Capstone Project
Hotel Booking Analysis

Rohit Pal

Points to Discuss:

- Agenda

- Data summary

- Univariate analysis

- Hotel wise analysis

- Booking cancellation analysis

- Some important questions

- Correlation heatmap

- Conclusion

Agenda

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

We'll be doing analysis of given data set in following ways :

- Univariate analysis

- Hotel wise analysis

- Booking cancellation analysis

- Timewise analysis

Data Summary

Given data set has different columns. 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(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.)

AI

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?

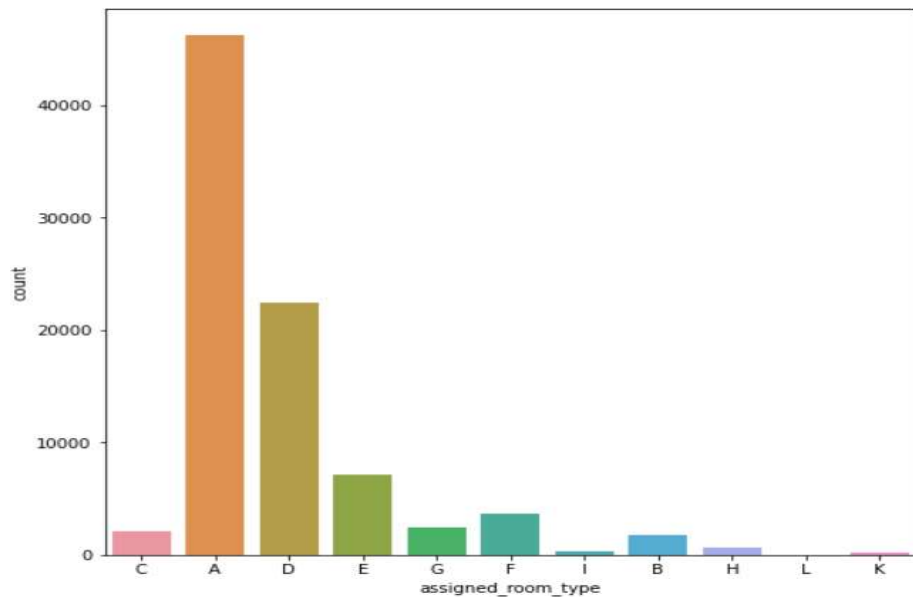(4) What is the most preferred meal by customers?

Correlation Heatmap
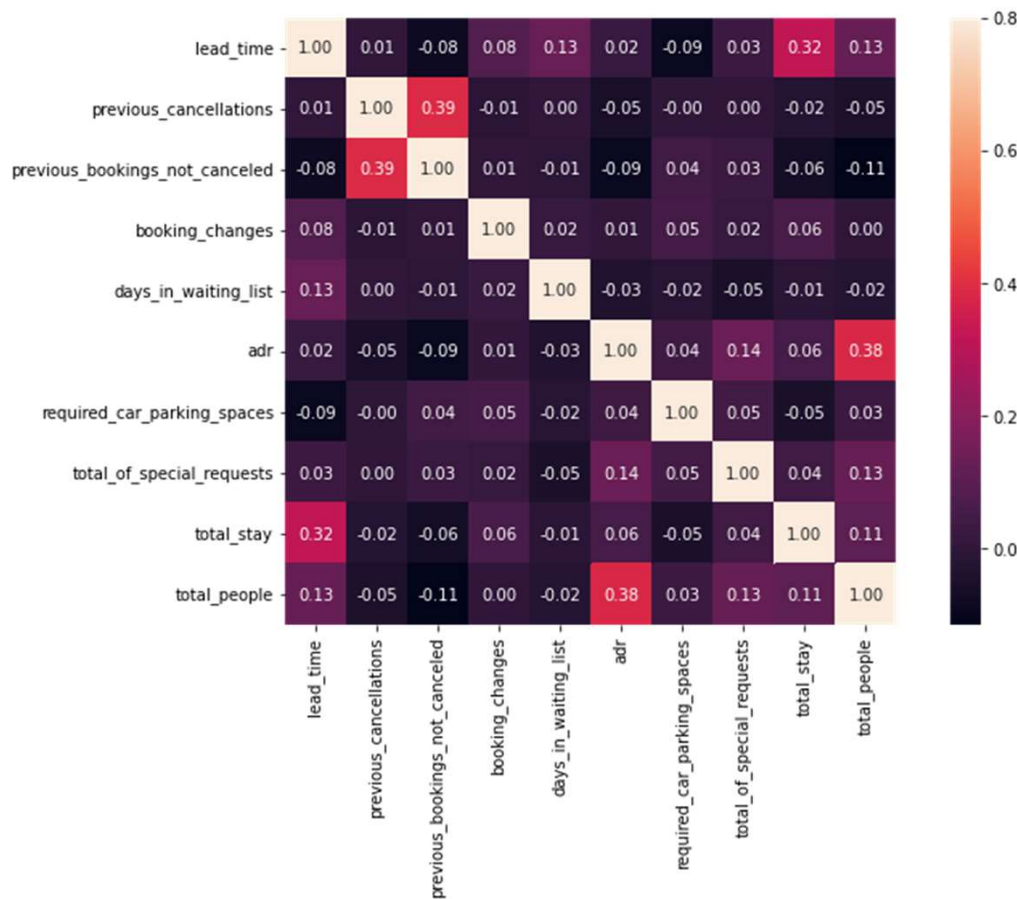
- Total stay length and lead time are slightly correlated. This may means that for longer hotel stays, people generally plan little before the actual arrival.

