

STUDY OF THE FACTORS THAT INVOLVE PRICING STRATEGY OF HOTELS IN INDIA

SANJAY HANSDAK

ABSTRACT

The purpose of this project is to analyze the pricing strategy of hotels in the Indian hotel industry. Many factors drive hotel room prices which are primarily of two types: external and internal. The objective of this project is to identify the factors that matter the most.

INTRODUCTION

The dataset tracks hotel prices on 8 different dates at different hotels across different cities.

DATA PREPROCESSING:-

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 Room Rent. 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

CityRank	Dummy	Rank order of City by Population (e.g. Mumbai = 0, Delhi = 1, so on
IsMetroCity	Dummy	'1' if CityName is {Mumbai, Delhi, Kolkatta, Chennai}, '0' otherwise
IsTouristDestination	Dummy	We use '1' if the city is <u>primarily</u> 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.

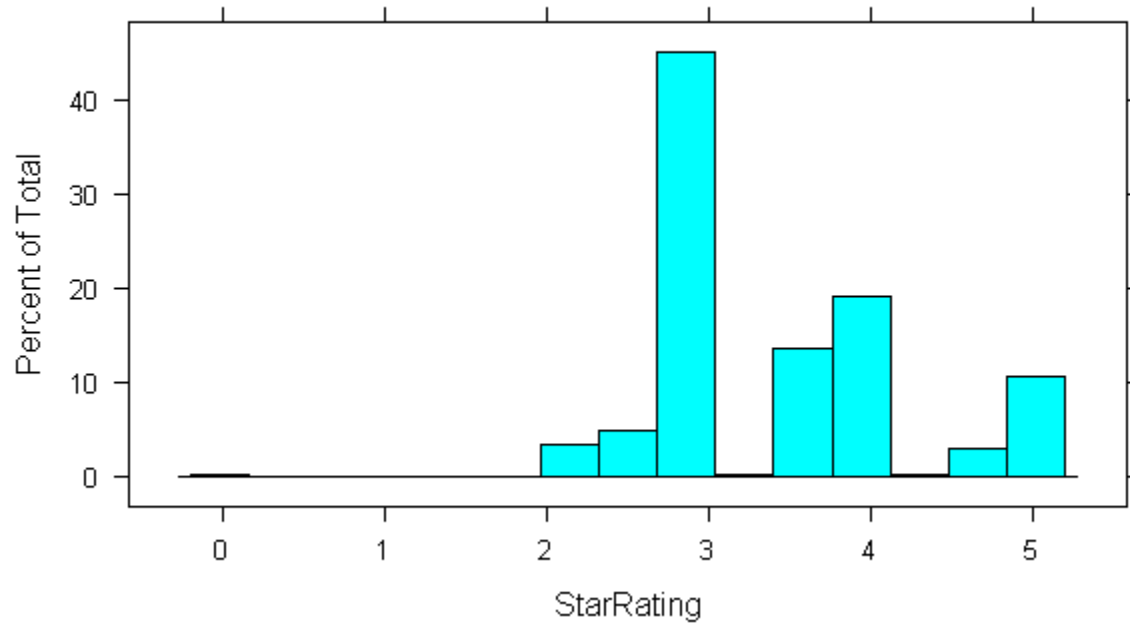
INTERNAL FACTORS

Many Hotel Features can influence the Room Rent. 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

