



Capstone Project

Project Title: - Hotel Booking Analysis

by
Kunal Wankhede

Discussion point

- Introduction
- Problem Statement
- Data summary
- Data Preparation, Cleaning And Adding
- Data Visualization
- Conclusion
- References

Introduction

To discuss the analysis of given hotel bookings data set from 2015-2017
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 disproportional high number of special requests? This hotel booking dataset can help you explore those questions!

Problem Statement

- In this 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:
 - Data Reading.
 - Data Preparation, Adding and Cleaning
 - Data Visualization

Data summary

- Given data set has different columns of variables crucial for hotel bookings. Some of them are:
- hotel: The category of hotels, which are two resort hotel and city hotel.
- is_cancelled : The value of column show the cancellation type. If the booking was cancelled or not. Values [0,1], where 0 indicates not cancelled.
- lead time : The time between reservation and actual arrival.
- meal : Meal preferences per reservation.(BB, FB, HB, SC, Undefined)
- Country : The origin country of guest. country of
- market segment : This column show how reservation was made and what is the purpose of reservation. Eg, corporate means corporate trip, TA for travel agency guest
- '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



- **distribution channel** : The medium through booking was made. (Direct, Corporate, TA/TO ,undefined ,GDS.)
- **is repeated guest** : Shows if the guest is who has arrived earlier or not. Values [0,1] --> 0 indicates no and 1 indicated yes person is repeated guest.
- **days in waiting list** : Number of days between actual booking and transact.
- **customer type** : Type of customers(Transient, group, etc.)

Data Preparation, Cleaning And Adding

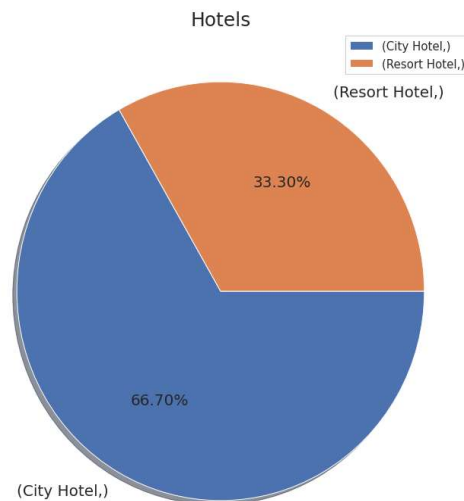
Unifying messy and complex data sets to a meaningful format for easy access and analysis. There are various processes designed to transform raw data into more readily used formats.

It includes following steps:

- Handling missing values.
- Removing duplicates data.
- Converting columns to proper data type format.
- Adding or removing columns for analysis.

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



Hotel bookings per year

While doing Hotel bookings per year of given hotel booking dataset, we answered following questions

1. Which year has the highest number of confirmed bookings?
2. Which year has the lowest number of confirmed bookings?
3. Which year has the highest number of canceled bookings?
4. Which year has the lowest number of canceled bookings?



In 2016, city hotels and resort hotels had the highest number of confirmed bookings as compare to 2015, 2017

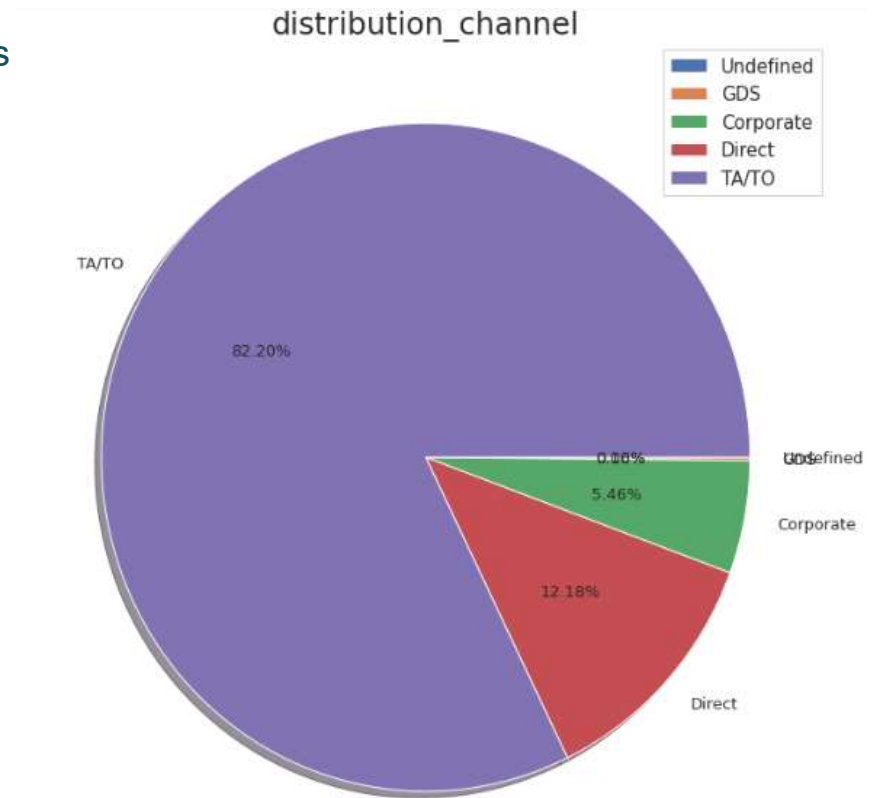
In 2015, city hotels and resort hotels had the lowest number of confirmed bookings as compare to 2016, 2017

