**A Study of factors influencing Hotel Rents, India**

**James Rohan Gangavarapu**

1. **Introduction**

The services sector in India has experienced impressive growth over the years. The services sector comprises of more than half of the national GDP and employs more than one third of the work population. The Indian hospitality sector has experienced impressive growth. The growth in Indian Hospitality sector can be attributed to the Indian tourism sector that has seen significant levels of growth in both Domestic and International Tourism. This is growth in tourism has in turn impacted the occupancy rates and room rents in the Hospitality Sector (IBEF, 2008a, b) (Mohsin, Asad, and Tim Lockyer).

There are many different categories of tourism such as cultural, business etc. Studying the link between the tourism sector and hospitality sector is beyond the purview of this study. This study primarily focuses on the pricing of the room rents based on certain Internal (Swimming Pool, Free WiFi, Free Breakfast, Hotel Capacity etc.) and External Factors (Date, New Year’s Eve, City Population etc.). However, this paper does explore the effect of social media ranking of different hotels due to the findings in *Dwivedi, Mridula, Shibu et al*.

1. **Data Set**

This study uses the data set found here: <https://drive.google.com/file/d/0B3oXSJHLqZqXeGxwMjJ6eDg2UTA/view>. It is named Cities42.csv and contains the Room Rent for hotels along with other information related to Internal and External Factors of the hotels across 42 different cities in India. This data set uses the population figures of cities from the 2011 National Census as shown in table 1. This data was collected from [www.hotels.in](http://www.hotels.in) website on the following dates, Dec 31, Dec 25, Dec 24, Dec 18, Dec 21, Dec 28, Jan 4, Jan 8. The data set was compiled on October 2016 by the website.

**2.1 Factors**

The data sets primarily indicates two types of factors. External Factors and Internal Factors. The hotels do not have any kind of control over the external factors whereas they can influence and control the internal factors.

The dependent factor is the Room Rent listed in Indian Rupee(INR).

**2.1.0 External Factors**

Factors that are beyond the control of an Individual hotel such as: Is it a Weekend, The Date of the Booking, Is it the New Year’s Eve, the name of the city/city-population/city-rank, is it a metro city, is it a tourist destination.

The rank of the City is based on the population. The Rank 0 indicates the city with highest population and Rank 42 indicates the city with the smallest population among the 42 cities listed in the Data Set (Table 1). The data types of these individual factors are listed in Table 2.

**2.1.1 Internal Factors**

Factors that the hotel has a great deal of control and influence over such as: The name of the hotel, it’s star rating on the website, the distance between the hotel and the airport, Hotel Address, the pin-code/zip-code of the hotel, hotel description, does the hotel have free Wifi, does the hotel have free breakfast, does the hotel have swimming pool and the capacity of the hotel. The data types of these variables are listed in Table 2.

**2.2 Findings**

The Summary Statistics for the table are listed in the Output file.

When we try to find the boxplot distribution for Room Rent based on City Ranking using the original dataset we get Figure 1. The range of room rents is too large. In order to stabilize the variance we take the log to the base 2 of the actual room rent. We shall refer to this new Data Vector as Log-Room-Rent. We use the new Log-Room-Rent and repeat the above process to output Figure 2. The variance in the Room Rent has been stabilized. From this we can see that we can perform a t-test/Chi-Square-test to see if the mean of Room Rent is equal or not equal as the City Rank changes. Since the City Rank is in turn related to the Population this should give us further insights based on the City Population. This Hypothesis is tested in **2.3.1 From Figure 1 and Figure 2**.

Using the Log-Room-Rent we can visualize the distribution of Log Room Rents based on the City being a Metro City or not. The mean of Room Prices between Metro and Non Metro Cities is about the same while the middle 50%(2nd and 3rd quartile) for Metro and Non Metro is distributed almost identically. The two Boxplots in general are almost identically distributed. Metro Cities have a relatively broader range of cheap and expensive rooms compared to non-metro cities and as a result Metro Cities have a much greater number of Room Rent outliers that fall above the 75% and below the 25% mark. Refer Figure 3. We can perform a t-test/Chi-Square to check whether the average room rent for a Metro and Non-Metro Cities is equal. The Hypothesis is tested in **2.3.2 From Figure 3**. It should be noted that only 28.42% of the rooms listed are in a Metro City and 71.58% of the other rooms are in a Non-Metro City.

We check to see if having a swimming pool has any effect on the room rents. Having a swimming pool indicates the rooms are priced higher in both Metro and Non-Metro areas and the rooms are priced lower in both Metro and Non-Metro areas without a swimming pool. Refer Figure 4.

Plotting room rents with respect to date we can see that room rents show a slight increase from December to January. Refer Figure 5(Note Data Scrubbing certain overlapping dates can result in a slightly better organized box plot). We can test this further by plotting the room rents with respect to whether that day is a weekend or not or whether it is a new year or not. The weekends and weekday price distribution is identical except that there are more rooms that are priced higher in the highest 25% price range during the weekend than the weekday. Refer Figure 6. Plotting room rent with respect to new year eve shows almost identical distribution with the mean room prices and the 75% quartile mark being slightly higher during the new year eve compared to other dates. Refer Figure 7.

Now we check to see if the place being a tourist destination has a role to play. Room rents in tourist destinations are priced higher especially in Non-Metro Cities. Refer Figure 8 and Figure 9.

We Check to see if the Distance to the airport has a role to play. The majority of the hotels in the data set are located less than 20-30 Km from the airport. Refer figure 10.

We check whether there exists a relationship between free Wi-Fi and room rent. 92.59% of the hotel rooms in the data set have free Wi-Fi service and the remaining 7.41% do not. Per Figure 11 the more expensive a room the more likely it is to have free Wi-Fi.

Checking to see if Free Breakfast determines the room rent. From the scatter plot matrix in Figure 12 we can observe that as the room rent increases so do the chances of getting a free breakfast. It must be noted that 64.91% of the hotels listed have free breakfast.

Comparing the hotel capacity to Log-Room-Rent we observe that as the capacity of the hotel gets larger the room rent also increases. See Figure 13.

**2.3 Hypothesis Testing**

**2.3.1 From Figure 1 and Figure 2:**

Hypothesis H0: The average Price of Room Rents is **equal** for different Cities based on Rank.

Hypothesis H1: The average Price of Room Rents is **not equal** for different Cities based on Rank.

Result: After performing a Chi-Square test we get a P-Value of less than 2.2e-16 . Hence we reject the Null Hypothesis. There is evidence to suggest the City Rank and in turn the City population plays a role in determining the Room Rent.

> RoomRent\_CityRank <- xtabs(hotels.df$LogRoomRent~hotels.df$CityRank)

> chisq.test(RoomRent\_CityRank)

Chi-squared test for given probabilities

data: RoomRent\_CityRank

X-squared = 179560, df = 41, p-value < 2.2e-16

**2.3.2 From Figure 3:**

Hypothesis H0: The average Price of Room Rents is **equal** for Metro and Non-Metro Cities.

Hypothesis H1: The average Price of Room Rents is **not equal** for Metro and Non-Metro Cities.

Result: A two-sided t-test results in a P-Value = 3.807e-09. Hence the Null hypothesis must be rejected. There is strong evidence to suggest that the mean of room rents between Metro and Non-Metro cities is not equal.

> t.test(hotels.df$LogRoomRent~hotels.df$IsMetroCity, alternative = c("two.sided"), var.equal= FALSE)

Welch Two Sample t-test

data: hotels.df$LogRoomRent by hotels.df$IsMetroCity

t = 5.8989, df = 8043, p-value = 3.807e-09

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

95 percent confidence interval:

0.07294274 0.14554940

sample estimates:

mean in group 0 mean in group 1

12.01644 11.90719

**2.3.3 From Figure 13:**

Hypothesis H0: The average Price of Room Rents is **equal** for hotels with different total number of rooms.

Hypothesis H1: The average Price of Room Rents is **not equal** for hotels with different total number of rooms.

Result: After performing a Chi-Square test we get a P-Value of less than 2.2e-16. Hence we reject the Null Hypothesis. There is strong evidence to suggest the average prices of room rent changes with the Hotel Capacity.

> RoomRent\_HotelCap <- xtabs(hotels.df$LogRoomRent~hotels.df$HotelCapacity)

> chisq.test(RoomRent\_HotelCap)

Chi-squared test for given probabilities

data: RoomRent\_HotelCap

X-squared = 863790, df = 253, p-value < 2.2e-16

**2.3.4 From Figure 8 and Figure 9:**

Hypothesis H0: The average Price of Room Rents is **equal** for hotels in tourist destination cities and hotels not in tourist destination cities.

Hypothesis H1: The average Price of Room Rents is **not equal** for hotels in tourist destination cities and hotels not in tourist destination cities.

Result: After performing a t-test we get a P-Value of less than 2.2e-16. Hence we reject the Null Hypothesis. There is strong evidence to suggest the average prices of room rent is not equal for Tourist and Non-Tourist destinations.

