Capstone project
Hotel Booking Analysis

Chilaka naveen

Points to Discuss:

- Agenda
- Data summary
- Univariate analysis
- Hotel wise analysis
- Distribution channel wise analysis
- Booking cancellation analysis
- Timewise analysis
- Correlation heatmap
- Conclusion

Agenda

To discuss the analysis of given hotel bookings data from 2015-2017.

We'll be doing 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'll try to find out key factors driving the hotel bookinh trends

Data Summary

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

Hotel: The category of hotels, which are two resorts 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,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 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.)

Univariate Analysis

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

1. Which agent made most of bookings?
2. Which room type is in most demand and which room type generates highest adr?
3. From which country most of the customers are coming?
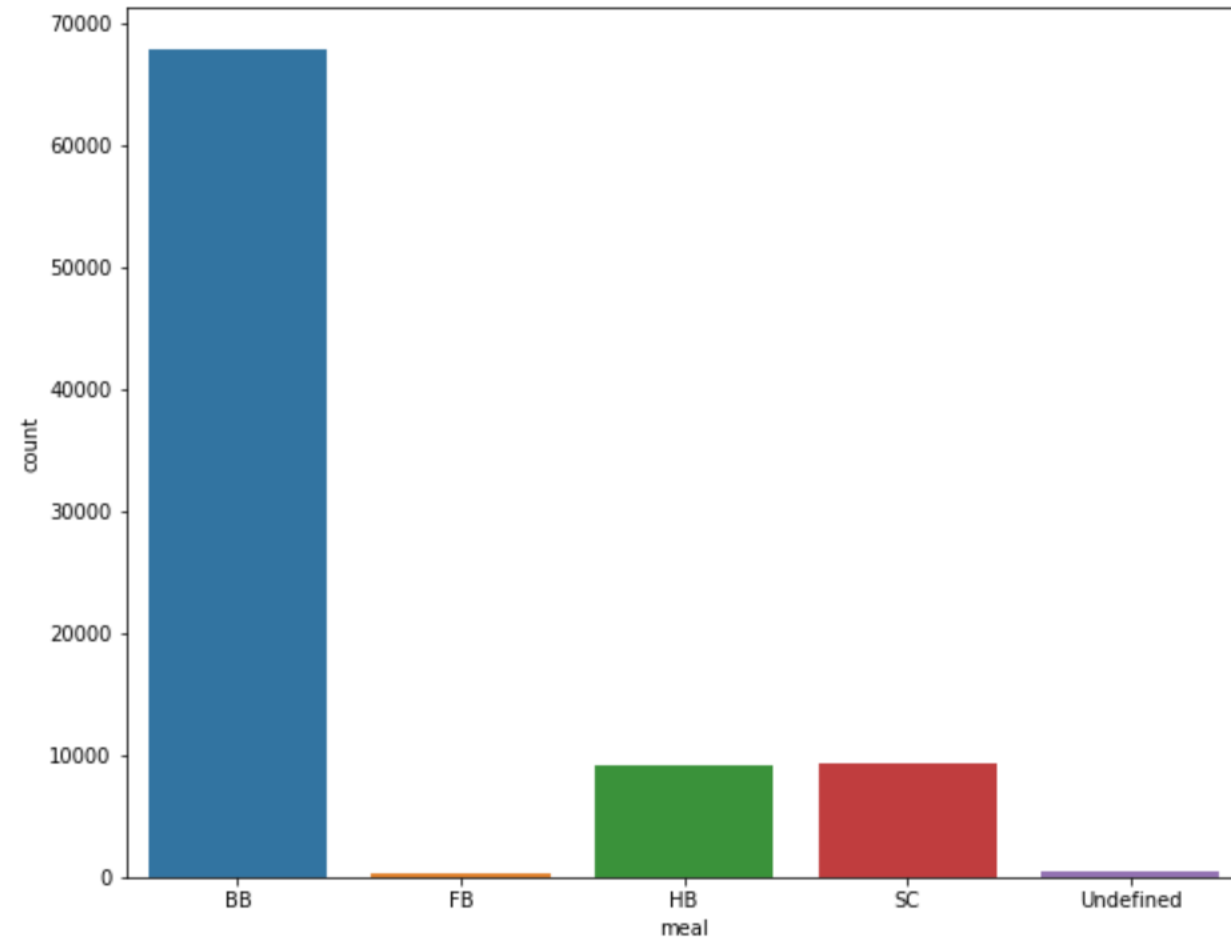4. What is the most preferred meal by customers?