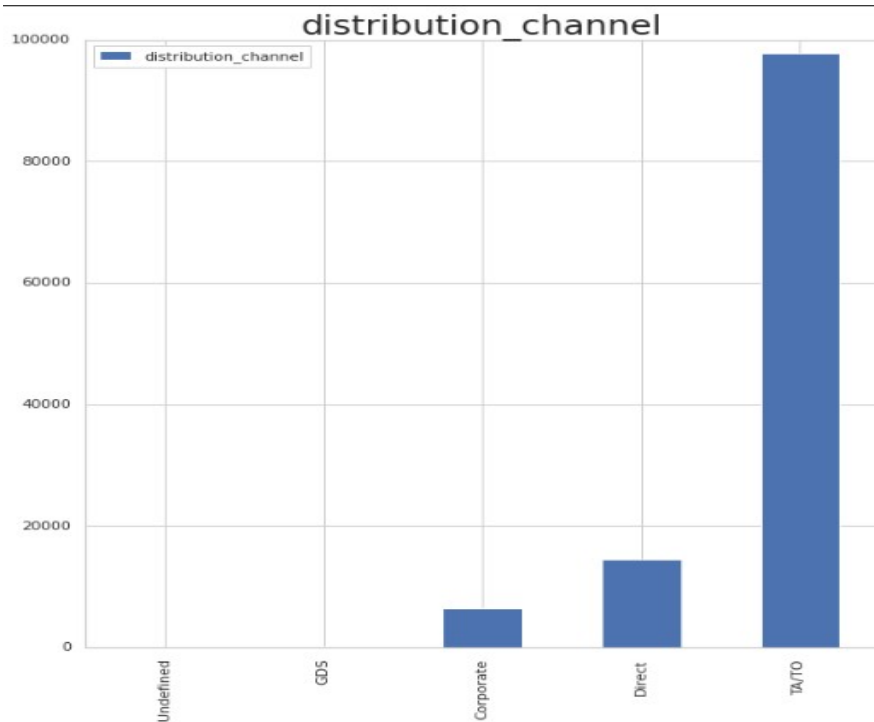
In 2016, city hotels and resort hotels had the highest number of canceled bookings as compare to 2015, 2017

In 2015, city hotels and resort hotels had the lowest number of canceled bookings as compare to 2016, 2017

Distribution channel

- Which Distribution channel had maximum bookings
- Which Distribution channel had minimum bookings





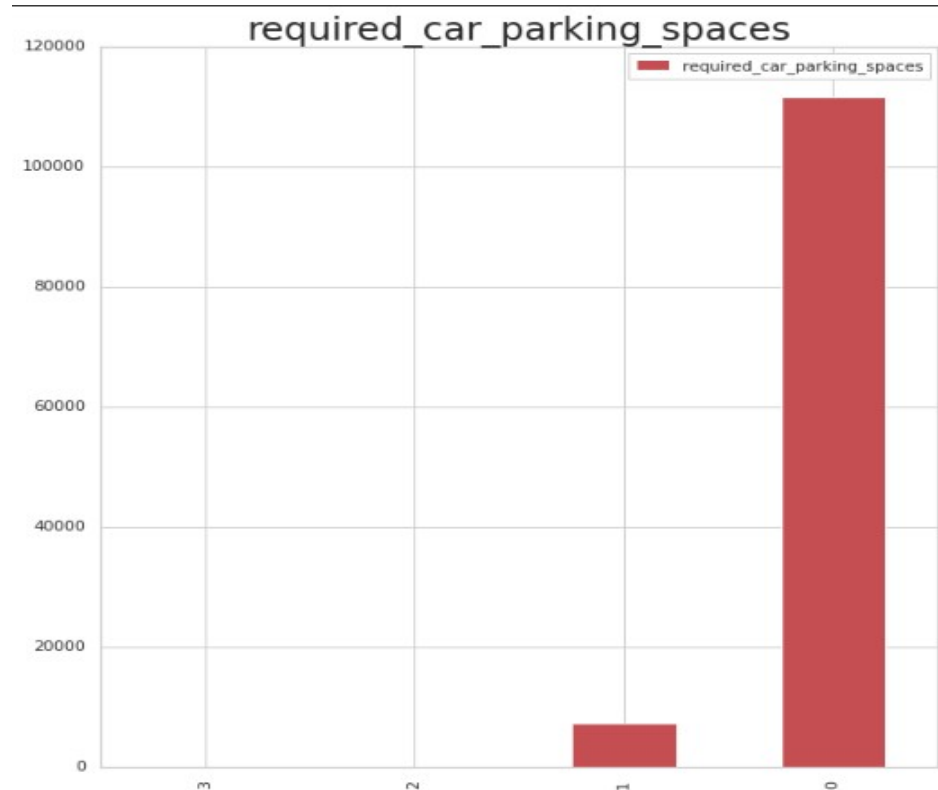
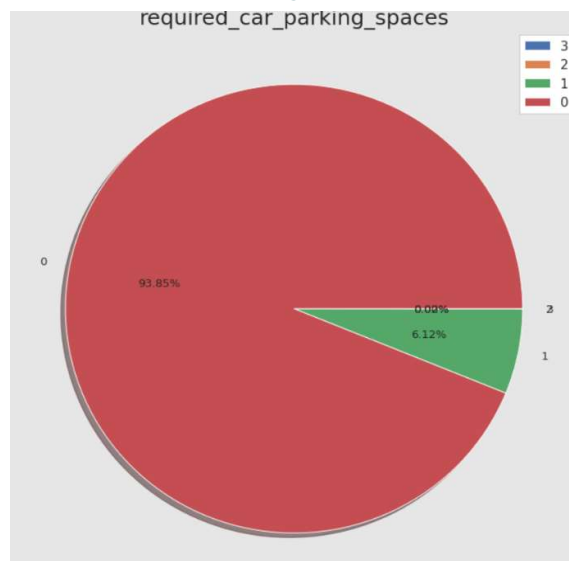
97730 and 82.19% is the highest bookings for TA/TO type Distribution Channel

