

CAPSTONE PROJECT

HOTEL BOOKING ANALYSIS
SRIKANYA SAHA

Points to Discuss:

- Agenda
- Data summary
- Univariate analysis
- Hotel wise analysis
- Distribution Channel wise analysis
- Booking cancellation analysis
- Timewise analysis
- Some important questions
- Correlation heatmap
- Conclusion

Agenda

To discuss and analyze the given hotel booking data set from 2015-2017. We'll be doing the analysis of given data set in following ways :

- Univariate analysis
- Hotel wise analysis
- Distribution Channel wise analysis
- Booking cancellation analysis
- Timewise analysis

By doing this we will be trying to find out some key factors driving the hotel booking trends.

Data Summary

Given data set has different columns of variables crucial for hotel booking.
Some of them are:

hotel: We have two type of hotels which are resort hotel and city hotel.

is_cancelled : The value of this column shows the cancellation type. When the booking

is cancelled the value is [1] and when the booking is not cancelled the value is 0.

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 guests.

Data Summary(contd..)

market_segment: This column shows how reservation was made and what is the purpose of the reservation. Eg, corporate means corporate trip, TA for travel agency etc.

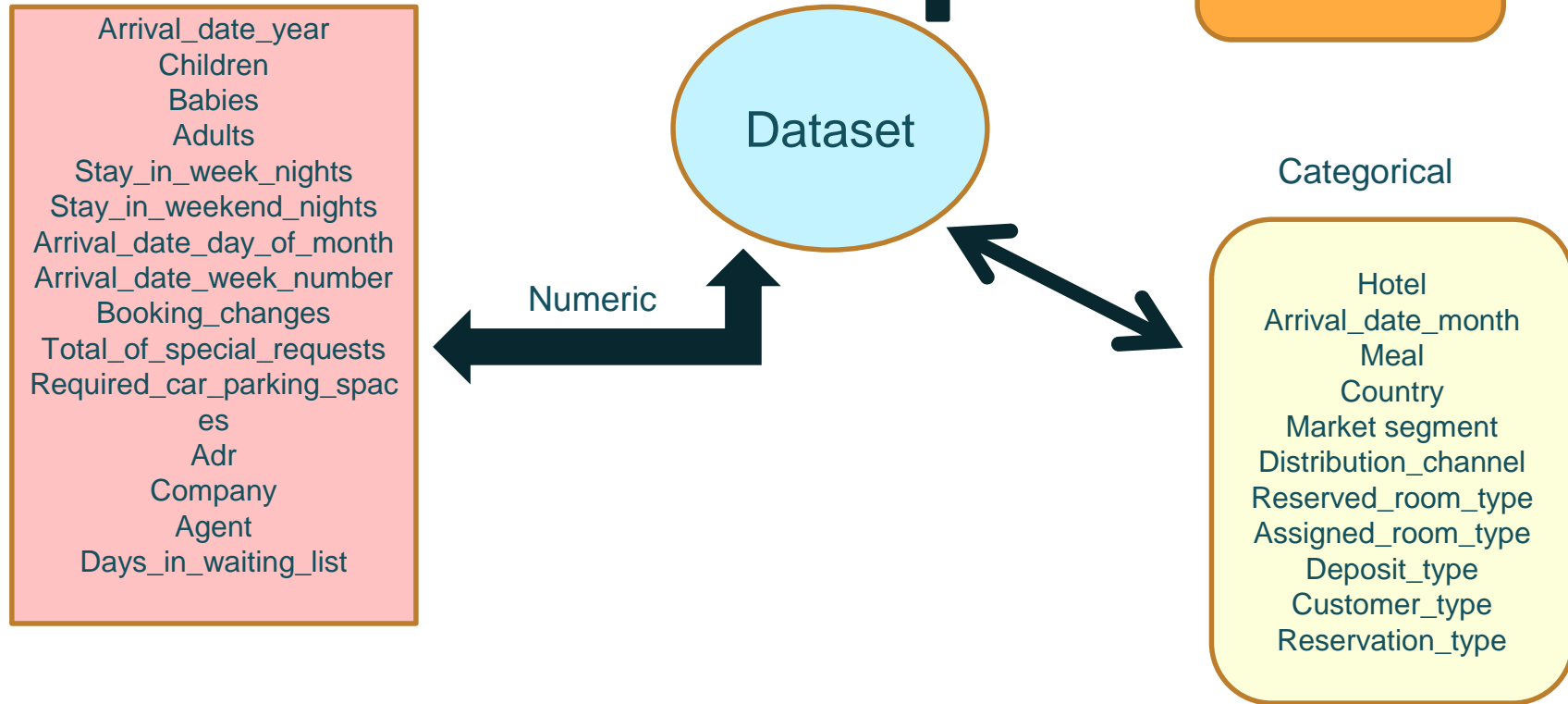
distribution_channel: It is the medium through which booking was made.[Direct,Corporate,TA/TO,undefined,GDS.]

Is_repeated_guest: Shows whether a guest has visited the hotel before or not.Values[0,1]-->0 indicates no and 1 indicates yes (the person is a repeated guest).

days_in_waiting_list: Number of days between the actual booking and the transaction.

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

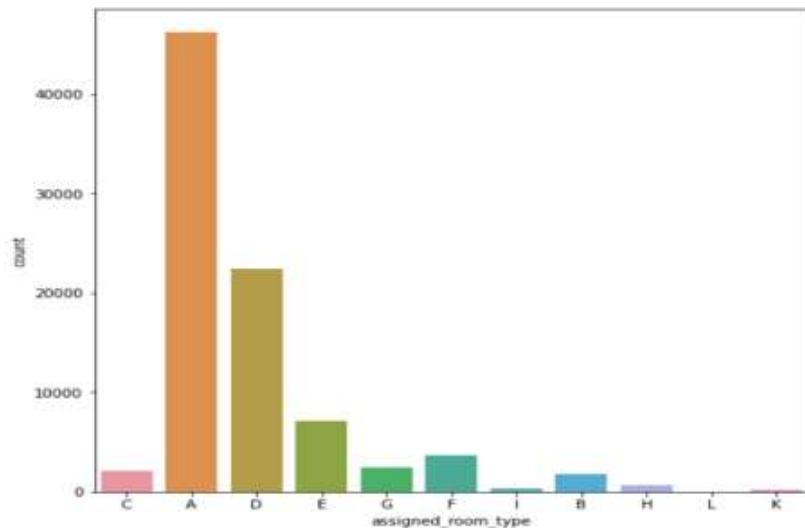
Data Summary



Univariate Analysis

While doing univariate analysis of the given hotel booking dataset, we will answer these following questions:

