**Hotel Booking Analysis**

Megha Dethe

Praful Gedam

**Data science trainees,**

**Alma Better, Bangalore**

**Abstract:**

For this project we will be applying hotel

Booking data. This data contains booking information for a city hotel and a resort hotel and including information such a when the booking was, length of stay, the number of adults, children and or babies, and the number of available parking space. Hotel industry is very volatile industry and the booking depends on the above factor and many more.

The main objective of behind this project

Is to explore and analyzing data to discover important factor that govern the booking and give insights to hotel management, which can perform various campaign to boost the business and performance

1. **Problem Statement**

This data contains of analysis of fallowing of question

1. Which agents makes most no. of booking
2. How many hotels cancelation rate
3. Which percentage of booking changes?
4. from which country visited the most
5. which year had the highest booking
6. in which month most of the booking happen
7. Most preferred room type of the customer
8. Count of repeated guest
9. Relationship between the repeated guest and previous booking not canceled?
10. Which hotel type has the highest ADR?
11. ADR across different market segment

**2. Introduction**

Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business

Feature Engineering, fifth part Model Building, sixth part Model Evaluation.

**3. Data Description**

The dataset contains workflow information (univariate analysis, bivariate analysis, and multivariate analysis),

* **Hotel:** Resort Hotel or City Hotel
* **Is\_canceled**: Value indicating if the booking was cancelled (1) or not (0)
* **Lead-time:** Number of days that elapsed between the entering date of the booking and the arrival date.
* **Arrival date year**: Year of arrival date.
* **Arrival\_date\_month:** Month of arrival date.
* **Arrival\_date\_week\_number:** Week number of year for arrival date.
* **Arrival\_date\_day\_of\_month:** Day of arrival date.
* **Stays\_in\_weekend\_nights:** Number of weekend nights.
* **Stays\_in\_week\_nights:** Number of weeknights.
* **Adults:** Number of adults.
* **Children:** Number of children.
* **Babies:** Number of babies.
* **Meal:** Type of meal booked.
* **Country:** Country of origin

## **4. Steps:**

**Data collection and understanding:**

Collecting data, it is very important to understand your data.so you had hotel booking analysis data which has 119390 rows and 32 column the values are not actually true so we can mislead by this. Needs to be converted to numerical format.

1. **Data description:**

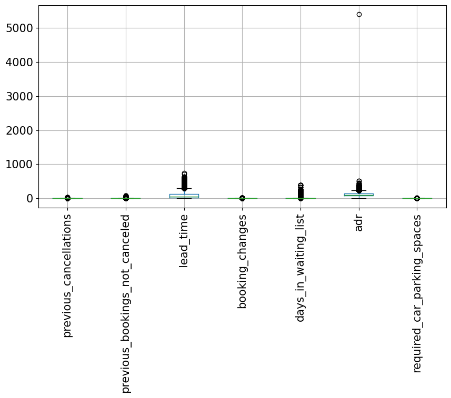
The objective of data description is expletory data analysis is to understand to trend and behavior of guest in hotel booking. For that first we will need to understand what every feature in data means. Our analysis starts with defining each column and our understanding

For each column mention below

* **Hotel: Resort** Hotel or City Hotel
* **Is\_canceled:** Value indicating if the booking was cancelled (1) or not (0)
* **Lead\_time:** Number of days that elapsed between the entering date of the booking and the arrival date.
* **Arrival\_date\_year**: Year of arrival date.
* **Arrival\_date\_month:** Month of arrival date.
* **Arrival\_date\_week\_number:** Week number of year for arrival date.
* **Arrival\_date\_day\_of\_month:** Day of arrival date.
* **Stays\_in\_weekend\_nights:** Number of weekend nights.
* **Stays\_in\_week\_nights:** Number of weeknights.
* **Adults:** Number of adults.
* **Children:** Number of children.
* **Babies:** Number of babies.
* **Meal:** Type of meal booked.
* **Country:** Country of origin
* **Market\_segment:** Market segment designation (TA/TO)
* **Distribution\_channel:** Booking distribution channel. (T/A/TO)
* **Is\_repeated\_guest:** is 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:**
* Number of previous bookings not cancelled by the customer prior to the current booking
* **Reserved\_room\_type:** Code of room type reserved.
* **Assigned\_room\_type:** Code for the type of room assigned to the booking.
* **Booking\_changes:** Number of changes made to the booking from the moment the booking was entered on the ¬
* PMS until the moment of check-in or cancellation
* **Deposit\_type:** No Deposit, Non-Refund, Refundable.
* **Agent:** ID of the travel agency that made the booking
* **Company:** ID of the company/entity that made the booking.
* days\_in\_waiting\_list: Number of days the booking was on the waiting list before it was confirmed to the customer
* **Customer\_type:** type of customer. Contract, Group, transient, Transient party.
* **Adr:** Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.
* **Reuired\_car\_parking\_spaces**: Number of car parking spaces required by the customer.
* **Total\_of\_special\_request:** Number of special requests made by the customer (e.g., twin bed or high floor)
* **Reservation\_status:** Reservation last status.

