Capstone Project

Hotel Booking Analysis

Jouher Lais Khan
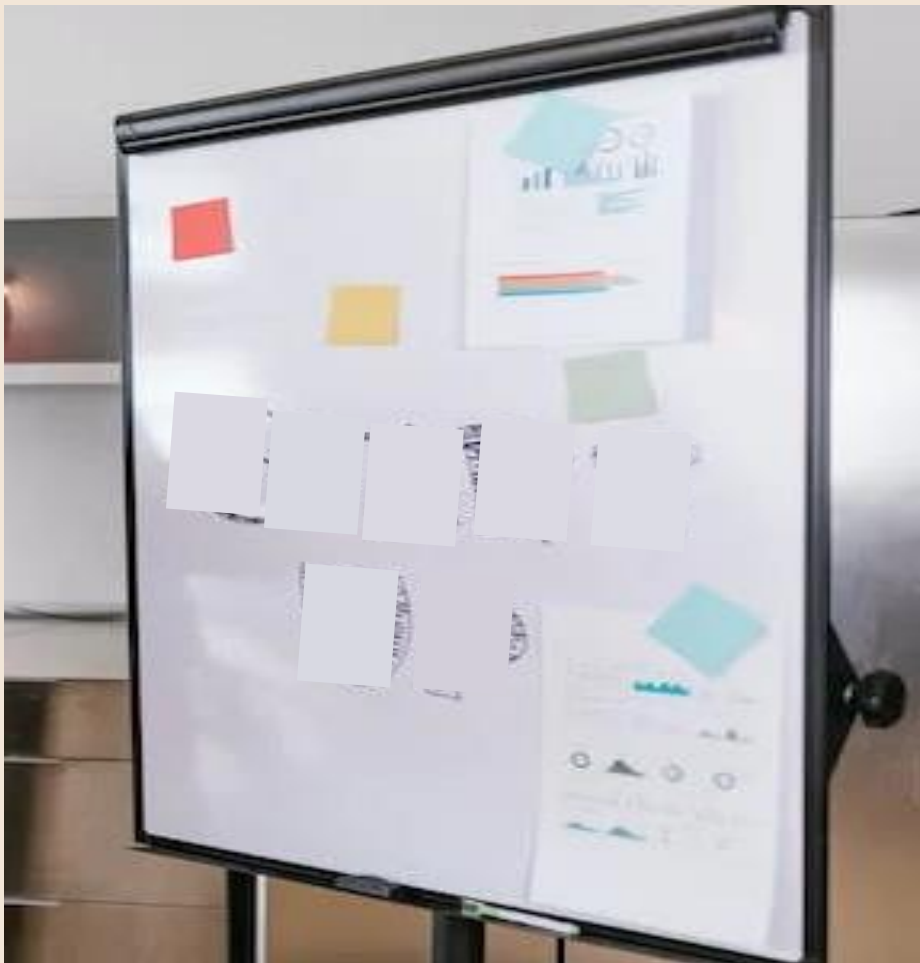
Hotel Booking
Analysis

Index
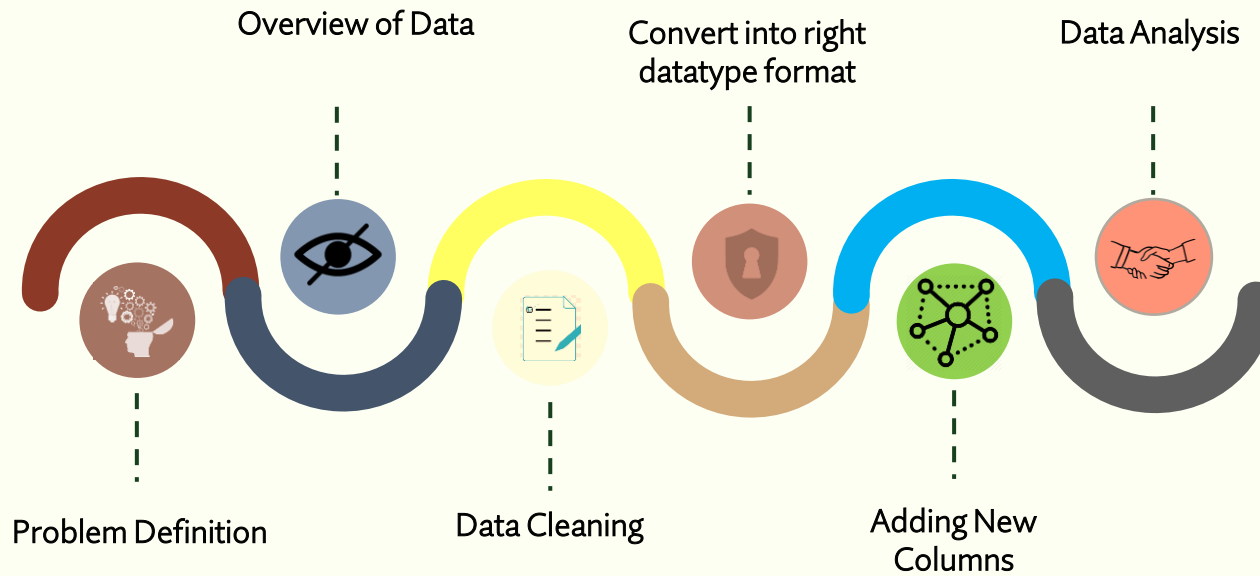
Purpose

➢ To discuss the analysis of given hotel bookings data set from 2015-2017
➢ The process of Data Analysis start's with following steps :

Overview of Data

Convert into right
datatype format

Data Analysis

Problem Definition

Data Cleaning

Adding New
Columns

Purpose

The Hotel Data involve following types of Analysis:

- **Univariate analysis**

- **Hotel wise analysis**

- **Marketing Channel wise analysis**

- **Booking cancellation analysis**

- **Timewise analysis**

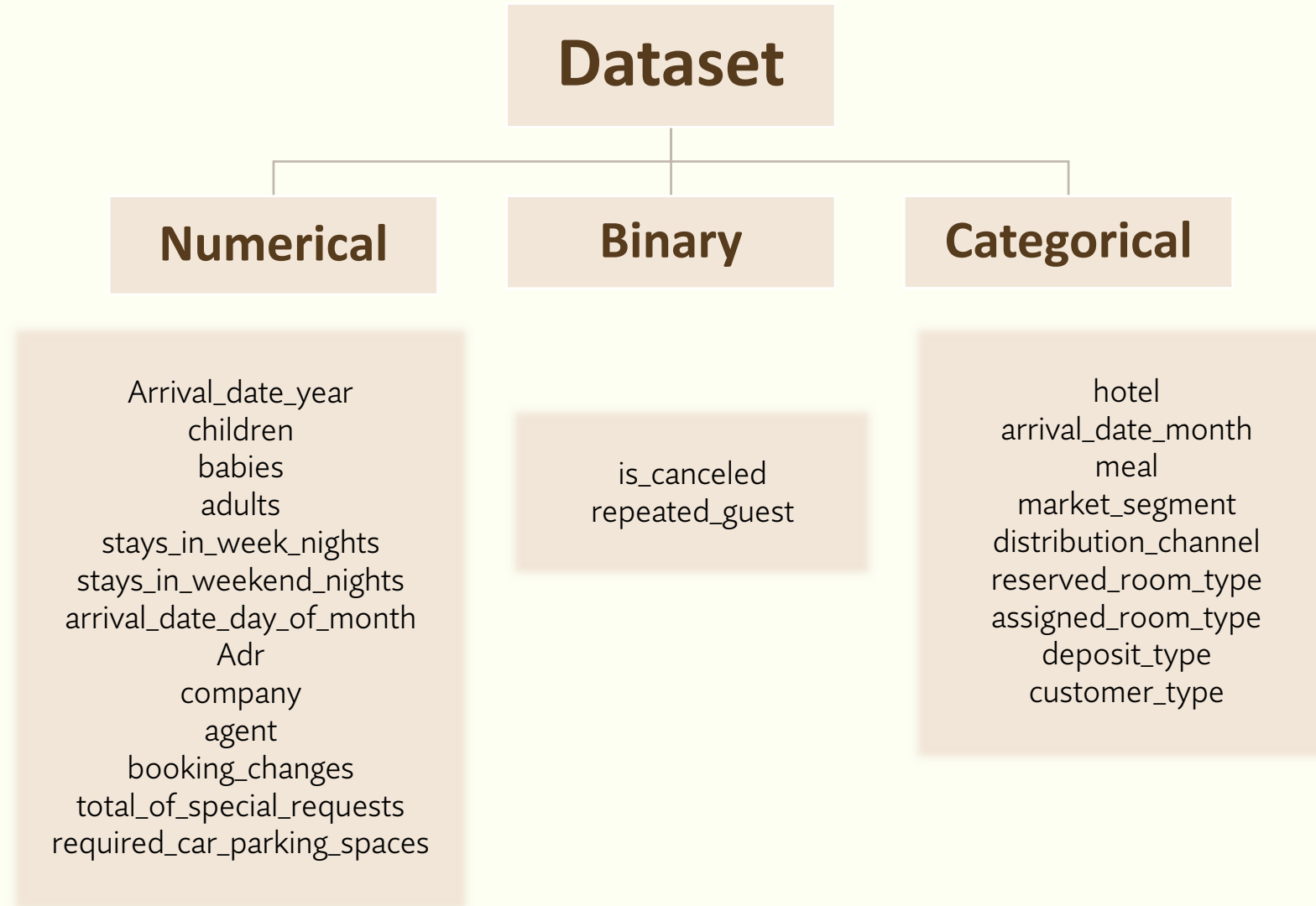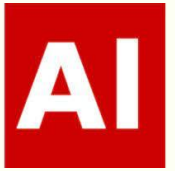By doing this analysis we will try to find trends in the given data

Data Summary

Our dataset has 119390 rows and 32 columns. The given dataset has some important columns with there variables . Some of them are:

- hotel:
  The hotel category, 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.

