

Hotel booking Analysis- EDA
Sandesh Gaikwad
Data Science Trainee,
AlmaBetter

Abstract :

For any hotel it is important to know how much booking the hotel is going to get in particular months. Knowing it partially helps the hotel to maintain resources so that the best service can be provided with optimal use of resources.

It is also essential for the hotel to know what type of booking it gets and which type of booking it should look for so that it can make more profit. How much deposit should be taken so that the people will avoid cancellation or even though the canceled hotel booking makes some money.

This dataset will help look into the parameters which make customers cancel their bookings.

Keywords: EDA, Seaborn, Hotel Booking, Customer Retention, Hotel profitability

Problem Statement :

Look at the data and study the booking pattern by month so that next month's resources can be managed prior.

Look for the cancellation rate and see if there is any pattern and what methods can be used to reduce the cancellation

About Dataset

- **Hotels:** data for two hotels is available Resort Hotel and City Hotel
- **Lead time:** Number of days booking is done before arrival date
- **Arrival date:** Three columns with arrival month, arrival date, arrival year
- **Stay days:** Weekend and weekdays
- **People:** Number of children, babies, and adults
- **Market segment:** type of customers grouped together

- **Repeated guest** - whether customers are repeated or not
- **The previous cancellation** - has the customer booked the hotel and canceled the booking
- **Room type** - which type of room was booked and which type of room was assigned
- **Deposit type**; whether customers asked to give any deposit
- **Avg Daily rate** - How much avg daily cost for staying
- Some other parameters such as Car parking space, meal-type, status date, special request

Steps involved in EDA:

1. **Null value treatment:** The country and children column was having a few missing values that were replaced with the mode of the column. The company and agent columns had significant data missing so both columns were dropped from further analysis
2. **Column-wise data analysis:** Each column was analyzed and visualization tools are used. some conclusions were made from the eda

Observations :

- **Hotels:** Out of all bookings 66.4 % of bookings are for the city hotel and 33.6 % of bookings are for the resort hotel. The city hotel also has the highest cancellation compared to the resort hotel. The city hotel has cancellations of 41 % on the other hand resort hotel has fewer cancellations
- **Repeated Customers:** Overall repeated guests are very few compared

to new guests Resort Hotel has 4.5 percent repeated guests while the city has only 2.7 percent repeated guests. The resort also has significantly lower cancellations compared to the city hotel. Cancellation for new customers is almost the same for both Resort and City hotel

- **Deposit Type:** We can see that most of the bookings are for the No deposit category. The cancellation is significantly higher for the non-refundable category. Refundable has very few bookings so it won't be fair to compare. It is very contrary to what should ideally happen but the good thing is the hotel is making a few bucks from these bookings. Probably hotel has its own system which checks the probability of cancellation and asks for a deposit
- **Previous Cancellation:** Repeated customers with no previous cancellation have a very less percentage of booking cancellations. People who have canceled once have more cancellations. The ones who have canceled twice have small cancellations but the numbers are small so it can't be trusted and seems counterintuitive. Compared with one cancellation overall people with more cancellations have a higher chance of canceling the booking.
- **Meal:** Most people book their room with Breakfast costlier items can be added to the menu
- **Distribution Channel:** Travel agents are responsible for bringing most of the bookings, accounting for 82.5 % of the total booking
The second and Third highest booking comes from direct bookings and corporate bookings
- **Lead time:** Bookings with more than 150 days of lead time mostly get canceled
- **Room Type:** Most of the bookings are for low-cost rooms. Most high-value

customers choose the resort hotel while low-value customers choose the city hotel

Summary and Suggestions :

- City Hotel gets higher booking than Resort hotel
- The resort hotel has a lower cancellation rate than the City hotel
- Weekends are more crowded than weekdays
- Resort hotel gets the most high-value customers
- Hotels should charge some booking deposit for the following types of people :
 - People who have a lead time greater than 100 days
 - Customers with previous cancellations
 - Large group bookings
 - Crowded months (July, August) so that full utilization can happen

References

- Stack overflow
- Google
- Official documentation
- Medium blogs