

Capstone Project

Hotel Booking Analysis

Soumabha Sarkar

Points to Discuss:

- Agenda
- Data summary
- Correlation heatmap
- Hotel wise analysis
- Distribution Channel wise analysis
- Booking cancellation analysis
- Analysis based on size of customer groups(Single, Couple, Family).
- Some important questions
- 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 :
 - Correlation between number type data columns.
 - Hotel wise analysis
 - Distribution Channel wise analysis
 - Booking cancellation analysis
 - Analysis based on size of customer groups(Single, Couple, Family).
- By doing this we'll try to find out key factors driving the hotel bookings trends.

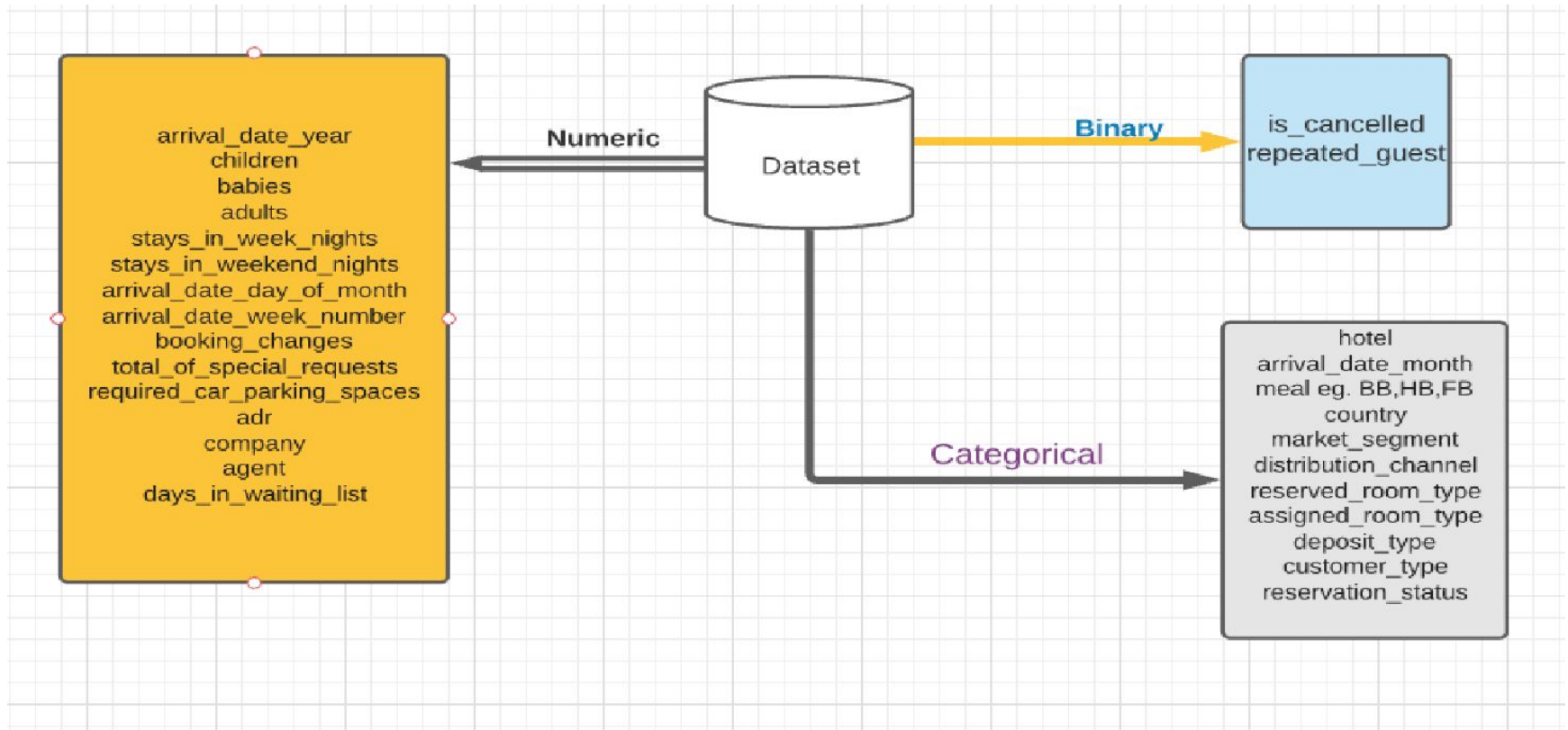
Data Summary

- 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_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.

