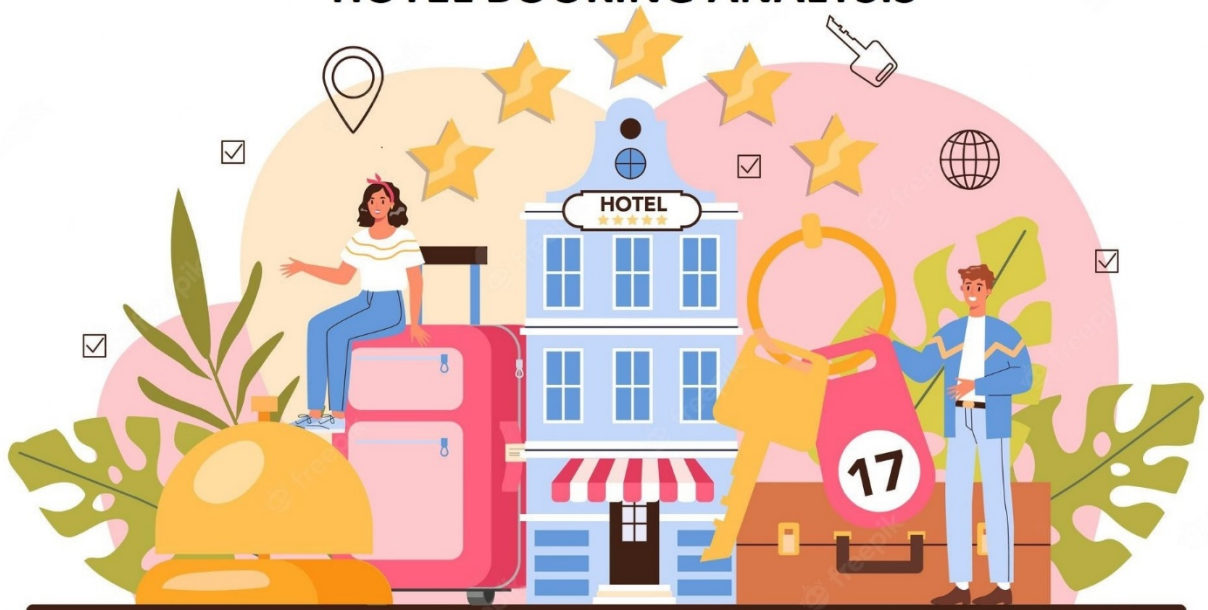# AlmaBetter

Capstone Project-1

HOTEL BOOKING ANALYSIS

Dr. Raj Kumar

Dileep Rawat

Prachi Jadhav

Shubham TIwari

Nowadays, hospitality business is on high demand as everyone prefer to travel with their family to tourist place and many of people have to stay outside of home due to their job or other reason. So, hotel owners want to their hotel is running with full capacity & getting huge revenue.

This project we have analyzed a dataset of Hotel Booking and try to find out useful factors which may help the owners to increase the booking. The dataset 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. Also there are other important information that we have figured out by analyzing this data.

We found that all possible conclusions which may help the hotelier to increase their revenue. There are various possible reasons of getting less revenue for a particular hotel viz. Hotel facilities, Hotel location, Service, Rate and others.
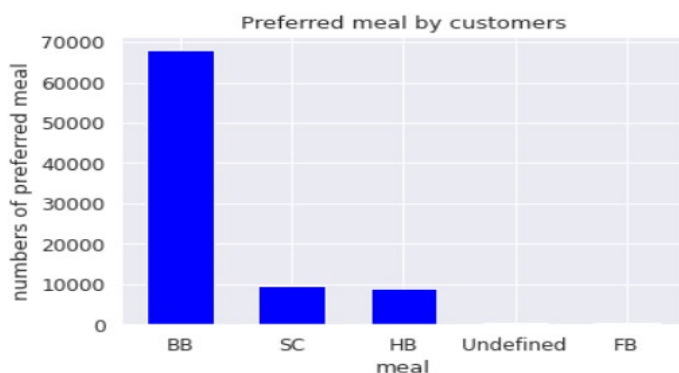
Table of Contents:

Exploratory Data Analysis

We have tried to analyse the following statements under the mentioned category.