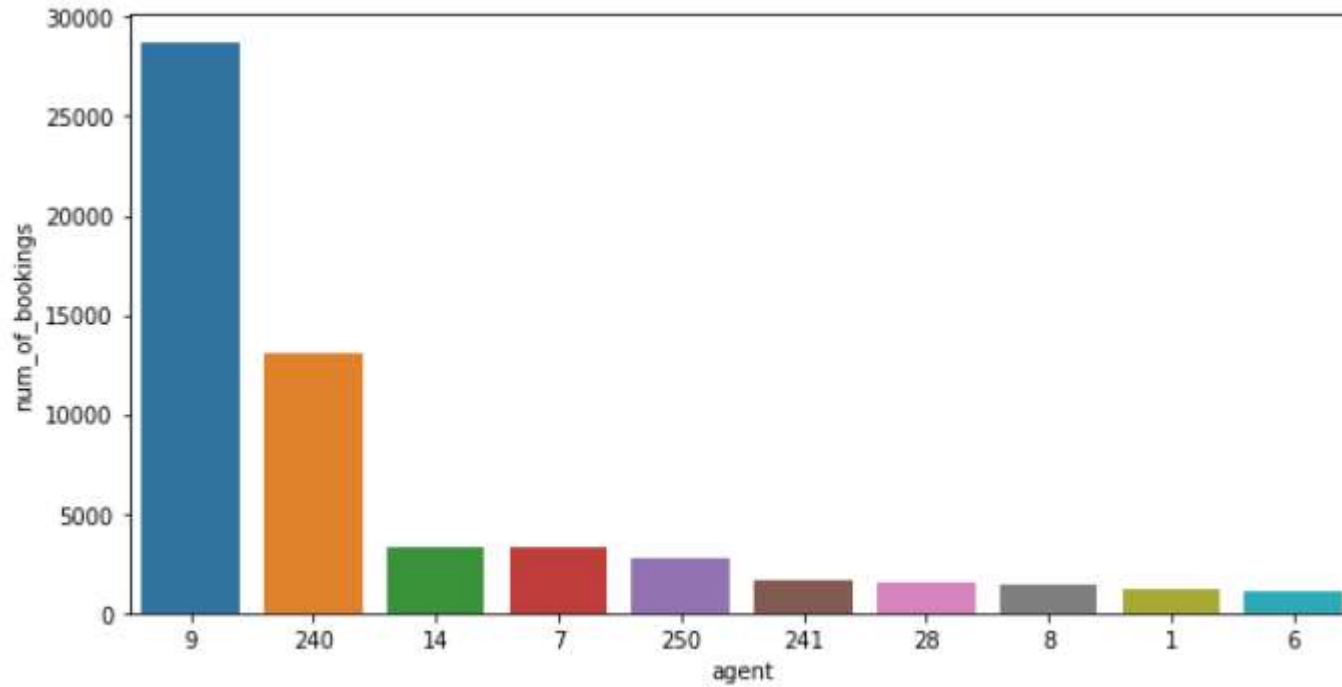
Agent no. 9 has made most no. of bookings.
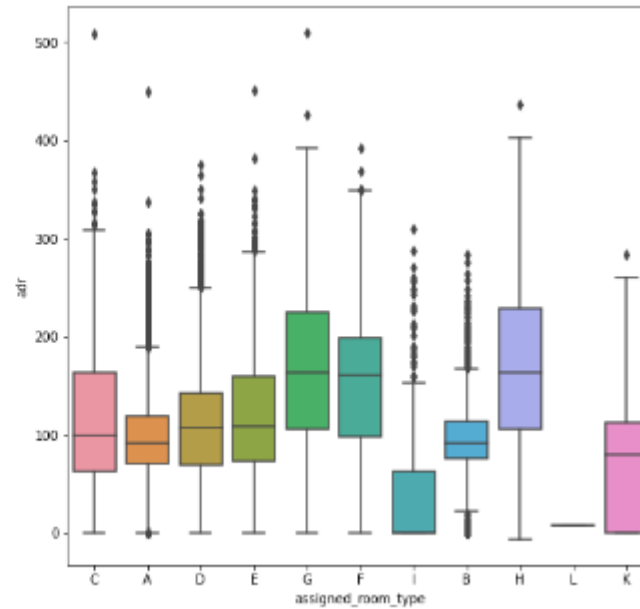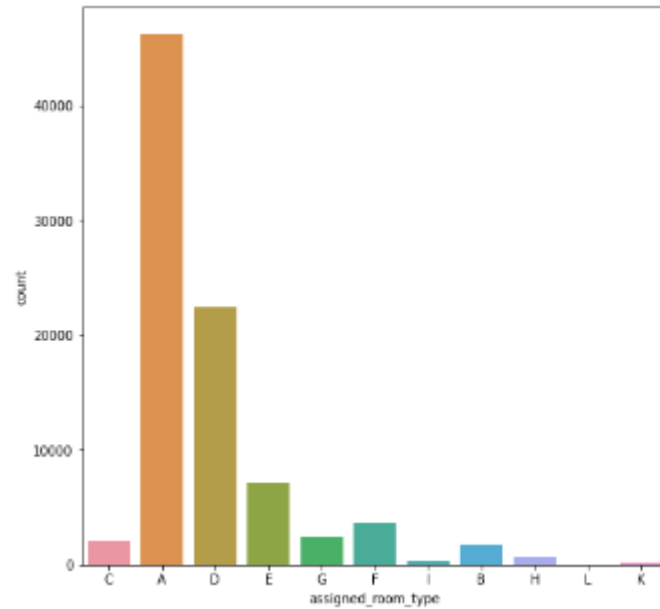
- Most demanded room type is A, but better adr room are of type H, G and C also. Hotels should increase the no. of room types A and H to maximise revenue.
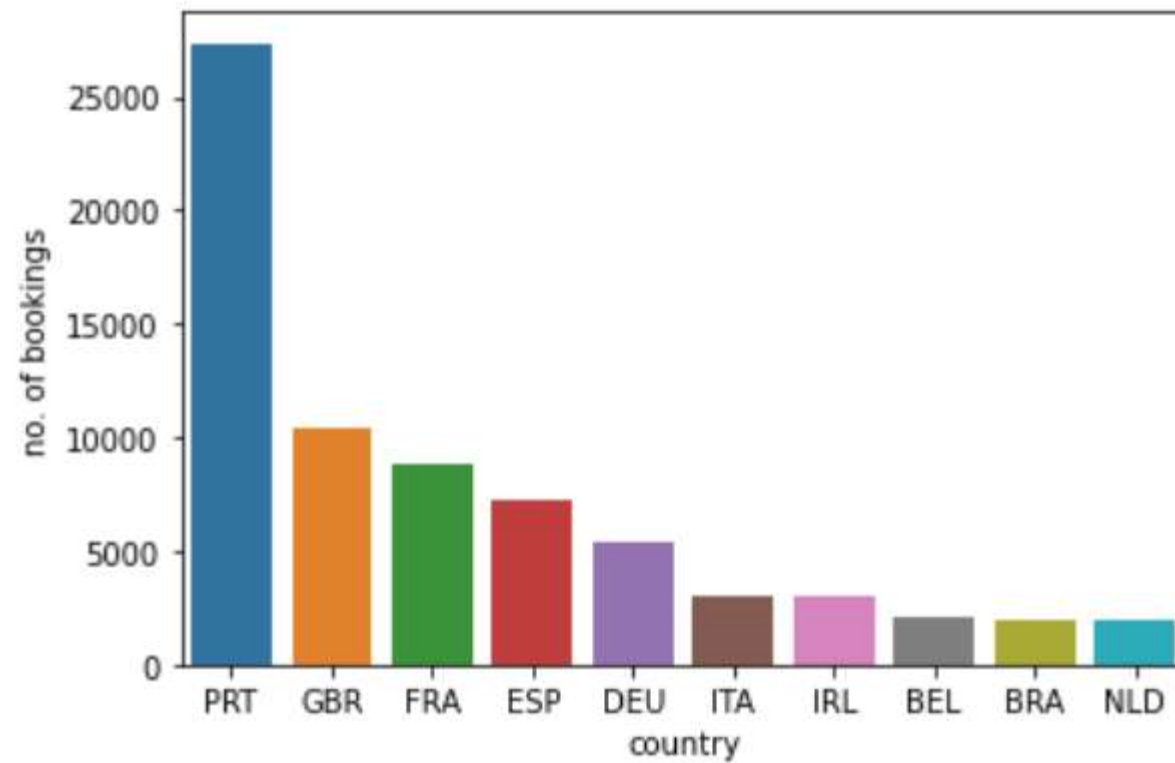
- Most of the customers comes from portugal, Great Britain, France and Spain
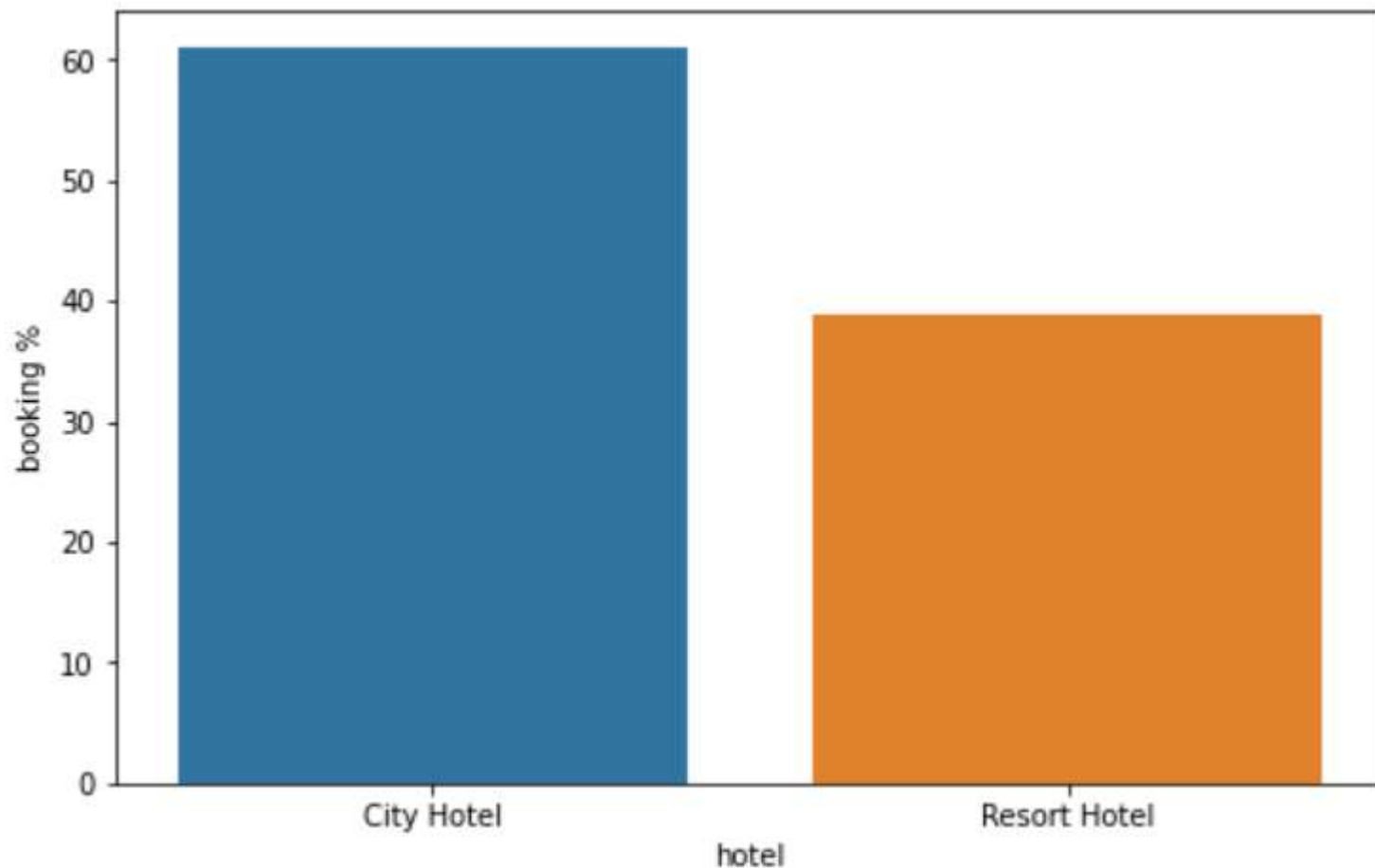
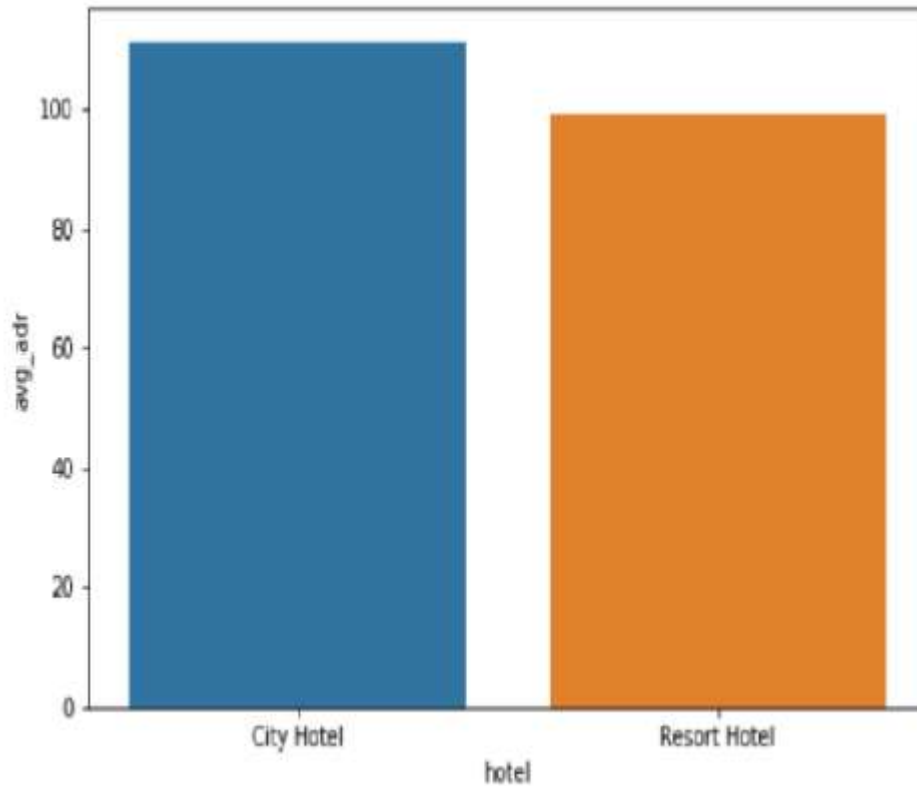- Most preferred meal type is BB(bed and breakfast)