- adr is slightly correlated with total_people, which makes sense as more no. of people means more service to deliver, therefore more adr.
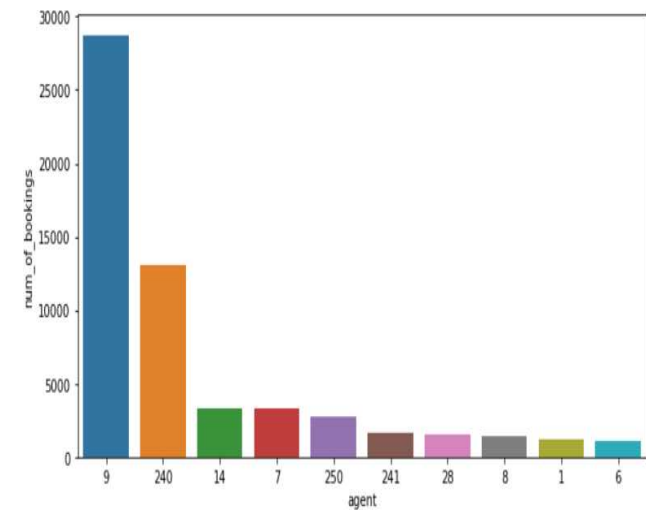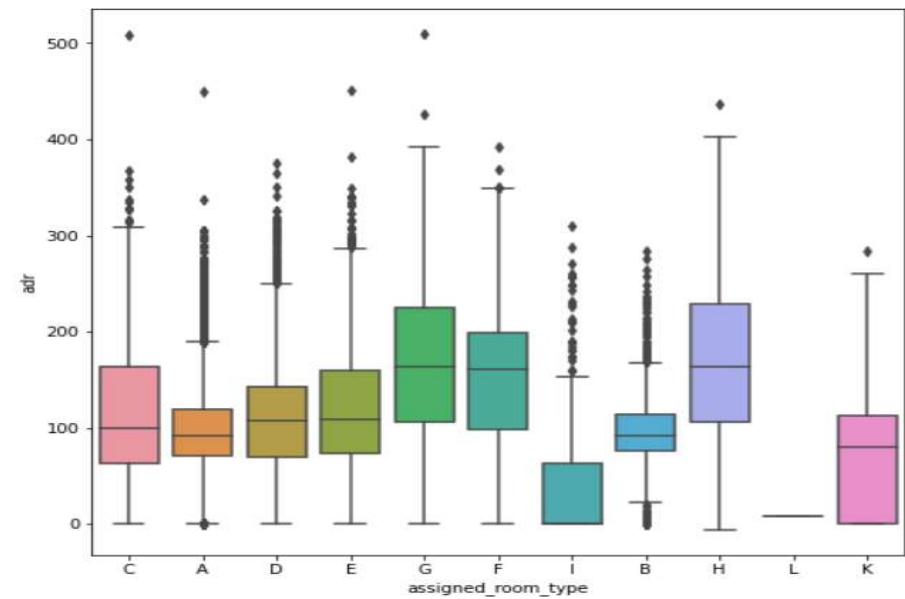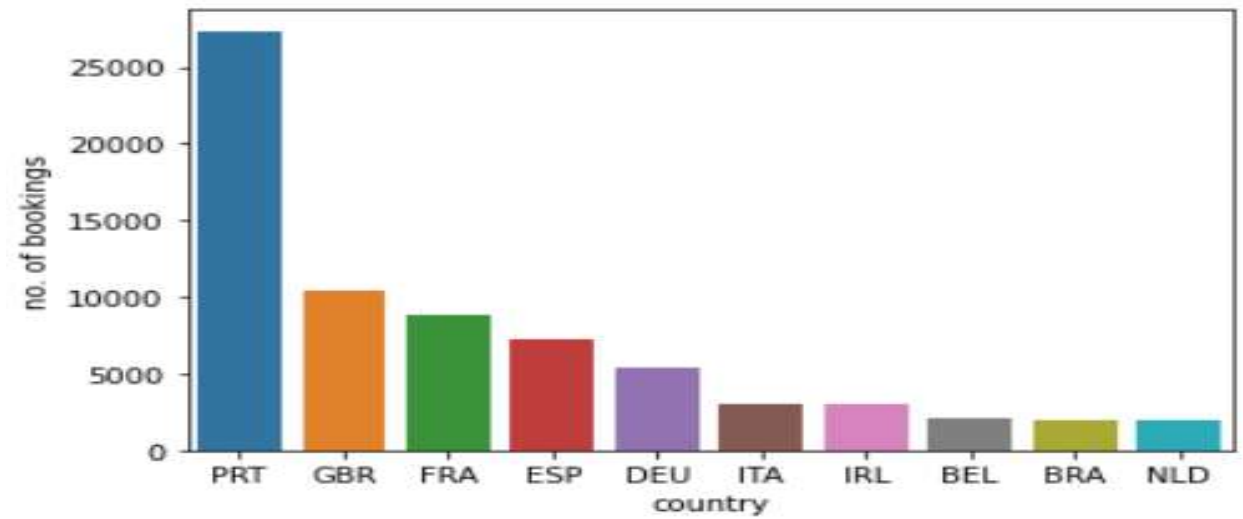
