Hotel Booking Analysis

Team Capstone Project - 1

Team Members:

Anushree V. Kajavadekar

Anjali Kasar

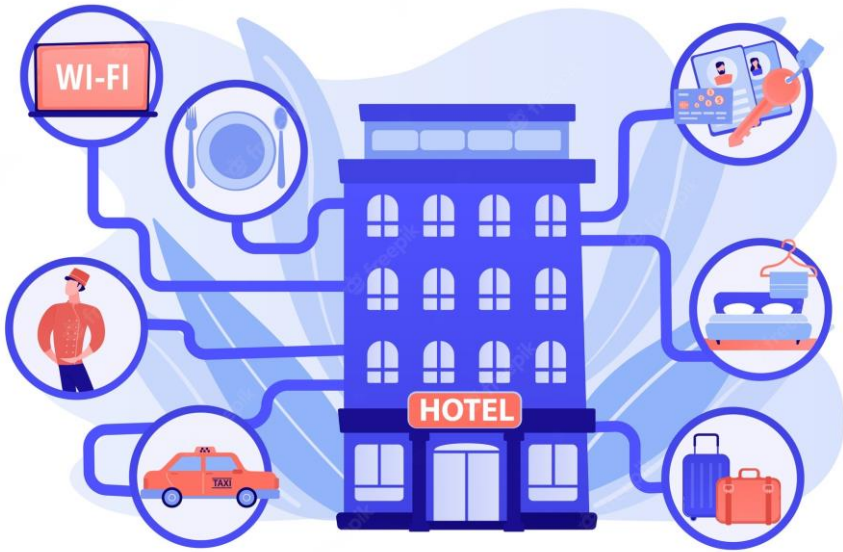
Prasanna R Kadlayyanavaramath

Satish Patil

Public

Outline:

1. Problem Statement
2. Data Summary
3. Data Preparation
4. EDA
5. Challenges
6. Conclusion



Public

Problem Statement:

Tourism is one of the world‘s most rapidly growing industries. Much of its growth is due to higher disposable incomes, increased leisure time and falling costs of travel.

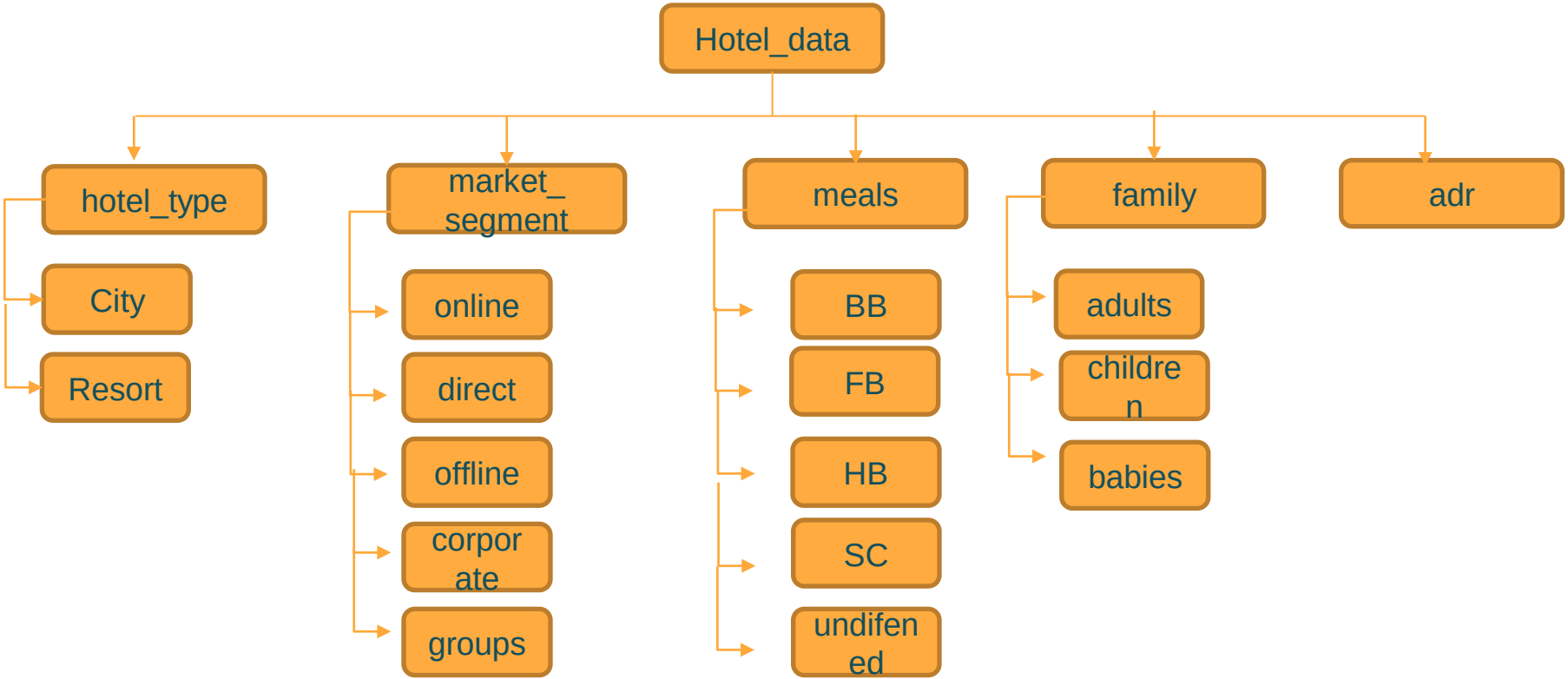
A hotel system manages information about rooms, reservations, customers, and customer billing. Hotel industry facing to analyze the problems like change in marketing trends and dynamics, housekeeping issues, customers’ expectations, Data security, mode of bookings etc...

