EDA
Capstone Project
Hotel Booking Analysis

Made By:

SHABANA ALI

AI

Content

➤ Cleaning data

➤ Data Summary

➤ Univariate analysis

➤ Hotel wise analysis

➤ Distribution Channel wise analysis

➤ Better Revenue Generating Deals for Hotels

➤ Booking Cancellation Analysis

➤ Time wise analysis

➤ Conclusion

Cleaning data

Cleaning data is vital step some time recently EDA because it will remove the ambigous data that can influence the result of EDA.

While cleaning data we will perform following steps:

1. Remove duplicate rows.

2. Handling missing values.

3. Convert columns to appropriate data types.

4. Adding important columns.

Data Summary

Given data set has different columns of variables vital 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.

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

market_segment: This column show how reservation was made and what is the purpose of reservation. Eg, corporate means corporate trip, **TA** for travel agent, **TO** for Tour Operator

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 Summary
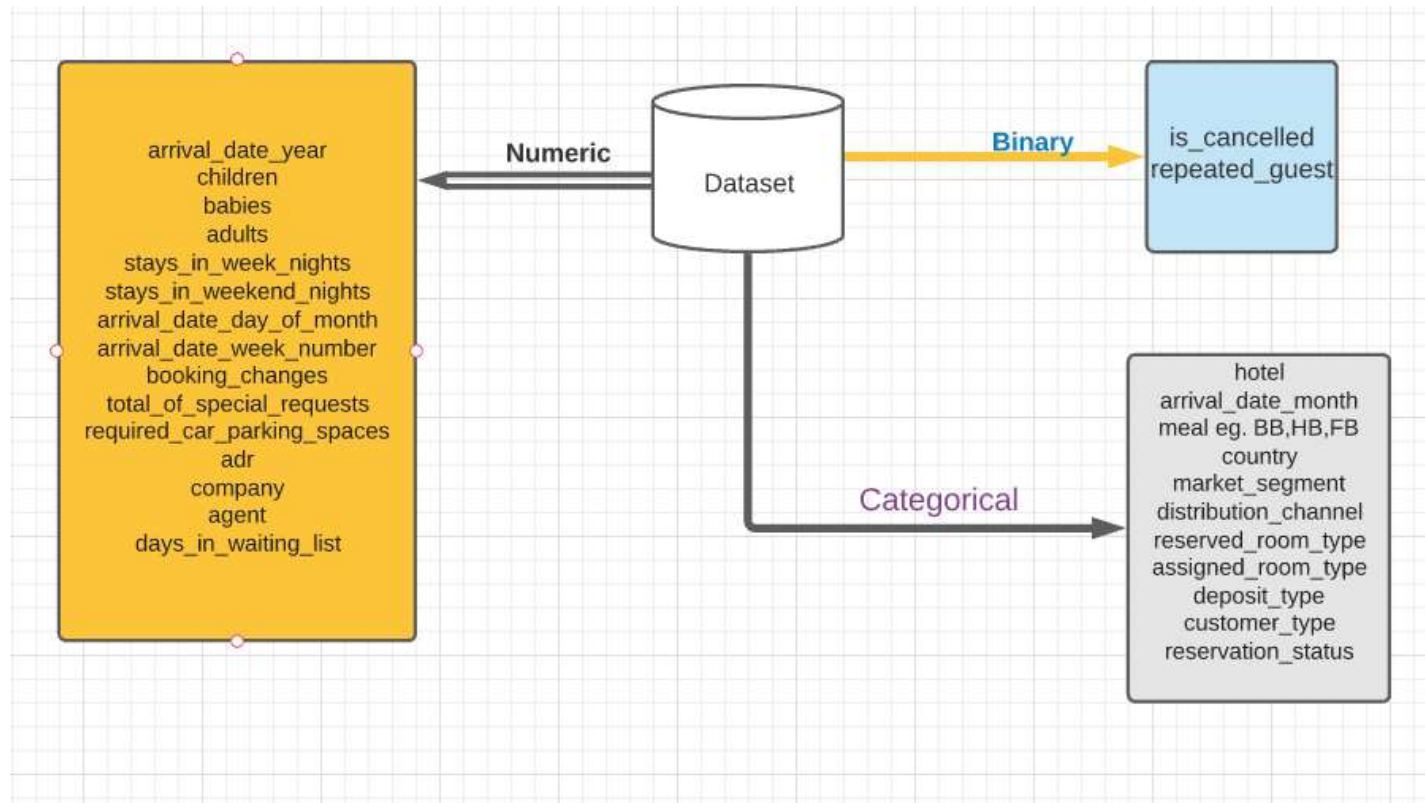
Data Summary

In univariate analysis of given hotel booking dataset, answered the following questions:

