

# Exploratory Data Analysis

## Capstone Project -1

### Hotel Booking Analysis

**Team Members:-**

1. Pankaj Rathod
2. Rahul Chouhan

**Data science trainees,**

**AlmaBetter, Bangalore**

---

#### **Abstract:**

The hotel and tourism industry plays an essential role in the economic growth of any country. In this project, we will do exploratory data analysis on the hotel booking dataset, in order to assist hotels in making more profits.

Through feature selection, data analysis, and prediction with data visualisation, we can understand what could be the factors behind the data

**1.Problem Statement:-** Helping Hotel business studying the past records and help the Hotel industry to grow exponentially

- Are you curious about hotel booking trends?
- When it comes to staying, how long do people stay?
- Is there a high rate of cancellations?
- When is the busiest time of year?

The purpose of this analysis is to examine these questions using a large dataset.

**2. Introduction:-** 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.

For this project, we used Python libraries like Pandas, Matplotlib, Seaborn, etc. to examine, clean, and analyze the "Hotel Booking Analysis" dataset

**3.Data Contain:-** The data of Hotel booking Was given for Exploratory Data Analysis. It contains Total 32 columns and 119390 rows.

Taking a look at the column with details (excluding irrelevant columns).

1. **hotel** – Hotel type(city hotel or Resort hotel)
2. **is\_canceled** – Booking cancellation
3. **lead\_time** - the number of days between the time a guest books their room and the time they are scheduled to arrive at the hotel.
4. **arrival\_date\_year** – year of arrival (2015,16 or 17)
5. **arrival\_date\_month** – Month of arrival(Jan to dec)
6. **arrival\_date\_week\_number** – Week no of arrival in year out of 52 or 53
7. **arrival\_date\_day\_of\_month** – arrival day in moth
8. **stays\_in\_weekend\_nights** – How many days people lived on weekend
9. **stays\_in\_week\_nights** – How many days people lived on week nights
10. **adults** – Adult members stayed in hotel
11. **children** – child members stayed in hotel
12. **babies** – Baby member stayed in hotel
13. **meal** – Meal type
  - 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)
14. **country** – Visitors country
15. **market\_segment** - “TA” means “Travel Agents” and “TO” means “Tour Operators”
16. **distribution\_channel** - The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
17. **is\_repeated\_guest** – repeated guest in hotel

18. **previous\_cancellations**
19. **previous\_bookings\_not\_canceled**
20. **reserved\_room\_type** – Categories in room type
21. **assigned\_room\_type** – Room type allotted to customer
22. **booking\_changes** – Charges of booking hotel
23. **deposit\_type** – No deposit, refundable, non-refundable
24. **agent** – Booking agent
25. **company** – Booking done by company
26. **days\_in\_waiting\_list** – Waiting days
27. **customer\_type** –
  - **Contract** — when the booking has an allotment or other type of contract associated to it;
  - **Group** — when the booking is associated to a group;
  - **Transient** — when the booking is not part of a group or contract, and is not associated to other transient booking;
  - **Transient-party** — when the booking is transient, but is associated to at least other transient booking
28. **adr** - Average Daily Rate
29. **required\_car\_parking\_spaces** – requirement of car parking spaces
30. **total\_of\_special\_requests** – Special requests
31. **reservation\_status** -Check in or check out
32. **reservation\_status\_date** – Check in date or check out date

## 4. Steps involved:-

- **Exploratory Data Analysis**

In order to gain a better understanding of the dataset, we performed EDA and visualized some insights column-by-column. For our simplicity, we merged the column babies and children under the kid's column and we again merged the column kids and adults under total members. We looked only at some

important columns which are going to make differences in the growth of the Hotels.

- **Null values Treatment**

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

## **5. Conclusion:**

1. Almost 67% Bookings are for City hotel and 33% Bookings are for Resort hotel, therefore City hotel is busier than resort hotel
2. Most of the customer prefer to stay one, two, three and four nights. Where number of bookings for two and three nights has high demand.

So if we try to analyse the optimal length to stay in order to get best daily rate, we can say that two to four days of accommodation can be more affordable since more people prefer the same, the hotels also launch the offers for mid term accommodation because long term stay offers will consume their rooms and new customers can not avail the rooms and hotels may fail to generate extra profits.

3. In total, 119000 customers checked in at the hotel, while 75000 canceled their reservations. In percentage terms, 37% of bookings were canceled, whereas 63% of customers checked in. As a result, we realize that no deposit policies may account for the high rate of cancellations
4. Hotel bookings seem to be at their highest level in 2016 so far. Bookings also increase around the middle of the year, with August being the highest followed by July and May. Around August, summer ends and autumn begins. The summer period seems to be the busiest time for hotel bookings. It appears the Winter period has the lowest number of hotel bookings, followed by November and December.

So for the Winter season price of hotel booking should be low or they should provide some special discount so that most of the customers are attracted to the hotel and also it can be the best time of year to book a hotel room. Due to less demand for rooms, the cost for rooms on daily basis is also minimum as compared to other months whereas the month of August has high demand for rooms so it is obvious that the cost of rooms is also at a peak.

5. The demand for city hotels is always higher than that for resort hotels. It means the majority of the hotels booked are city hotel. So definitely need to spend the most targeting fund on these hotels.

6. we can see that 75% of customer is of transient type which is 3/4 th of the customer based and 21% of customer is of transient-party type.
7. According to the representation, people interested in city hotel to stay for two, three, one and four days respectively where for resort hotel people interested to stay one, seven, two and three days respectively.
8. Most of bookings are created from Portugal and then Great Britain in second place and France at third place
9. It is seen that most number of customer are containing 2 adults customer and those who have more than 2 either containing adults, children & babies have the lowest number of customer.