We used the given data set to predict the future bookings using pandas data frame techniques.

We will be using the data set available to analyze the factors affecting the hotel bookings.

Public

Data Summary:



Public

hotel_type:In the given dataset hotels are classified into two types city and resort hotels.
Is_canceled:total number of bookings are canceled.
lead_time: the number of days between the time a guest books their room and the time they are scheduled to arrive at the hotel.
booking_changes:any change required by the Client in the travel and/ or arrival date, any increase or reduction in the number of rooms booked, and/ or any services required after the date of Confirmation.
market_segment:mode/platform of booking done by the customer.like,online, offline etc..
Adr:It stands for average daily rate, and it's used to measure the average revenue that a hotel receives for each occupied guest room per day. By measuring the ADR for your property, you're able to see the average rate that comes from all occupied rooms.
Stays_in_weekend_night:Number of people booked and stayed for weekend nights.
Reservation_type:describes about room type which customers are booked.ie,A,B,C,D,..

Public

Data cleaning:

Data cleaning is the very first important fundamental thing which every data scientist must know. It is the process of finding the inaccurate, incorrect and irrelevant or missing part of a data and then, modifying the data according to our necessity.

- **Steps:**

1. 1.Remove duplicate rows.
2. 2.Removing NULL values by replacing zero.
3. 3.Converting datatypes.
4. 4.Adding new columns if necessary.

Public

Data Visualization:

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

1) Bar Plot.
2) Histogram.
3) Scatter Plot.
4) Pie Chart.
5) Line Plot.
6) Heatmap.
7) Box Plot

Public

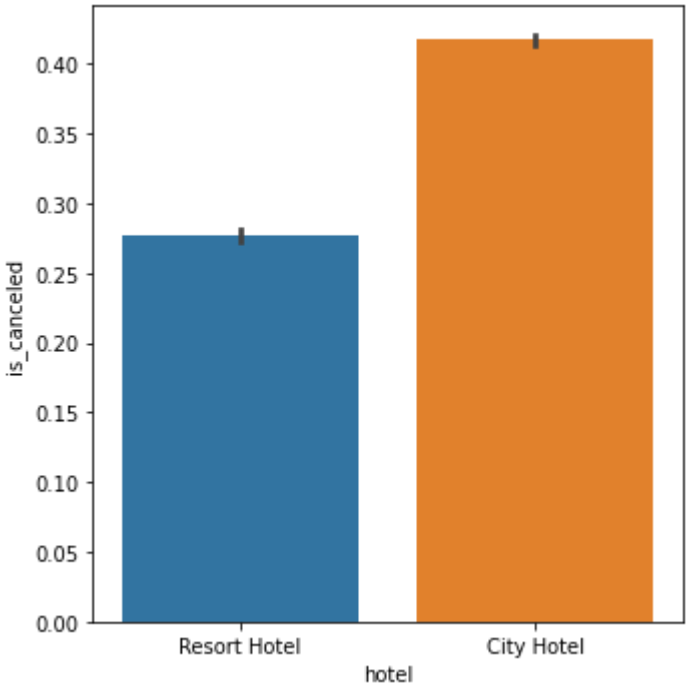
EDA:

Countries from which the customers visited hotel

- Arrival of guests is high in countries with code PRT,GBR,FRA,DEU,ESP. Portugal, Great Britain, France, Germany, Spain are the top countries ,most guests come from these 5 countries. We can conclude that country 'PRT' made huge number of bookings as compared to other countries.

Public

Hotels with the cancellation rates



- Using Seaborn, we have visualized the data for cancellations rates for the Resort and City Hotels.
- This plot shows us that City hotels has more cancellation than the Resort Hotels which is around 40%.
- So from this data, we can observe that customers city hotels should improve the services to avoid the cancellation.