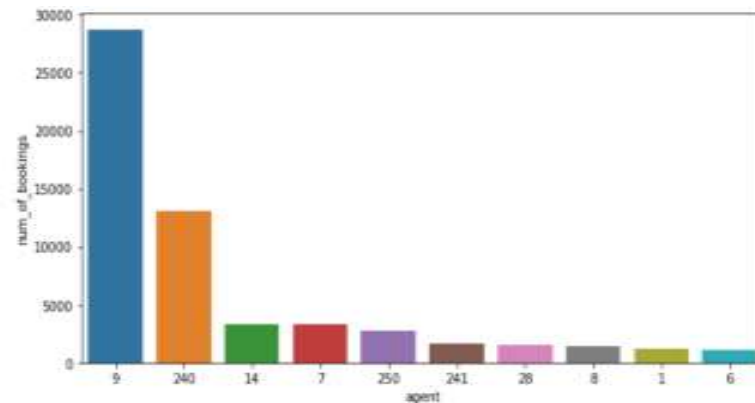
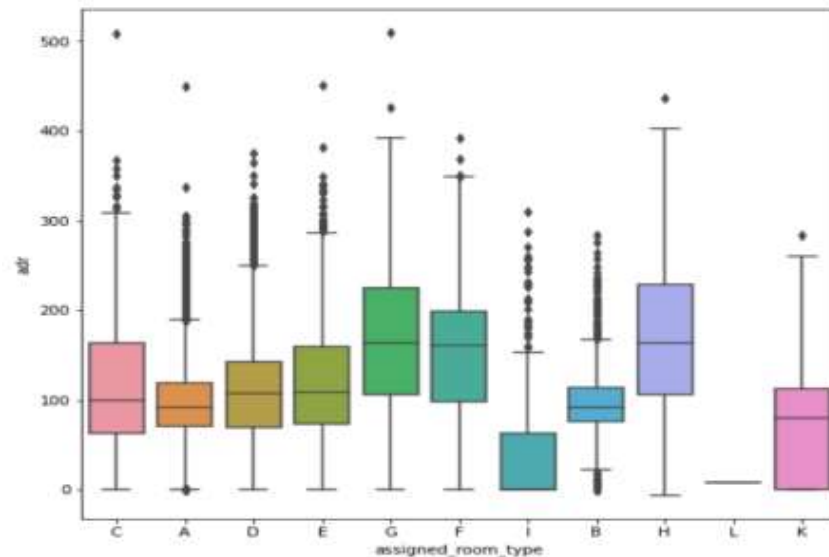
- (1) Which agent has made the most of the bookings?
- (2) Which room type is the most demanding one and which room type generates the highest adr?
- (3) From which country the most of the guests are coming?
- (4) What is the most preferred meal by the guests?



Type A room is the most demanding room type among guests.

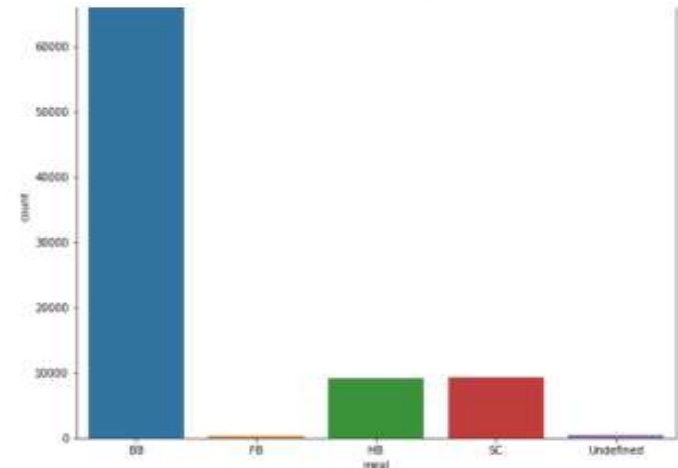
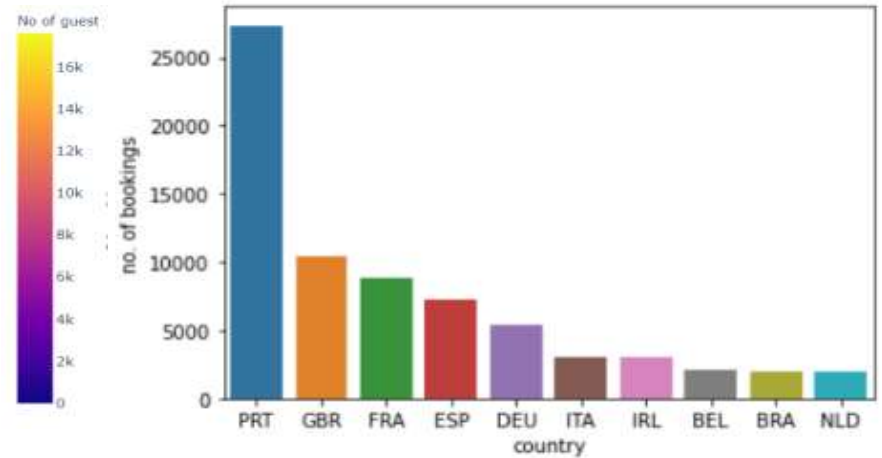
Room types C, G and H are some of the highest adr(average daily rate) generating rooms.

Agent with an id no. 9 has made the most of the bookings





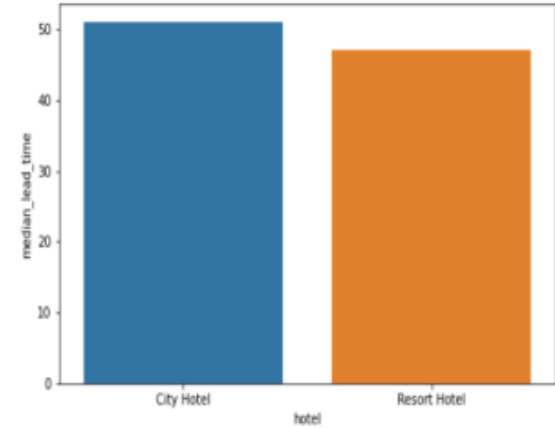
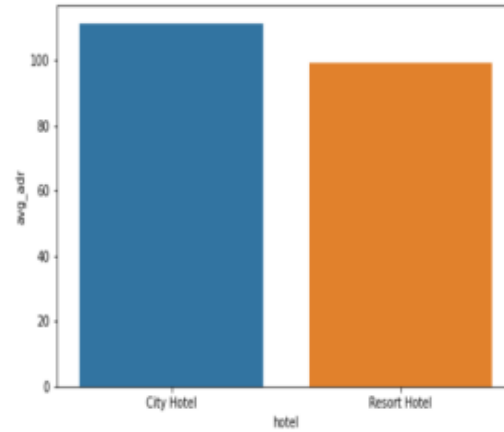
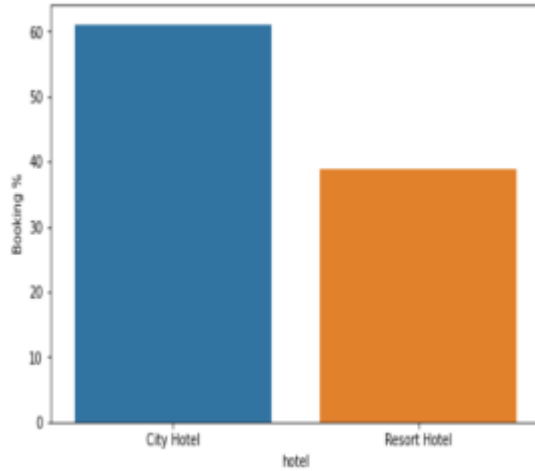
- Most of the customers from European countries like Portugal, Great Britain, France and Spain.
- Most preferred meal type is BB(Bed and breakfast