- stayed_in_weekend_nights:
  The number of weekend nights stay per reservation

- stayed_in_weekday_nights:
  The number of weekday nights stay per reservation.

- meal:
  Meal preferences per reservation.[BB , FB , HB , SC , Undefined]

- Country:
  The origin country of guest

Data Summary(contd..)

- 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.

- distribution_channel:
  The medium through booking was made.
  [Direct,Corporate,TA/TO,undefined,GDS.]

- Is_repeated_guest:
  shows if the guest has arrived previously or not.
  {0 means not arrived and 1 means has arrived}.

- days_in_waiting_list:
  Number of days between actual booking and payment.

- customer_type:
  Type of customers(Transient, group, etc.)

Data Summary

Dataset

Numerical

Arrival_date_year
children
babies
adults
stays_in_week_nights
stays_in_weekend_nights
arrival_date_day_of_month
Adr
company
agent
booking_changes
total_of_special_requests
required_car_parking_spaces

Binary

is_canceled
repeated_guest

Categorical

hotel
arrival_date_month
meal
market_segment
distribution_channel
reserved_room_type
assigned_room_type
deposit_type
customer_type

Cleaning the Data

Data cleaning is a crucial component of data analysis since unclean data, such as duplicates and Nan/Null Values, render our visualizations useless. The steps involved in Data Cleaning are:

1.Remove duplicate rows.(31994, 32)

2.Taking care of missing values. Changing the nan values to others or
in some case dropping the row. (company 82137, agent 12193, country 452, children 4)

3.Change the datatypes of the columns. (children , agent to int64 and reservation_status_date to datetime)

4. Adding important columns . (number of people , total stay , arrival date and arrival day)

Univariate Analysis

Q.1 Which agent has booked most hotel and generate more revenue ?

Q.2 Which meal type is mostly preferred ?
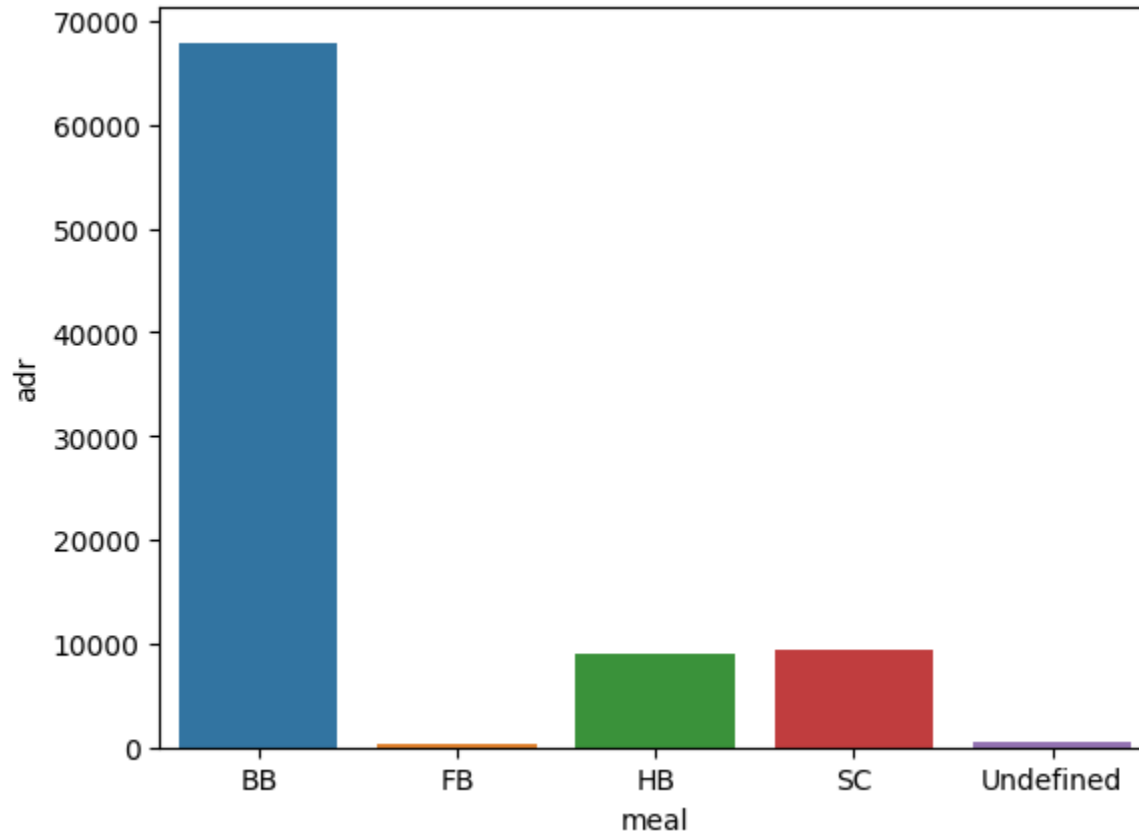
Q.3 What are the most visited country?

Q.4 Which type of Hotel is mostly booked ?

**Q.1 Which agent has booked most hotel and generate more revenue ?**
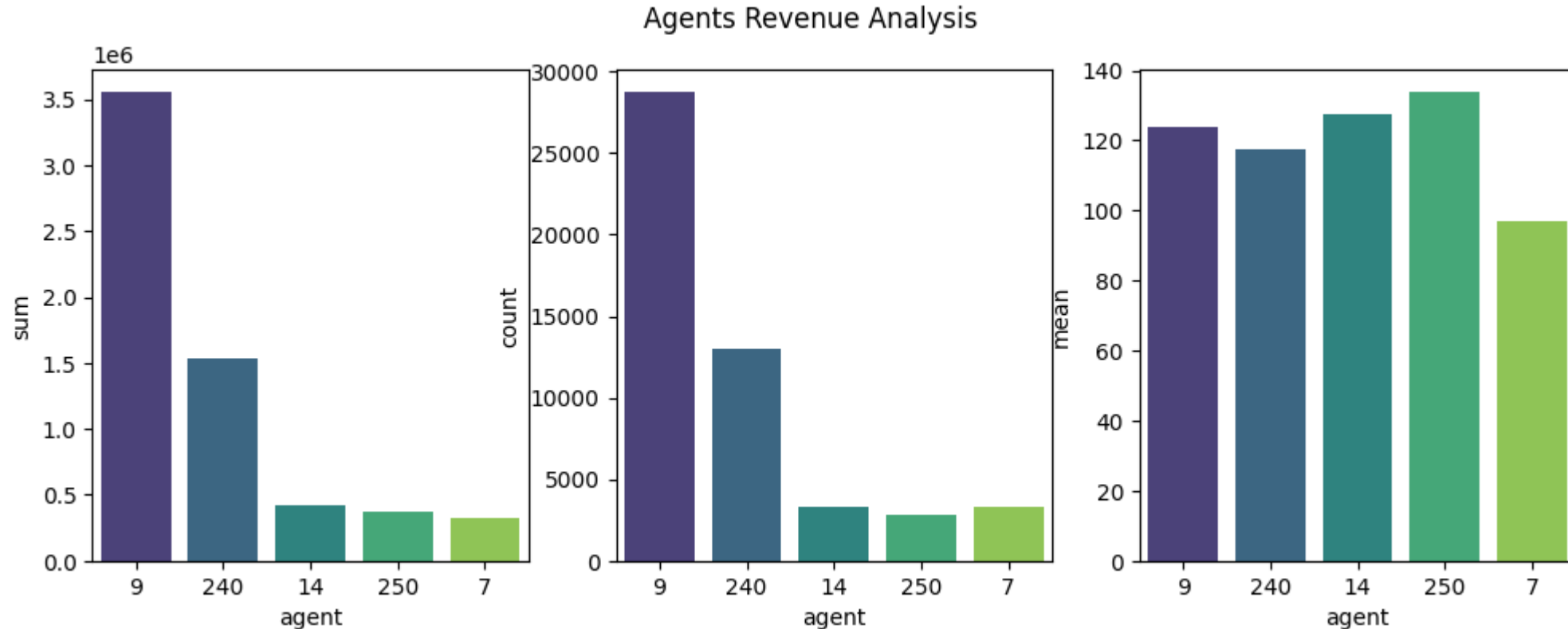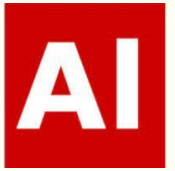
Agent 9 has most bookings and revenue,whereas agent 250 has highest Average Revenue

Q.2 Which meal type is mostly preferred ?

