**Hotel Booking Analysis**

**Nikhil Shinde**

* **Abstract:**

This data set contains booking information for a city hotel and a resort hotel and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, along with other things.

* Reasons for booking cancellation across various parameters
* Best time to book hotel
* Area wise booking

Booking cancellations are a key aspect of hotel revenue management. Therefore, in order to better manage the reservation system and apply appropriate cancellation policies, decision support systems based on data analysis.

1. **Problem Statement**

This project contains real world data record of hotel booking of a city and a resort hotel containing details like booking cancellation, area details etc. from 2015 to 2017. Main aim of the project is to understand and visualize the dataset from the hotel and customer point of view.

1. We will be analyzing some key metrics for hotel bookings like:

* The number of cancellations
* Most preferred meal types
* Country wise bookings
* New customers acquired
* Customer lifetime value of the existing customers
* Type of rooms preferred by customers
* Booking types,
* Hotels available for booking

1. We will be using various lenses to look through the data to analyze patterns associated with each segment such as:

* The type of hotel
* Day of week
* Type of customers
* Type of rooms

1. **Introduction**

Hotel industry is a very volatile industry and the bookings depend on a variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan well. Using the historical data, hotels can perform various campaigns to boost the business.

We will be using the data available to analyze the factors affecting the hotel bookings. These factors can be used for reporting the trends and predict the future bookings.

Description of features used in our dataset is as given below:

* After collecting data it’s very important to understand your data. So we had hotel Booking analysis data. Which had 119390 rows and 32 columns. So let’s understand these 32 columns.

Data Description:

**hotel** :Resort Hotel or City Hotel

**is\_canceled** : Value indicating if the booking was canceled (1) or not (0)

**lead\_time** : Number of days that elapsed between the entering date of the booking and the arrival date

**arrival\_date\_year** : Year of arrival date

**arrival\_date\_month** : Month of arrival date

**arrival\_date\_week\_number** : Week number of year for arrival date **arrival\_date\_day\_of\_month** : Day of arrival date

**stays\_in\_weekend\_nights** : Number of weekend nights

**stays\_in\_week\_nights** : Number of week nights.

**adults** : Number of adults

**children** : Number of children

**babies** : Number of babies

**meal** : Type of meal booked.

**country** : Country of origin.

**market\_segment** : Market segment designation. (TA/TO) **distribution\_channel** : Booking distribution channel.(T/A/TO) **is\_repeated\_guest** : is a repeated guest (1) or not (0)

**previous\_cancellations** : Number of previous bookings that were canceled by the customer prior to the current booking

**previous\_bookings\_not\_canceled** : Number of previous bookings not canceled by the customer prior to the current booking

**reserved\_room\_type** : Code of room type reserved.

**assigned\_room\_type** : Code for the type of room assigned to the booking.

**booking\_changes** : Number of changes made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

**deposit\_type** : No Deposit, Non Refund , Refundable.

**agent** : ID of the travel agency that made the booking

**company** : ID of the company/entity that made the booking .

**days\_in\_waiting\_list** : Number of days the booking was in the waiting list before it was confirmed to the customer

**customer\_type** : type of customer. Contract, Group, transient, Transient party.

**adr** : Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights

**required\_car\_parking\_spaces** : Number of car parking spaces required by the customer

**total\_of\_special\_requests** : Number of special requests made by the customer (e.g. twin bed or high floor)

**reservation\_status** : Reservation last status.

Description of variable used in our dataset is as given in table below:

| Variable | Description |
| --- | --- |
| is\_canceled\_counts | Count the number of bookings which are canceled by the customers. |
| booking\_percentage | Percentage of booking canceled by the customer versus percentage of customers checked-in to the hotel . |
| previous\_cancellations | Booking canceled by customers previously. |
| cancellation\_rate | Dictionary containing previous\_cancellations as key and current cancellation rate as values. |
| cancellation\_rate\_data | Transforming cancellation\_ rate to data fame |
| deposit\_type | Count the number of different types of deposit policies accepted by the customer. |
| year\_counts | Total number of bookings across different years |
| country\_counts | Top 10 countries of maximum customers. |
| total\_nights\_stays | Number of nights customers want to stay in the hotel. |
| single | One customer has booked the hotel. |
| couple | Two customers have booked the hotel. |
| family | Booking done for more than two customers. |
| names | After calculation contains the variable single, couple and family. |
| count | It is counting the customers of accommodation type. |
| count\_percent | It calculates the percentage of customers of accommodation type. |
| merged\_RJ\_df | Merging the data frames p1 and p2 (variables) |
| acc\_type | It is data frames of accommodation type and its count\_percent. |
| p1 | Counting the reserved room type |
| p2 | Counting the assigned room type |
| arrival\_date\_months\_count | Number of customers arrived in the hotel across all the  month. |
| market\_segment\_customers | Number of bookings across various market segments. |
| meal\_cat | Different types of meal preferred by the customers. |
| customer\_type\_ | Different types of booking done by the customers.  Contract - when the booking has an allotment or other type of contract associated to it;  Group - when the booking is associated to a group  Transient - when the booking is not part of a group or contract, and is not associated to other transient booking;  Transient-party - when the booking is transient but is associated to at least another transient booking |
| room\_type\_booking | Different types of rooms booked by the customers. |
| ADR | Average Daily Rate of Hotels across different months |

## **Year-wise Comparison**

We can see that 2016 seems to be the highest which is 47.4% while in 2015 only 18.3% bookings were made which shows an increase in reservation while in 2017 only 34.1% bookings were made which shows a decrease in reservation.