1. Which room type is in most demand and which room type generates highest adr?

2. What is the most preferred meal by customers?

3. Which agent made most of bookings?

4. From which country most of the customers are coming?
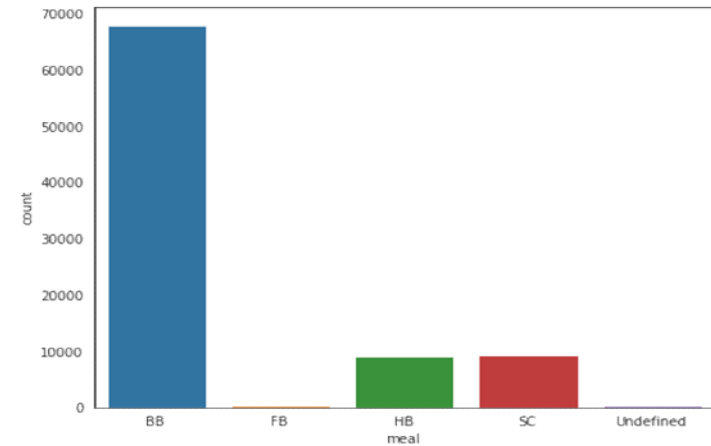
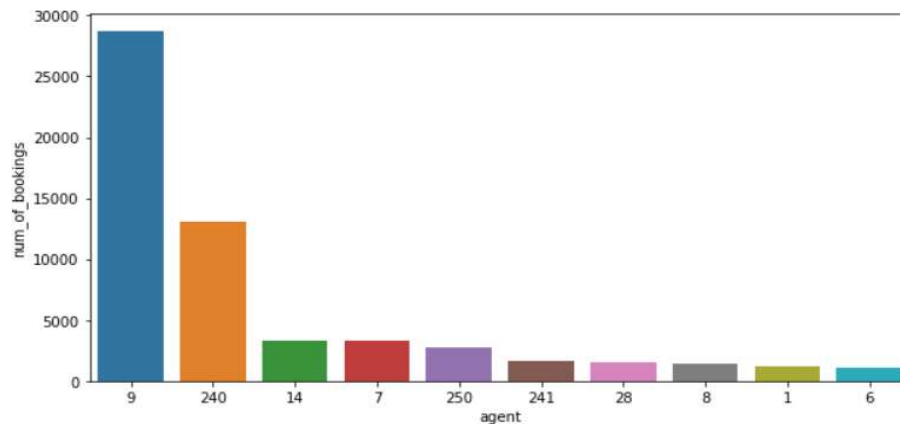# Most demand and which room type generates highest adr
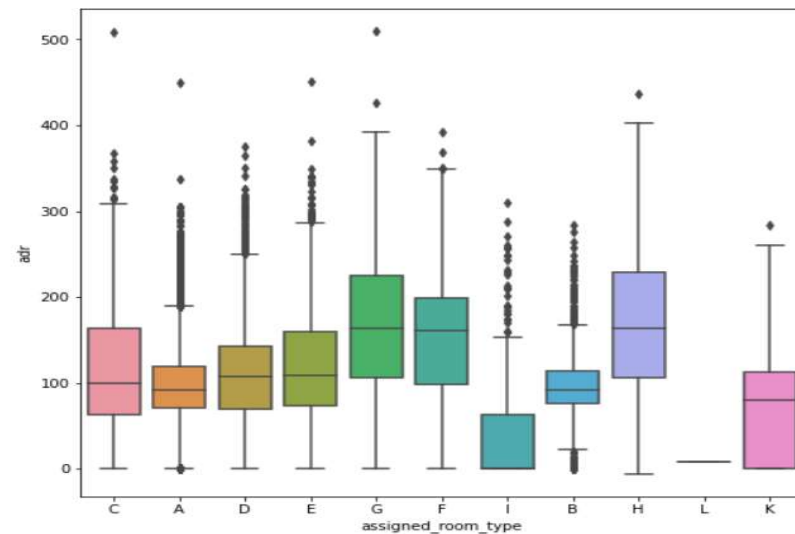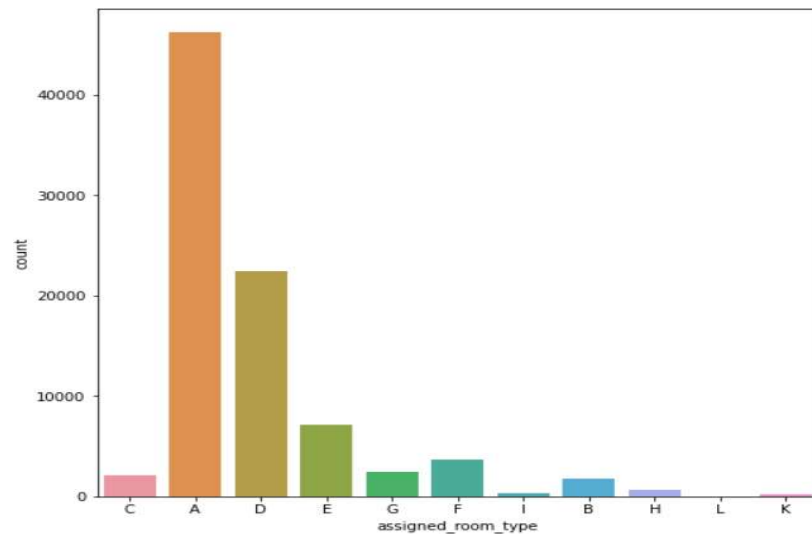
- Higher demand for type 'A' rooms.
- Room types C, G and H are some of the highest adr(average daily rate) generating rooms.
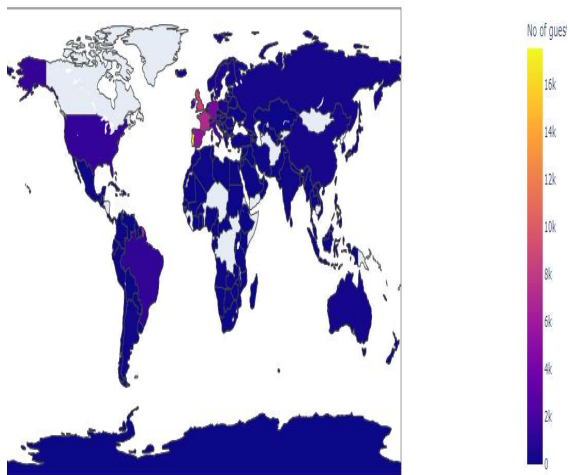- Hotels want to max revenue, then they need to increase the number of room types A and H

Most preferred meal by customers & agent made most of bookings
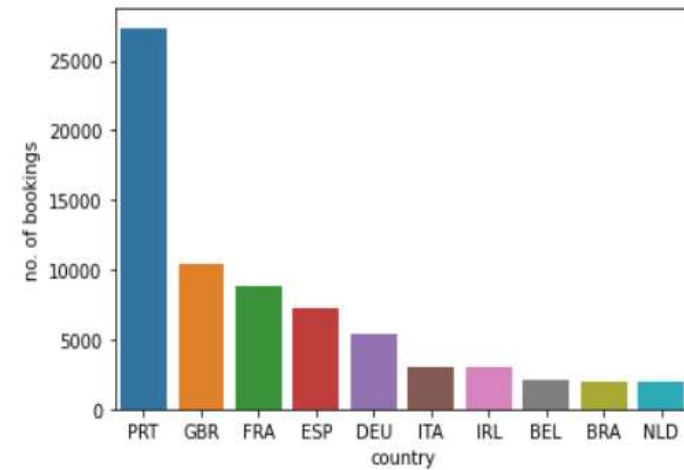
Most preferred meal type is BB (Bed and breakfast).

Agent Id number 9 agent made most of the bookings and 6 has least.

AI

Most customers are from Countries

Most of the customers from European countries like Portugal, Great Britain, France and Spain.


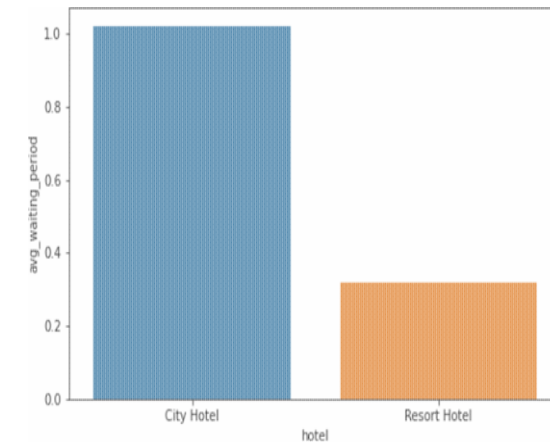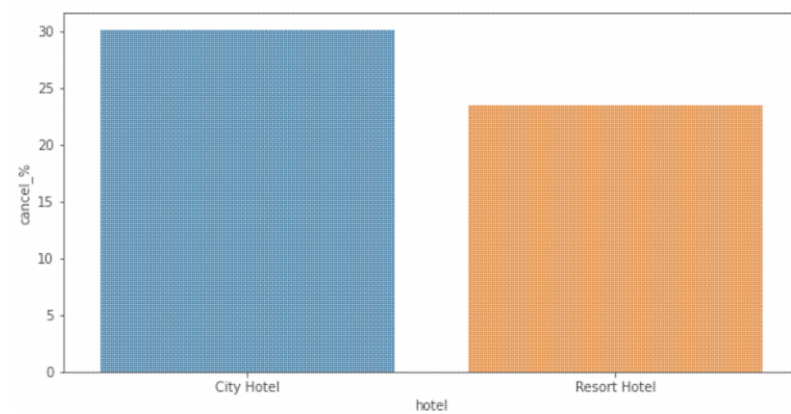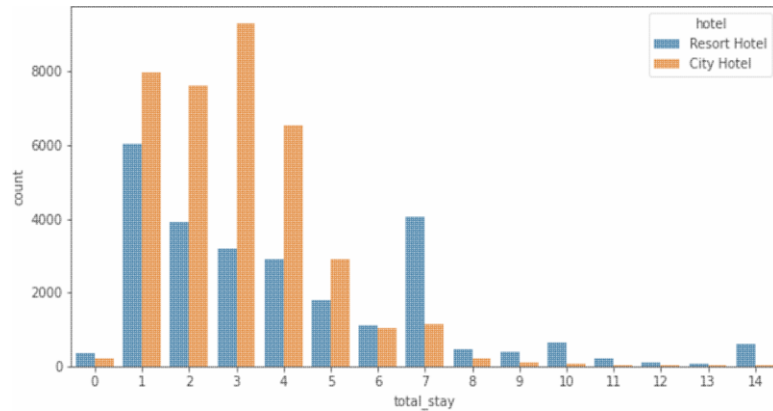Most guest are from Portugal and other Europian contries.