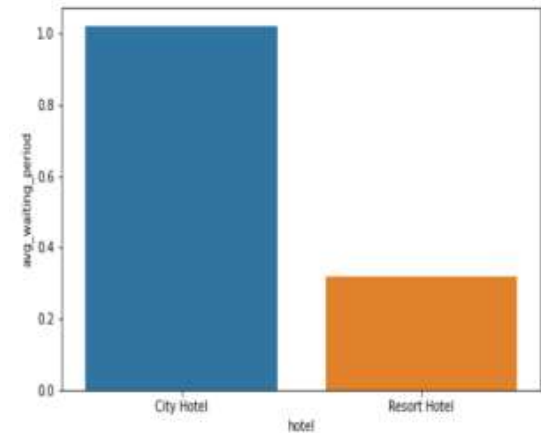
Hotel wise Analysis

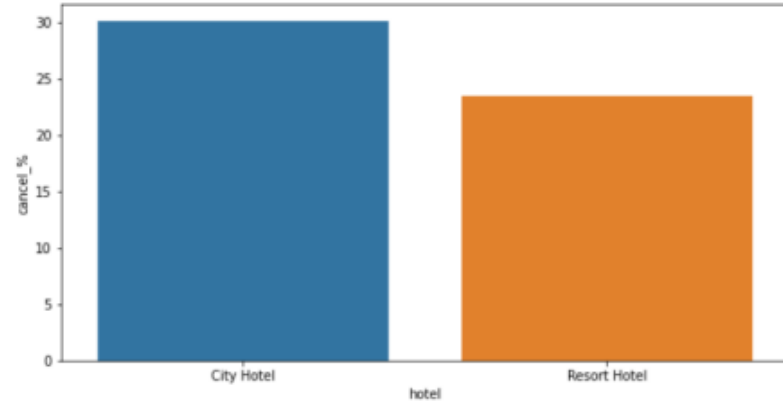
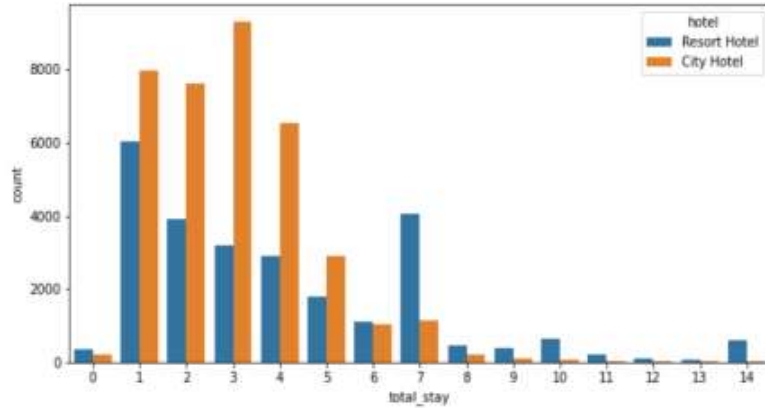
While doing hotel-wise analysis of the given hotel booking dataset, we are going to answer these following questions:

- (1) Percentage of bookings in each hotel?
- (2) Which hotel has made the highest revenue?
- (3) Which hotel has higher lead time?
- (4) What is the most preferred stay length in each hotel?
- (5) For which hotel, people have to wait longer to get the booking confirmation?
- (6) Which hotel has the highest booking cancellation rate?
- (7) Which hotel has the highest customer returning rate?

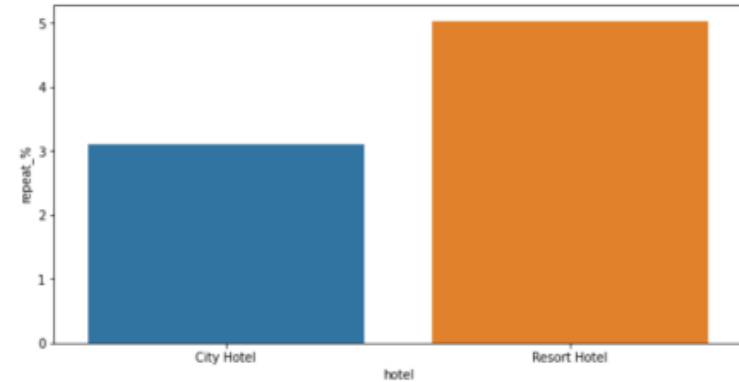


- Around 60% bookings are for the City hotel and 40% bookings are for the Resort hotel.
- Avg adr of Resort hotel is slightly lower than that of City hotel. Hence, City hotel seems to be making slightly more revenue.
- City hotel has slightly higher median lead time. Also median lead time is significantly higher in each case, that means customers generally plan their hotel visits way to early.
- City hotel has significantly longer waiting time, hence City Hotel is much busier than Resort Hotel





- Most of 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.
- Both the hotels have very small percentage of customers who return to stay again,. But Resort hotel has slightly higher repeat percentage than City hotel.

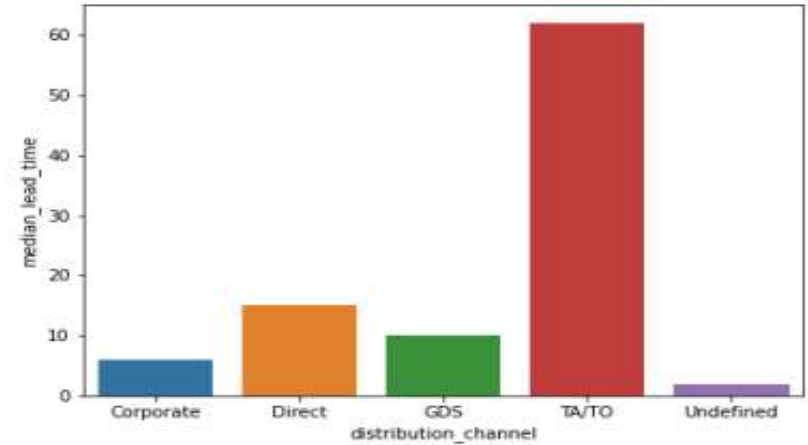
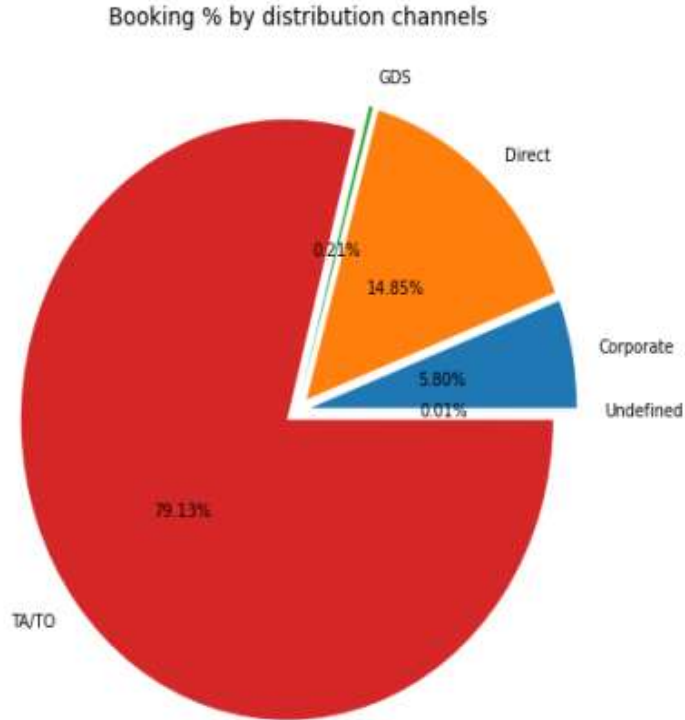