1. Univariate Analysis

I. What is the most preferred meal by customers?
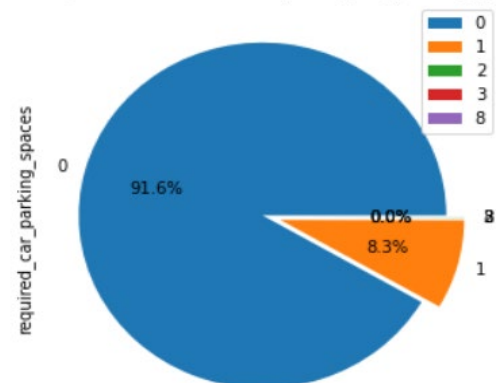

Preferred meal by customers

Most preferred Meal type is BB viz. Bed & Breakfast. On the other hand HB-

Half Board and SC-Self Catering are equally preferred.

ii. What is the percentage distribution of required car parking spaces?

Percentgae distribution of required_car_parking_spaces

From pie chart it is clear that 91.6% customers are not required parking and only 8.3% customer required 1 car parking space. So having less parking space or no space will not affect much on the revenue but Car Parking could be used as a perk to allure customers.

iii. What is the percentage of booking changes made by the customer?
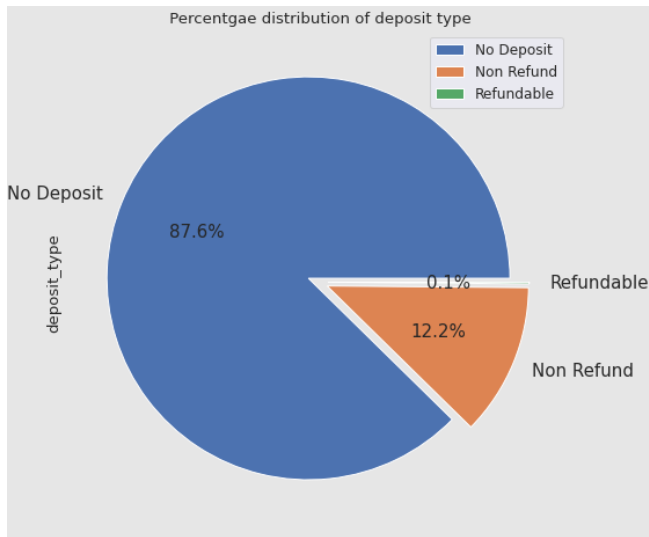

Percentage distributaion of booking changes

Approximate 80% of the bookings were not changed by guests.

0= 0 changes made in the booking

1= 1 changes made in the booking

### iv. What is Percentage distribution of Deposit type?

Percentgae distribution of deposit type

From the pie chart clearly visible that the most preferred booking type by the customers is No Deposit type (87.6%). As No deposit type is quite risky so there is need to increase other booking type as refundable type is least so there is need to focus on that to increase assured booking.
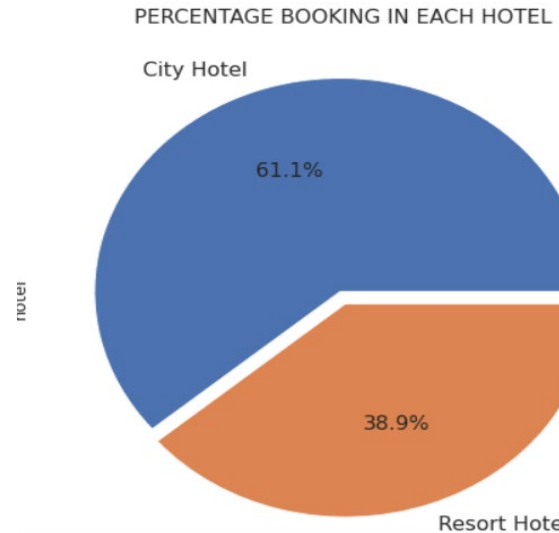
v. Which is the most preferred room type by the customers?

Percentage distributaion of assigned_room_type
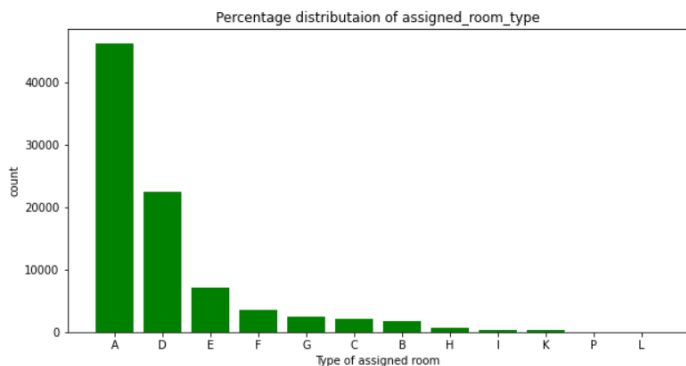
Room Type A is the most preferred there may be factors related to budget or no. of rooms. But this is playing a major role in getting most of the bookings.
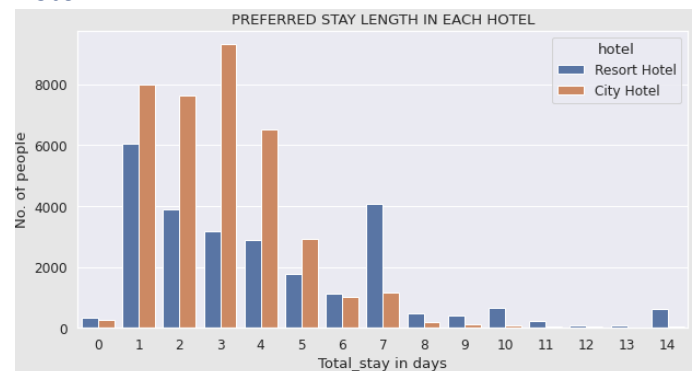
2. Hotel wise Analysis

