers beyond 300000 are present which indicate the above result.  **Relation between free wifi and free breakfast :**  freewifiandbf <- xtabs(~FreeWifi + FreeBreakfast,data=cities.df)  ftable(freewifiandbf)  FreeBreakfast 0 1  FreeWifi  0 606 375  1 4037 8214  **Relation with all factors above and holidays and swimming pool:**  freewifiandbfholiday <- xtabs(~FreeWifi + FreeBreakfast + HasSwimmingPool+  IsWeekend+IsNewYearEve,data=cities.df)  >ftable(freewifiandbfholiday)  IsNewYearEve 0 1  FreeWifi FreeBreakfast HasSwimmingPool IsWeekend  0 0 0 0 125 0  1 162 41  1 0 105 0  1 139 34  1 0 0 102 0  1 129 33  1 0 43 0  1 54 14  1 0 0 0 924 0  1 1242 311  1 0 574 0  1 789 197  1 0 0 2076 2  1 2706 671  1 0 1040 0  1 1376 343  library(vcd)  mosaic(freewifiandbfholiday, shade=TRUE, legend=TRUE, main="Relational factors") | |  | | |  | | --- | |  | |   C:\Users\Vidya\Pictures\Screenshots\Screenshot (468).png  CORRELATIONS:  Library(car)  scatterplotMatrix(~RoomRent+ Airport+StarRating +HotelCapacity,data=cities.df) |
|  |
| |  | | --- | |  | |



Library(car)  
scatterplotMatrix(~RoomRent+ FreeWifi+FreeBreakfast +HasSwimmingPool,data=cities.df)



Corrgram:



Matrix:

|  |
| --- |
| CityRank IsMetroCity IsTouristDestination IsWeekend  CityRank 1.0000000000 -0.5643937903 0.280713452 -0.007256477  IsMetroCity -0.5643937903 1.0000000000 0.176371706 0.001811801  IsTouristDestination 0.2807134520 0.1763717063 1.000000000 -0.019481101  IsWeekend -0.0072564766 0.0018118005 -0.019481101 1.000000000  IsNewYearEve -0.0006326444 0.0006464753 -0.002266388 0.292382051  RoomRent 0.0939855292 -0.0668397705 0.122502963 0.004580134  StarRating -0.1333810133 0.0776028661 -0.040554998 0.006378436  Airport 0.5059119892 -0.2073586125 0.194422049 -0.002724756  FreeWifi -0.1214309404 0.0868288677 -0.061568821 0.002960828  FreeBreakfast -0.0086837497 0.0513856623 -0.071692559 -0.007612777  HotelCapacity -0.2561197059 0.1871502153 -0.094356091 0.006306507  HasSwimmingPool -0.1029737518 0.0214119243 0.042156280 0.004500461  IsNewYearEve RoomRent StarRating Airport FreeWifi  CityRank -6.326444e-04 0.093985529 -0.133381013 0.5059119892 -1.214309e-01  IsMetroCity 6.464753e-04 -0.066839771 0.077602866 -0.2073586125 8.682887e-02  IsTouristDestination -2.266388e-03 0.122502963 -0.040554998 0.1944220492 -6.156882e-02  IsWeekend 2.923821e-01 0.004580134 0.006378436 -0.0027247555 2.960828e-03  IsNewYearEve 1.000000e+00 0.038491227 0.002360897 0.0004598872 2.787472e-05  RoomRent 3.849123e-02 1.000000000 0.369373425 0.0496532442 3.627002e-03  StarRating 2.360897e-03 0.369373425 1.000000000 -0.0609191837 1.800959e-02  Airport 4.598872e-04 0.049653244 -0.060919184 1.0000000000 -9.452368e-02  FreeWifi 2.787472e-05 0.003627002 0.018009594 -0.0945236768 1.000000e+00  FreeBreakfast -2.606416e-03 -0.010006370 -0.032892463 0.0242839409 1.582206e-01  HotelCapacity 1.352679e-03 0.157873308 0.637430337 -0.1176720722 -8.703612e-03  HasSwimmingPool 1.122308e-03 0.311657734 0.618214699 -0.1416665606 -2.407405e-02  FreeBreakfast HotelCapacity HasSwimmingPool  CityRank -0.008683750 -0.256119706 -0.102973752  IsMetroCity 0.051385662 0.187150215 0.021411924  IsTouristDestination -0.071692559 -0.094356091 0.042156280  IsWeekend -0.007612777 0.006306507 0.004500461  IsNewYearEve -0.002606416 0.001352679 0.001122308  RoomRent -0.010006370 0.157873308 0.311657734  StarRating -0.032892463 0.637430337 0.618214699  Airport 0.024283941 -0.117672072 -0.141666561  FreeWifi 0.158220597 -0.008703612 -0.024074046  FreeBreakfast 1.000000000 -0.087165446 -0.061522132  HotelCapacity -0.087165446 1.000000000 0.509045809  HasSwimmingPool -0.061522132 0.509045809 1.000000000 |
|  |
| |  | | --- | |  | |