Hotel wise Analysis

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

- ➢ What is the percentage of bookings in each hotels?

- ➢ Find the higher lead time hotel ?

- ➢ Most preferred stay length in each hotel?

- ➢ Which hotel makes more revenue?

- ➢ Which hotel have higher and how much customer returning rate?

- ➢ Find the higher booking cancellations rate?

- ➢ Which hotel, does people have to wait longer to get a booking confirmed?
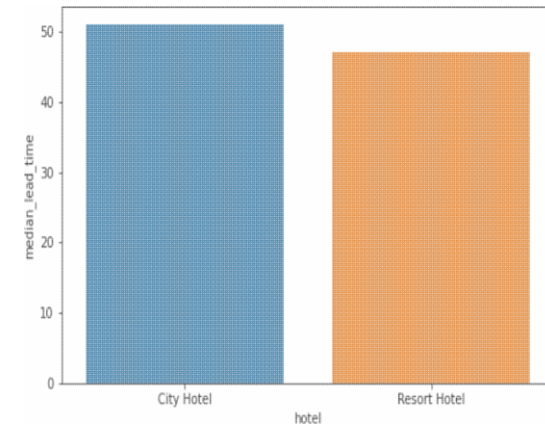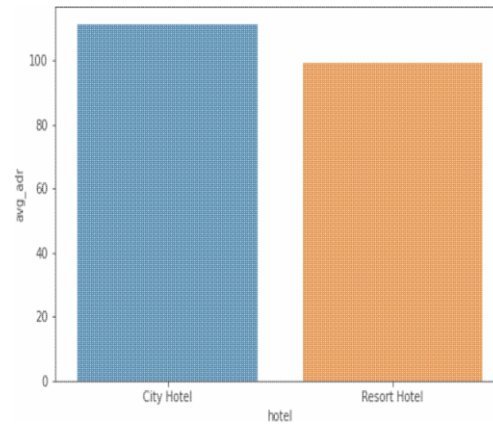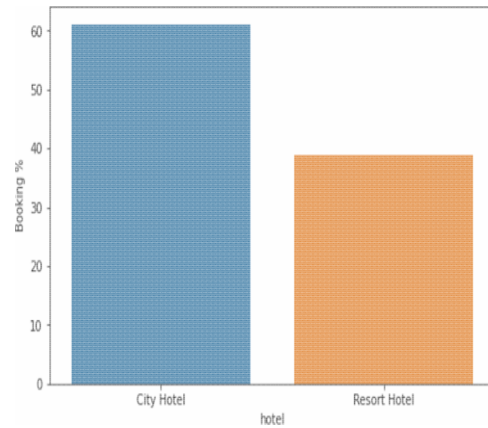
- About 60% bookings are for City hotel and 40% bookings are for Resort hotel.
- Avg adr of Resort hotel is somewhat lower than that of City hotel. Consequently, City hotel appears to be making marginally more revenue.
- City hotel has marginally higher median lead time. Moreover middle lead time is higher in each case, this implies customers for the most part arrange their hotel visits way to early.
- City hotel has longer waiting time, hence City Hotel is much busier than Resort Hotel.

- Most of the stays are less than 5 days. There are very few long stays at hotels but Resort Hotel is preferred for long stays.
- Almost 30 % of City Hotel bookings and 25 % of Resort hotel bookings got canceled anytime.
- Both hotels have exceptionally small persentage rate that customer will repeat, but Resort hotel has marginally higher repeat % than City Hotel.
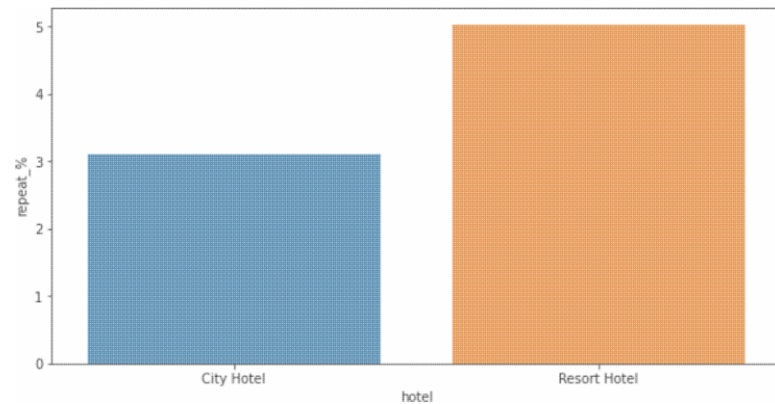
Distribution channel wise Analysis

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

- What is the most common method of booking hotels?

- What is the most popular channel for early booking of hotels?

- Which distribution channel brings better revenue generating deals for hotels?
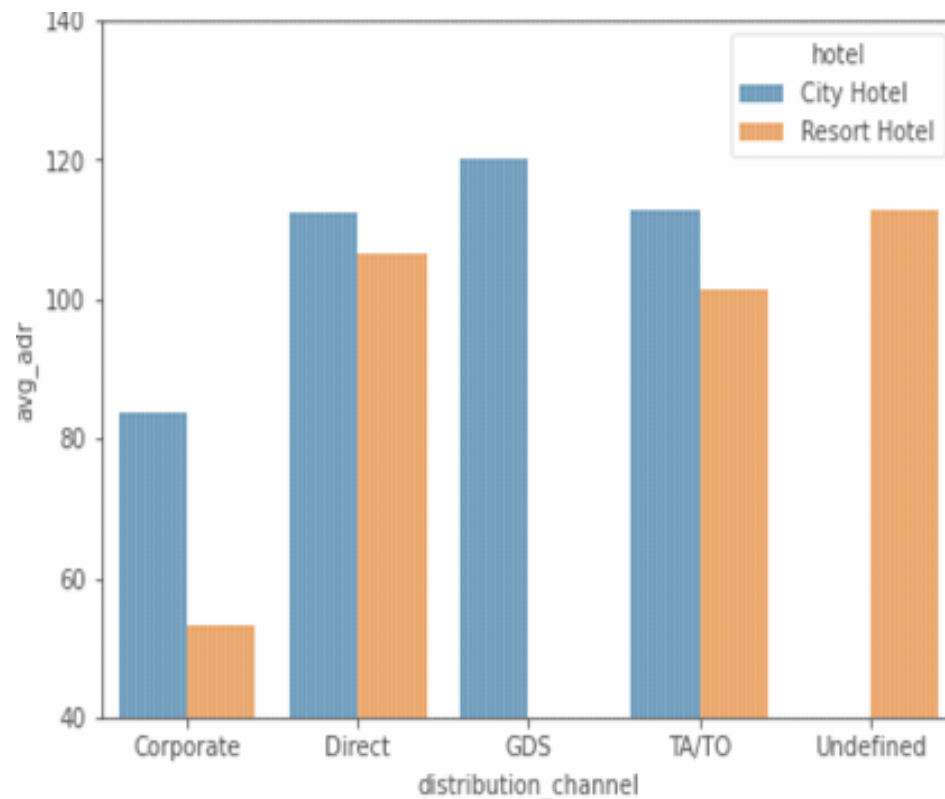
Distribution channel wise Analysis