i. Which type of hotel is mostly preferred by the guests?

PERCENTAGE BOOKING IN EACH HOTEL

City Hotel has secured 61% of the total bookings thus city hotel is quite popular there may be factors like budget, availability or short stay.

## ii. What is most preferred stay length in each hotel?

In resort hotel customer preferred to stay maximum number of nights viz. 7 however majority of the customers preferred to stay for 2 or more than 2 Nights in City hotel.

iii. Which hotel has higher lead time?

PREFERRED STAY LENGTH IN EACH HOTEL
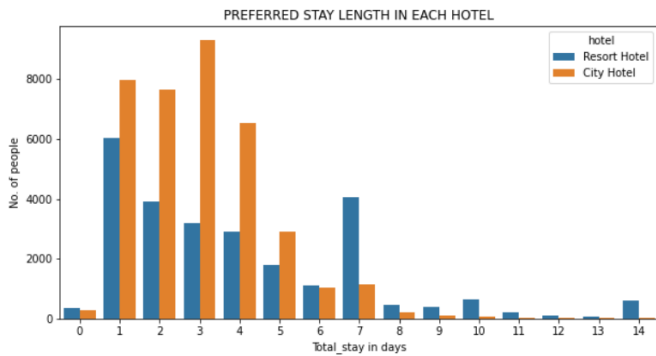
Lead time may help the hotel to predict the upcoming booking & they can make their arrangements as per this data. In the mentioned data most of the booking were made with lead time up to 5 days.

iv. Which hotel makes more revenue?
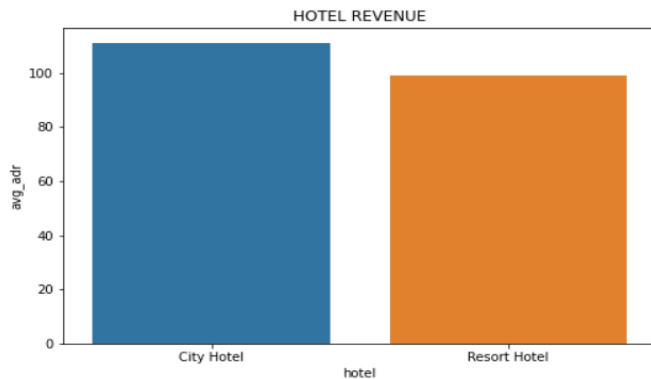
HOTEL REVENUE

It can be concluded that City hotel is contribution more to the overall revenue but as bookings count is way too high for city hotel so this might be the reason for its high revenue than Resort hotel which have got only 40% of the total booking.

v. Which hotel has the higher customer retention-rate?

CUSTOMER RETENTION RATE
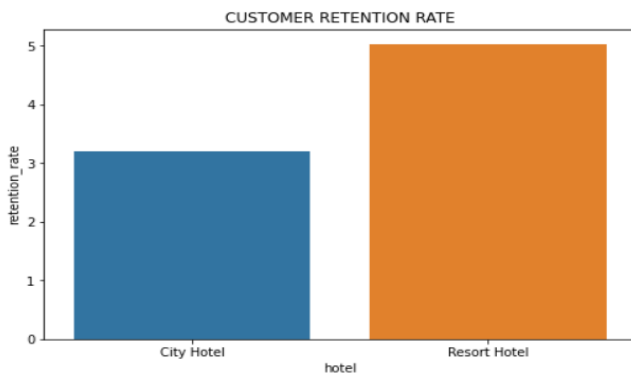
Resort hotel his getting most of his refereed customers as compared to city hotels which is a positive sign for resort hotel. We can predict that

customers are enjoying their stay in Resort hotel.

vi. For which hotel, does people have to wait longer to get a booking confirmed?