Distribution channel wise Analysis

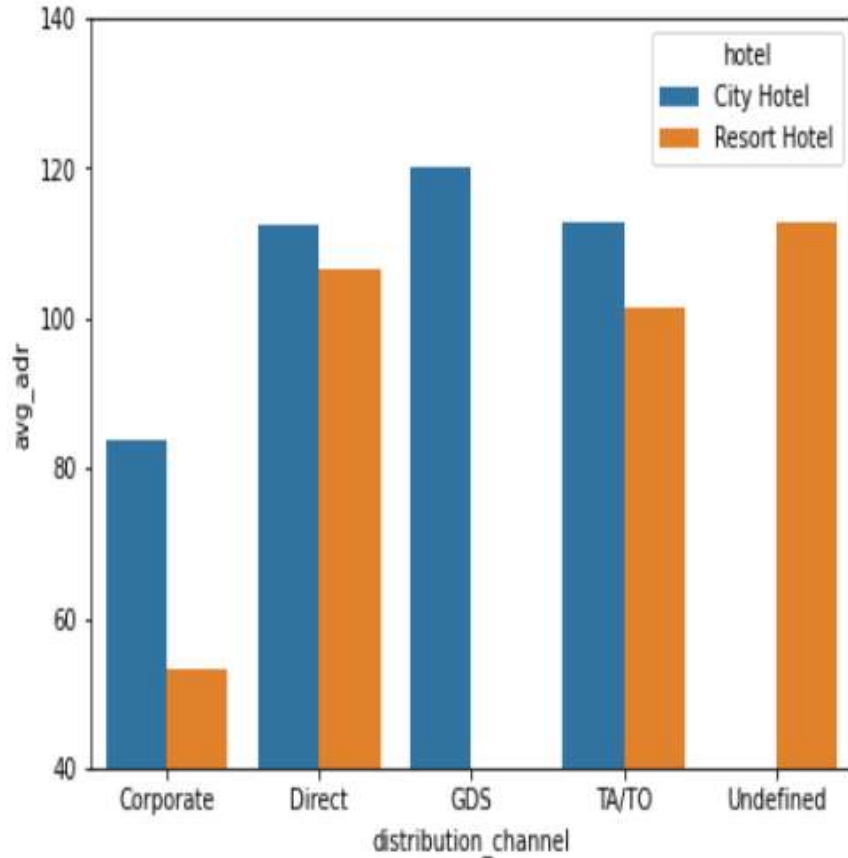
While doing Distribution channel wise analysis of given hotel booking dataset, we will answer following questions:

- (1) Which is the most common channel for hotel booking?
- (2) Which channel is mostly used for early booking of hotels?
- (3) Which distribution channel brings better revenue generating deals for hotels?

Distribution channel wise Analysis



- Here we can see that the most of guest are making reservation through TA/TO channels which is travel agency and tour operator.
- Than the second most used channel is direct.
- Channel which is mostly used for early booking of hotels is also TA/TO



- GDS channel brings the highest revenue

generating deals for the City hotel, in contrast to that the most of the bookings have come via TA/TO.

City Hotel should work to increase the outreach of GDS channels to get more higher revenue generating deals.

- Resort hotel has more revenue generating

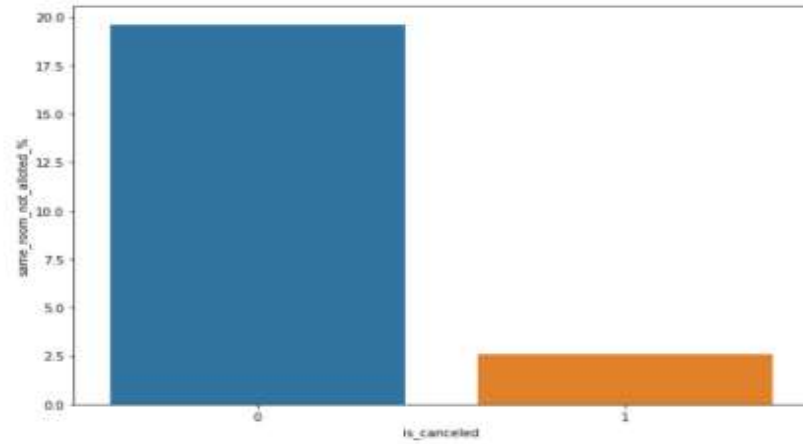
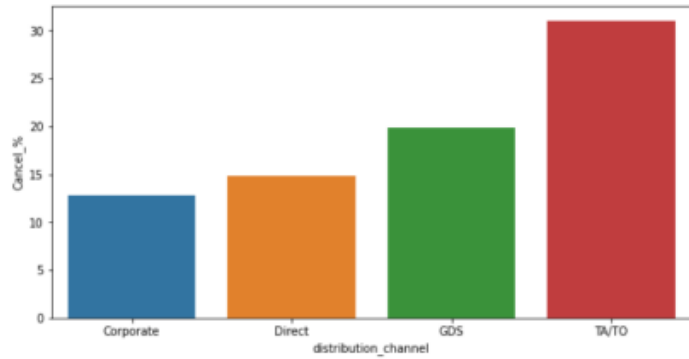
deals by direct and TA/TO channel. Resort

Hotel needs to increase its outreach of GDS channel to increase revenue.

Booking cancellation Analysis

Now we will analyze the following possible reasons for booking cancellations:

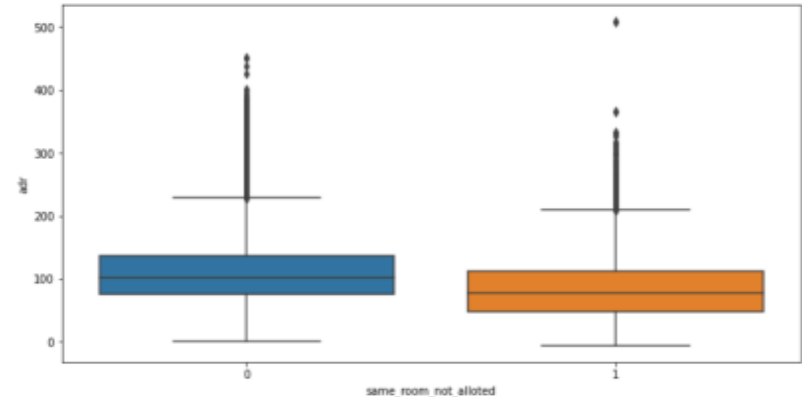
- (1) Which significant distribution channel has the highest cancellation percentage?
- (2) Longer lead time.
- (3) Longer time (in days) in waiting list.
- (4) Not getting the same room as demanded.
- (5) Is not getting the same room as demanded affecting the adr or not?