Time wise analysis
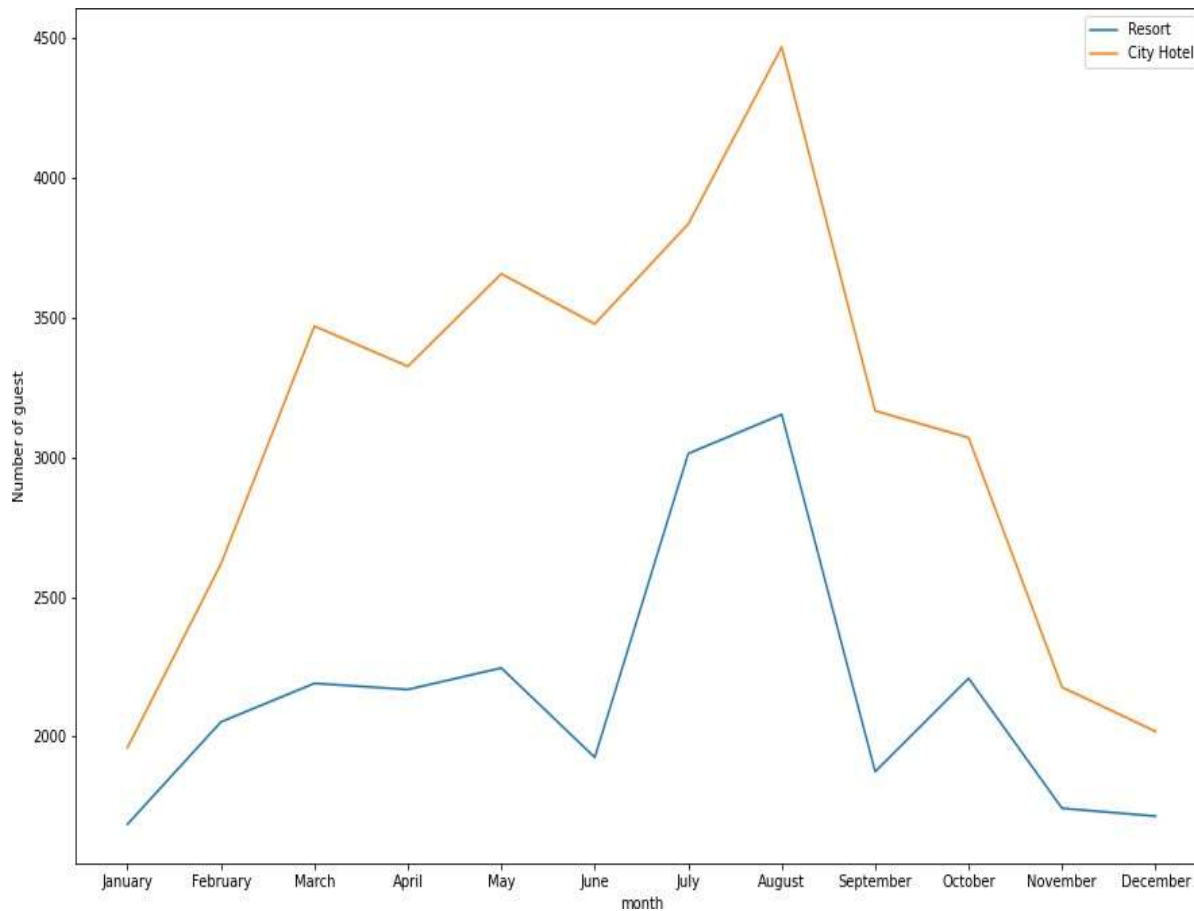
From the month of July to August the number of bookings increased and in August, City Hotel got most number of guests.
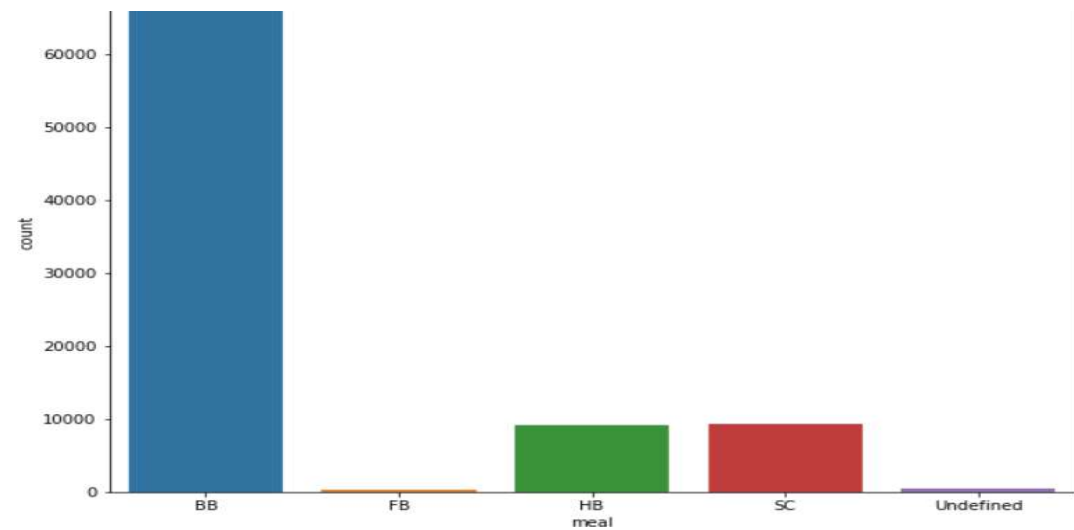
- Type A room is most demanded by customers.