Most of the visitor has opted for only Morning Breakfast(BB),there are also tourist who opted for HB,SC type meal but approximately no visitor has opted for Full Board(FB) type meal

Q.3 Where are the most tourist come from?
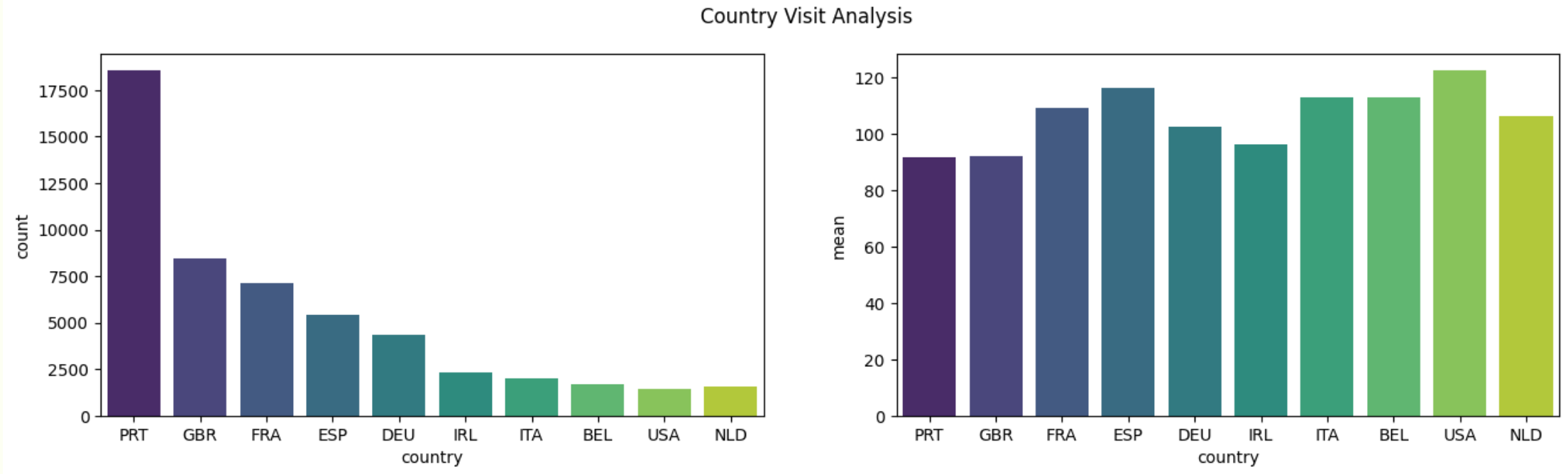
From this chart we can see most customer are from Portugal and in term of continent most customer are from Europe . From USA there are very least customer but it has highest Average Revenue.
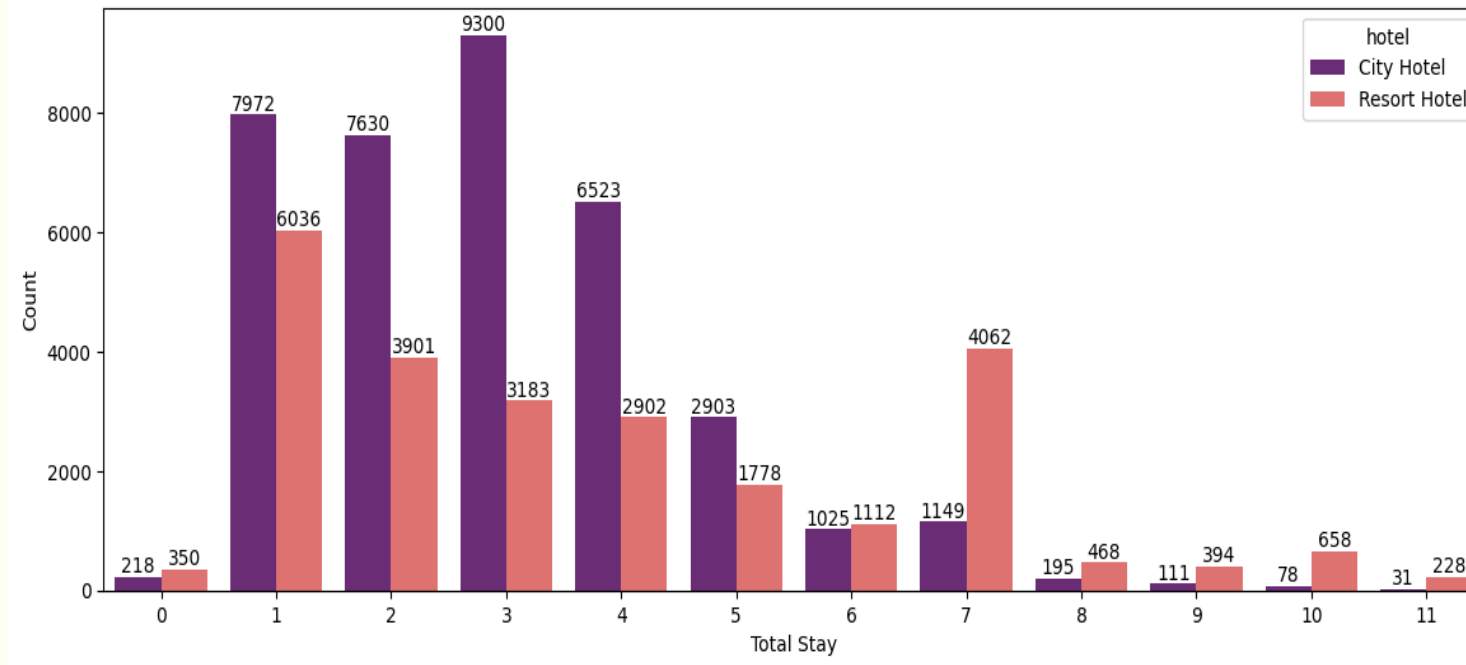
Q.4 Which type of Hotel is mostly booked ?
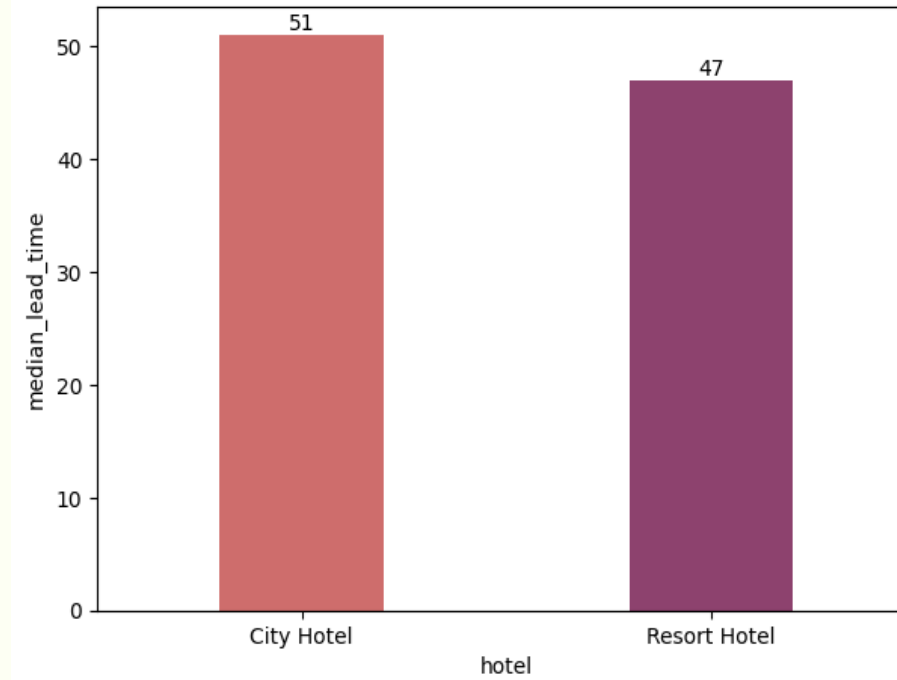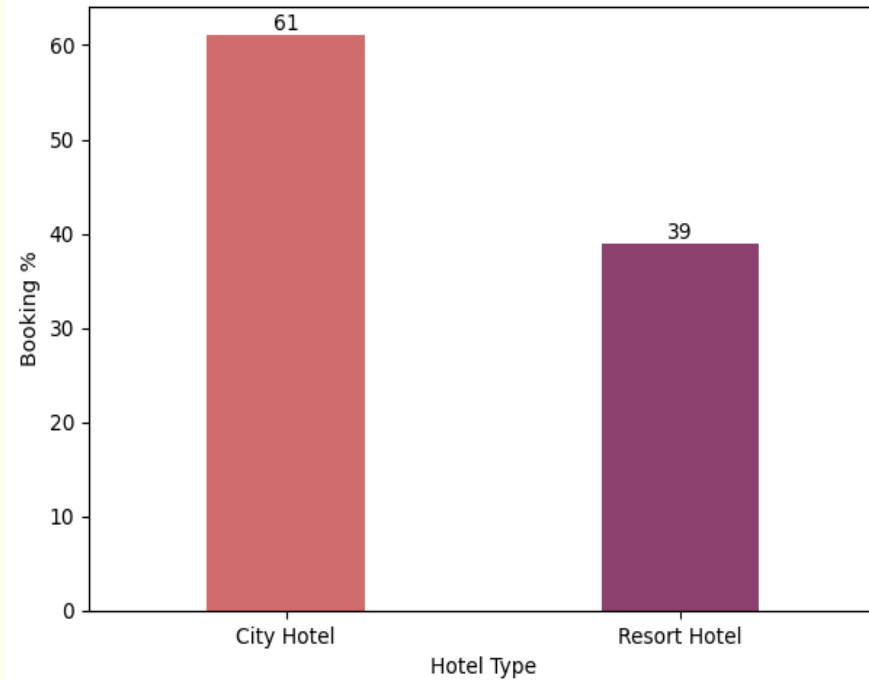
Customer booked City Hotel mostly(61%),whereas Resort Hotel is booked(39%)

Hotel Wise Analysis

What are number of days of stay in each hotel ?