Booking % by distribution channels
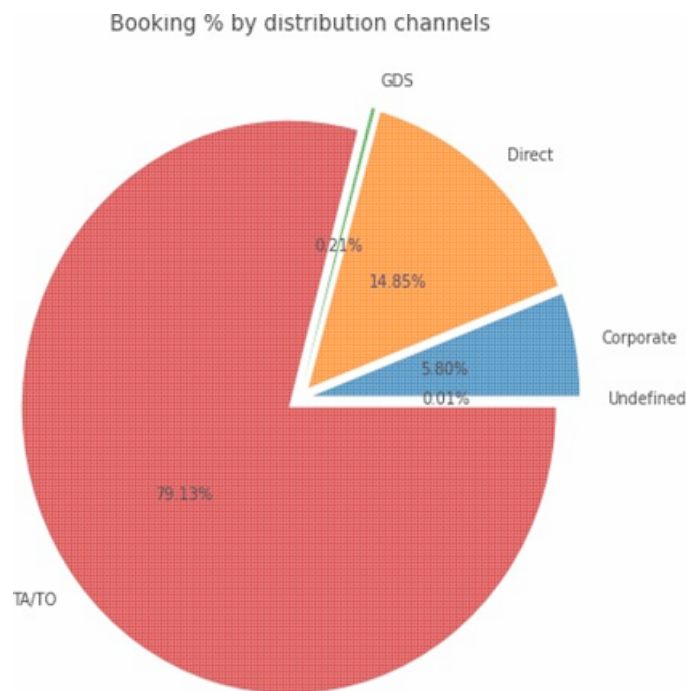
- The majority of customers make reservations through tour operators and travel agencies.
- Second most popular channel is Direct.
- Travel agents and tour operators are also used for early hotel bookings.
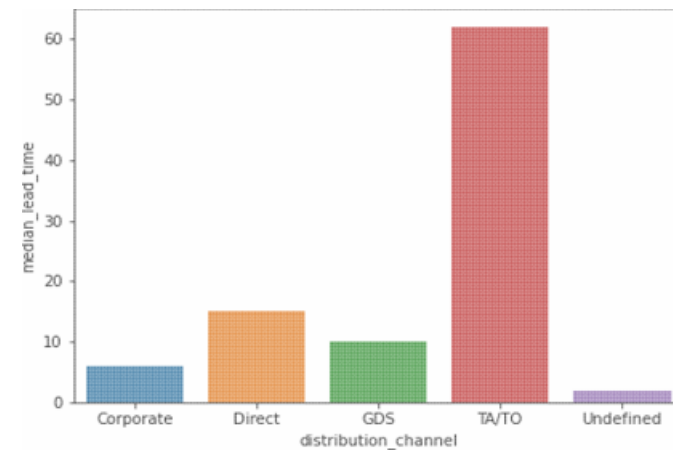
Better Revenue Generating Deals for Hotels

- There are more revenue-generating deals at resort hotels through direct and TA/TO channels. Resort hotel need to increase revenue, hotels need to increase outreach on GDS channels.

- GDS channel brings higher revenue generating deals for City hotel, in contrast to that most bookings come via TA/TO. City Hotel can work to increase outr each on GDS channels to get more higher revenue generating deals.

# Booking cancellation Analysis

We analysis the following possible reasons for booking cancellations:

- Which significant distribution channel has highest cancellation percentage?

- Longer time (in days) in waiting list.

- Longer lead time.

- Not getting same room as reserved.

