Capstone Project

Hotel Booking Analysis

Submitted by

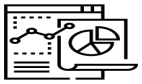
Abhishek Tiwari **Shubhankar Rathod** **Rashi Jaisinghani**

Sumedha Shejul **Dibakar Biswakarma**

Points to Discuss :


Agenda


Data Summery


Univariate analysis

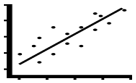
Hotelwise Analysis


Distribution Channel wise Analysis


Booking Cancelation Analysis


Time wise Analysis


Some Important Questions


Correlation heat Map


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

By doing this we'll try to find out key factors driving the hotel bookings trends.

Data Summery

Given data set has different columns of variables crucial 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_weekdays_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 Summery (cont…)

AI

# **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) What is the most preferred meal by customers?

Room type : most demanded & with highest adr

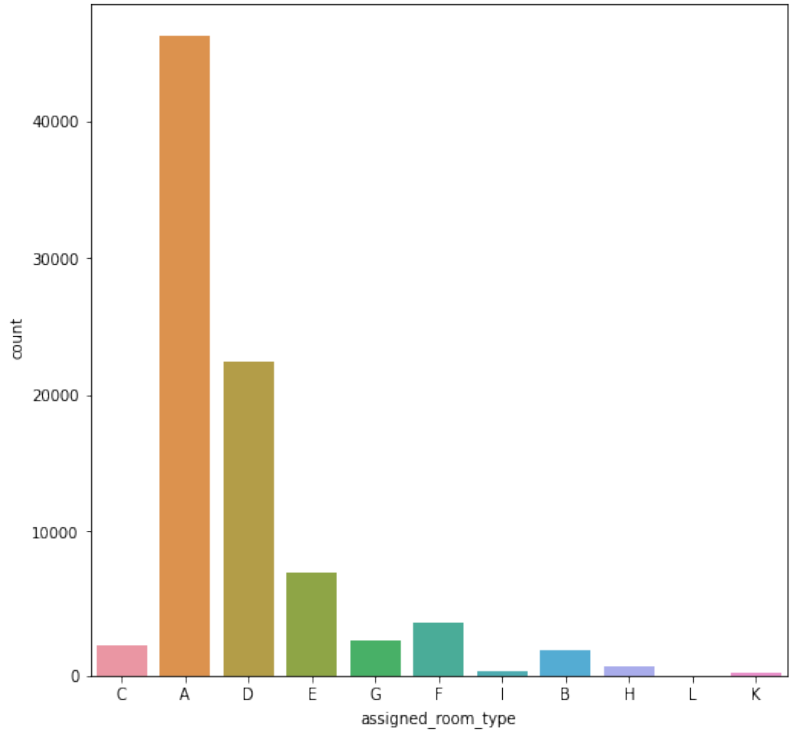


Fig.1

Type A room is most demanded by customers. (Fig.1)

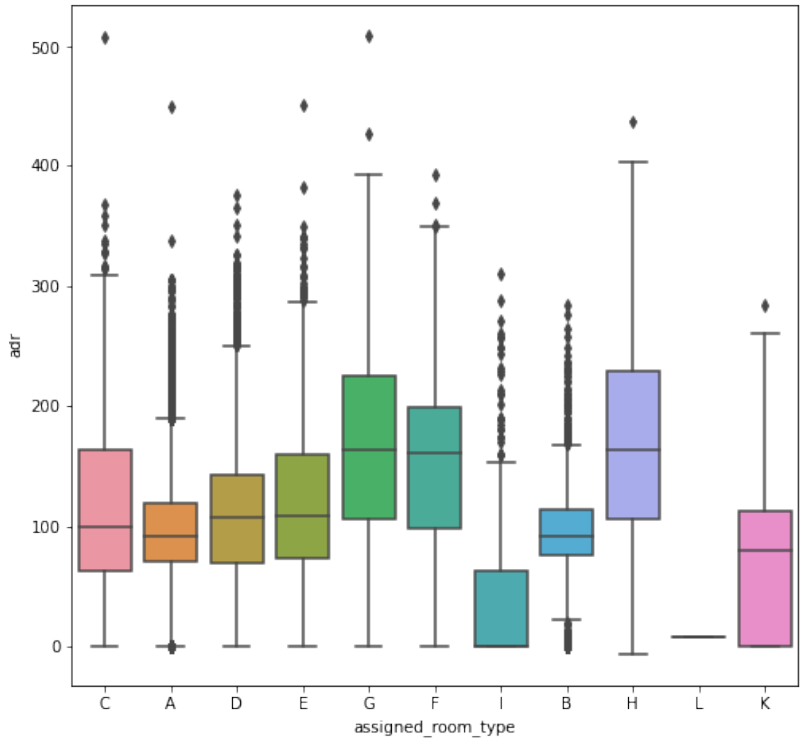


Fig.2

Room types C, G and H are some of the highest adr(average daily rate) generating rooms. (Fig.2)

Agent Bookings & Preferred Meal

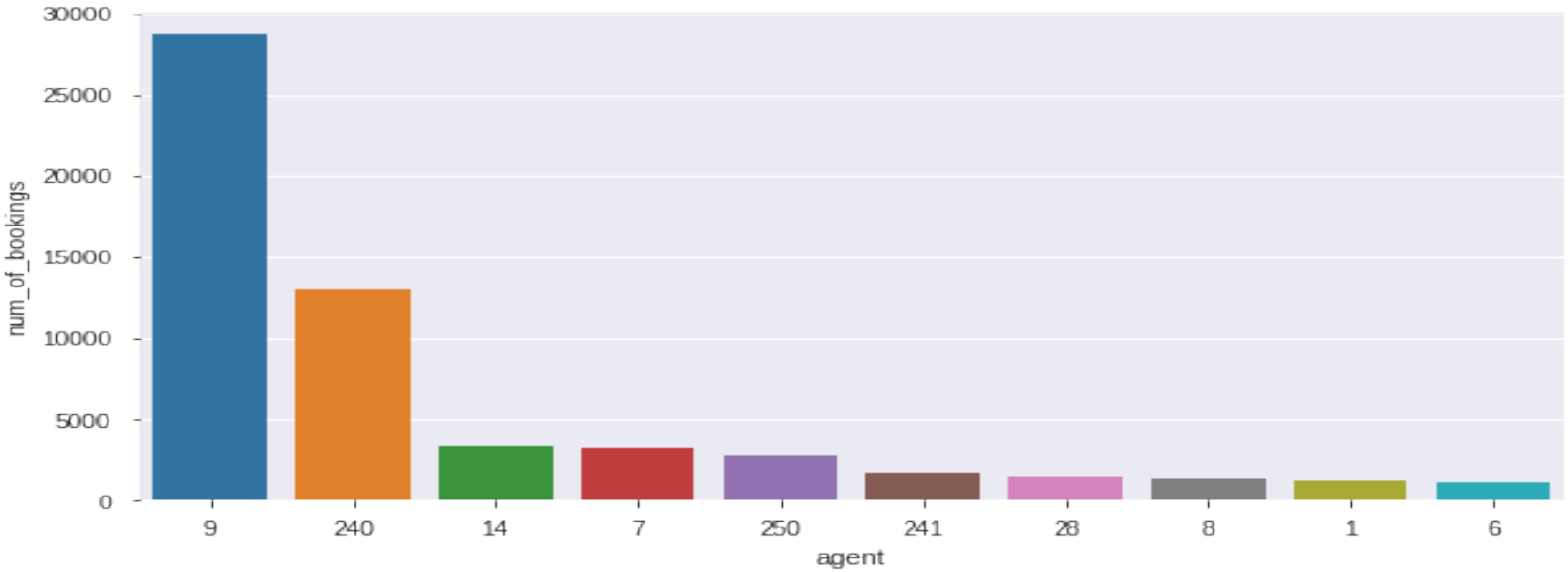


Fig. 3

Agent with Id no. 9 made most of the bookings. (Fig 3)

Most preferred meal type is BB (Bed and breakfast). (Fig 4)

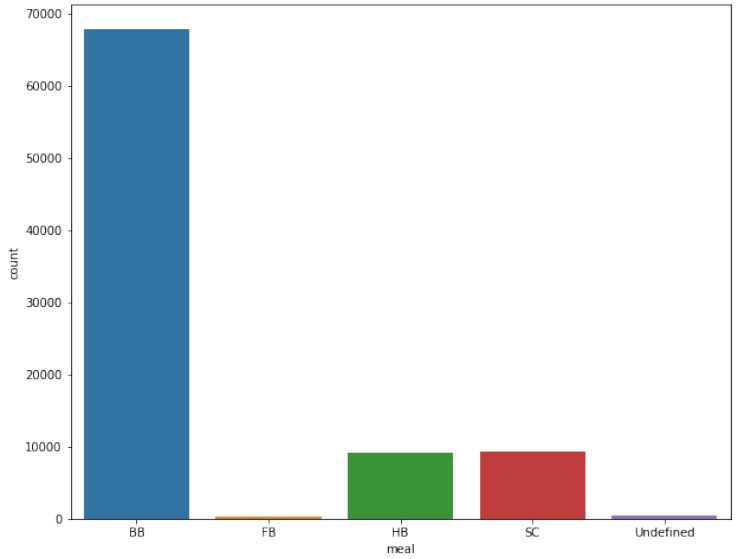


Fig. 4

Hotelwise 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) For which hotel, does people have to wait longer to get a booking confirmed?