> t.test(hotels.df$LogRoomRent~hotels.df$IsTouristDestination, alternative = c("two.sided"), var.equal= FALSE)

Welch Two Sample t-test

data: hotels.df$LogRoomRent by hotels.df$IsTouristDestination

t = -19.134, df = 9955.8, p-value < 2.2e-16

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

95 percent confidence interval:

-0.3646217 -0.2968574

sample estimates:

mean in group 0 mean in group 1

11.75481 12.08555

**2.4 Predictive Models**

**2.4.1 Linear Model 1(LM1)**

We begin by considering the following External Factors: Is Weekend, Is New Year, City Rank, Is Metro City, Is Tourist Destination; and the following Internal Factors: Airport, Free Wi-Fi, Free Breakfast, Hotel Capacity, Swimming Pool, Star Rating of the Hotel. We build a linear model and run the summary function on the model. From the output we see that New Year Eve, City Rank, Tourist Destination, Airport, Free Breakfast, Free Wi-Fi, Hotel Capacity, Has Swimming Pool, Star Rating play a significant role.

Result: Since the P-Value of the F-Statistic is less than 2.2e-16, we can reject the Null hypothesis that the Intercept only model is a better fit. Instead we have strong evidence to suggest Linear Model 1 is a better fit compared to an intercept only model.

|  |
| --- |
| > lm1 <- lm(logRoomRent ~ IsWeekend+IsNewYearEve+CityRank+  + IsMetroCity+IsTouristDestination  + +Airport+FreeWifi+FreeBreakfast+HotelCapacity+  + HasSwimmingPool+StarRating,data = hotels.df)  > summary(lm1)  Call:  lm(formula = logRoomRent ~ IsWeekend + IsNewYearEve + CityRank +  IsMetroCity + IsTouristDestination + Airport + FreeWifi +  FreeBreakfast + HotelCapacity + HasSwimmingPool + StarRating,  data = hotels.df)  Residuals:  Min 1Q Median 3Q Max  -2.9403 -0.5068 -0.0549 0.4260 5.2106  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 9.159e+00 4.885e-02 187.469 < 2e-16 \*\*\*  IsWeekend -1.750e-02 1.450e-02 -1.207 0.227518  IsNewYearEve 1.492e-01 2.129e-02 7.010 2.5e-12 \*\*\*  CityRank 1.139e-02 7.635e-04 14.916 < 2e-16 \*\*\*  IsMetroCity -7.235e-03 2.032e-02 -0.356 0.721840  IsTouristDestination 2.240e-01 1.716e-02 13.050 < 2e-16 \*\*\*  Airport 4.443e-03 3.485e-04 12.751 < 2e-16 \*\*\*  FreeWifi -9.189e-02 2.625e-02 -3.500 0.000466 \*\*\*  FreeBreakfast 1.725e-01 1.444e-02 11.950 < 2e-16 \*\*\*  HotelCapacity -6.891e-05 1.205e-04 -0.572 0.567333  HasSwimmingPool 5.029e-01 1.864e-02 26.972 < 2e-16 \*\*\*  StarRating 6.362e-01 1.293e-02 49.222 < 2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.7729 on 13220 degrees of freedom  Multiple R-squared: 0.4378, Adjusted R-squared: 0.4373  F-statistic: 935.7 on 11 and 13220 DF, p-value: < 2.2e-16 |
|  |
| |  | | --- | | > | |

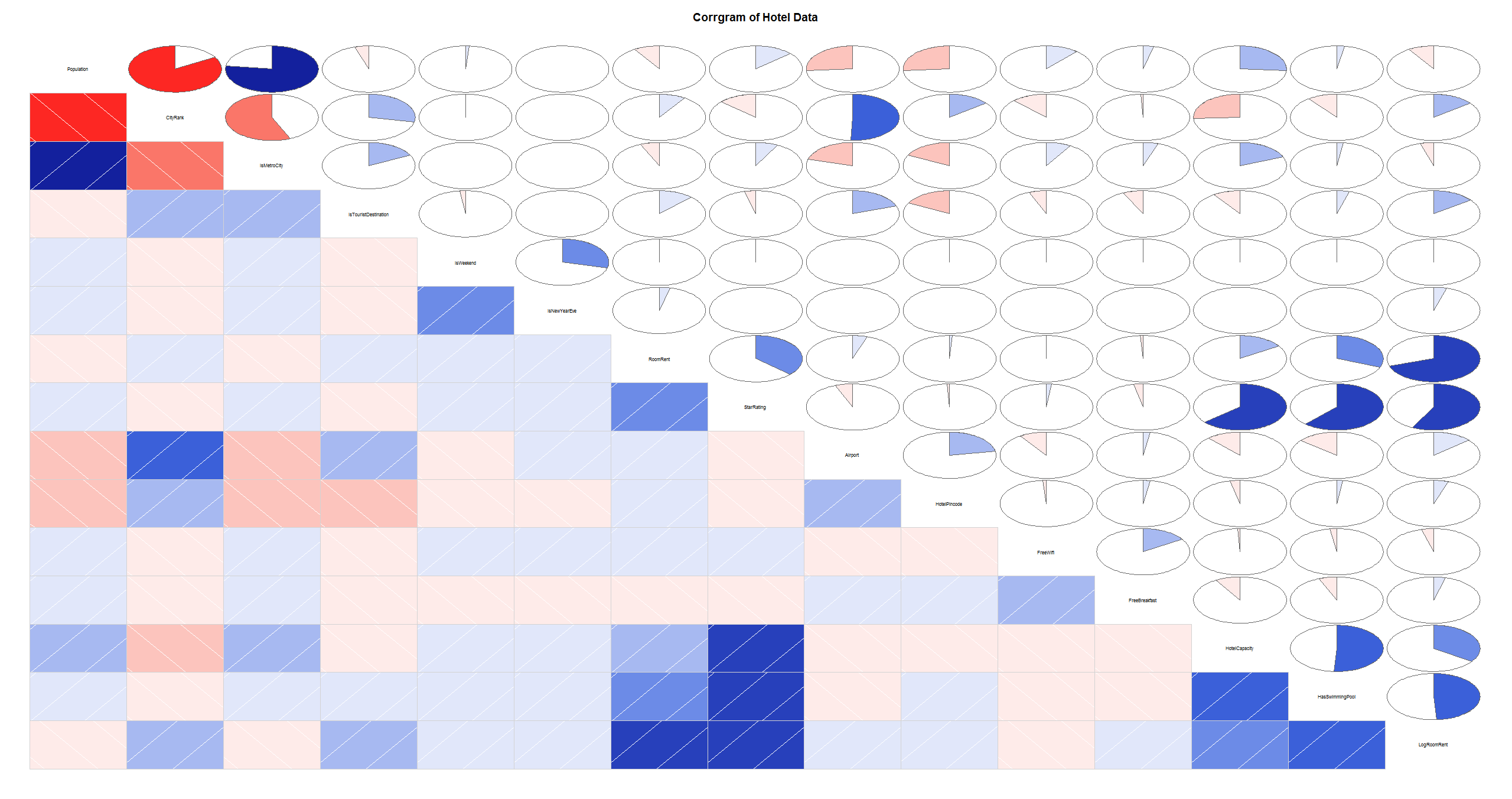
**2.4.2 Linear Model 2(LM2)**

From the output of Linear Model 1 we see that New Year Eve, City Rank, Tourist Destination, Airport, Free Breakfast, Hotel Capacity, Has Swimming Pool and Star Rating play a significant role. Hence, we build our second linear model using only these variables.

Result: Since the P-Value of the F-Statistic is less than 2.2e-16, we can reject the Null hypothesis that the Intercept only model is a better fit. Instead we have strong evidence to suggest Linear Model 2 is a better fit compared to an intercept only model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | > lm2 <- lm(logRoomRent ~ IsNewYearEve+CityRank+IsTouristDestination  + +Airport+FreeBreakfast+HotelCapacity+HasSwimmingPool  + +FreeWifi+StarRating,data = hotels.df)  > summary(lm2)  Call:  lm(formula = logRoomRent ~ IsNewYearEve + CityRank + IsTouristDestination +  Airport + FreeBreakfast + HotelCapacity + HasSwimmingPool +  FreeWifi + StarRating, data = hotels.df)  Residuals:  Min 1Q Median 3Q Max  -2.9426 -0.5068 -0.0524 0.4264 5.2062  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 9.147e+00 4.787e-02 191.079 < 2e-16 \*\*\*  IsNewYearEve 1.417e-01 2.036e-02 6.961 3.55e-12 \*\*\*  CityRank 1.155e-02 6.088e-04 18.978 < 2e-16 \*\*\*  IsTouristDestination 2.217e-01 1.542e-02 14.376 < 2e-16 \*\*\*  Airport 4.433e-03 3.476e-04 12.754 < 2e-16 \*\*\*  FreeBreakfast 1.722e-01 1.437e-02 11.981 < 2e-16 \*\*\*  HotelCapacity -7.393e-05 1.196e-04 -0.618 0.536598  HasSwimmingPool 5.036e-01 1.852e-02 27.192 < 2e-16 \*\*\*  FreeWifi -9.224e-02 2.624e-02 -3.515 0.000441 \*\*\*  StarRating 6.362e-01 1.293e-02 49.220 < 2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.7729 on 13222 degrees of freedom  Multiple R-squared: 0.4377, Adjusted R-squared: 0.4373  F-statistic: 1143 on 9 and 13222 DF, p-value: < 2.2e-16 | |  | | |  | | --- | | > | | |
|  |
|  |

1. **Conclusion**

Weekends, New Year Eve, Rank of the City and in turn its population, whether the City is a tourist destination, distance from the airport, whether the hotel has a free breakfast service, the hotel capacity, whether the hotel has a swimming pool, the Star Rating all play a role in determining the Room Rent. If we were to build a linear model solely based on the Corrgram the only dependent variables would be Hotel Capacity, Star Rating Data and if the Hotel has a Swimming Pool. This model would have an adjusted R-Value of 0.3636 but Linear Model 2 has a relatively higher adjusted R value making the model more accurate.

