**Analysis on Room Rent in hotels of various cities in India**

**1.INTRODUCTION**

This project is related to the analysis of factors that affect the room rents in hotel. As a common man, one would think that factors such as the rating of Hotels or the amenities and services have a great effect on the Hotel Room Rent. Therefore, an analysis was done to check which all factors would affect the most on the Room Rent.

These factors can be internal and external factors. External factors can be the population of city or whether data was collected on New Years Eve. It is also believed that places of tourist destination would have hotels of a higher room rent than with the hotels which are not tourist destinations.

Internal factors are those which affect the hotel room rent such as the various amenities such as whether the hotel has a swimming pool or whether the hotel have breakfast as complimentary.

The report is based on data collected from over 12000 hotels from 42 cities in India. The data was collected from range of Dec 2016 to Jan 2017.The dates that have been collected were off days and peak days such as Christmas Eve and New Year’s Eve.

After collection of data it is quite important to look for elements that can affect the factor in question.

A regression model was created that for predicting the value of Room Rent of the hotels if the factors are given. After the model was created it was check for residuals and plotting of the model for various factors were checked for the authenticity of the model.

For every plots or visual representations that were made was the interpretations and whether these interpretations were actually true is shown by the t-tests.

**2. STUDY OF HOTEL PRICING STRATEGY IN INDIA**

**2.1 HYPOTHESIS**

India’s tourism is an important source of India’s GDP growth. But this also affects the hotel pricing strategy in India too. The data that was collected had an approximate of 18 factors out of which 3 factors were considered as the most influencing factor.

Although this is not quite logical as defining a hotel rent based on just 3 factors has a lower possibility. But a model was created that was similar to the one where when inputs are provided the hotel rent was given therefore a closer looking model needs to be created.

Therefore for the model to work one needs 3 assumptions and then take an analysis on them to display their true meaning.

In our case it was considered that these 3 factors may affect the price of Room Rent the most.

1. **Hotel Rating**: Hotel Rating in this data is a continuous variable that gives rating to Hotels based on their hospitality, amenities, and various other factors that would affect a lot .Hotel Rating was given from 1-5 and there were given in points.5 was considered the highest. The assumption taken was if the hotel rating is high does it affect the room rent of hotels.

Secondly, if the lower rating has high room rents what are the advantages are provided to that set such that they had high rating. The flow of people also drives the Hotel Rating as it matters on what time people visit frequently to hotel. A service-less hotel usually have a lower rating than a hotel where amenities are good and hospitality awesome.