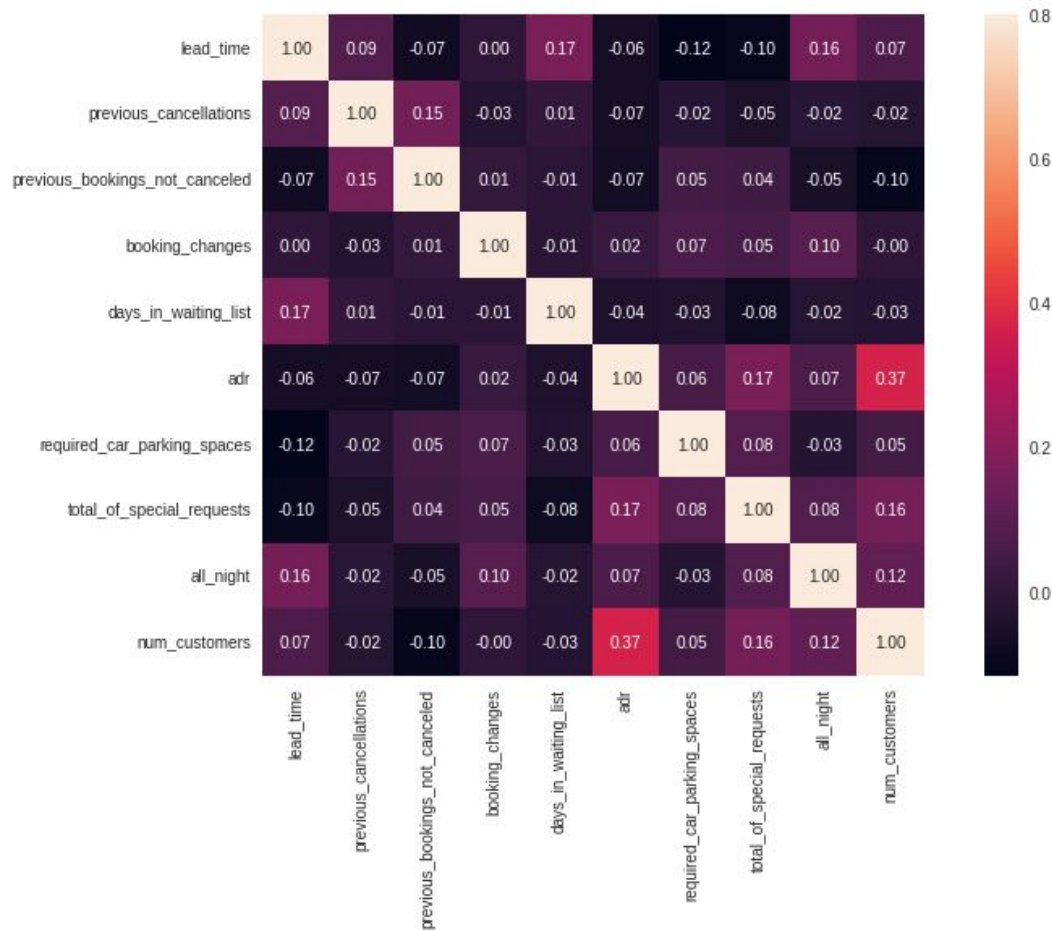
Data Summary

- **market_segment**: This column shows how reservation was made and what is the purpose of reservation. E.g., 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 indicates 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



Correlation between every columns with numerical values. AI

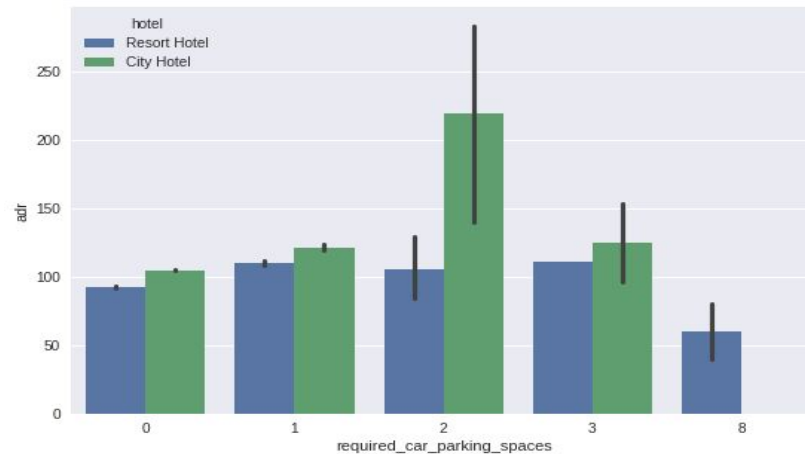
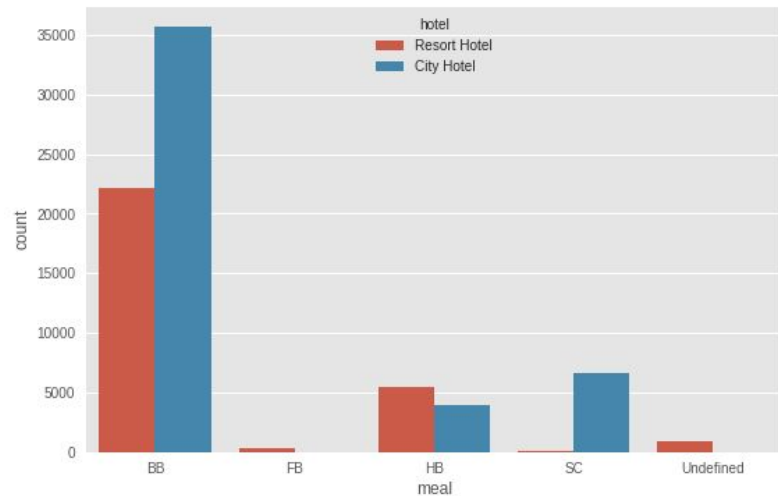


- ADR and number of customers is slightly correlated. This means for more number of customers, ADR is high as service requirement is high in case of more number of customers.