WAITING TIME FOR BOOKING
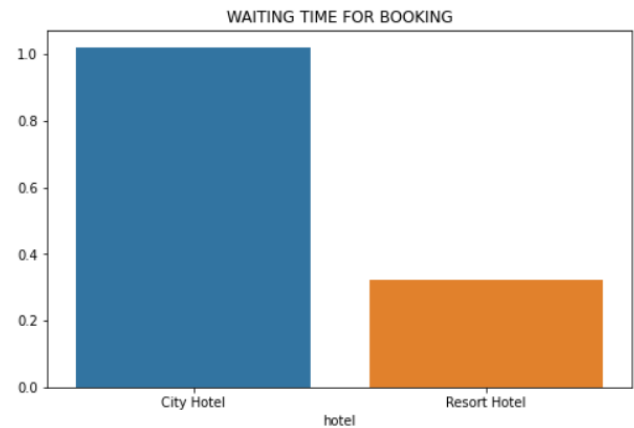
The reason for higher waiting time in the city hotel is due to most of the customers tends to book city hotel as compared to Resort Hotel.

vii. From which country most guest come?


NUMBER OF GUESTS FROM DIFFERENT COUNTRIES
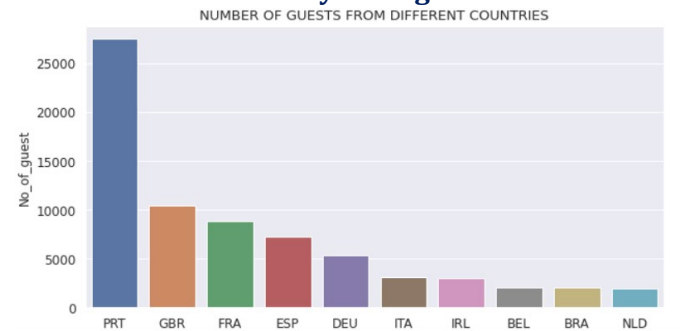
Most of the guest who booked their stay in the hotel are from Portugal so we may focus on this country in order to attract more customers from Portugal. But we need to work on other country as well to get more customers.

3. Distribution channel wise Analysis

i. Which Distribution Channel is contributing in most of the hotel bookings?
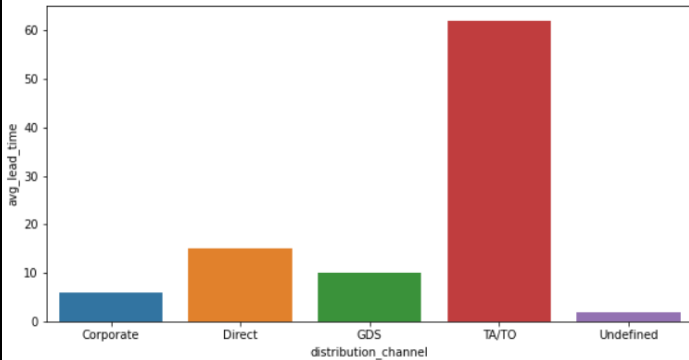
Booking % by Distribution Channels
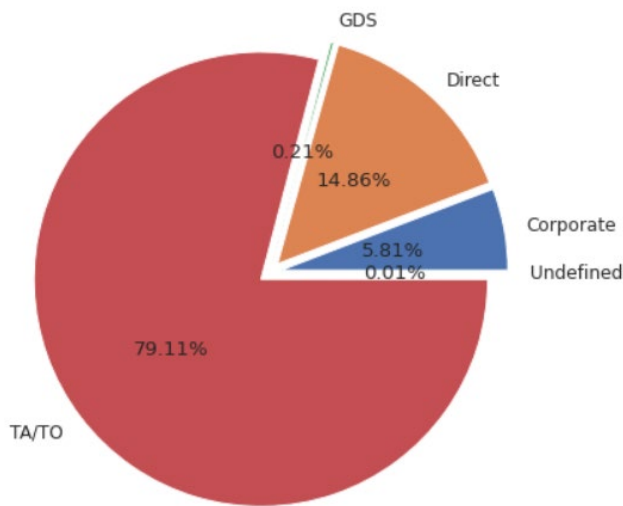
Highest Booking received by the hotels are through TA/OT so they are one of the most trusted booking provider. Also, Direct booking count is greater than GDS system might be customer does not have complete faith over online booking platforms or we may be not registered in major GDS platforms, we need to look into it.

