**Capstone Project**

**EDA on Hotel Booking Analysis**

**Rohit Madame**

**Abstract:**

This hotel booking data set is for analysing when the customer has booked the most hotels. City hotels or resort hotels of which the most are booked by customers. How long do customers prefer to stay in hotels.

Our EDA can help to understand which season is best for stays. How long

Should they stay.

**1. Problem Statement**

This dataset contains data from hotels from 2015 to 2017 and includes information about the city hotels and resort hotels. When has highest bookings, meal type, which agent has made most bookings, information about car parking.

is\_canceled : Value indicating if the booking was canceled (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 week nights.

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 in 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
* 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.

**2. Introduction:**

Do you know which is the best season for staying in hotels? There is always confusion in choosing city hotels or resort hotels. Do they have a parking space? Which is the preferred meal type. This EDA will give all the answers.

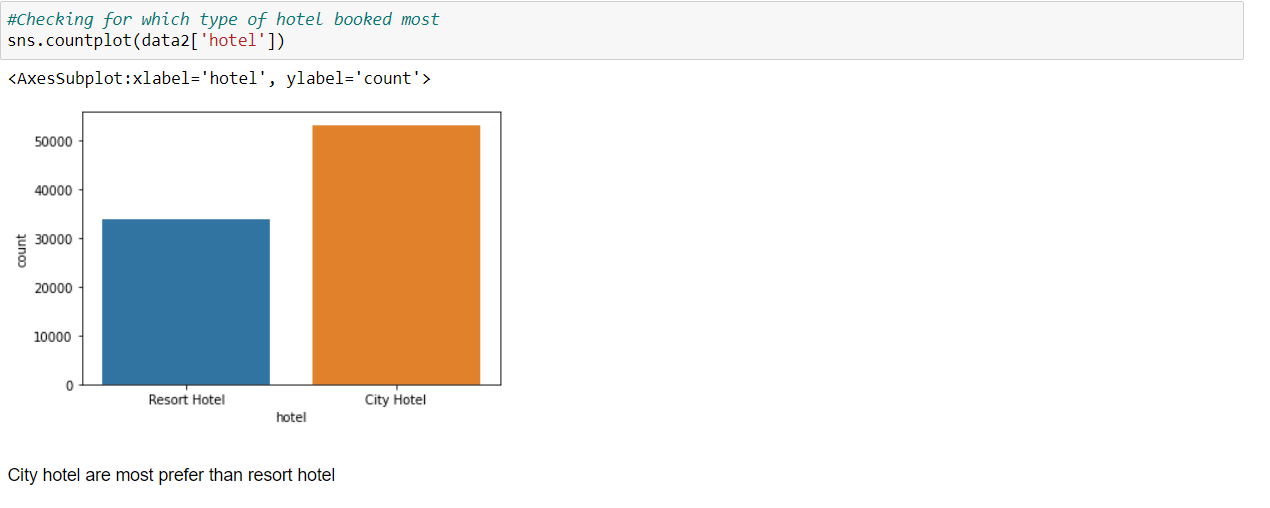
**3. Step Involved**

* Importing dataset and loading data of hotel booking.
* Data cleaning
* Data preparation
* Data Visualization
* Conclusion

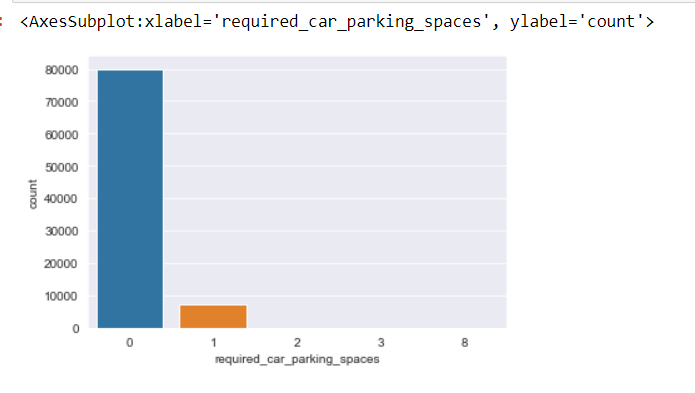
**Importing dataset:**Importing all the necessary libraries and hotel booking analysis datasets.

