**Hotel Bookings Exploratory Data Analysis**

**Objective**

**We have a dataset on hotel bookings.**

**Our main goal is to perform an EDA on the given dataset and gain useful insights into the overall trends in hotel bookings and how factors drive hotel bookings.**

**Dataset**

**We are given a dataset on hotel bookings. This dataset contains booking information for a city hotel and a resort hotel. It contains the following features.**

**- 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**

* **Creating the copy of the dataset**
* **checking for duplicate values**
* **checking the missing values**
* **Replace the null values by 0 and NA in columns**
* **Creating new column Total guest**

**Exploratory Data Analysis**

**Performed EDA and tried answering the following questions:**

* **Q1- Which is the most booked hotel**
* **Q2- which is busiest month in the year**
* **Q3- Which hotel has the most booking for each year.**
* **Q4- Which country has the highest booking and guest.**
* **Q5- What is proportion of non-cancelled and cancelled reservations.**
* **Q6- Which rooms in each hotel are the most frequently used.**

**The Matplotlib and Seaborn libraries were mostly used for pictorial representations, and the following graphs and plots were created:**

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

**Conclusion**.

**After a thorough analysis of the provided csv data file, I came to the conclusion that**

* **City hotel has more booking than resort hotel. Overall, there are 27.5% Cancelled and 72.5% Non-Cancelled reservations.**
* **In the month of August and July hotel has more booking so there is lack vacant rooms available if, want to visit come in December and November months as there is less booking compared to peak times.**
* **In 2015 resort hotel has more booking whereas in 2016 city hotel has more booking.**
* **Most of the guests came from PRT (Portugal), followed by GBR (United Kingdom), FRA (France), ESP (Spain) and DEU (Germany).**
* **Room A was the most used in both hotels, followed by Room D. However, room C at the resort hotel was unused while room B at the city hall was no longer in use.**

**Challenges**

* **There was a lot of duplicate data.**
* **Data columns had lot of 0 and NAN values.**
* **It was difficult to decide how to convey the data.**