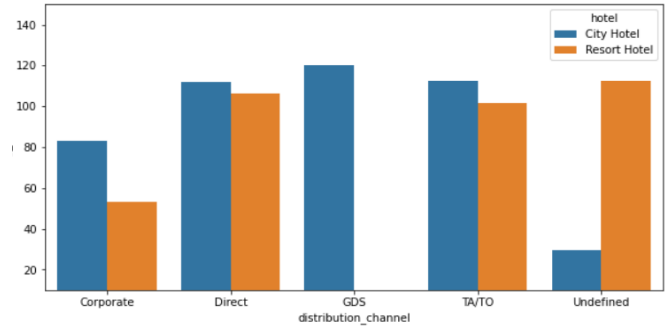## ii. Which channel is contributing most for early booking of the hotel?

We concluded that the most of the early bookings we have received from TA/TO. Other than undefined the least booking we have received from Corporate.

## iii. Which distribution channel brings better revenue generating deals for hotels?

1-In terms of revenue GDS is the most revenue generating Channel but its only for City hotel. For Resort Hotel its contribution is negligible as compared to other channels distribution.

2- Undefined can be associated to multiple channel distribution channels whose data is not provided so after undefined bookings from TA/TO are generating most revenue for the Resort Hotel.

3-Apart from other distribution channel Direct bookings are also playing the crucial role in terms of revenue generation but we need to focus more on other less revenue generating mediums in order to increase the overall revenue.

iv. Which is the most favorable Channel for Customers to book hotel?
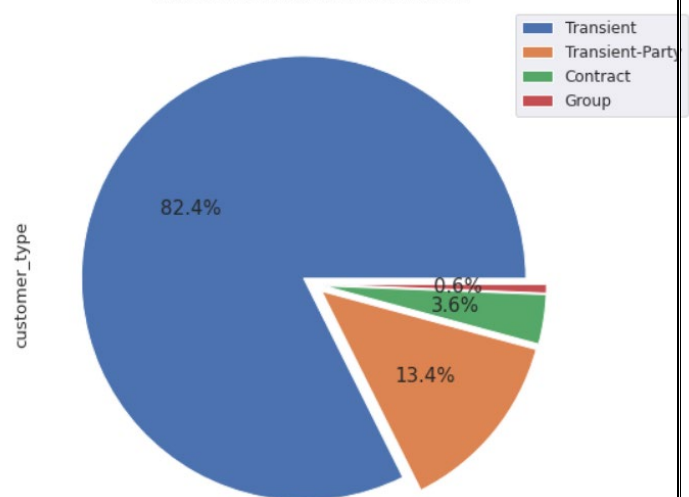
Transient Booking - These are walk-in bookings, last minute or bookers or simply people that require a very short term stay
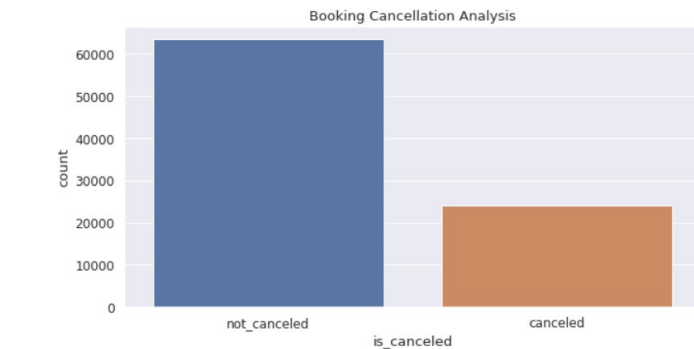
Transient Party - They are somewhere linked to Transient Group.

Contract Booking- These booking are on contract basis made by the hoteliers with their partners.

Group Booking- The booking made in groups of 1 or more than 1 room. Generally, it is of more than 1 room.

4. Booking cancellation Analysis

i. How many bookings were cancelled?


Booking Cancellation Analysis

Cancellation number is quite higher which is varying between 20 to 25% of the total bookings. We need to minimize it by marking out the suitable factors which are causing it.

ii. Which hotel has higher booking cancellation rate?


PERCENTAGE BOOKING CANCELLATION

The booking cancellation rate is quite high in the city hotel as compared to resort hotel there are chances that city hotel is getting majority of bookings so that's why their cancellations is also high off course there may be other factors that we need to figure out to reduce cancellations.

iii. Which significant distribution channel has highest cancellation percentage?

