Capstone Project – 1

Exploratory Data Analysis
Hotel Booking Analysis

AI

Team : Systematic Squad

- Kunal Badgujar
- Vivek Tangawade
- Aarti Gade

Points to Discuss :

- Introduction
- Understanding the problem
- Variable / Columns
- Maximum Booking
- Cancellation rate
- market segment
- Deposit policies
- Bookings across different years
- Preferences
  - Meals
  - Car parking
  - Room Type
- Guests
- ADR Across Distribution Channel
- Conclusion
- References

AI

Introduction

Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business.

We can do EDA to predict the future bookings, most engaged months of coming year, additional facilities which can attract more customers and based upon the data, we can raise the revenue

Understanding the Problem

While doing hotel-wise analysis of given hotel booking dataset, we answered following questions:

- Find the maximum booking in each types of hotel?
- Find the cancellation rate of booking of hotels?
- Find the total number of customers across various market segment ?
- Find the bookings across different years ?
- Find the preferences of customers according to meals they preferred, car parking required or not & room types they booked ?
- Find the country from which maximum number of guests are come?

Variable/Columns :

Let us try to analyze the type of data stored in various columns: 1. Columns/Variables having textual values(Categorical data):

i.   hotel: type of hotel ('Resort Hotel', 'City Hotel')

ii.   arrival_date_month : arrival calendar month.

iii. Meal : The values include - RO: Room only - BB: Bed & Breakfast - HB: Half Board (Breakfast and Dinner normally) - FB: Full Board (Breakfast, Lunch and Dinner) - SC: Self-catering

iv.   Country: Names of countries in short form

v.   market_segment : 'Direct', 'Corporate', 'Online TA'(Traveling Agency), 'Offline TA/TO', 'Complementary', 'Groups', 'Undefined', 'Aviation'

vi.   distribution_channel : It shows us the sector through which we got the bookin g like, 'Direct', 'Corporate', 'TA/TO', 'Undefined', 'GDS'.

vii. repeated_guest : guest repaeted or not. 1 for yes, 0 for No.

viii. reserved_room_type : Wing of the room like 'A', 'B", etc. for reserved rooms.

ix.   assigned_room_type: Wing of the room like 'A', 'B", etc. for assigned rooms.

x.   deposit_type : type of deposit like 'No Deposit', 'Refundable', 'Non Refund'.

xi. customer_type : type of the customer like transient, etc.

xii. reservation_status : values are 'Check-Out', 'Canceled', 'No-Show'.

xiii. reservation_status_date : date of reservation in textual format.

All other columns have data is in numerical form i.e. Numerical data. *'adr' stands for average daily rate, which measures the average rental revenue earned for an occupied room per day.

City Hotel and Resort Hotel

- It is Observed that City Hotel has more number of bookings compared to Resort Hotel.

- The obvious assumption for this trend would be because of monetary reasons.

- City hotels would be more affordable than resort hotels due to accessibility, reliability and lavishing factors.


maximum hotel bookings
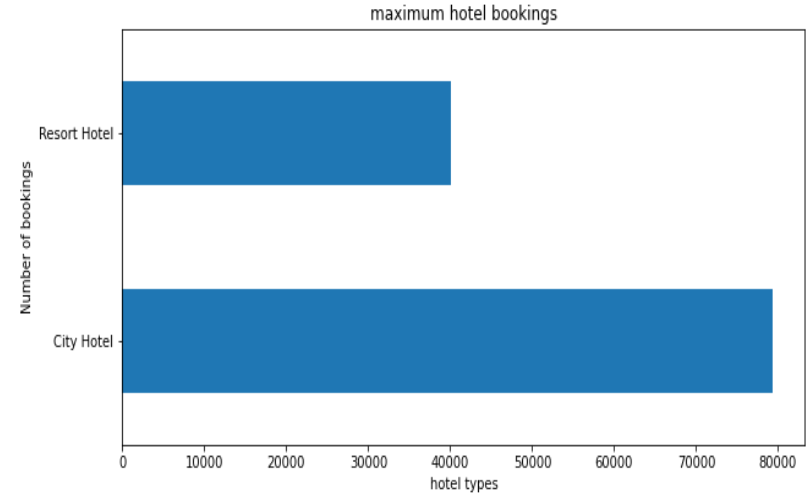
Cancelled Bookings

- It is Observed that out of 119390 bookings made, 75166 of them have checked in and 44224 of them has cancelled their bookings.

