**EDA on Hotel Booking Analysis**

**Ansh Srivastava**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

To understand the Hotel Booking firstly,

We have to know some parameters like the main few things I will usually consider include prices per night, distance of hotel from attractions and restaurants, availability of free breakfasts, scenery in hotel room, cleanliness of hotel room and of course, availability of free Wi-Fi. In this dataset, we are able to know different types of bookings (i.e. type of hotel, duration of stay, types of visitors, types of booking, etc).

**Keywords: EDA Hotel booking Analysis**

**Problem Statement**

For this project we will be analyzing Hotel Booking data. 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.

Hotel industry is a very volatile industry and the bookings depend on the above factors and many more.

The main objective behind this project is to explore and analyze data to discover important factors that govern the bookings and give insights to hotel management,which can perform various campaigns to boost the business and performance.

**Data description:**

The main objective of Exploratory data analysis is to understand trend and behaviour of guests in hotel bookings. For that first we will need to understand what every feature in data means. The data table consists of 119,390 rows and 32 columns. Our analysis starts with defining each column and our understanding for each column mentioned below:

* hotel: Hotel type (City hotels, Resort hotels)
* is\_canceled: value indicates if the booking is cancelled or not.
* lead\_time: How long in advance the booking was made.
* arrival\_date\_year: Customer arrival year.
* arrival\_date\_month: In which month of the year customer visited the hotel.
* arrival\_date\_week\_number: In which week of the year customer arrived.
* arrival\_date\_day\_of\_month: Date of the month customer visited hotel.
* stays\_in\_weekend\_nights: Customer stayed or booked to stay in hotel during weekend nights.
* stays\_in\_week\_nights: Customer stayed in hotel during week nights.
* adults: Number of adults.
* children: number of children.
* babies: Number of babies.
* meal: Type of meal booked.:
* country: Country of origin of customer.
* market\_segment:
* where the bookings came from
* distribution channel:
* Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
* is\_repeated\_guest: Value indicating if the booking name was from a repeated guest (1) or not (0).
* previous cancellations:

Number of previous bookings that were cancelled by the customer prior to the current booking.

* previous\_bookings\_not\_canceled:

umber of previous bookings that were cancelled by the customer prior to the current booking.

* reserved\_room\_type:
* Code of room type reserved. Code is presented instead of designation for anonymity reasons
* . • assigned\_room\_type:
* Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due
* booking\_changes:
* Number of changes/amendments made to the booking from the moment the booking was entered on the PMS.
* deposit\_type:

Indication on if the customer made a deposit to guarantee the booking.

* agent:
* ID of the travel agency that made the booking.
* company:
* ID of the company/entity that made the booking or responsible for paying the booking.
* days\_in\_waiting\_list:

Number of days the booking was in the waiting list before it was confirmed to the customer.

* customer\_type:

Type of booking, assuming one of four categories.

* adr:

Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.

* required\_car\_parking\_spaces: Number of car parking spaces required by the customer.
* total\_of\_special\_requests: Number of special requests made by the customer (e.g. twin bed or high floor).
* reservation\_status: Reservation last status, assuming one of three categories: Canceled –booking was cancelled by the customer;
* CheckOut:
* customer check out from hotel,
* No show:
* Customer did not check-in the hotel and informed the hotel with reason.
* reservation\_status\_date:
* Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when the booking was cancelled or when the customer checked out of the hotel.

**Data Cleaning:**

Cleaning data cleaning data is crucial step before EDA as it will remove the ambiguous data that can affect the outcome of EDA.

While cleaning data we will perform following steps:

1) Remove duplicate rows

2) Handling missing values.

3) Convert columns to appropriate datatypes.

4) Adding important columns

# Data Exploration

. **Data Wrangling:**

After loading the dataset, we performed this method by cleaning, organizing, and transforming raw data into the desired format which makes us to understand the data clearly. This process helped us to tackle the unwanted data, to produce accurate results, to make better decision.

**Null Value Treatment:**

Our data set contains a small number of null values; still we have treated the null values by filling with zeros in order to produce more accurate results.