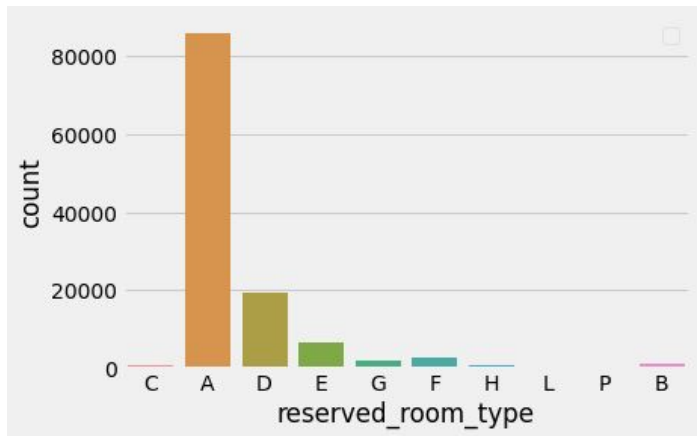
Some basic analysis

Most of the customers are from Portugal followed by Great Britain, France, Spain.

- Demand for BB i.e. Bed and Breakfast high in both of the hotels compared to another meal types.
- Demand for car parking with 2 spaces is high in City hotels.
- Demand for 1 and 3 parking spaces is high in resort hotels.

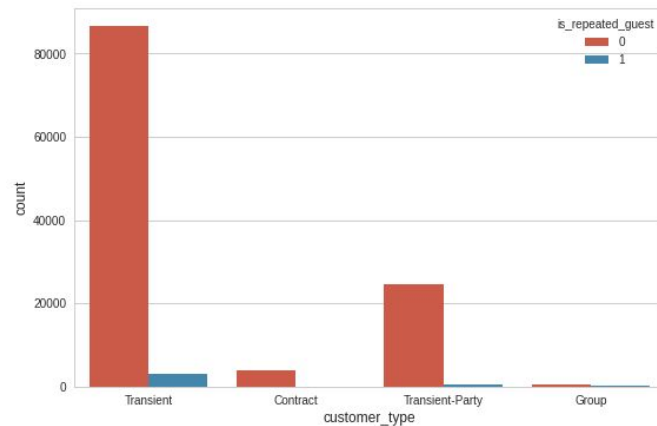
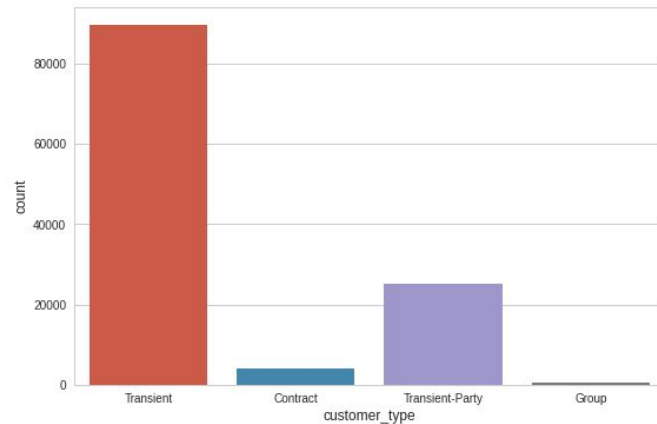


Some basic analysis(Continued)



Most of the customers prefer Type A hotel rooms

- Most of the customers are of Transient type. They are staying for short time periods.
- Though there is very nominal margin of repeated customers, most of them are transient type.



Hotel wise analysis.

- The booking ratio between Resort Hotel and City Hotel
- Average ADR variation Hotel wise.
- Different hotels booked month wise.
- Different hotels booked year wise.
- Staying at night in different hotels.

The booking ratio between Resort Hotel and City Hotel



- The percentage of booking in City hotel is above 61.5%.
- The percentage of booking in Resort hotel is 38.5%.

Average ADR variation Hotel wise.



- As we have seen in the previous chart that demand in City hotel is high so the adr average is also high. It can be said that City hotel is generating more revenues than Resort hotels.

Different hotels booked month wise.