DISTRIBUTION CHANNEL WISE CANCELLATION
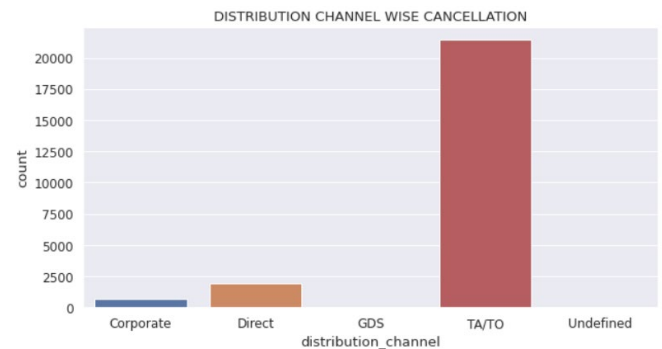
Cancellation of bookings are way to higher from the Travel Agents/Travel Operators we need to figure out that which particular TA/TO is constituting the most & how we can amend our policies in order to minimize cancellations without affecting our booking count.

iv. What is Year Wise Percentage of Cancellations?

Yearwise Cancellations
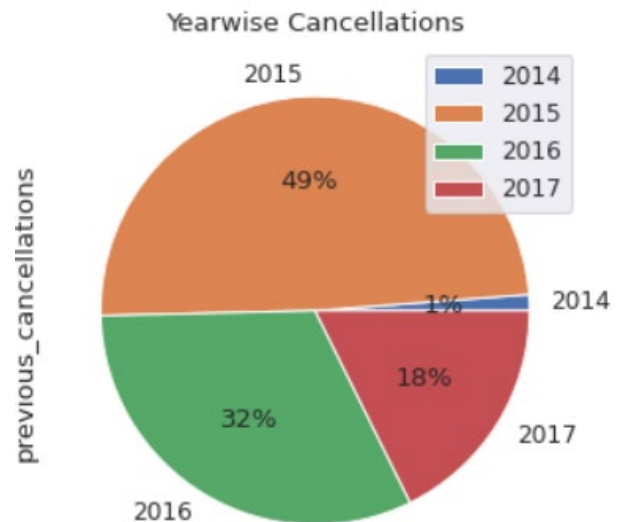
Hurrah!! We have some positive analysis that the number of cancellations thorough out the years are decreasing with a great pace so here pat your back.

v. Which period of year has longer lead time analysis?

Leadtime analysis of four years
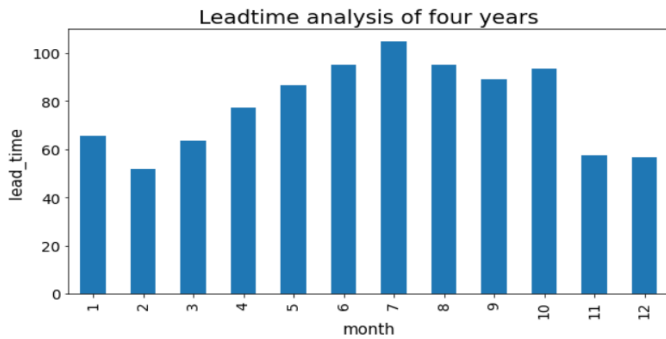
May to October have shown subsequently higher lead time as compared to rest of the year. These months have average lead time of 80 or more than 80 days. This may due to some festival or holidays were occurring during these months.

vi. What is Average Daily Rate Year wise and Month wise?
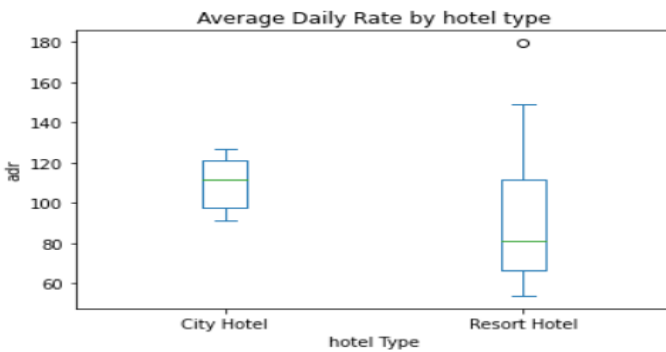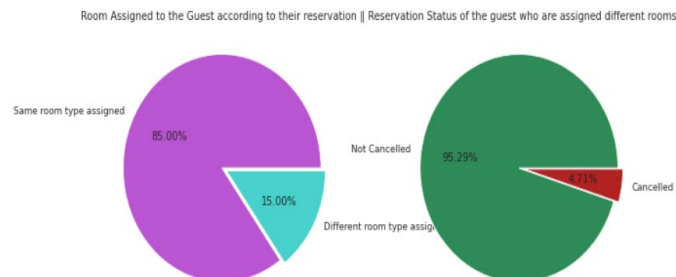

Average Daily Rate by hotel type

Here we can find that the Resort Hotel has highest Average daily rate for a particular booking but it is interesting to know that average/mean adr of city hotel is more than resort hotel.

5. Customer Centric Analysis

i. How not getting same room as reserved affects ADR?

