**Capstone Project**

**Hotel Booking Analysis**

**Krishna Kumar Verma,**

**.**

**Contents:**

* **Abstract**
* **Introduction**
* **Problem Statement**
* **Steps Involved**
* **Data Visualization**
* **Conclusion**
* **References**

**Abstract:**

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, among other things. All personally identifying information has been removed from the data.

The dataset describes two types of hotels, City and Resort hotel which

has a total 1,19,390 rows and 32 columns.

Data fields:

* 'hotel': Resort or City type hotel
* 'is\_canceled' : Cancelled or not.
* 'lead\_time': Time difference between booking date and date of arrival
* 'Arrival\_date\_year': Year of arrival
* 'arrival\_date\_month':Month of arrival
* 'arrival\_date\_week\_number',:Week of arrival
* ‘arrival\_date\_day\_of\_month': Day of arrival
* 'stays\_in\_weekend\_nights',:Total Stay on weekend
* 'stays\_in\_week\_nights',Total Stay on weekday
* 'adults',:No. Of adults in the room
* 'children',:No. Of children in the room
* 'babies',:No. Of babies in the room
* 'meal',: Type of meal
* 'country',Country of booking
* 'market\_segment',Segment of booking
* 'distribution\_channel',:The type of distribution
* 'is\_repeated\_guest',: Repeated guest or not.
* 'previous\_cancellations',: Customer previously cancelled
* 'previous\_bookings\_not\_canceled':Customer previous did not cancel
* 'reserved\_room\_type',:type of room type
* 'assigned\_room\_type',:type of room assigned
* 'booking\_changes':Any changes in booking
* 'deposit\_type':Type of deposit for booking
* 'agent',: Agent used for booking
* 'Company':Company of booking
* 'days\_in\_waiting\_list',: Waiting list days
* 'customer\_type',: Type of customer based on stay duration.
* 'required\_car\_parking\_spaces',Was parking required
* 'total\_of\_special\_requests',:No. of special guests
* 'reservation\_status',:Status of reservation
* 'reservation\_status\_date':Date of status of reservation.

Using various python features and with the help of pandas, numpy,statsmodels, seaborn and matplotlib libraries, we perform analysis.

**Introduction:**

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

**Problem Statement:**

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, among other things. All personally identifying information has been removed from the data.

Explore and analyze the data to discover important factors that govern the bookings.

We will tackle the problem statement in the following steps:

Step 1: Data Overview.

Step 2:Data Preparation and Cleaning.

Step 3: Visualising the Univariate and Bivariate features.

Step 4: Correlation Analysis.

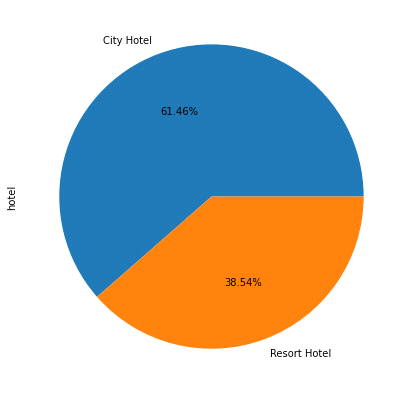
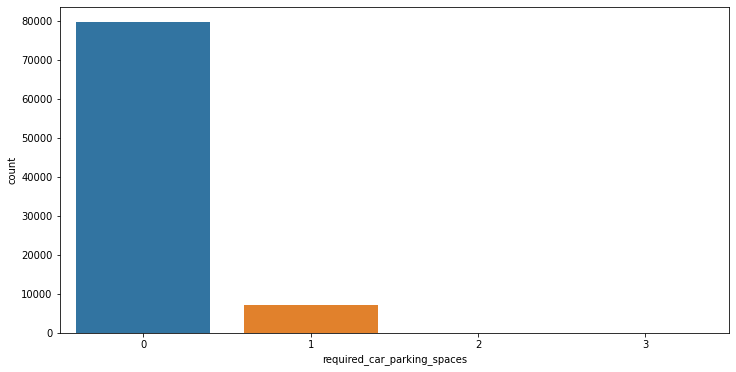
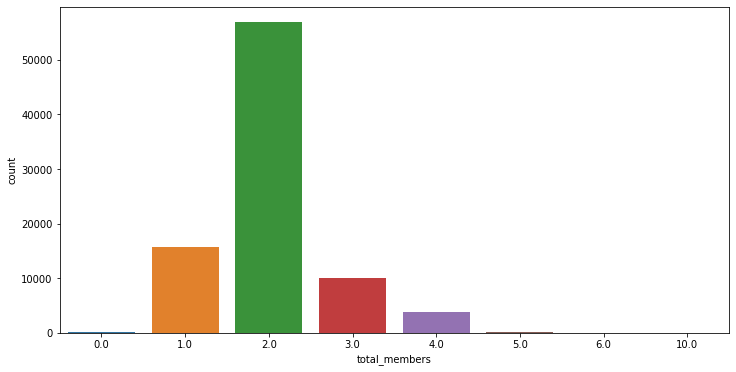
Step 5: Concluding Analysis.