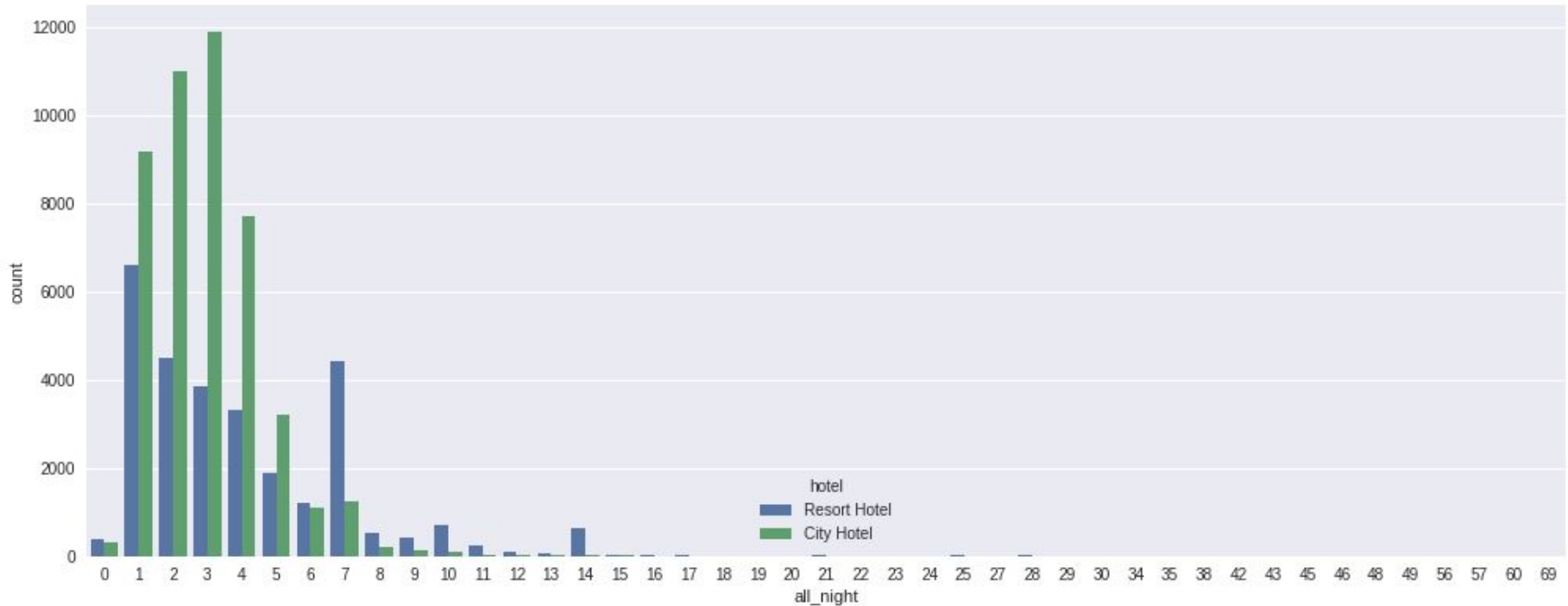
- City Hotel booking is high in July and August month.
- Resort Hotel booking is high in July and August month also but the demand is low compared to City hotel.

Different hotels booked year wise.



- Hotels were mostly booked in 2016
- Demand in City Hotels are high in every year.
- In 2016, which is the most demanded year, City hotel demand is higher by 1.75x compared to Resort hotel.

Staying at night in different hotels.

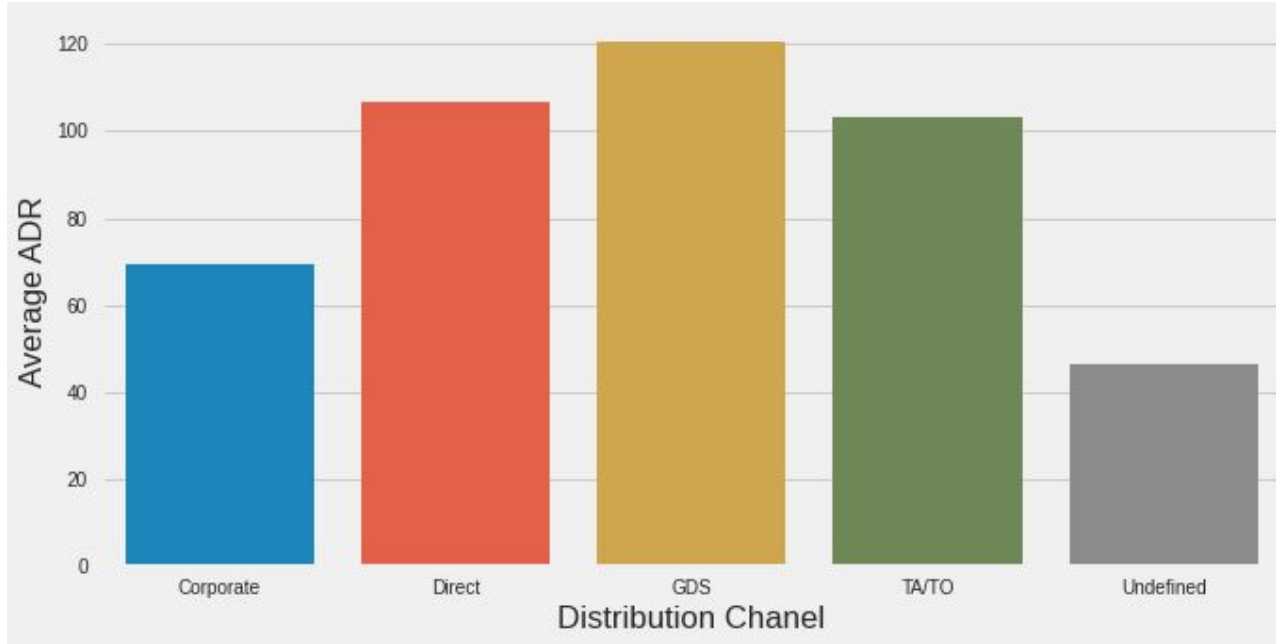


- Customers who are preferring City hotels, stays mostly multiple nights of range one to five.
- Customers who are preferring Resort hotels, stays mostly multiple nights of range one to seven.