- Does not getting same room as reserved effects adr?
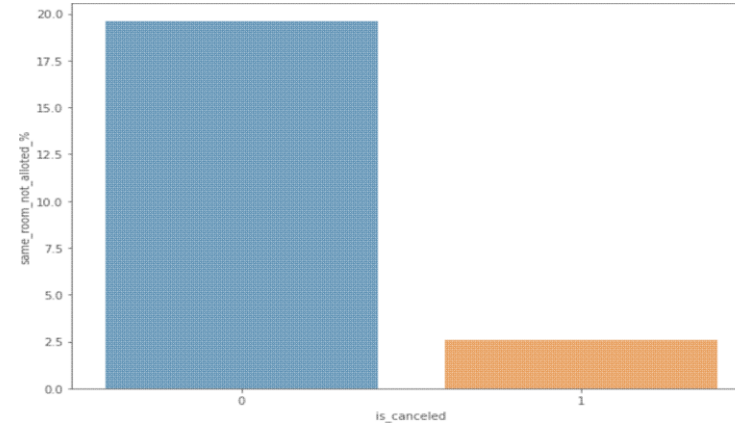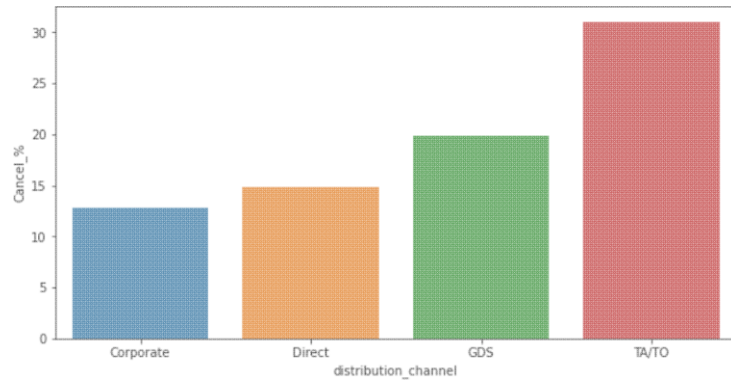
- Cancellation rate for TA/TO bookings is the highest. A total of 30% of TA/TO reservations will be canceled because of this.
- Not getting same room as demanded is not the case of cancellation of rooms. A significant percentage of bookings are not cancelled even after getting different room as demanded.
- Customers who didn't receive the same room have tended to pay slightly less than the advertised rate, with a few exceptions.
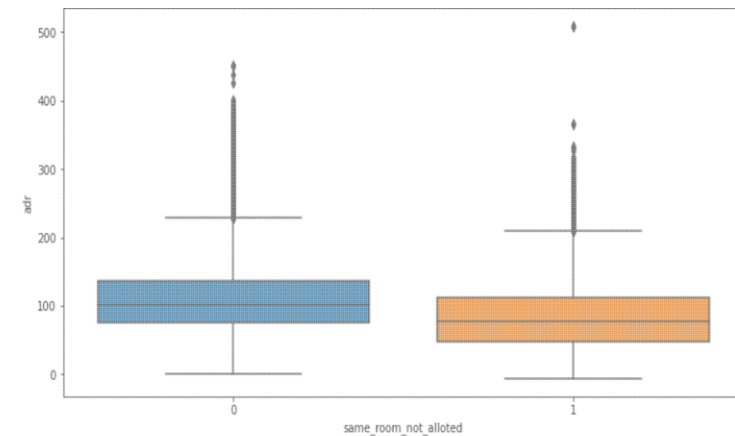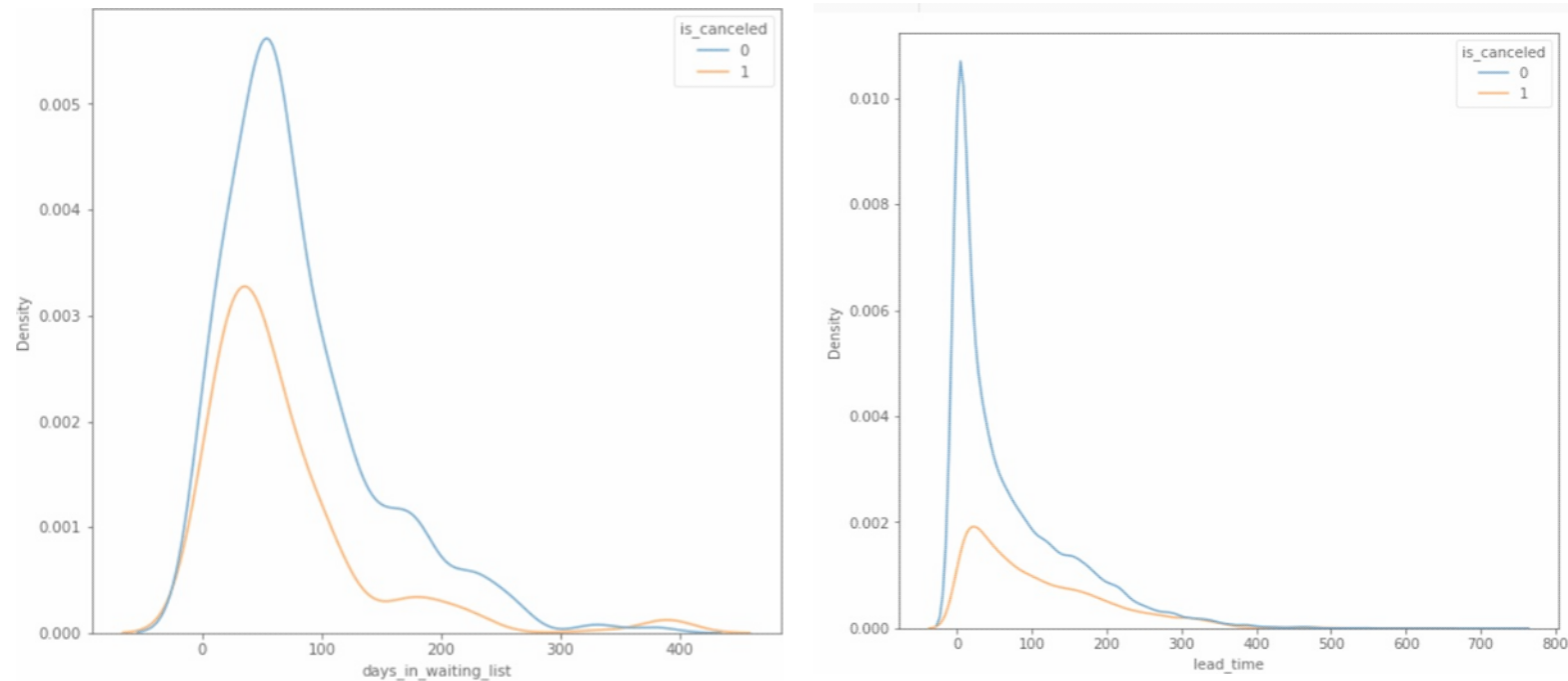
- The majority of cancelled bookings have a waiting period of less than 150 days, and the majority of bookings that are not cancelled also have a waiting period less than 150 days. There is no effect on cancellations due to the waiting period.
- Also, lead time has no effect on cancellation of bookings, as both curves of cancellation and not cancelation are similar for lead time too.
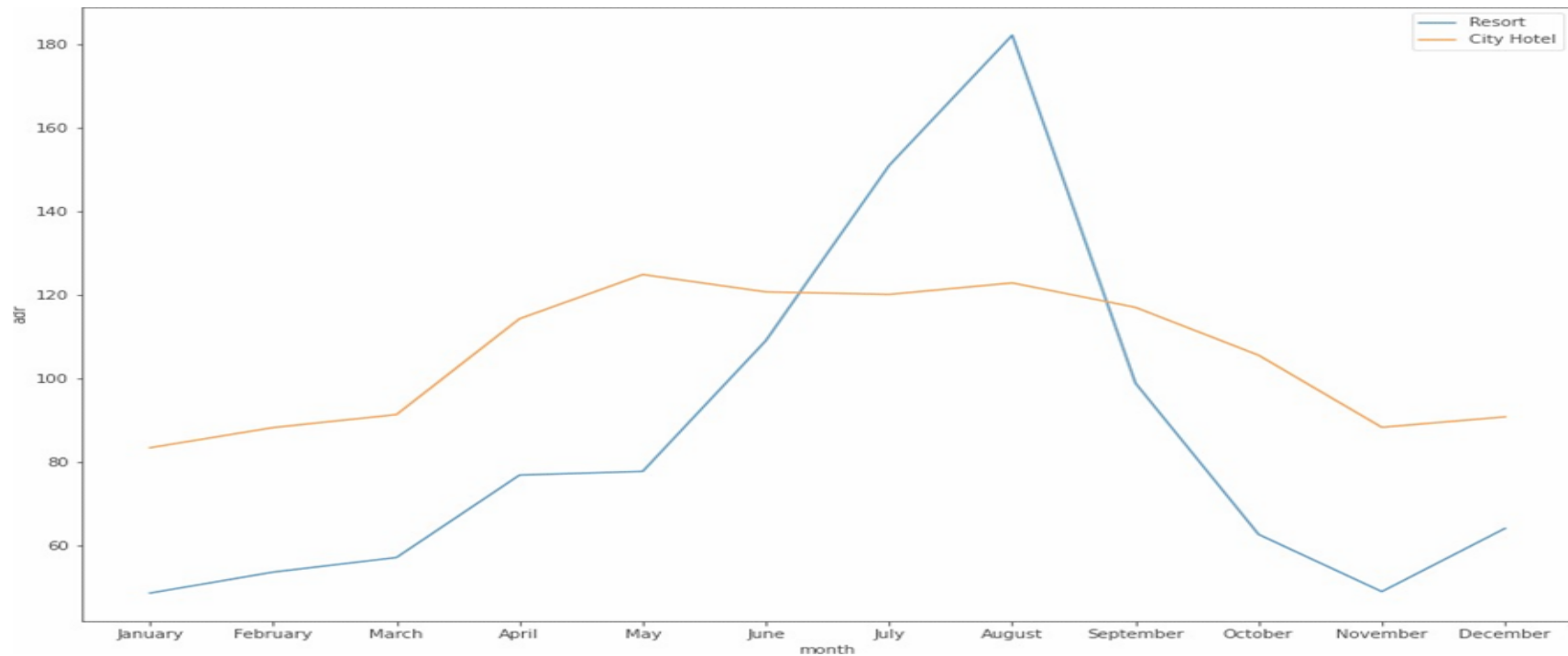
Time-wise Analysis

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

- What are the most busy months for hotels?

- In which months hotels charges higher adr?