- In city hotels 33102 bookings were cancelled from 79330 bookings and in resort hotels 11122 bookings were cancels from 40063 bookings

- High cancellation rate will adversely effect the hotel industry which reduces the yield of profit.
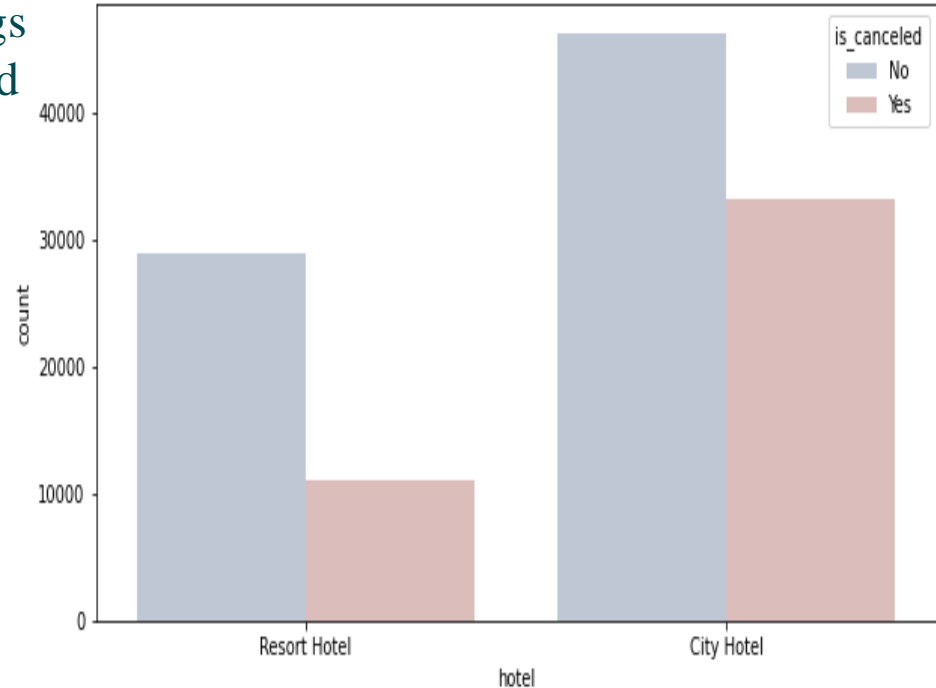