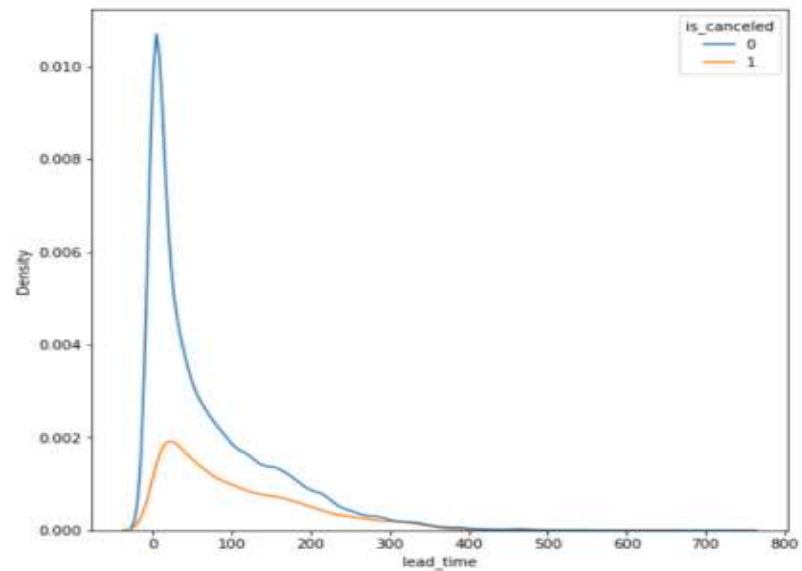
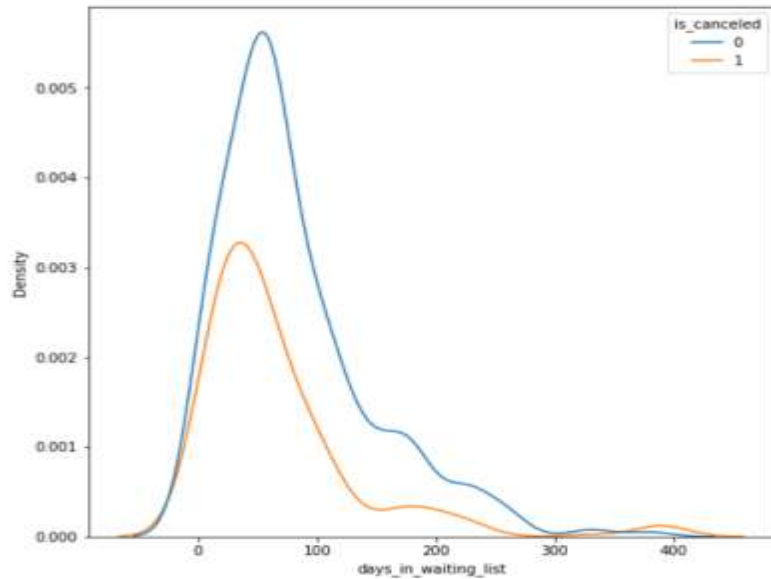


- TA/TO has the highest booking cancellation %. Therefore, a booking via TA/TO is 30% likely to get cancelled.

- Not getting the 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.

- But, customers who didn't get the same room as demanded have paid a little lower adr, except for a few exceptions.



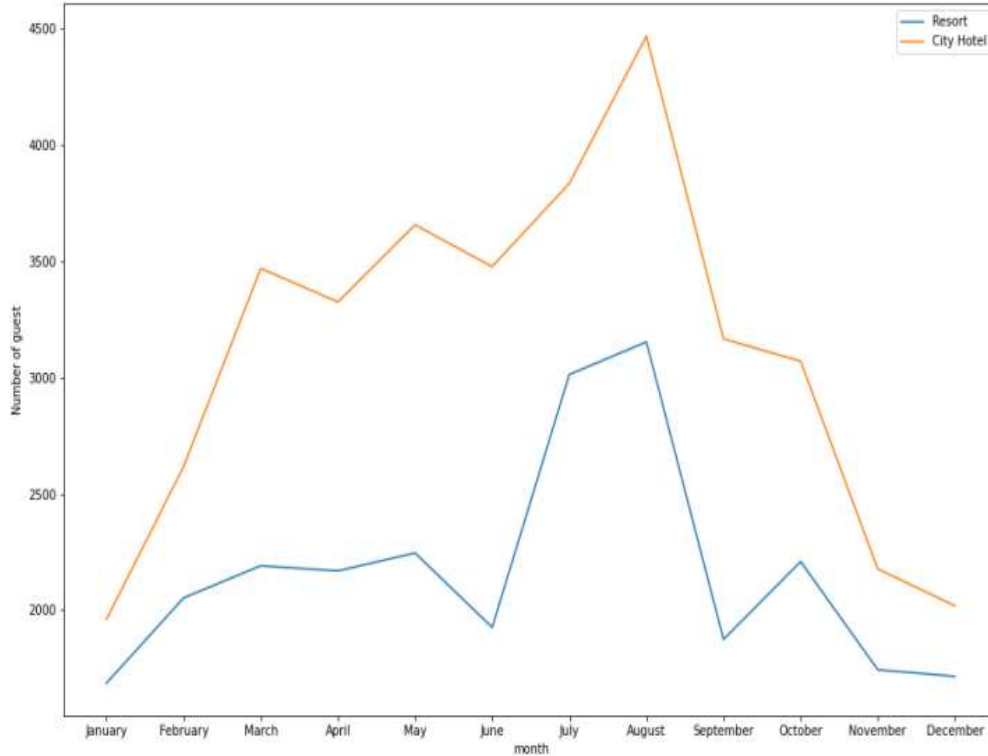


- Most of the bookings that are cancelled have waiting period of less than 150 days but also most of bookings that are not cancelled also have waiting period of less than 150 days. Hence this shows that waiting period has no effect on cancellation of bookings.
- Also, lead time has no effect on cancellation of bookings, as both curves are similar for lead time too.

