

# **Capstone Project Submission**

## **Hotel Booking Analysis**

### **Project Summary**

**Project Type** - Exploratory Data Analysis

**Contribution** - Individual Project

**Member** - Onkar Anil Pawar

**E-mail** – pawaronkar177@gmail.com

## Problem Statement:

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

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.

Explore and analyze the data to discover important factors that govern the bookings.

## Project Summary

---

For this hotel booking analysis, the goal was to explore the customer data of a hotel and identify any potential trends or correlations. This exploratory data analysis (EDA) aimed to explore the hotel booking data set and identify potential relationships between key variables.

The data set included customer booking information. As part of the analysis, descriptive statistics were calculated for each variable, and visualizations were created to explore the relationships between various variables. To get insight from the dataset, we built a variety of charts, including a count plot, bar plot, kdeplot, heatmap, pair plot, violin plot, and boxplot.

The data set was composed of over 119390 hotel bookings, each containing several variables such as 'hotel', 'is\_canceled', 'lead\_time', For this hotel booking analysis, the goal was to explore the customer data of a hotel and identify any potential trends or correlations. This exploratory data analysis (EDA) aimed to explore the hotel booking data set and identify potential relationships between key variables.

The data set included customer booking information. As part of the analysis, descriptive statistics were calculated for each variable, and visualizations were created to explore the relationships between various variables. To get insight from the 'arrival\_date\_year', 'arrival\_date\_month', 'arrival\_date\_week\_number', 'arrival\_date\_day\_of\_month', 'stays\_in\_weekend\_nights', 'stays\_in\_week\_nights', 'adults', 'children', 'babies', 'meal', 'country', 'market\_segment', 'distribution\_channel', 'is\_repeated\_guest', 'previous\_cancellations', 'previous\_bookings\_not\_canceled', 'reserved\_room\_type', 'assigned\_room\_type', 'booking\_changes', 'deposit\_type', 'agent', 'company', 'days\_in\_waiting\_list', 'customer\_type', 'adr', 'required\_car\_parking\_spaces', 'total\_of\_special\_requests', 'reservation\_status', and 'reservation\_status\_date'.

Dataset variables are in int64, float64, and object datatypes. There are 32 variables: 12 variables are objects, 16 are int64, and 4 are float64. 31994 duplicate values were removed. The variables country had 452, children had 4, agent had 12193, and company had 82137 null values. We replaced the null value with the mode of each variable (country, children, agent) for these variables, but the variable "company" had more than 50% null value, so we removed it. Further, we removed outliers from lead\_time and adr. The final dataset had 87396 observations.

I also changed the data types of variables children, agent, and reservation\_status\_date to int64, int64, and datetime64, respectively. We performed some feature engineering for more convenience and created new variables: total\_stays, total\_people, total\_childrens, reserved\_room\_assigned, guest\_category, and lead\_time\_category. Now total\_people and total\_childrens are in the floated 64 datatypes, so we converted them to int64. Now I removed it from the observation because having total\_people at 0 made no sense.

## Conclusion:

1. Top Hotel - City Hotel. Top meal - Bread and Breakfast. Top Agent - Agent No. 9. Top room type - A
2. One out of every three bookings is canceled.
3. People prefer to tour more in August.
4. Most preferred meal is BB(Bread and Breakfast.
5. Online marketing is the best way to attract customers.
6. People don't want to pre-deposit the money for booking.
7. Only 10% of people require parking space.
8. Most of the visitors are couples.
9. Resort hotel is preferred mostly for a longer stay, and daytime stays. and when the parking space is needed.
10. More than 15 days advance bookings have a high chance of cancellation.
11. Assigning a different room is not a reason for cancellation.
12. Direct bookings have very less cancellation%.
13. Best time to book a hotel is in January.
14. Average days in advance booking: 71 days
15. Average nights spent by visitors: 3
16. Most visitors are from these countries: Portugal, Britain, France, Spain, and Germany.
17. Total Special requests and the revenue depends more on the total number of members who arrived.

Github Link:- <https://github.com/Onkar-TAE/Hotel-Booking-Analysis.git>