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

- Agent with id no. 9 made most of the bookings.

- 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) Which hotel has higher lead time?

(4) What is most preferred stay length in each hotel.

(5) Which hotel has higher booking cancellations rate?
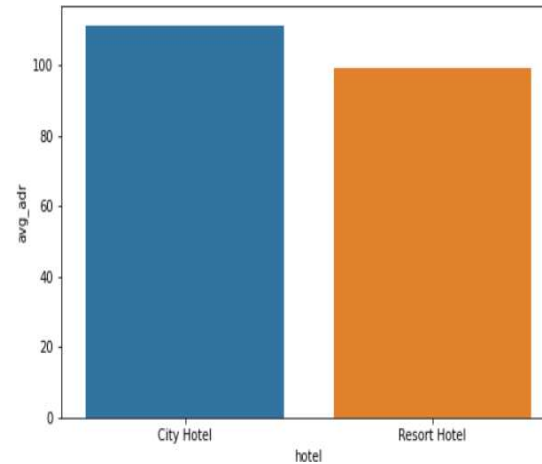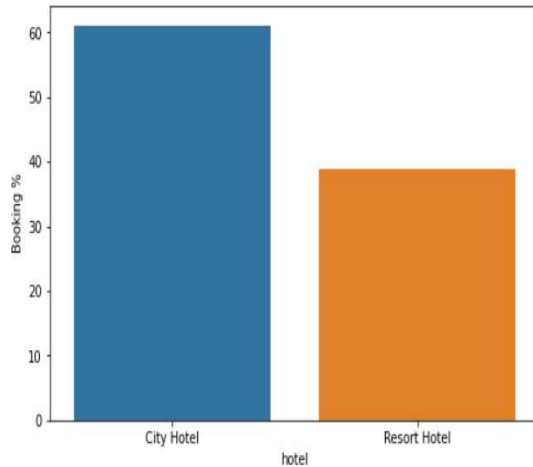
- 60% bookings are for City hotel and 40% bookings are for Resort hotel.
- Avg adr of Resort hotel is slightly lower than that of City hotel. Hence, City hotel seems to be making slightly more revenue.
- City hotel has slightly higher median lead time. Also median lead time is significantly higher in each case, this 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.
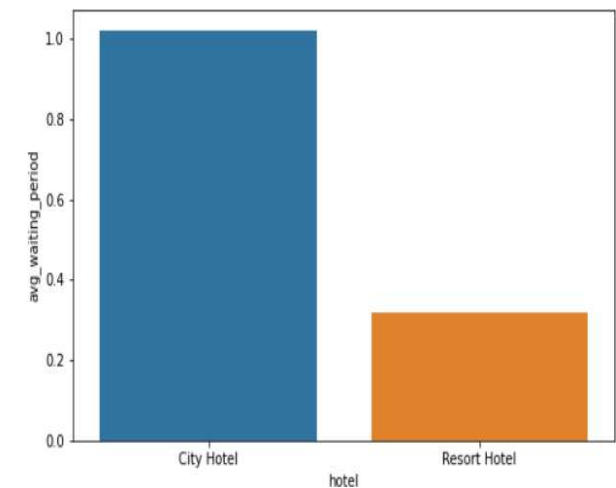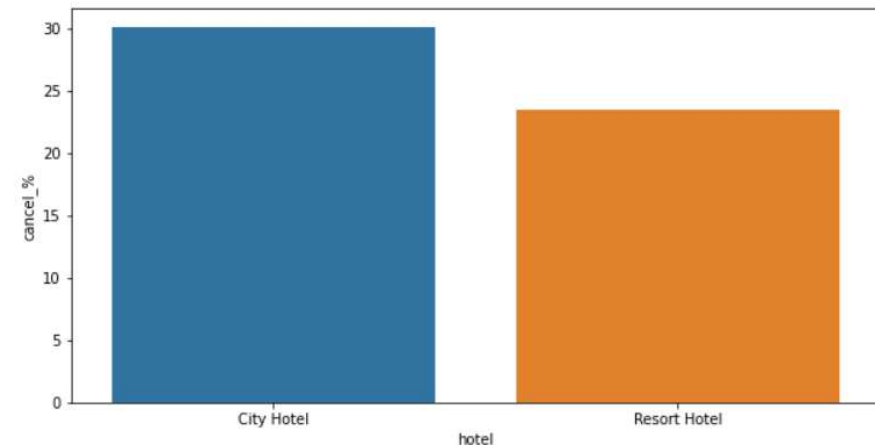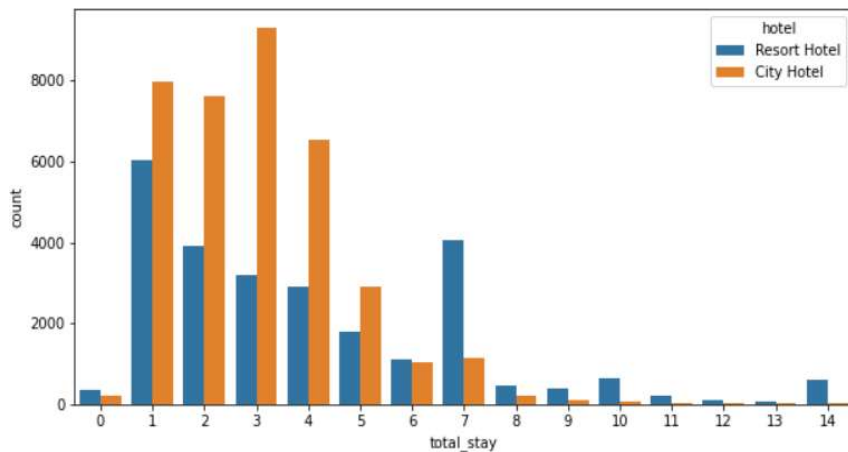
- Most of stays are less than 5 days. There are very few long stays at hotels but Resort Hotel is preferred for long stays.

