**A Study of Hotel Room Pricing in the Indian Market**

India is among the top tourist destinations in the world with around 8.89 million foreign tourists arriving in India in 2016. This brings huge revenue earnings for the country and to the hotel industry in India. This study analyzed the status of Indian hotel room rates, differentiated by hotel characteristics. Using a hedonic pricing method, this study based its analysis on the data drawn from [www.hotels.in](http://www.hotels.in) of 1671 hotels which spawns across 42 cities in India and contains 13232 data points from the year 2016. The dataset tracks hotel prices on 8 different dates at different hotels across different cities.

The model we built can explain the price variations in room prices by almost 45%. Further, the findings showed that the room rates charged by hotels with 4 or more star rating, with swimming pool and in a tourist destination are way higher than those who aren’t in this category. This study also noted in particular how lesser distance to the airport had the room prices lower as usually the airports are situated outside the city limits. This made us come to the conclusion that any hotel in India in tourist destinations, well rated (>4) with airport distance more than 15 kms on a new year eve would charge room rates much higher than other hotels.

1. **Introduction**

India is a large market for travel and tourism. It offers a diverse portfolio of niche tourism products - cruises, adventure, medical, wellness, sports, MICE, eco-tourism, film, rural and religious tourism. India has been recognized as a destination for spiritual tourism for domestic and international tourists.

Total contribution by travel and tourism sector to India’s GDP is expected to increase from US$ 136.3 billion in 2015 to US$ 275.2 billion in 2025. Travel and tourism is the third largest foreign exchange earner for India. A sum of US$ 1.76 billion was earned under foreign exchange through tourism during the month of September 2016.

India’s rising middle class and increasing disposable incomes has continued to support the growth of domestic and outbound tourism. Online hotel bookings in India are expected to double by 2016 due to the increasing penetration of the internet and smart phones.

The hotel market in India has hence become highly competitive with many hotels attempting to improve the competency of their resource allocations and market strategies, as well as the enhancement of their loyalty customer programs.

Room pricing decisions is one of the key factors for a hotel marketing strategy because the hotel room rate is one of the significant determinants when customers decide to select their accommodation. The hotel price also affects the customer’s perception of the hotel service quality because, between hotels, a room price difference can signal service quality differences (Hung et al., 2010; Thrane, 2007). Many previous studies (Chen & Rothschid, 2010; Hung et al., 2010; Israeli, 2002; Thrane, 2007) attempted to investigate the factor of room pricing in capital cities because of the profound variability from one city to another, or even within the same capital city.

Therefore, a hedonic pricing analysis, a suitable ordinary least square (OLS) technique, has been widely used in logging studies to analyze hotel room rate variation because it can assist in uncovering the hotel attributes or characteristics for which guests are willing to pay (Chen & Rothschid, 2010; Hung et al., 2010; Thrane 2007). Hedonic prices not only come into effect when quality differentiated hotel goods are sold in the competitive market, but they can also be categorized under the rubric of nonmarket because goods and services in the hotel industry sometimes have qualities not provided by the market.

The purpose of this study is to analyze the pricing strategy of hotels in the Indian hotel industry. Many factors drive hotel room prices and the objective of this study is to identify the factors that matter the most.

1. **Background**

The hedonic pricing model was introduced by the Waugh’s study (1928) of price differences in agricultural products. It also said that the observed room rate in a market is affected by the particular hotel characteristics (Espinet et al., 2003).

In relation to previous studies on the city hotel, Thrane (2007) analyzed the determinants on the room rates of hotels in Oslo, the capital of Norway, by using a seemingly unrelated regression method (SUR). This technique revealed that hotel chain and the distance to the city center both have a major influence on Oslo hotel room rates. The results showed that in Oslo any room belonging to a hotel chain will be more expensive than the non-hotel chain room. Furthermore, the results also showed that the closer a room is to the city center, the more expensive it will be.

Chen & Rothschild (2010), on the other hand, studying the influence of different attributes on hotel room rates in Taipei, the capital of Taiwan, Republic of China, found that distance to the city center has a negative relationship on the hotel price, implying that the expensive hotels in Taipei are located far from the center because they are typically of a resort type, providing more amenities to customers.

A similar study of hotels in Taiwan, by Hung et al. (2010) used a quartile regression to discern the determinants of hotel prices in terms of different quartiles and found that both distance to the city center and hotel chain-affiliation have no impact on Taipei room rates. In terms of expense, an older hotel in Taipei is cheaper than a newer one. In contrast, Lee & Jang (2012) conducted research on the Chicago hotel market with the data obtained from online websites and found that hotels located in the central business district have a higher demand season, prices, and profits. More importantly, the chain hotel room rate in Chicago, during both high and low seasons, is still higher than those which do not belong to the chain. Consistently, a recent study in Beijing by Zhang et al. (2011) with a geographically weighted regression found that hotel characteristics such as star rating, hotel size, hotel age, and location have an important influence on hotel room rates.