(6) Which hotel has higher booking cancellations rate?

(7) Which hotel have higher and how much customer returning rate?

Booking %, More Revenue, Cancelation & Waiting

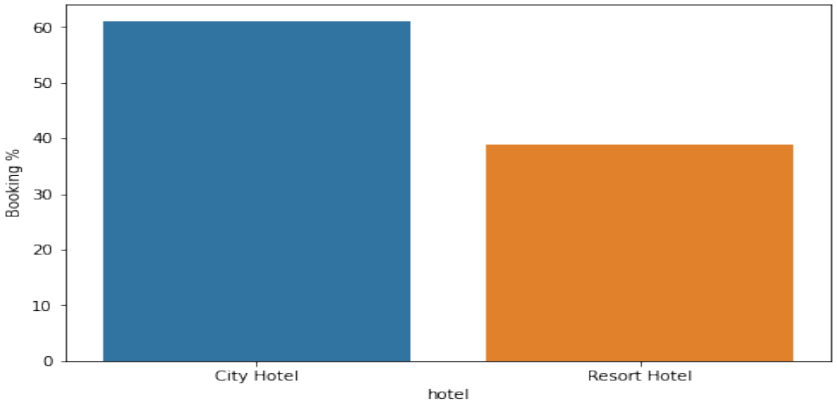


Fig. 5

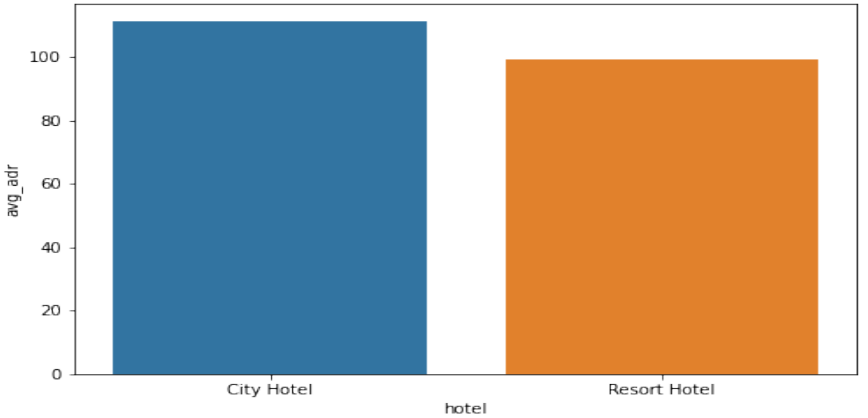


Fig. 6

Around 60% bookings are for City hotel and 40% bookings are for Resort hotel. (Fig. 5)

City hotel seems to be making slightly more revenue. (Fig. 6)

Almost 30 % of City Hotel bookings and 25 % of Resort hotel bookings got canceled. (Fig. 7)

City hotel has significantly longer waiting time, hence City Hotel is much busier than Resort Hotel. (Fig. 8)

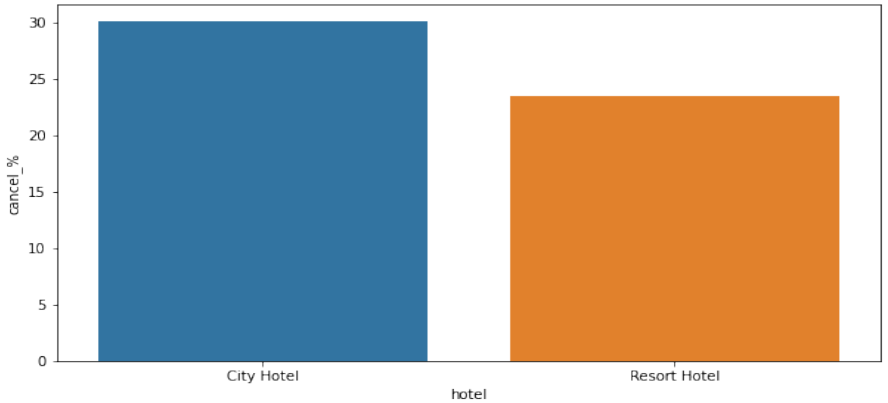


Fig.7

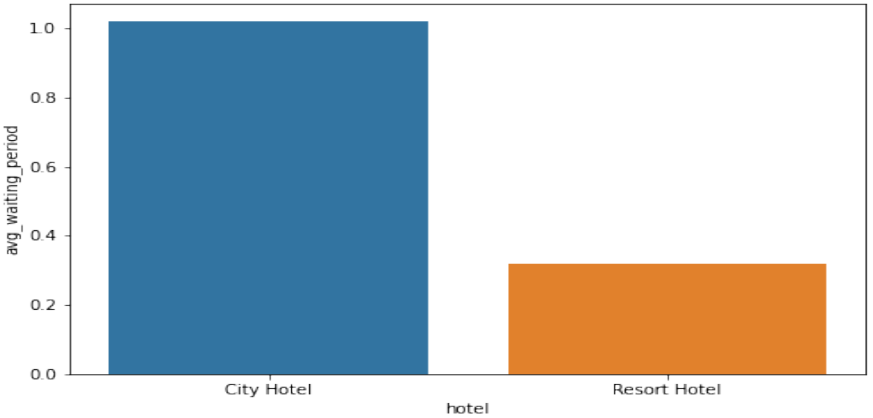


Fig.8

Higher lead time, Hotel type to stay, Return rate

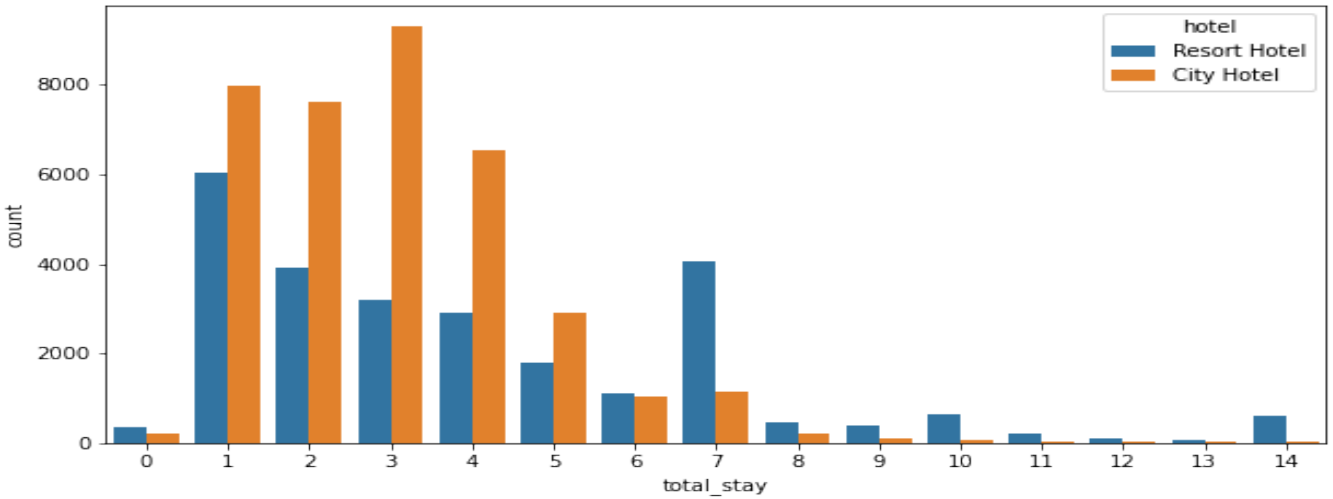


Fig. 9

Most of stays are less than 5 days. There are very few long stays at hotels but Resort Hotel is preferred for long stays. (Fig.9)

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. (Fig.10)

Both hotels have very small percentage that customer will repeat, but Resort hotel has slightly higher repeat % than City Hotel. (Fig.11)