Hotel wise analysis

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

1. Percentage of bookings in each hotels?

2. Which hotel makes more revenue?

3. What is the preferred stay length in each hotel?

4. Which hotel has higher booking cancellations rate?

- In this we see around 60% booking are for city hotel and 40% booking are for 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.
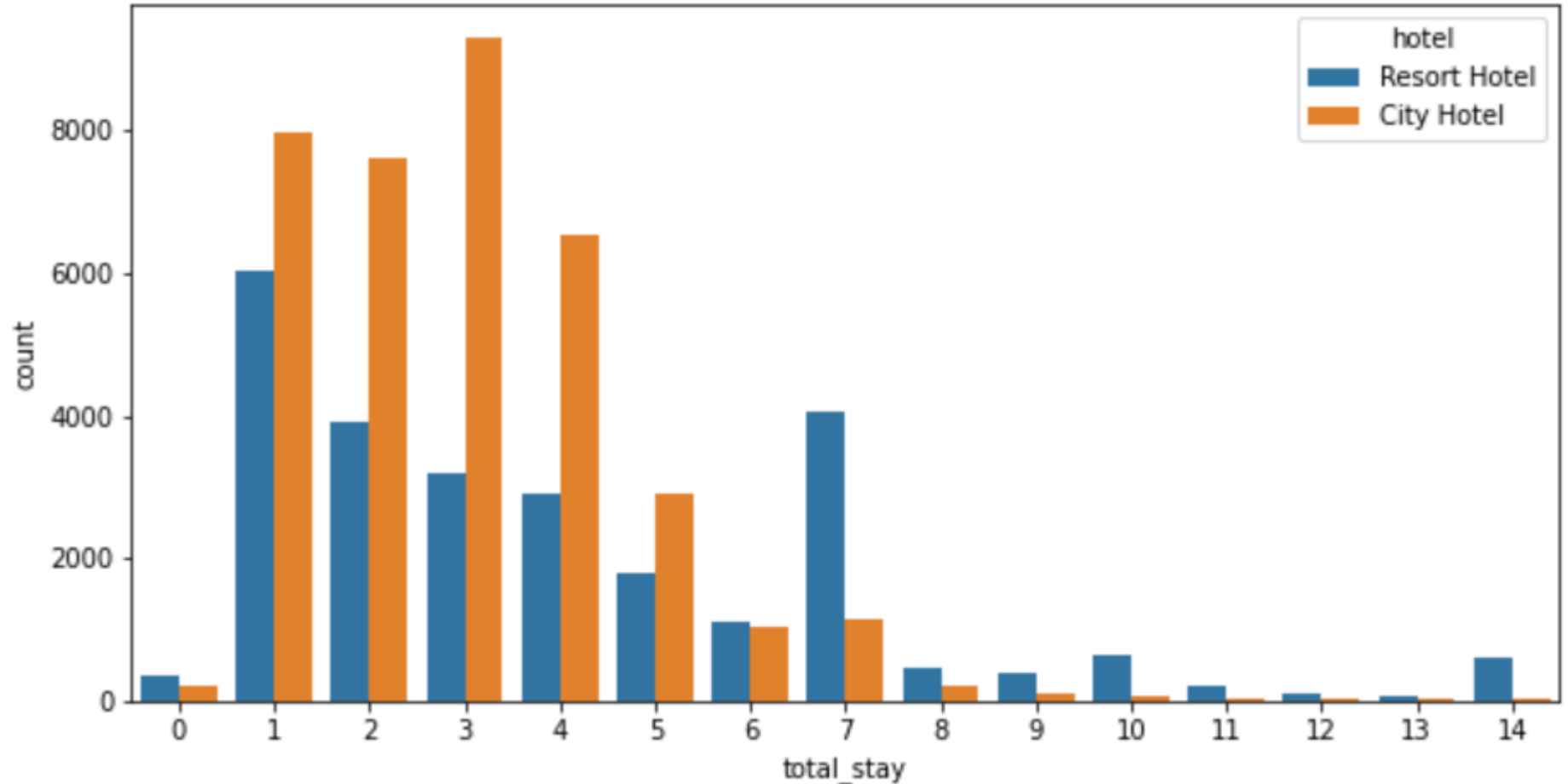
- We observe that most common stay length is less than 4 days and generally people prefer city hotel for short stay, resort hotel is preferred.
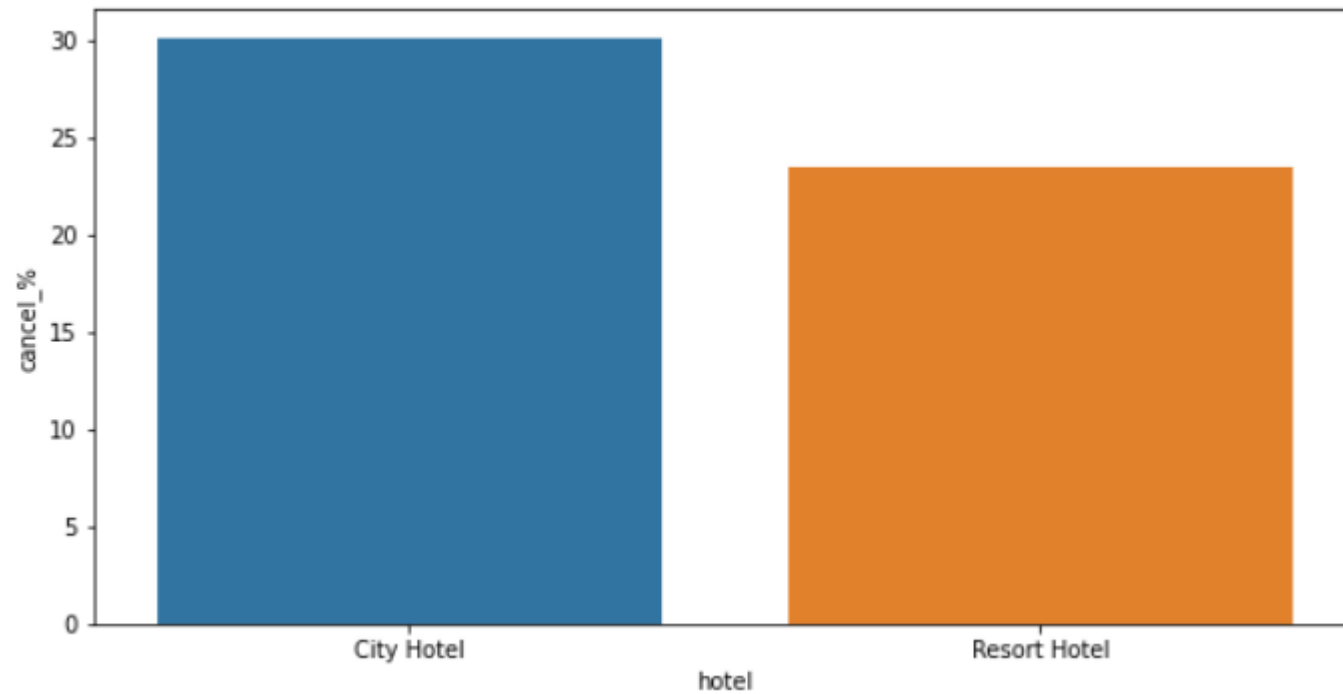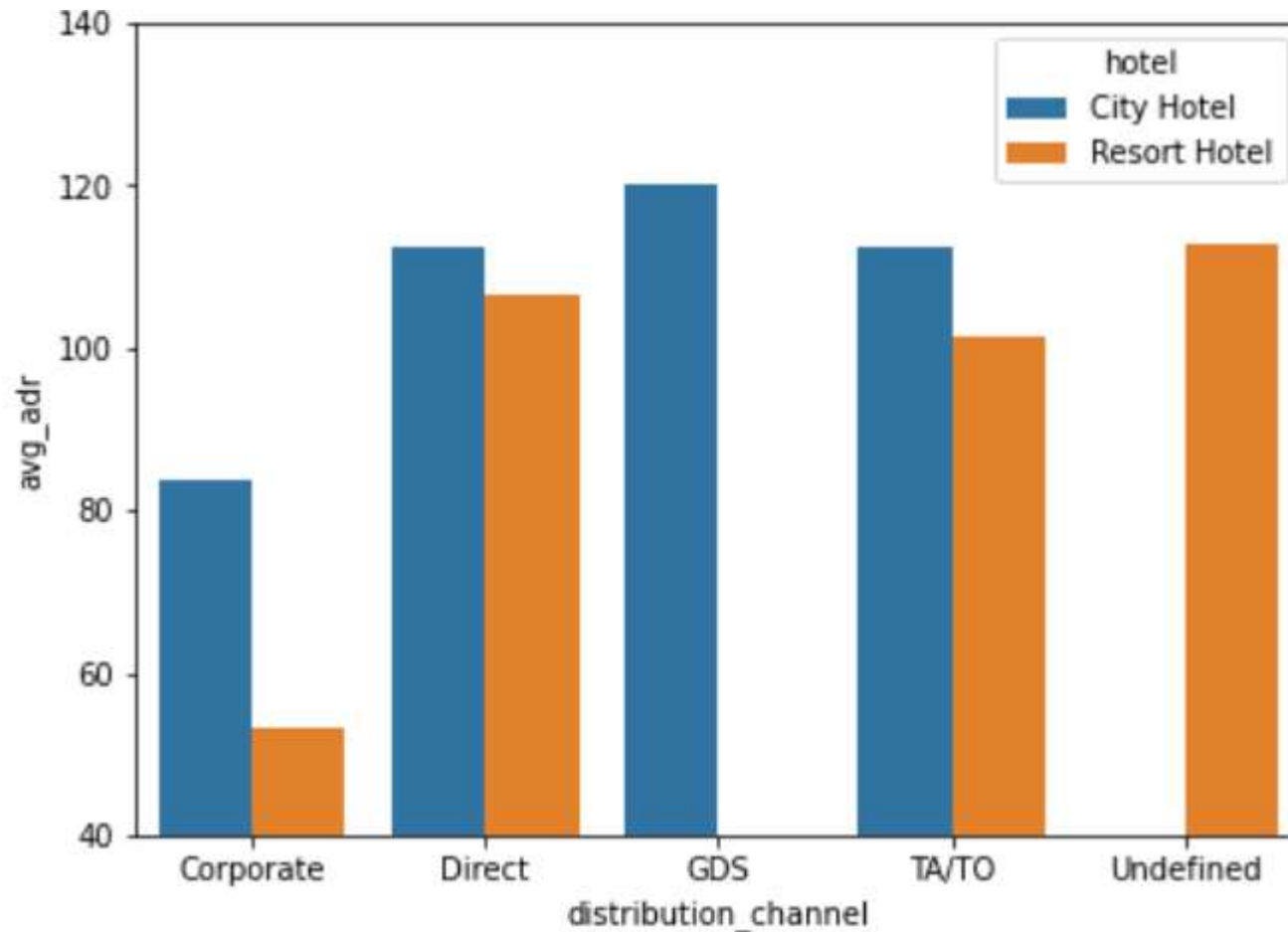
- Almost 30% of city hotel booking s and 25%of resort hotel bookings got cancelled.