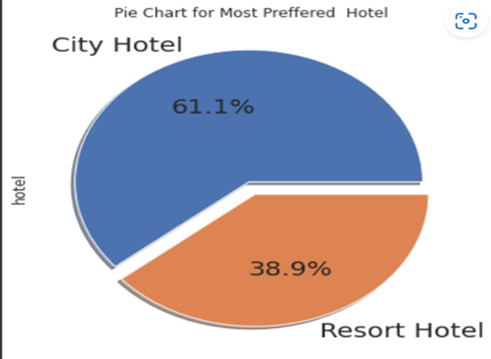
**EDA (Exploratory Data Analysis):**

After loading the dataset, we performed this method by comparing our target variable that is booking analysis with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

Mainly performed using Matplotlib and Seaborn library and the following graph and plots had been used:

* Bar Plot.
* Histogram.
* Scatter Plot.
* Pie Chart.
* Line Plot.
* Heatmap.
* Box Plot
* Type of hotel people booking more
* Find out which month people book the hotels?
* find the year in which most booked hotel?.
* Calculate the percentage of the nan values in the data?.
* Taking Necessary Columns Only.
* What is the percentage of cancellation?.
* Find the top 10 countries from where most people booked the hotels?.
* Make the dataframe top 10 countries?.
* Which is the most preferred room type by the customers?.
* What is the percentage distribution of "Customer Type"?.
* What is the percentage distribution of required\_car\_parking\_spaces?.
* What is the Percentage distribution of Deposit type ?.
* Which type of food is mostly preferred by the guests?.
* Correlation of the columns
* Which Hotels have the most repeated guests?.
* Relationship between the repeated guests and previous bookings not canceled?.
* Which distribution channel has the highest cancellation rate?.
* Which Market Segment has the highest cancellation rate?.
* Which Hotel type has the highest ADR?
* Which distribution channel contributed more to adr in order to increase the income?.
* ADR across different market segment

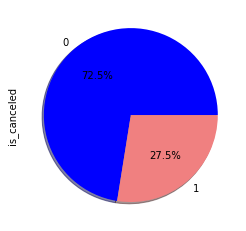
# type of hotel people are booking more :-

****

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

**Cancelled bookings**:-

According to the pie chart, 72.5% of bookings were not cancelled and 27.5% of the bookings were cancelled at the Hotel.



**Type of meals booked:-**

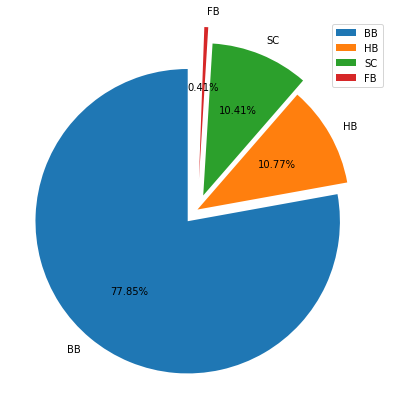
BB - (Bed and Breakfast)

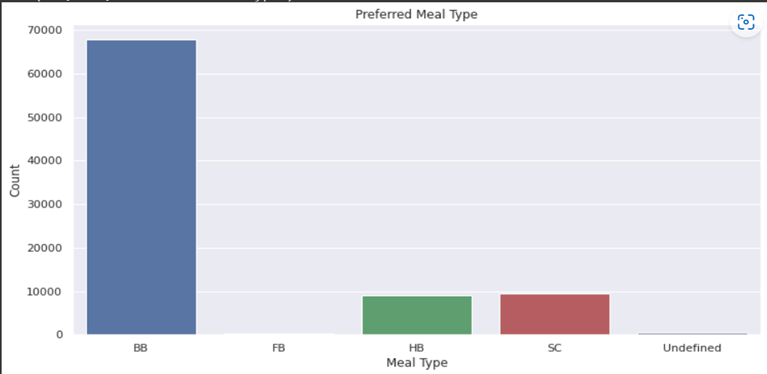
HB- (Half Board)

FB- (Full Board)

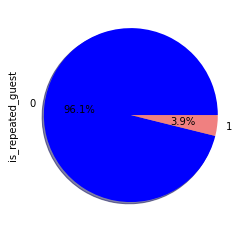
SC- (Self Catering)

So the most preferred meal type by the guests is BB( Bed and Breakfast)

HB- (Half Board) and SC- (Self Catering) are equally preferred.****

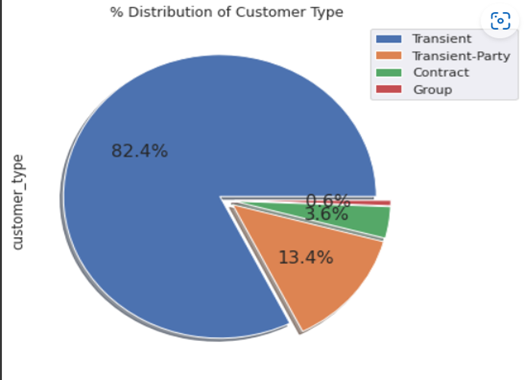
****

**Repeated guest:-**

****

Almost 4% of guest are coming repeatedly.