1. **An empirical field study of Hotel room prices in the Indian Market**
   1. **Overview**

The specific objective of this study was to investigate the pricing strategy employed by hotels in India. This study analyzed hotel prices in 42 cities of India which are given in Table 1 of Appendix A. Our goal was to compare prices of hotels rooms with swimming pool, with the prices of rooms without a view, similarly those which are present in tourist destination and those which aren’t, those which have more than 4 star ratings and those below. The rationale behind this is the fact that India is one of the sought-after tourist destinations of the world as well as that India has undergone a digital revolution due to which customers have become highly aware and the services have become competitive. Thus, amenities like swimming pools or distance to airport or city rank may have effect on the hotel room prices which is analyzed in this study.

In this Study, we compared the prices of hotel rooms with/without swimming pools, prices on a weekend/weekday, prices in tourist destinations and not in tourist destinations, prices of hotels which are above 4 stars than those which aren’t etc. Accordingly, we constructed the following hypotheses:

1. **Hypothesis H1:** *The average prices of hotel rooms on weekends are higher than the prices of hotel rooms on weekdays.*
2. **Hypothesis H1:** *The average prices of hotel rooms with swimming pools are higher than the prices of hotel rooms without swimming pools.*
3. **Hypothesis H1:** *The average prices of hotel rooms in tourist destinations are higher than the prices of hotels not in tourist destinations.*
4. **Hypothesis H1:** *The average prices of hotel in 4 star and above rated hotels are higher than the prices of hotels in below 4 star rated hotels.*
5. **Hypothesis H1:** *The average prices of hotel vary with city ranks.*
6. **Hypothesis H1:** *The average prices of hotel with higher hotel capacities are higher than the prices of hotels with lower hotel capacity.*
7. **Hypothesis H1:** *The average prices of hotel with greater distance than 15 kms to the airport are higher than the prices of hotels with distance lesser than 15 kms. (This is because airports are usually outside the city area.)*
   1. **Data**

Dependent Variable

|  |  |  |
| --- | --- | --- |
| DECISION VARIABLE | UNITS | MEANING |
| RoomRent | Rupees | 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. |

External Factors

Many external factors can potentially influence the RoomRent. The dataset captures some of these external factors, as explained below.

|  |  |  |  |
| --- | --- | --- | --- |
| 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 A1 below) |  |
| 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. |  |

Internal Factors

Many Hotel Features can influence the RoomRent. The dataset captures some of these internal factors, as explained below.

|  |  |  |
| --- | --- | --- |
| 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 |

* 1. **Regression Model**

In order to test Hypothesis 1 to 7, we proposed the following log linear models with the below given hypothesis:

**Hypothesis I** - H0: The variables ISWeekend, IsTouristDestination, HasSwimmingPool, StarRating, CityRank, HotelCapacity, Airport, IsNewYear, IsMetroCity collectively have no effect on RoomRent.

H1: ISWeekend, IsTouristDestination, HasSwimmingPool, StarRating, CityRank, HotelCapacity, Airport, IsNewYear, IsMetroCity together affect RoomRent

**Model 1 :** log(RoomRent) ~ B0 + B1\*IsWeekend + B2\*IsTouristDestination + B3\*HasSwimmingPool + B4\*StarRating + B5\*CityRank + B6\*HotelCapacity + B7\*Airport + B8\*IsNewYearEve + B9\*IsMetroCity

**Hypothesis** II - H0: The variables IsTouristDestination, HasSwimmingPool, StarRating, CityRank, Airport , IsNewYearEve collectively have no effect on RoomRent.

H1 : IsTouristDestination, HasSwimmingPool, StarRating, CityRank, Airport , IsNewYearEve together affect RoomRent

**Model 2:** log(RoomRent) ~ B0 + B1\*IsTouristDestination + B3\*HasSwimmingPool+ B4\*StarRating+ B5\*CityRank+ B6\*Airport + B7\*IsNewYearEve

We established the effect of start ratings, having swimming pool, being in a tourist destination, city rank, distance to airport, and being a new year eve on the price of a hotel room with the second model. We regressed on . We estimated model, using log linear least squares.

**3.4 Results**

The regression analysis **for model 1** is as below:

Here, the model equation looks like this -  
log(y)=B1\*x1 +B2\*x2 ..  
log(RoomRent) = 6.3734848 + 0.0102293\* IsWeekend + 0.1341822\*IsTouristDestination + 0.3496599\*HasSwimmingPool + 0.4400102 \*StarRating + 0.0083505\*CityRank -0.0001228\*HotelCapacity + 0.0032378\*Airport + 0.1025027\*IsNewYearEve + 0.0138648\*IsMetroCity + e  
  
Multiple R-squared: 0.4214, Adjusted R-squared: 0.421   
F-statistic: 1046 on 9 and 12931 DF, p-value: < 2.2e-16  
  
Since the model's p value is < 2.2e-16, hence we can reject the null hypothesis that the variables collectively do not explain the variation in the RoomRent.  
  