1. **Additional Figures and Tables**

**4.1 Tables**

Table 1

**City Rank (based on 2011 City Population)**

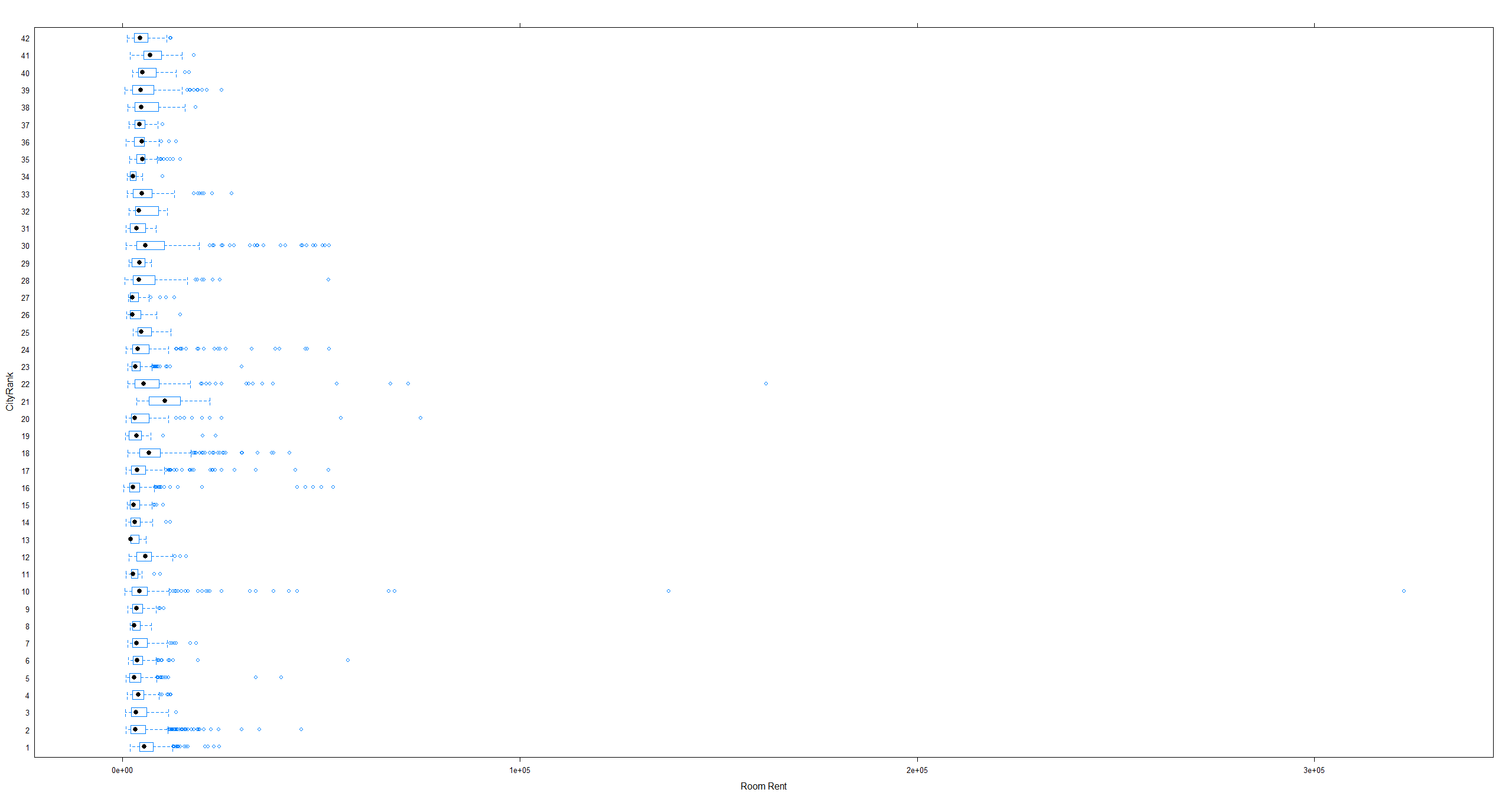
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CITYRANK** | **CITYNAME** | **IsHolidayDestination** | **Number of Hotels listed on Hotels.com** | [**City Population(2011)**](https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population#cite_note-Cities1Lakhandabove-3) |
| 0 | Mumbai | 0 | 405 | 12,442,373 |
| 1 | Delhi | 0 | 871 | 11,034,555 |
| 2 | Bangalore | 0 | 450 | 8,443,675 |
| 3 | Chennai | 0 | 287 | 7,088,000 |
| 4 | Hyderabad | 0 | 237 | 6,731,790 |
| 5 | Ahmedabad | 0 | 136 | 5,577,940 |
| 6 | Kolkata | 0 | 192 | 4,496,694 |
| 7 | Surat | 0 | 20 | 4,467,797 |
| 8 | Pune | 0 | 205 | 3,124,458 |
| 9 | Jaipur | 1 | 286 | 3,046,163 |
| 10 | Thrissur | 0 | 36 | 2,975,440 |
| 11 | Lucknow | 0 | 37 | 2,817,105 |
| 12 | Kanpur | 0 | 13 | 2,765,348 |
| 13 | Amritsar | 1 | 72 | 2,490,891 |
| 14 | Indore | 0 | 49 | 1,960,631 |
| 15 | Agra | 1 | 102 | 1,760,285 |
| 16 | Madurai | 1 | 21 | 1,465,625 |
| 17 | Goa | 1 | 626 | 1,457,723 |
| 18 | Rajkot | 0 | 26 | 1,286,678 |
| 19 | Varanasi | 1 | 60 | 1,201,815 |
| 20 | Srinagar | 1 | 57 | 1,180,570 |
| 21 | Jodhpur | 1 | 81 | 1,033,918 |
| 22 | Chandigarh | 0 | 117 | 960,787 |
| 23 | Thiruvathipuram | 0 | 128 | 957,730 |
| 24 | Guwahati | 0 | 12 | 957,352 |
| 25 | Mysore | 1 | 58 | 887,446 |
| 26 | Bhubaneswar | 0 | 29 | 837,737 |
| 27 | Kochi | 1 | 188 | 595,575 |
| 28 | Mangalore | 0 | 13 | 499,487 |
| 29 | Udaipur | 1 | 113 | 451,735 |
| 30 | Pondicherry | 0 | 42 | 241,773 |
| 31 | Haridwar | 1 | 73 | 228,832 |
| 32 | Puri | 1 | 24 | 201,026 |
| 33 | Shimla | 1 | 58 | 169,578 |
| 34 | Panchkula | 0 | 118 | 140,925 |
| 35 | Darjeeling | 1 | 32 | 132,016 |
| 36 | Rishikesh | 1 | 107 | 102,138 |
| 37 | Gangtok | 1 | 30 | 98,658 |
| 38 | Ooty | 1 | 64 | 88,430 |
| 39 | Jaisalmer | 1 | 82 | 65,471 |
| 40 | Nainital | 1 | 85 | 41,377 |
| 41 | Munnar | 1 | 108 | 38,471 |
| 42 | Manali | 1 | 80 | 8,096 |