Public

Problem with Cancellations

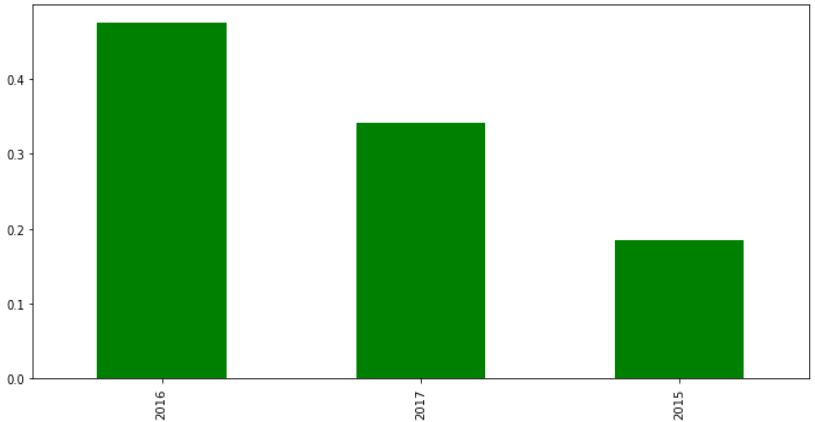
- Customer accustomed to free cancellation policies
- Operational Problems
- Non accurate forecast
- Non-optimized Occupancy
- Poor Management
- Revenue Loss

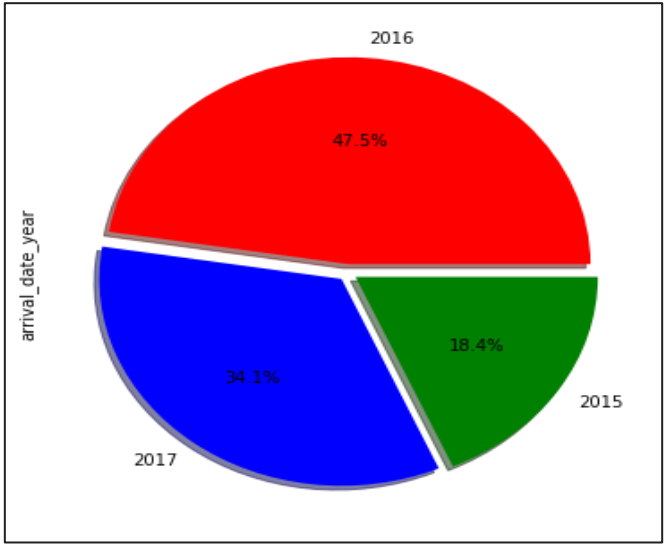
Approach: In order to fight the negative impact of cancellations, hotels need to be able to identify which bookings are likely to be cancelled.

Public

Busiest year and month

Busiest Year

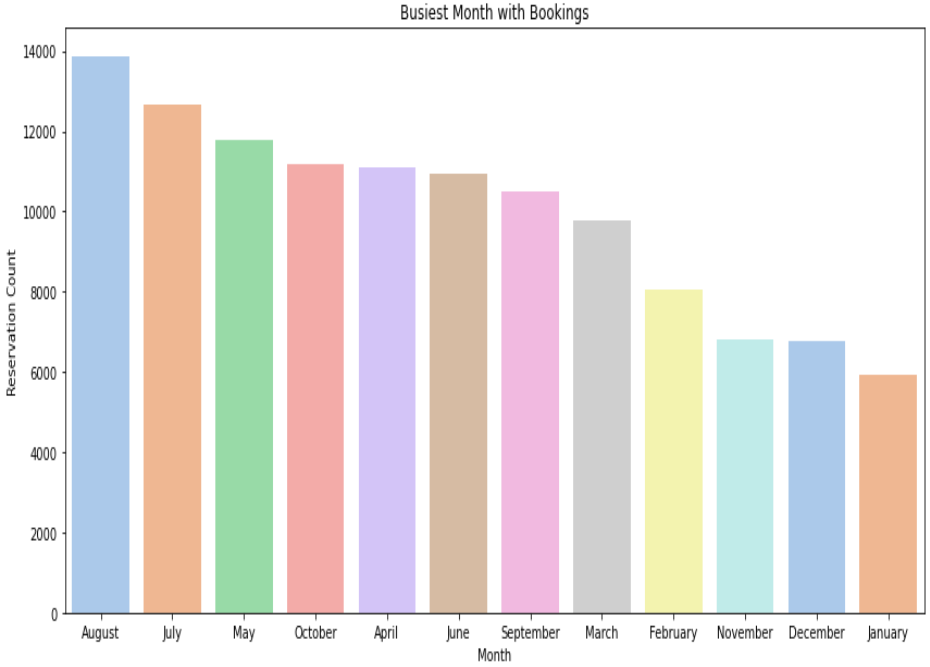




- Dataset contains booking data of 3 different years(2017,2016,2015)
- We can see from the data that maximum booking took place in 2016 and 2015 has the least bookings.