The results also show that the variable HasSwimmingPool is significant in controlling for the variable RoomRent (p = 2e-16), as are the rest of them except Hotel Capacity,IsMetroCity and IsWeekend.

AIC = 20761.41 , BIC = 20843.56, Multiple R-squared: 0.4214, Adjusted R-squared: 0.421

**For Model 2 –**

Here, the model equation looks like this -  
log(y)=B1\*x1 +B2\*x2 ..  
log(RoomRent) = 6.3780901 + 0.1409580\*IsTouristDestination + 0.3448546\*HasSwimmingPool + 0.4358485 \*StarRating + 0.0081063\*CityRank + 0.0032428\*Airport+ 0.0971002\*IsNewYearEve + e  
  
Multiple R-squared: 0.4212, Adjusted R-squared: 0.421   
F-statistic: 1569 on 6 and 12934 DF, p-value: < 2.2e-16  
  
All variables are significant in this model.  
Since the model's p value is < 2.2e-16, hence we can reject the null hypothesis that the variables collectively do not explain the variation in the RoomRent.

AIC = 20758.84 , BIC = 20818.59, Multiple R-squared: 0.4212, Adjusted R-squared: 0.421

Based on these AIC, BIC values & R2 values where R-squared measures the percent of variation in Y explained by variation in X (or combination of Xs), we would choose model\_1 as it has better AIC and BIC values while R2 is the same. Also, by using regsubset, we had removed the insignificant variables which might have resulted in these slightly better results.

* 1. **Partial F-Test**

According to model\_1, we are including the variables IsTouristDestination,HasSwimmingPool,StarRating,CityRank,HotelCapacity,Airport in our model and we will be testing whether the airport, IsTouristDestination are significant after taking HotelCapacity,HasSwimmingPool,StarRating,CityRank into consideration by doing partial F-Test.

**Reduced model**  
reduced = lm(log(RoomRent) ~ IsTouristDestination+HasSwimmingPool+StarRating+CityRank, data=hotel)

**Full Model**  
full = lm(log(RoomRent) ~ IsTouristDestination+HasSwimmingPool+StarRating+CityRank+Airport+IsNewYearEve,data=hotel)   
  
anova(reduced, full)

The output shows the results of the partial F-test. Since F=112.2 (p-value< 2.2e-16) we reject the null hypothesis ( B5 = B6 = 0) at the 5% level of significance. It appears that the variables IsNewYearEve and Airport do contribute significant information to the room rent prices once the other variables from the model like IsTouristDestination ,HasSwimmingPool , StarRating ,CityRank have been taken into consideration.

**Another Partial F-Test -  
   
*Reduced model***  
reduced = **lm**(**log**(RoomRent) ~ IsTouristDestination+Airport+HasSwimmingPool+StarRating+CityRank, data=hotel)

***Full Model***  
full = **lm**(**log**(RoomRent) ~ IsTouristDestination+HasSwimmingPool+StarRating+CityRank+Airport+IsNewYearEve,data=hotel)

Similarly, the output shows the results of the partial F-test. Since F=45.627 (p-value=1.492e-11)  
We reject the null hypothesis ( B6 = 0) at the 5% level of significance.

It appears that the variables IsNewYearEve do contribute significant information to the room rent prices once the other variables from the model like IsTouristDestination ,HasSwimmingPool , StarRating ,CityRank,IsTouristDestination have been taken into consideration.

**3.6 Confidence & Prediction levels of the Regression Model**

Taking the regression model as :

results = **lm**(**log**(RoomRent) ~ IsTouristDestination+HasSwimmingPool+StarRating+CityRank+Airport+IsNewYearEve,data=train)

We predicted the results by taking the observations as:

**predict**(results,**data.frame**(IsTouristDestination=0,HasSwimmingPool=0,StarRating=2,CityRank=3,Airport=9.6,IsNewYearEve=0),interval="confidence")

*We obtained that a 95% confidence interval is given by (1448.426, 1528.968)*

*With a 95% confidence interval, the results after exponentiation that we get are really close to the actual room price of 1468. Thus, our model seems to be doing good.*

**For prediction intervals -**

**predict**(results,**data.frame**(IsTouristDestination=0,HasSwimmingPool=0,StarRating=2,CityRank=3,Airport=9.6,IsNewYearEve=0),interval="prediction")

A 95% prediction interval is given by (521.0573, 4344.544).  
The results obtained for prediction have a wider range than confidence interval indicating that the variation about the mean is fairly large.