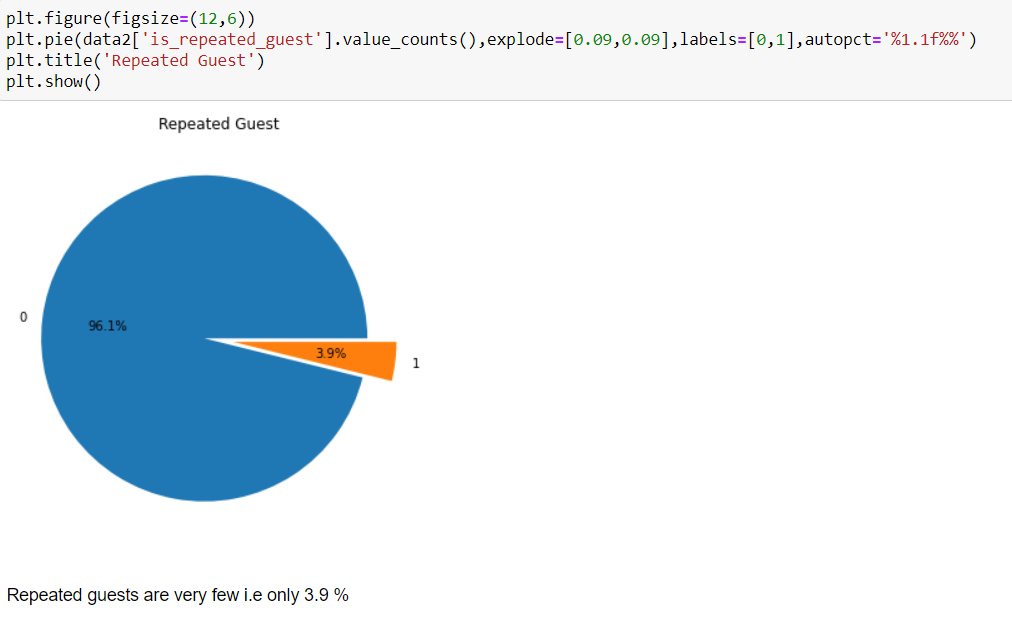
**Data Cleaning:** I have searched for null values and duplicate values. Removing it all gives us clean data.

**Data Visualization:** In this, I have used seaborn and matplotlib library to visualize one variable first i.e univariate then plotting two or more variable graph.

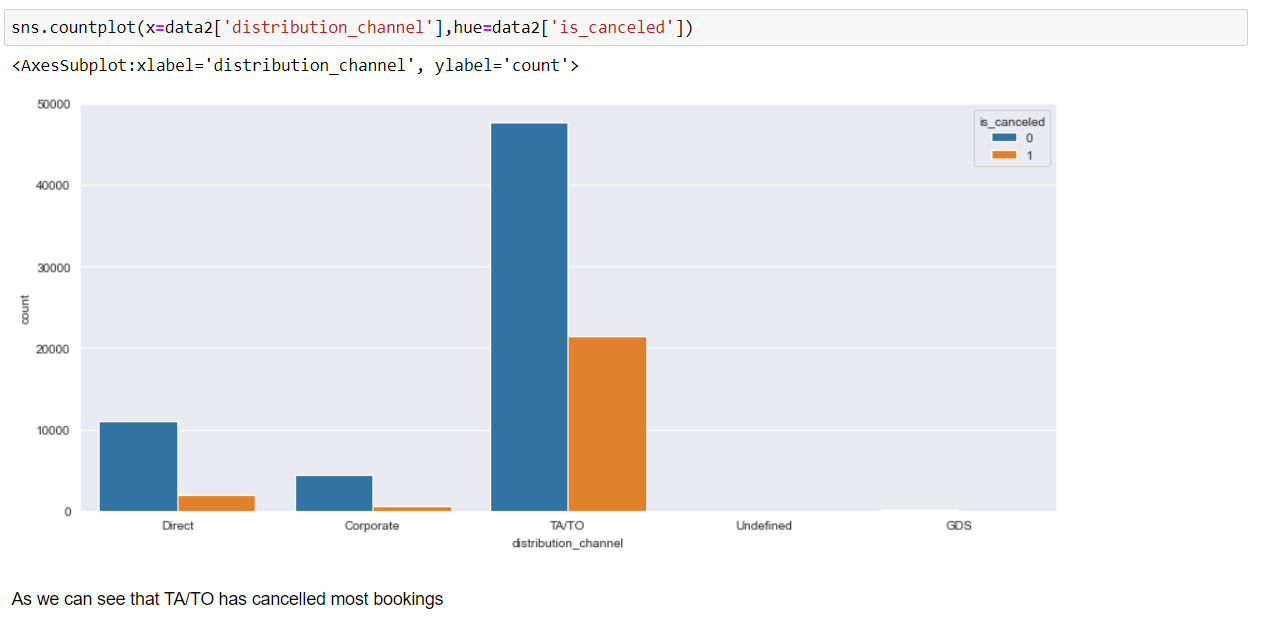
* City hotel is more liked by customers over resort hotels.
* 27.5% customers cancelled their bookings.