- Almost 30 % of City Hotel bookings and 25 % of Resort hotel bookings got canceled.

- Both hotels have very small percentage that customer will repeat, but Resort hotel has slightly higher repeat % than City Hotel.
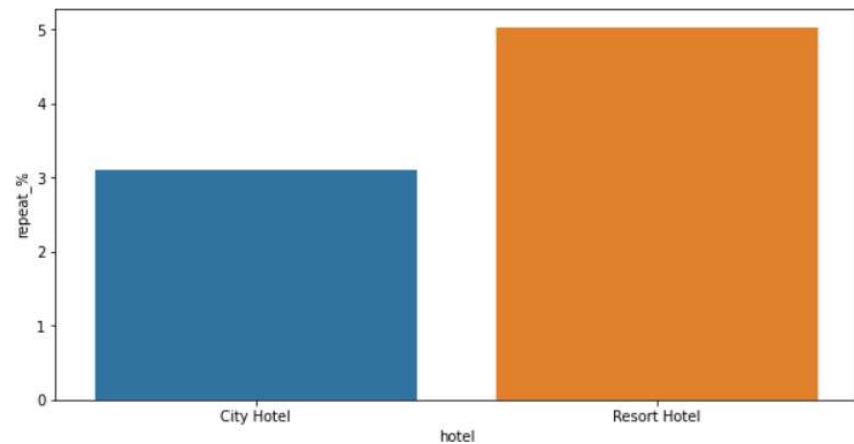
Distribution channel wise Analysis