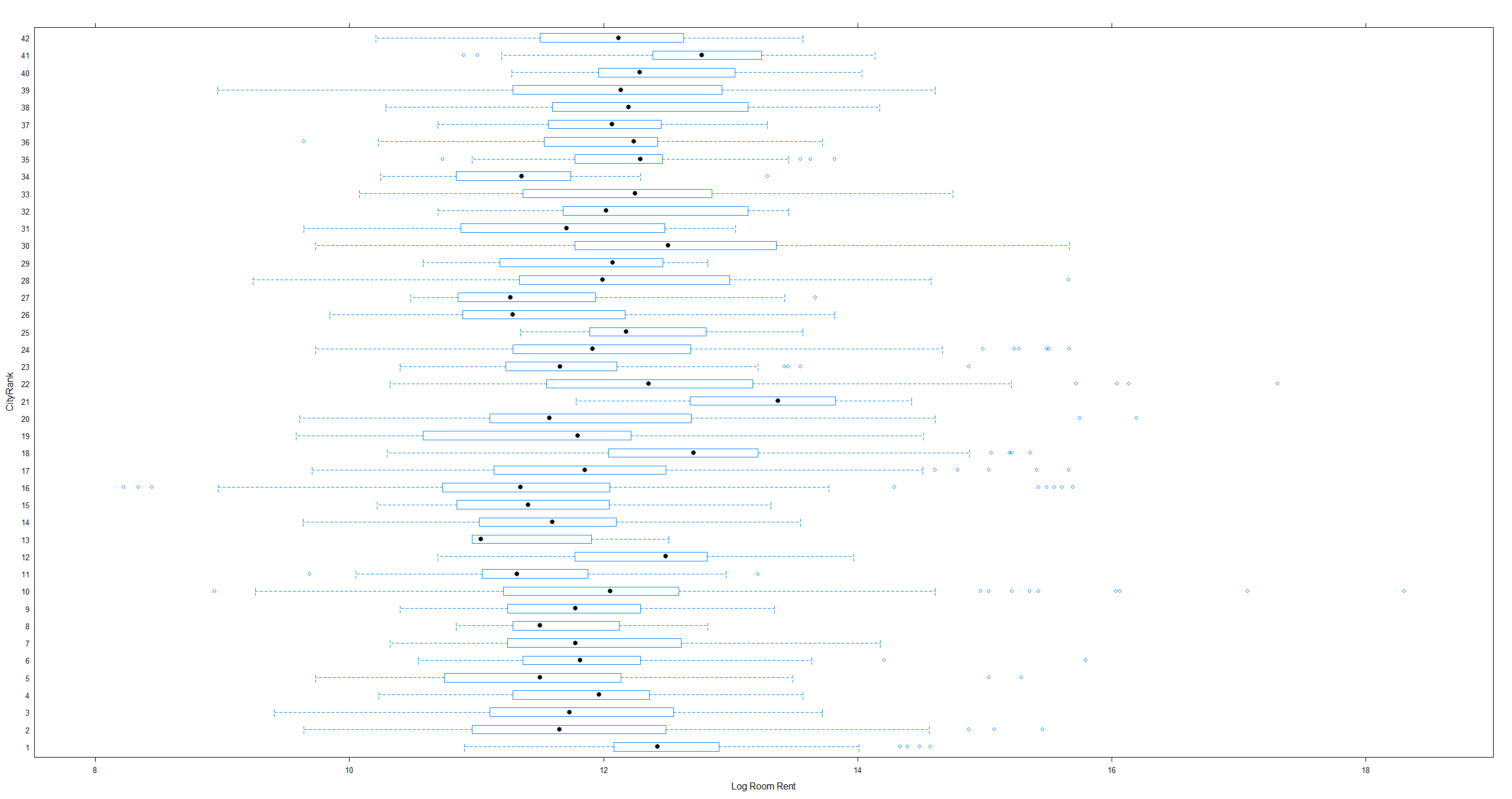
Table 2

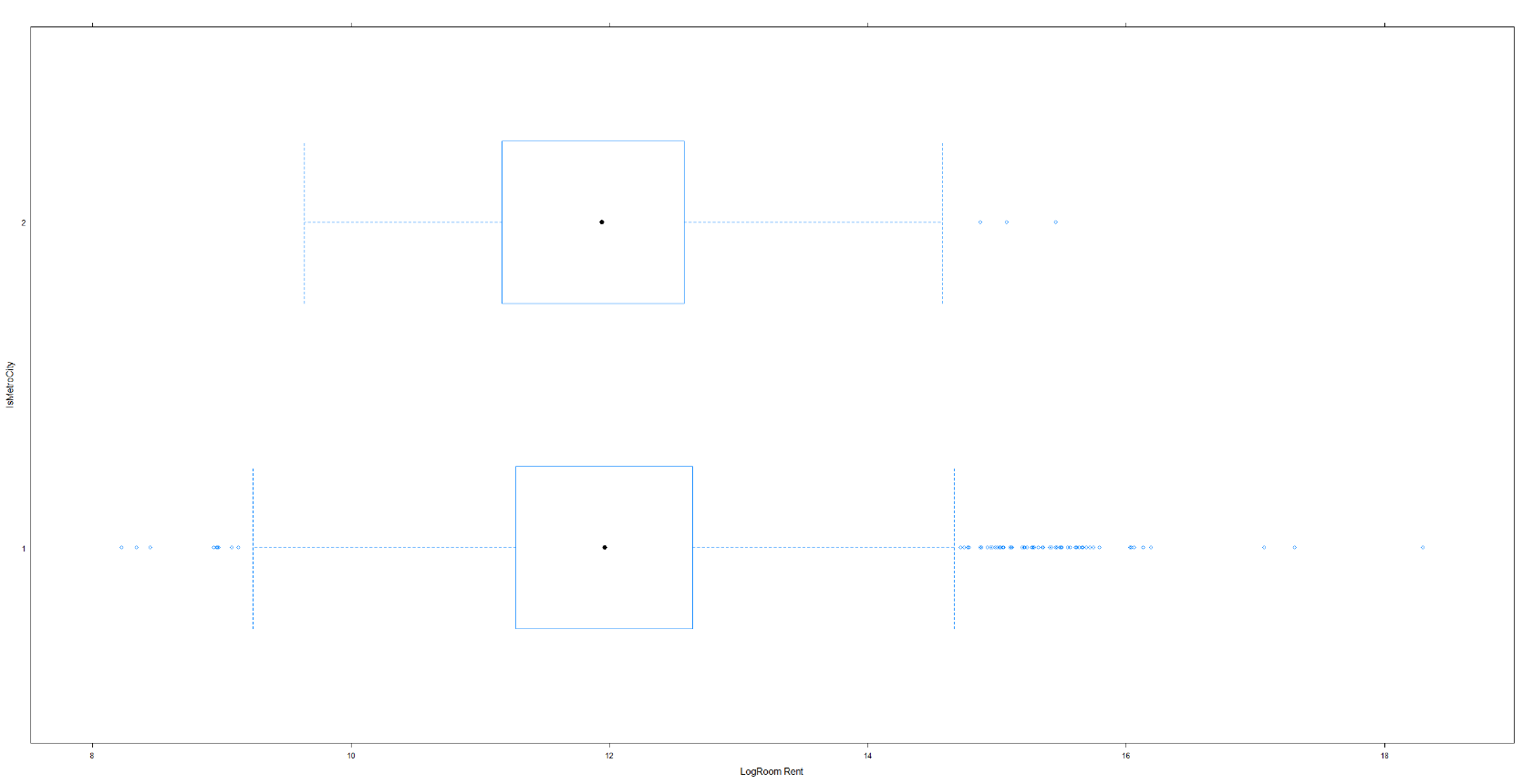
|  |  |  |
| --- | --- | --- |
| **VARIABLE** | **UNITS** | **MEANING** |
| **Date** | Text | We have hotel room rent data for the following 8 dates for each hotel:  {Dec 31, Dec 25, Dec 24, Dec 18, Dec 21, Dec 28, Jan 4, Jan 8}  If a hotel is sold out on a given date, assume that the price of the hotel room on the date it is sold out is the maximum price from the sample of dates for which prices are available. |
| **IsWeekend** | Dummy | We use ‘0’ to indicate week days, ‘1’ to indicate weekend dates (Sat / Sun) |
| **IsNewYearEve** | Dummy | ‘1’ for Dec 31, ‘0’ otherwise |
|  |  |  |
| **CityName** | Text | Name of the City where the Hotel is located   e.g. Mumbai` |
| **Population** | Number | Population of the City in 2011 (See Table 1 above) |
| **CityRank** | Dummy | Rank order of City by Population (e.g. Mumbai = 0, Delhi = 1, so on); (See Table A1) |
| **IsMetroCity** | Dummy | ‘1’ if CityName is {Mumbai, Delhi, Kolkatta, Chennai}, ‘0’ otherwise |
|  |  |  |
| **IsTouristDestination** | Dummy | We use ‘1’ if the city is primarily a tourist destination, ‘0’ otherwise. For example, Goa and Agra are primarily tourist destinations. We assume that most people who visit Goa and Agra and stay in their hotels are in these cities primarily for tourism. |