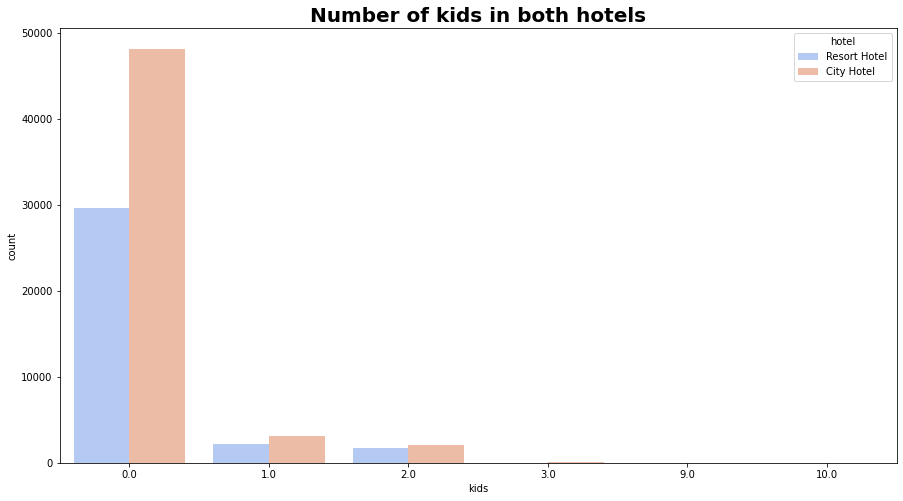
**Steps Involved:**

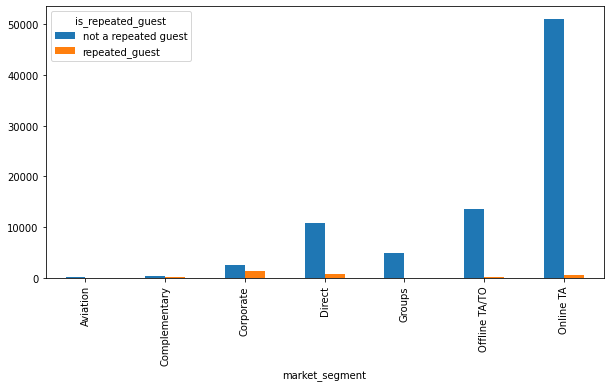
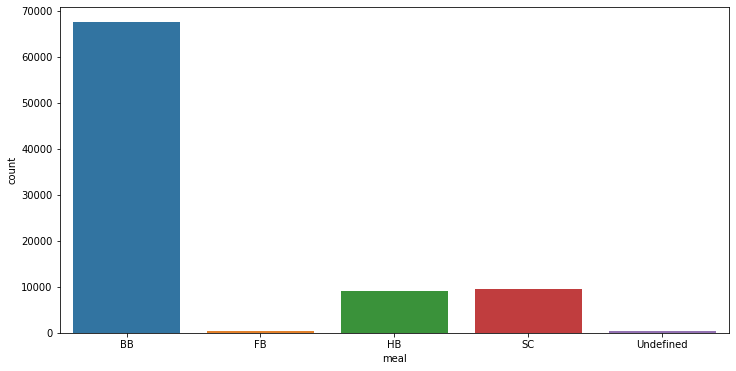
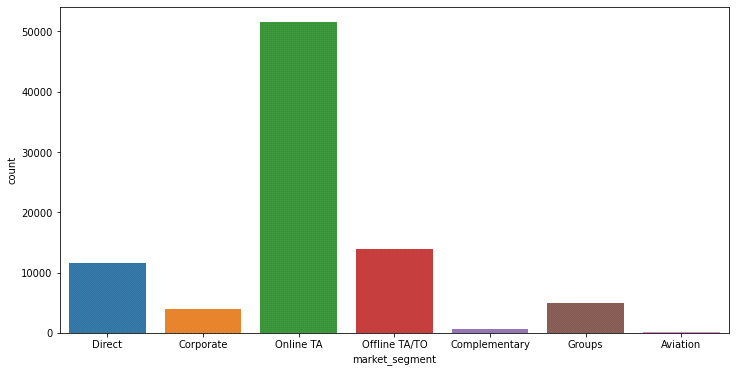
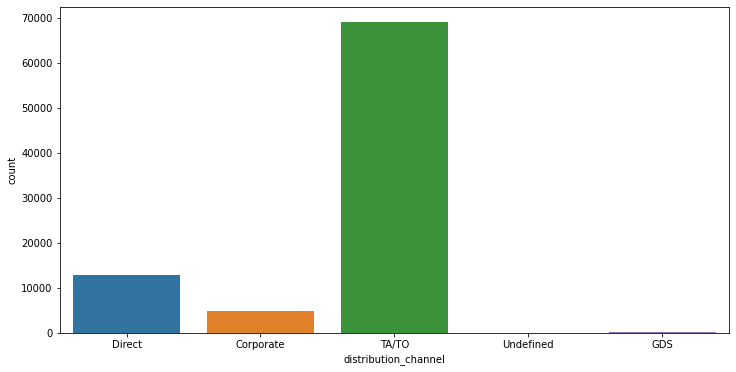
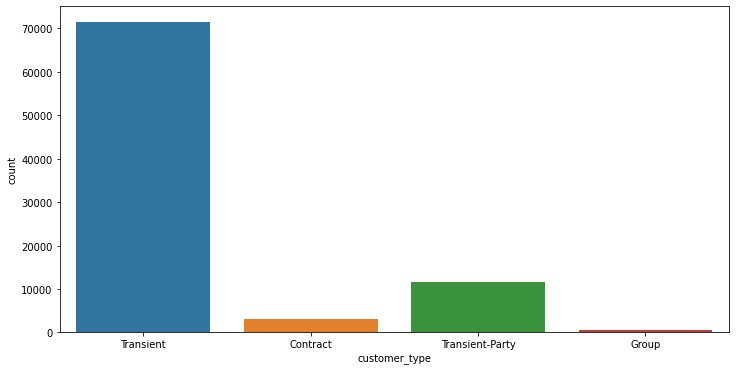
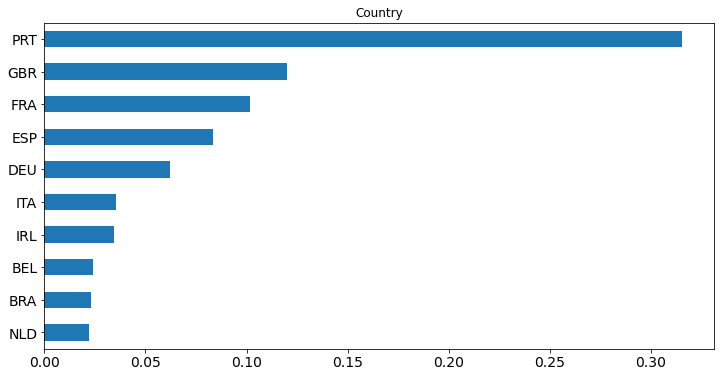
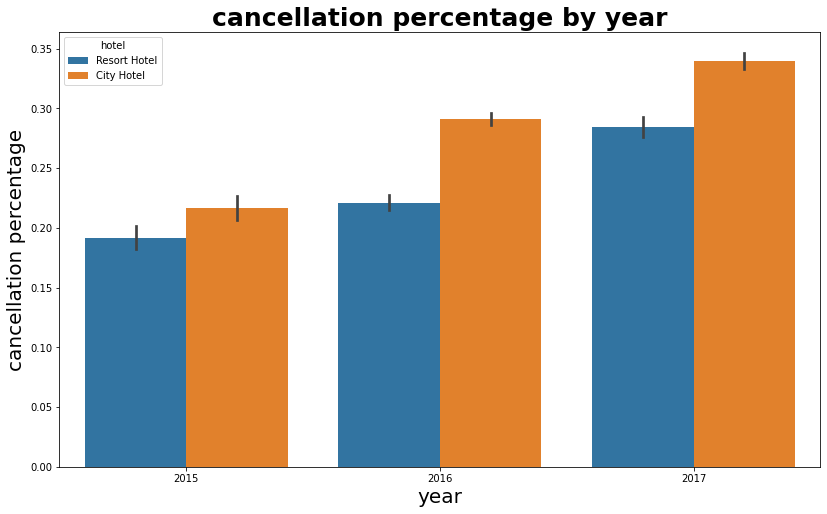
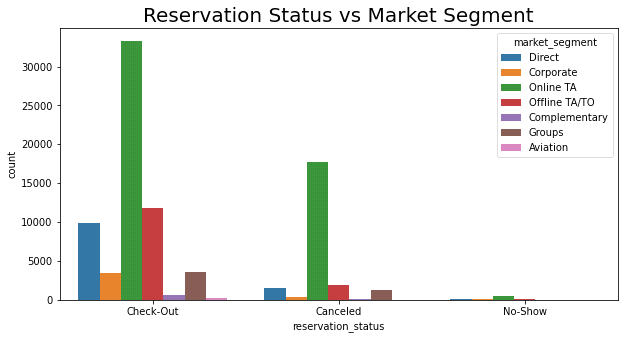
* **Importing Packages:** Importing the various libraries that will help us analyze our dataset properly with visual graphs.
* **Data Overview :** We load the data and go over the basic features, shape and datatypes of various variables.
* **Data Preparation and Cleaning:** We use various features of python to create combined features of date and time and other variables which can be simplified. We also drop columns and rows which have a lot of null values. Finally we take care of outliers.
* **Univariate and Bivariate Analysis:** We use seaborn and matplotlib to plot graphs starting with one variable graphs and then plotting two or more variable graphs to understand the variables and it spread and range with its frequency
* **Correlation Analysis:** We plot a correlation graph which basically gives us the correlation of every variable with one and other. This tells us which variable has what effect on the booking process.
* **Concluding Analysis:** Here after plotting and analyzing all the data we finally make predictions and remarks about our analyzes.