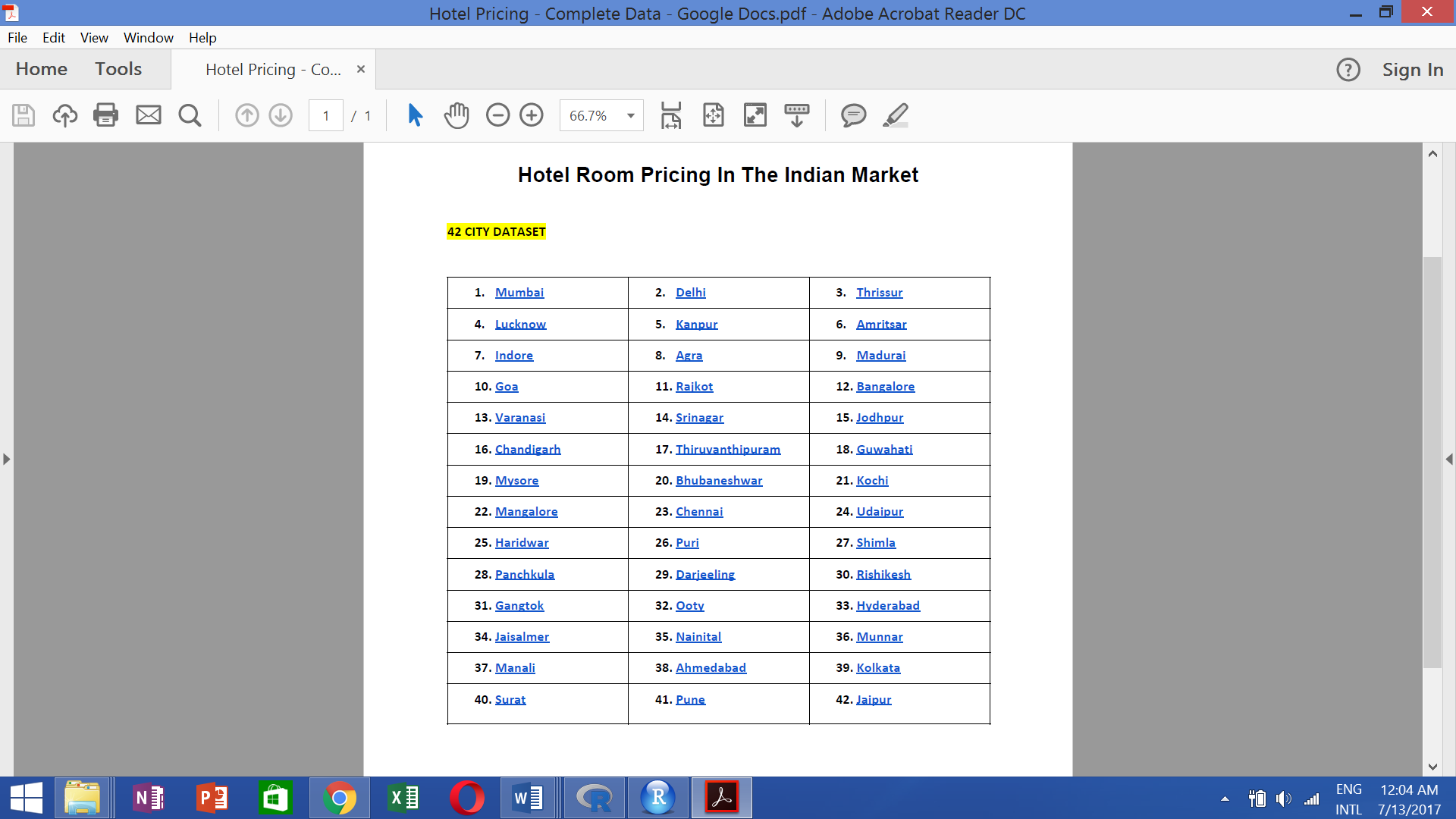
1. **Airport Distance**: Airport Distance can be considered one of the factors for the change in the room rents of the hotels because it is preferred that hotels that are nearer to the hotel has all the facilities so people who live for short durations can have easier transport to the airport. If an airport is far away from the hotel what would be factors would be available that makes people drive to that hotel. A hotel that is nearer to the airport is a boon for working professionals as it is usually a one or 2 day trip. As dates were considered when it is both off-season and on-season travels.
2. **Tourist Destination**: A hotel which is available at tourist places have high rates when it is peak time of travel such as Christmas Eve or New Years Eve. It was expected that most of the tourists would be travelling during this time and staying in these hotels. These tourist destinations plus the services they provide such has complimentary breakfast or free Wifi or whether these hotels have a swimming pool or not is a factor can be a plus point to these destinations.

H1: The Room Rent of Hotels are higher when it is a tourist destination.

**2.2 DATA**

This project displays the strategy of hotels in the perspective of pricing. Room Rent is quite varied in every area and it is quite important that these data do create a meaning to the effect on the Room Rent. There is approximately 17 variables that consists of both internal factors and external factors.

The data set is based on 42 cities in India. Each of these have hotels that have ratings ranging from 1-5 and with that each has several factors to be considered. Below is the collection of each city from which data is collected.

