**Hotel Booking Analysis**

**Akash Kumar**

**Abinash Kanungo**

**Kunal Kamble**

**Zueb Ansari**

**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, among other things. All personally identifying information has from the data.

We will perform exploratory data analysis with python to get insight from the data.

***Keywords: Exploratory data analysis, Hotel booking analysis***

**Introduction:**

The hotel booking data collection consists of binary, categorical, and numeric information. We were able to gain significant insights from the data set thanks to columns like the hotel type, is canceled, arrival date year, arrival date month, stays in weekend nights, country, market\_segment, distribution\_channel, etc. Here, our goal is to comprehend the crucial elements that influence hotel reservations

**Problem Statement**

**1. What types of hotels in market and how many numbers of hotels in each type?**

There are two types of hotels.

* City Hotels
* Resort Hotels

Out of 119390 Hotel bookings City hotel bookings which is 66.4% whereas Resort bookings makes 33.6%. Here by Count of City Hotel Bookings are morewhen compared to Resort Hotel.

**2. How Many bookings were cancelled for different types of hotels?**

There are four bars in graph which represent city hotels cancelled, city hotels not cancelled, resort hotels cancelled, resort hotels not cancelled.

we can say that city hotels bookings are

more cancelled than resort hotels.

* **Year wise comparison of booking of hotels.**

In this we have used count plot method which is used in seaborn for comparison between years.

we can say that in year 2016 most numbers of guests visited the both type of hotels.

**3. Which is the busiest month for hotels?**

Busiest month means that which month has most numbers of booking.

we can say that **August** is the busiest month for hotels, which means august month has the greatest number of bookings**.**

* **Now we will check comparatively for both types of hotels which month is busiest.**

For this we have used count plot imported from seaborn from this comparison also August has busiest month among all.

**4. Which was the most booked accommodation type (Single, Couple, Family)?**

According to analysis there are three types of people (guests) come to book hotels those are single, couple and family.

From the analysis we can say thatCouples with no children has greatest number of bookings.

**5. Which type of meal booked?**

Categories are presented in standard hospitality meal packages:

* Undefined/SC — no meal package
* BB — Bed & Breakfast
* HB — Half board (breakfast and one other meal — usually dinner)
* FB — Full board (breakfast, lunch and dinner)

In these categories of meals, guests usually like orders BB i.e., Bed & Breakfast.

Here we have used pie chart to represent different types of meals order by the guests.

BB i.e., Bed & Breakfast**is most preferable meal type of the guests which is 77.4%**

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

In this we have firstly fetch the top 10 countries from the given dataset.

Then rename the columns name such as country, number of bookings, percentage.

Then draw graph using bar plot from there we can say that most of the guests come from **PRT i.e., Portugal**.

**7. Which type of market segment is used for bookings hotels?**

There are eight type of market segment from guests can book their hotels;

Hotels with respect to number of bookings

Online TA 56408

Offline TA/TO 24182

Groups 19791

Direct 12582

Corporate 5282

Complementary 728

Aviation 235

Undefined 2

we can say that **Online TA** market segment is mostly used to book a hotel.

**8. How much guests pay per night for each type of rooms?**

In this we have used box plot to present the types of rooms.

we can say that **'G'** category room of city hotel are much costlier than other.

* **Assigned room type**

In this we have used bar plot to present the assigned room type from reserved room type.

we can say that **'A'** type rooms are most assigned room type among all.

**Steps involved:**

* **Exploratory Data Analysis**

After loading the dataset we performed this method by comparing our target variable that is Hotel booking analysis with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

* **Null values Treatment**

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

## **Check for the wrong entry**

Our dataset contains a there are 3 categories of guests which are children, adults and babies**.**

**If hotel is booked than there should be any entries in children, adults or babies’ columns, it should not zero at a time. Hence we drop them at the beginning if our project in order to get a better result.**

* **Data analysis and Data visualisation**

### Here we do data analysis and data visualization by using matplotlib and seaborn to the graph for better understanding of the insights.

**Conclusion:**

That's all there is to it! We have reached the end of our exercise. The data has been loaded, null values have been treated, categorical columns encoded, and major reasons that govern hotel bookings have been identified, along with steps to increase them.

* About 66% of the reservation are for the city hotel and 34% of the booking are for the Resort Hotel, so City hotel is busier or the resort Hotel. So does the overall adr.
* Both hotels have significantly higher booking cancellations rates and very few guests less than 3% return for any other booking at the City Hotel. 5% of visitors stay at the Resort hotel again.
* The majority of the visitors were from European nations, with the majority being from Portugal.
* Customers book through many channels, with TA/TO being the most popular option.
* Nearly 30% of reservation made via TA/TO are cancelled.
* The busiest and most lucrative months for both hotels is August.
* Since couples are the most frequent hotel customers, hotel can design services to meet their needs in order to boost income.
* Adults and adjacent market segment booking typically have a sufficient number of customised requests.

**References-**

* Pandas user guide: <https://pandas.pydata.org/docs/user_guide/index.html>
* Matplotlib user guide: <https://matplotlib.org/3.3.1/users/index.html>
* Seaborn user guide & tutorial: <https://seaborn.pydata.org/tutorial.html>
* Tutorials: 1) [https://stackoverflow.com](https://stackoverflow.com/) , 2) [https://www.w3schools.com](https://www.w3schools.com/)