Time-wise Analysis

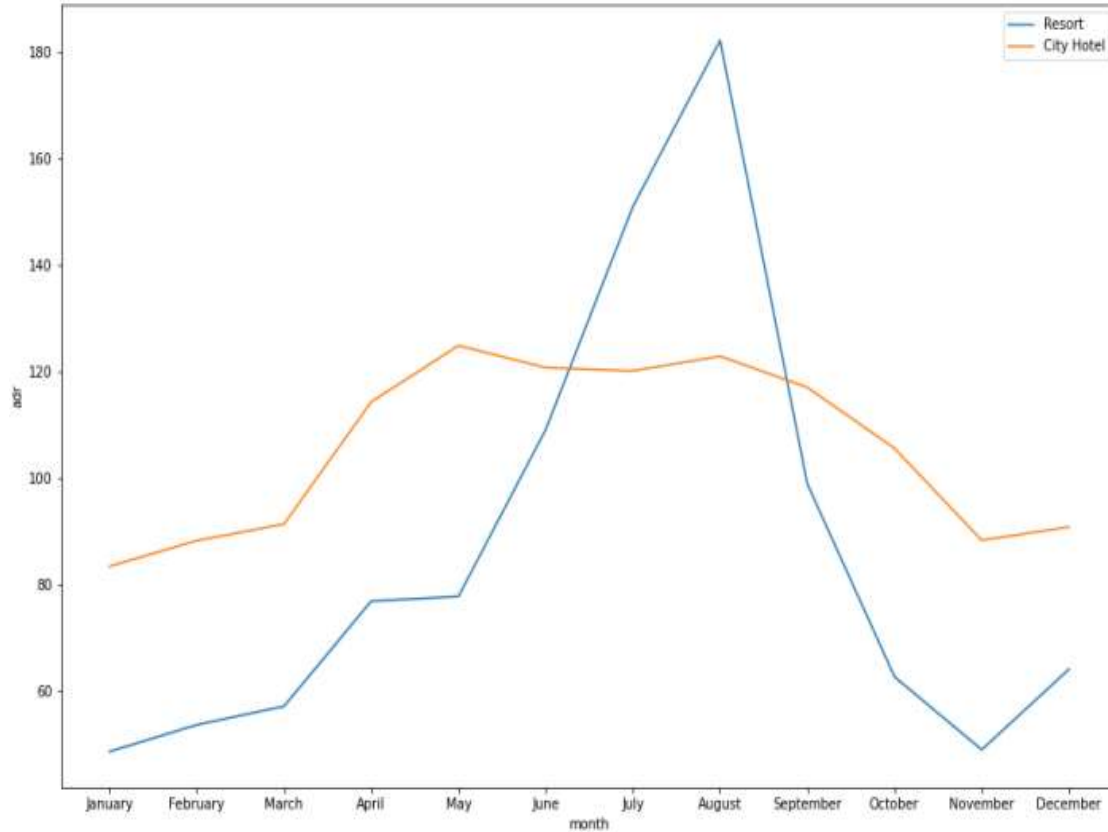
While doing time-wise analysis of given hotel booking dataset, we will answer these following questions:

- (1) What are the most busy months for the hotels?
- (2) In which month of the year hotels charge the highest adr?
- (3) How booking numbers and adr change along with months?
- (4) How bookings vary along with the months of the year for different types of customers?



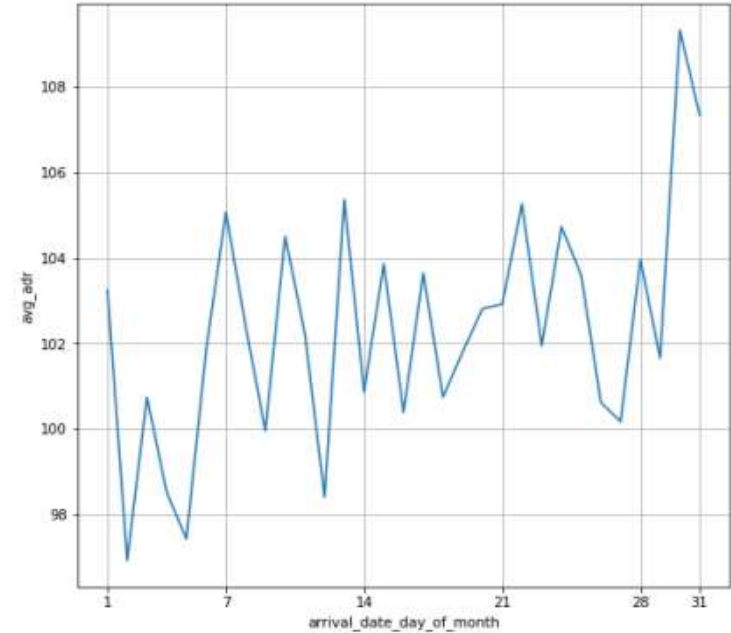
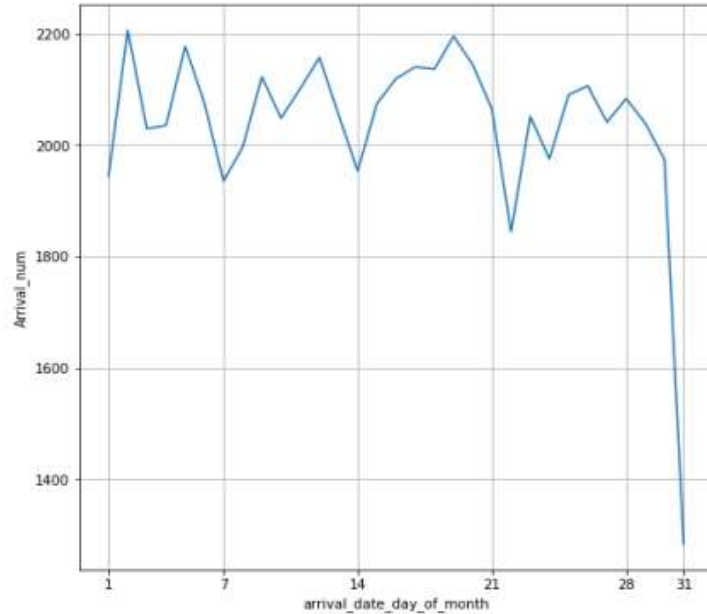
From the month of July to August the number of bookings increased and in August, City Hotel got the most number of guests.

And Resort hotel also got its highest no of guests around the month of august.