T Tests:

t.test(RoomRent,StarRating)

Welch Two Sample t-test

data: RoomRent and StarRating

t = 85.813, df = 13231, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

5345.575 5595.491

sample estimates:

mean of x mean of y

5473.991838 3.458933

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| t.test(RoomRent,Airport)  Welch Two Sample t-test  data: RoomRent and Airport  t = 85.535, df = 13231, p-value < 2.2e-16  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  5327.875 5577.792  sample estimates:  mean of x mean of y  5473.99184 21.15874  t.test(RoomRent,FreeWifi&FreeBreakfast)  Welch Two Sample t-test  data: RoomRent and FreeWifi & FreeBreakfast  t = 85.858, df = 13231, p-value < 2.2e-16  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  5348.413 5598.329  sample estimates:  mean of x mean of y  5473.9918380 0.6207678   |  | | --- | | t.test(RoomRent,IsWeekend|IsNewYearEve)  Welch Two Sample t-test  data: RoomRent and IsWeekend | IsNewYearEve  t = 85.858, df = 13231, p-value < 2.2e-16  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  5348.411 5598.327  sample estimates:  mean of x mean of y  5473.9918380 0.6229595  t.test(RoomRent,HotelCapacity)  Welch Two Sample t-test  data: RoomRent and HotelCapacity  t = 84.882, df = 13234, p-value < 2.2e-16  alternative hypothesis: true difference in means is not equal to 0  95 percent confidence interval:  5286.515 5536.445  sample estimates:  mean of x mean of y  5473.99184 62.51164  **Regression Analysis**  model1<- RoomRent~ IsMetroCity + IsTouristDestination+IsWeekend+IsNewYearEve+  Date+StarRating+FreeWifi+FreeBreakfast+HotelCapacity+HasSwimmingPool+Airport  > fit1 <- lm(model1,data =cities.df)  > summary(fit1)  Call:  lm(formula = model1, data = cities.df)  Residuals:  Min 1Q Median 3Q Max  -11915 -2333 -702 1086 309816  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -10516.696 1162.758 -9.045 <2e-16 \*\*\*  IsMetroCity -1723.081 135.153 -12.749 <2e-16 \*\*\*  IsTouristDestination 2174.771 135.085 16.099 <2e-16 \*\*\*  IsWeekend 433.877 442.216 0.981 0.3265  IsNewYearEve 1841.441 1420.686 1.296 0.1949  Date21-Dec-16 504.627 1476.646 0.342 0.7326  Date24-Dec-16 144.000 1408.875 0.102 0.9186  Date25-Dec-16 -17.205 1408.875 -0.012 0.9903  Date28-Dec-16 517.127 1476.646 0.350 0.7262  Date31-Dec-16 -1637.918 2000.819 -0.819 0.4130  Date4-Jan-16 1800.487 1611.633 1.117 0.2639  Date4-Jan-17 908.959 2132.457 0.426 0.6699  Date8-Jan-16 1535.482 1549.631 0.991 0.3218  Date8-Jan-17 488.544 2086.359 0.234 0.8149  DateDec 18 2016 880.485 1014.481 0.868 0.3855  DateDec 21 2016 1495.610 1099.763 1.360 0.1739  DateDec 24 2016 1544.629 1014.448 1.523 0.1279  DateDec 25 2016 1466.613 1014.449 1.446 0.1483  DateDec 28 2016 2011.824 1100.593 1.828 0.0676 .  