Public

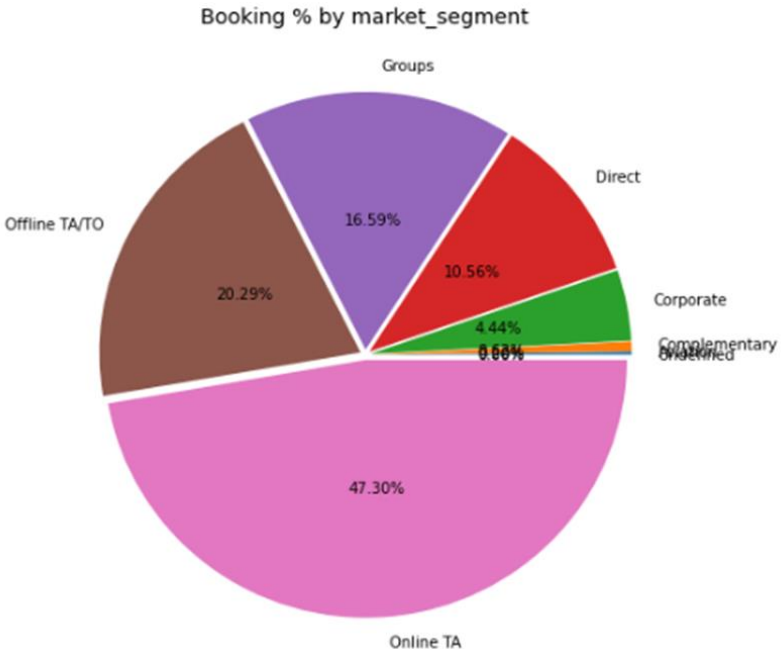
Busiest Month



- From the plot above, we can see July and August are the busiest months for both the City and Resort Hotels whereas January is month with least bookings.
- Hotel management need to apply some strategies to improve the rate of bookings for the least busy months.

Public

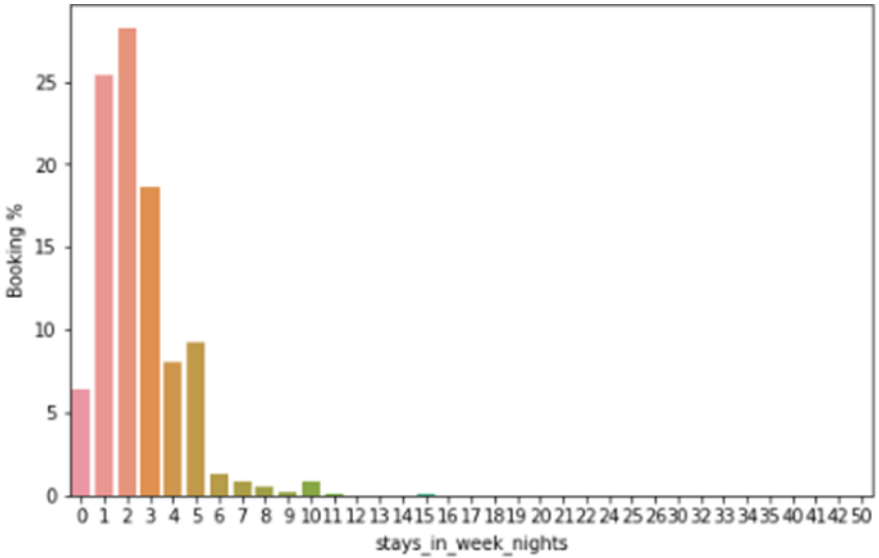
Mode of booking:

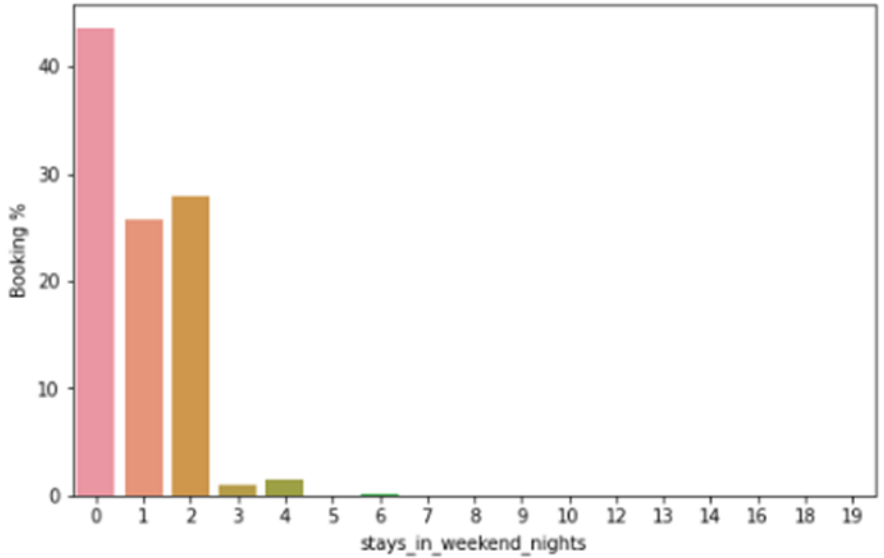


- Over half of the bookings made by online TA mode.
- Here, 70% of booking done using three methods are: Online, Offline and groups compared to direct and other agents booking.