Distribution Channel wise analysis

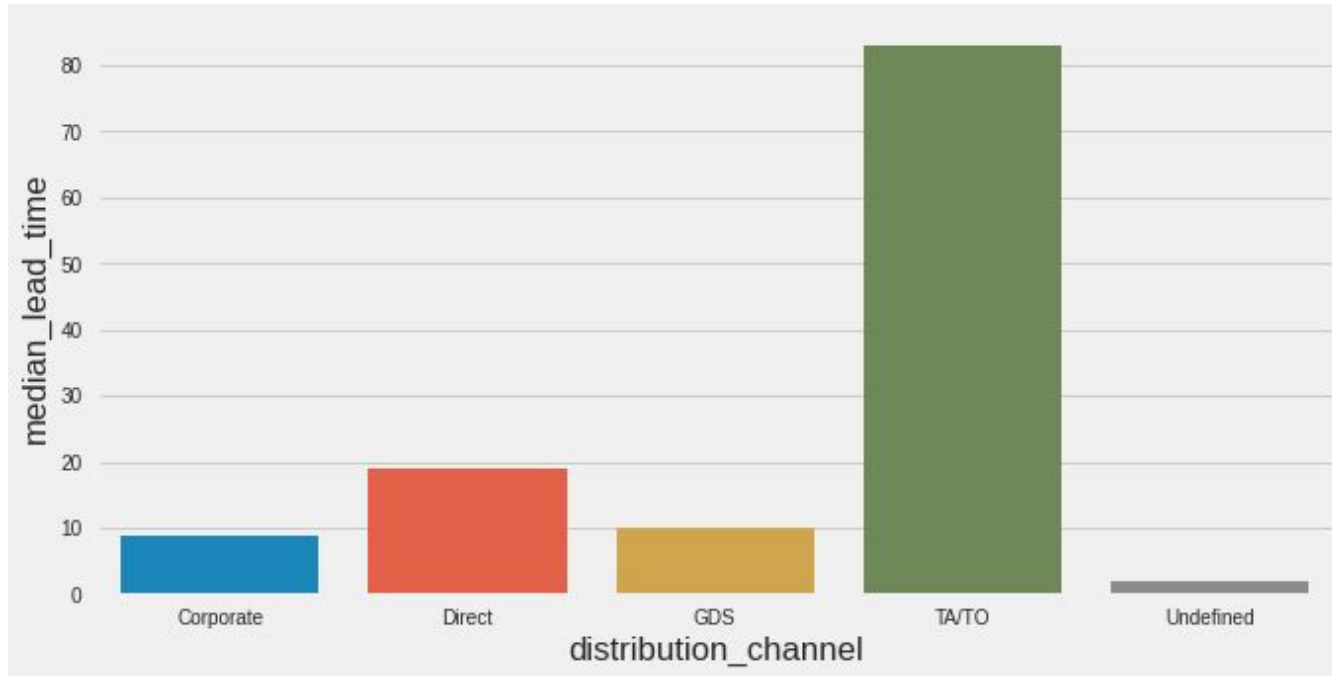
- Distribution channel wise average ADR.
- Distribution channel wise lead time variation.

Distribution channel wise average ADR.



- Average ADR is highest in GDS channel, followed by Direct and TA/TO.

Distribution Channel wise lead time variation.