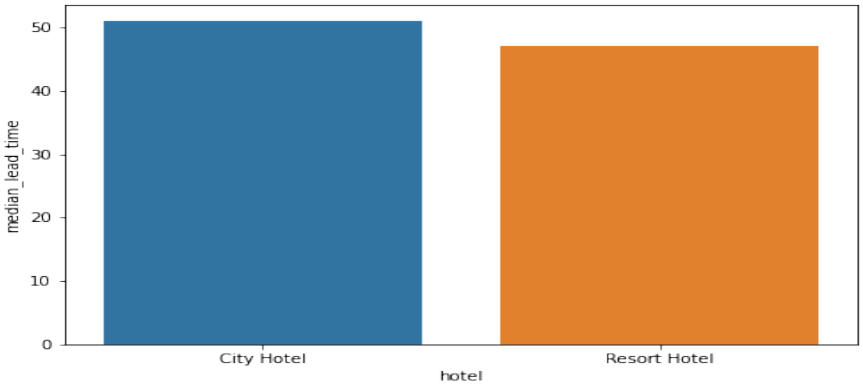


Fig. 10

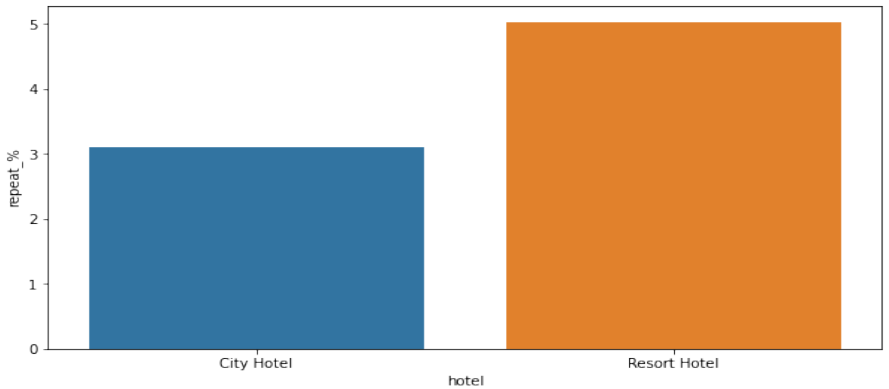


Fig. 11

Distribution Channel wise Analysis

While doing Distribution channel wise analysis of given hotel booking dataset,

we answered following questions:

(1) Which is the most common channel for booking hotels?

(2) Which channel is mostly used for early booking of hotels?

(3) Which channel has longer average waiting time?

(4) Which distribution channel brings better revenue generating deals for hotels?

Booking Hotels: common channel & used for early booking

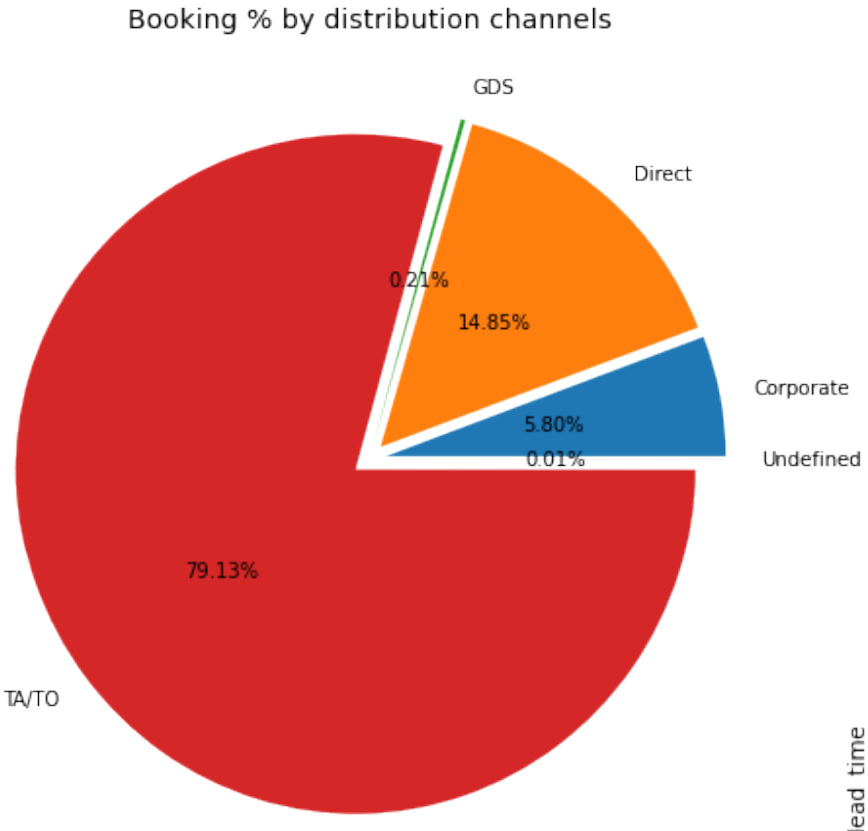


Fig. 12

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. (Fig.12)

Channel which is mostly used for early booking of hotels is also TA/TO. (Fig.13)