**Data Visualization:**

Lets understand various relation among target and other variables

* There are only 2 hotels i.e. 'City Hotel' and 'Resort Hotel' and City Hotel has more bookings as compared to the Resort Hotel.
* Most bookings don’t require car spacing spaces.
* Majority of bookings have 2 guests
* Majority of bookings don’t have any kids accompanied.



* After complimentary, corporate has the most proportion of repeated guests.
* The most common meal is breakfast in bed.
* People tend to book for Online TA among various segments. 
* TA/TO has the highest booking among different channels.
* Most people have Transient travel booking, meaning they stay for a short time.
* Portugal has the highest booking followed by Great Britain and France.
* Average daily rate is highest in August followed by July for resorts but then sharply drops below City hotels for most of the year.
* 
* Most of the booking either in the canceled status or in check-out status are done by the online travel agencies (online TA).

**Conclusion:**

So we have reached the end of the analysis and some noteworthy things we found are:

* The majority of guests come from western europe countries.
* The majority of reservations are for city hotels.
* The number of repeated guests is too low.
* The majority of reservations convert into successful transactions.
* For resort hotels, the average daily rate is more expensive during August, July.
* From market\_segment and distribution\_channel we can observe that City Hotel is more into 'Online TA'.

**References:**

* GeeksforGeeks
* Stackoverflow
* Almabetter
* Youtube
* Github