1 is the lowest bookings for Undefined type Distribution Channel

Second lowest bookings is GDS type Distribution Channel

Booking Of Required Car Parking Spaces

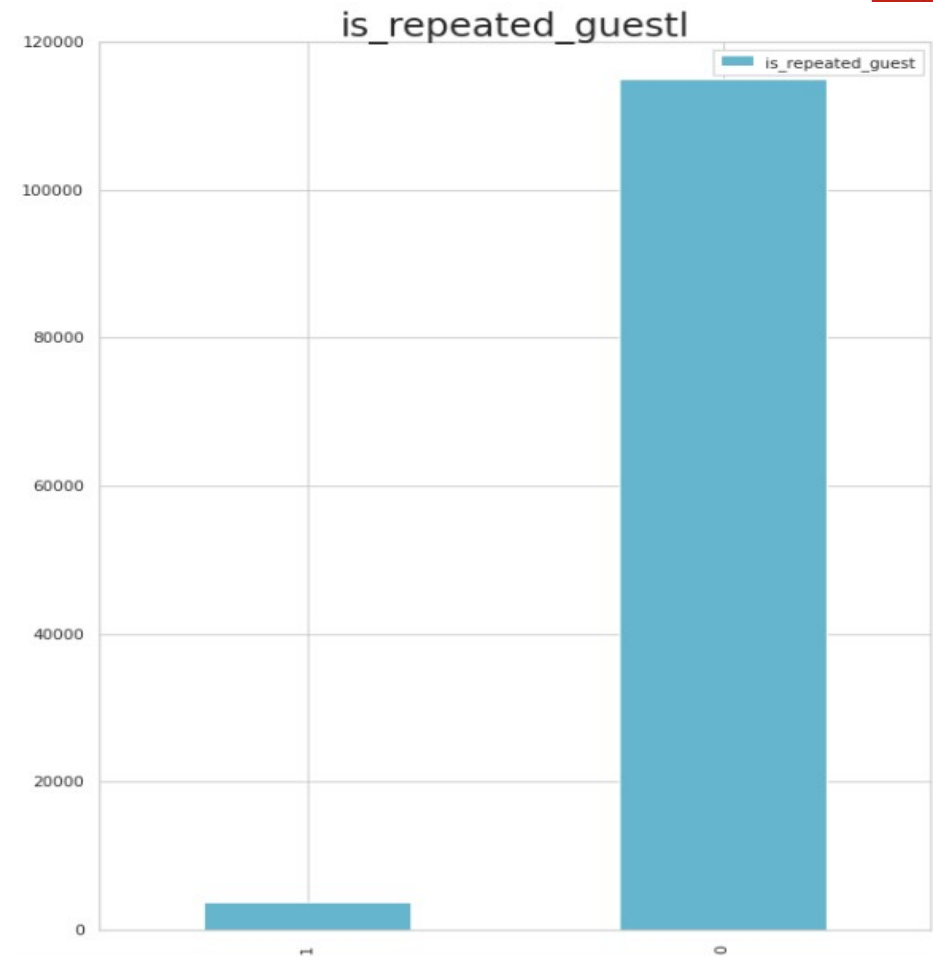
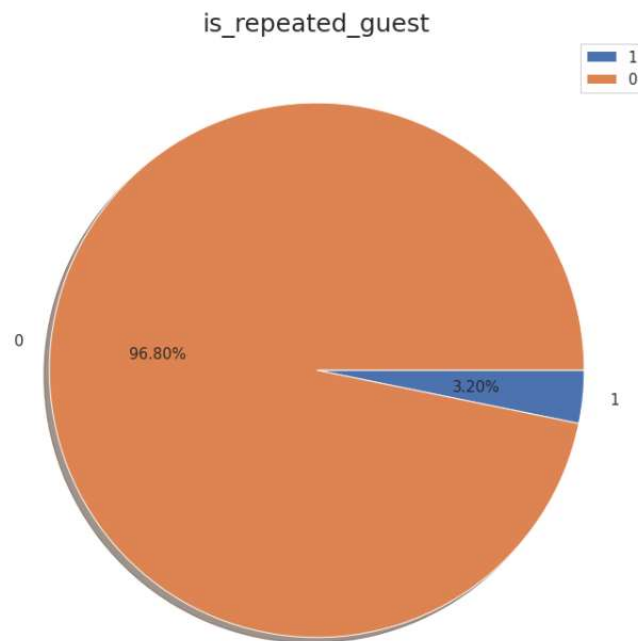
- In this we analysis highest bookings for 0 Required Car Parking Spaces
- 3 and 28 is the lowest bookings for 3 and 2 Required Car Parking Spaces