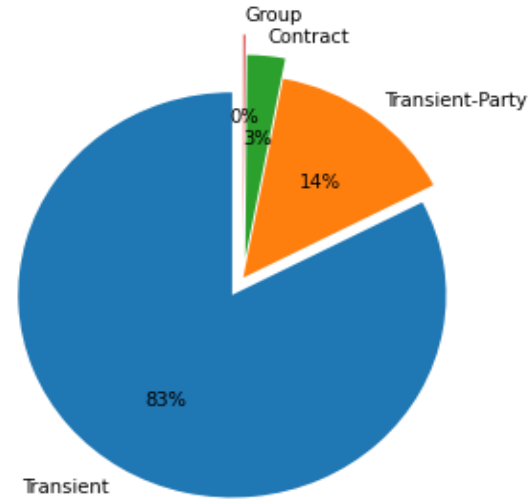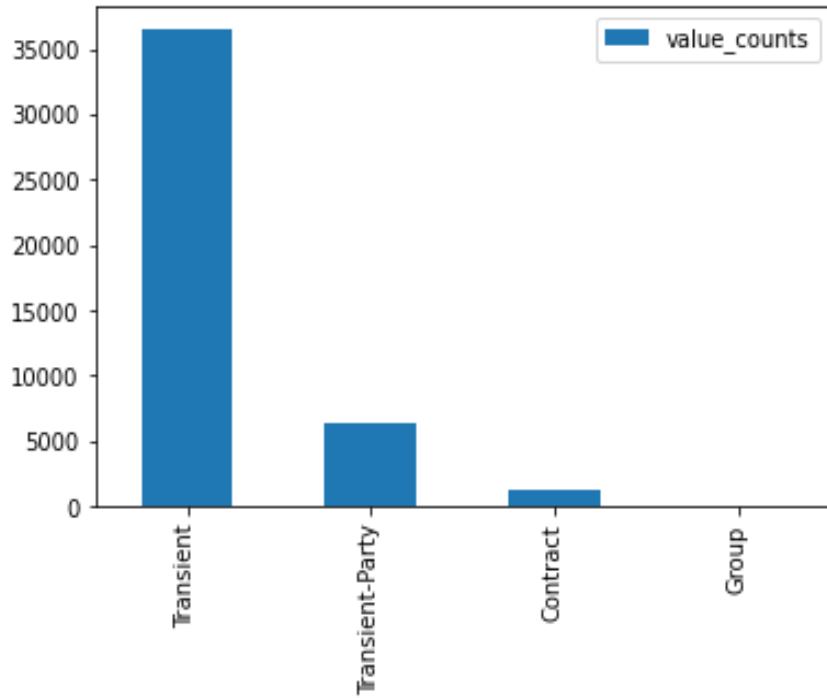
It is observed that Transient customers are more likely to cancel the booking.
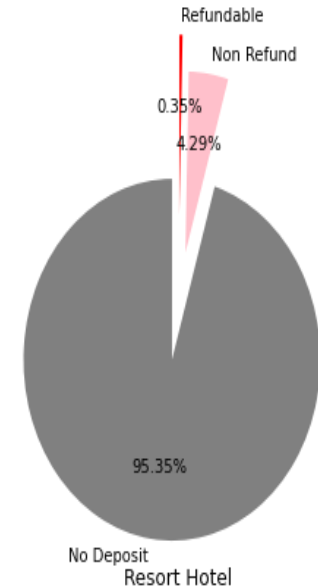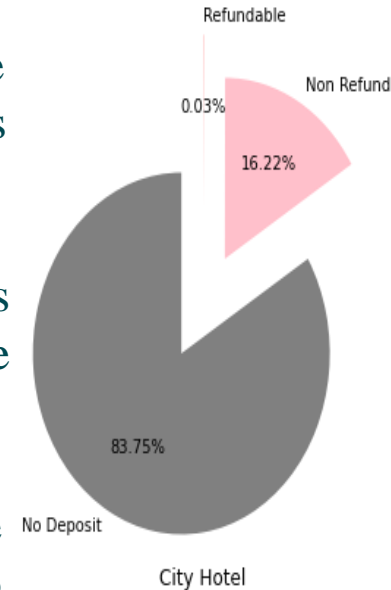
Total Number of Customers across Various Market Segments