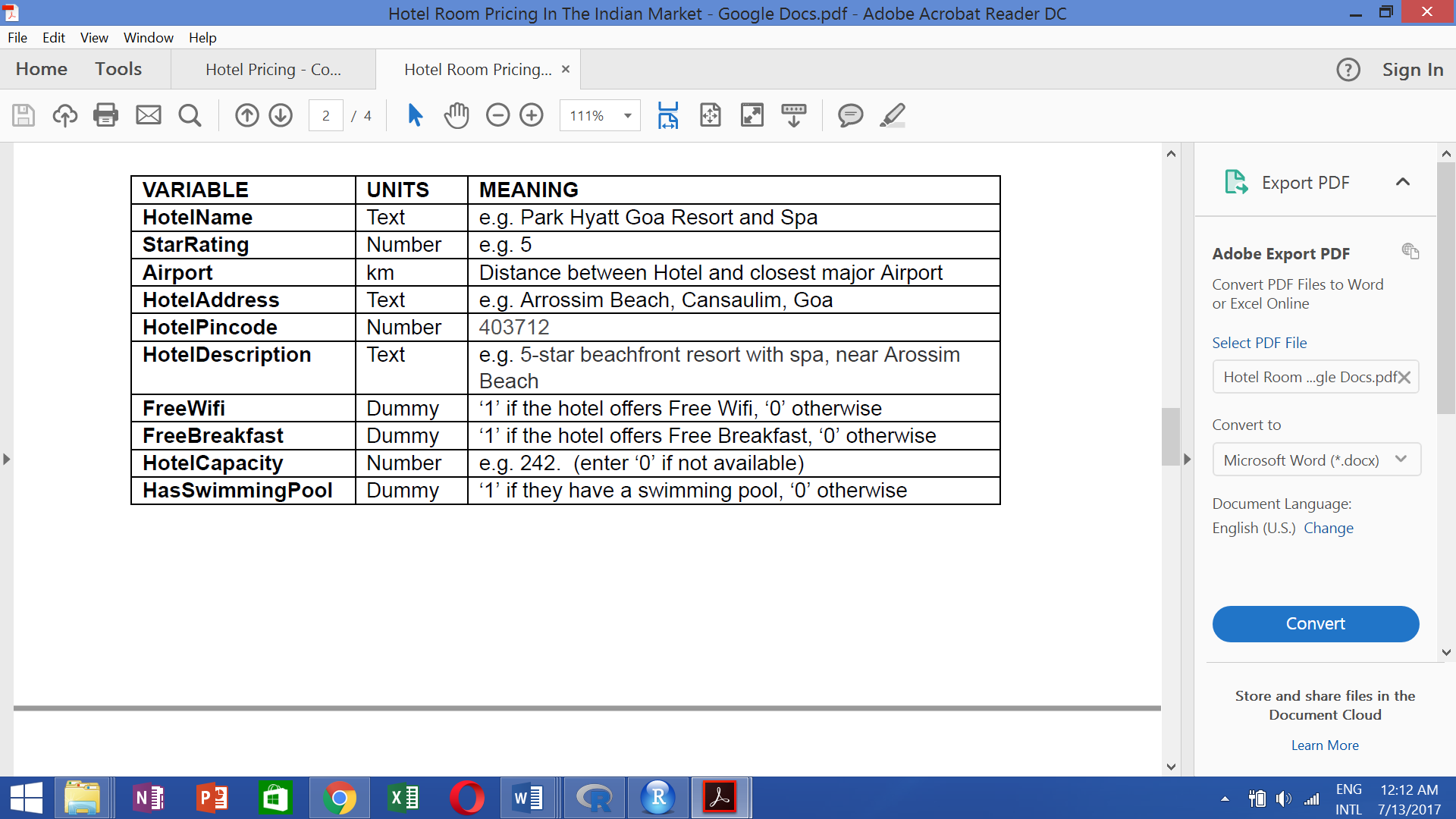


The factors that these data are based on are of two types: **Internal Factors** and **External Factors**

**Internal Factors** are those factors that are the amenities that are present or available in the hotel such as complimentary breakfast or whether these hotels have a swimming pool or not.