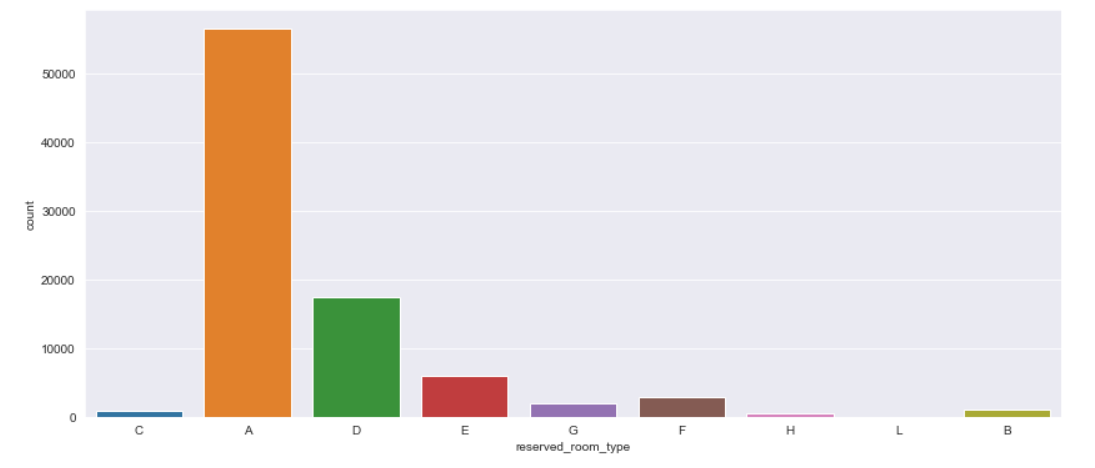
* Most customer don’t need car parking. Few customers need only one parking space.
* Only 3.9 are repeated customers.



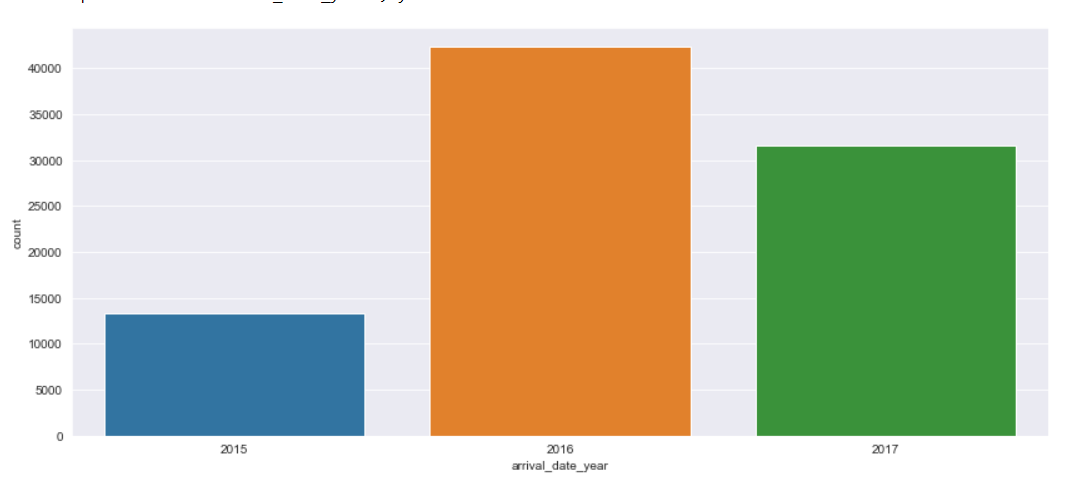
* Breakfast in Bed is most preferred meal .Where HB and SC are at same level.
* TA/TO has highest cancelation ratio.