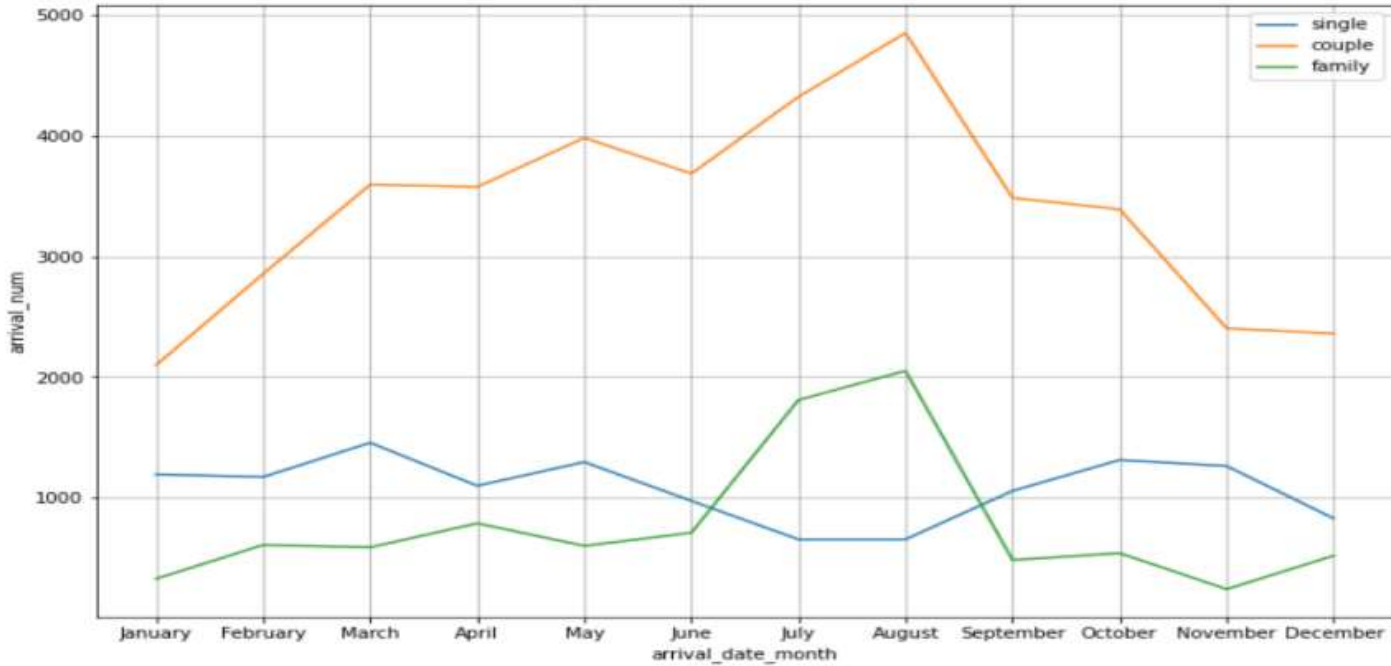


The revenue aspect looks different. The Resort Hotel receives more revenue with respect to City Hotel.

There is a rapid increase in adr from May to August . August recorded the highest adr for resort hotel.



We can see that graph Arrival_num has small peaks at regular interval of days. This can be due to increase in arrival in weekends. Also, the avg adr tends to go up as the month ends. Therefore charges are more at the end of month



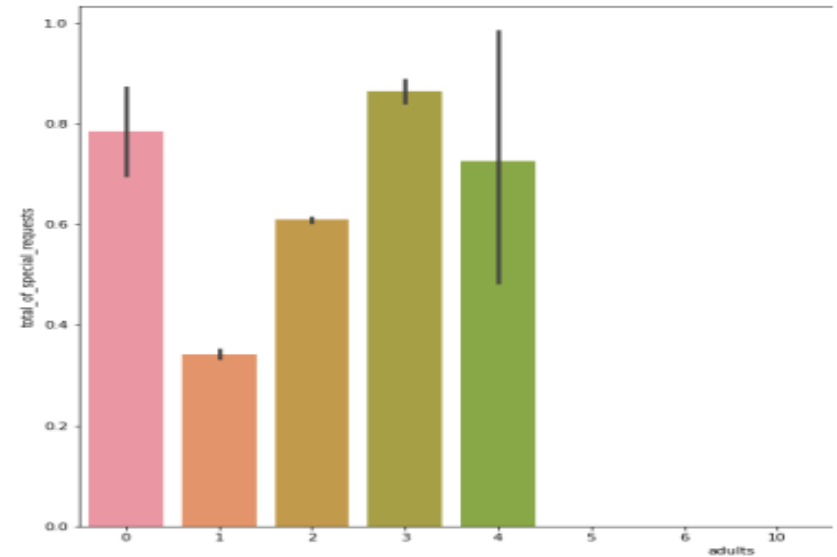
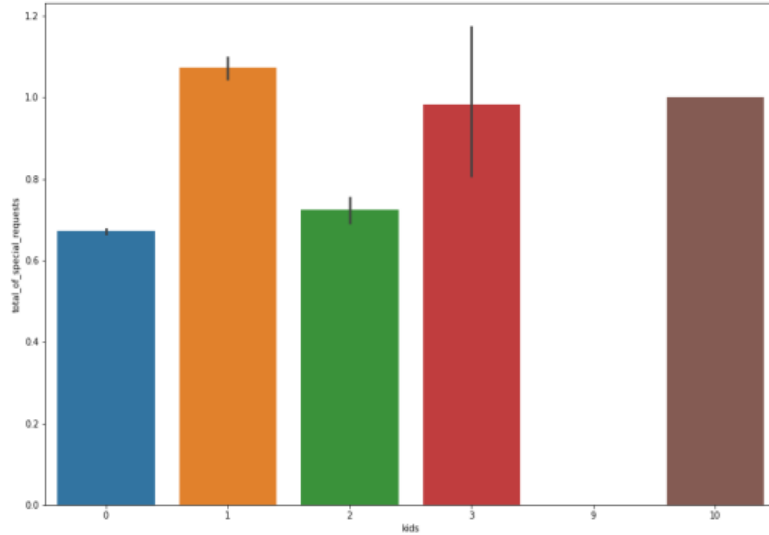
Mostly bookings are done by couples. It is clear from graph that there is a sudden surge in arrival num of couples and family in months of July and August. So better plans can be implimented accordingly at that time for these type of customers.

Some important questions

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

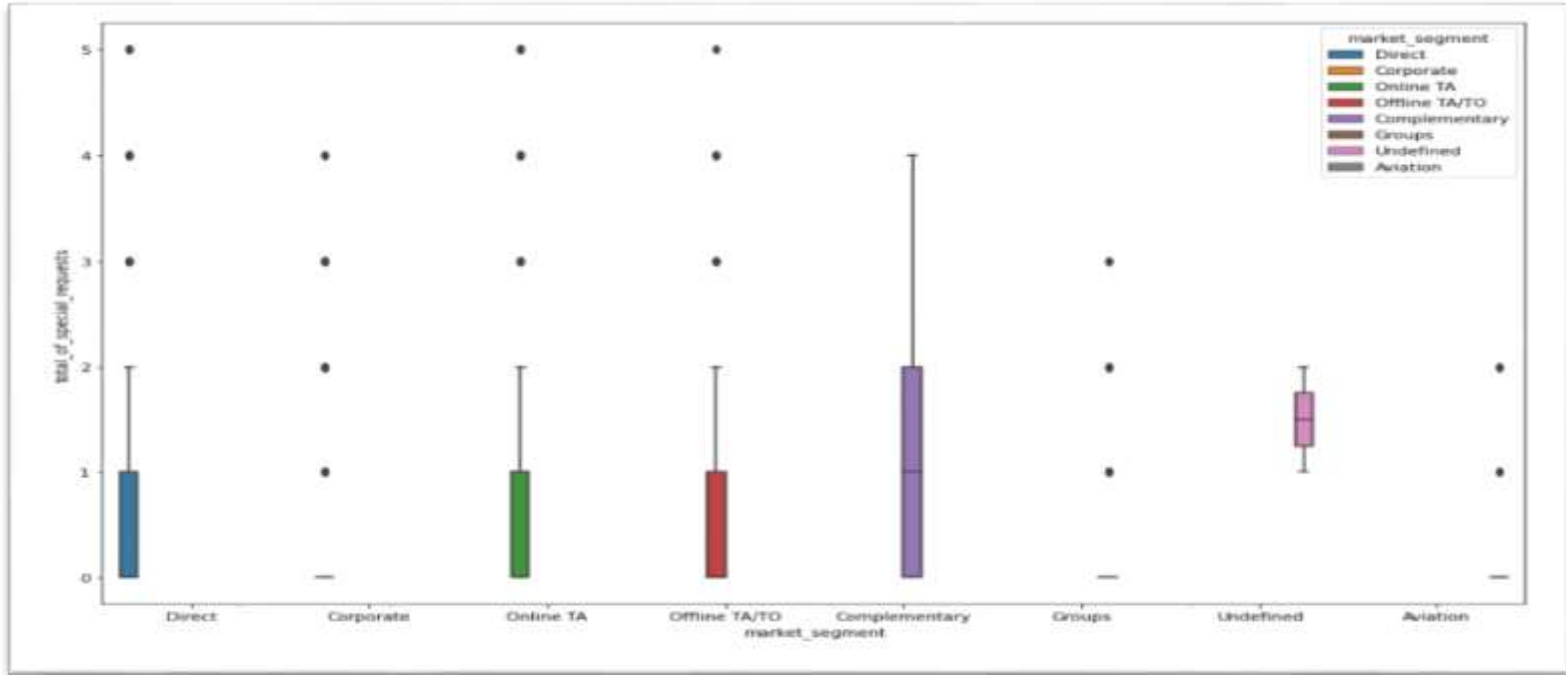
- (1) What are the different reasons for special requests?
- (2) What is the optimal stay length of customers for better deal?
- (3) How adr is affected by total staying period in hotels.