- Most of the people stay for upto 4 days.
- For large days people prefer Resort Hotel.
- There is sharp increase in number of booking in Resort Hotel for 7 days , this maybe due to the reason that is customer get a good deal for a week booking in Resort Hotel.

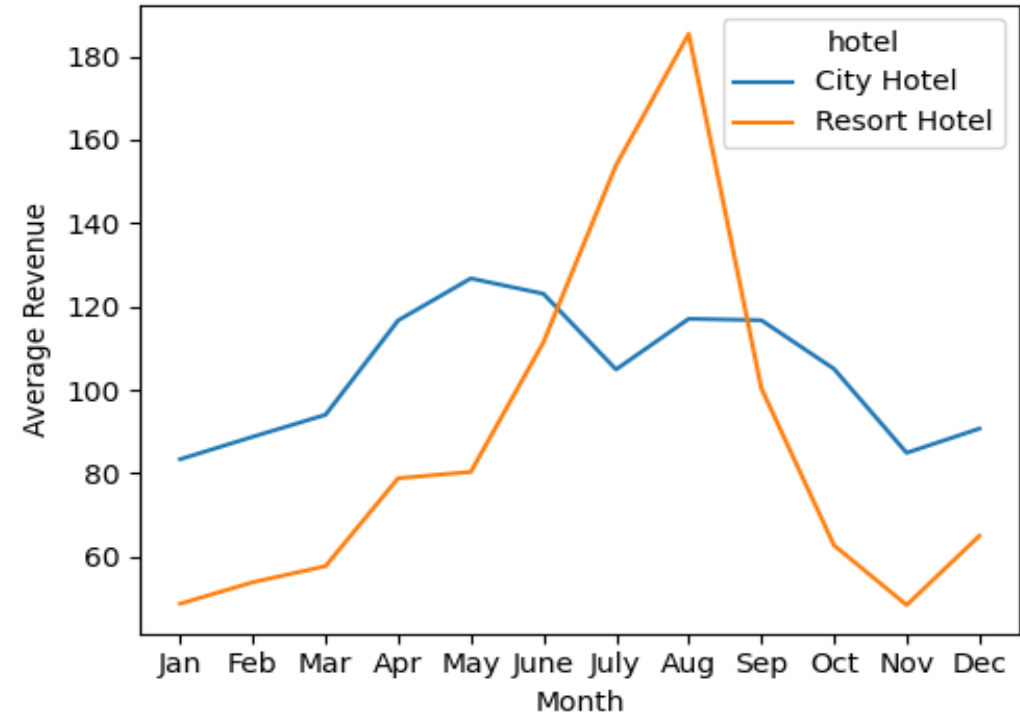**Which group has stayed the most often at each type of hotel?**
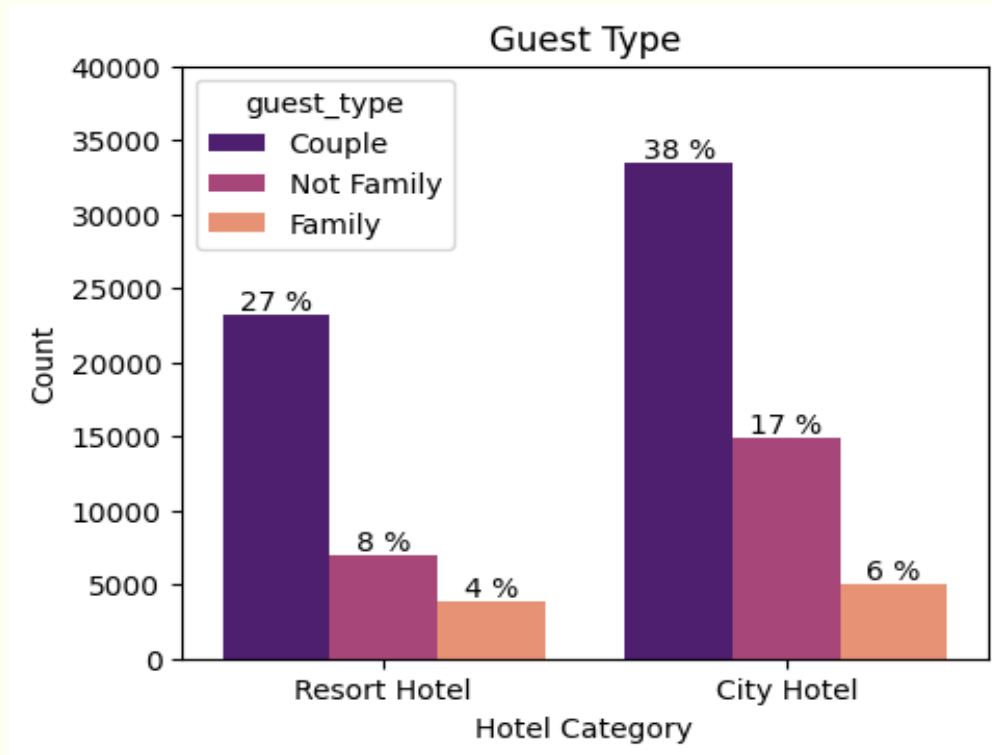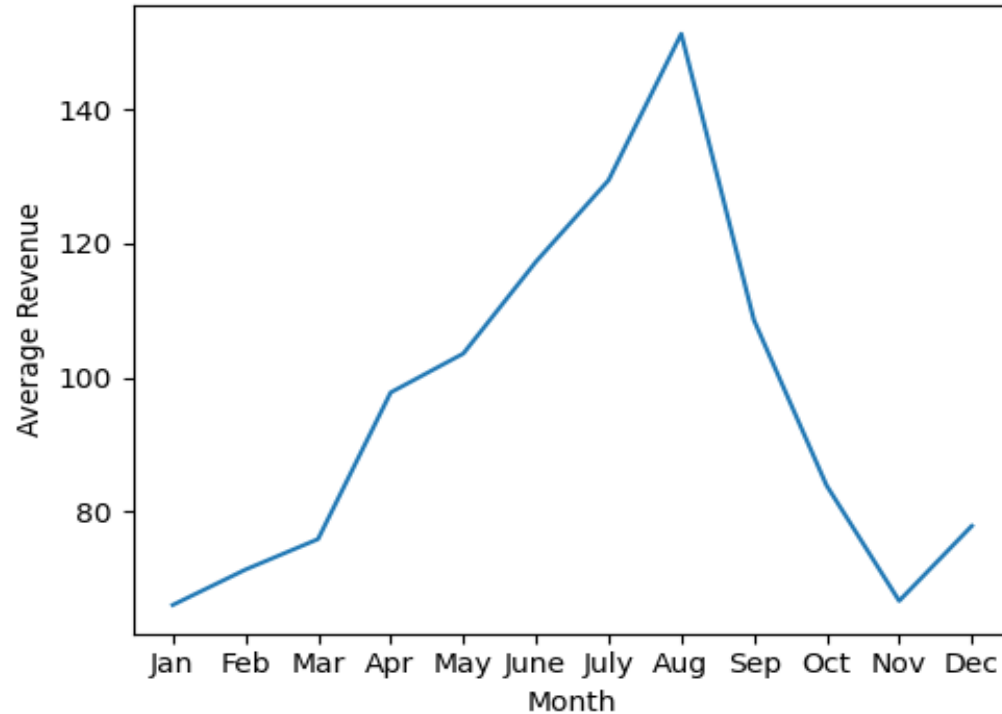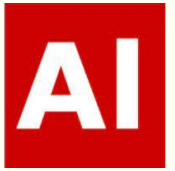

- Most of the booking's are done by couples.
- City Hotel compromise of 38% couple and Resort Hotel compromise of 27% couple.