Room Assigned to the Guest according to their reservation || Reservation Status of the guest who are assigned different rooms

From here we can calculate that those who were not assigned same room as reserved does not affects ADR (Average Daily Revenue. From pie chart we can analyse that only 5.39% guest cancelled their reservation after assigning different rooms.

ii. Overview of arrival period.
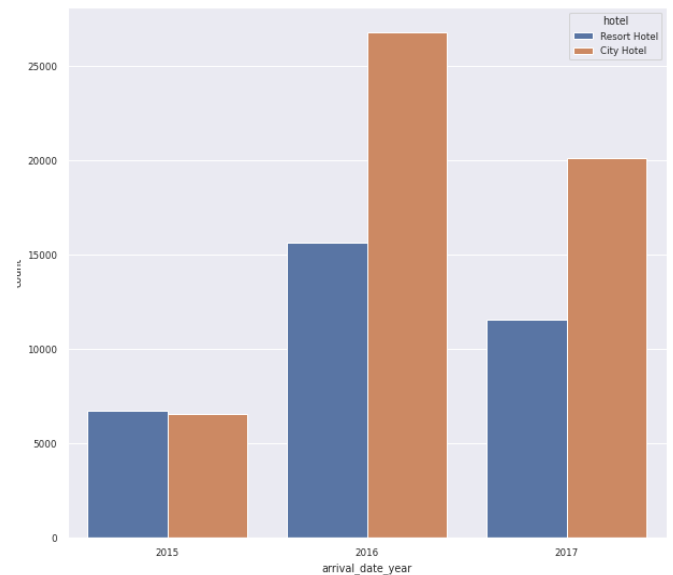
The year 2016 was the most successful year in terms of hotel booking we had created a new benchmarks & the booking numbers are way too higher than the previous year. But there is a fall in next year booking count we need to figure out the reasons behind it.

iii. Stay duration over a weekend and weekdays?

WEEKDAYS STAY

Number of stays on weekday night
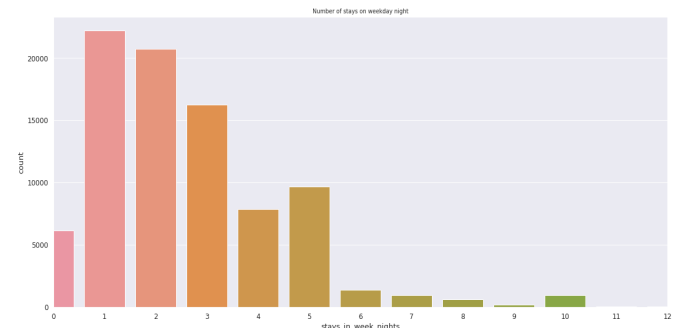
Most of the people preferred to stay for 2-4 Nights during the weekdays however there are some bookings whose stay was as long as of 10 days. Few of the customers have checked out on the same day that's why it's showing 0 in the duration.

WEEKENDS STAY

Number of stays on weekend nights

Graph represents that during weekends
most of the stays duration is less than
1 days which means they stayed for few
hours & checked it out after that.
However as compared to weekdays
weekends stay is not that for longer
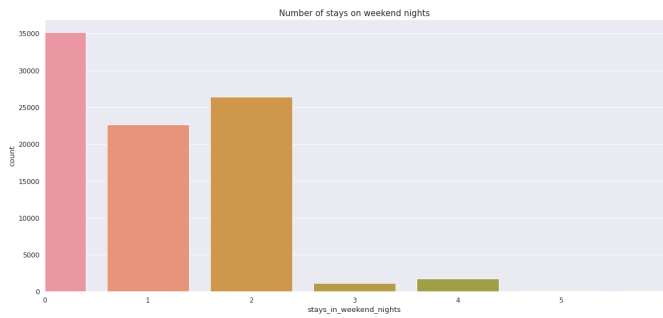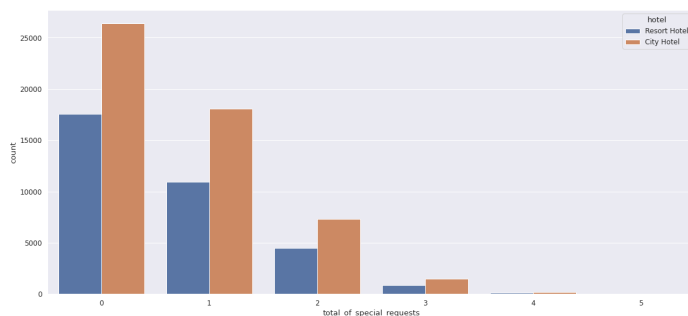duration most of the people preferred
to stay for 1-2 nights.

iv. Analysis based on special requests

In most of the bookings customers
didn't ask for any type of special
requests. Few customers added 1 or 2
special request only.