**External Factors** are those that are not caused by the hotels but by for example whether it was a New Years Eve etc.

Internal Factors: Each Factor has its own meaning and can affect the data highly



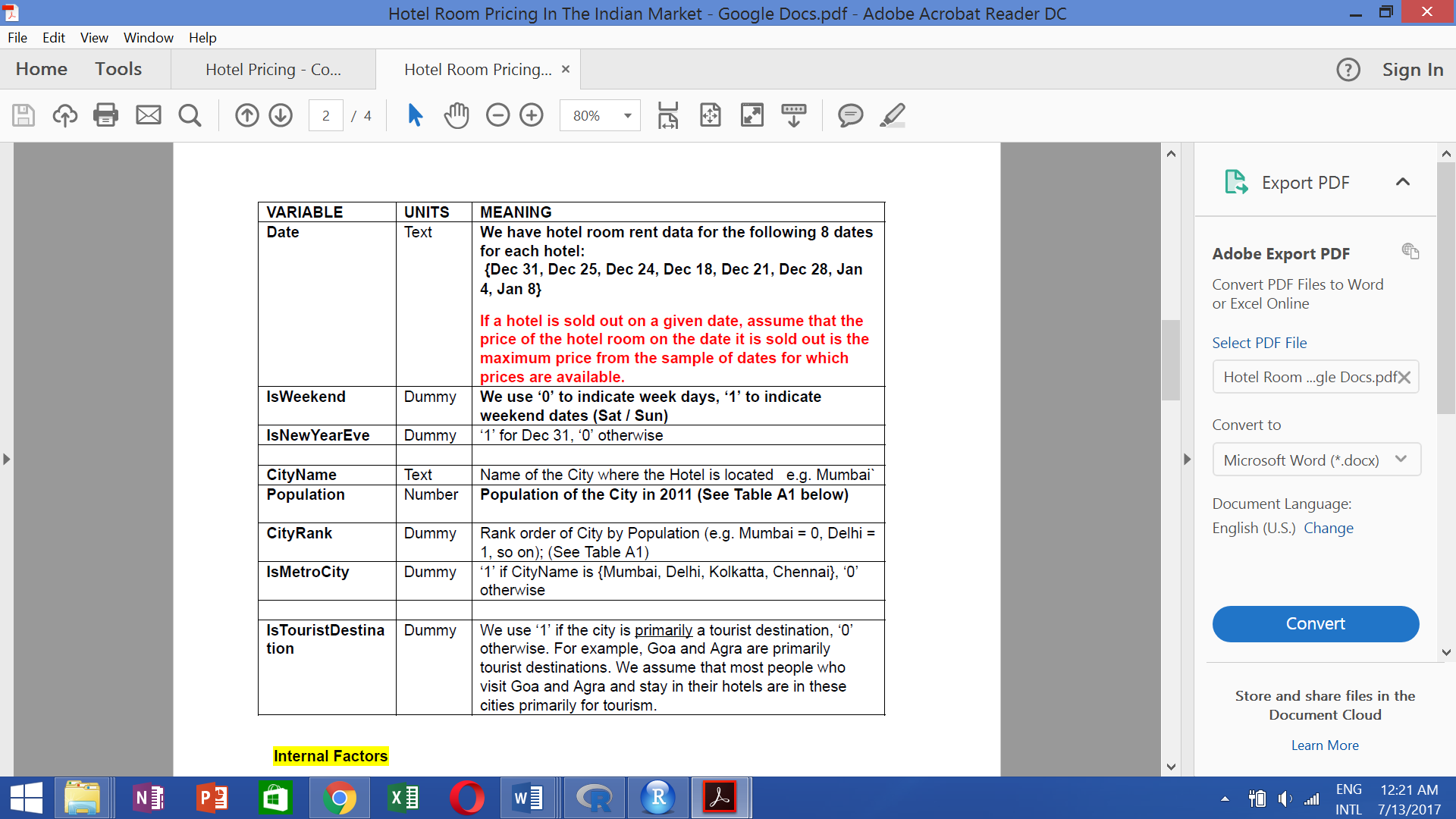
Factors like Hotel Name,Hotel Address,Hotel Description and Hotel Pincode are factors that we are assuming may not be influencing to the data available.

As For the Hotel Name,it matters that the hotel rating and with that the hotel also becomes famous.

Hotel Address may not be needed as airport distance can be a better factor for the analysis hence was not considered. Similar could be thought of Pincode.

Instead of Hotel Description factors such as Star Rating,Airport distance,Wifi requirement, Hotel Capacity and availability of swimming pools was considered better variables as most of these variables are continuous.

