Capstone Project
Hotel Booking Analysis

Ashok Ravindra Kondhalkar,

Data science trainee, AlmaBetter,Bangalore.

Contents:

Abstract:

This hotel booking dataset contains booking information for city and resort hotels. Both datasets share the same structure, with Database having 119390 Rows and 32 columns. 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.

Feature Description:

- **'hotel'**: Hotel(Resort Hotel or City Hotel)
- **'hotel1'**:Copy of hotel
- **'is_canceled'**: Booking was 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 origin
- **'market_segment':** Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"

- **'distribution_channel'**:The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- **'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
- **'adr'**:average daily rate
- **'required_car_parking_spaces'**:parking required
- **'total_of_special_requests'**:No. of special guests
- **'reservation_status'**:Status of reservation
- **'reservation_status_date**:Date of status of reservation
- **'total_stay'**:'stays_in_week_nights'+'stays_in_weekend_nights'
- **'total_peoples'**='babies'+'adults'+'children'

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

Introduction:

Most of the industries data are shared to grow business, identity choice of customer, how much they are willing to pay and so more. It help a lot to overcome if any limitation. Also, the dataset is used to build prediction models classify a hotel bookings and lot more.

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay? Which distribution channel has longer average waiting time? From which country most guest come? And more. 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.

Let, Explore and analyse the data to discover important factors that govern the bookings.

We will tackle the problem statement in the following steps:

Step 1: Data Reading and Inspection

Step 2: Data Cleaning.

Step 3: Correlation Analysis.

Step 4: Data Analysis and Visualization

Step 5: Concluding Analysis.

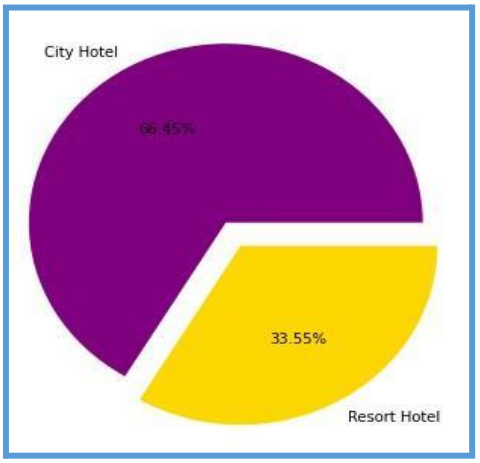
Steps Involved:

- **Importing Packages:** Importing the various libraries that will help us in analyse our dataset in visual format.
- **Data Reading and Inspection:** Here, We load the data and apply random analysis like .head (), .shape, .describe, .info, and more.etc.
- **Data Cleaning:** We are use Various features of pythons like finding null values in columns, unique values, create new columns by combining two columns, replace null values by zero, remove duplicate rows
- **Correlation Analysis:** We plot a correlation graph which basically gives information about which variable has what effect on the Hotel market.
- **Data Analysis and Visualization:** Here**,** We use seaborn and matplotlib to plot graphs. We are using pie chart, count plot, bar plot, boxplot and plot bar for better visualization.
- **Concluding Analysis:** Here after plotting and analysing all the data we finally make predictions and remarks about our analyses.