# Distribution channel wise Anaiysis

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

1. Which distribution channel give better generating deal for hotel?

2. Which is the most common channel for booking hotels?

- 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 outreach on GDS channels to get more higher revenue generating deals.

- Resort hotel has more revnue generating deals by direct and TA/TO channel. Resort Hotel need to increase outreach on GDS channel to increase revenue

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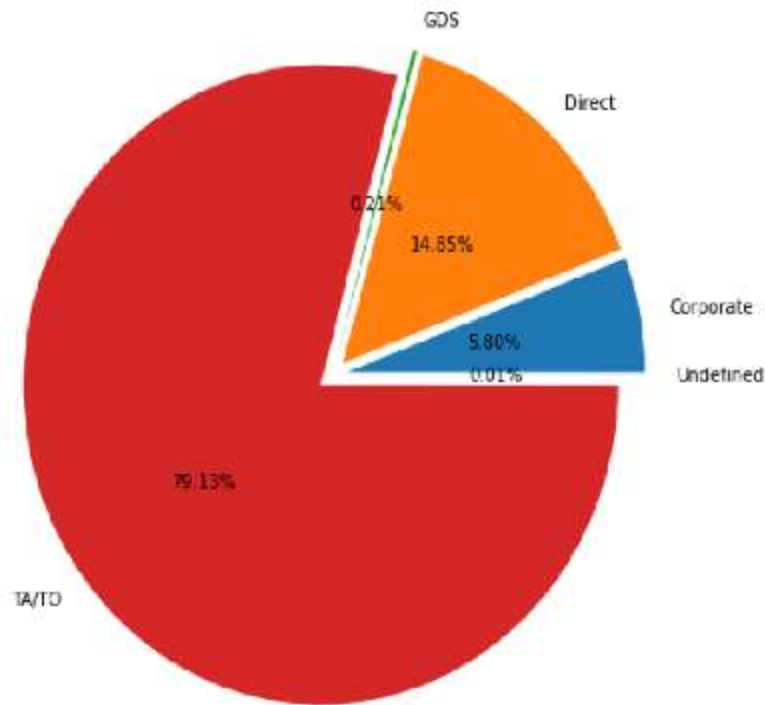
- Here we can see that the most of guest are making reservation through TA/TO channels which is travel agency and tour operator.
- Than the second most used channel is direct.
- Channel which is mostly used for early booking of hotels is also TA/TO.
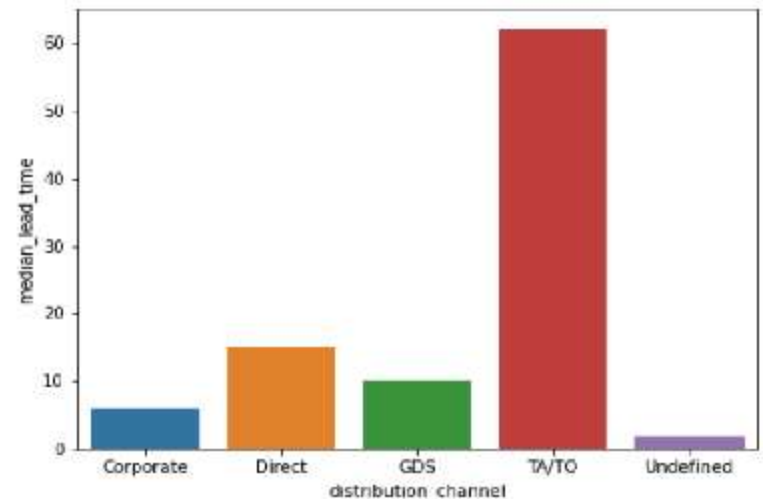
Booking cancellation Analysis

We analysis the following possible reactions for booking cancellations:

1. Which significant distribution channel has highest cancellation percentage?

2. Not getting same room as reserved.

3. Does not getting same room as reserved effects adr?
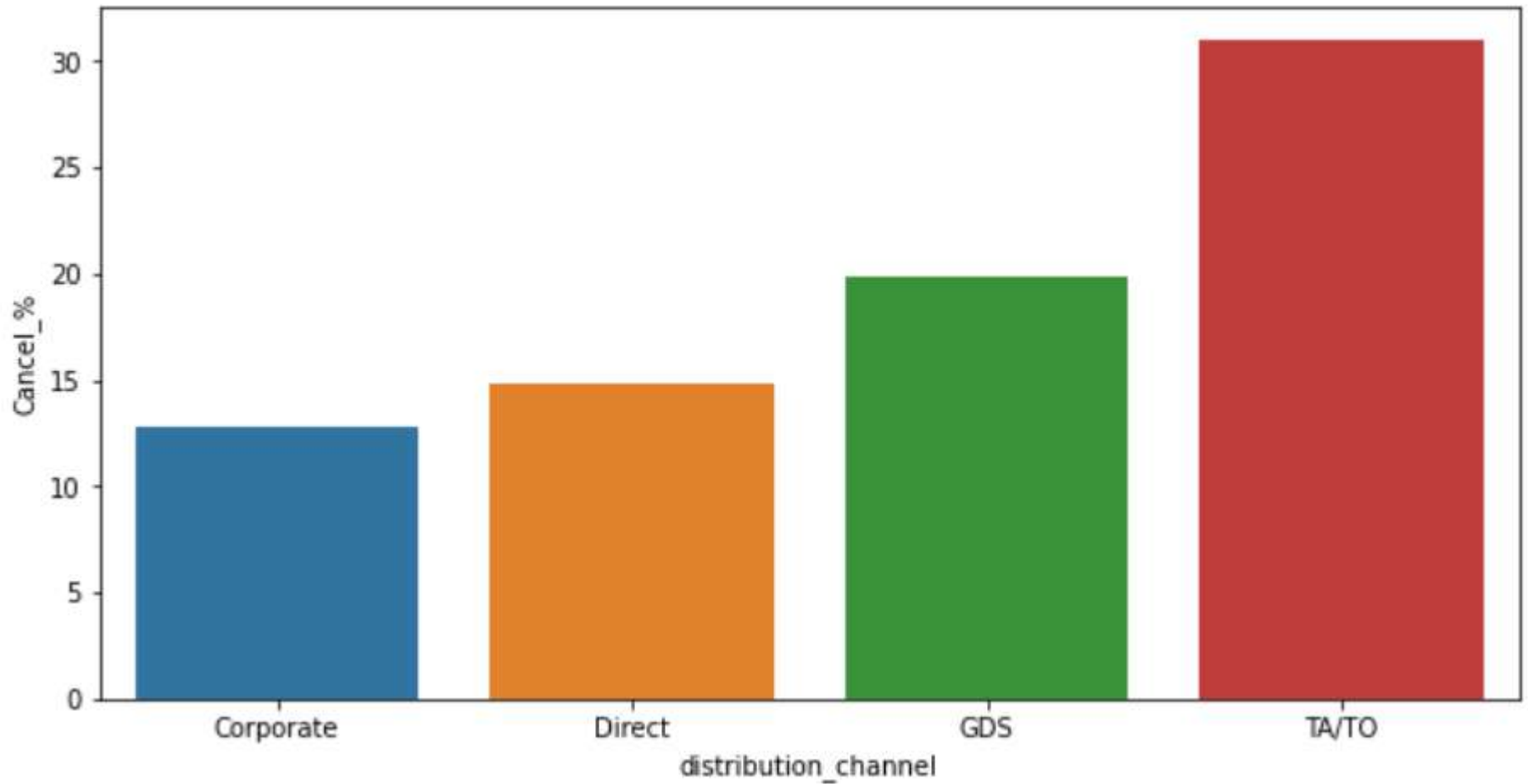
- TA/TO has highest booking cancellation %. Therefore, a booking via TA/TO is 30% likely to get cancelled.
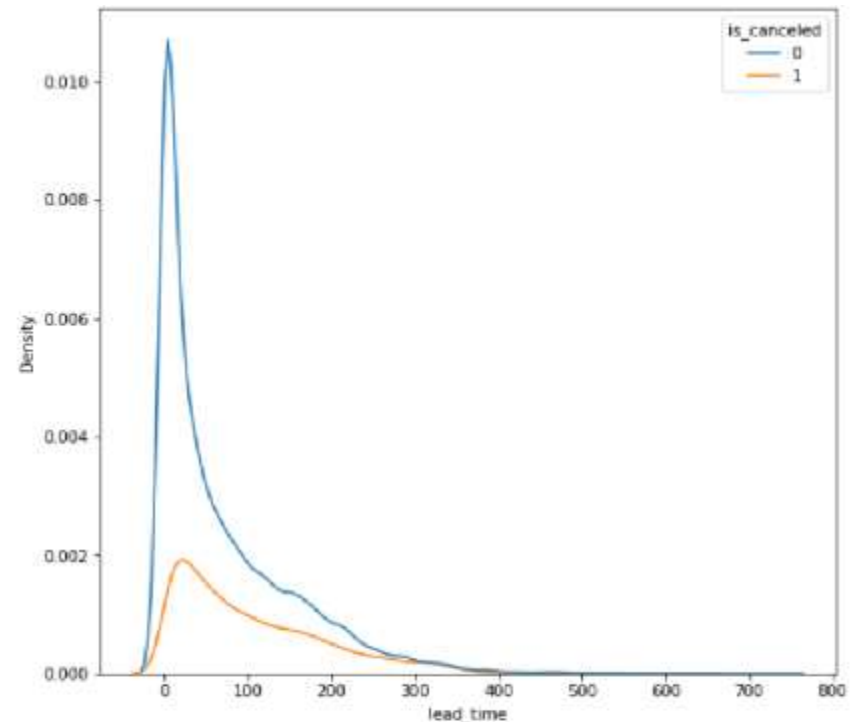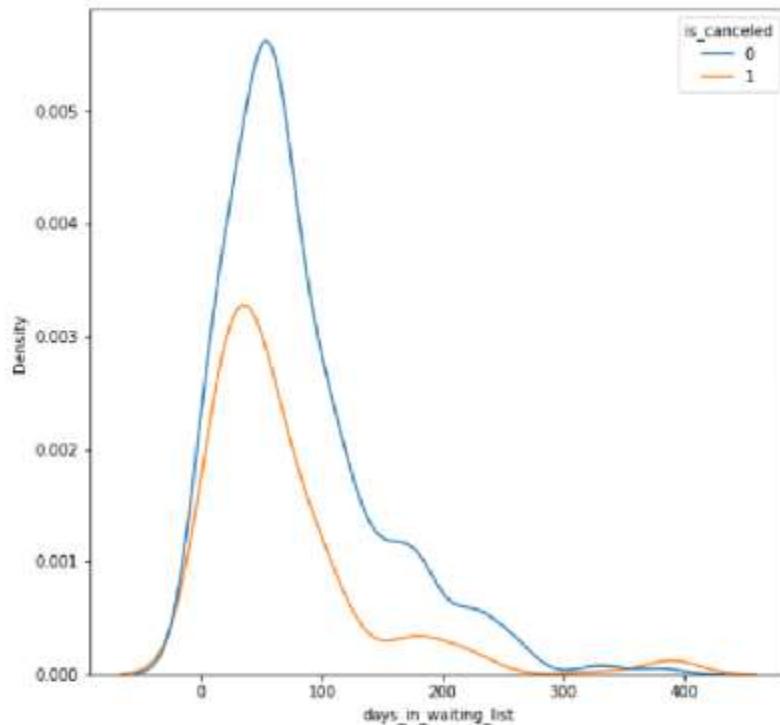
- Most of the bookings that are cancelled have waiting period of less 150 days but also most of bookings that are not cancelled also have waiting period of less than 150 days. Hence this shows that waiting period has no effect on cancellation of bookings.

- Also, lead time has no effect on cancellation of bookings, as both curves of cancellation and not cancelation are similar for lead time too.
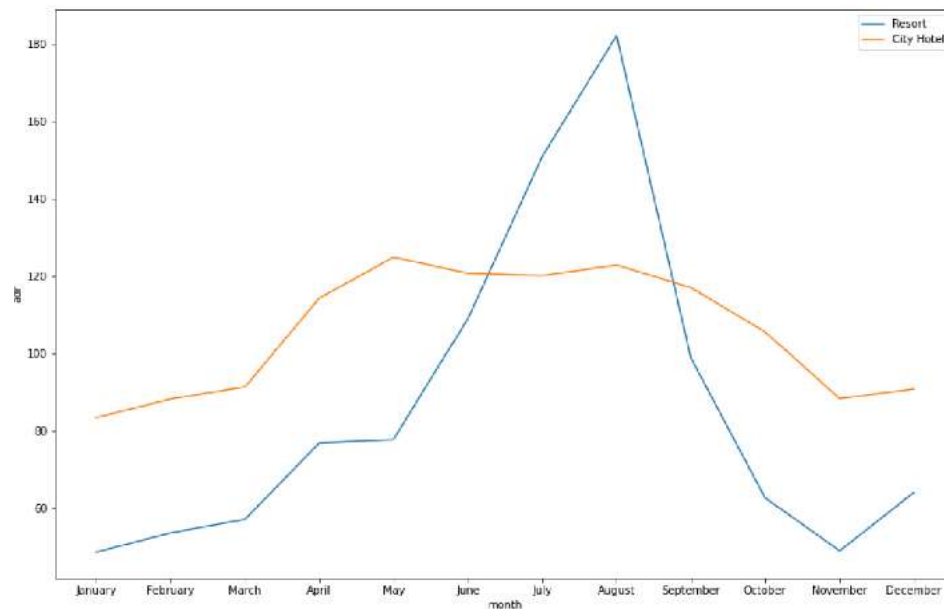
Time-wise Analysis

