***Hotel Bookings***

***Exploratory Data Analysis***

***Objective***

*We are provided with a hotel bookings dataset.*

*Our main objective is perform EDA on the given dataset and draw useful conclusions about general trends in hotel bookings and how factors governing hotel bookings interact with each other.*

***Dataset***

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

***Data Description:***

*1)* ***hotel*** *: Hotel(Resort Hotel or City Hotel)*

*2)* ***is\_canceled*** *: Value indicating if the booking was cancelled (1) or not (0)*

*3)* ***lead\_time*** *: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date*

*4)* ***arrival\_date\_year*** *: Year of arrival date*

*5)* ***arrival\_date\_month*** *: Month of arrival date*

*6)* ***arrival\_date\_week\_number*** *: Week number of year for arrival date*

*7)* ***arrival\_date\_day\_of\_month*** *: Day of arrival date*

*8) stays\_in\_weekend\_nights : Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel*

*9)* ***stays\_in\_week\_nights*** *: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel*

*10)* ***adults*** *: Number of adults*

*11)* ***children*** *: Number of children*

*12)* ***babies*** *: Number of babies*

*13)* ***meal*** *: Type of meal booked. Categories are presented in standard hospitality meal packages:*

*14)* ***country*** *: Country of origin.*

*15)* ***market\_segment*** *: Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”*

*16)* ***distribution\_channel*** *: \*Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”\**

*17)* ***is\_repeated\_guest*** *: Value indicating if the booking name was from a repeated guest (1) or not (0)*

*18) previous\_cancellations : Number of previous bookings that were cancelled by the customer prior to the current booking*

*19)* ***previous\_bookings\_not\_canceled*** *: Number of previous bookings not cancelled by the customer prior to the current booking*

*20)* ***reserved\_room\_type*** *: Code of room type reserved. Code is presented instead of designation for anonymity reasons.*

*21)* ***assigned\_room\_type*** *: Code for the type of room assigned to the booking.*

*22)* ***booking\_changes*** *: Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation*

*23)* ***deposit\_type*** *: Indication on if the customer made a deposit to guarantee the booking.*

*24)* ***agent*** *: ID of the travel agency that made the booking*

*25)* ***company*** *: ID of the company/entity that made the booking or responsible for paying the booking.*

*26)* ***days\_in\_waiting\_list*** *: Number of days the booking was in the waiting list before it was confirmed to the customer*

*27)* ***customer\_type*** *: Type of booking, assuming one of four categories*

*28)* ***adr*** *: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights*

*29) required\_car\_parking\_spaces : Number of car parking spaces required by the customer*

*30)* ***total\_of\_special\_requests*** *: Number of special requests made by the customer (e.g. twin bed or high floor)*

*31)* ***reservation\_status*** *: Reservation last status, assuming one of three categories*

*\* Cancelled – booking was cancelled by the customer*

*\* Check-Out – customer has checked in but already departed*

*\* No-Show – customer did not check-in and did inform the hotel of the reason why*

*32)* ***reservation\_status\_date*** *: Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking cancelled or when did the customer checked-out of the hotel*

⬤ Total number of rows in data: 119390

⬤ Total number of columns: 32

***Data Cleaning***

***1) Removing Duplicate rows***

*All duplicate rows were dropped.*

***2) Handling null values***

*Null values in columns company and agent were replaced by 0.*

*Null values in column country were replaced by 'others'.*

*Null values in column children were replaced by the mean of the column.*

***3) Converting columns to appropriate data types***

*Changed data type of children, company, agent to int type.*

*Changed data type of reservation\_status\_date to date type.*

***4) Creating new columns***

*Created a new column total\_stay by adding stays\_in\_weekend\_nights+stays\_in\_week\_nights.*

*Created a new column total\_people by adding adults+children+babies.*

***Exploratory Data Analysis***

*Performed EDA and tried answering the following questions:*

***Univariate Analysis:***

*Performed univariate analysis and made following conclusions*:

*1) Which type of hotel is mostly prefered by the guests?*

*2) Which Agent made the most bookings?*

*3)What is the percentage of cancellation?*

*4) What is the Percentage of repeated guests?*

*5) What is the percentage distribution of "Customer Type"?*

*6)What is the percentage distribution of required\_car\_parking\_spaces?*

*7)What is the percentage of booking changes made by the customer?*

*8)What is Percentage distribution of Deposit type ?*

*9) Which type of food is mostly preferred by the guests?*

*10) Which is the most preferred room type by the customers?*

*11)In which month most of the bookings happened?*

*12) Which Distribution channel is mostly used for hotel bookings?*

***Bivariate and Multivariate Analysis****:*

*We tried to answer following questions*

*1) Which Hotel type has the highest ADR?*

*2) Which hotel type has the more lead time?*

*3) Which hotel has the highest percentage of booking cancellation?*

*4) Which hotel has a longer waiting time?*

*5) Which Hotels have the most repeat guests?*

*6) ADR across the different months.*

*7) Which distribution channel has the highest cancellation rate?*

*8)Which Market Segment has the highest cancellation rate?*

*9) Do the guests alloted the same room type which was reserved by them?*

*10)What is the Optimal stay length in both types of hotels ?*

***Conclusion***

*1) City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.*

*2) 27.5 % bookings were cancelled out of all the bookings.*

*3) Only 3.9% of people were revisited the hotels. Rest 96.1 % were new guests. Thus retention rate is low.*

*4) The percentage of 0 changes made in the booking was more than 82 %. Percentage of Single changes made was about 10%.*

*5) Most of the customers (91.6%) do not require car parking spaces.*

*6) 79.1 % bookings were made through TA/TO (travel agents/Tour operators).*

*7) BB( Bed & Breakfast) is the most preferred type of meal by the guests.*

*8) Average ADR for city hotels is high as compared to resort hotels. These City hotels are generating more revenue than the resort hotels.*

*9) Booking cancellation rate is high for City hotels which is almost 30 %.*

*10) Average lead time for resort hotels is high.*

*11) Waiting time period for City hotels is high as compared to resort hotels. That means city hotels are much busier than Resort hotels.*

*12) Resort hotels have the most repeated guests.*

*13) In "TA/TO", City hotels have a higher cancellation rate compared to resort hotels.*

*14 )In "direct" both the hotels have almost the same cancellation rate.*

*15) Online T/A' has the highest cancellation in both type of cities*

*16) Almost 19 % of people did not cancel their bookings even after not getting the same room which they reserved while booking a hotel. Only 2.5 % of people cancelled their booking.*

*17) Optimal stay in both the type hotel is less than 7 days*

***Challenges***

*(1) There was a lot of duplicate data.*

*(2) Data was present in the wrong datatype format.*

*(3) Choosing appropriate visualisation techniques to use was difficult.*

*(4) A lot of null values were there in the dataset.*