Public

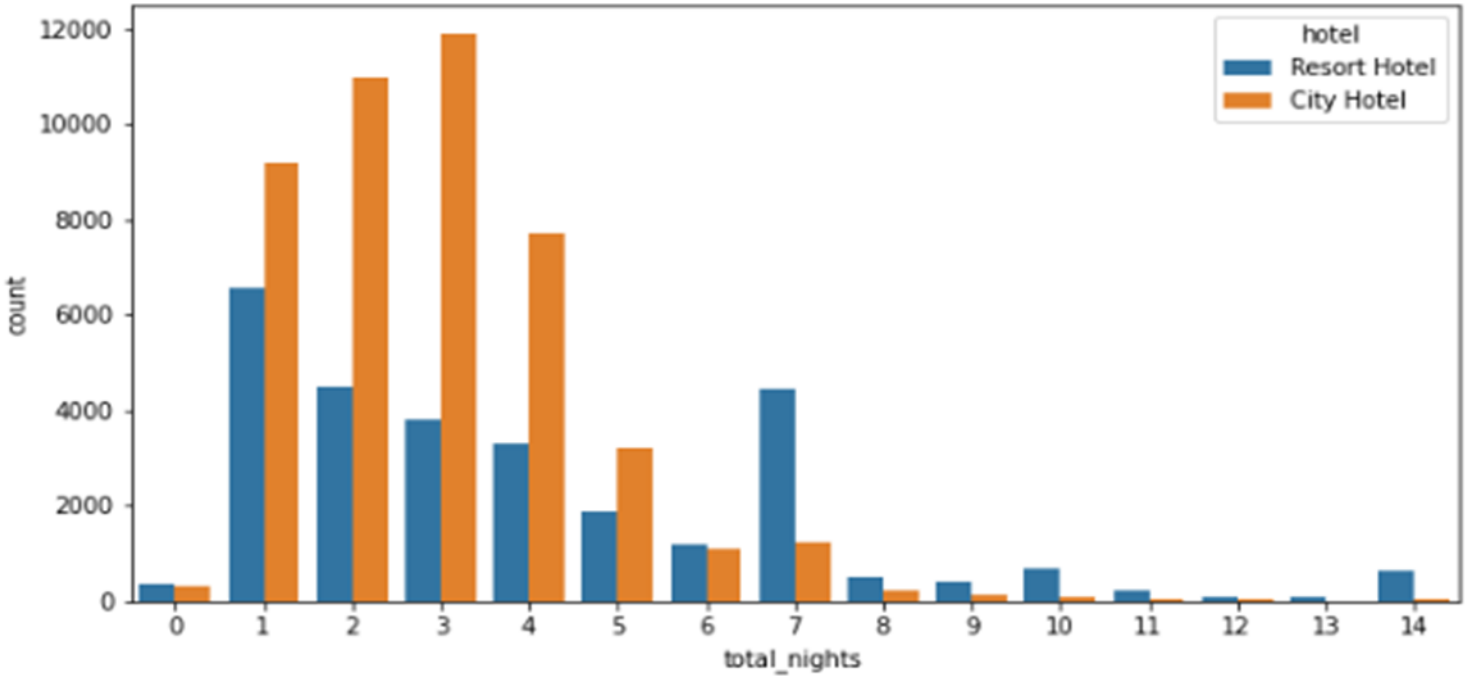
Comparing stays in weeknights, weekend nights and total nights in hotel:





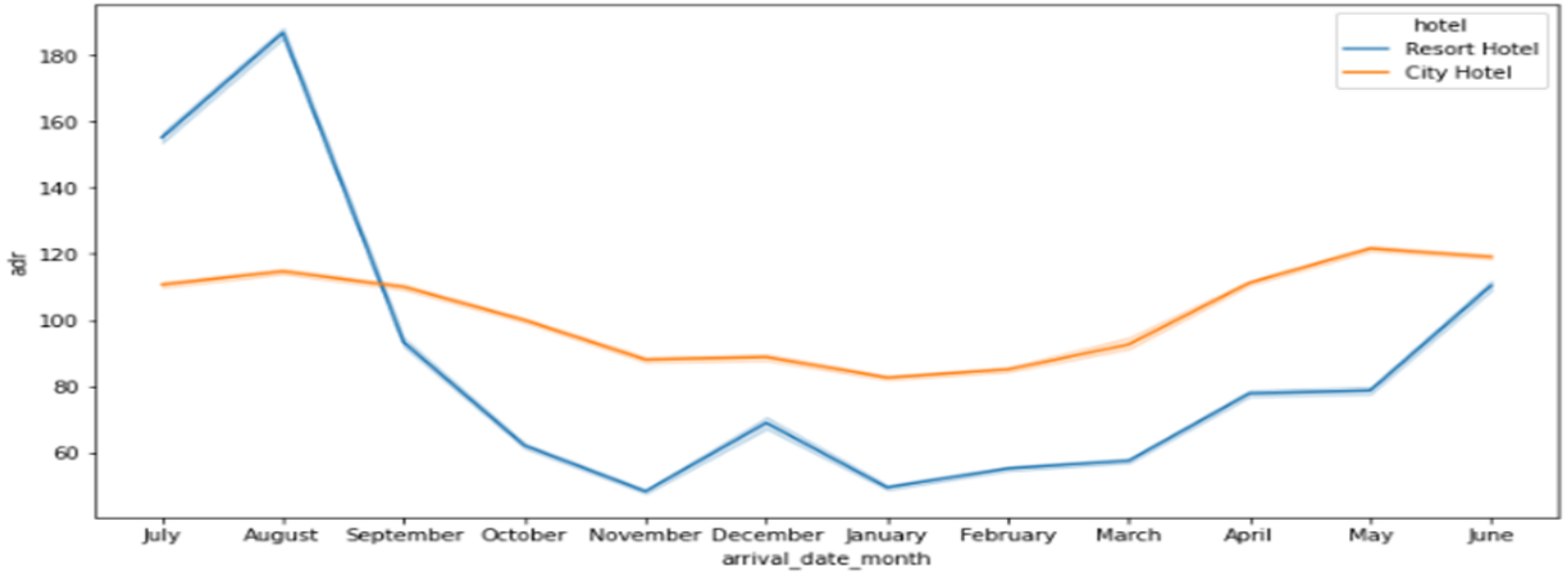
Public

Total Nights stayed in Hotel:



Public

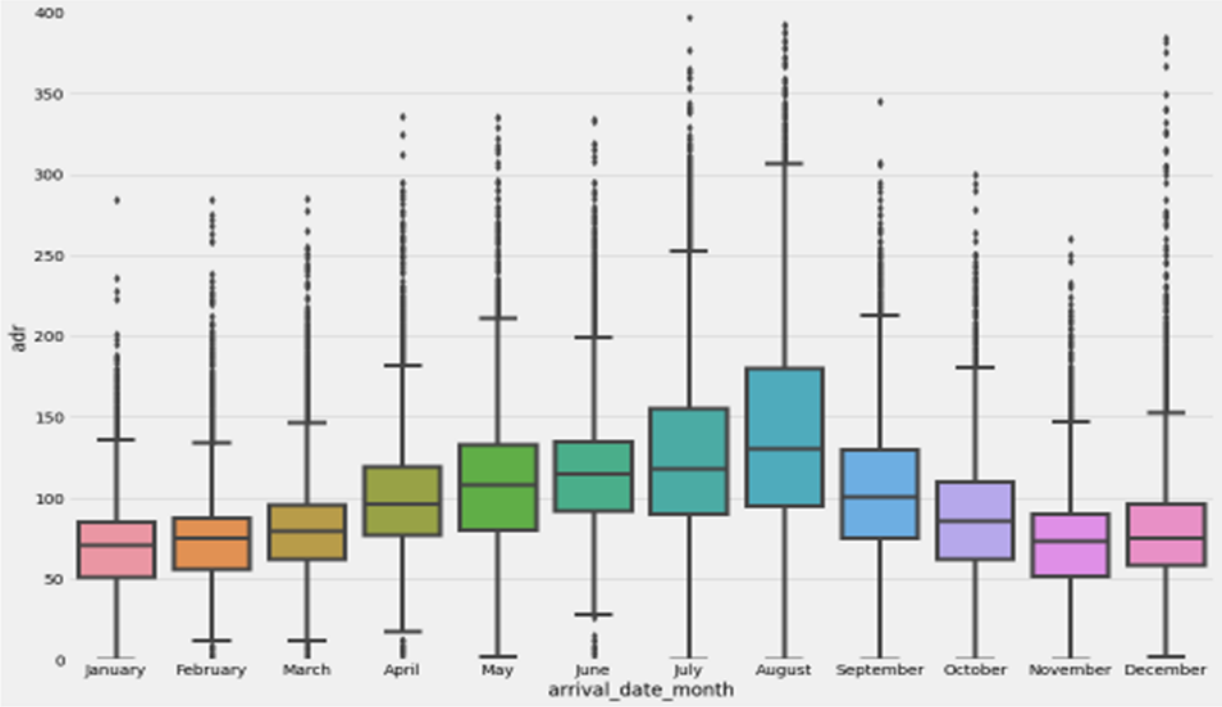
Average daily rate(adr) Analysis:



- Resort hotel type Adr was very expensive during July and august .
- During May,June and August Adr of city hotel slightly more compare to resort hotel.

