**Analysis of Hotel Room Pricing In the Indian Market**

**Name: Raghu Ram Chinthakrinda**

**E-mail:** [**edm14b010@iiitdm.ac.in**](mailto:edm14b010@iiitdm.ac.in)

**College: IIITDM KANCHEEPURAM**

**Mobile No: 7382998506**

**1. Abstract**

This analysis is about the hotel room rent pricing in the Indian market considering many factors and how they are affecting them.

In this analysis we will take insights from different graphs, histograms and plots drawn between the room rent of hotel versus different other factors such as whether the hotel has facilities like swimming pool, free Wi-Fi, free breakfast or not, whether the hotel capacity has an impact on the price of room etc.

We also evaluate by how much supply of rooms in hotels affects prices for hotel rooms, i.e., hotels with high capacity charges less as compared to the one with low capacity.

The data collected is a field study which empirically investigates the pricing of hotel rooms located in 42 different cities of India during the time period of December 2016 to January 2016. We estimate a regression of hotel room prices in a mixed-model framework. Our model accounts for both fixed-effects and random-effects, controlled for unobserved heterogeneity. We estimate it using the Restricted Maximum Likelihood (REML) methodology. Our analysis reveals a significant “price of ratings, swimming pool and capacity” embedded in hotel room rent among hotels in different cities of India.

**2. An empirical field study of hotel pricing strategy in India**

**2.1 correlation**

we study how the price of a room is effected by using Boruta package.

Boruta performed 11 iterations in 3.361225 mins.

10 attributes confirmed important: Airport, CityRank, FreeBreakfast,

FreeWifi, HasSwimmingPool and 5 more;

2 attributes confirmed unimportant: IsNewYearEve, IsWeekend;

After this we had done correlation to test room rent with other variables and found the most three important variables. They are as follows

1. Star ratings
2. Hotel capacity
3. Has swimming pool

We found the independent variables which room rent does not depend upon. They are as follows:

1. Population
2. ismetrocity
3. freebreakfast

**2.2 Hypotheses**

We study how the price of a room at a hotel is affected by external and internal factors. We assume out of 18 factors the 3 most influencing factors are Star Rating, availability of swimming pool and Hotel capacity. We are taking these three factors on the basis of their strong correlation with room rent.

We will frame our hypothesis based on these three factors as well as with other binary functions which may affect the room rent of a hotel.

**H1:** The average room rent in hotel having swimming pools is more than that which don’t have swimming pools.

**ans:** by t.test, it is a null hypothesis

**H2:** The average Room Rent in hotels with high star rating is high as compared to one which has less star rating.

Ans: by t.test , we cannot predict the p value due to 5 levels of star rating. So by doing chisq.test, it is null hypothesis

**H3:** The average Room Rent in hotels with more capacity is lower than the Room Rent with low hotel capacity.

Ans: by t.test, it is a null hypothesis

**H4:** The average Room Rent in hotels providing Free Breakfast is more than that which don’t provide breakfast.

Ans: by t.test, it is a null hypothesis

**H1:** The average Room Rent in hotels in metro cities is more than that of hotel in non-metro cities.

Ans: by t.test, it is a null hypothesis

**2.3 Data**

The purpose of this project is to analyze the pricing strategy of hotels in the Indian hotel industry. Many factors drive hotel room prices. The objective of this project is to identify the factors that matter the most. The dataset tracks hotel prices on 8 different dates at different hotels across different cities.

this project, our dataset is based on hotels located in forty two Indian cities (Mumbai, Delhi, Bangalore, Chennai, Hyderabad, Ahmedabad, Kolkata, Surat, Pune, Jaipur, Thrissur, Lucknow, Kanpur, Amritsar, Indore, Kanyakumari, Agra, Madurai, Goa, Rajkot, Varanasi, Srinagar, Jodhpur, Chandigarh, Thiruvathipuram, Guwahati, Mysore, Bhubaneswar, Kochi, Mangalore, Udaipur, Pondicherry, Haridwar, Puri, Shimla, Panchkula, Darjeeling, Rishikesh, Gangtok, Ooty, Jaisalmer, Bodh Gaya, Nainital, Munnar, Manali) India. We collected data from the well-known website [www.hotels.in](http://www.hotels.in) that aggregates the hotel prices on 8 different dates at different hotels across different cities.