- Online Travel agency segment gives the high amount of leads for the hotel booking than that of any other sources of Marketsegments.

- We can report that we need to target our marketing area on online TA websites or apps and focus majorly on online TA.

- The following majority market segments are offline travel agencies, groups and direct customers.
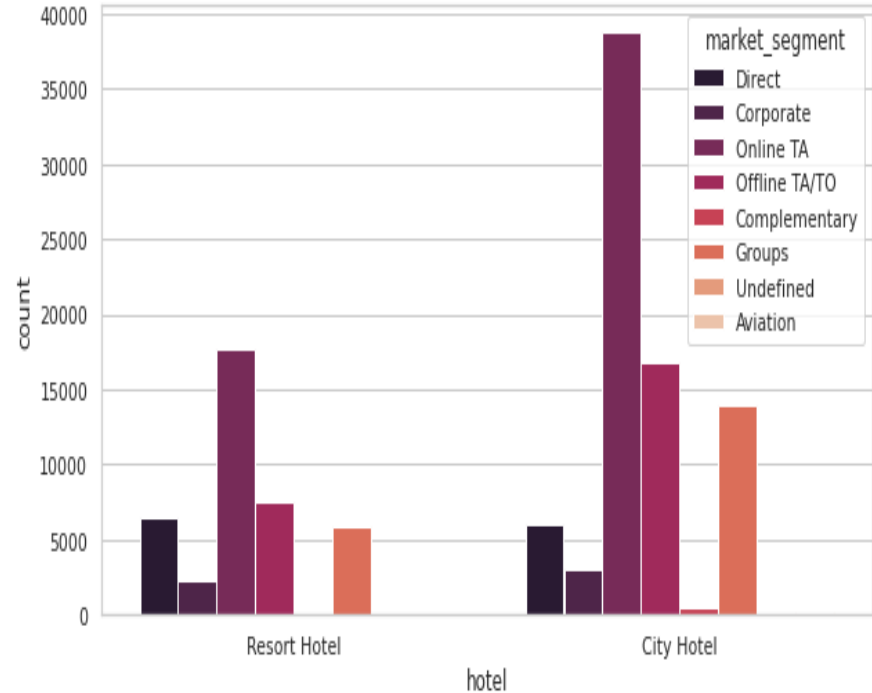
Deposit Policies of Hotel

- Most number of the hotel bookings are made in 'No Deposit, type of transaction which is the reason for the high cancelation rate.

- 'No Deposit' type of transaction means without paying any advance money for the booking.

- In order to reduce the cancelation rate, we need to find suitable alternative for the 'No Deposit' type of transaction.

City Hotel: Refundable 0.03%, Non Refund 16.22%, No Deposit 83.75%

Resort Hotel: Refundable 0.35%, Non Refund 4.29%, No Deposit 95.35%

# Number of Bookings Across Different Years

- Most Number of the bookings are done in the year 2016 following 2017 and 2015 for the dataset given comprising the data for these 3 years.