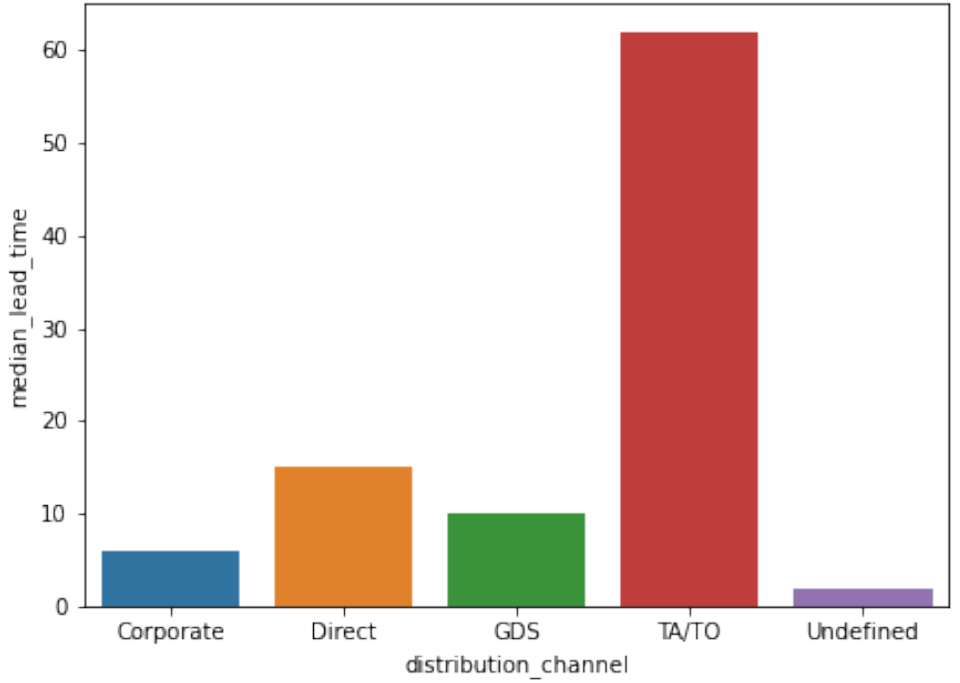


Fig. 13

Time to confirm bookings & Revenue deals

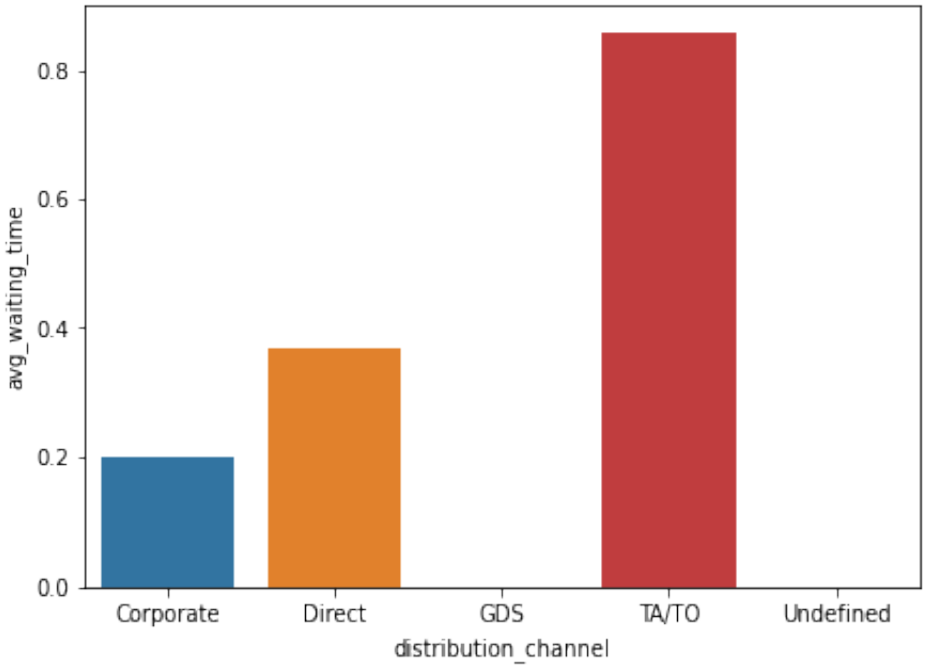


Fig. 14

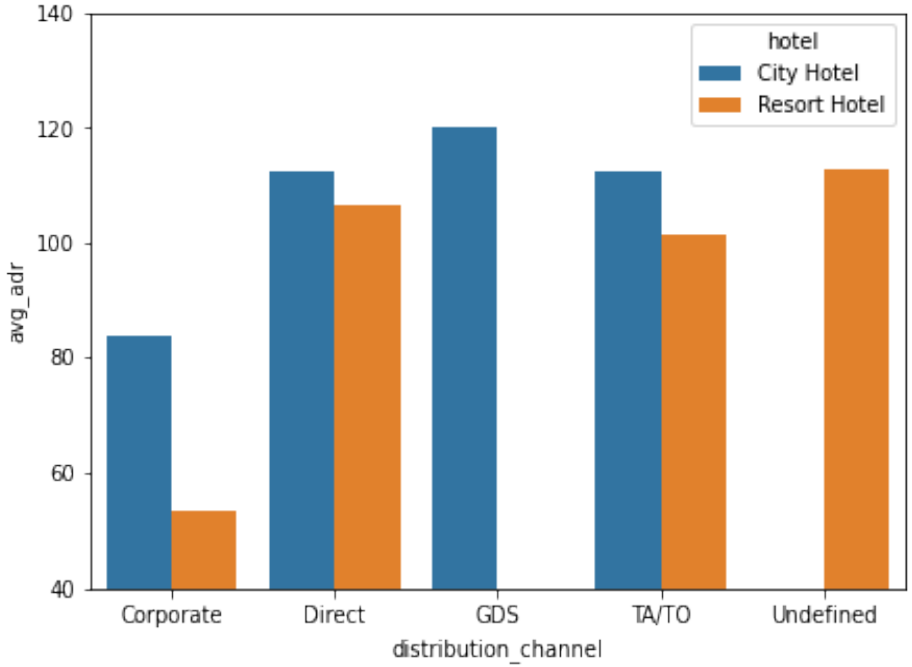


Fig. 15

Hotels booking via TA/TO one may have to wait a little longer to confirm booking of rooms. (Fig.14)

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. (Fig. 15)

Booking cancellation Analysis

We analyze the following possible reasons for booking cancellations:

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

(2) Longer lead time.

(3) Longer time (in days) in waiting list.

(4) Not getting same room as reserved.

(5) Does not getting same room as reserved effects adr?

Cancelation %, Same room allocation

TA/TO has highest booking cancellation %. Therefore, a booking via TA/TO is 30% likely to get cancelled. (Fig.16)

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. (Fig.17)

But, customers who didn't got same room have paid a little lower adr, except for few exceptions. (Fig.18)

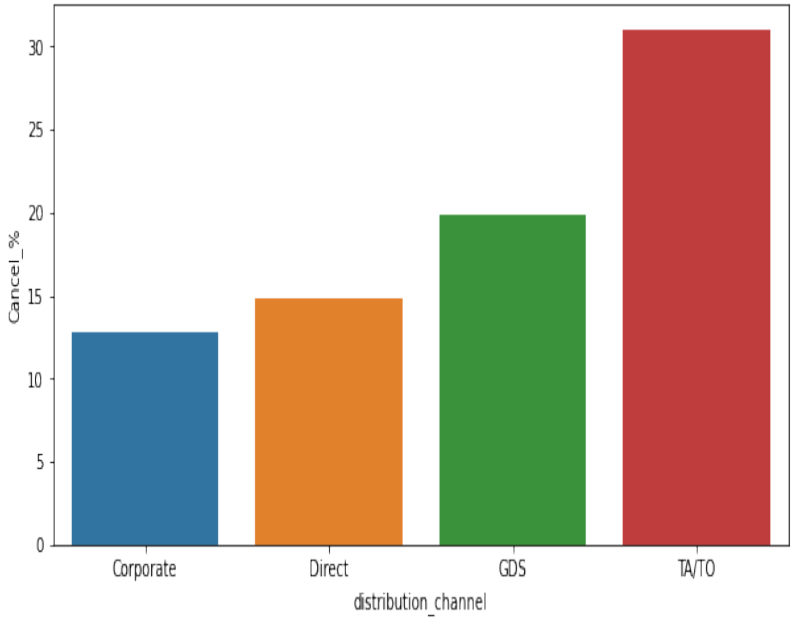


Fig. 16

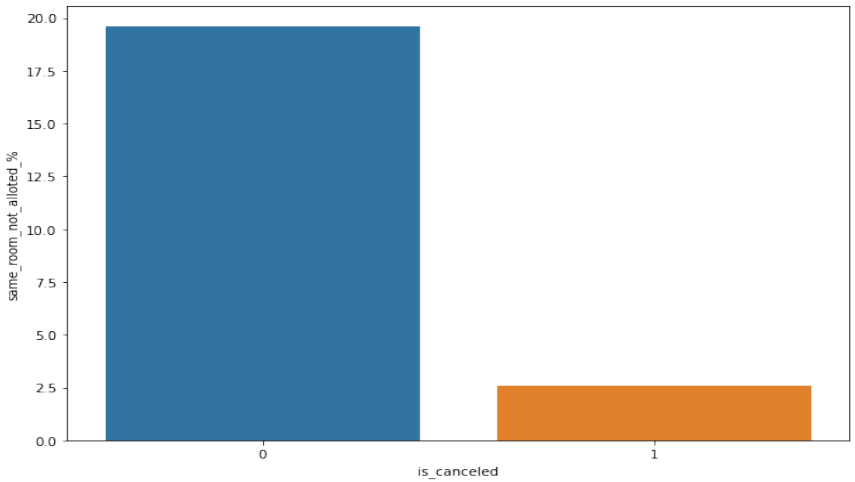


Fig. 17

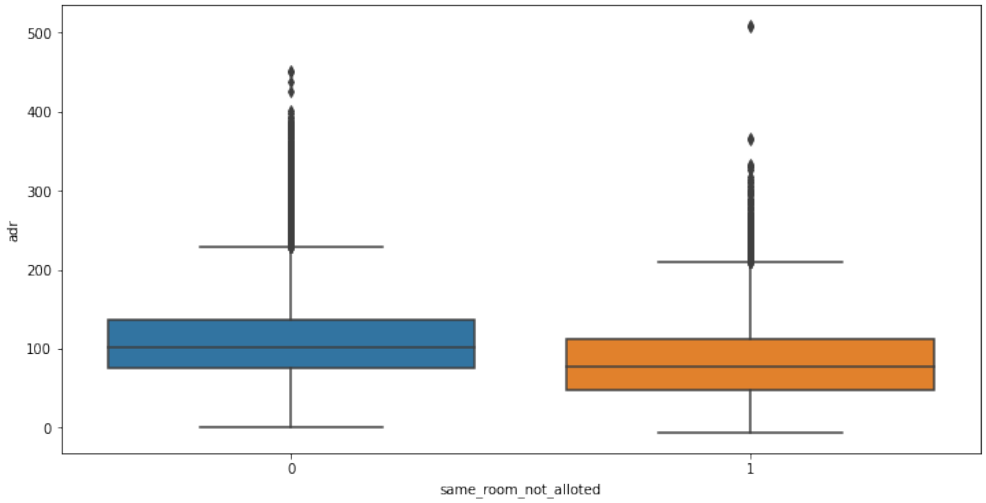


Fig. 18

Longer lead time & Waiting

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.