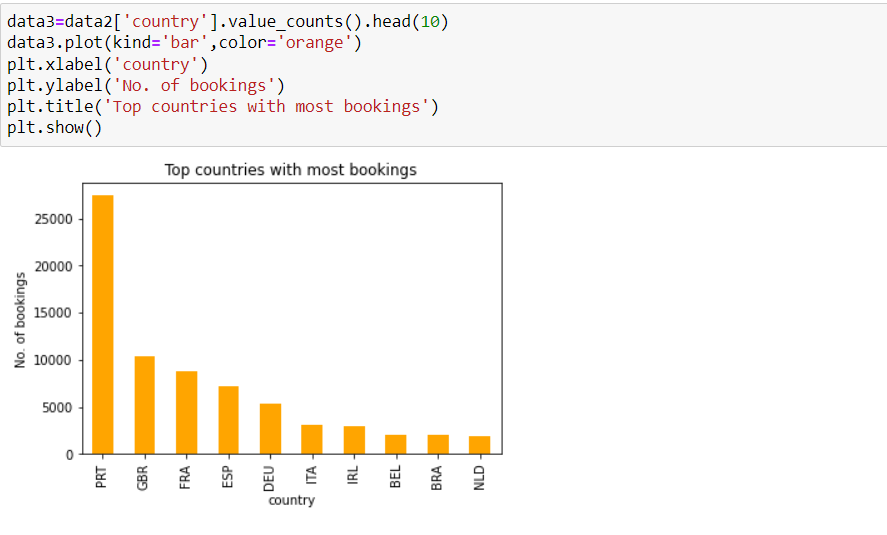
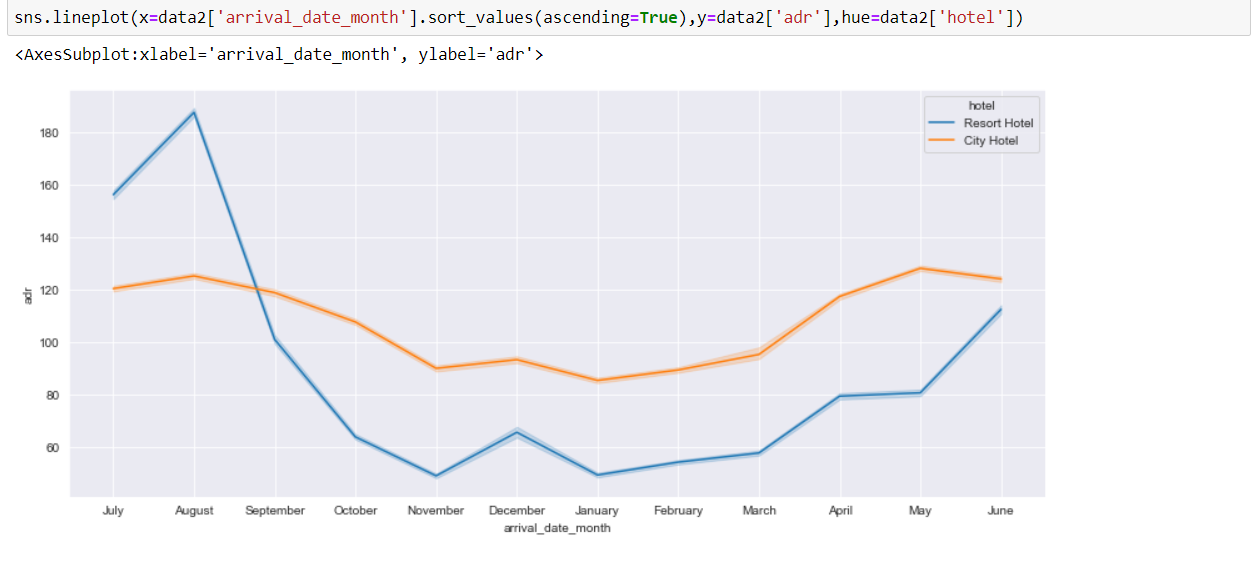


* Most reserved room type is A.



* 2016 has most bookings.



* Portugal has most bookings. After that Great Britain and France have most booking.
* August month has highest adr compare to all.

**4. Conclusion :**

* Booking for city hotels is more than resort hotels.
* New guests are more as compared to a repeated guest.
* ADR is highest for august month
* Breakfast in Bed is most preferred meal
* Most people don’t need car parking. Few want parking for one car only.

**5. References:**

* Stackoverflow
* Almabetter