External Factors: These factors can be considered as important too.



**2.2 MODEL**

The model that was created were of types:-

1. A model with 3 factors in assumptions
2. A model with all factors taken in excluding variables that contained text and information on the data.

The model that has been created using the linear regression model. Linear Regression is part of supervised learning. In a simple regression model there are only 2 factors similar to a linear equation where one variable is an intercept and the other is the variable with the slope that provides the predictor variable.  A linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression.

Further going in, a multiple regression model is quite complex as compared to the linear regression model it depends on varied factors. In the end there is an end term known as the residual error which contains the error on how far the equation is away from the data.

Having 3 variables helps to rule out the correlation data that affects the data very much and results in error prone R^2 hence is not a true representation of the data.

Linear Model 1 created in R:

lm(formula = RoomRent ~ StarRating + Airport + IsTouristDestination,   
 data = HotelData.df)

Linear Model 2 created in R:

lm(formula = RoomRent ~ Population + IsMetroCity + IsTouristDestination +   
 IsNewYearEve + StarRating + Airport + FreeWifi + HotelCapacity +   
 HasSwimmingPool, data = HotelData.df)

1. **ANALYSIS REPORT**

Each of the analysis of the data that was collected from the dataset of the hotels of 42 cities. An analysis was done taking each factor in question.

1. Population

Population had a varied data with population going as high as 12442373 in Mumbai to the 8096 in Manali. This shows that the data that has been collected from varies regions and the data would be collected in that way. Hence ,the uniformity in data is reduced.

1. Room Rent

The predictor variable and the y part in the function scenario. If one considers the outliers too the maximum room rent was 322500 at RamBagh Palace to as minimum as 299 at Backpacker Panda.

1. Airport Distance

Airport distance varied from 0.20 km to 124 km.Analysing the room rent was done further afterwards.

1. Individual Variables  
    (a)Histogram of Population   
    Most cities have population in the range of 0-4000 and only few cities have population over 12000---These could be metropolitan cities   
   (b)Histogram of Airport Distance  
    Positively skewed graph with most of the hotels in the range of 0-40 km  
    (c)Histogram of HotelCapacity  
    Positively skewed graph with capacity of hotels in the range of 0-300  
    (d)Histogram of StarRating of hotels  
    Most of hotel ratings are in the range of 2.5-3.5
2. Comparison of Function Variable to other variables  
    (a)Boxplot of Room Rent and Star Rating of Hotel  
    Room Rent Increases with Increase in Star Rating  
    (b)Box Plot of Room Rent and whether city is a Metro City  
    Non Metro cities have a higher room rent than Metro Cities  
    (c)Boxplot of Room Rent and whether the city is a tourist destination   
    Those with tourist destinations have higher room rents than those which are not tourist destination  
    (d)Boxplot of Room Rent and whether it was New Year's Eve  
    No considerable difference in room rent and New Year's Eve  
    (e)Boxplot of Room Rent and whether hotel has a free Wi-fi  
    Room Rent are higher when Free Wi-fi is available  
    (f)Boxplot of Room Rent and Whether hotel has swimming pool  
    Room Rents are considerably high when Swimming pools are available  
    (g)Boxplot of Room Rent and Whether hotel provides breakfast  
    There was no considerable difference in hotel rent and breakfast barring one or two outliers
3. Plots for continuous variables  
   (a)Plot for Room Rent and Hotel Capacity  
    Most of the hotel room rents are higher when hotel capacity is less than 300.Slow increase in room rent   
    (b)Plot for Room Rent and Star Rating  
    Considerable Increase in the rates of Hotel rent as the ratings are increased  
    (c)Plot for Room Rent and Airport Distance  
    Slight Increase in the difference of Room rents and distance of hotel from airport
4. A value that reaches near correlation value 1 has positive correlation while that has value going towards -1 has negative value.  
    (a)Star Rating has a correlation value of 0.60 with Hotel Rating   
    (b)Star Rating has a correlation value of 0.62 with HasSwimmingPool  
    (c)Hotel Capacity has a correlation value of 0.51 with HasSwimmingPool  
    (d)Rest have low correlation values
5. Variables taken  
    StarRating,IsTouristDestination,Airport  
    Experimenting the variables with the RoomRent  
    Inferences: (a)All the correlations were not too high as they reduced homoscedasticity  
    (b)The pairs function showed interesting plots with respect to the RoomRent as only IsTouristDestination contained the categorical variable rest both of them were continuous
6. T test Results