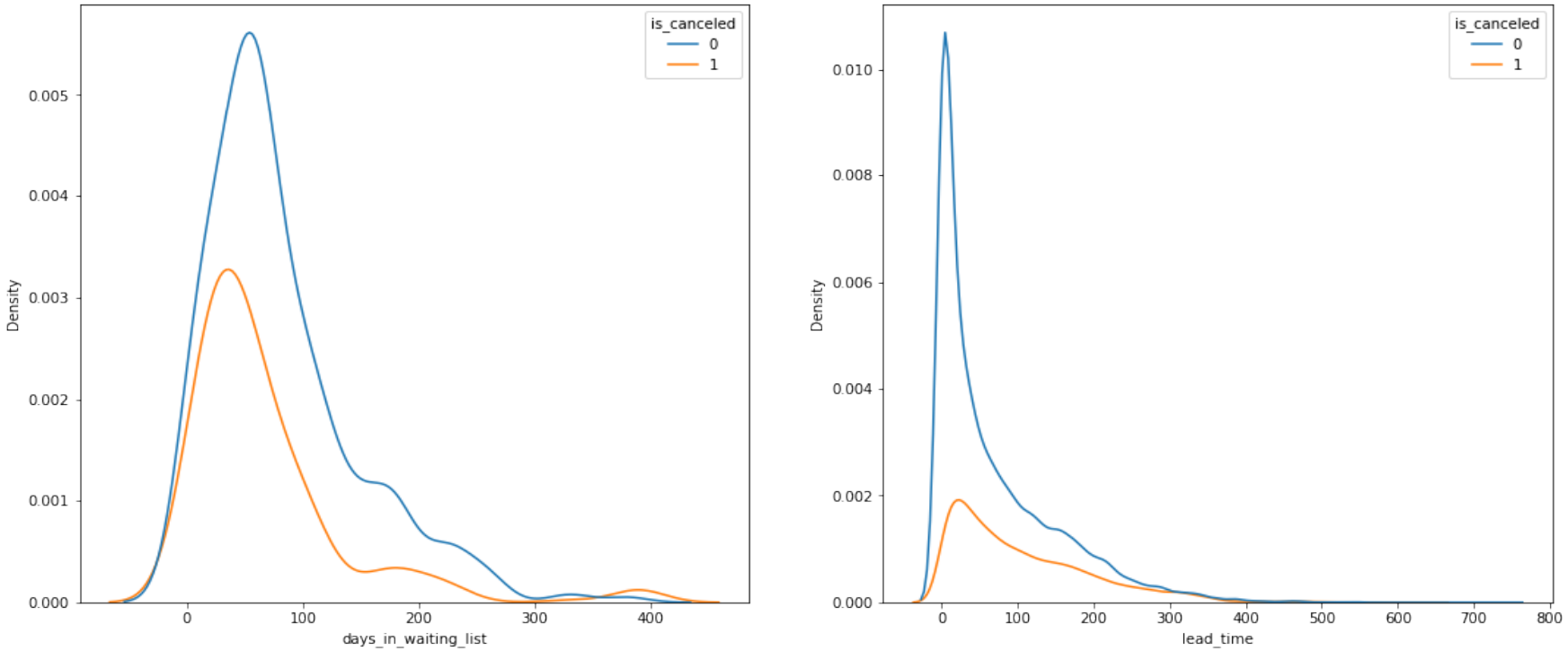


Fig. 19

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?

(3) How does booking numbers and adr changes within a month?

(4) How does bookings varies along year for different types of customers?

Busy Month, Higher adr

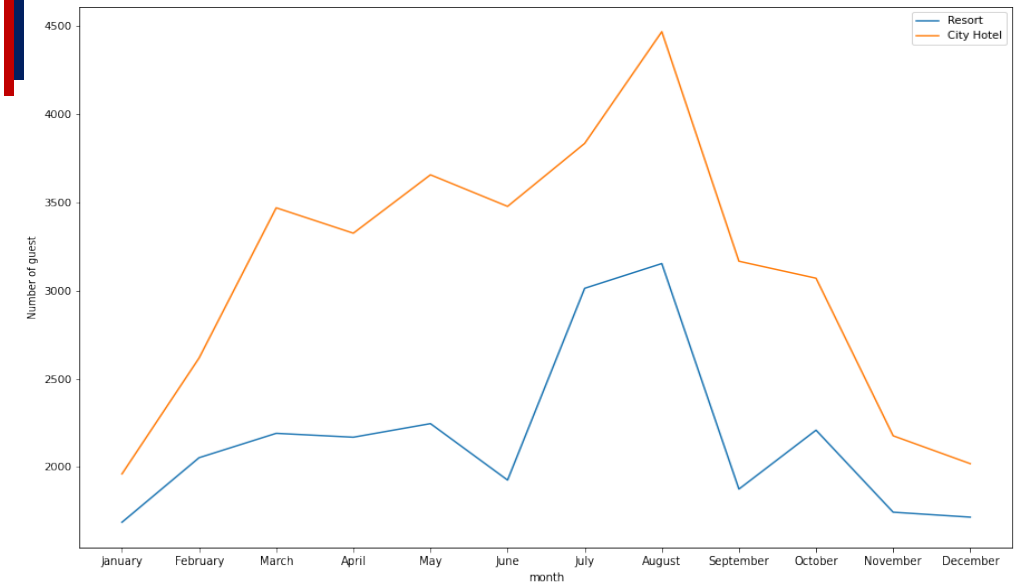


Fig.20

From the month of July to August the number of bookings increased and in August, City Hotel got most number of guests. (Fig.20)

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. (Fig. 21)