- When, sub plotted, we can see that City hotel has high demand rather than Resort hotels due to its reliability and price difference compared to resort hotels.
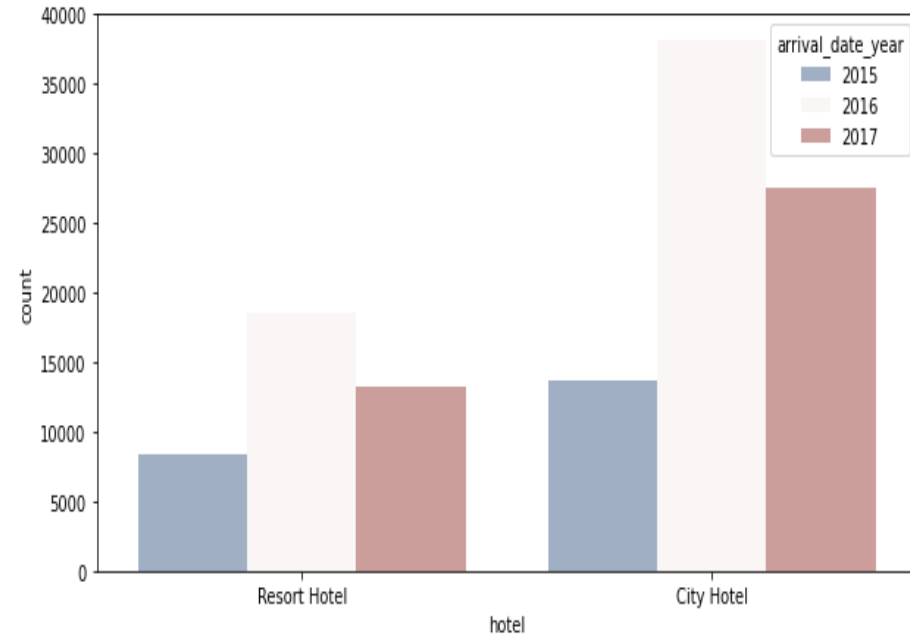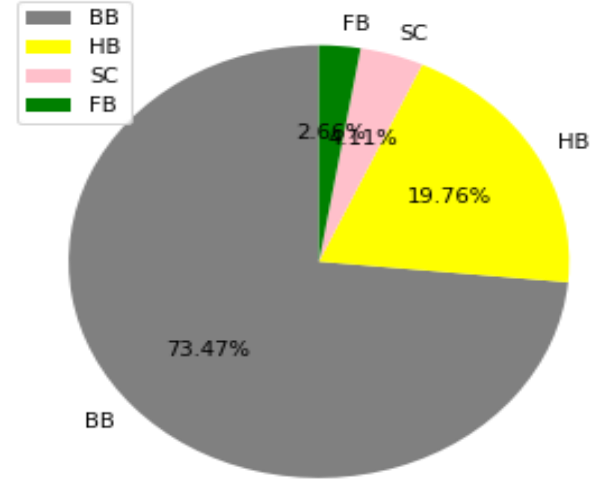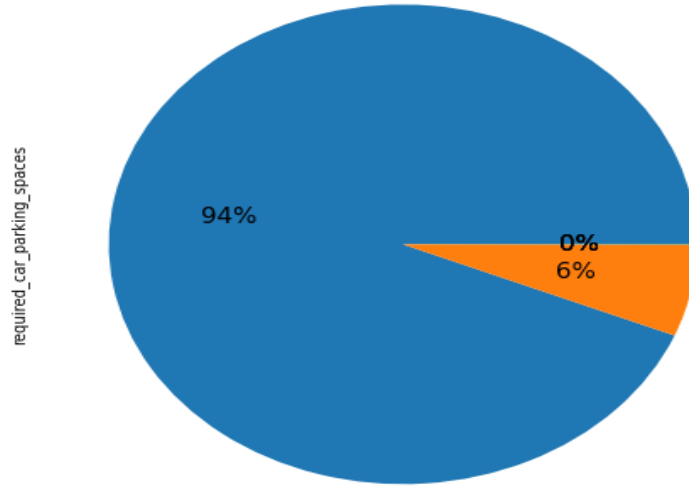
Meals preference
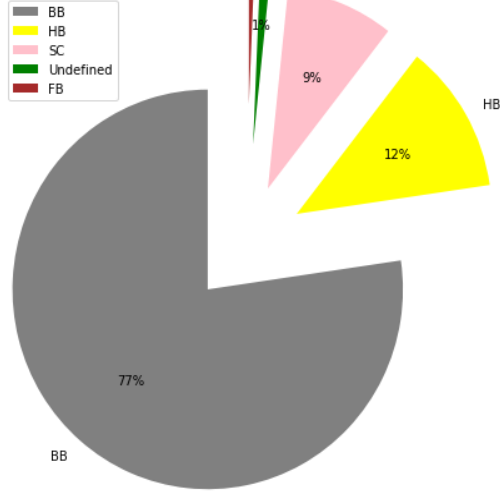
Guests fav.
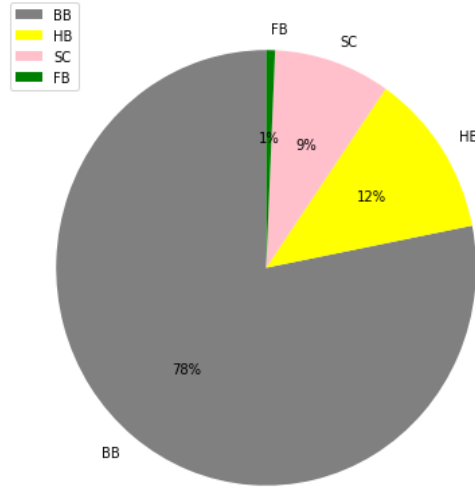
adults fav.

Babies fav.