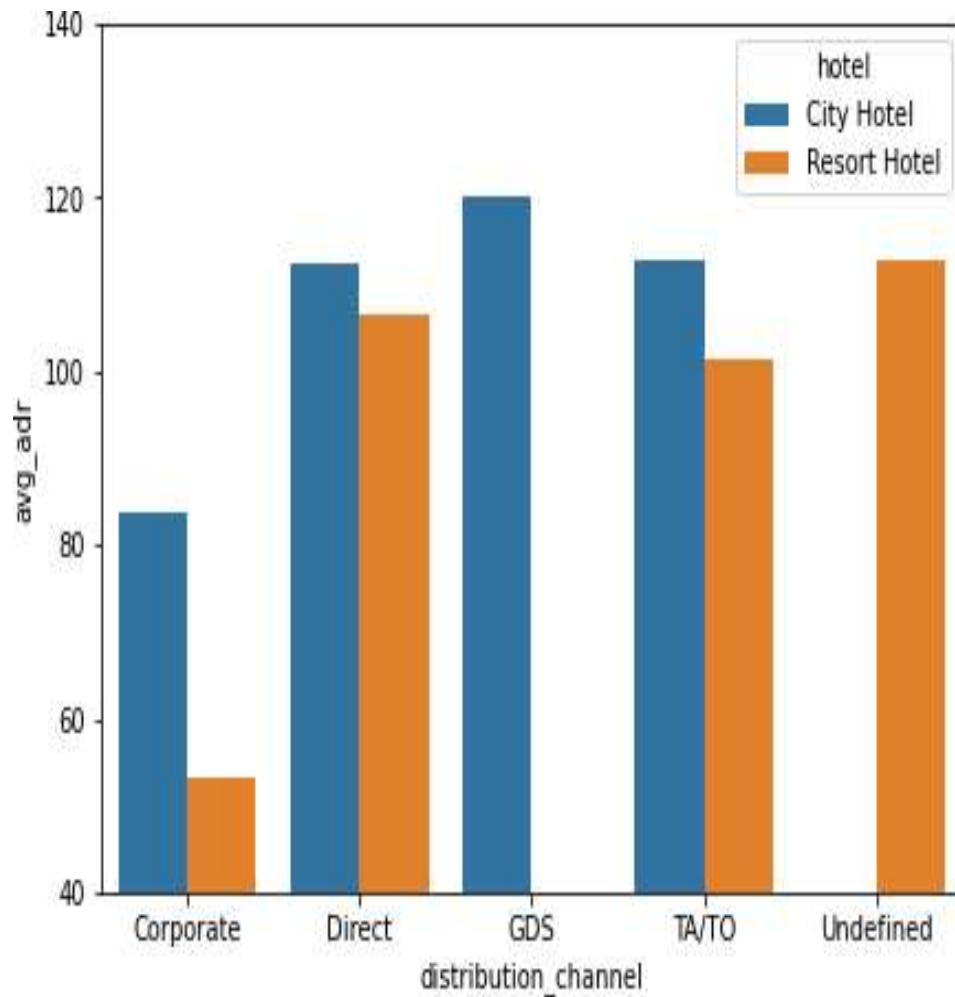Booking % by distribution channels
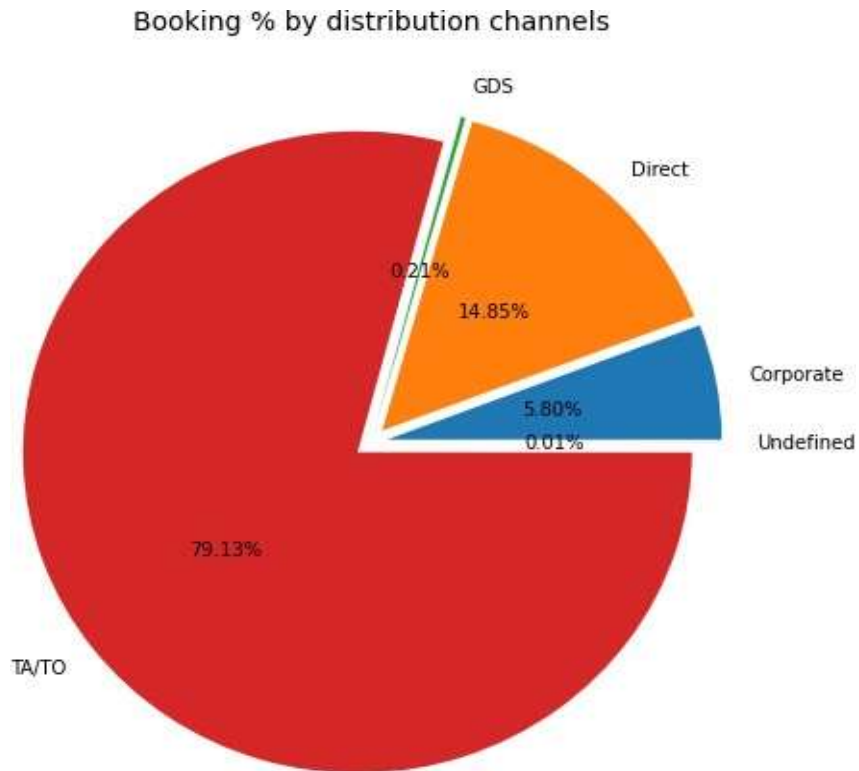
- Here we can see that the most of guest are making Than the reservation through TA/TO channels which is travel agency and tour operator.
- second most used channel is direct.
- Channel which is mostly used for early booking of hotels is also TA/TO.
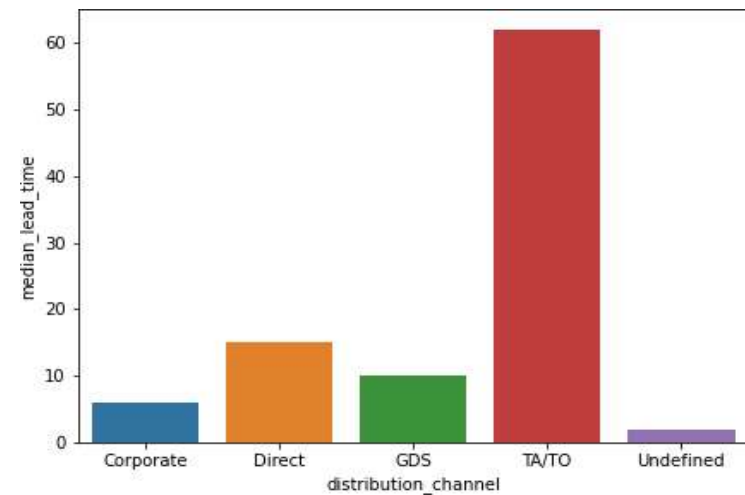
- GDS channel brings higher revenue generating deals for City hotel, in contrast to that most bookings come via TA/TO. City Hotel can work to increase outreach on GDS channels to get more higher revenue generating deals.

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

Booking cancellation Analysis

We analyze the following possible reasons for booking cancellations:

(1) Which significant distribution channel has highest cancellation percentage?

(2) What causes the cancelation of bookings of rooms by customers.?

