**Hotel Bookings Exploratory Data Analysis**

**Objective**

We are provided with a hotel bookings dataset.

Our main objective is to perform EDA on the given dataset and draw useful insights about general trends in hotel bookings and how factors govern hotel bookings.

**Dataset**

We are given a hotel bookings dataset. The dataset contains booking information for both City hotel and Resort hotel. The below are columns and their classifications.

- hotel: City Hotel and Resort Hotel.

- is\_canceled: Whether the booking is cancelled or not (0 for not cancelled and 1 for cancelled).

- lead\_time: time (in days) between booking transaction and actual arrival.

- arrival\_date\_year: Year of arrival.

- arrival\_date\_month: Month of arrival.

- arrival\_date\_week\_number: Week number of arrival date.

- arrival\_date\_day\_of\_month: Day of month of arrival date.

- stays\_in\_weekend\_nights: No. of weekend nights spent in a hotel.

- stays\_in\_week\_nights: No. of weeknights spent in a hotel.

- adults: No. of adults in single booking record.

- children: No. of children in single booking record.

- babies: No. of babies in single booking record.

- meal: Type of meal chosen (BB, HB, SC ….).

- country: Country of origin of customers (PRT, GBR, FRA, ESP, DEU….).

- market\_segment: Booking through medium (TA/TO, Direct, Corporate, ….).

- distribution\_channel: Medium through which bookings were made (TA/TO, Direct, Corporate, ….).

- is\_repeated\_guest: Whether the guest visited before(0 for No and 1 for Yes)

- previous\_cancellations: No. of previous cancelled bookings.

- previous\_bookings\_not\_canceled: No. of previous non-cancelled bookings.

- reserved\_room\_type: Room type reserved by guests.

- assigned\_room\_type: Room type assigned to the guests.

- booking\_changes: No. of booking changes done by guests.

- deposit\_type: Type of deposit at the time of making a booking (No deposit/ Refundable/ No refund).

- agent: I’d of agent for booking.

- company: I’d of the company making a booking.

- days\_in\_waiting\_list: No. of days on waiting list.

- customer\_type: Type of guests(Transient, Group,…..).

- adr: Average Daily rate.

- required\_car\_parking\_spaces: No. of car parking asked during booking rooms.

- total\_of\_special\_requests: total no. of special request made by guests.

- reservation\_status: Whether guest checked out, cancelled bookings or not failed to show up.

- reservation\_status\_date: Date of making reservation status.

* Total number of rows in data: 119390
* Total number of columns: 32

**Data Cleaning and Feature Engineering**

1. **Removing Duplicate rows**

All duplicate rows were dropped.

Total number of rows cleaned: 31994

1. **Handling null values**

* Null values in column country was replaced by “na”.
* Null values in column agent was replaced by 0.
* Null values in column company was replaced by 0.

1. **Removing columns not required in analysis**

* Columns adults, children, babies.
* Columns stays\_in\_week\_nights and stays\_in\_weekend\_nights.

1. **Creating new columns**

* Column guests adding adults, children, babies.
* Created new column total\_hotel\_stay by adding stays\_in\_week\_nights and stays\_in\_weekend\_nights.

**Exploratory Data Analysis**

Performed EDA and tried answering the following questions:

Q1) Which is the popular of both the hotels?

Q2) Are bookings more compared to cancellations?

Q3) What is the total cancellation for both the hotels and which hotel has higher cancellation rate?

Q4) What is the preferred duration for stay in the hotels?

Q5) Which hotel has highest chance of guests repeating their stay?

Q6) Which of the distribution channels should the hotels prefer?

Q7) What is the classification of guest arriving the hotel?

Q8) Which are the countries the hotels receive most guests from?

Q9) Which are the busiest months for the hotel?

Q10) Which has been the busiest of years for both the hotels?

Q11) Which rooms are the most utilized in both the hotels?

Q12) What is the most preferred meal choice opted by guests at the time of booking for both the hotels?

Q13) Which of the hotels guests demand special requests?

Pictoral representations were mainly performed using Matplotlib and Seaborn library and the following graphs and plots had been used:

* Bar Plot.
* Count Plot.
* Pie Chart.
* Line Plot.
* Heatmap.

**Analysis:**

Performed univariate analysis and made following conclusions:

1.) City hotel is the popular of both the hotels.

2.) There are more number of bookings compared to cancellations. The ratio for booking is 72.5%.

3.) City hotel has 30% cancellation as compared to 23% of Resort hotel.

4.) Guests mostly prefer upto 4 nights package stay.

5.) Resort hotel has guests repeating often.

6.) TA/TO, direct and corporate.

7.) Transient guests are more in both the hotels.

8.) Guests from Western European countries frequent the hotels more.

9.) May, July & August for City hotel and July and August for Resort hotel are the busiest.

10.) 2016 has been the busiest year for both the hotels.

11.) Rooms A, D, E and F are the most utilised in both the hotels.

12.) BB is most popular for both the hotels.

13.) City hotel guests make a lot of special requests.

**Conclusion**

1. For a guest to get a good hotel deal, they should visit either of the hotels before or after the month of August and September. City hotel prices remanin constant throughout the year, so **\*\*Resort hotel is the best pick to get the best deals.\*\***

2. We can see the we have a higher ADR in the middle of the year (ie July, August) and lowers down towards the end of the year. But after November again the ADR sees a increasing trend. A higher ADR means good revenue for the hotels.

3. City hotel has high no of special requests by guests.

4. City hotel is the most visited hotel although it has 30% of booking cancellations compared to 23% of cancellation of Resort hotel.

5. Most of the guests prefer a four night stay at the hotels which seems like a ideal period. So to increase the revenue of the hotels special packages can be derived to increase the rush in the hotels.

6. There is also a surprising trend. Resort hotel guests tend to repeat more than City hotel. It is conclusive that City hotel needs to look into key hospitality parameters so that they have more guests frequenting their hotel.

7. Also most of the bookings come from TA/TO ie tours and travels, followed by Direct clients who are walkins at the hotel and then Corporate bookings So the hotels need to incentivise the TA/TO(tours and travel operators) and the transient customers more to gain in the no of guest volume.

8. Most of the guests are from Western Europe, namely Portugal, France, UK. Thereby we should concetrate on advertising in those regions so that better business can be achieved.

9. Rooms A, D, E, F are most utilised. The rest of the rooms are not frequently used. Further analysis needs to be done to increase the yield from those rooms.

10. Further analysis needs to be done as to why 2017 there was a dip in volume of guests for both the hotels.

**Challenges**

1. There was a lot of duplicate data.

2. Data columns had lot of 0 and NAN values.

3. Choosing which data to use was difficult.

4. How to present the data was a challenge.