Timewise Analysis

**What are the average revenue collection in each month ?**
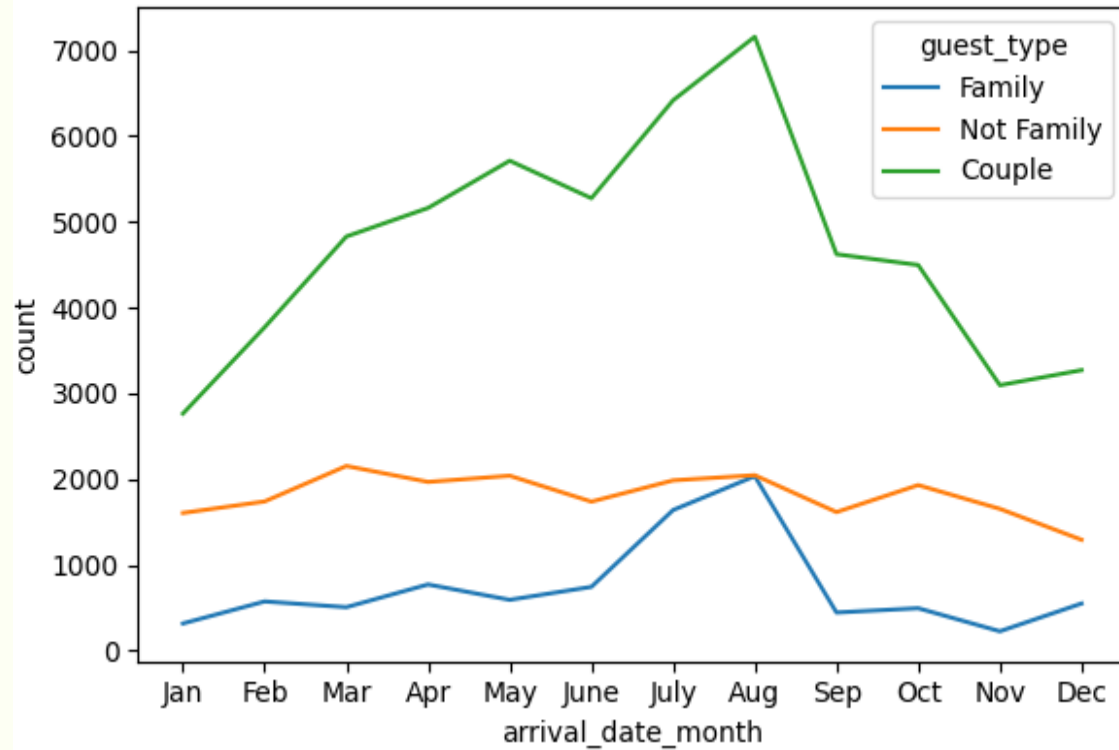

Hotel booking increases from Jan to August , after that it started to fall.

Most of the booking are done in July and August this may be due to the summer holidays.

Resort Hotel provides highest revenue in July and August. This means in July and August resort Hotels are very high in demand.

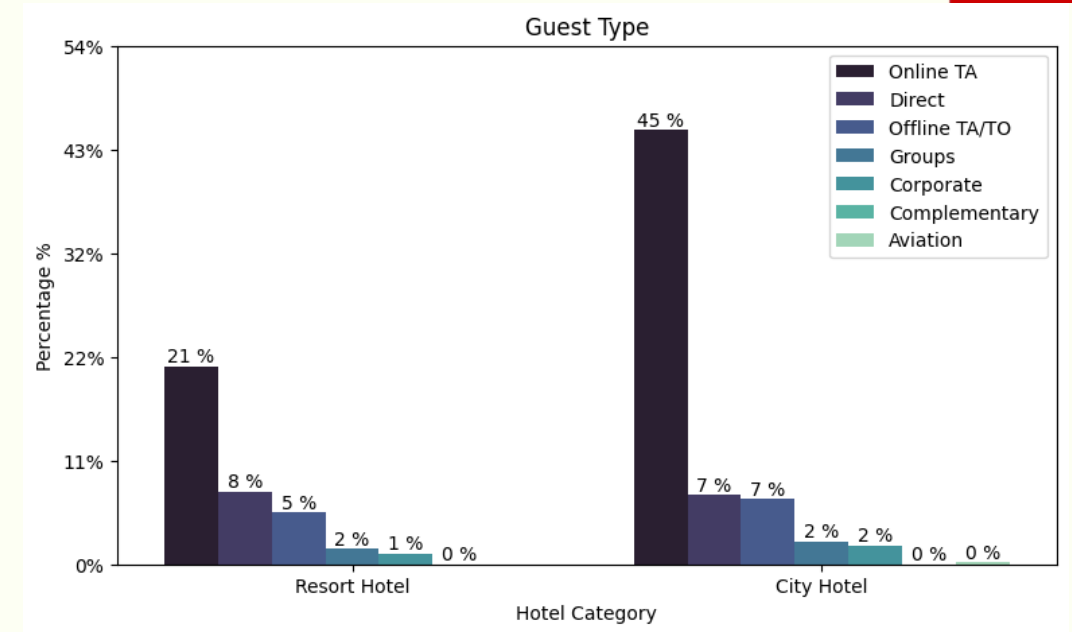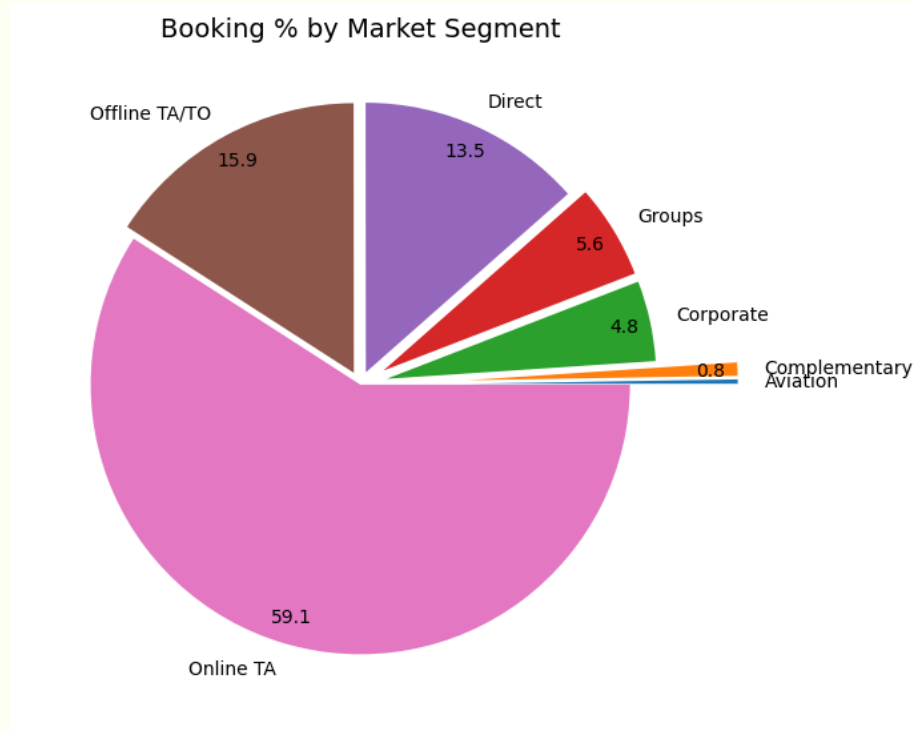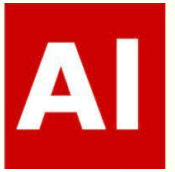**Which category visit most in specific months?**
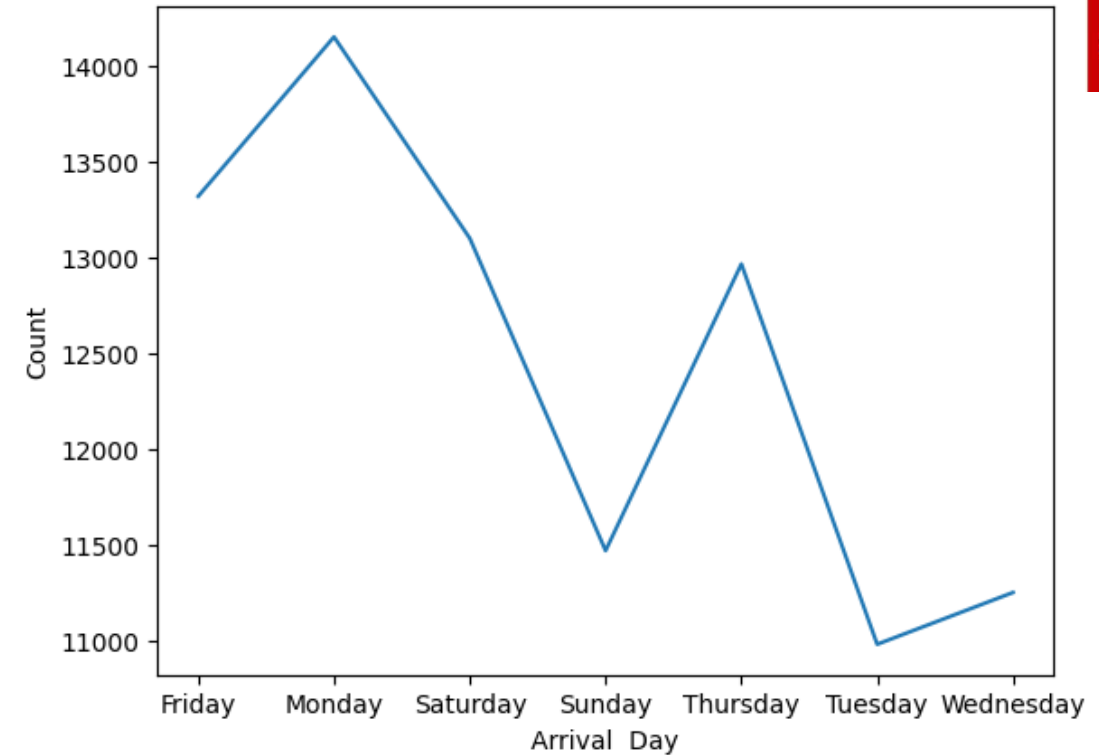
Couple has highest visit monthly .
Monday and Friday has highest Arrival .The highest number of arrival in Friday is due to the weekend holidays.

Market Segment

Which market segment attract more customer ?