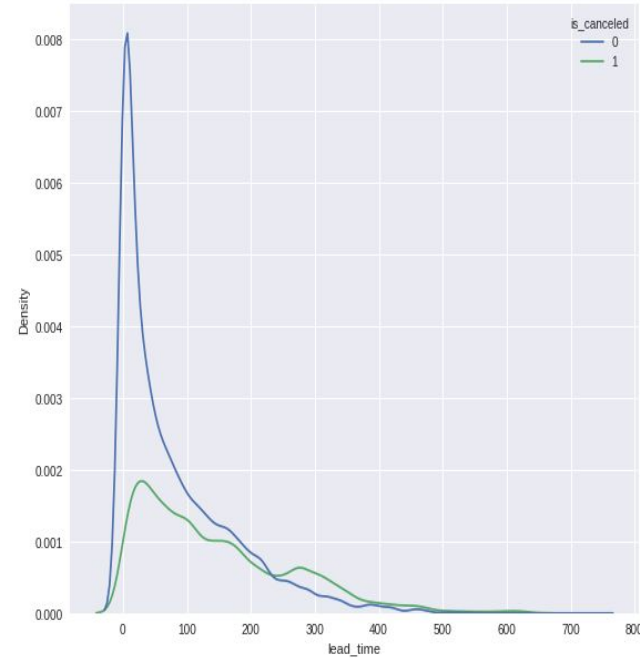
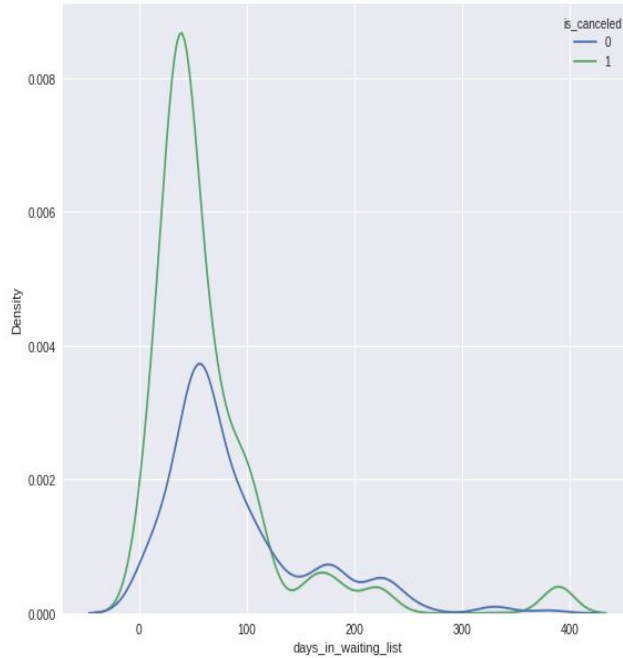


- For TA/TO i.e. Travel Agencies/ Travel Operator lead time is high. If any customer is booking hotels through TA/TO they should book the hotel several days before compared to other booking channels.

Booking cancellation analysis

- Cancellation based on lead time and waiting time.
- Hotel cancellation on various months.
- Cancellation of hotels in various distribution channels.

Cancellation based on lead time and waiting time. AI



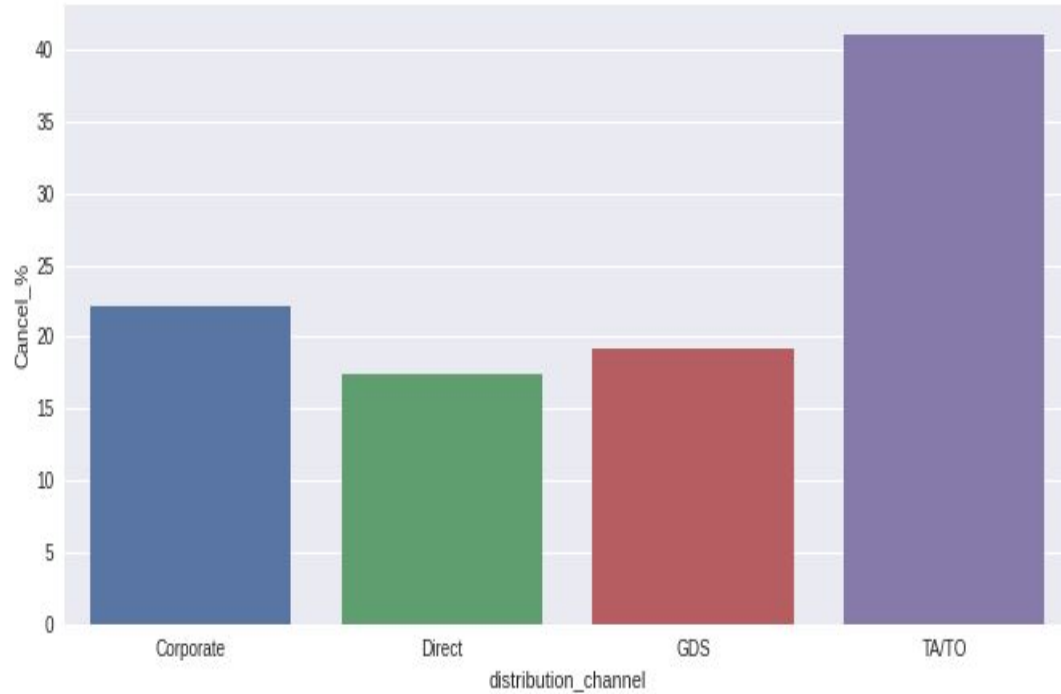
- In both of the above plot it is seen that customers who cancelled and those who did not cancelled the bookings have lead time and waiting list mostly in the range of 150 days. So we can say lead time and waiting days have no relation with booking cancellation.

Hotel cancellation on various months.



- It can be concluded that cancellation of booking is highest in August.
- April, May, June, July have almost same number of cancellation count.

Cancellation of hotels in various distribution channels.