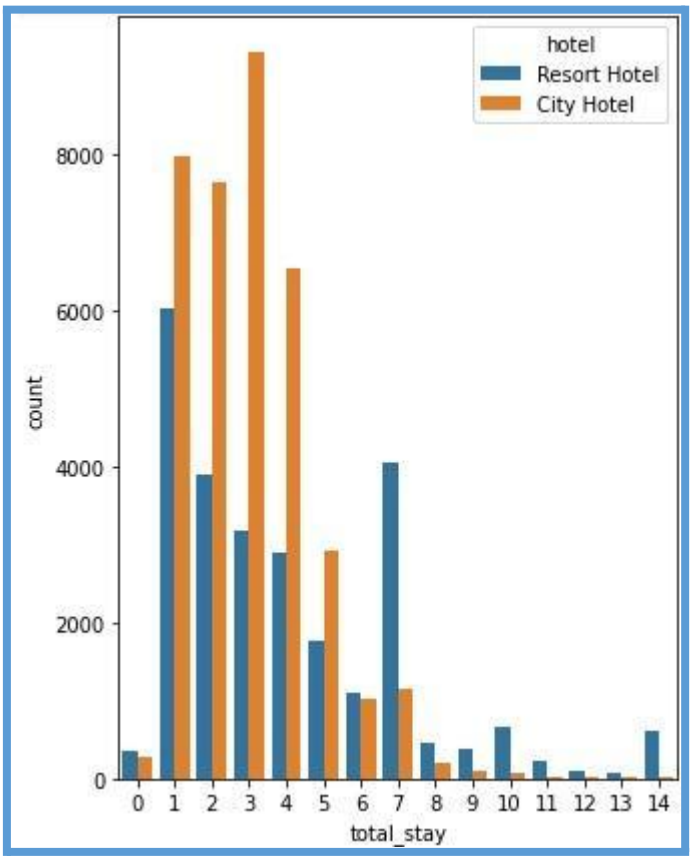
Data Visualization:

Let’s understand how each factor effect on one another in Hotel market.

1. There are only 2 hotels i.e. 'City Hotel' and 'Resort Hotel' and City Hotel has more bookings as compared to the Resort Hotel.