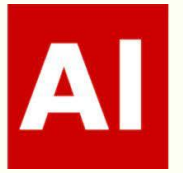Booking % by Market Segment

Guest Type

Almost (59%) of people belong to online TA market segment whereas Offline TA/TO has (15.9%) market segment followed by Direct(13.5%).
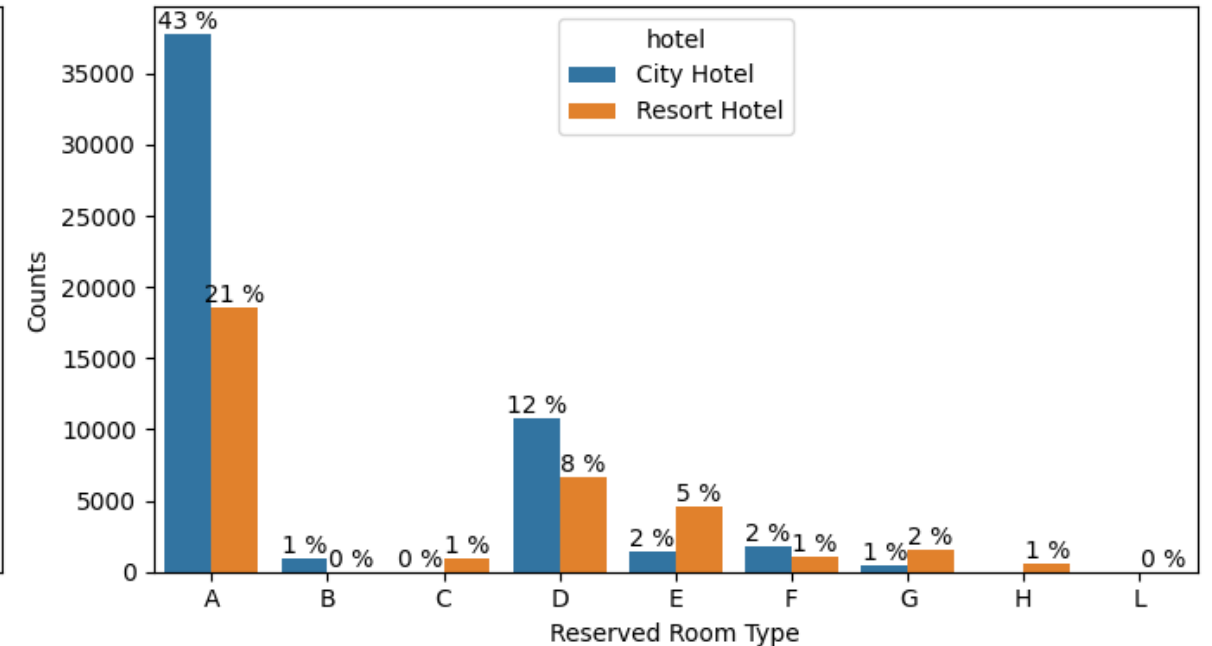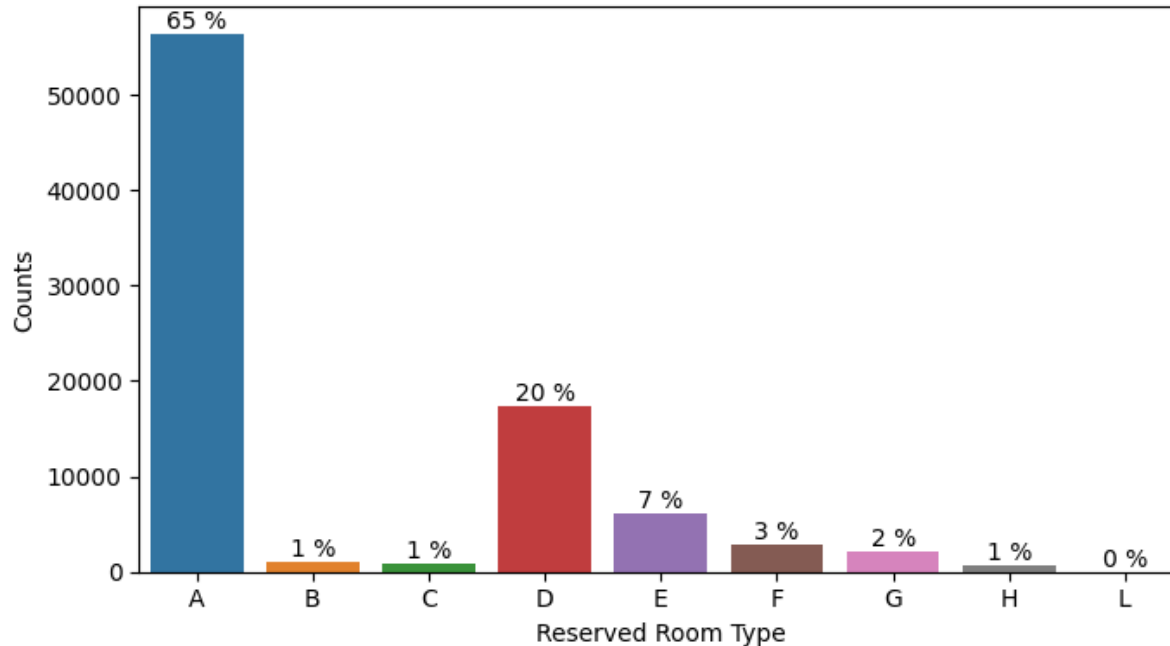
Most of the revenue comes from Online TA.

City Hotel Online TA compromise 45% of the revenue
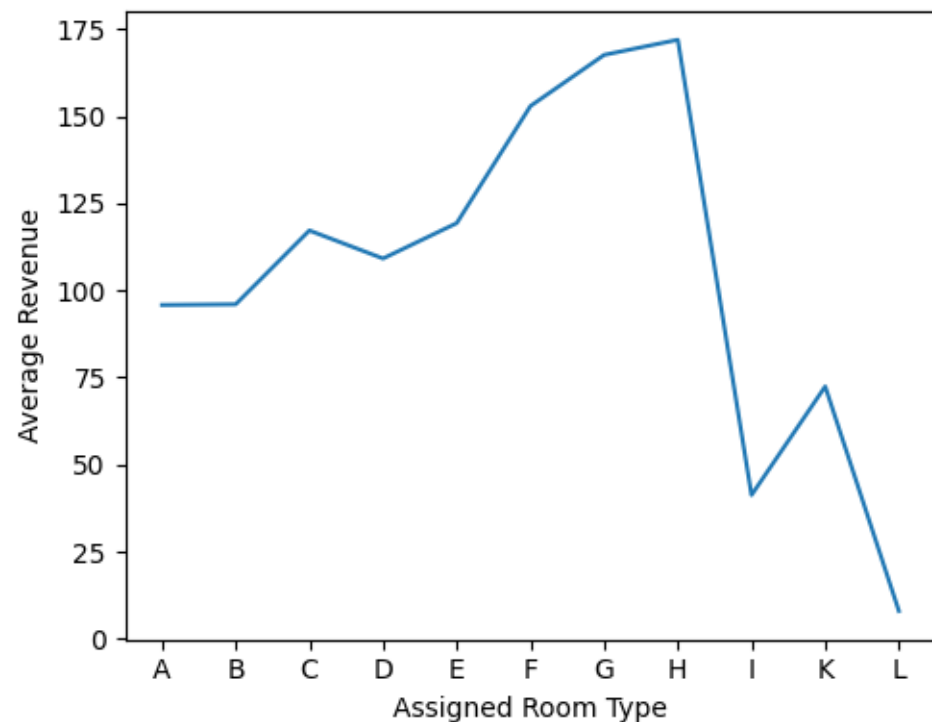
Reserved type and Assigned Type
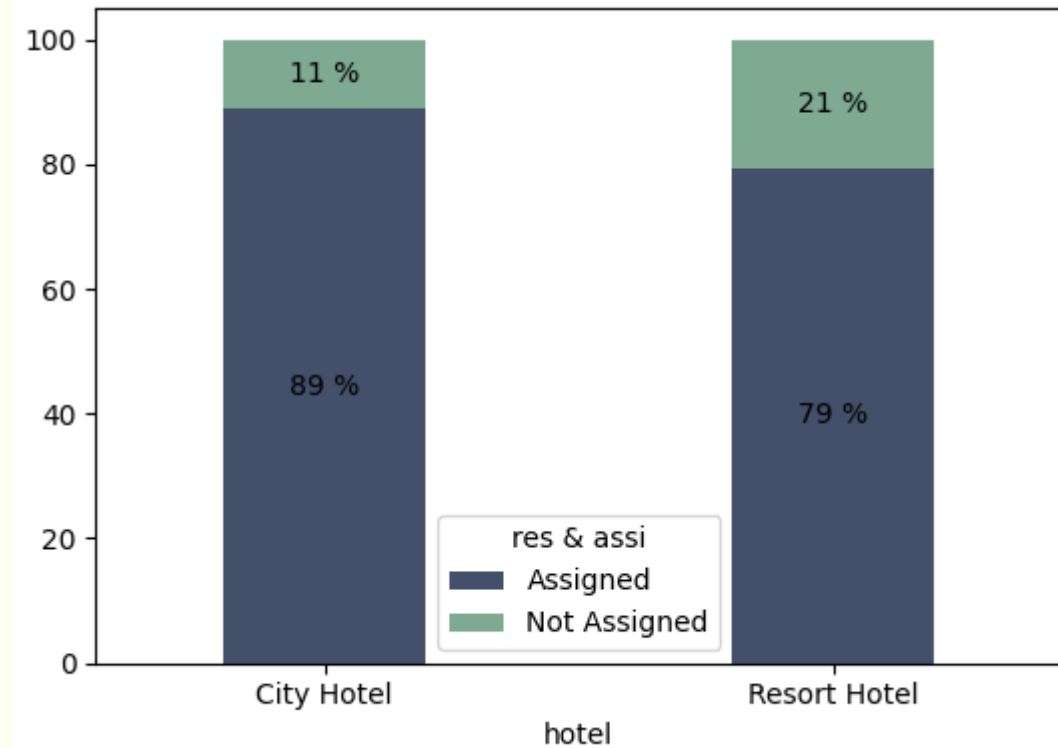
Which type of room is mostly demanded ?