It is indeed probable that many factors govern the rent of hotel rooms. Any meaningful empirical analysis will need to control for factors. For example, factors such as whether the hotel is rated as a five star hotel, how many rooms does it have, has swimming pool are all likely to influence hotel prices.

**Price:**

We collected data from 18 Dec 2016 to 08 Jan 2017. We used Room Rent to denote the average price of a room at a hotel. We measured Room Rent, rent for the cheapest room, double occupancy, in Indian Rupees. Some hotels have more than one type of double occupancy room. For simplicity, we picked the cheapest room with double occupancy at hotel k in city j.

**Star Rating**:

In India, the Ministry of Tourism has formulated a scheme for classification of operational hotels using a “Star” rating. Hotels are rated as either 5 Star, 4 Star, 3 Star, 2 Star or 1 Star. Accordingly, we classified the hotels in our dataset using their star rating. The reason for doing this is that the star rating of a hotel has a direct, strongly positive correlation with the price of its hotel rooms. Therefore, it is important to control for price variation because of the star rating. We used the variable Star Rating to denote the star rating of hotel k in city j.

**Hotel Capacity:**

We recorded the total number of rooms in hotel k in city j as Hotel Capacity. Ultimately, the number of rooms in a hotel denotes the available supply and it is expected that this will keenly influence the price that a hotel will set. Accordingly, we used Hotel Capacity as a control variable to account for the possibility that the room price set by a hotel may depend upon the supply of available rooms.

**Swimming Pool:**

The amenities and facilities provided within a hotel can also potentially influence the price of a room. The greater the amenities, the higher should be the price of the hotel room. To partially control for such factors, we recorded whether a hotel had a Swimming Pool or not. We used HasSwimmingPool to denote the presence or absence of a Swimming Pool at hotel k in city j.

**2.4 Model**

We analyzed the research question using one model.

**Model:**

We established the effect of Star Rating, Hotel Capacity and availability of Swimming Pool on the price of a room in a hotel with the simplest model we could come up with. We regressed the room rent on the variables Star Rating, Hotel Capacity and whether hotel had a swimming pool, in our second model the previous three variables remained and we added IsTouristDestination and IsWeekend as factors and lastly we added Airport distance from the hotel in the basic three variable of the beginning to propose a better model, they are as follows.

We estimated four different Models, described above using linear least squares.

The benefit of having the three regressions outlined in Model was that it helped us rule out some alternate explanations for the variation in hotel room rent. For example, it is well-known that five-star hotels are more expensive than four-star hotels. Including the star rating as a regression, permitted us to investigate the effect of other variables on hotel room rent, after controlling for price variation due to the star rating. We expected to find the coefficient for Star Rating to be positive (B1>0). Similarly, having a dummy variable has Swimming Pool or not for each hotel, permitted us to control effect of availability of swimming pool on rent of hotel rooms and the same way about Hotel capacity, whether the place is a tourist destination, what is the distance of hotel from the airport, the sold out of all rooms in hotel depends on weekend or not, etc.

**2.4 Results**

**Model:**

The analysis of Model also yielded statistical support for our hypotheses H1s. Recall that Model includes three to six independent variables, as shown in equations above. We found that the average room prices with higher ratings and having swimming pool were higher than the prices with low ratings and no swimming pool. This regression analysis yielded B1 >0, B2 >0, with p <0.05, as shown in Table 2. As expected, we additionally observed a negative relationship between the average hotel room prices and the hotel capacity, B3>0, with p < 0.05. But, we found that the Model that we have taken doesn’t have very good R², so there may exist models better than the model we have taken with other variables, in explaining the relationship between hotel pricing strategies.