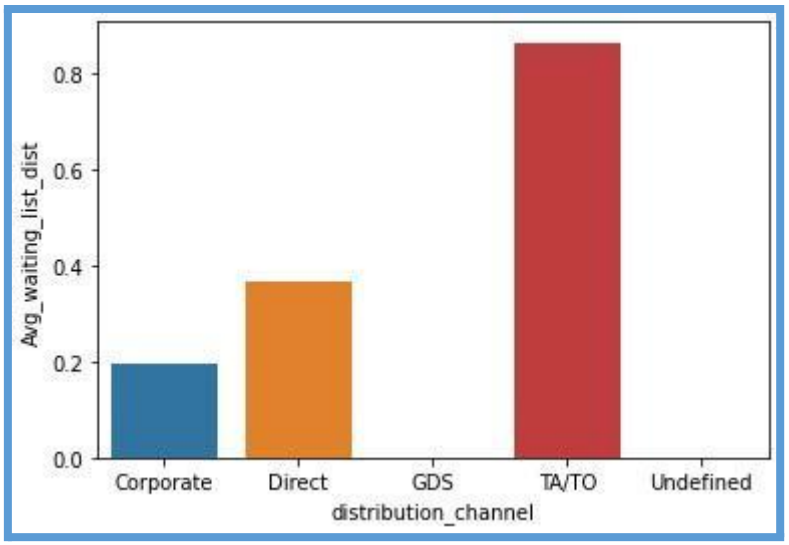


2. Lead time of city hotel is more than Resort Hotel Resort hotel.

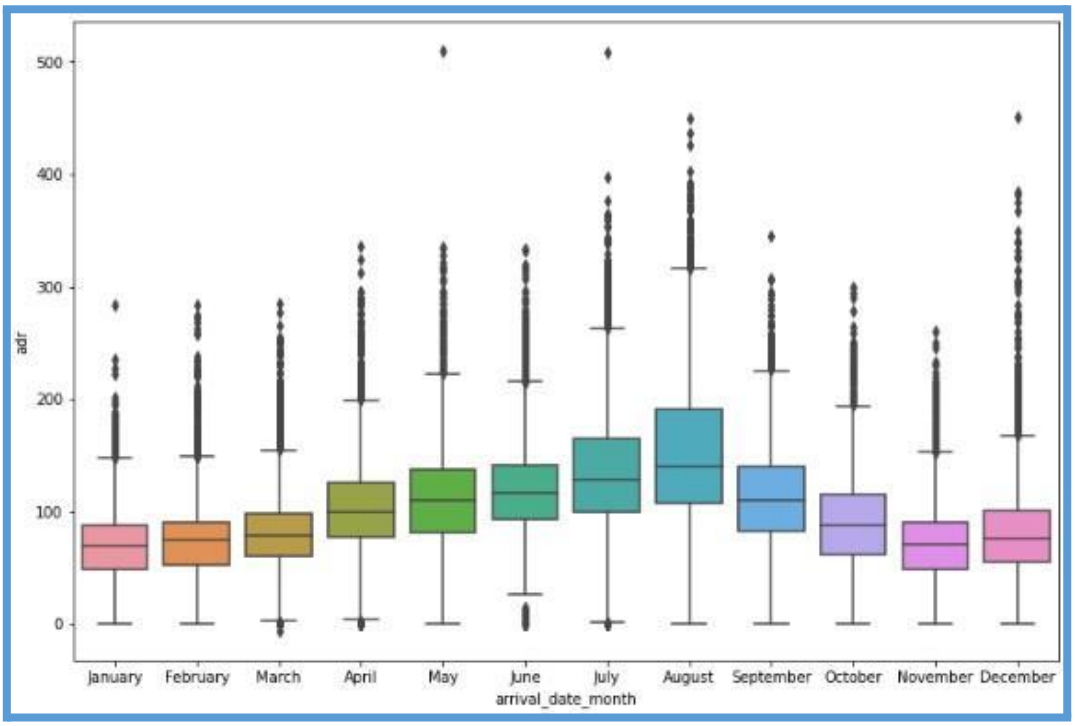


3. City Hotel bookings are more canceled than Resort hotel

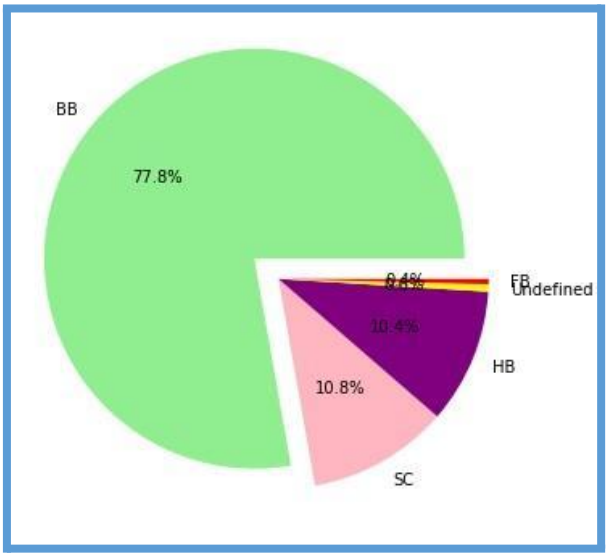
4. TA/TO distribution channel having longer waiting time compared to Corporate and Direct. Here “TA” means “Travel Agents” and “TO” means “Tour Operators”