Data cleaning and Manipulation:

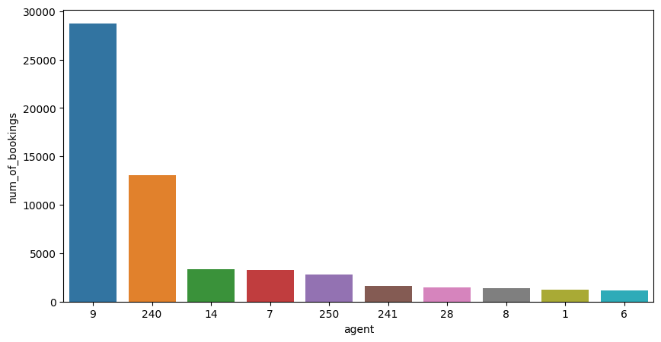
Data cleaning is most important before EDA and they are following steps od data cleaning.

* Removing duplicates rows
* Handling null values
* We there 4 column company, agent, country and children with null values changing the value
* One outlier was found in the ADR column then dropped it
* 

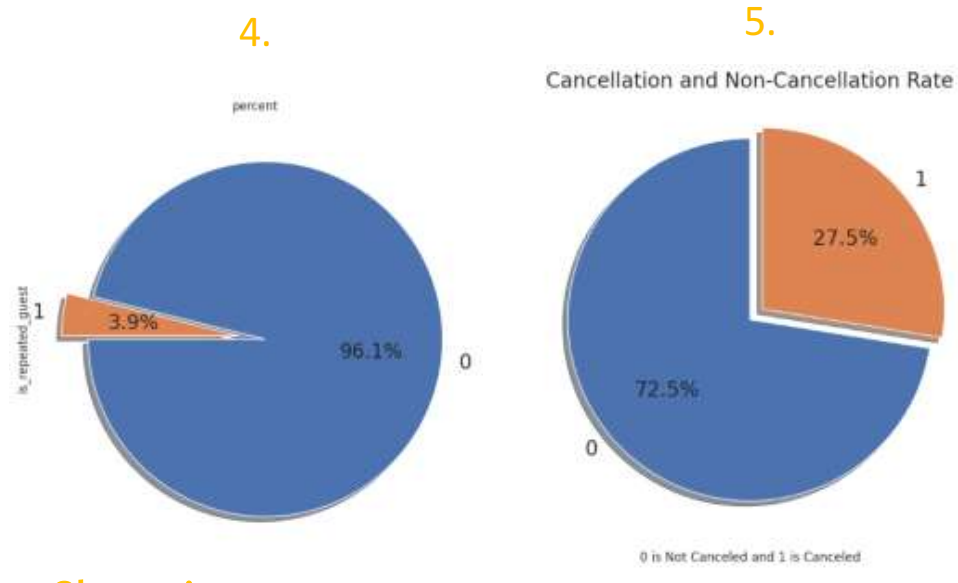
**Exploratory Data Analysis**

In this step we will analyze dataset by the use of Data visualization by using different graphs. Get idea about features available in dataset.

* **Agents makes no of booking:**

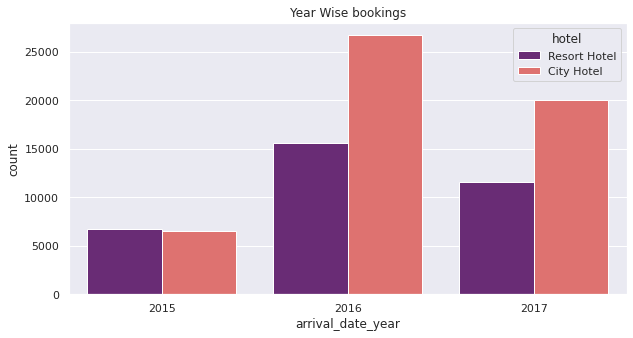
****

Above plot show that agents make no of booking in the dataset. Which the largest agents 9 has the booking done as per dataset.