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

1.   What are the most busy months for hotels?

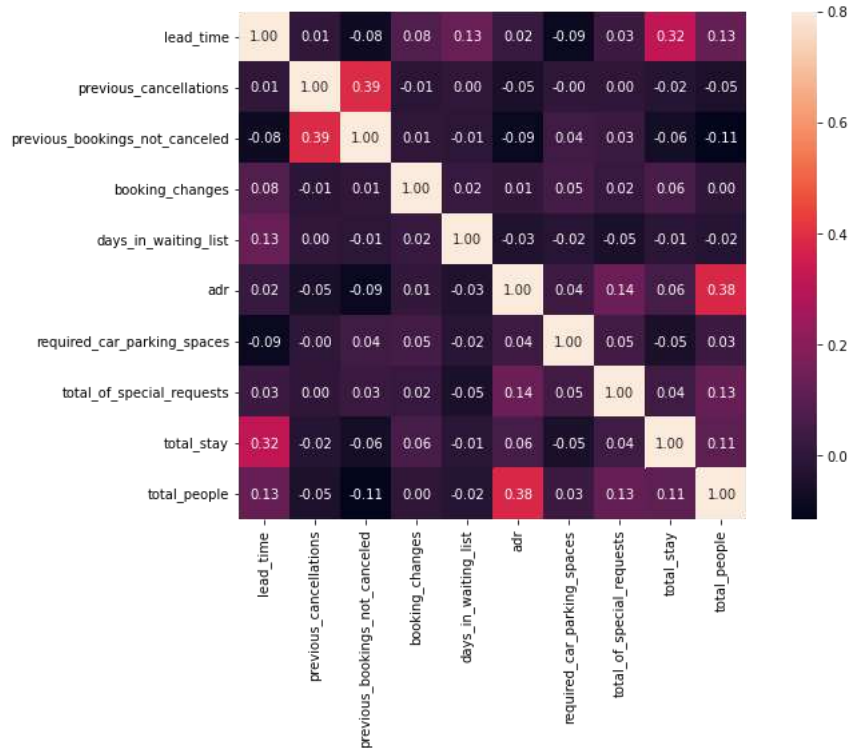2.   In which months hotels charges higher adr?
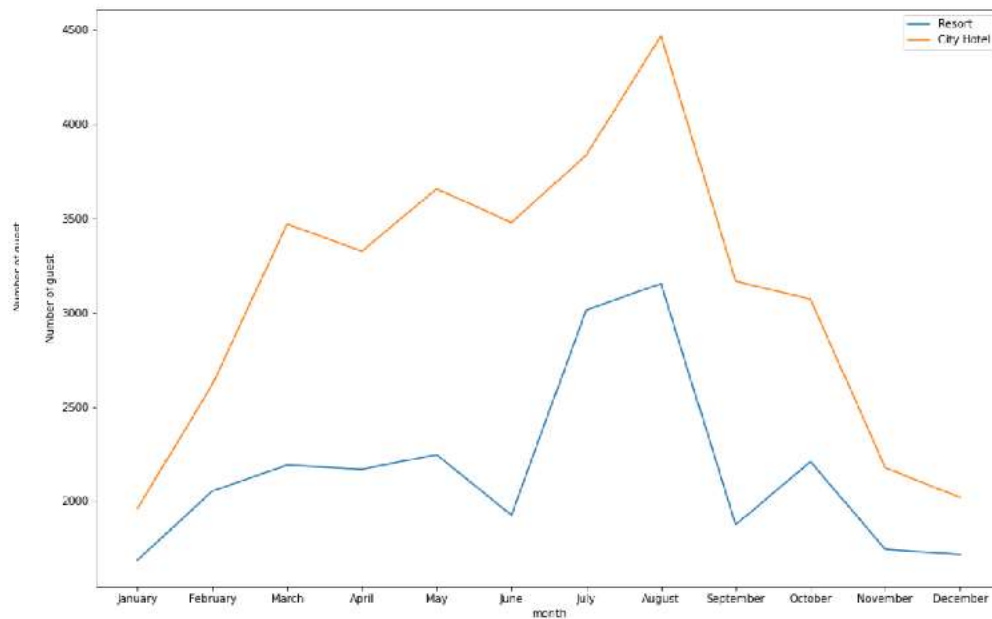
From the month of July to August the number of bookings increased and in August, City Hotel got most number of guests.

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.

Correlation heatmap

- total stay length and lead time have slightly correlation. this may means that for longer hotel stays people generally plan little before the actual arrival.

- Adr is slightly correlated with total_people means more revenue, therefore more adr.

Conclusion

- I conclude and observe thatin this hotel booking are as follows.
- Most of the market_segments customers are used online TA and offline TA/TO.
- In this booking 72% guests book the rooms type A but only 62% is actually assigned the same room.
- In this they are only 3% guests are repeated.
- In this booking 36% reservation are canceled by guests
- Only 88% guests have reserved and assigned same room.
- They have 82% distribution_channel in TA/TO.
- 37% booking has been canceled
- From reservation_status it is clear that though City Hotel is 64% but Resort Hotel has more number of guests and City Hotel has more number of Cancellation as compare to Resort Hotel

Conclusion(contd…)

- From market_segment and distribution_channel we can observe that City Hotel is more into 'Online TA' and Resort Hotel is more into 'Direct' segment.
- From meal we can observe that City Hotel provide 'BB' and Resort Hotel provide 'HB'.
- City Hotel is more non-refund as compare to Resort Hotel.
- Most of the guests are from PRT and they happen to choose Resort Hotel more than City Hotel
- It is observed arrival_date_week_number and arrival_date_year are 54% negative correlated.
- Company and agent are positive correlated by 35%.
- adr and children are positive correlated by 32%
- Guest using 'Complementary' segment are the one who is repeated followed by 'Corporate'.
- Guest with babies and children likely to make special request as comapre to guest without babies or children.
- 'Undefined' and 'Group' are the top segemnt where guest usually cancel the booking foolowed by 'Online TA'.
-)Colusion have been made that customer type 'Group' has more number of guests i.e 89% and 'Transient' has more number of canceled guests 39.5%.
- We have concluded that having different room then desired room is not the major cause of cancelation.

Thank you