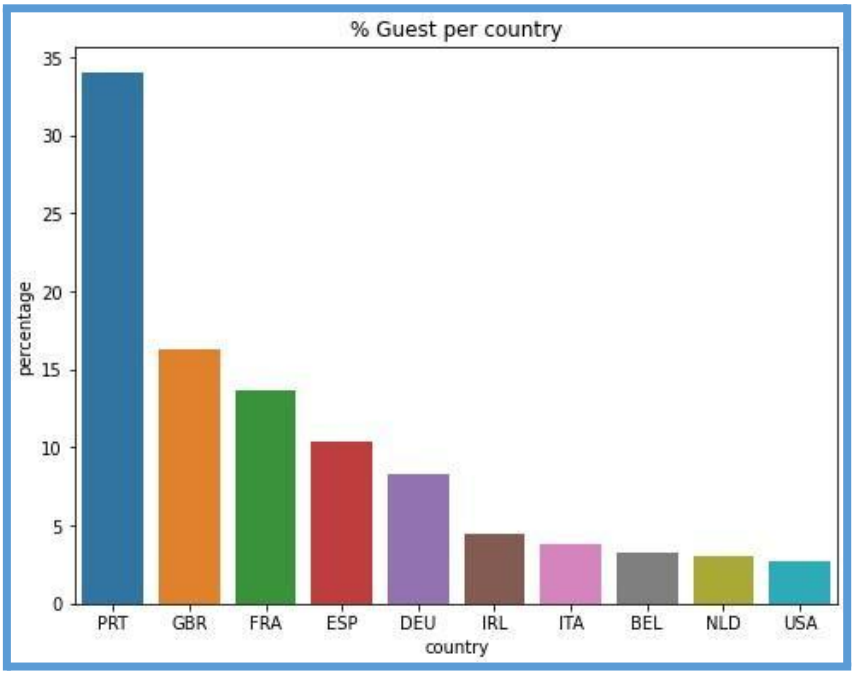
5. Less peoples are visited to Hotel in January month.so, revenue having huge cut off



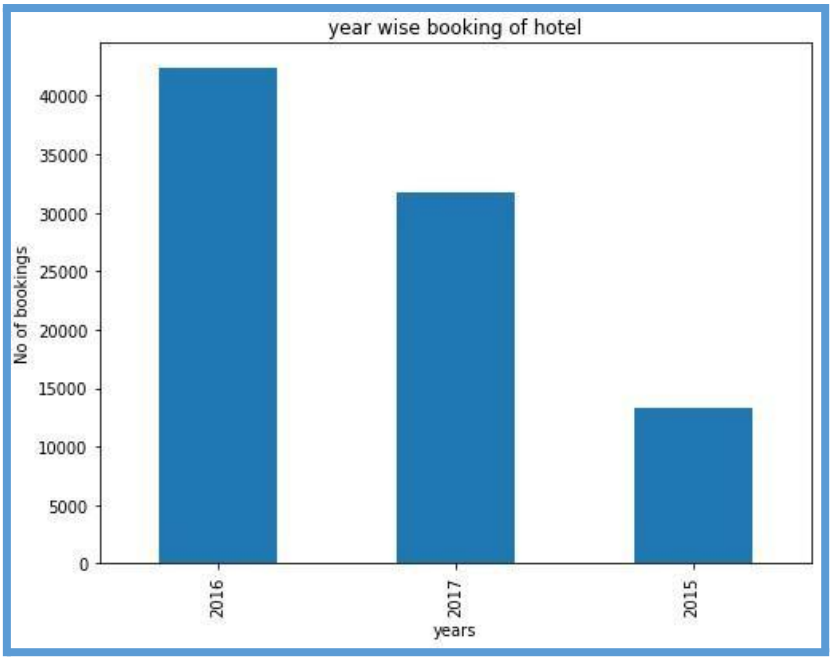
6. BB. (I.e. Bed & Breakfast) is most preferable type of meal for 77.8% guest



7. Most of the peoples are come from PRT (i.e. Portugal)



8. Increment of booking with alternate years.i.e. 2016 had higher bookings compared to 2017 and 2015.



Conclusion:

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

- Average daily rate (adr) is directly proportional to totle_peoples.No peoples increases then revenue must be increased.
- The percentage of city hotel is 66.45%.while the percentage of resort hotel is 33.55% is use to stay.So,City hotel connect more no of peoples and having higher lead time
- For longer stay peoples are choose Resort hotel and for short stay choose city hotel
- City Hotel bookings are more canceled
- Here,TA/TO distribution channel having longer waiting time
- More people visit hotels in August and less people visit in January.
- BB.(i.e. Bed & Breakfast) is most preferable type of meal for 77.8% guest
- Most of the peoples are come from PRT (i.e. Portugal)
- Year 2016 having higher bookings compared to year 2017 and 2015

References:

- Alma better
- Geeksforgeeks
- Stack overflow
- YouTube