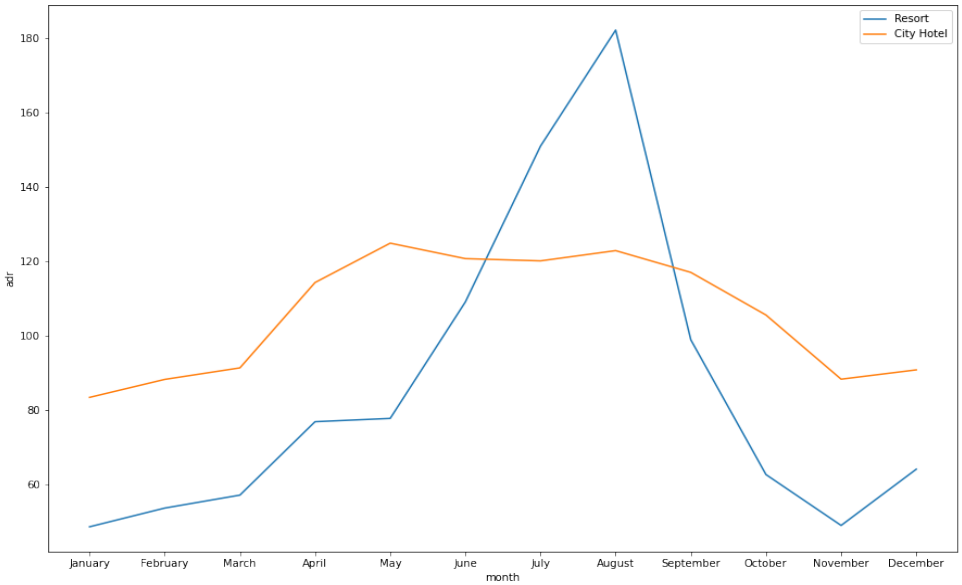


Fig. 21

Booking Numbers & adr within a month

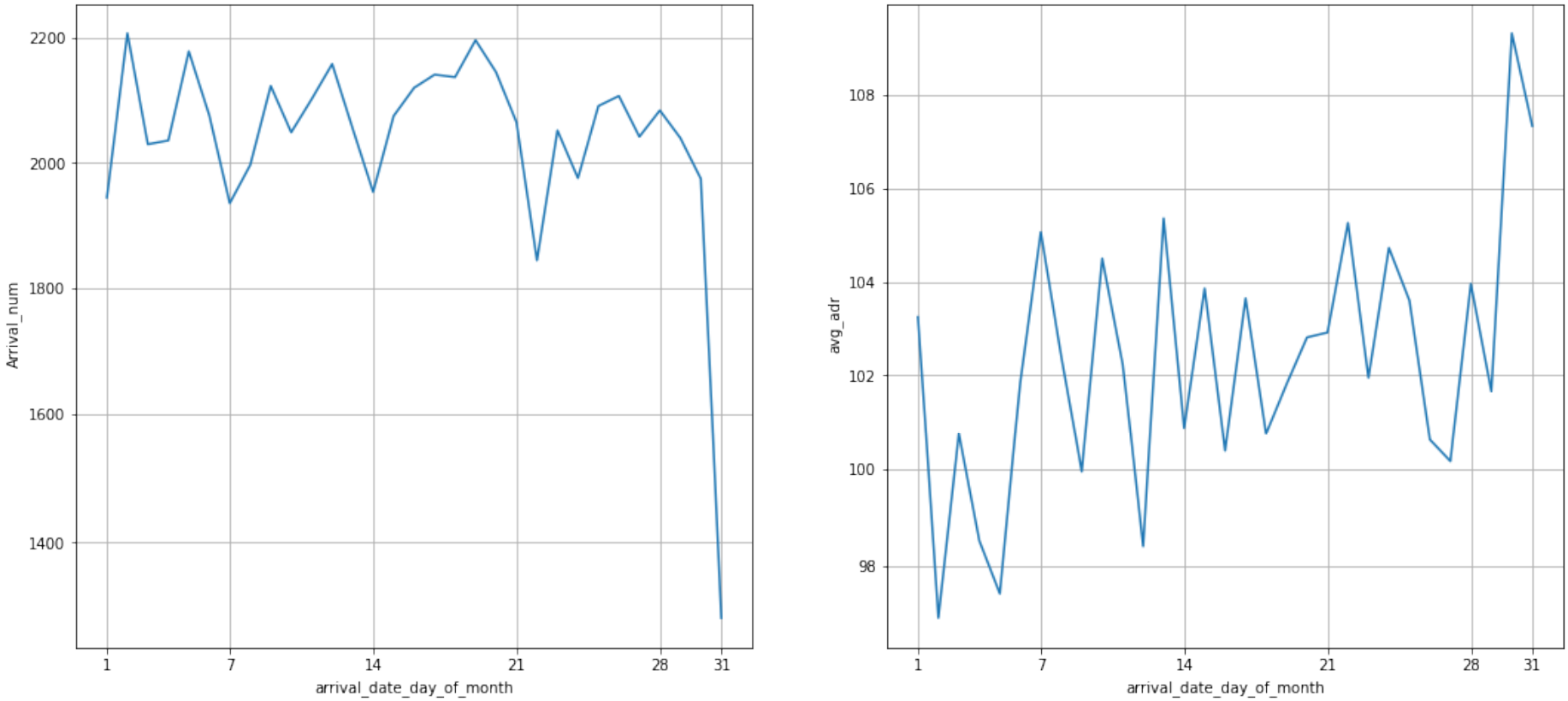


Fig. 22

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

Customer Type

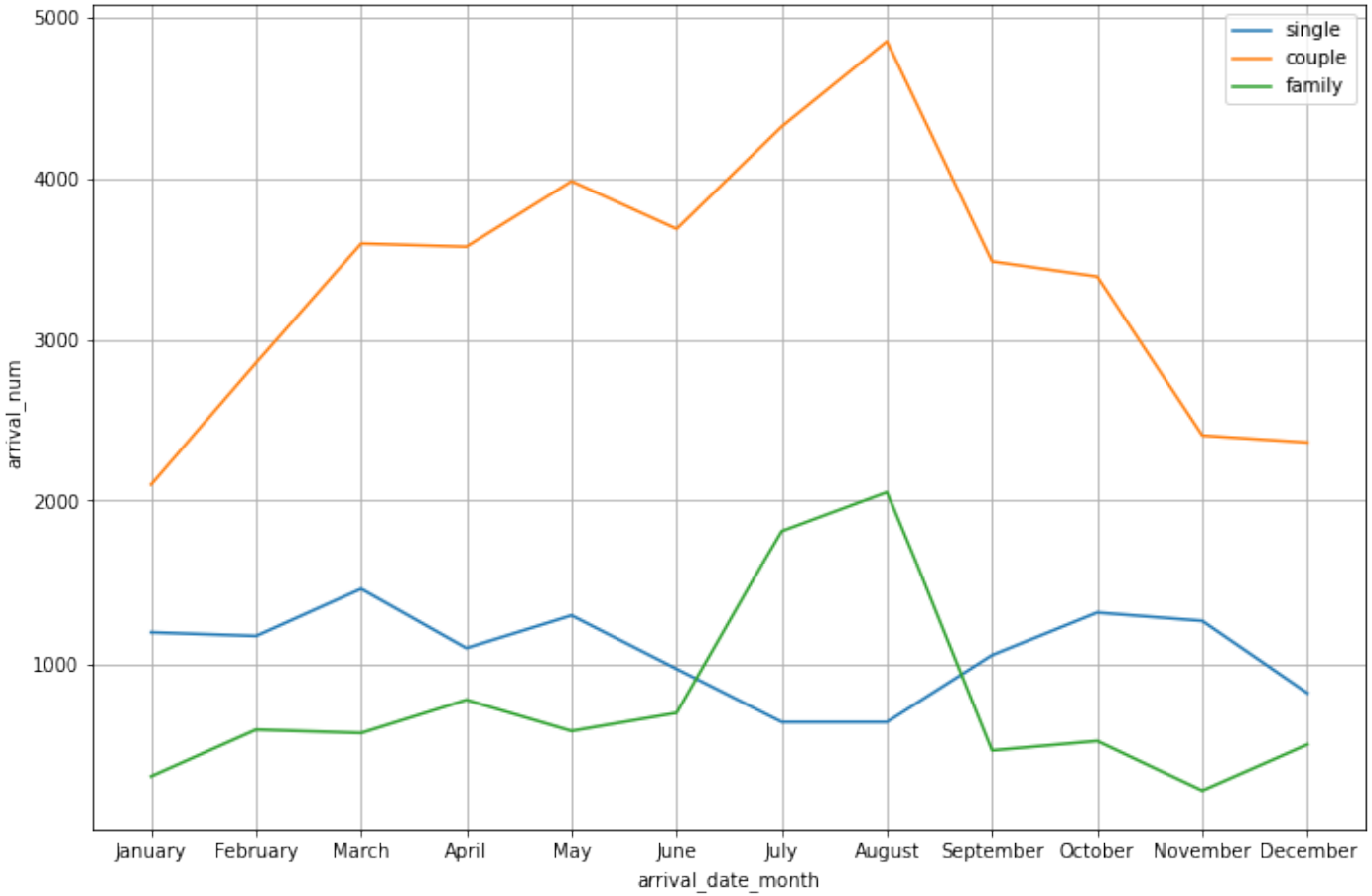


Fig. 23

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:

(1) Form which market segment gets special requests?

(2) What are the different reason for special requests?

(3) From where the most guests are coming ?

(4) How long do people stay at the hotels?

Market Segment: Special request

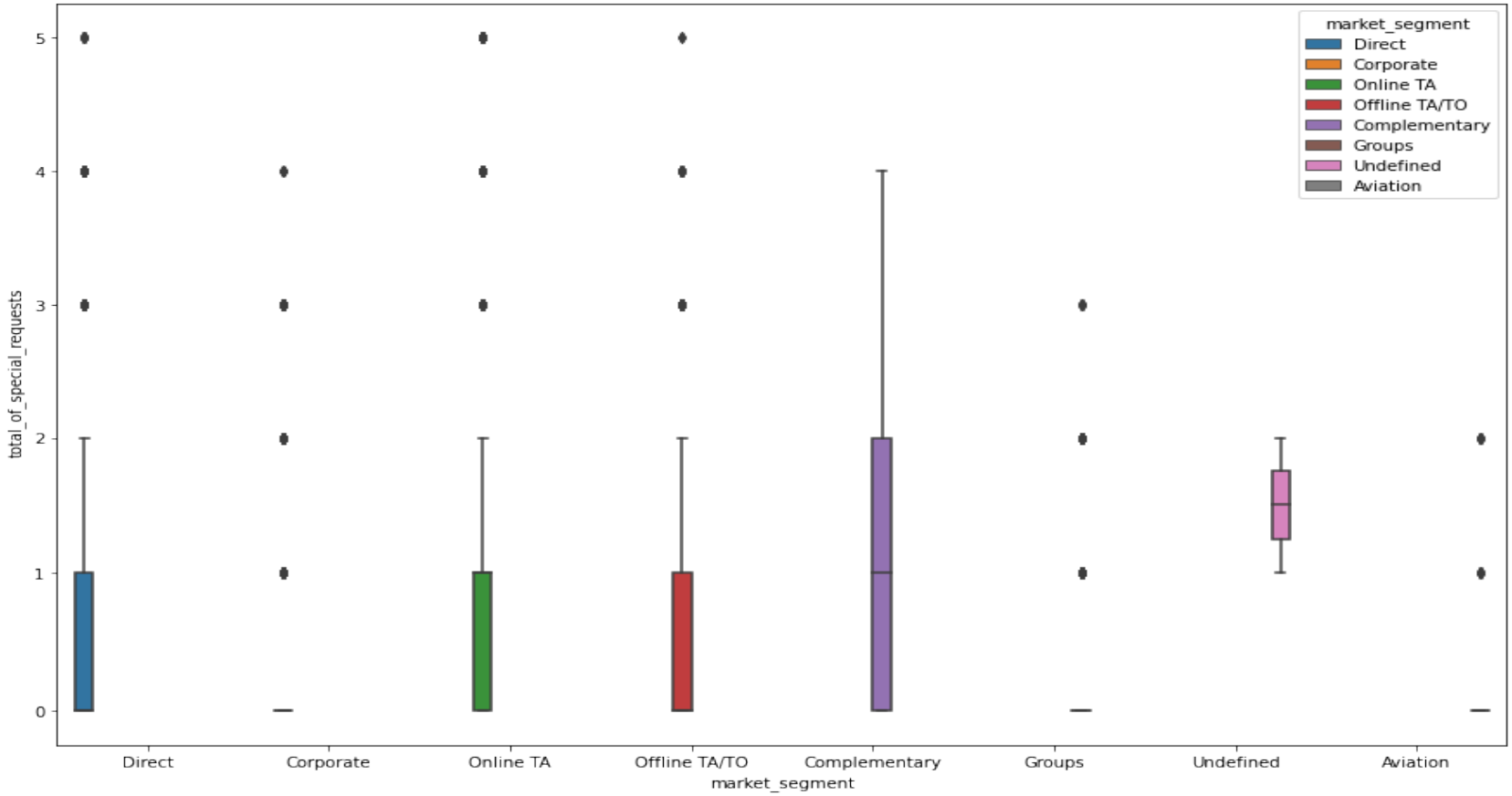


Fig. 24

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.