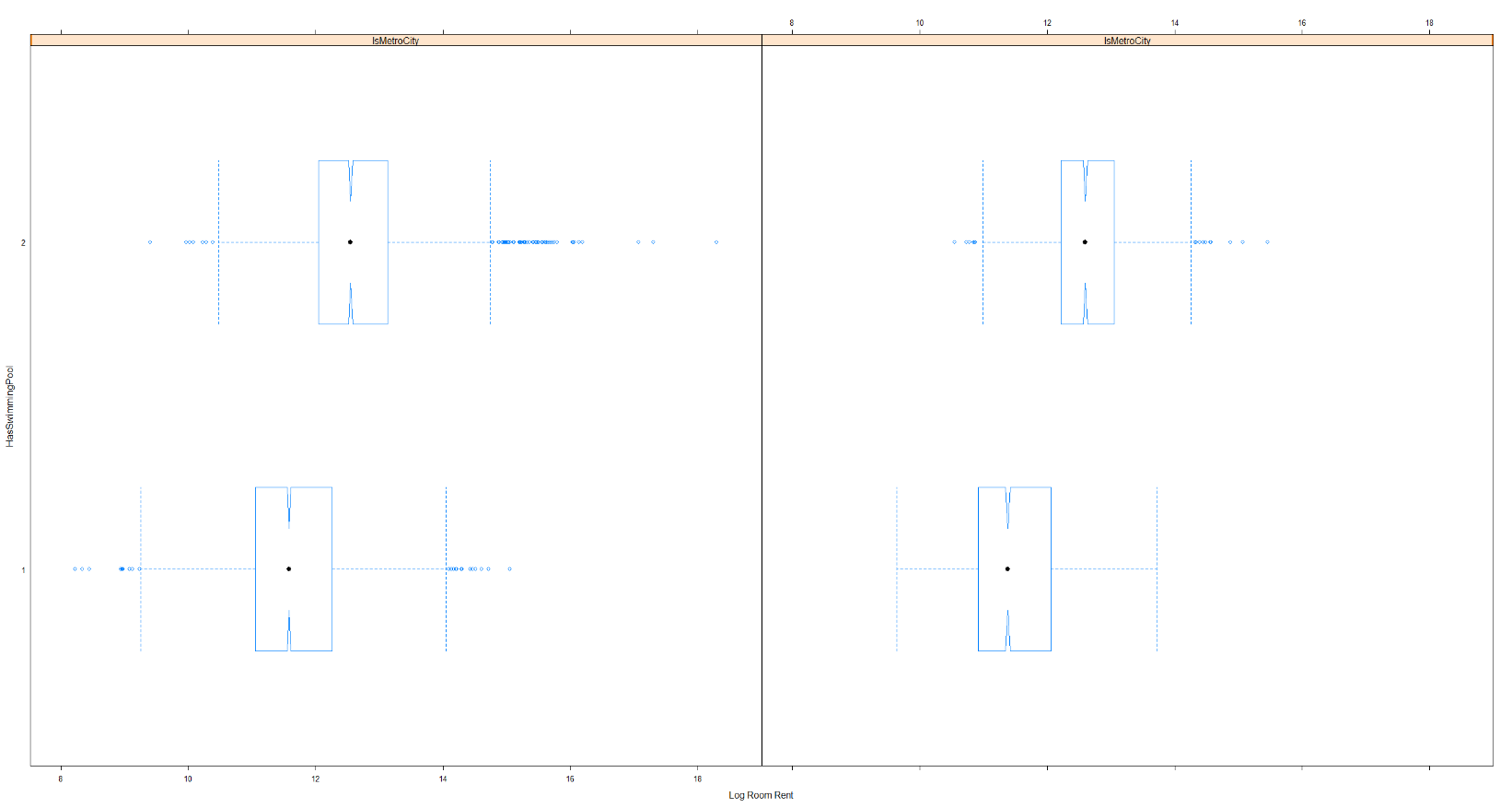
|  |  |  |
| --- | --- | --- |
| **VARIABLE** | **UNITS** | **MEANING** |
| **HotelName** | Text | e.g. Park Hyatt Goa Resort and Spa |
| **StarRating** | Number | e.g. 5 |
| **Airport** | km | Distance between Hotel and closest major Airport |
| **HotelAddress** | Text | e.g. Arrossim Beach, Cansaulim, Goa |
| **HotelPincode** | Number | 403712 |
| **HotelDescription** | Text | e.g. 5-star beachfront resort with spa, near Arossim Beach |
| **FreeWifi** | Dummy | ‘1’ if the hotel offers Free Wifi, ‘0’ otherwise |
| **FreeBreakfast** | Dummy | ‘1’ if the hotel offers Free Breakfast, ‘0’ otherwise |
| **HotelCapacity** | Number | e.g. 242.  (enter ‘0’ if not available) |
| **HasSwimmingPool** | Dummy | ‘1’ if they have a swimming pool, ‘0’ otherwise |