Reasons for special requests



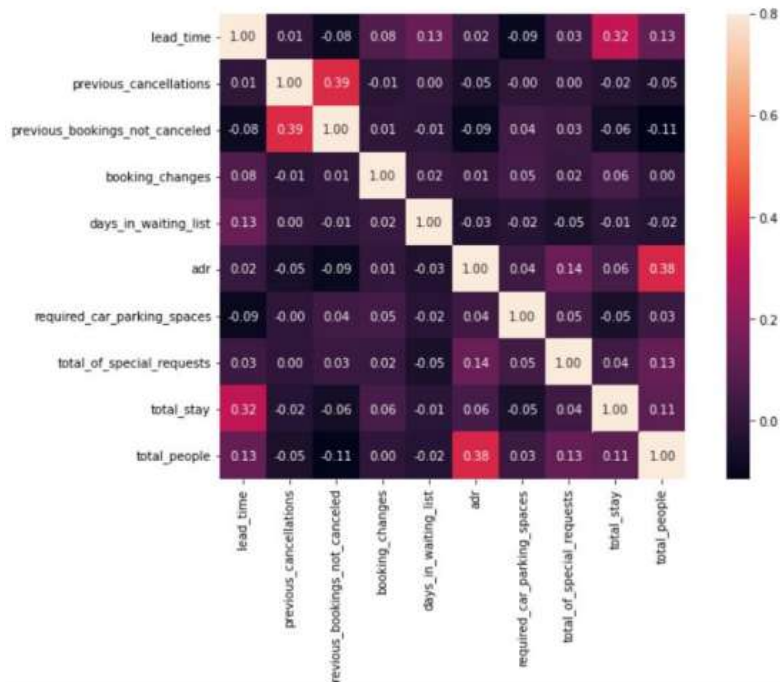
The number of special request are almost the same in the kids section. In adult section we can see that if the adults are 3 in number, then there are more chances that hotels will receive more special requests.

Reasons for special requests(cont.)



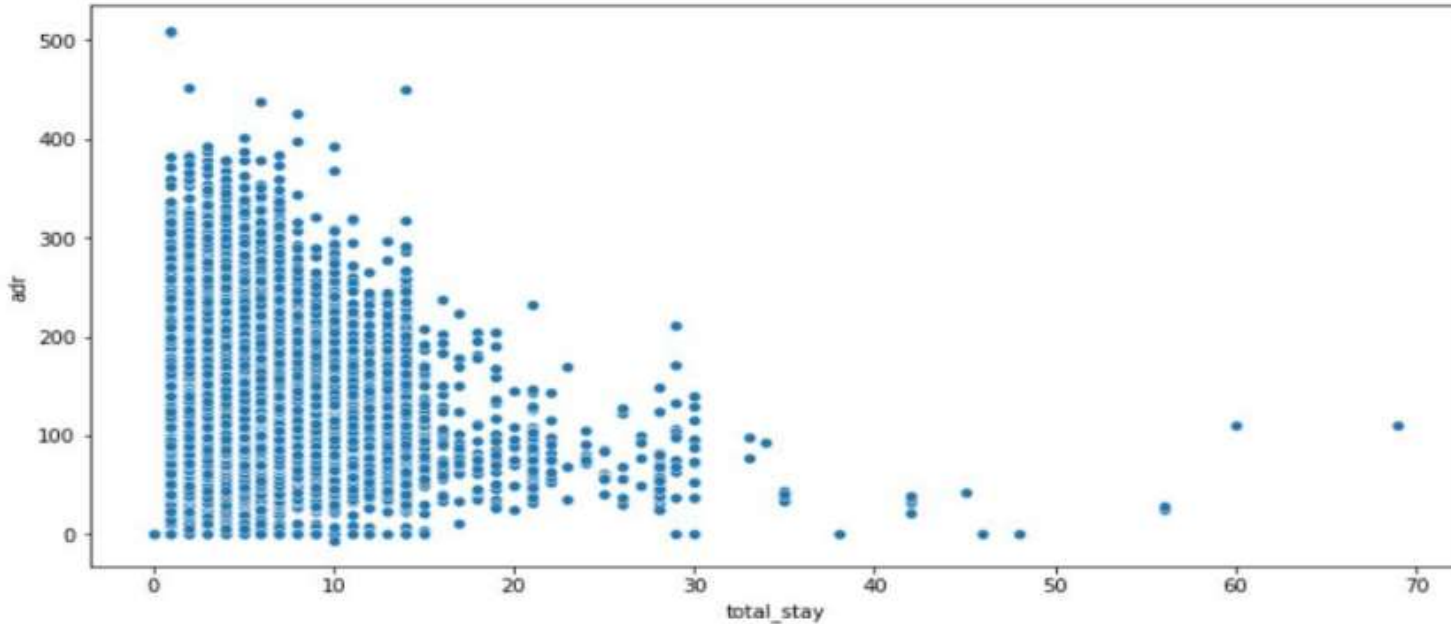
Here we can see that all market segment mostly have special request. There is one segment which is complementary, having more than average number of special request.

Correlation Heatmap



- Total stay length and lead time are slightly correlated. This may mean that people generally plan a little before for longer stays than the actual arrival.
- adr is slightly correlated with total_people, which makes sense as more no. of people means more service to deliver, therefore more adr.

Optimal stay length for better deals



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

- Around 60% bookings are for City hotel and 40% 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 5 days in the hotels and for longer stays Resort hotel is preferred.
- Both the hotels have significantly higher booking cancellation rates. For city hotel very few guests (less than 3 %) return to stay again. 5% guests return for an another 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 the most preferred way is TA/TO.
- Higher adr deals come via GDS channel for city hotel., So city hotel should increase its popularity on this channel. And for Resort hotel, higher adr deals come via direct and TA/TO channel.
- Almost 30% of bookings via TA/TO are cancelled.

- 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 generally have more special request.
- For customers, generally the longer stays (more than 15 days) result in better deals in terms of adr.

THANK YOU