**Hotel Booking Analysis**

Vikash Kumar

Data Science Trainees

AlmaBetter, Bengalore

**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 been removed from the data.

**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:**

Hotel Booking is ruled by way of many elements consisting of the time of the year, wide variety of guests, distribution channel, hotel type, etc.

The predominant thing of this undertaking is to operate Exploratory information evaluation and draw insights to recognize all the vital elements that govern the Hotel bookings.

**Understanding the data:**

The given data set has 3 types of data:

* Numerical data
* Categorical data
* Binary Data

**Data Summary**

The provided data set has following different columns of variables necessary for hotel bookings:

* hotel: The category of hotels, which are two city hotel and resort hotel.
* is\_cancelled : The value of column shows if the booking was canceled or not.
* Values[0,1], where 0 indicates not canceled and 1 indicates the cancellation.
* lead\_time : The time between reservation and actual arrival.
* stayed\_in\_weekend\_nights: The number of weekend nights stay per reservation.
* stayed\_in\_weekday\_nights: The number of weekday night stays per reservation.
* meal: Meal preferences per reservation, which are
* Country: The origin country of guest.
* market\_segment: This column shows how reservation was made and what is the purpose of reservation.
* distribution\_channel: The medium through booking was made. [Direct, Corporate, TA/TO, undefined, GDS.]
* Is\_repeated\_guest: Shows if the guest is who has arrived earlier or not. Values[0,1]-> 0 indicates no and 1 indicated yes person is repeated guest.
* days\_in\_waiting\_list: Number of days between actual booking and transact.
* customer\_type: Type of customers (Transient, group, etc.)
* Undefined/SC – no meal package and dinner)
* distribution\_channel: The medium of booking was [Corporate, Direct,
* GDS, TA/TO, undefined]

**Data Summary:**

1. Undefined/SC – no meal package
2. BB – Bed & Breakfast
3. HB – Half board (breakfast and one other meal – usually dinner)
4. FB – Full board (breakfast, lunch and dinner)
5. TA –Travel agency
6. TO –Tour operator
7. GDS – Global Distribution System.

**Types of Hotels:**

1. Resort Hotel
2. City Hotels

**How hotel booking works?**

Hotel reservations are made by guests before they arrive at the hotel. Bookings are received from the following distribution channels:

1. Direct
2. Corporate
3. TA/TO
4. GDS

Depending on availability and demand, the hotel assigns a room. If the requested room is unavailable a different room is assigned.

The Hotel receives no deposit, or refundable deposit or non-refundable deposit against the bookings.

**Steps involved:**

**Importing important libraries:**

During this step, our main goal was to import all the necessary libraries to help us explore

the problem statement and perform EDA to draw conclusions from the data.

**Understanding the data set:**

Next, we worked on checking the data set. We need to know how many rows and columns are available and what columns could be important in solving the problem statement.

**Null values Treatment:**

Our data set contains a large number of null values which might tend to disturb our insights. As a result, we replaced them with '0' for numerical data and 'undefined' for categorical data.

**Exploratory Data Analysis:**

After treating the null values, we started with the EDA. We performed EDA.

The following analysis was used to solve the problem statement during EDA.

**Uni-variate Analysis:**

We answered the following questions when performing univariate analysis on a hotel booking data set:

1. Which distribution channel gave most of the bookings?
2. Most customers are from Portugal.
3. Type A room is in most demand.
4. Which room type is in most demand?
5. From which country most of the customers are coming?
6. What is the most popular meal among customers?

**Hotel wise Analysis:**

While doing hotel-wise analysis of the given hotel booking data set, we answered following questions:

1. Which hotel type was more engaging and in most demand?
2. Which hotel type receives more guests?
3. Which hotel type makes more revenue?
4. Which type of guests have the most check-ins?
5. What is the most preferred length of stay in each hotel?
6. Which hotel has higher and how much customer returning rate?

**Time wise Analysis:**

We answered the following question based on time-wise analysis of given hotel booking data:

1. What are the busiest months for hotels?
2. Some other questions:
3. Which hotel type received the most special requests?
4. What is the effect of deposit type on profit?
5. How many special requests were received?

**Observations:**

As a result of the univariate analysis, hotel-wise analysis, and time-wise analysis, the following conclusions were reached:

**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.

1. About 66% of the reservations are for the City Hotel and 34% of the bookings are for the Resort Hotel, so City Hotel is busier than the Resort Hotel. So does the overall adr.

2. The majority of guests stay less than 7 days at the hotel and for longer stays Resort Hotel is preferred.

3. Both hotels have significantly higher booking cancellation 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.

4. The majority of the visitors were from European nations, with the majority being from Portugal.

5. Customers book through many channels, with TA/TO being the most popular option.

6. Hotels should improve their appeal on the GDS channel because here is where they can find higher adr deals.

7. Nearly 30% of reservations made via TA/TO are cancelled.

8. Booking cancellations are unaffected by not receiving the exact accommodation that was reserved, lengthier lead times, and waiting times. Although adr is reduced by varied room allocation.

9. The busiest and most lucrative months for both hotels are July and August.

10. A month's worth of adr steadily rises toward the end of the month, with brief spikes on weekends.

11. Since couples are the most frequent hotel customers, hotels can design services to meet their needs in order to boost income.

12.There are more special requests when there are more guests present.

13. Adults and adjacent market segment bookings typically have a significant number of customised requests.

14. Customers can typically find better deals with low adr for longer stays (more than 15 days).