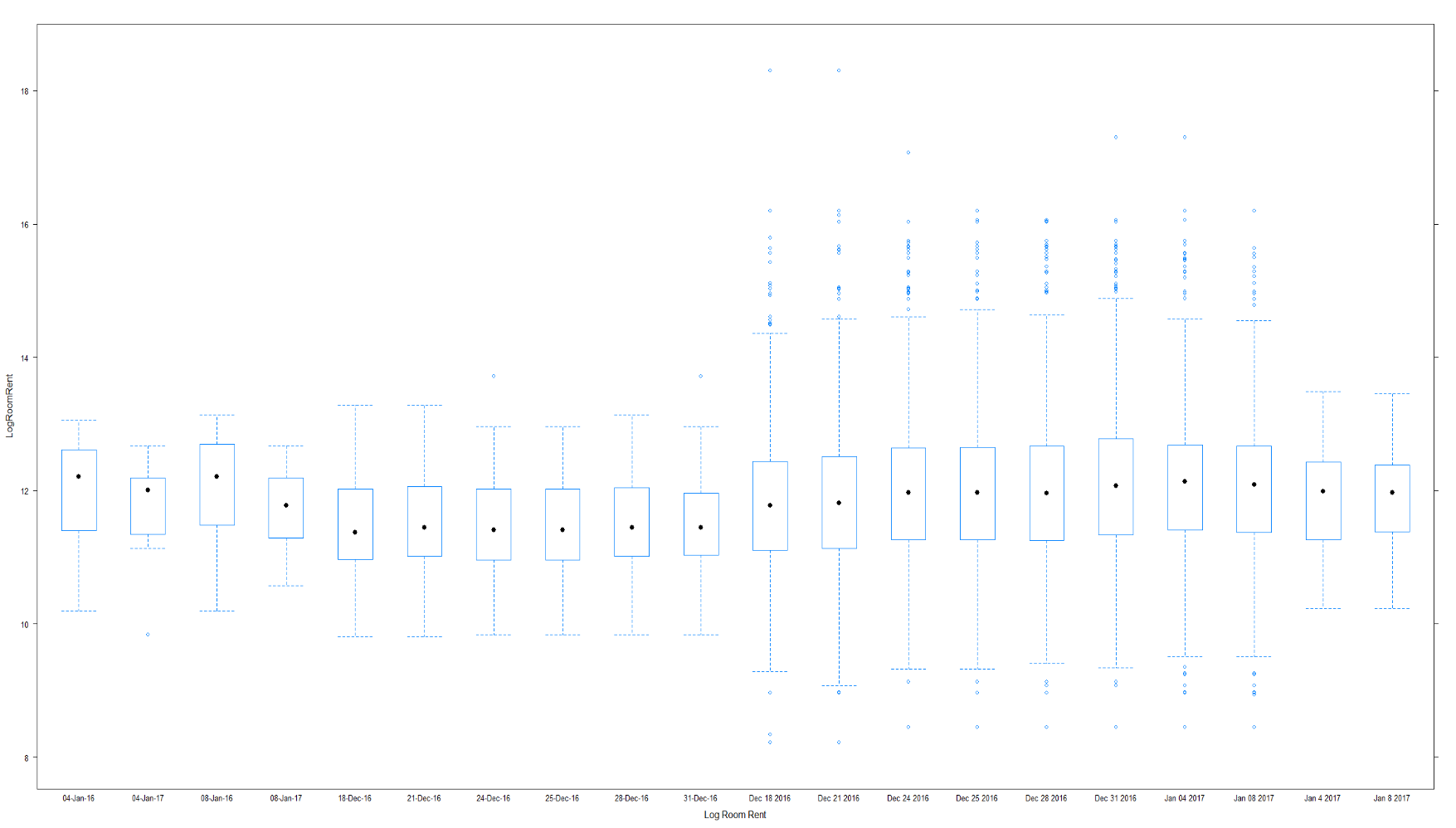
**4.2 Figures**

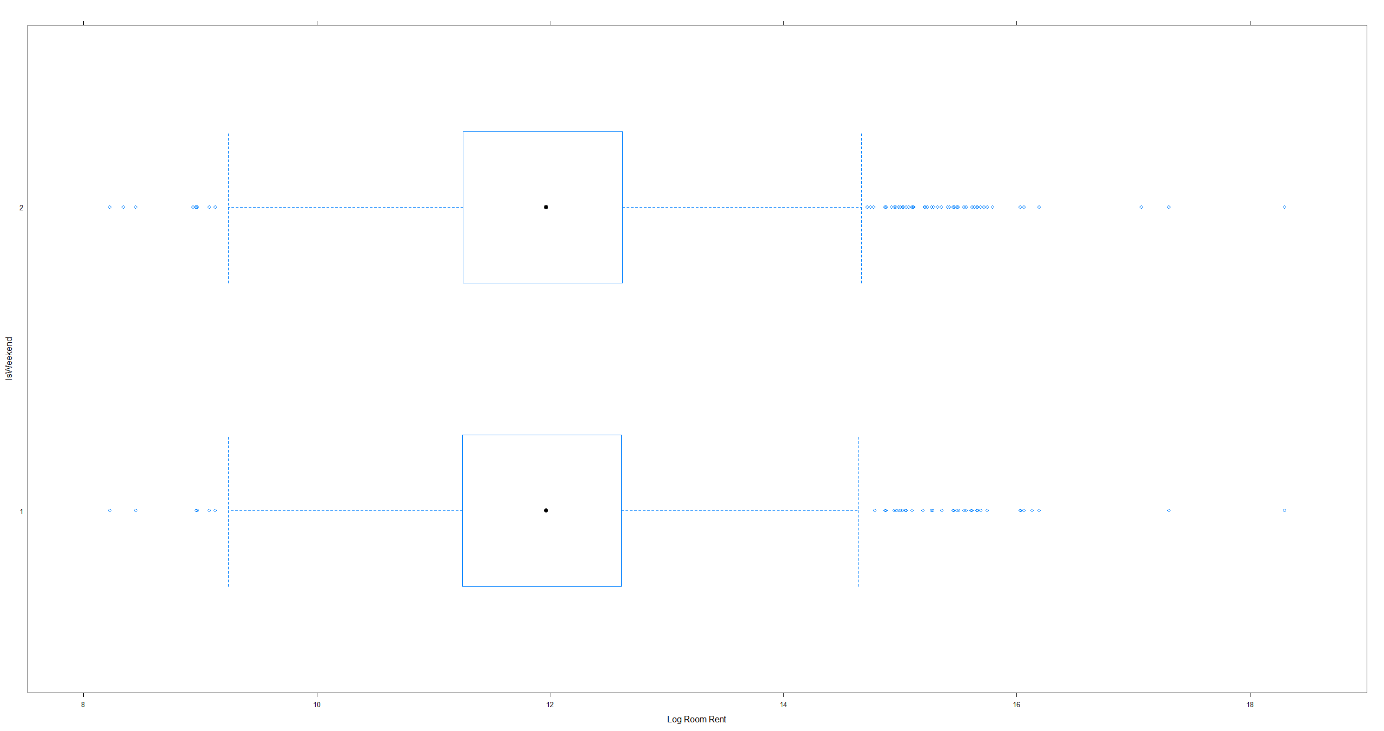
Figure 1

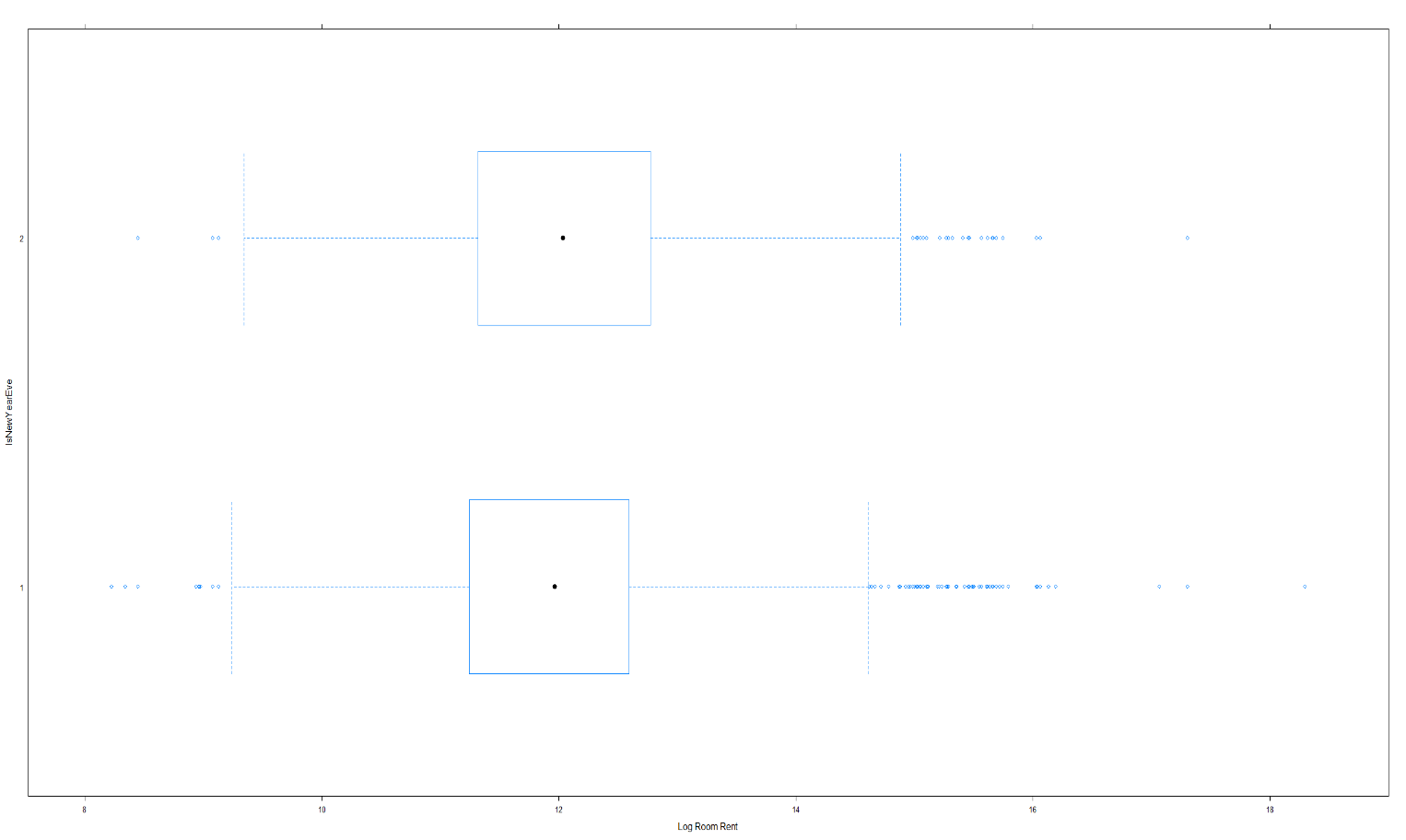
Figure 2

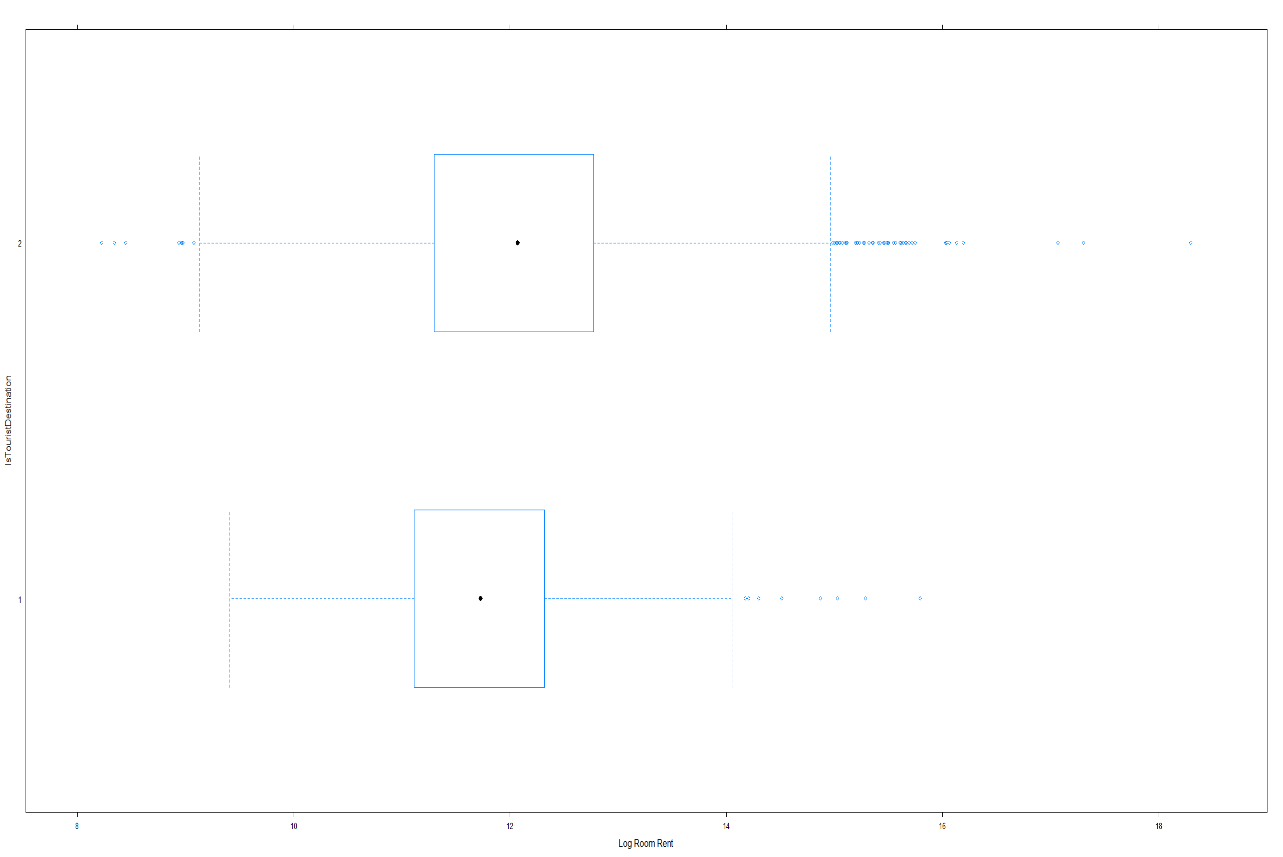
Figure 3

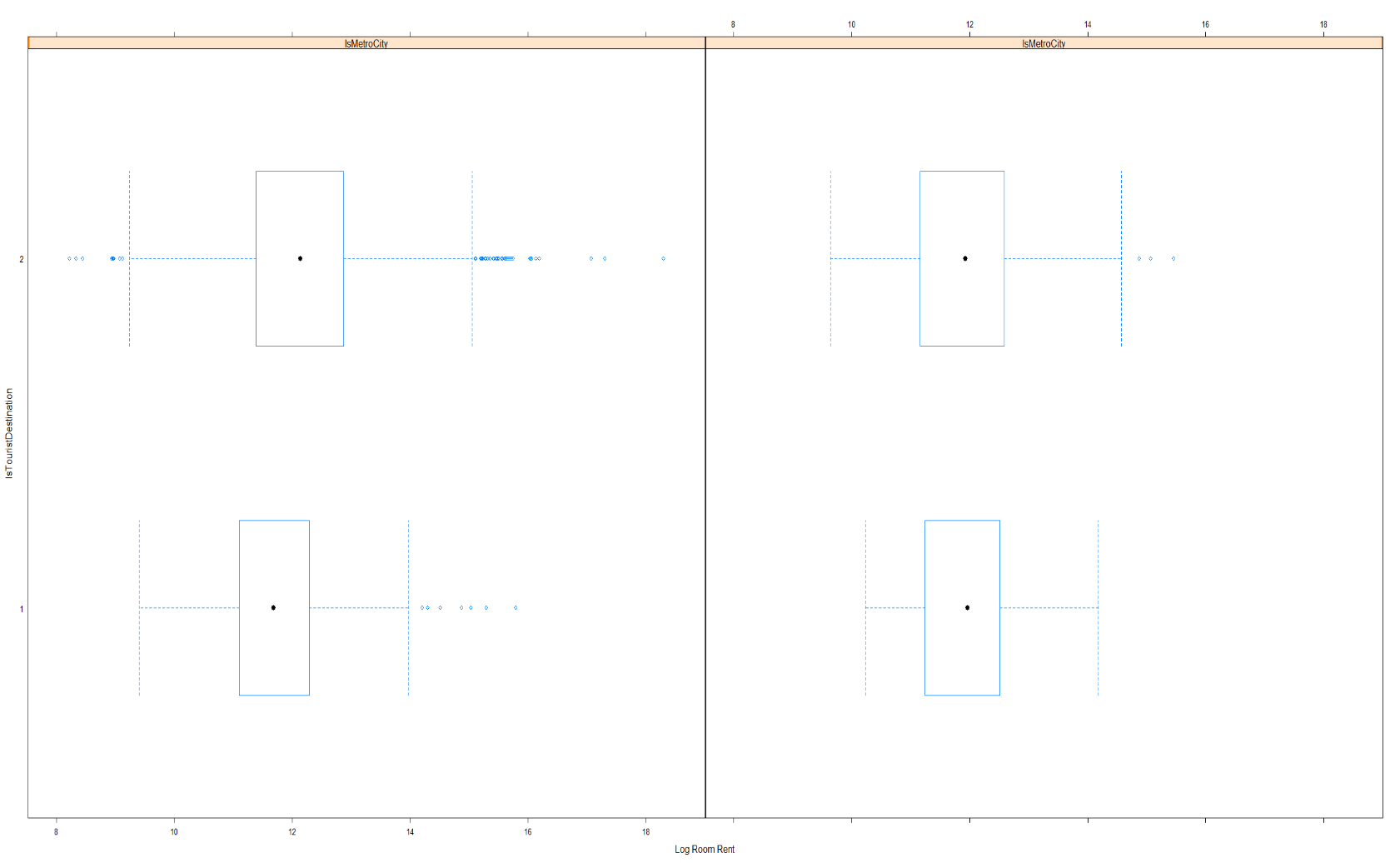
Figure 4

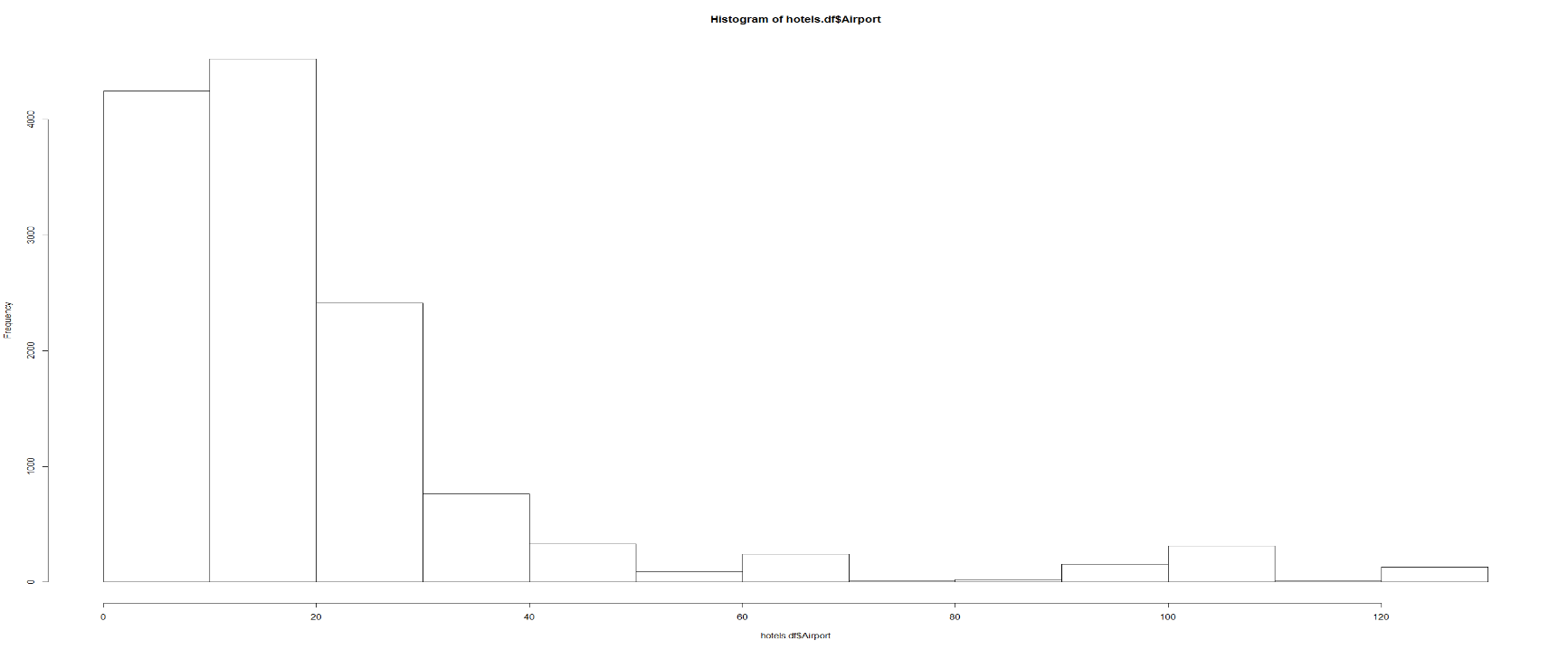
Figure 5

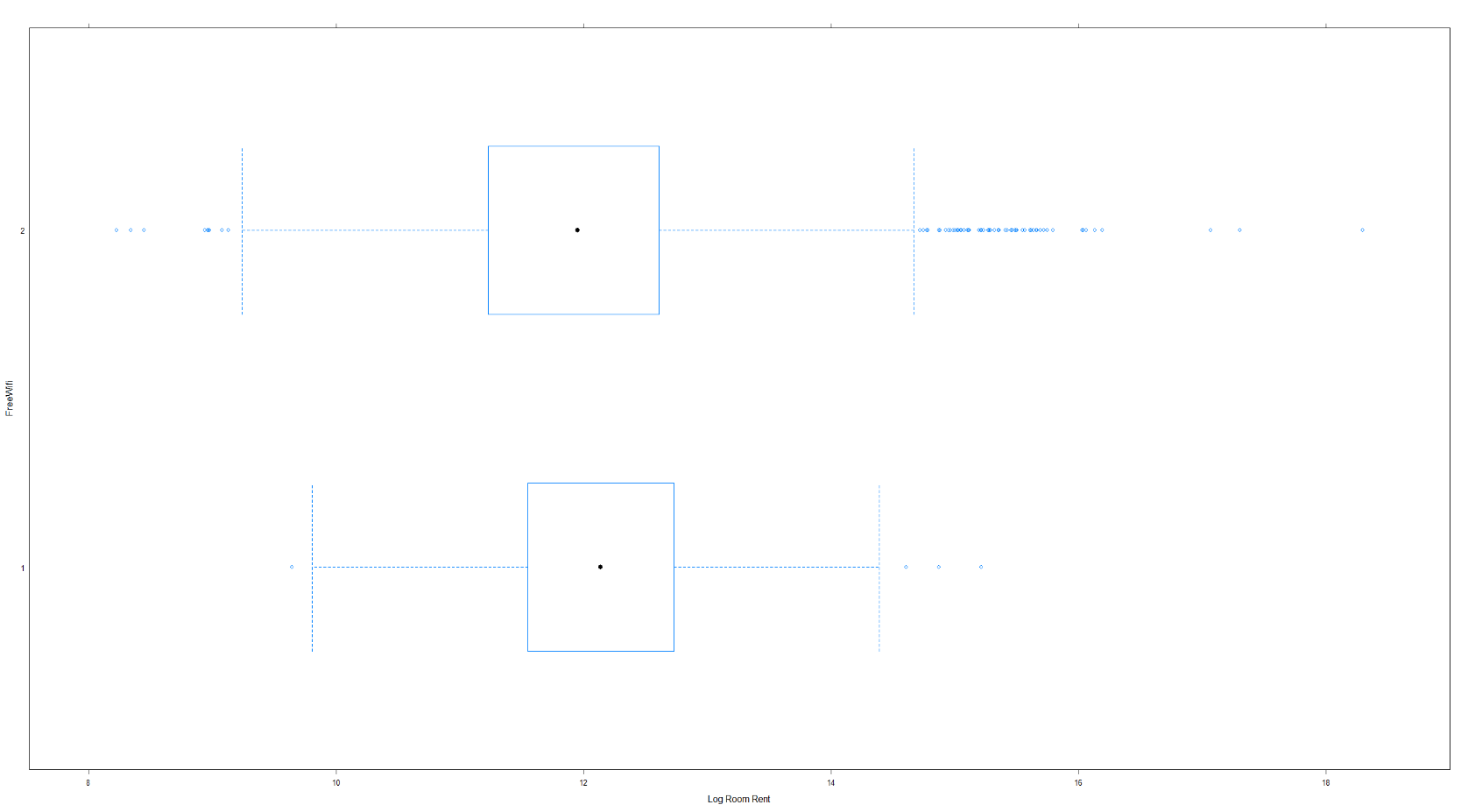
Figure 6

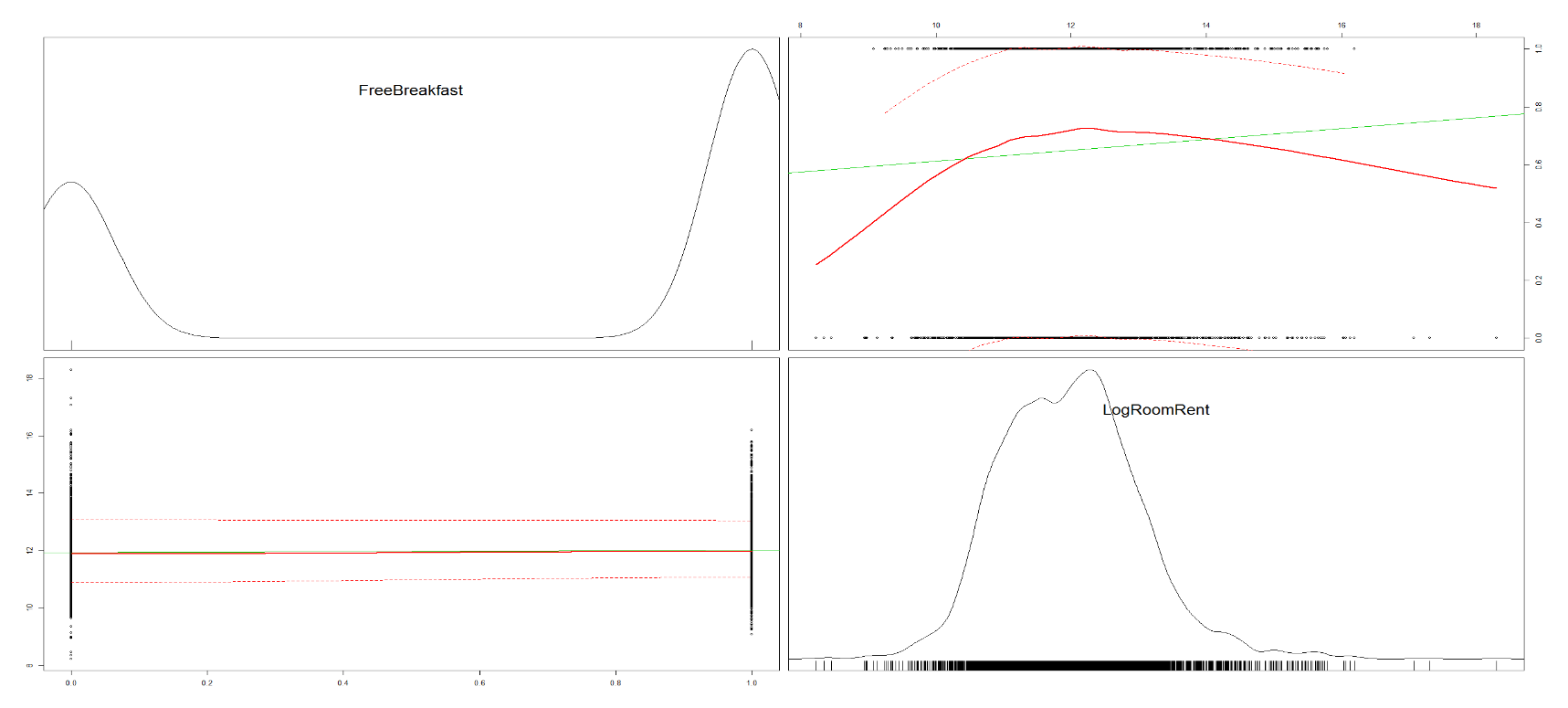
Figure 7

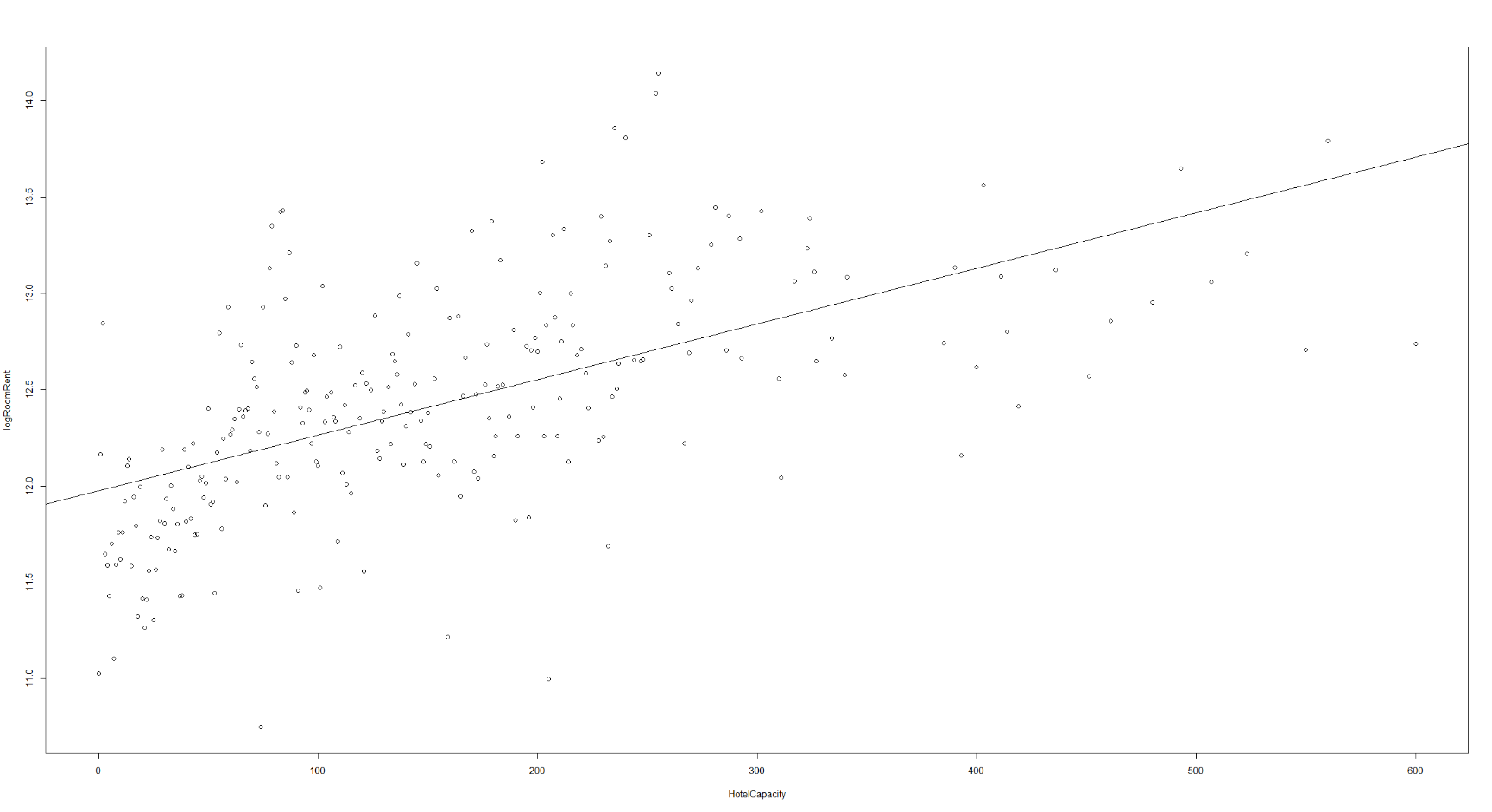
Figure 8

Figure 9

Figure 10

Figure 11

Figure 12

Figure 13

**5.0 References**

India Brand Equity Foundation (IBEF) (2008a), *India at a Glance. An Initiative of the Ministry of Commerce and Industry*, Government of India, New Delhi.

India Brand Equity Foundation (IBEF) (2008b), *Tourism and Hospitality Industry. An Initiative of the Ministry of Commerce and Industry*, Government of India, New Delhi.

Mohsin, Asad, and Tim Lockyer. "Customer perceptions of service quality in luxury hotels in New Delhi, India: an exploratory study." *International Journal of Contemporary Hospitality Management* 22.2 (2010): 160-173.

Dwivedi, Mridula, T. P. Shibu, and Umashankar Venkatesh. "Social software practices on the Internet: Implications for the hotel industry." *International journal of contemporary hospitality management* 19.5 (2007): 415-426.