Bookings Of Repeated Guest

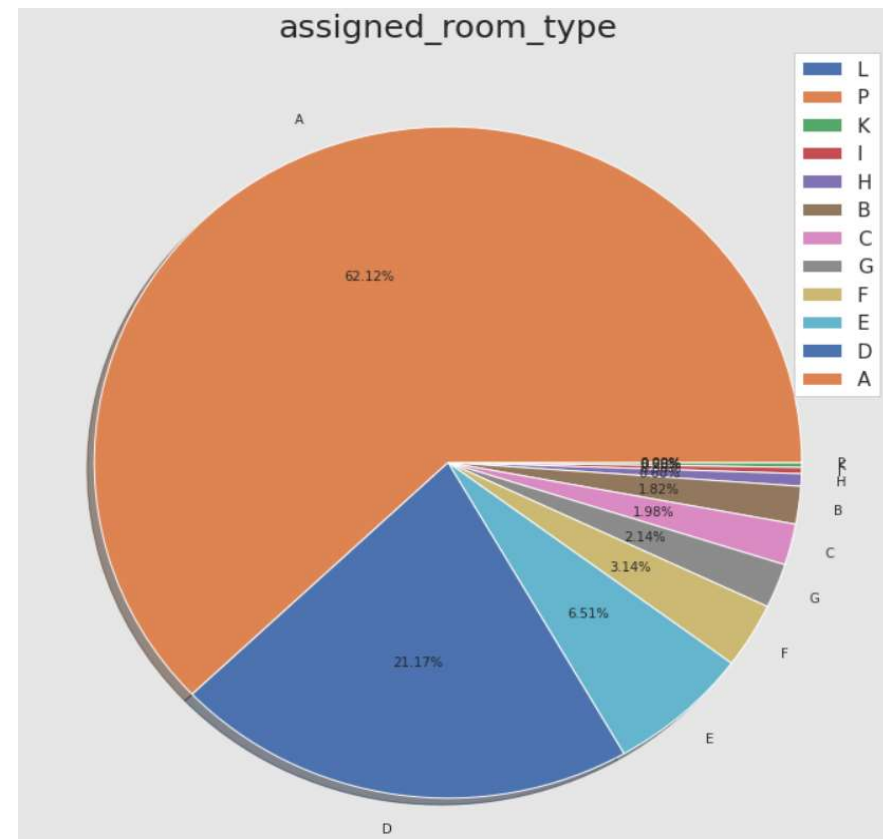
In this we analysis 115092 and 96.80% is the highest bookings for not type Repeated Guest.

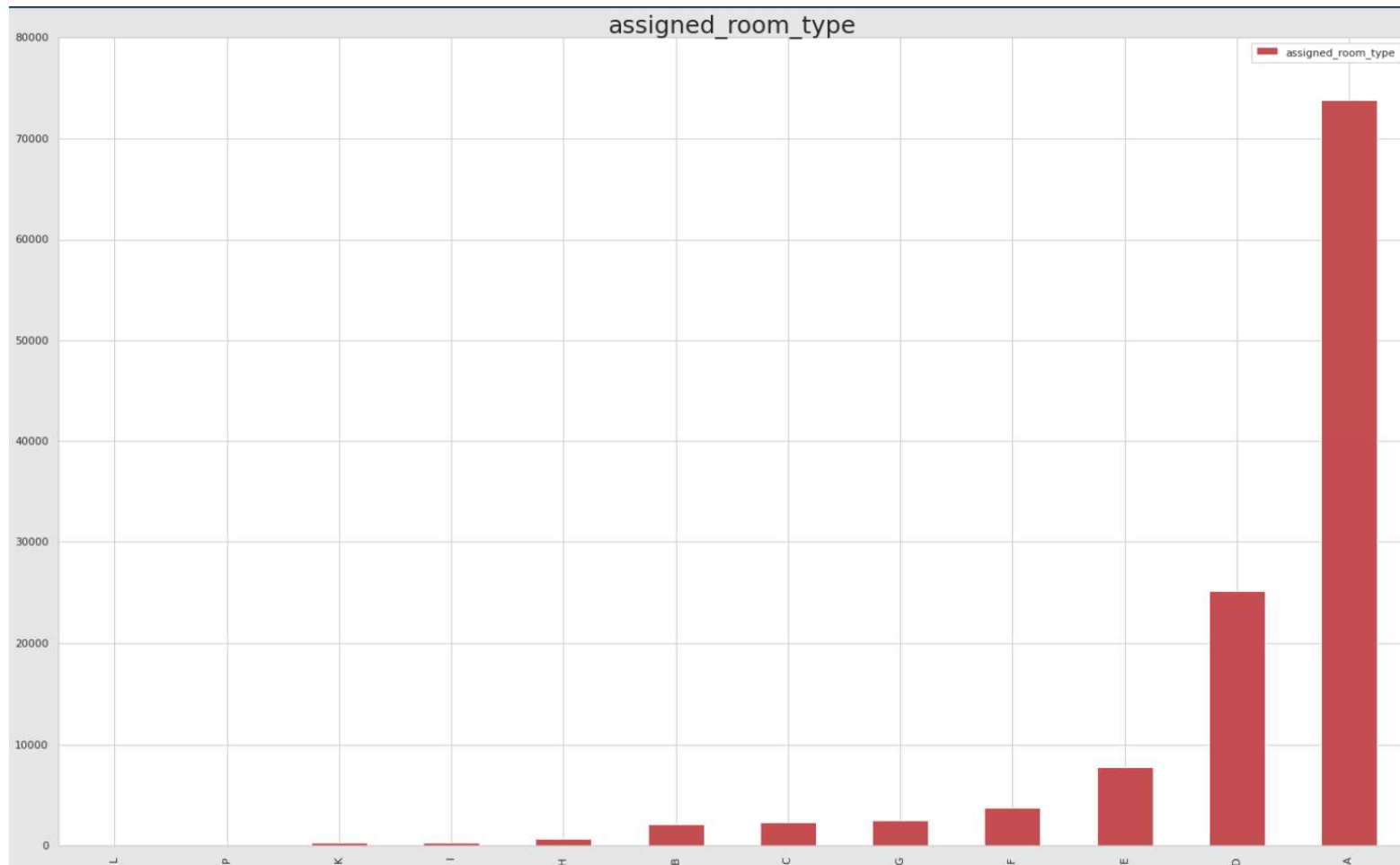
3806 and 3.20% is the lowest bookings for repeated g