**distribution of customer type:**

****

Most of the customers/guests were Transient type(82.4%). Transient party were 13.4% and 0.6% belonged to the group. Remaining guests belongs to Contract type.

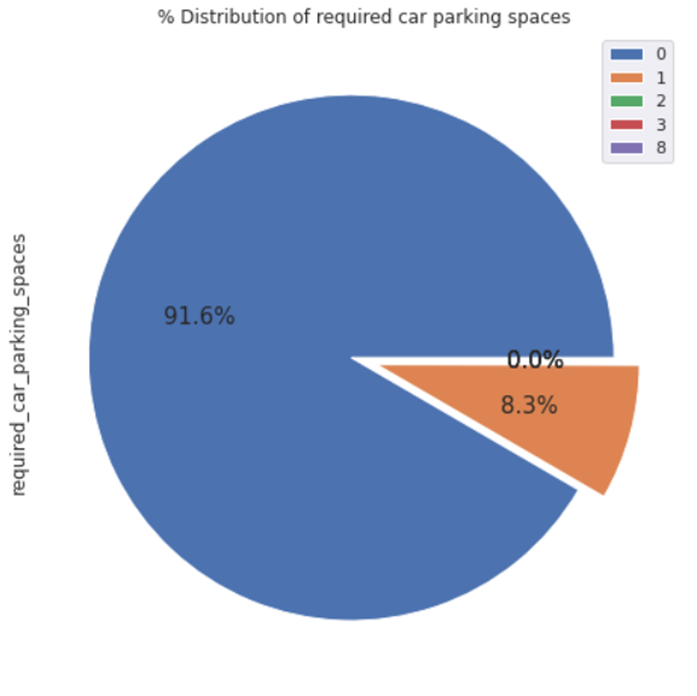
**Contract-**when the booking has an allotment or other type of contract associated to it

**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

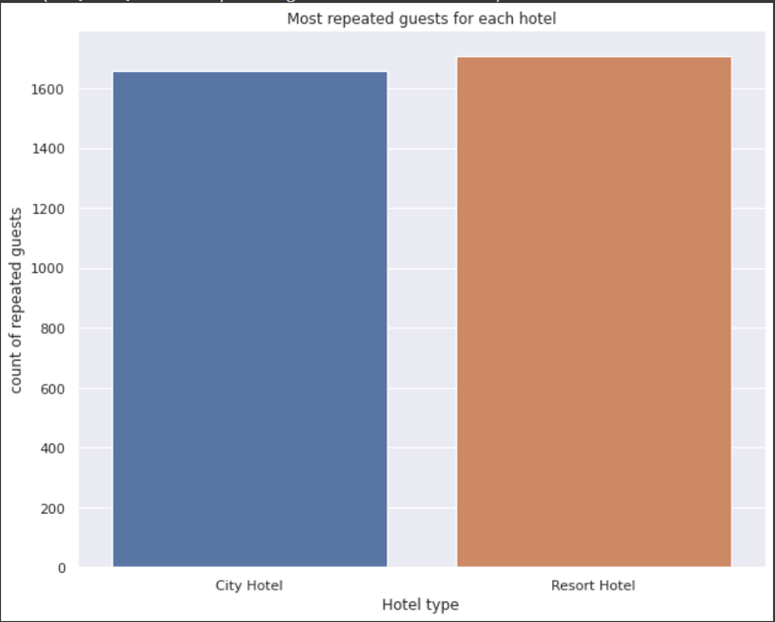
**Transient-party-**when the booking is transient, but is associated to at least other transient bookings.

**distribution of required car parking spaces:**

****

91.6 % of guests did not require the parking space. only 8.3 % of guests required only 1 parking space.