(1)T test for when the average of Room Rent is higher when It is a tourist Destination

H0: *The average room rent is equal when it is a tourist destination.*

H1: *The average room rent is higher when it is a tourist destination*

RESULT: The null hypothesis was rejected and H1 hypothesis was accepted and the average of room rent did matter when it was a tourist destination

(2)T test for when the average of Room Rent is higher when Hotel has a swimming pool

H0: *The average room rent is equal when it has a swimming pool*

H1:*The average room rent is higher when it has a swimming pool*

RESULT: The null hypothesis was rejected and H1 hypothesis was accepted and the average room rent is higher when the hotel has a swimming pool

1. Showing results when average of room rent was achieved when different factors

(a)The Room Rent doesn't quite depend on the StarRating linearly but a 4.8 rating starred hotel has average of 46000 Room Rent  
(b)Rates are quite high when it is closer ranging from 3000-24000  
(c)Room Rents are higher when it is a tourist destination  
(d)On New Years Eve the rates of Room Rent is higher as compared to a regular day

**4. RESULTS AND CONCLUSION**

**For Linear Model 1**

The model that was created had high significance factors, this meant that each model created has its own significance and cannot be removed as they help in getting a R^2   
value although it was low. R^2 and adjusted R^2 value are respectively 0.1575 and 0.1574. However due to huge RSE value this model is not a correct adaptation of the RoomRent. But these factors shouldn't be eliminated also. The outlier test for the model provided the outliers that may affect the model and its overall value but it wasn't much so. VIF stands for Variation Inflation Factor. This shows the variation.Any Variation that is greater than 10 creates a problem but as displayed most of them were 1. gvlma model stands for Global Validation of Linear Models Assumptions. This model has been created for checking the validation of the model with respect to GlobalStat, Skewness, Kurtosis, Link Function, Heteroscedasticity. Plotting a linear model with respect to the values 4 graphs were created. In Residuals Vs Fitted shows a linear relationship that means this is not a correct model. The Normal Q-Q model shows that a linear model upto a value but huge outliers create a problem. The fourth plot has the cook's distance where the outliers that surpass the 0.5 range are rejected. However it is shown that these outliers didn't matter.

## Call:

## lm(formula = RoomRent ~ StarRating + Airport + IsTouristDestination,

## data = HotelData.df)

##

## Residuals:

## Min 1Q Median 3Q Max

## -9076 -2427 -863 904 310970

##

## Coefficients:

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

## (Intercept) -8942.499 296.290 -30.182 < 2e-16 \*\*\*

## StarRating 3660.485 77.563 47.194 < 2e-16 \*\*\*

## Airport 15.355 2.625 5.849 5.05e-09 \*\*\*

## IsTouristDestination 2051.468 129.893 15.794 < 2e-16 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 6731 on 13228 degrees of freedom

## Multiple R-squared: 0.1575, Adjusted R-squared: 0.1574

## F-statistic: 824.6 on 3 and 13228 DF, p-value: < 2.2e-16

1. For every 1 unit of star rating the room rent changes to Rs. 3660.485.

2. For every 1 unit of Airport Distance the room rent was changed for a factor of Rs. 15.535.

3. If it is a tourist destination a factor of Rs. 2051.468 was changed.   
**4.1 CONCLUSION**

This report was motivated by the need for research that could improve our understanding of how different external and internal factors influences the pricing strategies in the hotel industry of India. The unique contribution of this analysis is that we investigated the price premium charged by hotels according to the facilities they provide and also where it is situated.

This research has some important managerial implications. We found that not only supply affects prices but also there many other factors which can influence our pricing strategy. When consumer sees good ratings and reviews about hotel and gets better amenities, it prompts an increase in quality perceptions, purchase intentions and willingness-to-pay.