Bookings In Assigned Room type

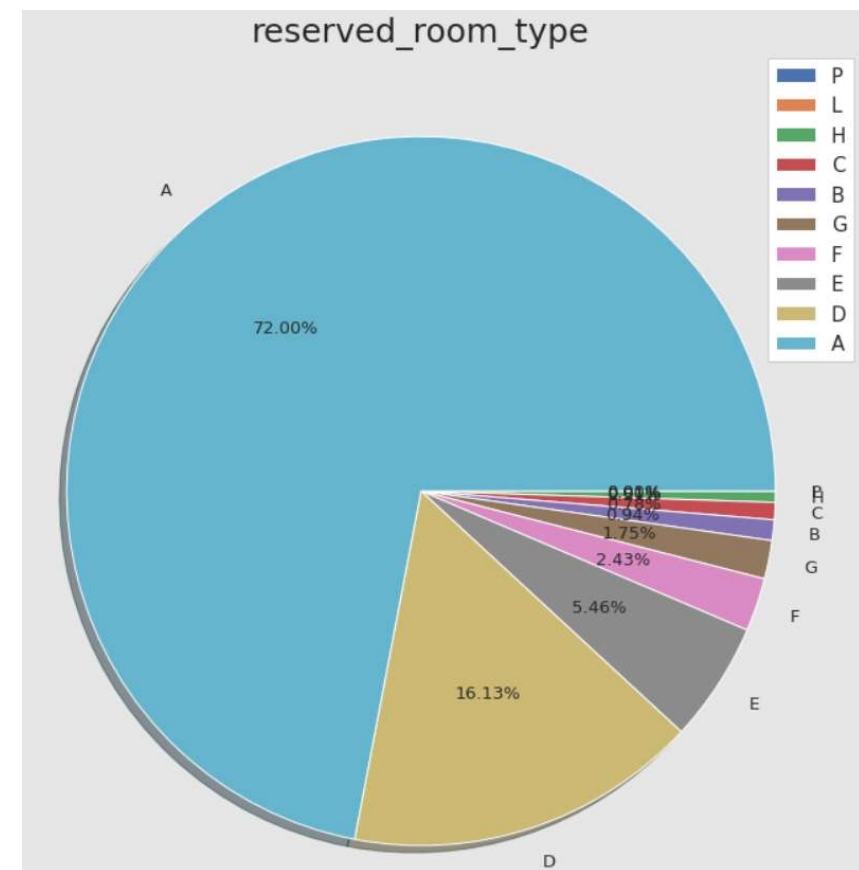
- In this we analysis 73863 and 62.12% is the highest bookings for A type Assigned Room
- 1 and 2 is the lowest bookings for P and L type Assigned Room