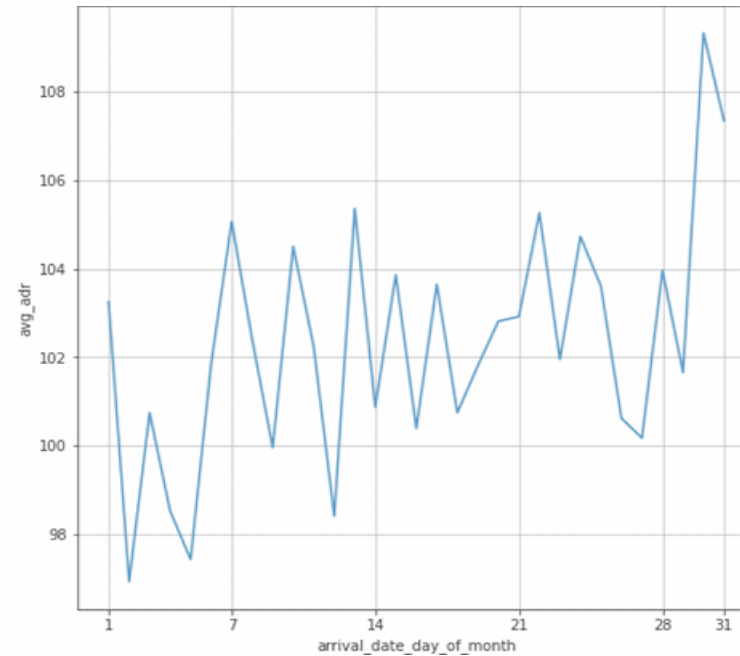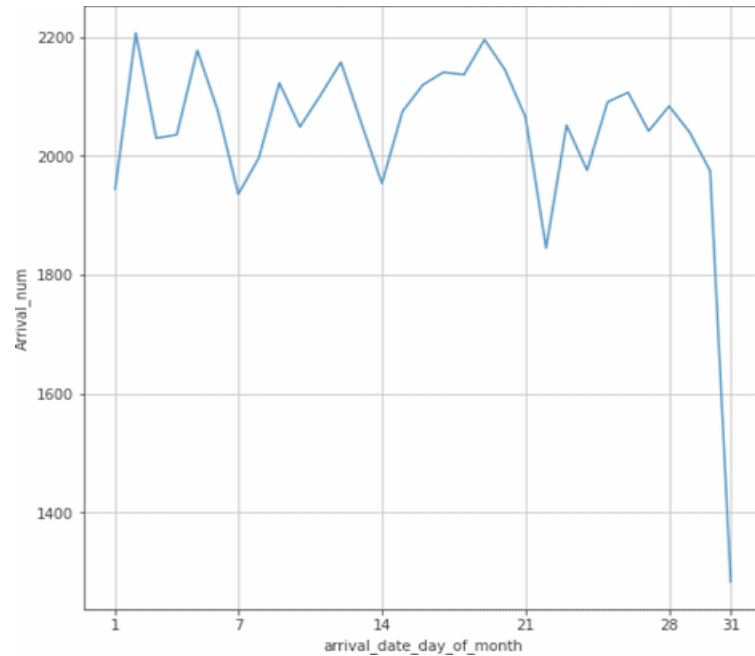- How does booking numbers and adr changes within a month?
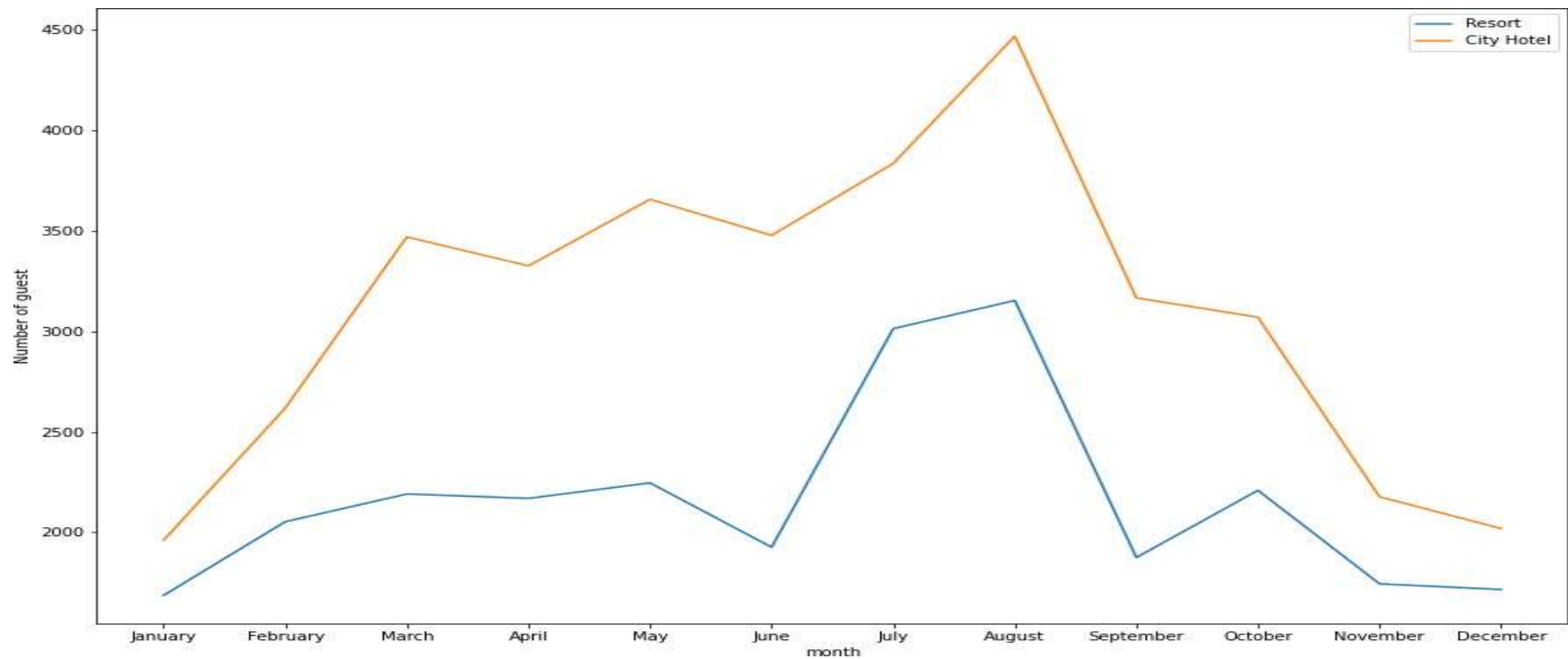
From the month of July to August the number of bookings increased and in August, City Hotel got most number of guests.

The revenue aspect looks different, the Resort Hotels receives more revenue with respect to City Hotel.

From May to August there was rapid increase in adr. August recorded the highest.

We can see that graph Arrival_num has small peaks at regular interval of days. This can be due to increase in arrival weekend.

Also, the average adr tends to goes upwards in month ends. Therefore charges are more at the end of months.
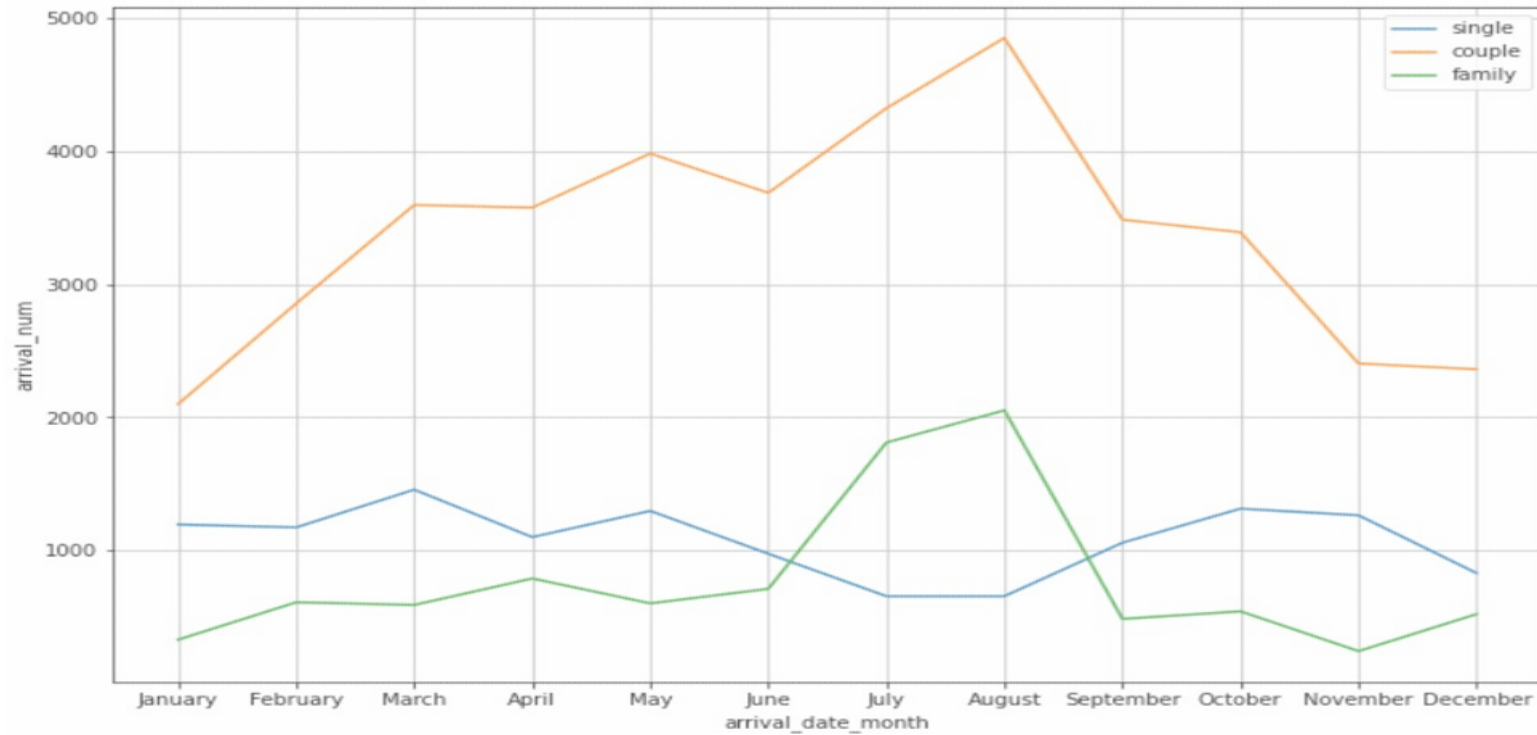
Mostly bookings are done by couples.