DateDec 31 2016 385.626 1734.852 0.222 0.8241  DateJan 04 2017 2095.440 1100.414 1.904 0.0569 .  DateJan 08 2017 1349.159 1015.317 1.329 0.1839  DateJan 4 2017 2021.634 1387.662 1.457 0.1452  DateJan 8 2017 1401.229 1290.970 1.085 0.2778  StarRating 3608.197 110.289 32.716 <2e-16 \*\*\*  FreeWifi 435.973 223.667 1.949 0.0513 .  FreeBreakfast 224.609 123.223 1.823 0.0684 .  HotelCapacity -10.998 1.026 -10.719 <2e-16 \*\*\*  HasSwimmingPool 2103.854 157.289 13.376 <2e-16 \*\*\*  Airport 12.213 2.725 4.482 7.47e-06 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 6608 on 13203 degrees of freedom  Multiple R-squared: 0.1897, Adjusted R-squared: 0.1879  F-statistic: 110.4 on 28 and 13203 DF, p-value: < 2.2e-16  **To find the best predictors**  library(leaps)  leap1 <- regsubsets(model1,data =cities.df,nbest=1)  plot(leap1,scale ="adjr2")  C:\Users\Vidya\Pictures\Screenshots\Screenshot (472).png  **Refined Model**  **model2<- RoomRent~ IsMetroCity + IsTouristDestination+StarRating+HotelCapacity+**  **HasSwimmingPool + Airport + CityRank**  **fit2 <- lm(model2,data =cities.df)**  **summary(fit2)**  Call:  lm(formula = model2, data = cities.df)  Residuals:  Min 1Q Median 3Q Max  -11511 -2334 -710 1041 309145  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -8156.507 350.268 -23.29 <2e-16 \*\*\*  IsMetroCity -1643.989 133.337 -12.33 <2e-16 \*\*\*  IsTouristDestination 2211.235 129.383 17.09 <2e-16 \*\*\*  StarRating 3617.338 110.180 32.83 <2e-16 \*\*\*  HotelCapacity -11.271 1.022 -11.03 <2e-16 \*\*\*  HasSwimmingPool 2103.621 157.284 13.38 <2e-16 \*\*\*  Airport 11.074 2.701 4.099 4.17e-05 \*\*\*  CityRank 28.070 6.517 4.307 1.67e-05 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 6618 on 13226 degrees of freedom  Multiple R-squared: 0.1857, Adjusted R-squared: 0.1854  F-statistic: 603.3 on 5 and 13226 DF, p-value: < 2.2e-16  **Visualising the beta coefficients and their confidence intervals from Model 2**  library(coefplot)  coefplot(fit2,intercept = FALSE,outerCI =1.96,coefficients= c("IsMetroCity",  "Istourist","StarRating","HotelCapacity","Haspool","Airport"))  **C:\Users\Vidya\Pictures\Screenshots\Screenshot (473).png**  **MODEL 2 fits better than MODEL 1, as indicated by Adjusted R**  **Squared and AIC values**  summary(fit1)$adj.r.squared  [1] 0.1891081  summary(fit2)$adj.r.squared  [1] 0.1863897  AIC values  AIC(fit1)  [1] 270362.8  AIC(fit2)  [1] 270344.1  Thus Model 2 is the best fit model that helps us to predict the factors that are  responsible for the room rent across various places in India. | |  | | |  | | --- | |  | | |
|  |
| |  | | --- | |  | |