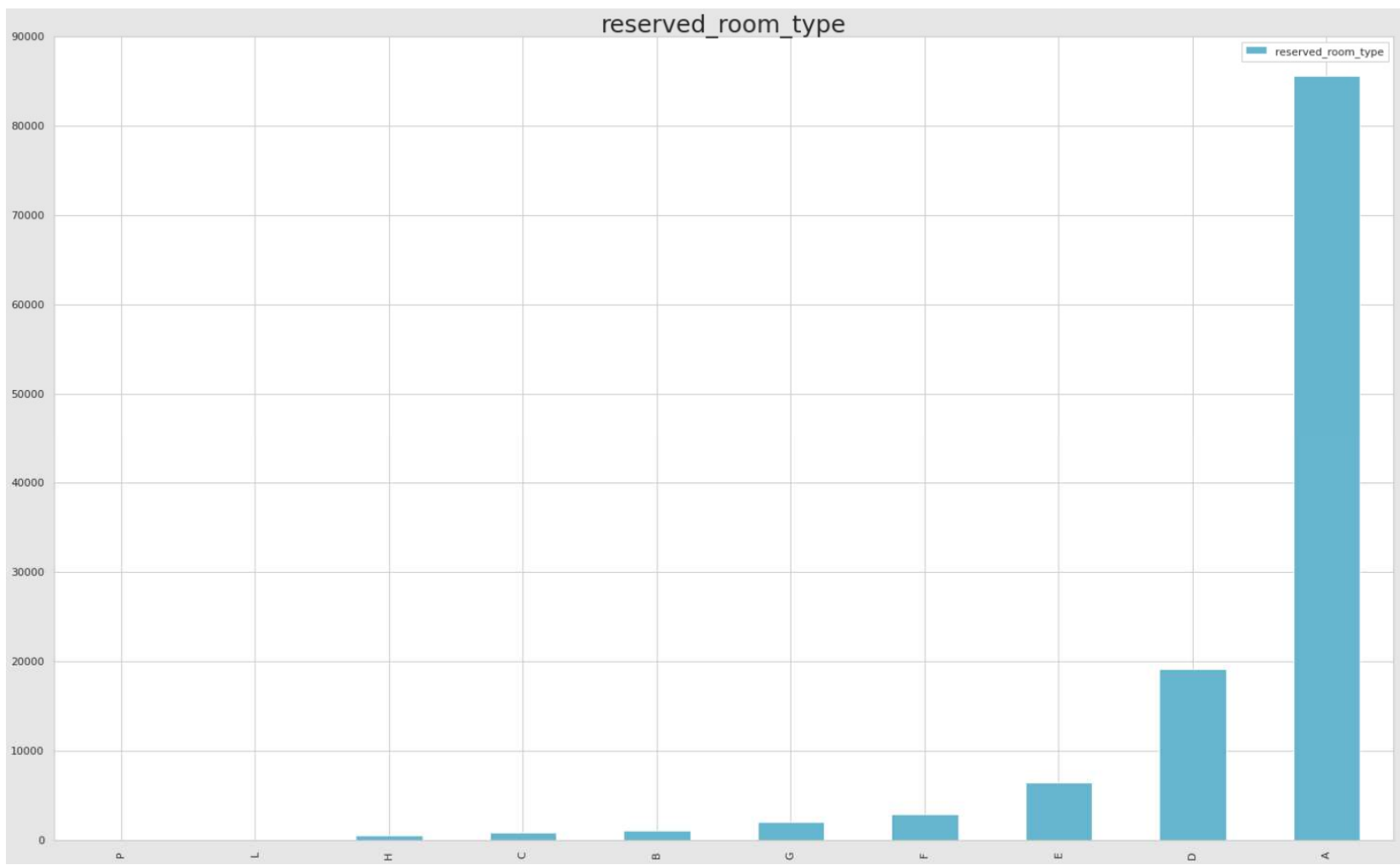




Bookings In Reserved Room type

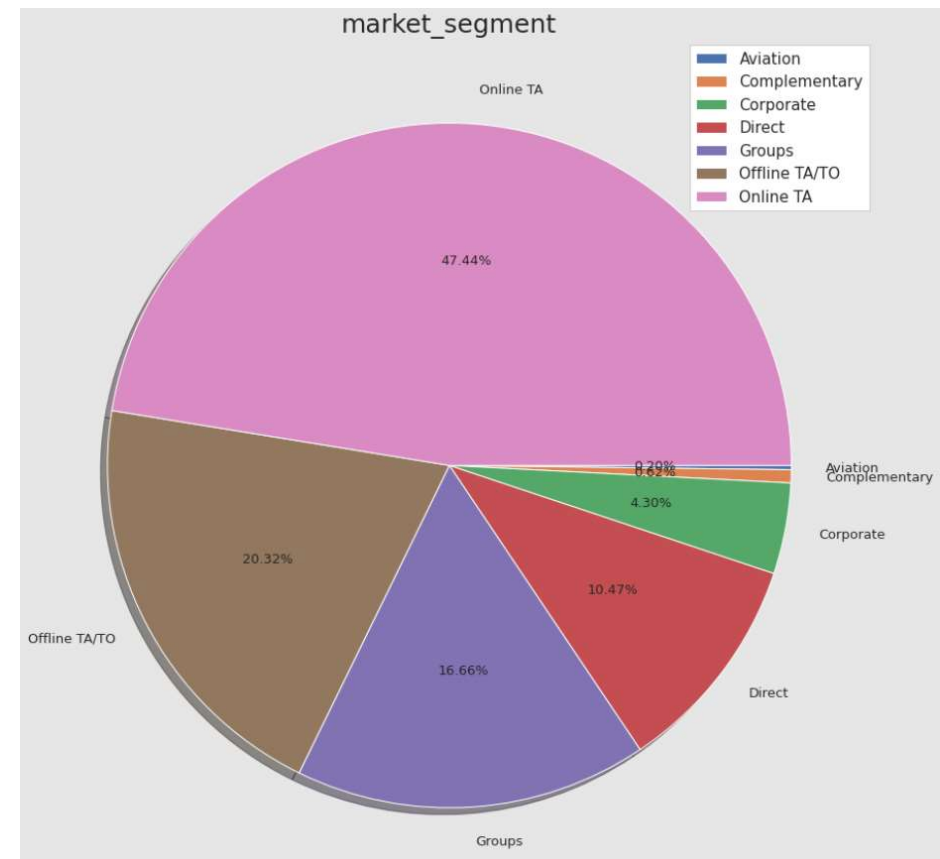
- In this we analysis 85601 and 71.99% is the highest bookings for A type Reserved Room
- 2 and 6 is the lowest bookings for P and L type Reserved Room