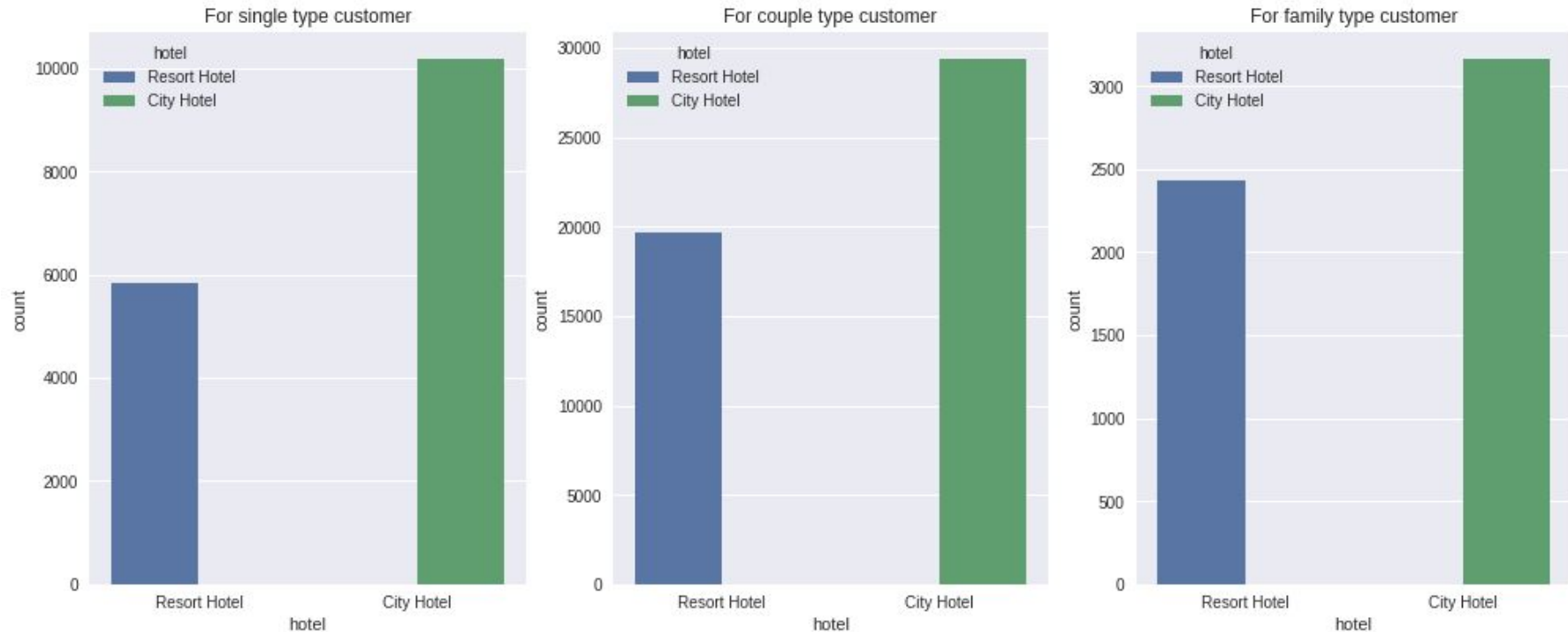


- Cancellation percentage of cancellation of hotels is higher in Travel Agencies/ Travel Operators.

Analysis based on size of customer groups(Single, Couple, Family).

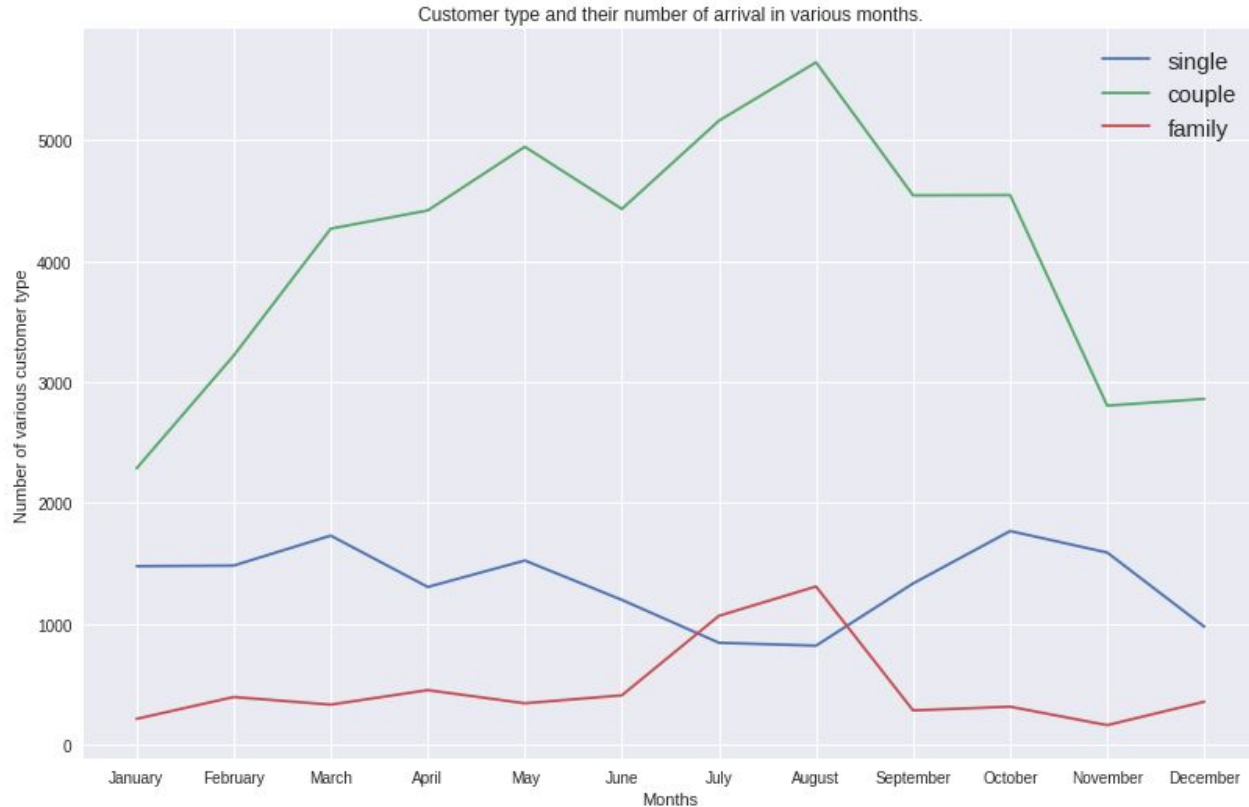
- Various customer type and their hotel preferences.
- Various customer types and their preferable booking months.