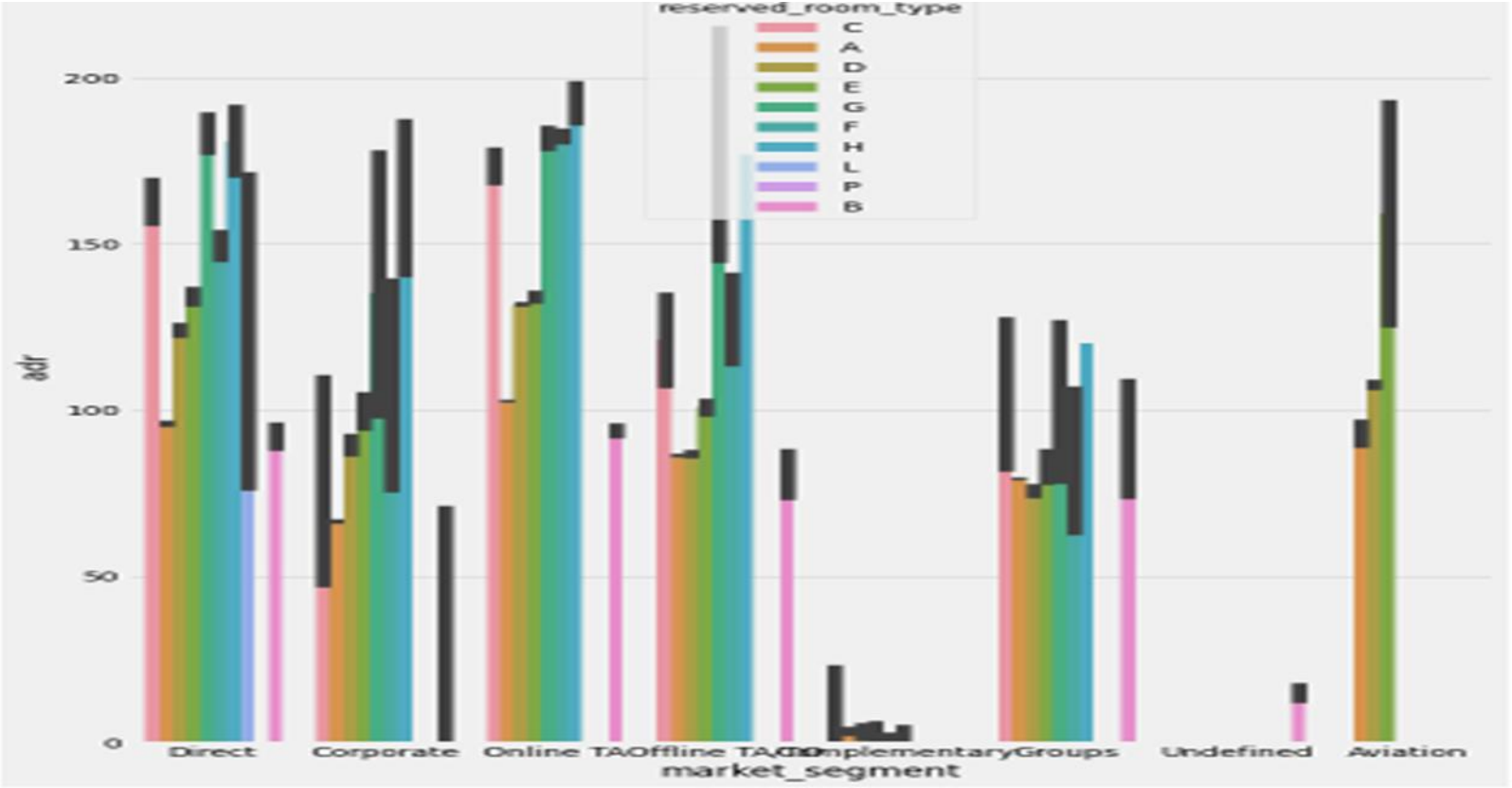
Public

Adr analysis using Box Plot:



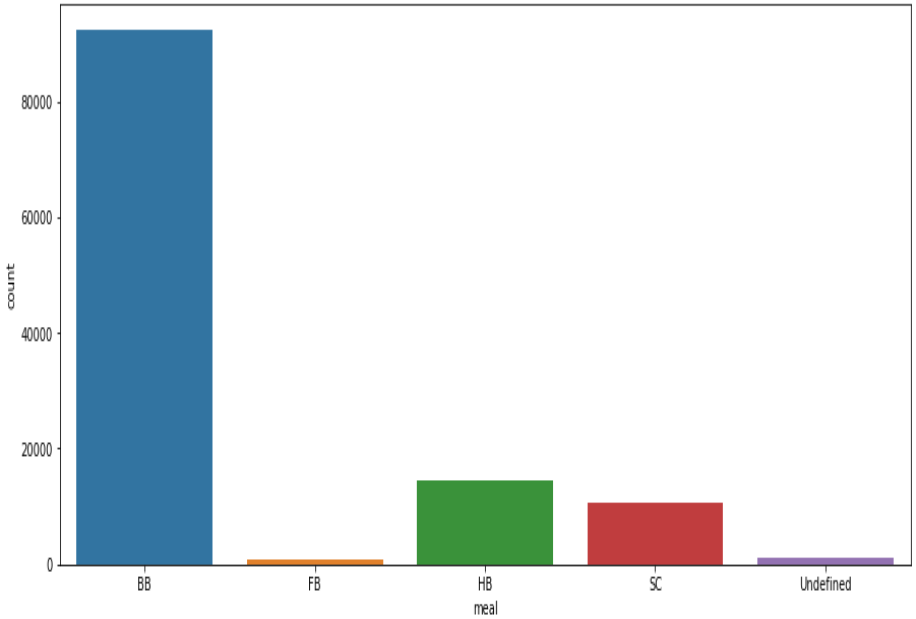
Public

Prices of Room types



Public

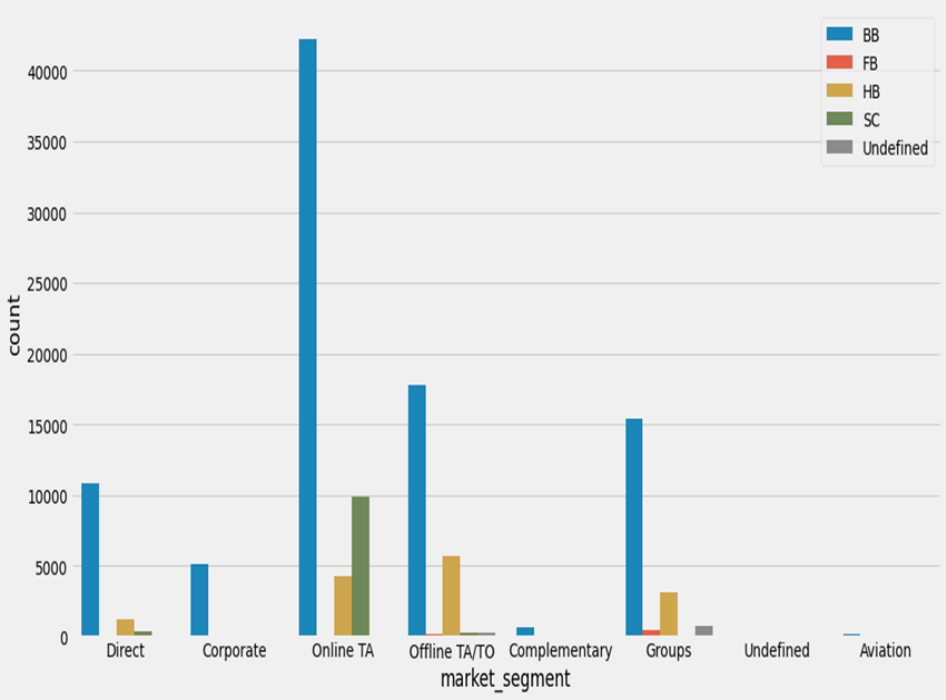
Type of Meal Preference by customers



- Out of total meal options, type BB meal has been availed by most customers followed by HB, SC & FB.
- As BB type of meal availed most by customers, hotel must be prepared with adequate supply of same to avoid any dissatisfaction to customer services or last-minute chaos.
- Meals also is a source of revenue to hotel, if customers only opting for BB this means either most of them checking out after 1 night stay or lack of engaging activity at resort. Resort could propose package or engaging activity which will increase preference of meal type FB

Public

Meal type preference by Market segment



- Most bookings done through online mode, but it has low contribution to meal type FB
- Group has significant contribution to meal type FB.
- Corporate has low/nil at availing meal services.
- Aviation has only contribution to type of meal BB.
- From above graph it can be seen online mode is contributing more to number of hotel booking & meals preference is BB while Groups has significant contribution in availing FB type of meals

Public

Challenges:

- Data cleaning is the difficult thing to resolve in a short time.
- It was hard to find which graph technique to use.
- Removing null values.
- Analyzing the data and visualizing box plot and scatter plot made confused.

Public

Tips to Improve:

- Show great care for your employees because your staff is the hallmark of your service
- Prioritize technology.
- Personalize customer service.
- Value Customer Service Surveys.

Public

Conclusion:

1.City hotels' and 'Resort hotels' are two types of hotels present in the data.
2.The cancellation rate of city hotel is higher than the Resort Hotels. Also Over 62.96% of total bookings were not cancelled irrespective of hotel type and 37.04% of total bookings cancelled.
3.Dataset contains booking data of 3 different years(2017,2016,2015),out of which, maximum hotel bookings took place in 2016 & 2015 witnessed the least number of hotel bookings.
4.Out of all months, 'August' witnessed highest number of hotel bookings whereas 'January' witnessed the least.
5.Among all the countries in the dataset, PRT(Portugal) has got the maximum number of hotel bookings.
6.The Online mode of hotel booking is preferred by majority of customers compare to other offline methods like TA/TO. Also lot of customers bookings done by online TA/TO are preferred 'BB' as a first choice and we can see 'FB' has least choice.
7.Average ADR of 'City hotels' is less than that of 'Resort hotels'.
8.Most Families preferred Resort hotels compare to city hotels
9.Customers prefer City hotel for longer stays over the Resort hotel.

Public

Q & A

Public

THANK YOU

Public