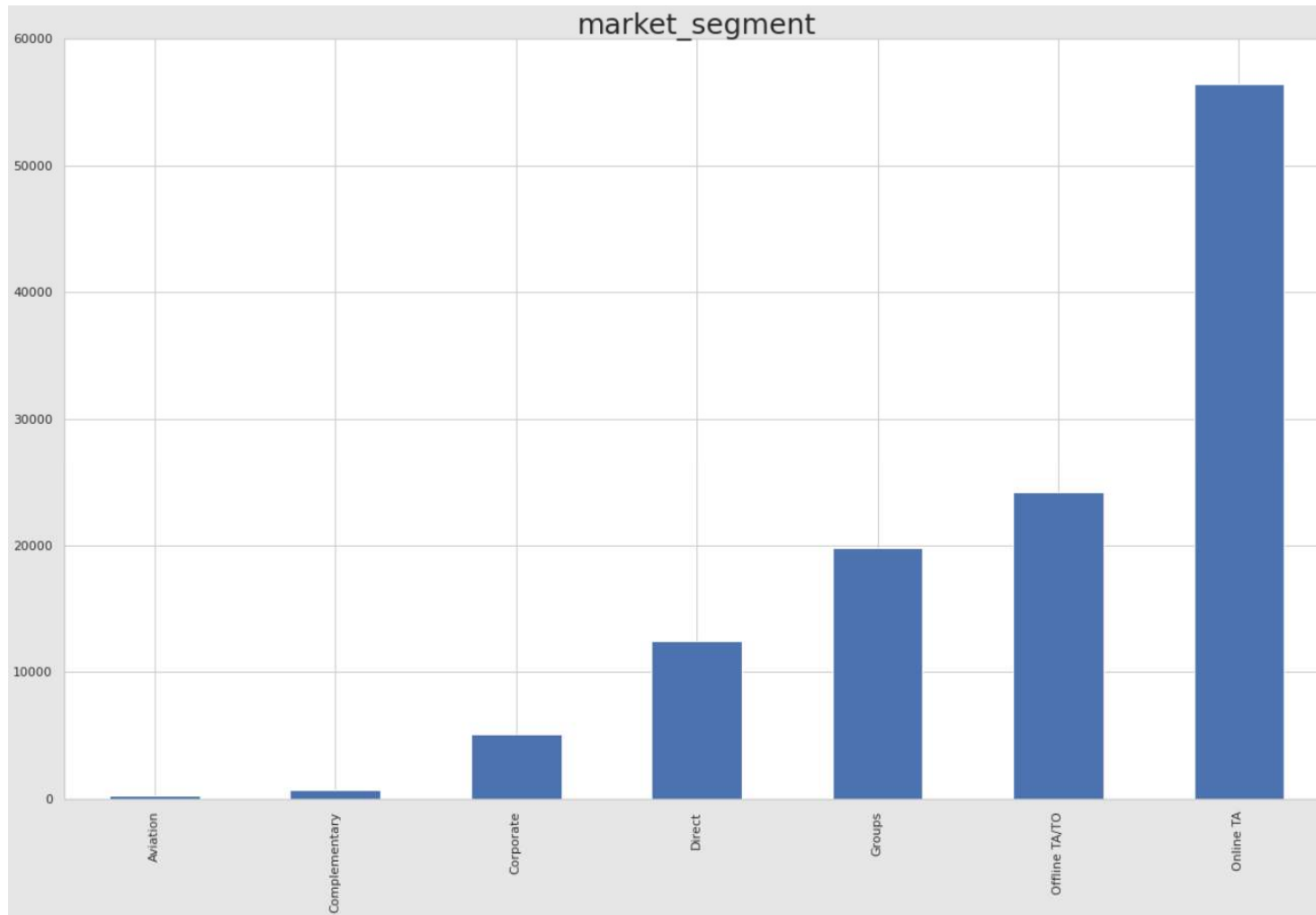




Bookings In the Market Segment

- In this we analysis 56402 and 47.43% is the highest bookings for Online Travel Agents type Market Segment
- 237 and 0.20% is the lowest bookings for aviation type Market Segment

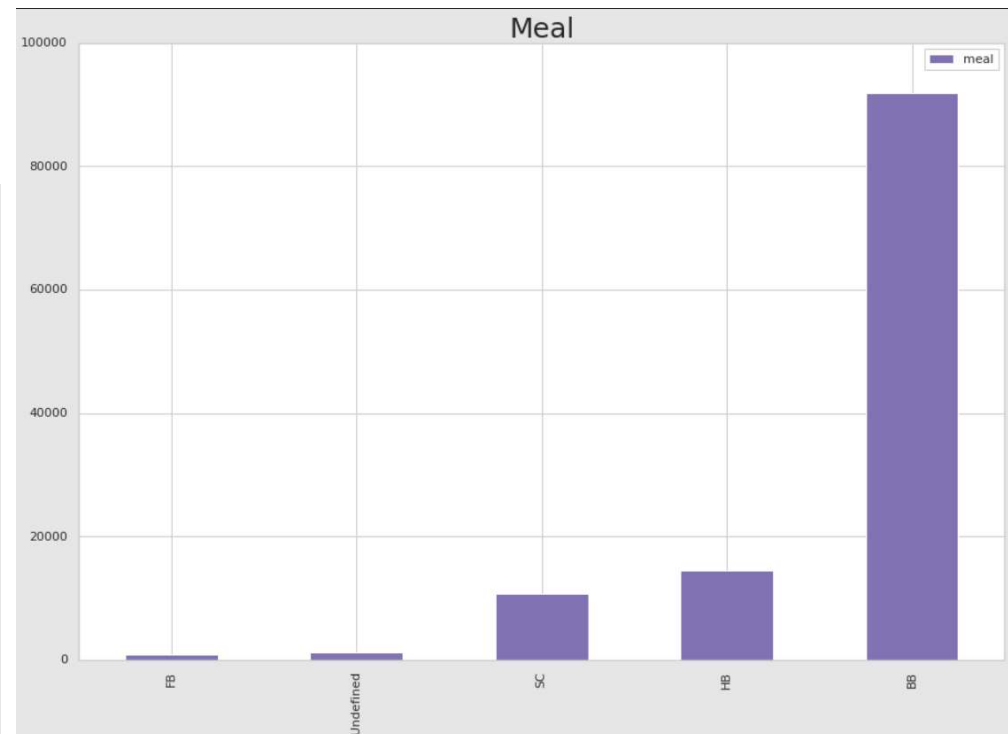
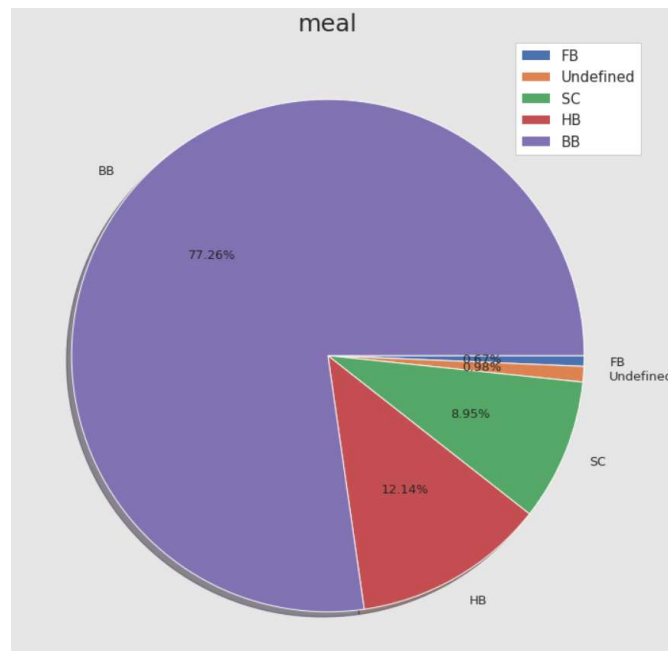




