***HOTEL BOOKING***

***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 Agent made the most no. bookings?

# 2) What is the Percentage of repeated guests?

# 3) What is the percentage of cancellation?

# 4) Which months have cheaper booking rates?

# 5) Which countries should be targetted for promotion?

# 6) No. of customers repeated their bookings.

# 7) Does people with children have a particular preference for the type of hotel?

# 8) Which month is the most occupied and which is the least occupied?

# 9) Which type of hotel is mostly prefered by the guests?

# 10) What is the percentage of booking changes made by the customer?

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

# 12) What is the percentage distribution of required\_car\_parking\_spaces?

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

*We tried to answer following questions*

# 1) Which hotel type has the highest ADR

# 2) which hotel has longer waiting time?

# 3) Which hotel type has the more lead time?

# 4) What is the Optimal stay length in both types of hotels ?

# 5) Do the guests alloted the same room type which was reserved by them?

# 6) Which Market Segment has the higest cancellation rate?

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

# 8) ADR across the different months.

# 9) Which Hotels have the most repeat guests?

# 10) Which hotel has highest the percentage of booking cancellation?

***Conclusion***

1. Agent no. 9 has made most no. of bookings.

### 2.Only 3.9% of people were revisited the hotels. Rest 96.1 % were .new guests. Thus retention rate is low.

### 3.27.5 % bookings were cancelled out of all the bookings.

### 4.For both city and resort hotels, Nov to Jan have cheaper average daily rates.

### 5.More visitors are from western europe, namely Portugal,France, Great Britain, and Spain being the highest.

### 6.Most customers are not repeating their booking.

### 7.Families with children have no particular preference for the hotel type.

### 8.August is the most occupied month with 12.88% bookings and January is the least occupied month with 5.37% bookings.

### 9.City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.

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

### 11.uly and August months had the most Bookings. Summer vaccation can be the reason for the bookings.

### 12.Most of the customers (91.6%) do not require car parking spaces.

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

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

### 15.Average lead time for resort hotels is high.

### 16.Optimal stay in both the type hotel is less than 7 days

### 17.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

### 18.In "direct" both the hotels have almost the same cancellation rate.

### 19.Resort hotels have the most repeated guests.

### 20.Booking cancellation rate is high for City hotels which is almost 30 %.

***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.*

### 3.27.5 % bookings were cancelled out of all the bookings.

### 4.For both city and resort hotels, Nov to Jan have cheaper average daily rates.

### 5.More visitors are from western europe, namely Portugal,France, Great Britain, and Spain being the highest.

### 6.Most customers are not repeating their booking.

### 7.Families with children have no particular preference for the hotel type.

### 8.August is the most occupied month with 12.88% bookings and January is the least occupied month with 5.37% bookings.

### 9.City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.

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

### 11.uly and August months had the most Bookings. Summer vaccation can be the reason for the bookings.

### 12.Most of the customers (91.6%) do not require car parking spaces.

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

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

### 15.Average lead time for resort hotels is high.

### 16.Optimal stay in both the type hotel is less than 7 days

### 17.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

### 18.In "direct" both the hotels have almost the same cancellation rate.

### 19.Resort hotels have the most repeated guests.

### 20.Booking cancellation rate is high for City hotels which is almost 30 %.