From the above point plot and bar plot we can say that 27.5 people canceled their booking. We have to find out their reason why this people canceled their booking also feedback cab be taken there from guest canceled there booking

.

* **Highest year wise booking:**

In the above plot which shows the when the booking is year wise counted. When the booking is done 2015 is less than 7000 booking and the 2016 is the highest booking is done.in the analysis when the overall city hotel is the most of the booking is done in the above graph

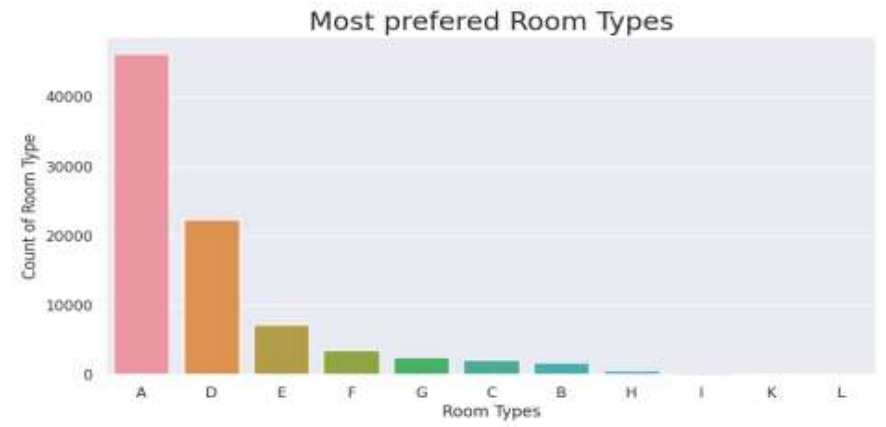
**Highest month wise booking**:



plot shows that in when the no of booking is done by is month wise

and July and august are most of the booking done. The summer vacation can be the reason of the booking.

* **Most preferred room type customer**

****

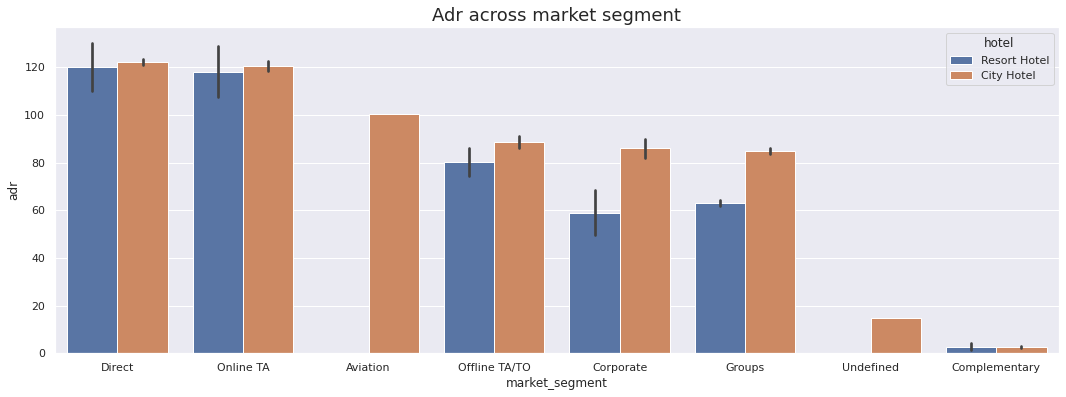
in hotel booking when the customer are chooses are various type of rooms. But above bar plot show that type A are most preferred room in customer to choose.

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

****

In the analyzing of hotel booking ADR is the most important criteria. in the above plot two type of hotel city hotel and resort hotel. The city hotel is the highest ADR. This means city hotel are generating more revenue than the resort hotel. More the ADR more the revenue

**ADR across different market segment**

****

in the above plot show that ADR different market. Direct and online TA are contributing the most in both type of hotel.

Aviation segment should be focus on increasing the booking of city hotel

**Conclusion**

* It helps a hotel in up scaling their business by their technique, technologies, and many small requirements which guest require during their stay in this hotel
* It also helps people to understand how things work in real life and the need and importance of data science in their real world.
* It increases the competition and productivity among various hotel to provide a better facility and comfort that the customer needed

**References-**

1. <http://medium.com>
2. <http://jovian.com>
3. <http://matplotlib.org>