As per the above Graph, we can conclude that BB (i.e. Bed & Breakfast) is more than others i.e. Half Board(HB), Full Board(FB), Self Catering(SC)
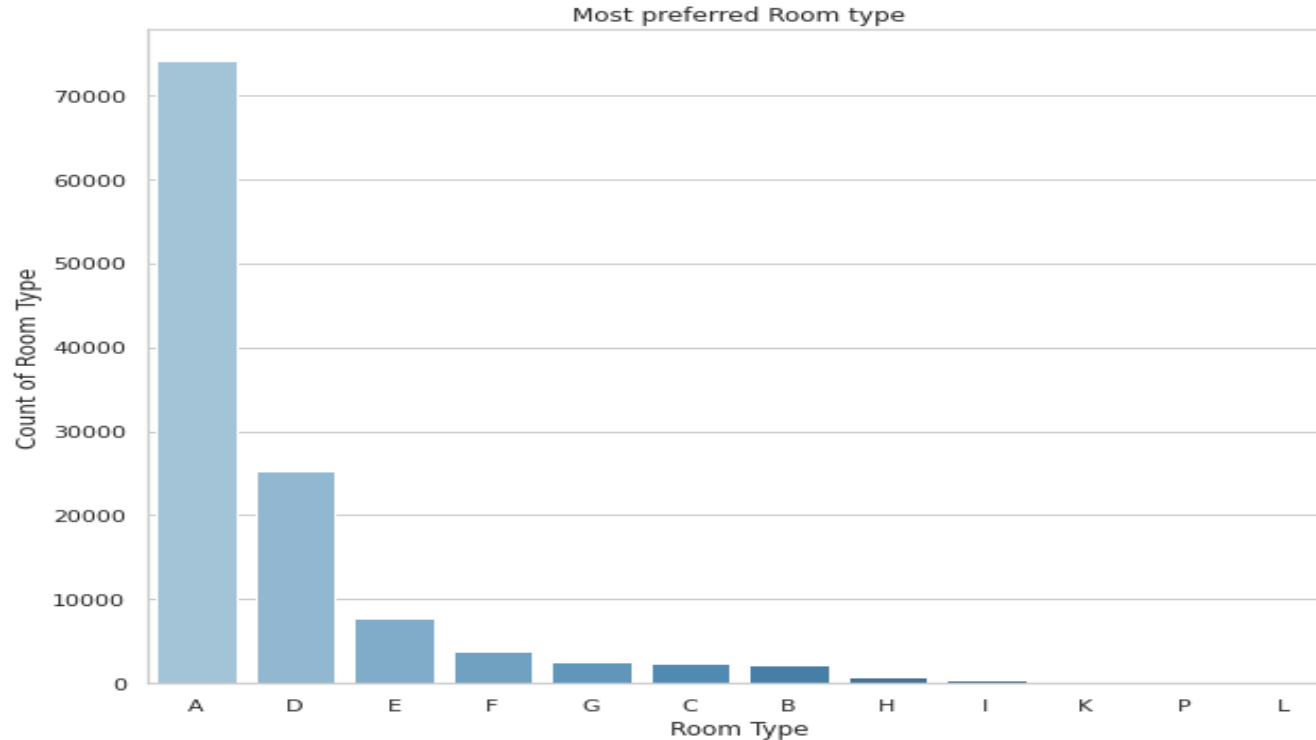
Car Parking Space Analysis

% Distribution of required car parking spaces

- Most of the hotels have zero car space parking.

- Among which have car space parking has one car spacing and the trend decreases drastically as number of car spacing increases.

Preferred room type by guests :

Maximum number of guest Preferred 'A' type rooms.

