**Hotel Booking Analysis**

**Gaurav Gade, Hitesh Verma, Anand Gend**

**Data science trainees,**

**AlmaBetter, Bangalore.**

* **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 dataset from hotel and customer point of view.

1. We will be analysing 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 analyse patterns associated with each segment such as:

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

**2. Introduction**

Hotel industry is a very volatile industry and the bookings depend on 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 better. 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 variable used in our dataset is as given in table below:

| **Variable** | **Description** |
| --- | --- |
| is\_canceled\_counts | Count 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 . |
| deposit\_type | Count 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 hotel. |
| 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\_ | 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 room booked by the customers. |
| ADR | Average Daily Rate of Hotels across different months |

## 

## **3.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 period is a peak period for hotel booking.** We can also see that January month has lowest number of customers followed by November and December. It seems Winter period is at lowest peak for hotel booking.

## **4.Reasons for cancellation**

Out of 119000 customers ,75000 of customers checked-in to the hotel while 44000 of customers cancelled their bookings. If we talk about percentage, 37% of bookings got cancelled Whereas 63% of customers did check-in. So, we realise that the high rate of cancellations can be due to **no deposit** policies.

# **5.Best time to book hotel**

# **Since January month has the lowest amount of booking, it can be the best time of year to book a hotel room. Due to less demand of rooms, the cost for room on daily basis also minimum 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.**

**6. 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 making some insights, finding out reasons for variation of bookings across different years with types of hotels, different countries and so on.

**7. 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 out 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.

**8. References**

1. Geeks for Geeks
2. Tableau
3. Sea born Doc.

4. Analytics Vidhya

5. Matplotlib doc.

6. Pandas Doc.