Tools & Libraries

  I. Pandas
 II. Numpy
III. Seaborn
IV. Matplotlib
 V. Folium
VI. Plotly Express

Challenges Faced

While exploring the provided dataset we found about
the challenges we are going to face while analyzing it.
But we have tried our best to get rid of all such
challenges in order to provide you the best analyzed
through taking few mentioned steps before analyzing it.

i. Duplicate Rows:

The data set have almost 32000 duplicate rows which
and such a huge amount is enough of duplicate data
is sufficient to manipulate our analysis.

ii. Missing Values:

Although most of the columns have none missing
values but there are 2 or 3 columns which have
sufficient amount of null values especially "Company"
column which have 82137 missing values.

iii. Conversion of datatypes:

Thera are columns in the data set whose datatype is
not appropriately mapped so in a way to get the
desired value & analysis we have convert the
columns into appropriate datatypes.

iv. Addition of Columns:

We were need to add some required columns which
were important for analysis this data. So we have
added some columns in the dataset using the pre-
defined columns in order to expand our analysis.

    Apart from all the above mentioned challenges
there are many other challenges we have faced while
analyzing the dataset but we successfully came out of
this using the help of Alma Better Lectures &
materials & some other trusted source like google,
stack overflow, pandas.pydata.org etc. also we.

Conclusions:

We have successfully cleaned our data in term of replace
**duplicates, Null values, and convert inappropriate
datatype to appropriate datatype.** We have
successfully find out the relevant insights from this
dataset as follow:

Univariate Analysis:

- ✓ Type 'A' room and type BB(Bed & Breakfast)
  meal have been preferred by most of customers
  followed by room type 'D' and meal type HB/SC,
  respectively.
- ✓ Since 98.7 % of the guests prefer No deposit type
  of stay. The high rate of cancellations can be due
  to high no deposit policies. About 91.6 %
  customers did not required the parking space
  and approximately 8.3 % customer's required
  only 'one' parking space.

Hotel wise Analysis:

- ✓ City Hotel is most preferred by customers and significantly longer waiting time. Thus city hotels have slightly more revenue but much busier than Resort Hotel.
- ✓ The time taken between when a customer makes a reservation and their actual arrival is called the lead time. Resort hotels have slightly high avg lead time, i.e., customers plan their trips very early to reserve resort hotels.
- ✓ Most common stay length is less than 4 days and generally people prefer city hotel for shorter stay, but for longer stay resort hotel is preferred.
- ✓ Resort hotel has higher retention rate compare to city hotel that means customers are willing to stay again in resort hotel and highest number of customers are from Western Europe, namely Portugal, UK and France. So, marketing team have to be target these regions.

Booking Cancellation Analysis:

- ✓ Maximum cancellation has been observed by Travel Agent in 2015 and City hotels have higher booking cancellation rate of approximately 25%.
- ✓ In month of August, longer lead time can be seen, i.e., customers had been booking their rooms so early and average daily rate is more in Resort hotel than city hotel.

Distribution channel wise Analysis:

- ✓ Highest Booking received by the hotels are through TA/OT so they are one of the most trusted booking provider. Thus most of the bookings we have received from TA/TO.
- ✓ GDS is the most revenue generating channel but its only for city hotel. For resort hotel its contribution is negligible as compared to other channels distribution.
- ✓ Undefined can be associated to multiple distribution channels whose data is not provided so after undefined bookings from TA/TO be generating most revenue for the Resort Hotel.
- ✓ The majority of booking channel is from Transient and Transient Party having 82.4% and 13.4% contribution respectively. Transient parties are somewhere linked to Transient Group.

Customer centered analysis:

- ✓ Figure (a) depicts the distribution of room assigned to the guest according to their reservation and result shows that 85% people got same room type which is assigned and 15% customers got different room. Figure (b) shows reservation status of the guest who are assigned different room and result shows that only 4.71% customers were cancelled their reservation.
- ✓ Year-wise, month-wise and day-wise hotel booking data are represent in Figures (a, b & c), respectively. It is clearly visible in figure (a), highest booking in city hotel as well as resort hotel were in 2016. From figure (b) depicts highest booking in July and August. Summer ends in Aug followed by autumn, so it seems that summer period is a peak for hotel booking. From figure(c) shows trend for the arrival day of month has been roller coaster.
- ✓ Majority of the stays are over the weekday's night. Whatever we saw for the chart on day of the month was random.
- ✓

Special Requests:

- ✓ More of special request is for city hotel.