(3) Longer waiting period longer lead time causes the cancellation of bookings?
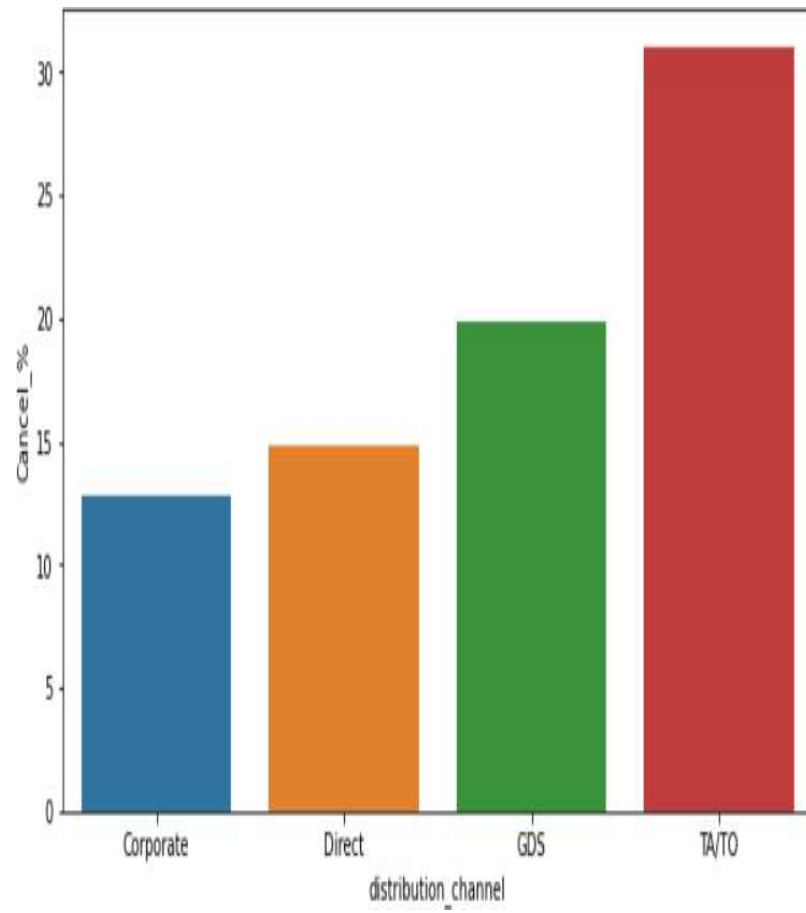
- TA/TO has highest booking cancellation %. Therefore, a booking via TA/TO is 30% likely to get cancelled.
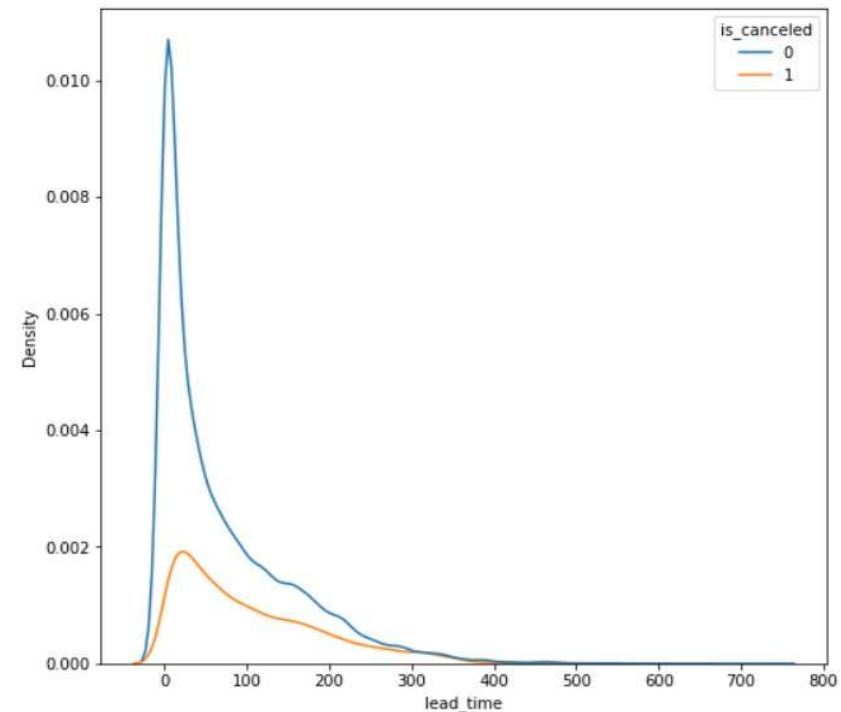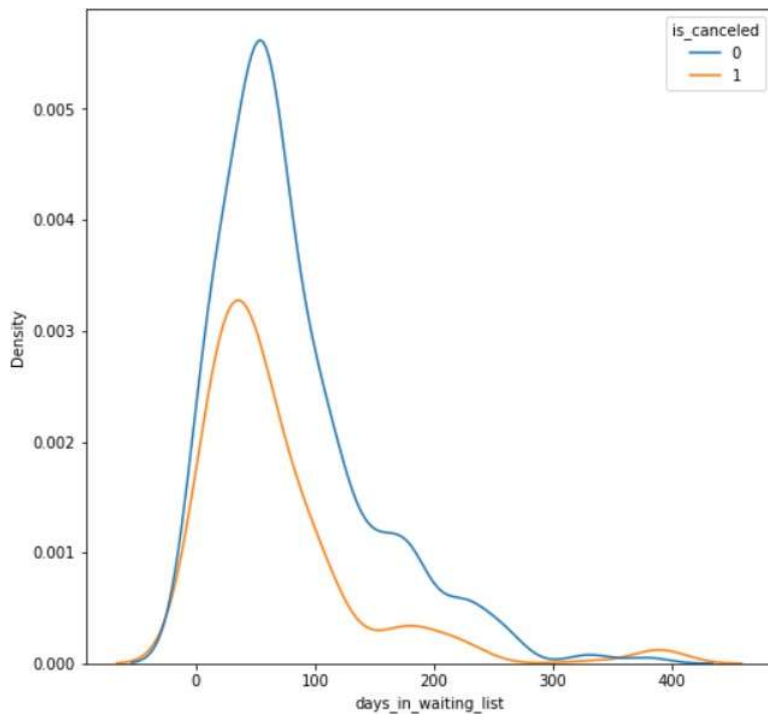
- Most of the bookings that are cancelled have waiting period of less 150 days but also most of bookings that are not cancelled also have waiting period of less than 150 days. Hence this shows that waiting period has no effect on cancellation of bookings.