## **Peak Seasons**

We can see that 2016 seems to be the year where hotel booking is at its highest**.** We also see an increasing trend in booking around the middle of the year, with August being the highest followed by July and May. Summer ends around August, followed straight by autumn. It seems that summer is a peak period for hotel booking**.** We can also see that January has the lowest number of customers followed by November and December. It seems the end of the year is at its lowest peak for hotel booking.

## **Reasons for cancellation**

Out of 119000 customers, 74745 of customers checked-in to the hotel while 44157 of customers canceled their bookings. If we talk about percentage, 37.2% of bookings got canceled whereas 62.8% of customers did check-in. So, we realize that the high rate of cancellations can be due to no deposit policies.

# **Best time to book hotel**

# Since January has the lowest amount of booking, it can be the best time of year to book a hotel room. Due to less demand for rooms, the cost for rooms on a daily basis is also minimal as compared to other months whereas month of August has higher demand comparatively for room so it is obvious that the cost of room is also at peak.

# **Most reserved and assigned room type**

Out of all room types “A” type was most preferred by customers and also assigned to customers. 41.6% of the customers were assigned their preferred room and 5.4% of the customers were assigned a different room from their preferred room. And room type “P” and “L” are the least preferred and assigned room type.

1. **From which country most guests come?**

Portugal, UK and France, Spain and Germany are the top countries from which most guests come; more than 73.16% come from these 5 countries and more than 84.77% come from top 10 countries.

1. **How long do customers stay in hotels?**

Most people stay for one, two, three and four nights whereas; number of bookings for two and three nights has higher demand comparatively. If we try to optimize the length of stay it will be around two to four nights. For resort hotels, the most popular stay duration is one, two, seven (week) and three days respectively. While for city hotels, the most popular stay duration is two, three, one and four days respectively.

1. **Most booked accommodation Type (Single, Couple, Family / Friends)**

Couple (or 2 adults) is the most popular accommodation type. After that Single is the second most popular accommodation type. So hotels can make plans accordingly to promote their activities.

1. **Effect of Meal Type**

### Meal types provided by hotel are:

* Undefined/SC – no meal package
* BB – Bed & Breakfast
* HB – Half Board (Breakfast & one other meal – usually Dinner)
* FB – Full Board (Breakfast, Lunch & Dinner)

### Out of the meals, BB (Bed & Breakfast) is the most ordered meal which is around 77.2%, followed by HB (Half Board), SC (no meal package), Undefined and FB (Full Board).

1. **Market Segment and Distribution Channel**

Market Segments are:

* Online TA
* Offline TA/TO
* Groups
* Direct
* Corporate
* Complementary
* Aviation
* Undefined

And, Distribution Channels are:

* TA/TO
* Direct
* Corporate
* GDS
* Undefined

Majority Distribution channels and Market segment were TA/TO (Travel Agencies / Travel Operator) whether offline / online. So our focus should be to improve this.

1. **Deposit Type**

Deposit types are:

* No Deposit
* Non Refund
* Refundable

Most of the bookings made by customers are No deposit which is 87.6%. No deposit may lead to cancellation.

1. **Customer Type**

Customer types are:

* Transient
* Transient – Party
* Contract
* Group

Transient types of customers are more than 75%.

1. **Does hotel have regulars or not**

Most of the bookings were made for the first time which is 96.8% (115096) and customers who are regulars or repeated their bookings are 3.2% (3806). Hotels can introduce new policies to increase their regular customer base.

1. **Number of Car Parking Spaces and Special Request made by customer**

According to data 93.8% of customers don't need any parking spaces and mostly, one parking space was required by 6.1% of customers. Most of the customers didn’t make any special request which is 58.8% and out of these 28% of customers canceled their bookings and at least one special request was made by 27.8% of customers and so on.

1. **Effect of Lead Time on Cancellation**

It was seen that higher lead time has a higher chance of cancellation of the booking. And lead time has a positive correlation with cancellation of bookings.

1. **Steps involved:**

* **Explore The Dataset**

After loading the dataset, we explored the data and divided the project into three different categories as – 1. Hotel wise, 2. Booking wise, and remaining part 3. Type of rooms, meal, customer, market segment, countries etc.

* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our operations hence we replaced them at the beginning of our project in order to get a better result.

* **Exploratory Data Analysis**

In this section, we tried to make some insights, finding out reasons for variation of bookings across different years with types of hotels, different countries and so on.

1. **Conclusion**

At last but not the least we reached the end of our exercise.Starting with loading the data so far, we have done EDA, null values treatment, data cleaning, processing and finally with analysis we found some insights such as reason for booking cancellation, Peak season, Types of hotels customer preferred, the time to get best rate of booking on daily rate basis etc.