Booking Of the Meal

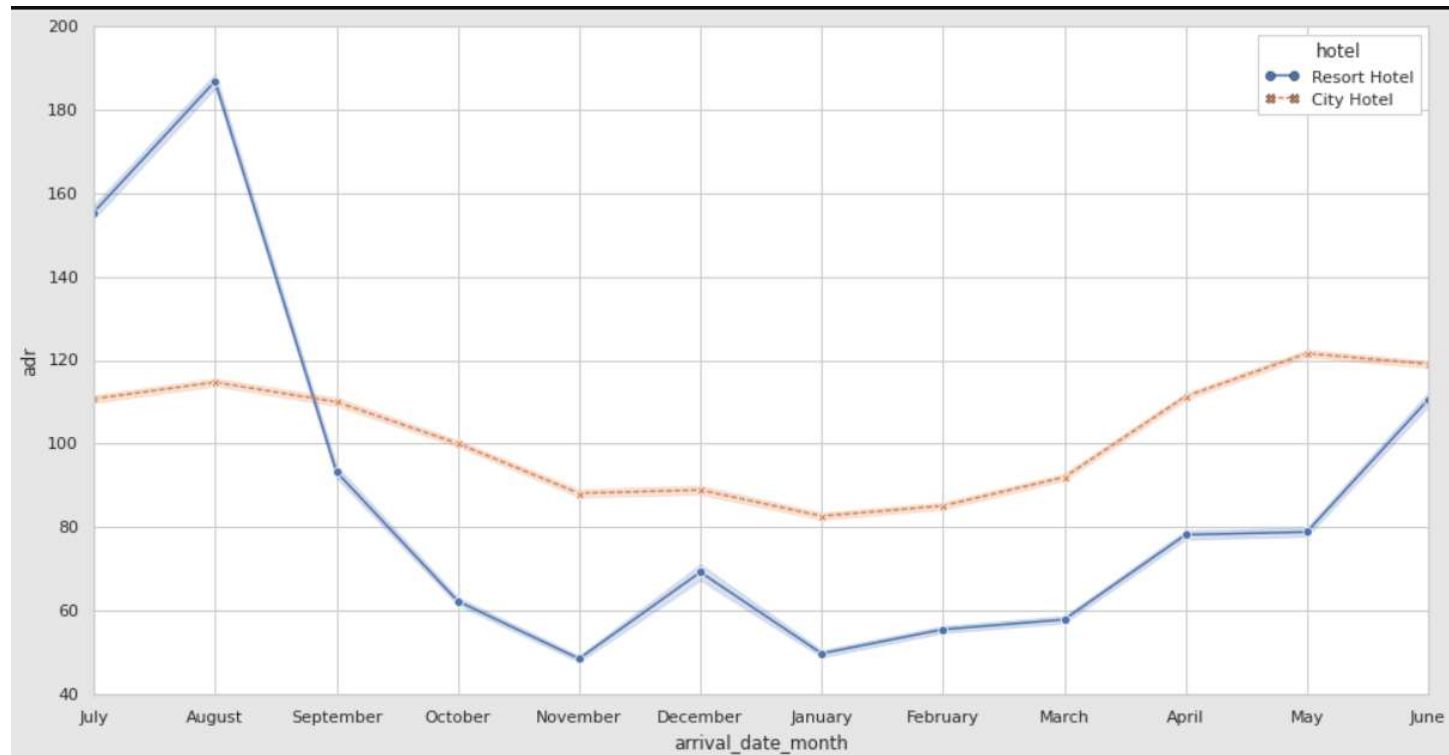
In this we 91863 and 77.26% is the highest bookings for bed & breakfast type meals

798 and 0.67% is the lowest bookings for Full board type meals



ADR : Average Daily Rate

Average daily rates at resort hotels are more expensive in August, July, June and May.



Conclusion

- 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 converts into successful transactions.
- Set Non-refundable Rates, Collect deposits, and implement more rigid cancellation policies.
- Encourage Direct bookings by offering special discounts
- Monitor where the cancellations are coming from such as Market Segment, distribution channels, etc.

References

- pandas.pydata.org
- seaborn.pydata.org
- matplotlib.org
- GeeksforGeeks
- Stackoverflow
- Almbetter
- Youtube
- Github



Thank you