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

Optimal stay length for better deals in adr
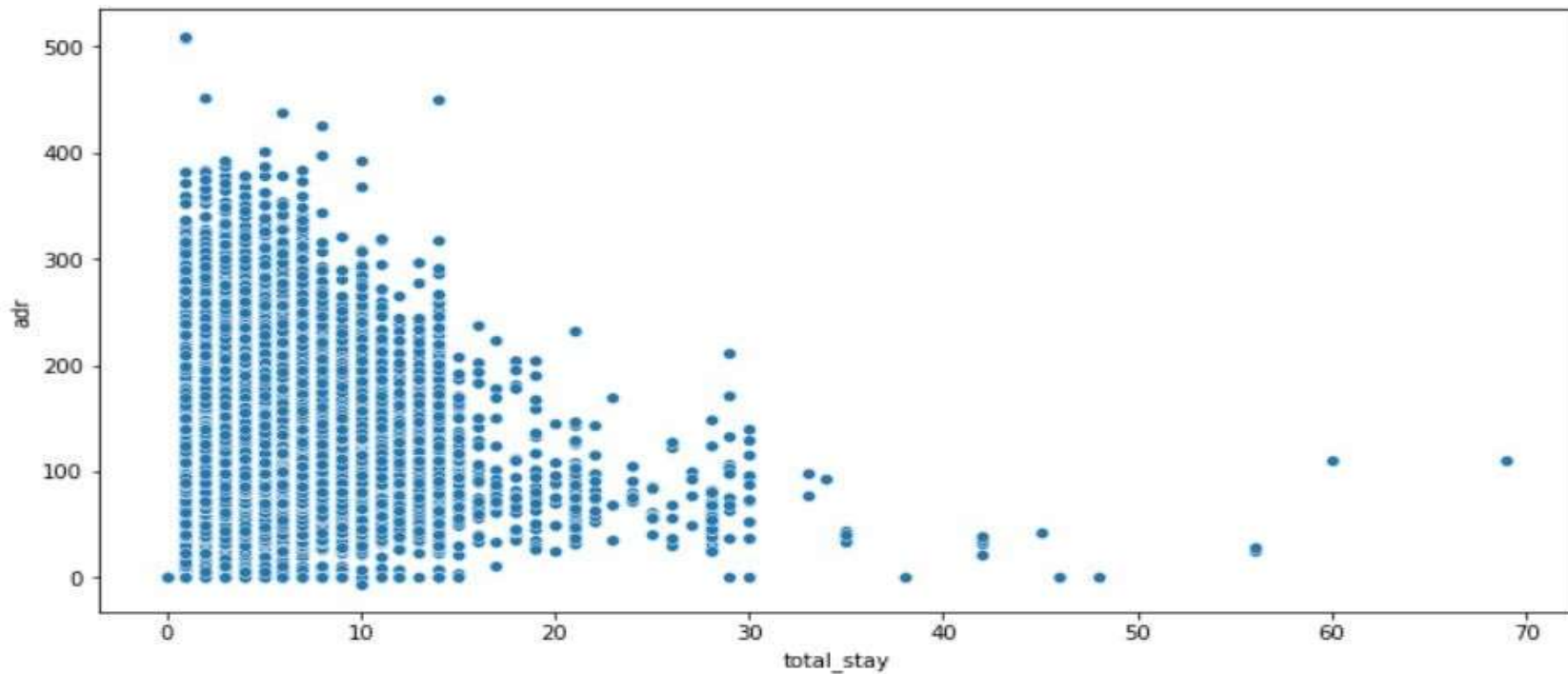
For shorter stays the adr(average daily rate varies greatly) but for longer stays (> 15 days) adr is comparatively very less. Therefore, customers can get better deal for longer stays more than 15 days.

Conclusion

- 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 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 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.

AI

Thank You