It is clear from graph that there is a sudden surge in arrival num of couples and family in months of July and August.

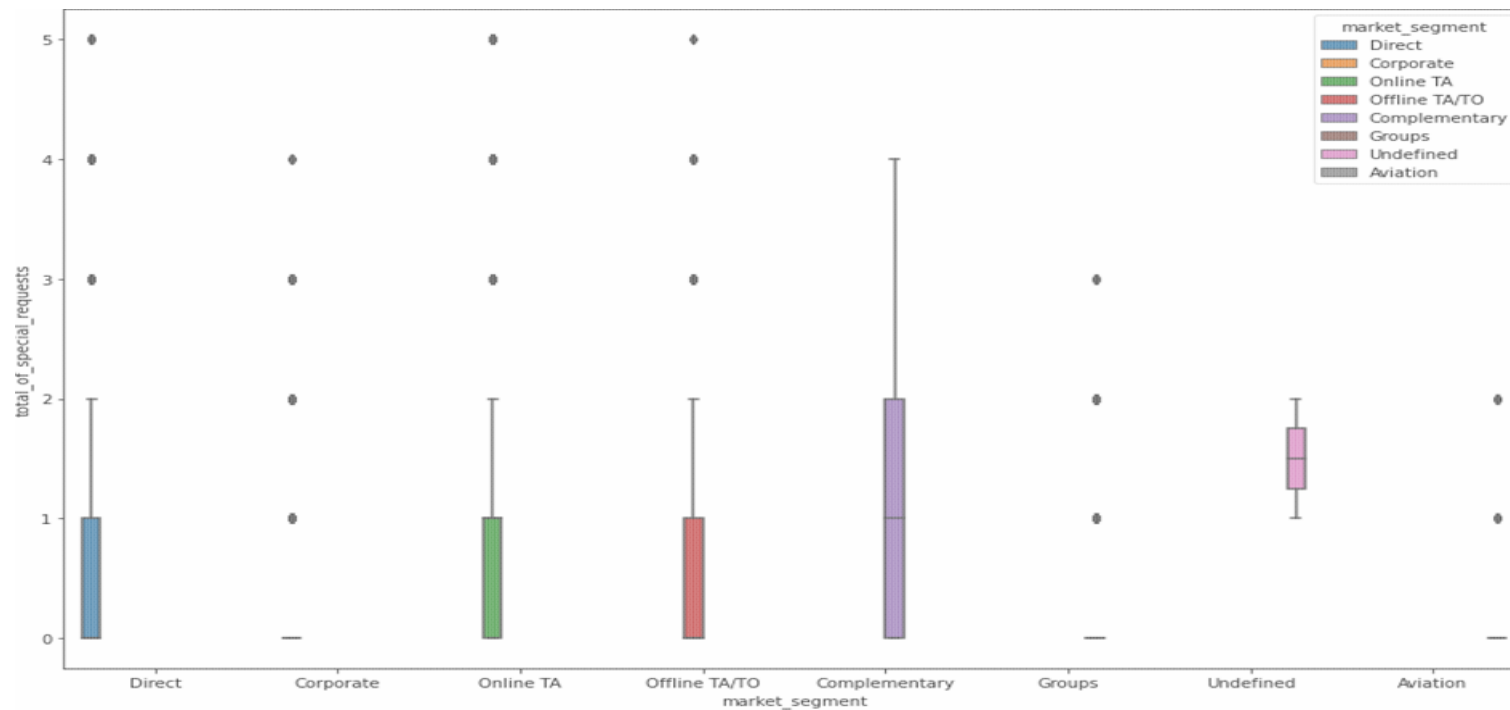So better plans can be planned accordingly at that time for these type of customers.

Some important questions

Some other analysis are also done, which are as follows:

- What are the different reason for special requests.
- What is the optimal stay length for better deal for customers.
- How adr is affected by total staying period in hotels.

Reasons for special requests
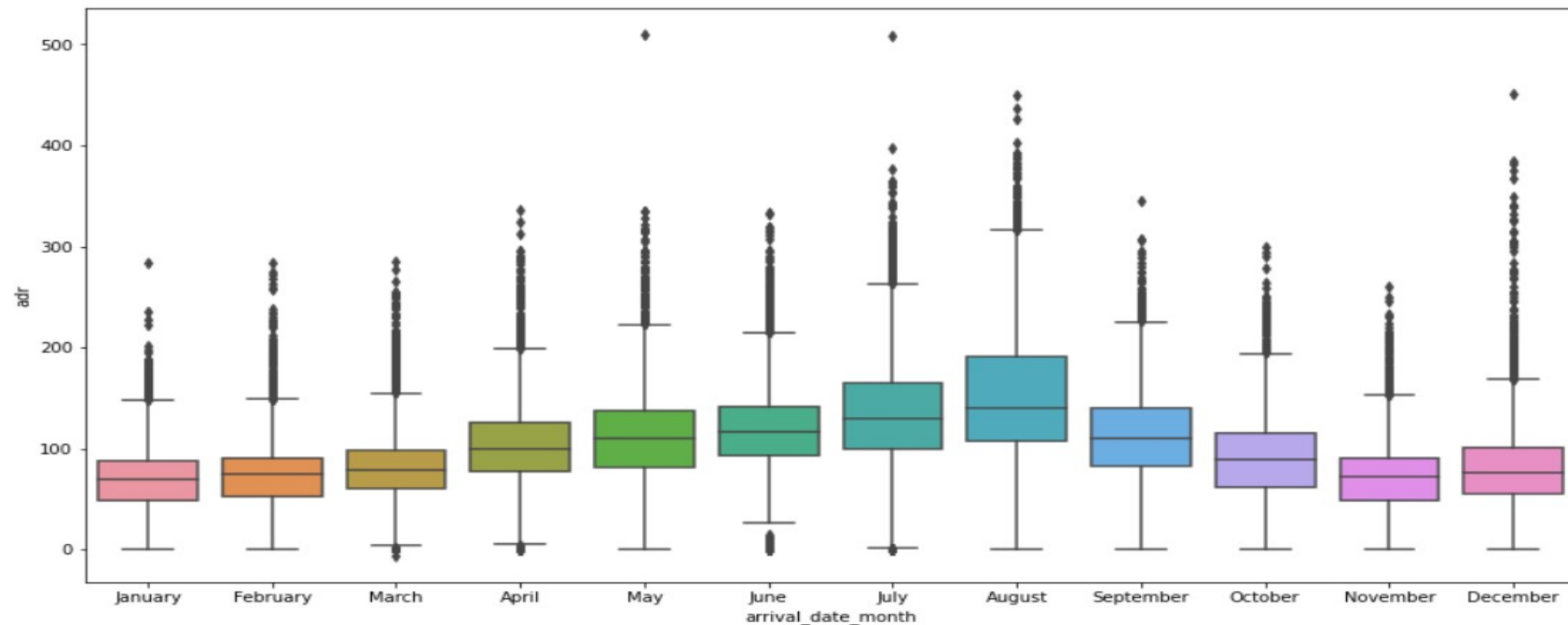

Here we can see that all market segment mostly have special request.

