

Hotel Booking Analysis

Adil Khan, Sunil Panigrahi, Shubham Kumar, Vivek Singh

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.

2. Introduction

The Hotel booking data set includes Numerical, Categorical and Binary data. The data set has columns like the hotel type, is_canceled, arrival_date_year, arrival_date_month, stays_in_weekend_nights, stays_in_week_nights, country, market_segment, distribution_channel, etc. which helped us draw major insights from the data set. Our aim here is to understand the important factors that governs the hotel bookings.

In this project, we will perform exploratory data analysis with python to get insight from the data.

1.Problem Statement

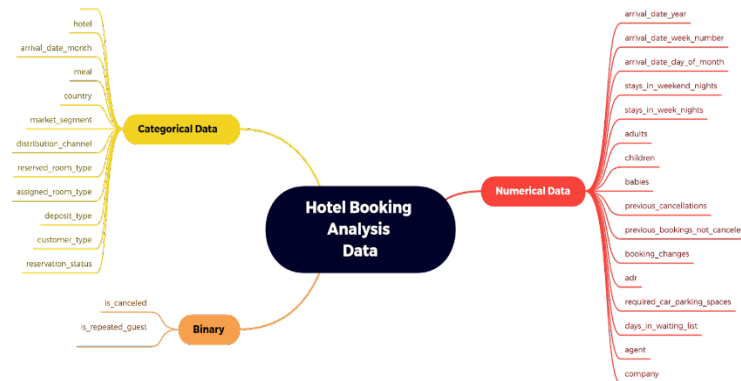
Hotel Booking is governed by many factors including the time of the year, number of guests, distribution channel, hotel type, etc.

The main aspect of this project is to perform Exploratory data analysis and draw insights to understand all the important factors that govern the Hotel bookings.

3. Understanding the data

The given data set has 3 types of data:

- Numerical data
- Categorical data
- Binary Data



3.1 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:
 - 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)
- **Country:** The origin country of guest
- **market_segment:** This column shows how reservation was made and what is the purpose of reservation. Eg, corporate means corporate trip, TA for travel agency and TO for tour operators.
- **distribution_channel:** The medium of booking was [Corporate, Direct, GDS, TA/TO, undefined]
- **Is_repeated_guest:** Shows if the guests have arrived the hotel before or not. Values[0,1]-->0
- indicates no, and 1 indicates yes.

- **days_in_waiting_list:** Number of days between actual booking and the transaction.
- **customer_type:** Type of customers (Contract, group, transient, transient party, etc.)

3.2 Data Summary

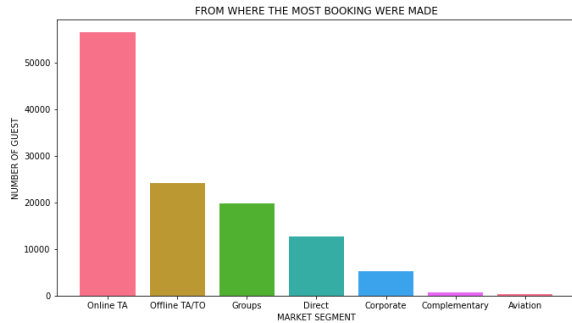
- Undefined/SC – no meal package
- BB – Bed & Breakfast
- HB – Half board (breakfast and one other meal – usually dinner)
- FB – Full board (breakfast, lunch and dinner)
- TA –Travel agent
- TO –Tour operator
- GDS – Global Distribution System.

4. Types of Hotels

- Resort Hotel
- City Hotels

5. How hotel booking works?

- The Hotel booking is done by the guests before visiting the hotel. The hotel receives the booking from the following distribution channels:
 - Direct
 - Corporate
 - TA/TO
 - GDS

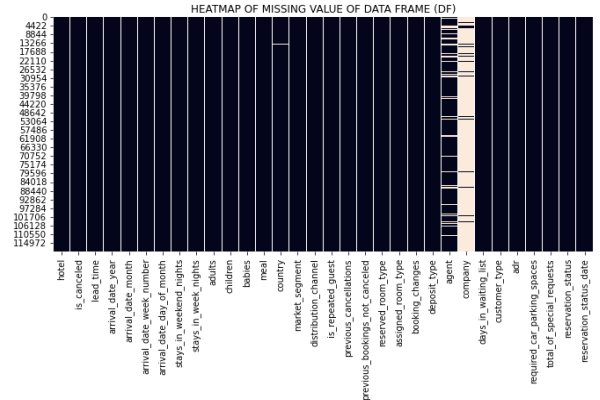


- The Hotel then assign a room as per the availability and demand. If the demanded room is unavailable the hotel assigns a different room.
- The Hotel receives no deposit, or refundable deposit or non-refundable deposit against the bookings.

6. Steps involved:

- **Importing important libraries**
Our main motive through this step was to import all the important libraries to help us explore the problem statement and perform EDA to draw conclusion on the basis of the data set.
- **Understanding the data set**
Next, we worked on checking the data set. How big the data set is? How many rows and columns are available? What could be the important columns to solve the problem statement? How many null values we have in the data set? We imported the important libraries along with our data set.
- **Null values Treatment**
Our dataset contains a large number of null values which might tend to

disturb our insights. Hence, we replaced them with '0' for numerical data and 'undefined' for 'categorical data' to get a better result.



• **Exploratory Data Analysis**

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

7. Exploratory Data Analysis:

While doing EDA we used the following analysis to solve the problem statement:

Univariate Analysis:

While doing univariate analysis of given hotel booking dataset, we answered the following questions:

- Which distribution channel gave most of the bookings?
- Which room type is in most demand?
- From which country most of the customers are coming?
- What is the most preferred meal by customers?

Hotel wise Analysis:

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

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

Time wise Analysis:

- While doing time-wise analysis of given hotel booking dataset, we answered the following question:
- What are the busiest months for hotels?

Some other important questions:

- How is the deposit type affecting the profit?
- Which hotel type received more special requests?
- How many special requests were received?

8. Observations:

After performing the univariate analysis, Hotel wise analysis, Time wise analysis and finding answers to the other important questions, we observed the following:

- Type A room is in most demand.
- Most bookings were made from online travel agents.
- Most customers are from Portugal.
- Most customers prefer the Bed and breakfast meal type.
- City hotels are more engaging and in demand. Hence, city hotels are more profitable.
- City hotels receive a higher number of guests than the resort hotels. However, the cancellation rate is more for city hotels.
- City hotels have a higher number of repeat customers. However, the ratio of repeat customers is more for resort hotels as the city hotels receive a higher number of guests than the resort hotels.
- Non-refundable deposits tend to convert to a successful transaction.
- May is the busier and most profitable month for the hotels in 2017.
- On an average, May-June is the busier and most profitable month for the hotels.
- City hotels have received the most special requests.
- 1 number of requests is the highest.
- 1 car parking was needed for most of the guests.

8. Conclusion:

That's it! We reached the end of our exercise.

Starting with loading the data so far, we have done EDA, null values treatment, encoding of categorical columns, and found out major reasons that governs the hotel booking and steps on how we can increase it.

References-

1. Stack overflow
2. X-mind