65% customer want's the room type A and 20 % wants room type D.
So,room type A and D are most demanded rooms.

Which type of room give highest revenue ?

What are the percentage of customer assigned ,the reserved room ?

The Average Revenue increases from Room type A to Room type H,after that it falls drastically. In Resort Hotel,21% are not Assigned the reserved room . This means in Resort Hotel there are shortage of specific room types which are high in demand

Not assigned demanded room effects the revenue or not ?
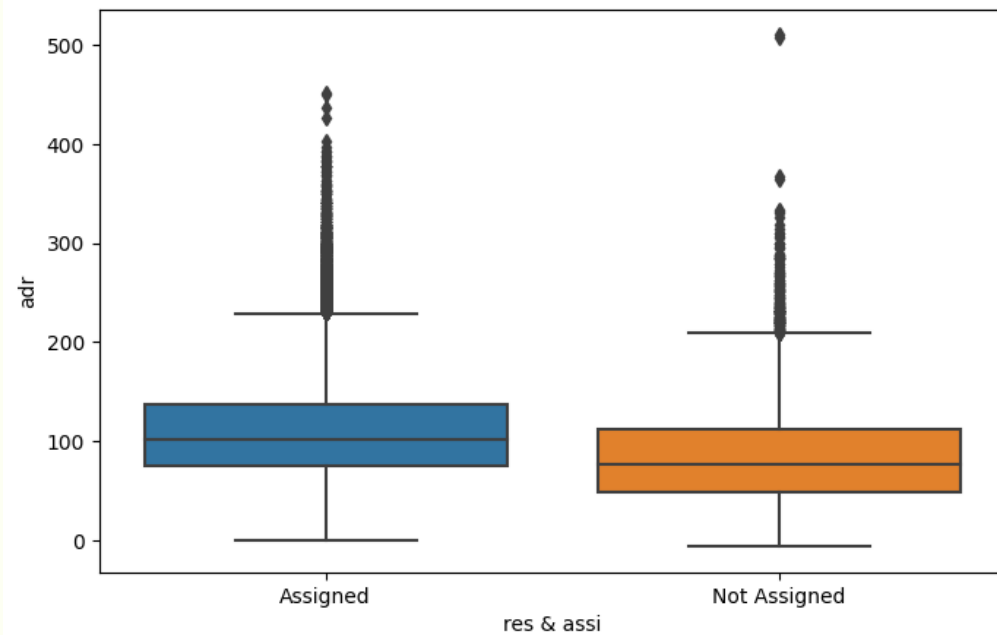
Does not assigned the same room effects cancellation ?
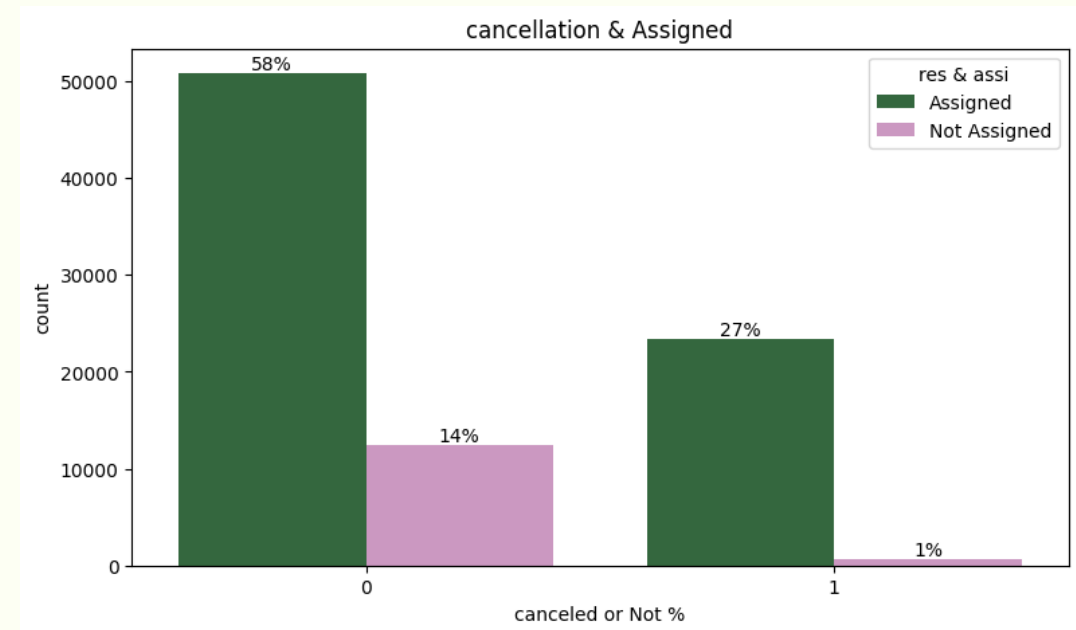
Not Assigned the same room will lead to decrease in revenue and also good deals.
Not Assigned the same room does not lead to cancellation.

Cancellation
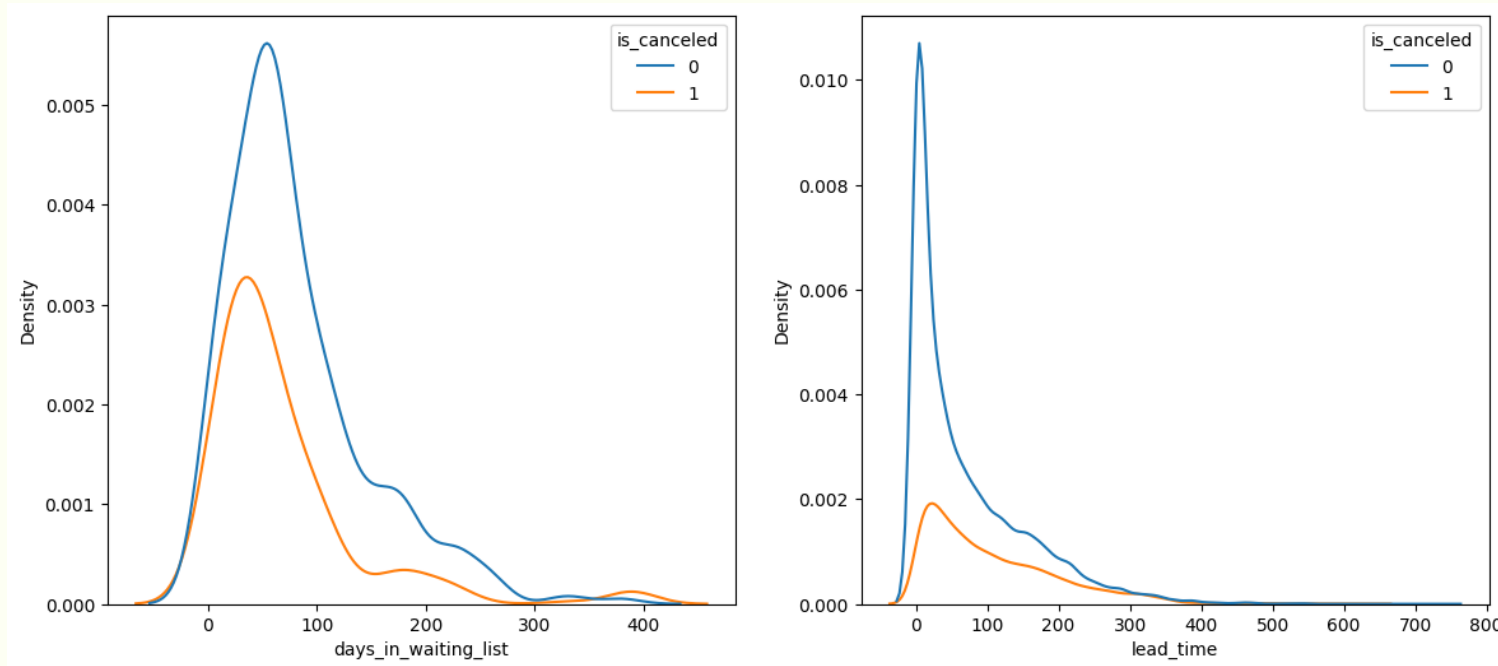
While it is evident that the majority of cancelled reservations have waiting periods under 150 days, the majority of confirmed reservations also do.
This demonstrates that the waiting period has no affect on booking cancellations.
Booking cancellations are unaffected by lead time since the cancelation and not-cancellation curves are equal for lead time as well.