Special request : Age type

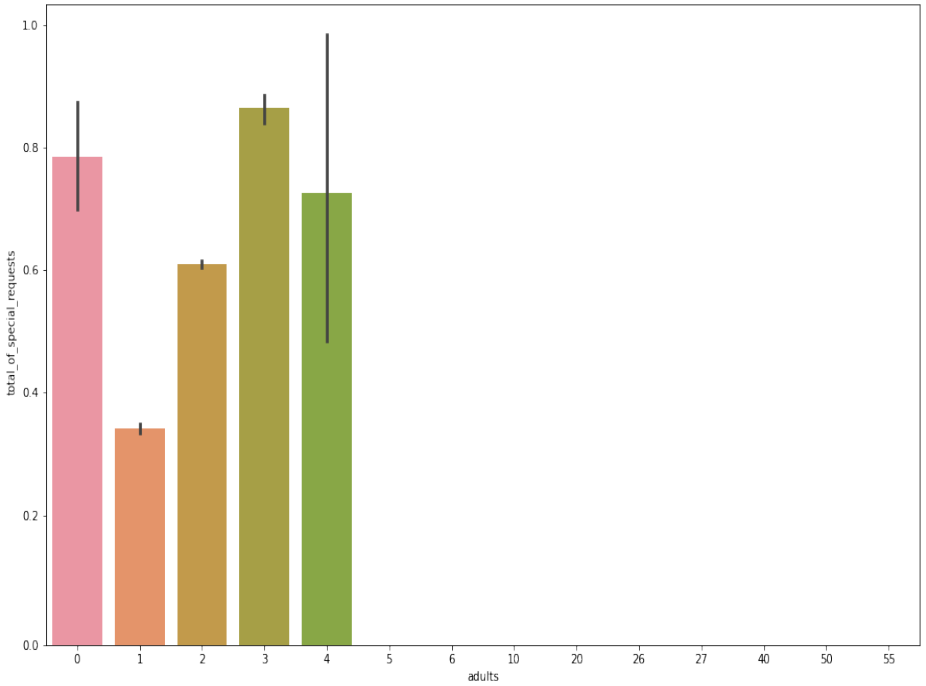


Fig. 25

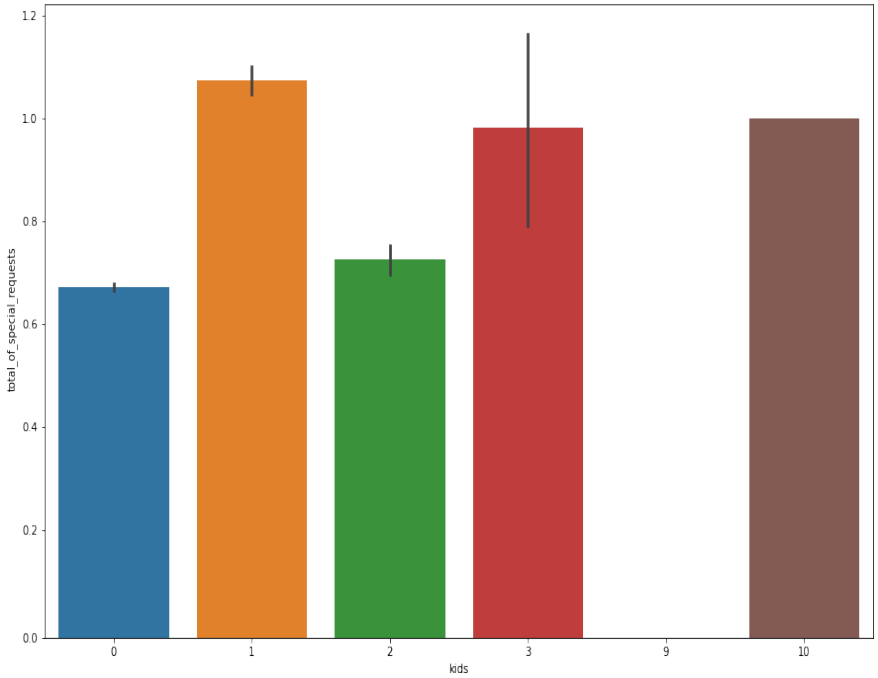


Fig. 26

The number of special request are almost the same in the kids section.

But, we can see that if the adults are more than 2 there are more chances that hotels will receive more special requests.

Guests from different countries, Stay length

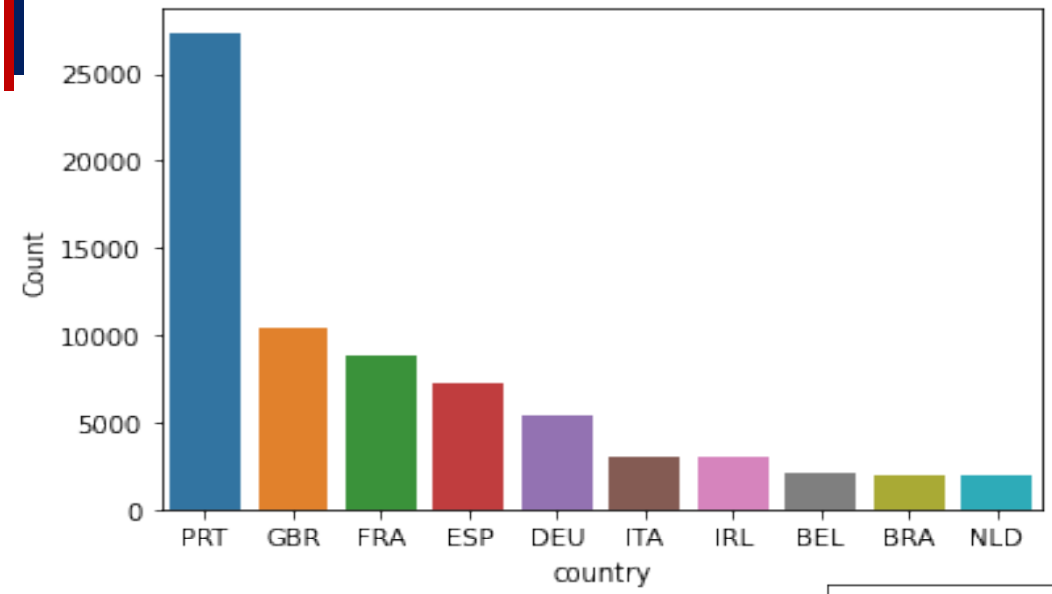


Fig. 27

Most of the customers from European countries like Portugal, Great Britain, France and Spain. (Fig.27)

Most people prefer to stay at the hotels of <=5 nights. (Fig.28)