Top 10 Countries with highest number of Bookings

Number of guests from diffrent Countries
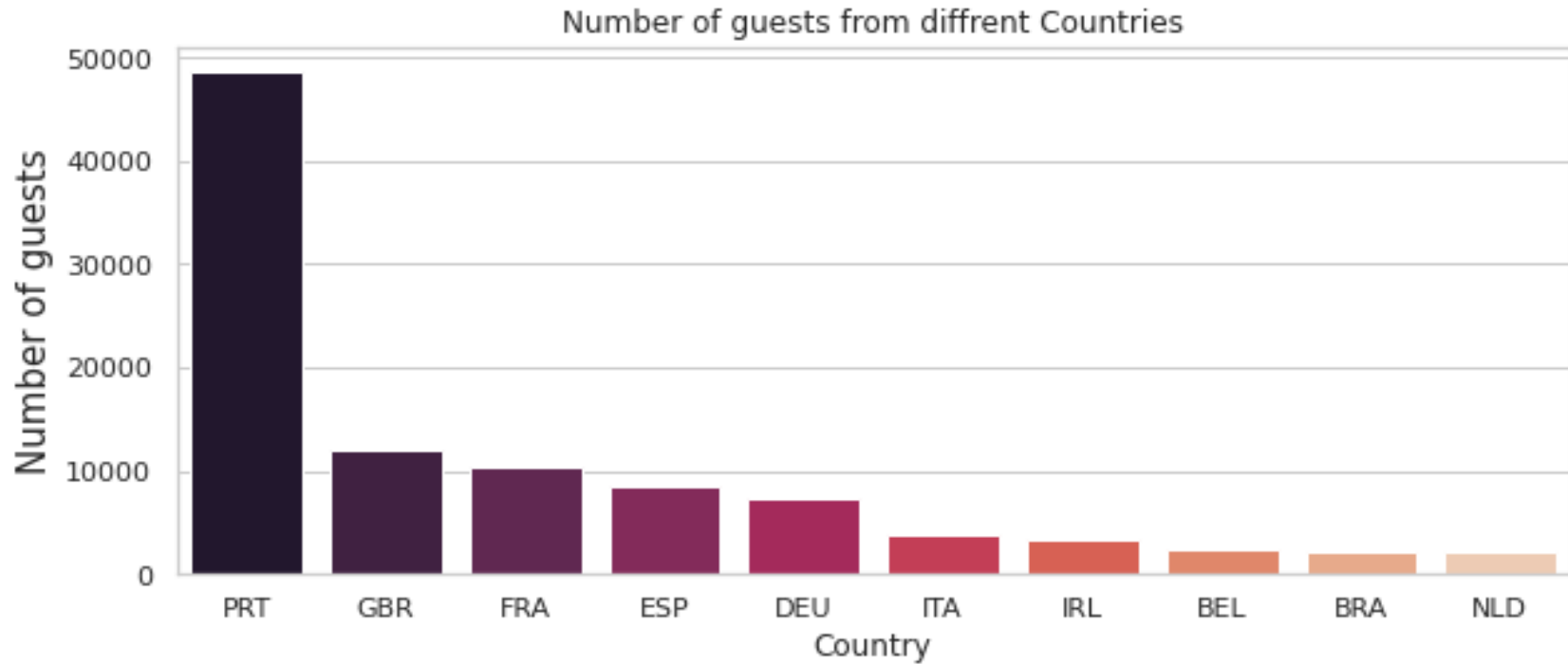
- After analyzing the dataset, we found that Portugal tops the position with 48590 customers followed by UK with 12129, France with 10415, Spain with 8568 and Germany with 7287 customers.

- Netherlands sits back with least number of customers among the top 10 Countries.