**most repeated guest for each hotel:**

****

Resort hotels have the most repeat guests. In order to get increase the count of repeated guests, hotel management needs to take the valuable feedbacks from the guests and try to give good service. Exploratory Data Analysis (EDA)

**people from top traveling countries:**

1.PRT- Portugal

2.GBR- United Kingdom

3.FRA- France

4.ESP- Spain

5.DEU - Germany

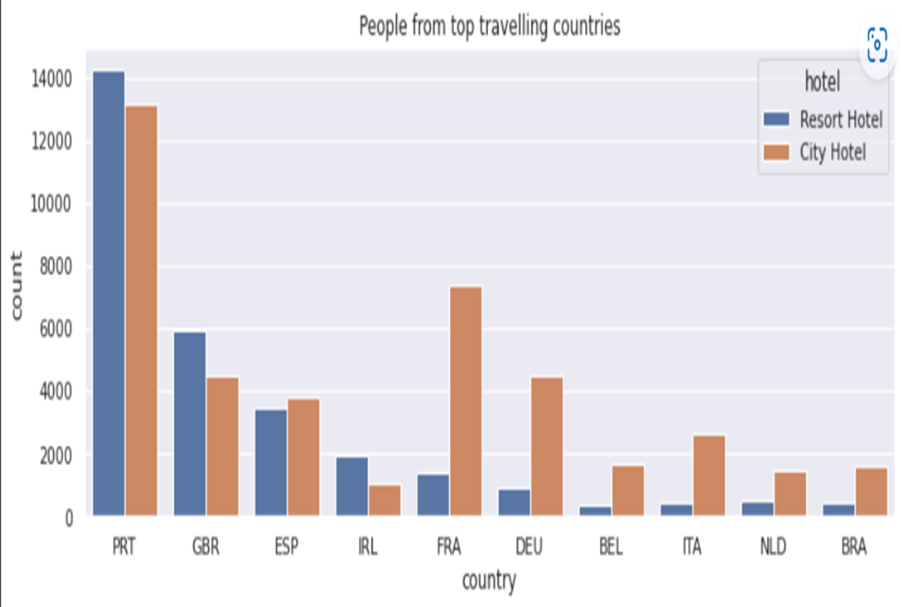
6.ITA -Italy

7.IRL - Ireland

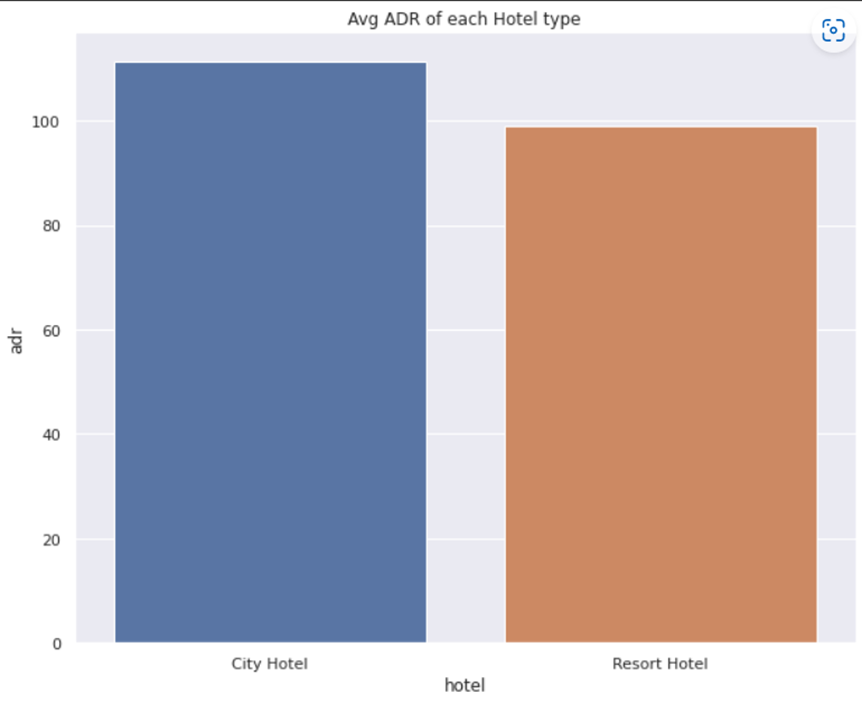
8.BEL -Belgium

9.BRA -Brazil

10.NLD-Netherlands

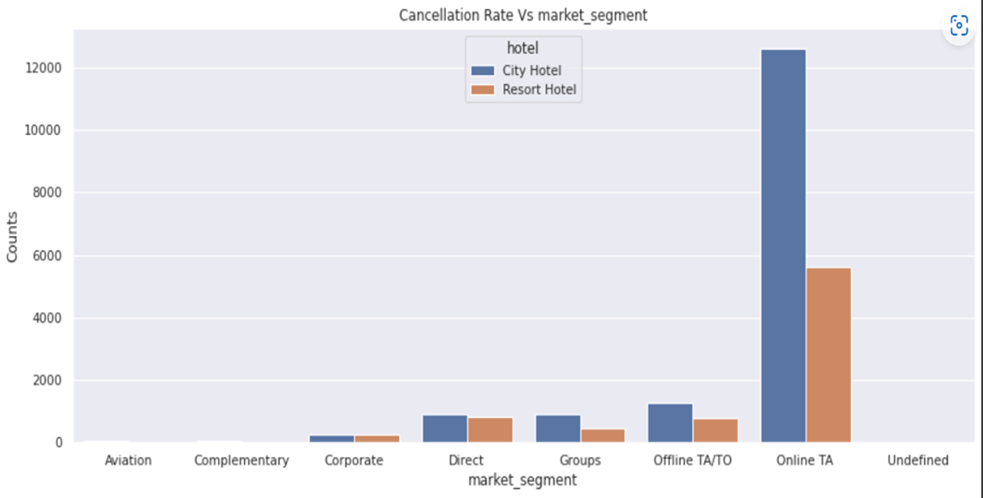
****

**Average ADR of each hotel type:**

****

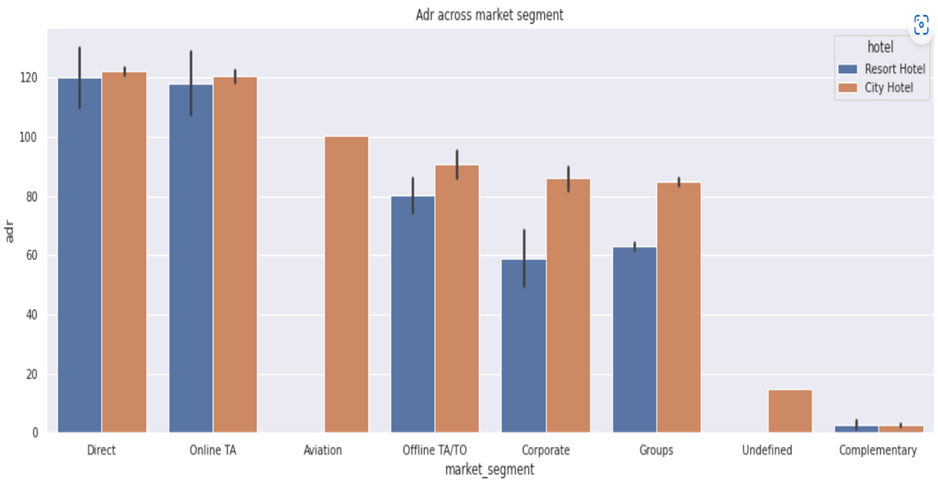
City hotel has the highest ADR. That means city hotels are generating more revenues than the resort hotels. More the ADR more is the more revenue.

**Market Segment:**



‘Online TA/TO’ market segment has highest cancellations for city hotels

**Adr Across market segment:**

****

Direct' and 'Online TA' are contributing the most in both types of hotels.

Aviation segment should focus on increasing the bookings of 'City Hotel'

**Conclusion:-**

Conclusions- Out of all months, 'August' witnessed the highest number of hotel bookings whereas 'January' witnessed the least. Among all the countries in the dataset, PRT(Portugal) has got the maximum number of hotel bookings. It’s observed that 'City hotels' were more canceled as compared to 'Resort hotels'. Coming to the analysis of the market segment, 'Online TA' brings maximum bookings. Considering all the three years, 'August' has got the highest average ADR in each year

**References-**

1. Pandas user guide: <https://pandas.pydata.org/docs/user_guide/index.html>
2. Matplotlib user guide: <https://matplotlib.org/3.3.1/users/index.html>
3. Seaborn user guide & tutorial: <https://seaborn.pydata.org/tutorial.html>
4. AnalyticVidya :- <https://www.analyticsvidhya.com/blog/2022/04/exploratory-data-analysis-eda-in-python/>