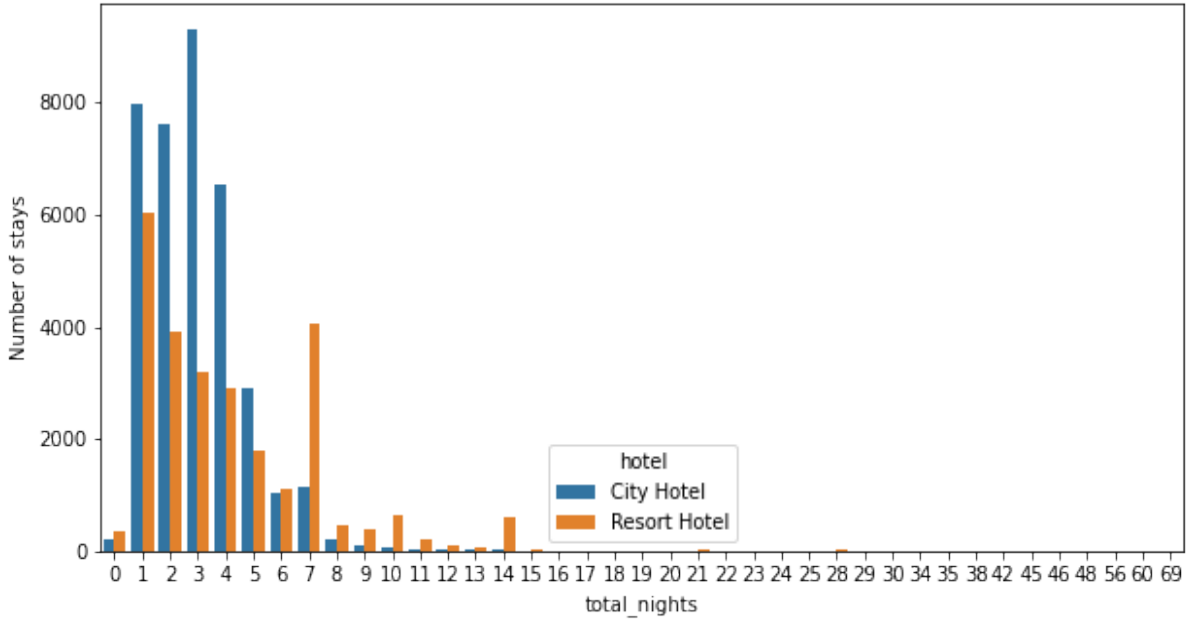
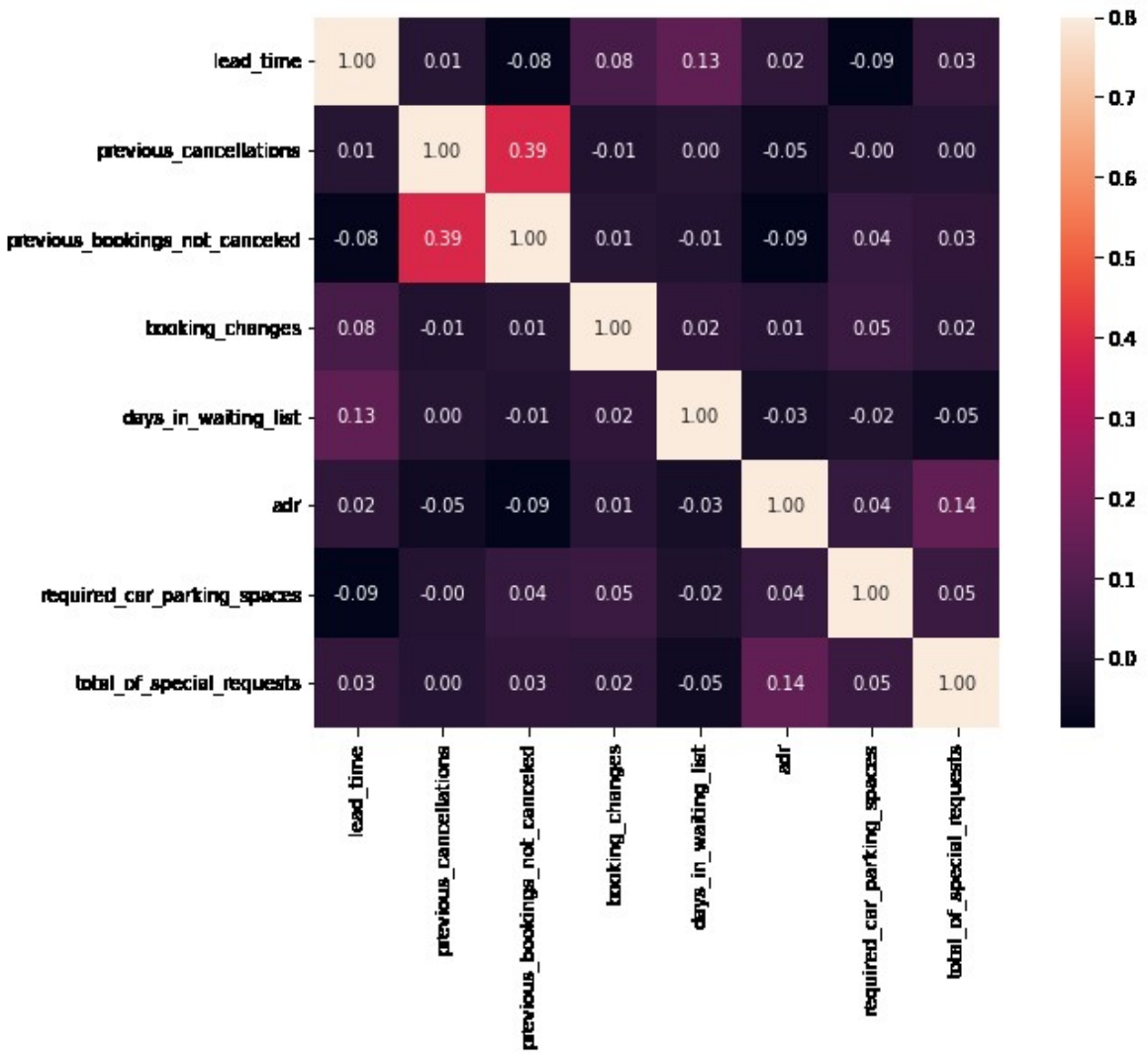


Fig. 28

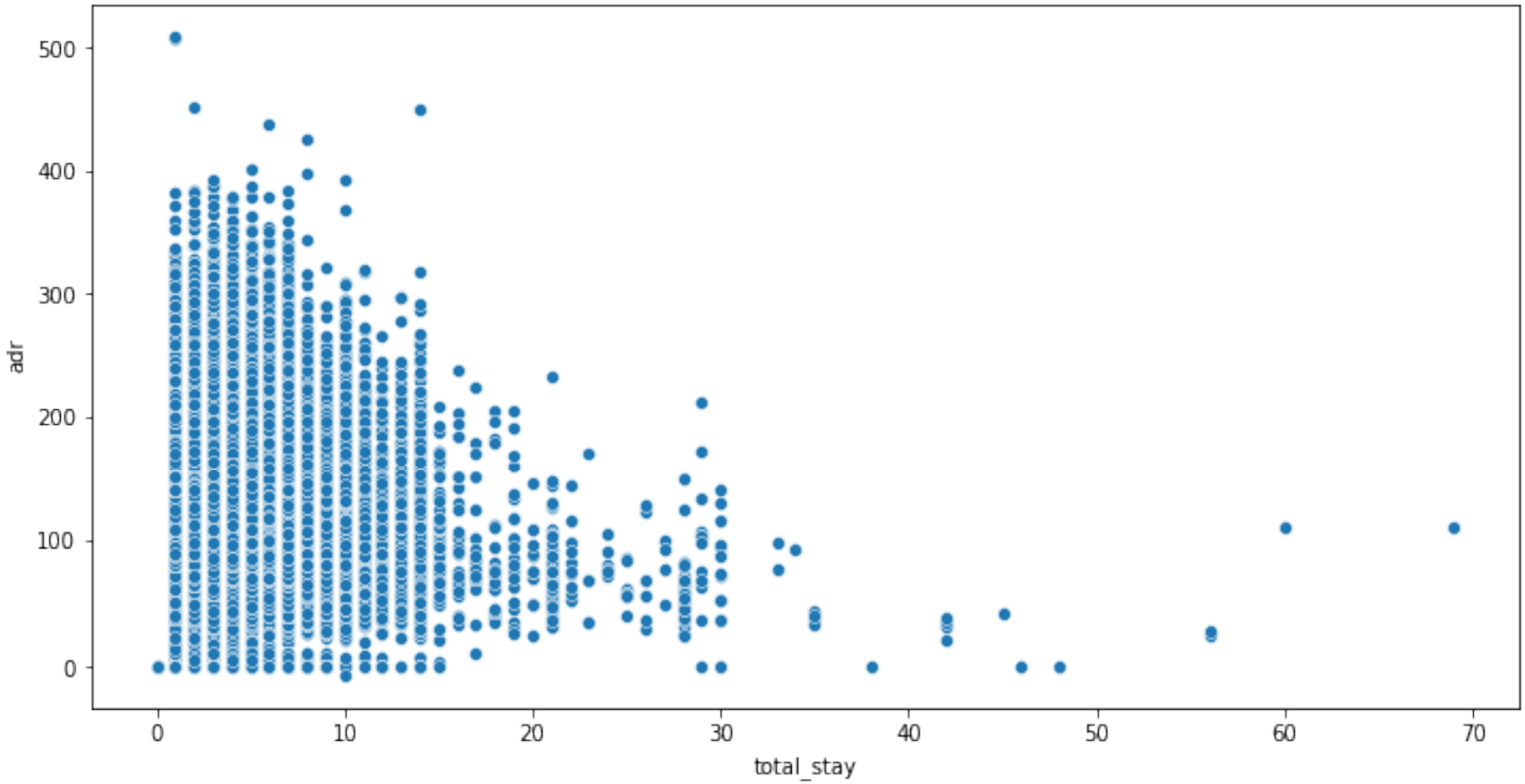
Correlation Heat Map



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.

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

- Number of bookings: Majority of the hotels booked are city hotel. Also the overall adr of City hotel is slightly higher than Resort hotel.
- Duration of stay: Mostly guests stay for less than 5 days in hotel and for longer stays Resort hotel is preferred. For customers, generally the longer stays (more than 15 days) can result in better deals in terms of low adr.
- Booking Cancelation: Both hotels have significantly higher booking cancellation rates.
- Customer returning rate : Very few guests return for another booking in City hotel while return for stay in Resort hotel is more.
- Foreign customers : Most of the guests came from European countries, with most no. of guest coming from Portugal.
- Cancelation via channel: Almost 30% of bookings via TA/TO are cancelled.

Conclusion

- Preferred booking channel: 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.
- Factors affecting cancellation: 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.
- Monthwise review: July- August are the most busier and profitable months for both of hotels. Within a month, adr gradually increases as month ends.
- Type of customer : Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.
- Special requests : Bookings made via complementary market segment and adults have on average high no. of special request.

Thank you