ADR Across Distribution Channel

ADR across Distribution channel
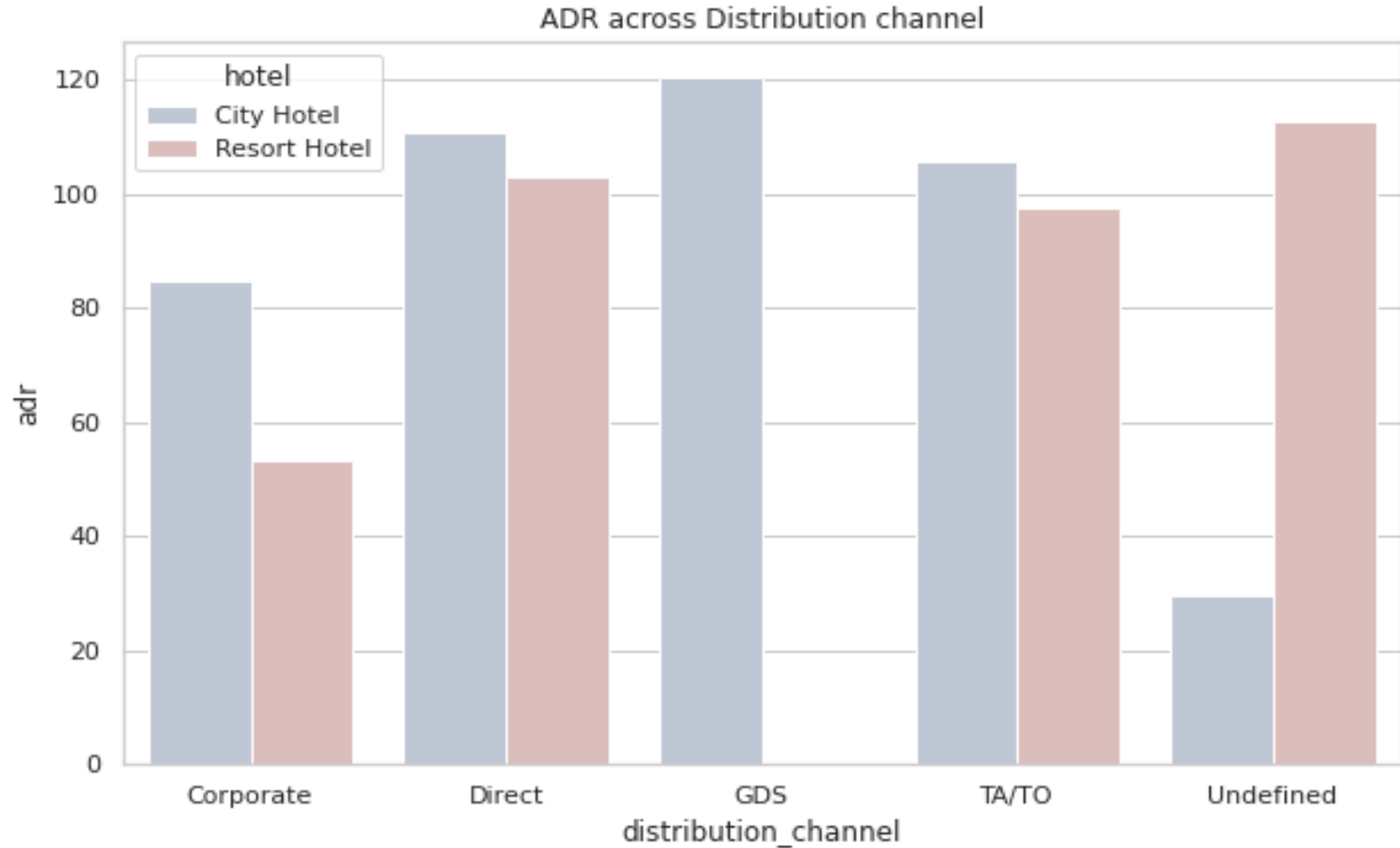
Conclusion

- 'City hotels' and 'Resort hotels' are two types of hotels present in the dataset, out of which, 'City hotels' are more preferred by the customers than the latter. (66.4% customer prefers 'City hotels' whereas 33.6% customerprefers 'Resort hotels')

- Out of 119000 customer dataset, 75166 customers checked in hotel while 44224 customers cancelled their bookings, that is about 37% of the booking got cancelled whereas 63% customers checked in the hotel.

- Majority of the deposit type is of **'No deposit'** type, which itself concludes the high rate of cancellation rate.

- From the given dataset, we can see 2016 is the year in which hotel bookings are the highest.

- City hotels are of in high demand compared to resort in all aspects due to its reliability of majority of the population.

- Online Travel agencies provides high amount of leads and customers following Offline TA, Groups, Direct, etc., So, we an conclude that we need to focus our marketing area on online TA since majority of the visitors tend to reach out to them

- After analyzing the meal data, its found that 77.3% of customers prefers BB(Bed & Breakfast).

- Among the top 10 countries with respect to number of bookings, most of them are European countries which belongs to the list of which Portugal stands first with 48.5k bookings.

References

1) https://pandas.pydata.org/

2) https://matplotlib.org/

3) https://seaborn.pydata.org/

4) Geek for geeks

5) Almabetter Study Materials

AI

Thank You