**The main conclusions about the model are:**

1. The R-squared is 0.4213, which means that 42% of the variance in our dependent variable can be explained by the set of predictors in the model; at the same time, the adjusted R-squared is not far from that number, meaning that the original R-squared has not been artificially increased by adding variables to the model. Note that the R-squared can only increase or stay the same by adding variables, whereas the adjusted R-squared can even decrease if the variable added doesn't help the model more than what is expected by chance;
2. All the variables are statistically significant (p < 0.05) and the most significant predictor is the StarRating as seen from the coefficient plot. The advantage of doing a log transformation is that, if the regression coefficient is small (i.e. -0.1 to 0.1), a unit increase in the independent variable yields an increase of approximately coeff\*100% in the dependent variable.
3. To be clear, the coefficient of the Star Rating is 0.4369466. It means that a unit increase in the Star Rating (i.e., increasing the rating by 1 point), increases the predicted amount of Room Rent by approximately 44%. One can always exponentiate to get the exact value. By the same token, if a hotel has a swimming pool, the predicted RoomRent increases by 35%)
4. Both the RMSE and MAE have significantly decreased when compared with the baseline model, which means that this linear model, despite all the linearity issues and the fact that it predicts negative values of rain in some days, is still much better, overall, than our best guess.

**4. Conclusion**

This study was motivated by the to understand the factors behind hotel room pricing in the Indian context. Research in this area could improve the understanding for the consumers as well as lay light on the revenue management systems of hotels. We found that hotel bookings in 4 or more star rated hotels with swimming pools in a tourist destination means paying a hefty “price while travelling”.

1. **References**
2. Hung, W.-T., Shang, J.-K., & Wang, F.-C. (2010). Pricing determinants in the A Hedonic Pricing Analysis of Hotel Room Rates in Bangkok 15 hotel industry: Quantile regression analysis. International Journal of Hospitality Management, 29, 378-384.
3. Thrane, C. (2005). Hedonic price models and sun-and-beach package tours: the Norwegian case. Journal of Travel Research, 43, 302-308.
4. Chen, C.-F., & Rothschid, R. (2010). An application of hedonic pricing analysis to case of hotel room in Taipei. Tourism Economics, 16(3), 685-694.
5. Israeli, A. A. (2002). Star rating and corporate affiliation: their influence on room price and performance of hotels in Israel. Hospitality Management, 21(4), 405-424.
6. Zhang, H., Zhang, J., Lu, S., Cheng, S., & Zhang, J. (2011). Modeling hotel room price with geographically weighted regression. International Journal of Hospitality Management, 30, 1036-1043.
7. Lee, S. K., & Jang, S. (2012). Premium or discount in hotel room rates? The dual effects of a central downtown location. Cornell Hospitality Quarterly, 53(2), 163-173.

**Appendix A**

**Table A1:  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 | 1 | 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 | Kanyakumari | 1 | 4 | 1,870,374 |
| 16 | Agra | 1 | 102 | 1,760,285 |
| 17 | Madurai | 1 | 21 | 1,465,625 |
| 18 | Goa | 1 | 626 | 1,457,723 |
| 19 | Rajkot | 0 | 26 | 1,286,678 |
| 20 | Varanasi | 1 | 60 | 1,201,815 |
| 21 | Srinagar | 1 | 57 | 1,180,570 |
| 22 | Jodhpur | 1 | 81 | 1,033,918 |
| 23 | Chandigarh | 0 | 117 | 960,787 |
| 24 | Thiruvathipuram | 0 | 128 | 957,730 |
| 25 | Guwahati | 0 | 12 | 957,352 |
| 26 | Mysore | 0 | 58 | 887,446 |
| 27 | Bhubaneswar | 0 | 29 | 837,737 |
| 28 | Kochi | 1 | 188 | 595,575 |
| 29 | Mangalore | 0 | 13 | 499,487 |
| 30 | Udaipur | 1 | 113 | 451,735 |
| 31 | Pondicherry | 0 | 42 | 241,773 |
| 32 | Haridwar | 1 | 73 | 228,832 |
| 33 | Puri | 1 | 24 | 201,026 |
| 34 | Shimla | 1 | 58 | 169,578 |
| 35 | Panchkula | 0 | 118 | 140,925 |
| 36 | Darjeeling | 1 | 32 | 132,016 |
| 37 | Rishikesh | 1 | 107 | 102,138 |
| 38 | Gangtok | 1 | 30 | 98,658 |
| 39 | Ooty | 1 | 64 | 88,430 |
| 40 | Jaisalmer | 1 | 82 | 65,471 |
| 41 | Bodh Gaya | 1 | 20 | 45,349 |
| 42 | Nainital | 1 | 85 | 41,377 |
| 43 | Munnar | 1 | 108 | 38,471 |
| 44 | Manali | 1 | 80 | 8,096 |