The coefficients and linear model of the above three models mentioned in 2.3are as follows: -

Model1: salary = b0 + b1\*StarRating + b2\*HasSwimmingPool+ b3\*HotelCapacity  
 b0 = -1(assumption), b1 = 1396.874562, b2=3719.6943, b3= -7.659814  
 Model: salary = -1 + 1396.874562\*StarRating + 3719.6943\*HasSwimmingPool -7.659814\*HotelCapacity

Model1: salary = b0 + b1\*StarRating + b2\*HasSwimmingPool+ b3\*HotelCapacity +b4\*IsWeekend(0) + b5\*IsWeekend(1) + b6\*IsTouristDestination  
 b0 = -1(assumption), b1 = 3635.819, b2=2285.132, b3= -13.965, b4=-8396.67457, b5=-8325.09152,b6=1878.94395  
 Model: salary = -1 + 3635.819\*StarRating + 2285.132\*HasSwimmingPool -13.965\*HotelCapacity - 8396.67457\*IsWeekend(0) - 8325.09152\*IsWeekend(1) + 1878.94395\*IsTouristDestination

Model1: salary = b0 + b1\*StarRating + b2\*HasSwimmingPool+ b3\*HotelCapacity +b4\*Airport + b5\*Date  
 b0 = -1(assumption), b1 = 1248.426988 , b2=3903.736921, b3= -6.743354, b4= 18.869726  
 Model: salary = -1 + 1248.426988\*StarRating + 3903.736921\*HasSwimmingPool -6.743354\*HotelCapacity + 18.869726\*Aiport

The coefficients of model 1 implies:

* When Star rating increases by 1-unit Hotel Room rent increases by

Rs 1396.87

* When availability of swimming pool changes then Hotel room rent increases by Rs 3719.69
* When Hotel capacity changes by 1 unit then Hotel room rent decreases by Rs 7.659814

The coefficients of model 2 implies:

* When Star rating increases by 1-unit Hotel Room rent increases by Rs 3635.81
* When availability of swimming pool changes then Hotel room rent increases by Rs 2285.13
* When Hotel capacity changes by 1 unit then Hotel room rent decreases by Rs 13.965
* When there are no weekend days the charges of room rent increases by Rs 8396.67 and on weekends it increases the room rent by

Rs 8325.09.

* Is hotel being near a tourist destination then room rent increases by

Rs 1878.94.

The coefficients of model 3 implies:

* When Star rating increases by 1-unit Hotel Room rent increases by Rs 1248.42
* When availability of swimming pool changes then Hotel room rent increases by Rs 3903.73
* When Hotel capacity changes by 1 unit then Hotel room rent decreases by Rs 6.7433.
* When distance of airport from any hotel increases by 1 unit then Hotel room rent increases by Rs 18.8697

The coefficients of model 4 implies:

* With increase in star rating by 1 unit, room rent increases by Rs 2005
* With availability of swimming pool, room rent increases by Rs 3170
* With increase in hotel capacity by 1 unit, room rent decrease by Rs 6.74
* With free breakfast, room rent decreases by Rs 1924
* With free wifi, room rent decreases by Rs 1359
* When distance of airport from any hotel increases by 1 unit then Hotel room rent increases by Rs 10.8697

**3. Conclusion**

This paper 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 paper 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.

**Table 1:** Summary Statistics of Hotels pricing strategy study

Room Rent

Min. 1st Qu. Median Mean 3rd Qu. Max.

299 2436 4000 5474 6299 322500

Star Rating

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 3.000 3.000 3.459 4.000 5.000

Swimming Pool

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0000 0.0000 0.0000 0.3558 1.0000 1.0000

Hotel Capacity

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 16.00 34.00 62.51 75.00 600.00

Is a Tourist Destination

Min. 1st Qu. Median Mean 3rd Qu. Max.