Country wise Analysis
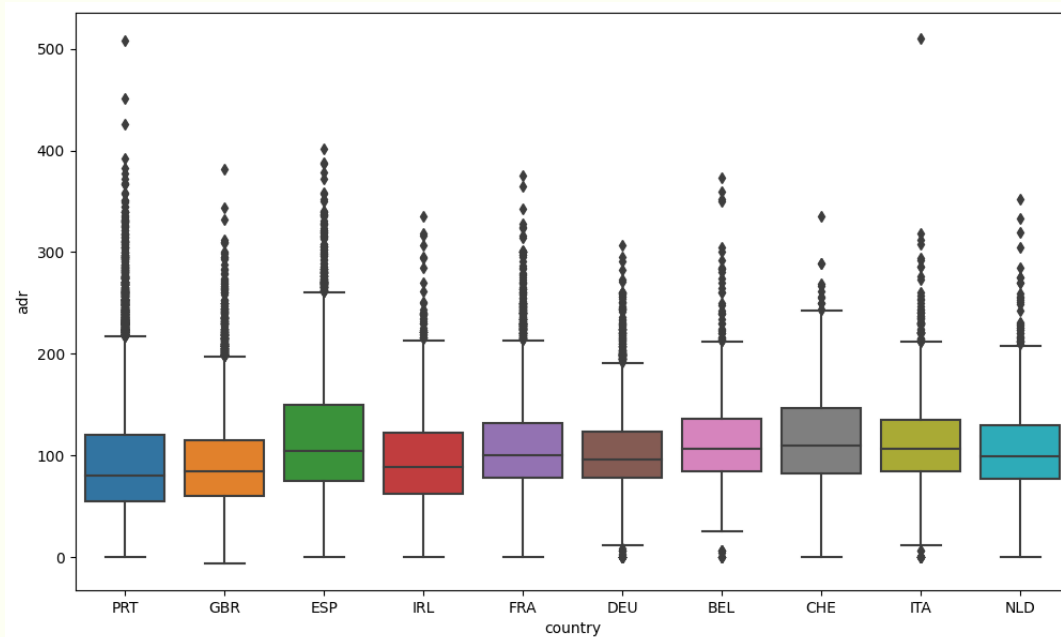
Switzerland(CHE),Belgium(BEL) and Spain(ESP) has highest Average Revenue but the best deals are mostly from Spain and Portugal

Correlation Heatmap
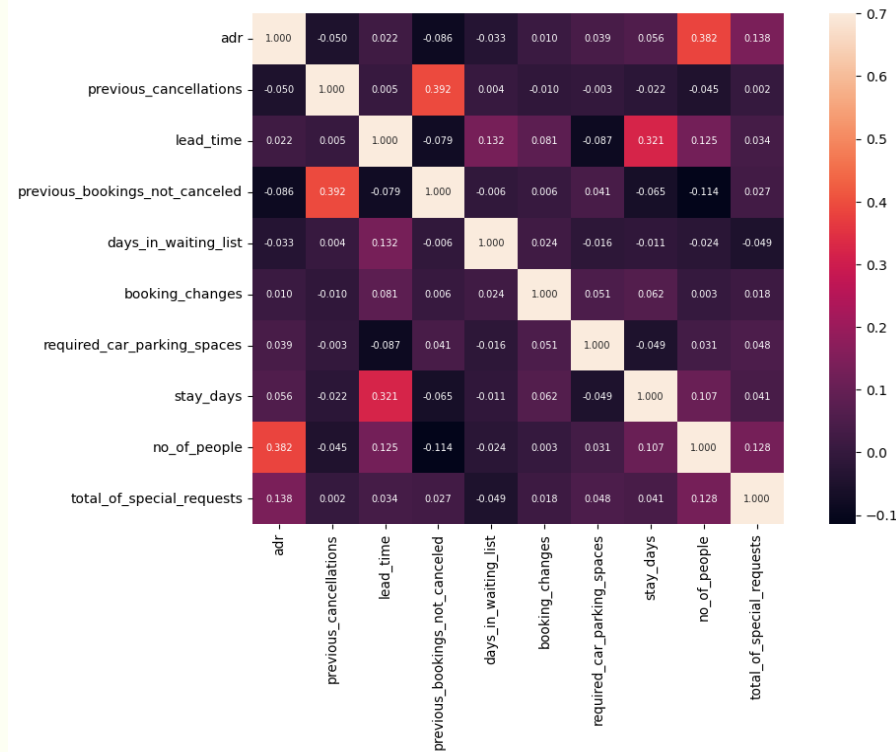
From correlation Heatmap we can see,there are three correlation which are high(30 to 40 %).
1. adr vs no_of_people (0.382 +ve)
2. stay_days vs lead_time(0.321 +ve)
3. previous cancellations vs previous_bookings _not cancelled

    Medium correlation(10 to 15 %)
1.   no_of_people vs total_of special_requests
2. stay_days vs no_of_people
3. lead_time vs day_in_waiting_list
4. lead_time vs no_of_people
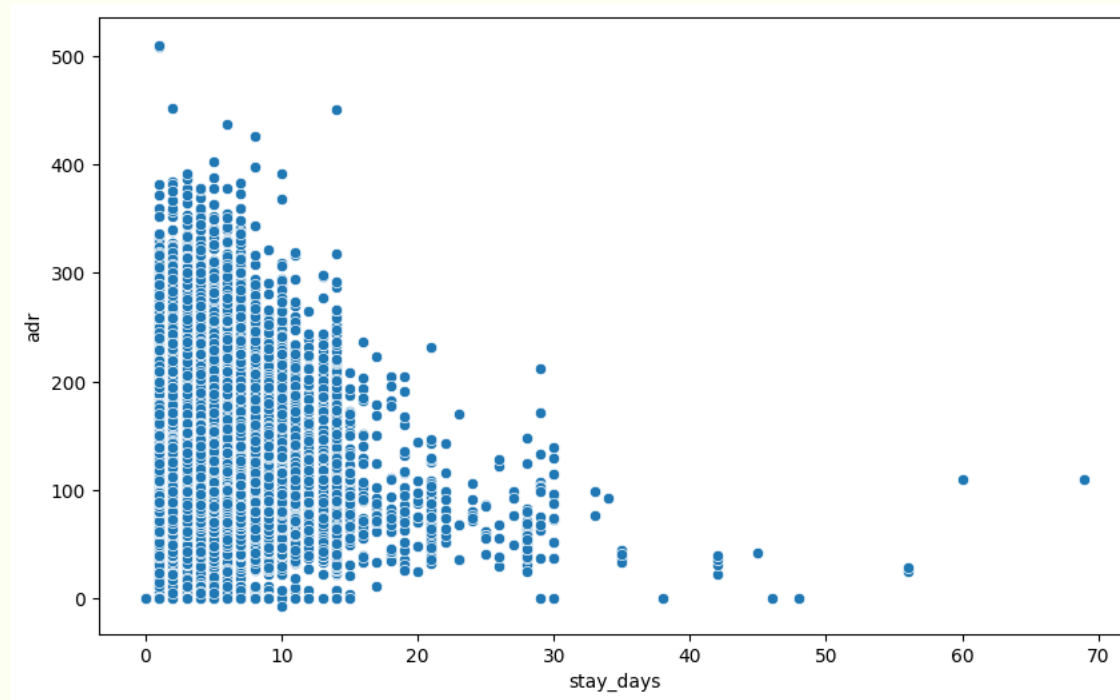
For shorter stays the adr(average daily rate varies greatly) but for longer stays (> 15 days) adr is comparatively very less. Therefore, customers can get better deal for longer stays more than 15 days.

Conclusion

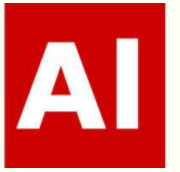(1) Agent 9 has highest no of booking and revenue ,but in term of Average Revenue Agent 250 and
    Agent 14 are the highest

(2) Increase in total_request lead to increase in Revenue.

(3) Most of the visitor has opted for only Morning Breakfast(BB),there are also tourist who opted
    for HB,SC type meal but
     approximately no visitor has opted for Full Board(FB) type meal

(4) For Resort Hotel there is drastically increase increase in number for One Week ,this may be
    because there is
     better deal for a week in Resort Hotel.

(5) People mostly stayed for upto 4 days .Highest number of days of stay is 1. For long days people
    mostly prefer
     Resort Hotel and for short days ,City Hotel.

(6) Most(65%) of the visitor type are couple's and after most of the visitor are Not family(25%)

(7) August and July is the most visited month of the year.This is maybe mostly summer holidays lie
    in July and August.

Conclusion

(8) Customer mostly arrive in Monday and Friday. The high number of visitor in Friday may be due to weekends.

(9) Almost (59%) of people belong to online TA market segment whereas Offline TA/TO has (15.9%) market segment followed by Direct(13.5%)

(10) In term of revenue Online TA city hotel has highest revenue and also highest Average Revenue.

(11) The Average Revenue increases from Room type A to Room type H , after that it falls drastically.

(12) In city Hotel,11% are not Assigned the reserved room whereas in resort Hotel,21% are not Assigned the reserved room. This means in Resort Hotel there are shortage of specific room types which are high in demand

(13) Not Assigned the same room , not only decrease the revenue but it also decrease the chances of best deals.

(14) Switzerland(CHE),Belgium(BEL) and Spain(ESP) has highest Average Revenue but the best deals are mostly from Spain and Portugal

Thank *You*