There is one segment which is complementary, having more than average number of special request.

Average or Adr

Average adr rises from beginning of year upto middle of year and reaches peak at August and then lowers to the end of year. But hotels do make some good deals with high adr at end of year also.

Correlation Heatmap

AI

1. Total_stay and lead_time have slight correlation. This may means that for longer hotel stays people generally plan little before the actual arrival.

2. adr is slightly correlated with total_people, which makes sense as more no of people means more revenue, therefore more adr.
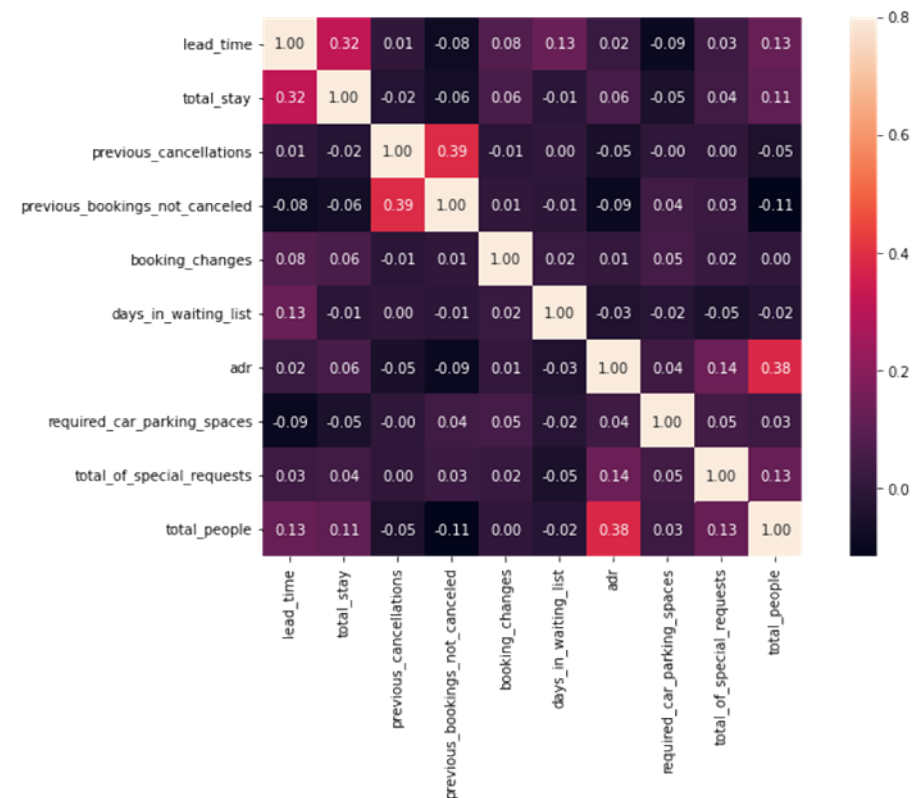
Optimal stay length for better deals in adr
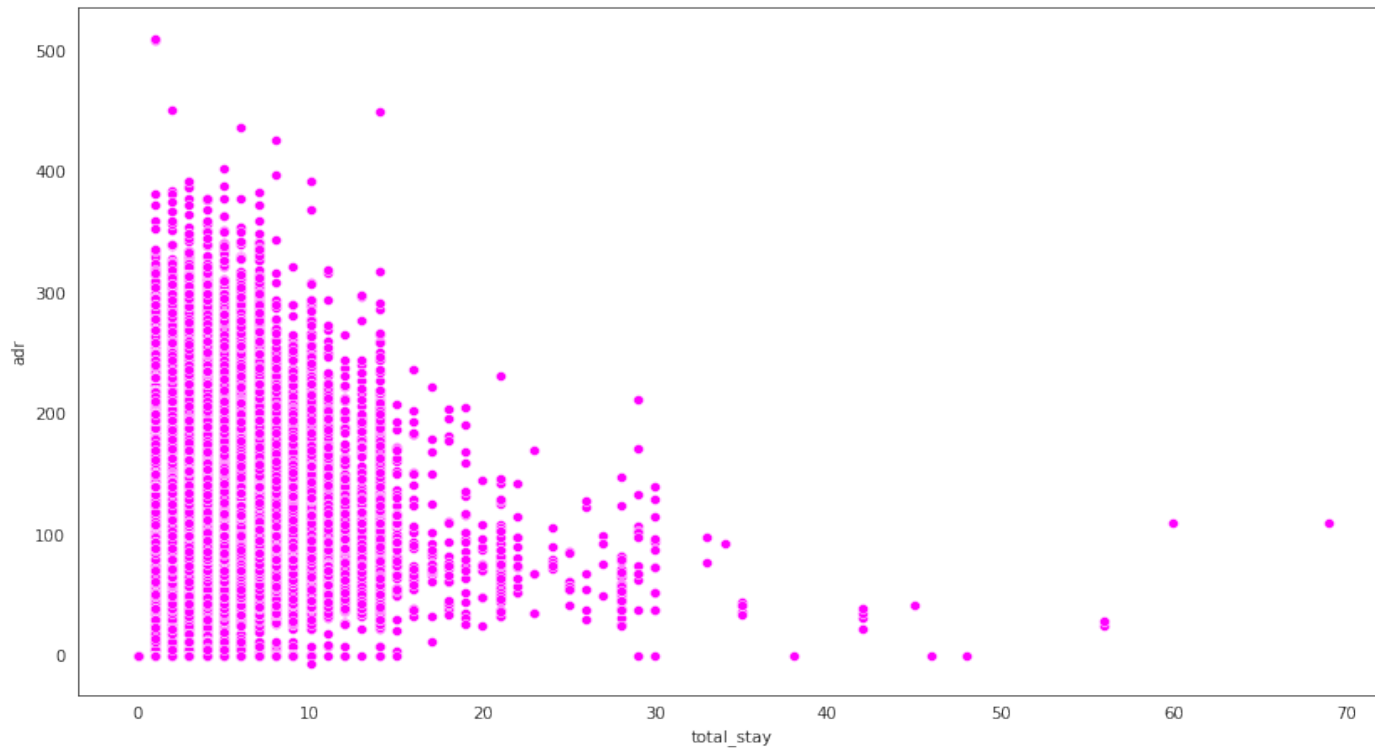
From the scatter plot we can see that as length of total_stay increases the adr decreases. This means for longer stay, the better deal for customer can be finalised.

Conclusion

- About 66% bookings are for City hotel and 34% bookings are for Resort hotel, therefore City Hotel is busier than Resort hotel. Also the overall adr of City hotel is slightly higher than Resort hotel.
- Mostly guests stay for less than 7 days in hotel and for longer stays Resort hotel is preferred.
- Both hotels have significantly higher booking cancellation rates and very few guests less than 3 % return for another booking in City hotel. 5% guests return for stay in Resort hotel.
- Most of the guests came from european countries, with most no. of guest coming from Portugal.
- Guests use different channels for making bookings out of which most preferred way is TA/TO.
- For hotels higher adr deals come via GDS channel, so hotels should increase their popularity on this channel.
- Almost 30% of bookings via TA/TO are canceled.
- Not getting same room as reserved, longer lead time and waiting time do not affect cancellation of bookings. Although different room allotment do lowers the adr.
- July- August are the most busier and profitable months for both of hotels.
- Within a month, adr gradually increases as month ends, with small sudden rise on weekends.
- Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.
- More number of people in guests results in more number of special requests.
- Bookings made via complementary market segment and adults have on average high no. of special request.
- For customers, generally the longer stays (more than 15 days) can result in better deals in terms of low adr.

Thank You