1. 0.0000 0.0000 0.2842 1.0000 1.0000

Is a Weekend when all rooms sold out

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0000 0.0000 1.0000 0.6972 1.0000 1.0000

Airport distance from the hotel

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.20 8.40 15.00 21.16 24.00 124.00

**Table 2:** Regression Analysis in the Hotels pricing strategy study

**Model 1:**

Call:

lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity -

1, data = hotel.df)

Residuals:

Min 1Q Median 3Q Max

-8039 -2448 -1249 461 312401

Coefficients:

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

StarRating 1396.8746 26.1320 53.455 < 2e-16 \*\*\*

HasSwimmingPool 3719.6943 148.7835 25.001 < 2e-16 \*\*\*

HotelCapacity -7.6598 0.9415 -8.136 4.44e-16 \*\*\*

---

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

Residual standard error: 6813 on 13229 degrees of freedom

Multiple R-squared: 0.4457, Adjusted R-squared: 0.4456

F-statistic: 3546 on 3 and 13229 DF, p-value: < 2.2e-16

P-values of individual independent variables are also < 0.05, So we reject all the null hypotheses and hence, Star Rating, availability of swimming pool and Hotel Capacity all significantly affects Room Rent.

R2 is very small so it may not be good model and some other factors which are not included might affect Room rent.

**Model 2:**

Call:

lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity +

IsWeekend + IsTouristDestination - 1, data = hotel.df)

Residuals:

Min 1Q Median 3Q Max

-8326 -2517 -1212 463 312480

Coefficients:

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

StarRating 1258.9558 44.4985 28.292 < 2e-16 \*\*\*

HasSwimmingPool 3670.2511 148.8411 24.659 < 2e-16 \*\*\*

HotelCapacity -6.1769 0.9658 -6.396 1.65e-10 \*\*\*

IsWeekend -509.6479 119.1618 -4.277 1.91e-05 \*\*\*

IsTouristDestination 1053.0394 124.7325 8.442 < 2e-16 \*\*\*

---

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

Residual standard error: 6792 on 13227 degrees of freedom

Multiple R-squared: 0.4493, Adjusted R-squared: 0.4491

F-statistic: 2159 on 5 and 13227 DF, p-value: < 2.2e-16

**Model 3:**

Call:

lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity +

Airport - 1, data = hotel.df)

Residuals:

Min 1Q Median 3Q Max

-8240 -2380 -1224 384 312742

Coefficients:

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

StarRating 1248.4270 33.2220 37.578 < 2e-16 \*\*\*

HasSwimmingPool 3903.7369 150.6728 25.909 < 2e-16 \*\*\*

HotelCapacity -6.7434 0.9482 -7.112 1.20e-12 \*\*\*

Airport 18.8697 2.6157 7.214 5.73e-13 \*\*\*

---

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

Residual standard error: 6800 on 13228 degrees of freedom

Multiple R-squared: 0.4479, Adjusted R-squared: 0.4477

F-statistic: 2683 on 4 and 13228 DF, p-value: < 2.2e-16

Model 4:

lm(RoomRent~StarRating+HasSwimmingPool+HotelCapacity+FreeBreakfast+Airport+Population+FreeWifi-1, data = hotel.df)

lm(formula = RoomRent ~ StarRating + HasSwimmingPool + HotelCapacity +

FreeBreakfast + Airport + Population + FreeWifi - 1, data = hotel.df)

Residuals:

Min 1Q Median 3Q Max

-8980 -2443 -1022 689 311696

Coefficients:

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

StarRating 2.006e+03 6.958e+01 28.825 < 2e-16 \*\*\*

HasSwimmingPool 3.171e+03 1.562e+02 20.294 < 2e-16 \*\*\*

HotelCapacity -6.179e+00 9.938e-01 -6.218 5.18e-10 \*\*\*

FreeBreakfast -1.925e+02 1.245e+02 -1.546 0.122038

Airport 1.014e+01 2.670e+00 3.796 0.000147 \*\*\*

Population -1.932e-04 1.490e-05 -12.966 < 2e-16 \*\*\*

FreeWifi -1.396e+03 2.030e+02 -6.875 6.47e-12 \*\*\*

---

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

Residual standard error: 6735 on 13217 degrees of freedom

Multiple R-squared: 0.4589, Adjusted R-squared: 0.4586

F-statistic: 1601 on 7 and 13217 DF, p-value: < 2.2e-16