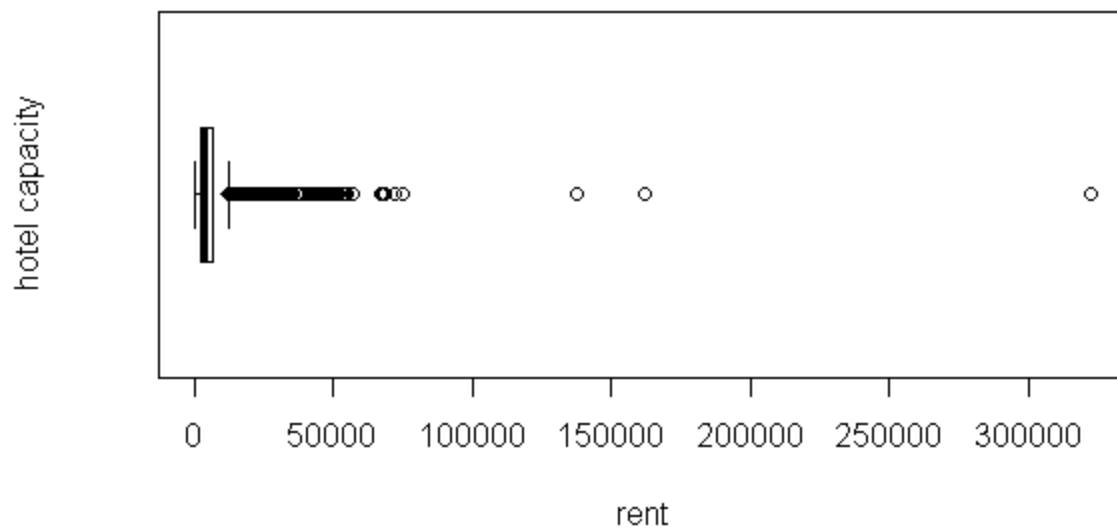
METHOD

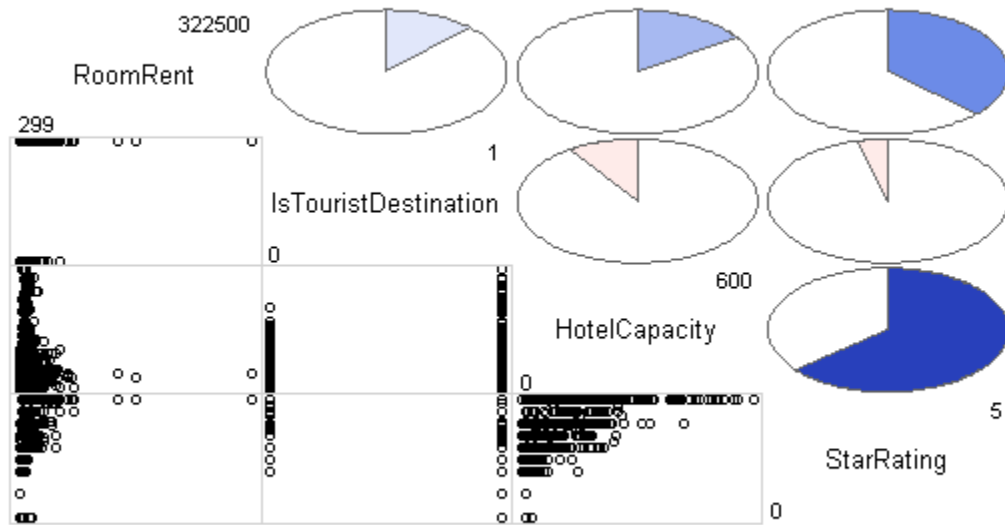
The dataset was read into R. The data was summarized to understand the mean, median, standard deviation of each variable. The problem was formulated as $Y = F(x_1, x_2, x_3)$. The Dependent Variable(s) (i.e. the Y in the $Y = F(x)$) in the Dataset was identified as RoomRent. The three most important Independent variables (i.e. x_1, x_2, x_3) in the dataset were taken as StarRating, HotelCapacity and IsATouristDestination. Some visualizations have been shown below to understand the correlation between these parameters.



Percentage of Hotels having 0-5 Star ratings

Boxplot showing distribution of hotel capacity vs rent





Corrgram in R involving the Independent and Dependent Variables

The dataset was then fitted by a linear regressor on a training set which consisted of 80% of the sample and predictions were made on the test set which contained 20% of the sample.

OBSERVATIONS

To get the optimal model the adjusted R-Squared value was looked at. The model which gave the highest R-Squared value was chosen as final. First all the external factors were clubbed together and then the internal factors. Those features which showed a significantly less p value than 0.05 were taken as statistically significant and the final model result is shown below.

Call:

```
lm(formula = RoomRent ~ IsTouristDestination + HasSwimmingPool +
    IsNewYearEve + IsMetroCity + StarRating + HotelCapacity,
    data = training_set)
```

Residuals:

```
Min    1Q  Median    3Q   Max
-11995 -2373  -711  1049 308998
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) -8615.856  409.016 -21.065 < 2e-16 ***
IsTouristDestination 2269.917  150.651  15.067 < 2e-16 ***
HasSwimmingPool 2112.919  183.645  11.505 < 2e-16 ***
IsNewYearEve 702.754  203.505  3.453 0.000556 ***
IsMetroCity -1660.269  154.920 -10.717 < 2e-16 ***
StarRating 3730.666  128.298  29.078 < 2e-16 ***
HotelCapacity -11.630  1.175 -9.894 < 2e-16 ***
```

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

Residual standard error: 6955 on 10764 degrees of freedom

Multiple R-squared: 0.1795, Adjusted R-squared: 0.179
F-statistic: 392.4 on 6 and 10764 DF, p-value: < 2.2e-16

CONCLUSION

The most significant factors include the location of a hotel whether it is in a tourist area or in a metropolitan city, the date of booking falls on a special occasion like New Year Eve, the review it has in terms of rating and the total capacity of the hotel that determines the price of a room.

REFERNECES

www.RBloggers.com

The final project report to be submitted under the internship of Prof. Sammer Mathur (IIM Lucknow, CMU) as a part of his data analytics internship in R.