Various customer type and their hotel preferences.



- There are mostly couple type customers i.e. two adults.
- For any type of customers (Single, Couple, Family) preferences is higher in booking City hotels than Resort hotels.

Various customer types and their preferable booking months.



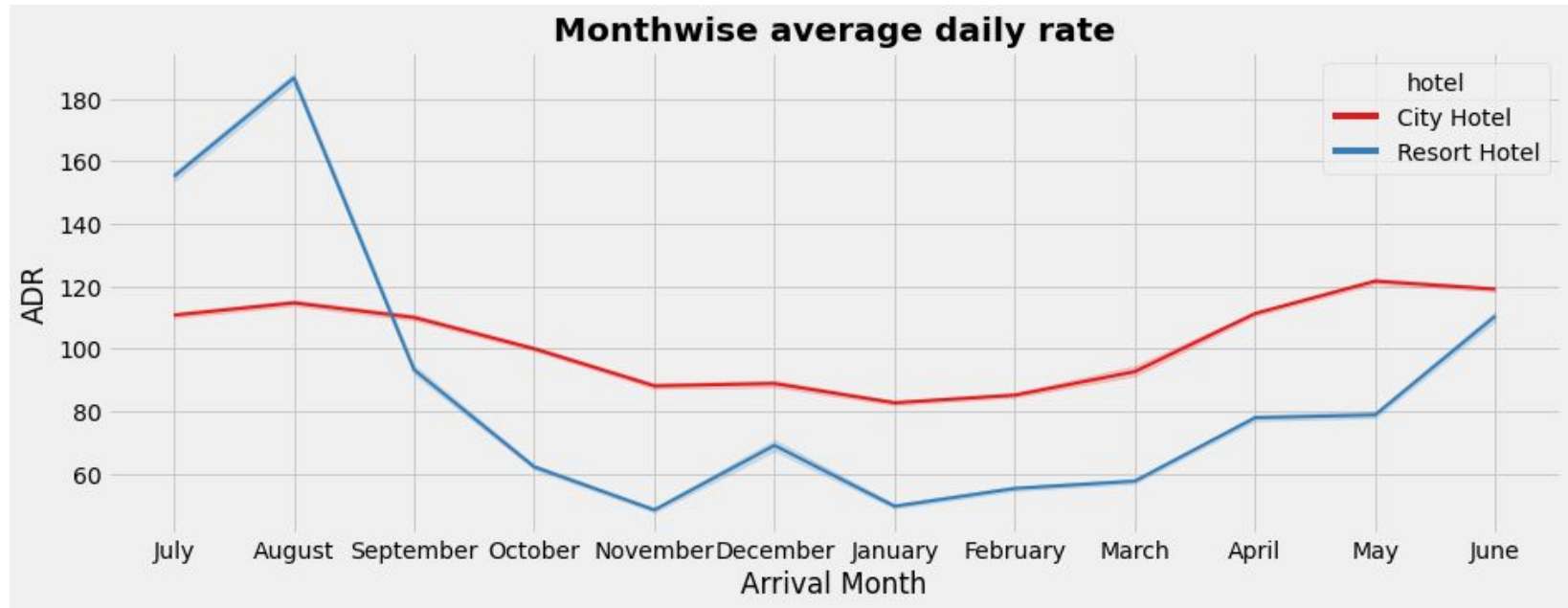
- In case of Couple and Family type customers preferences are high in booking hotels in July and August.
- But in case of Single type customers they are preferring March and October mostly.

- We can predict that Singles are preferring Summer and Winter holidays mostly.

Some important questions

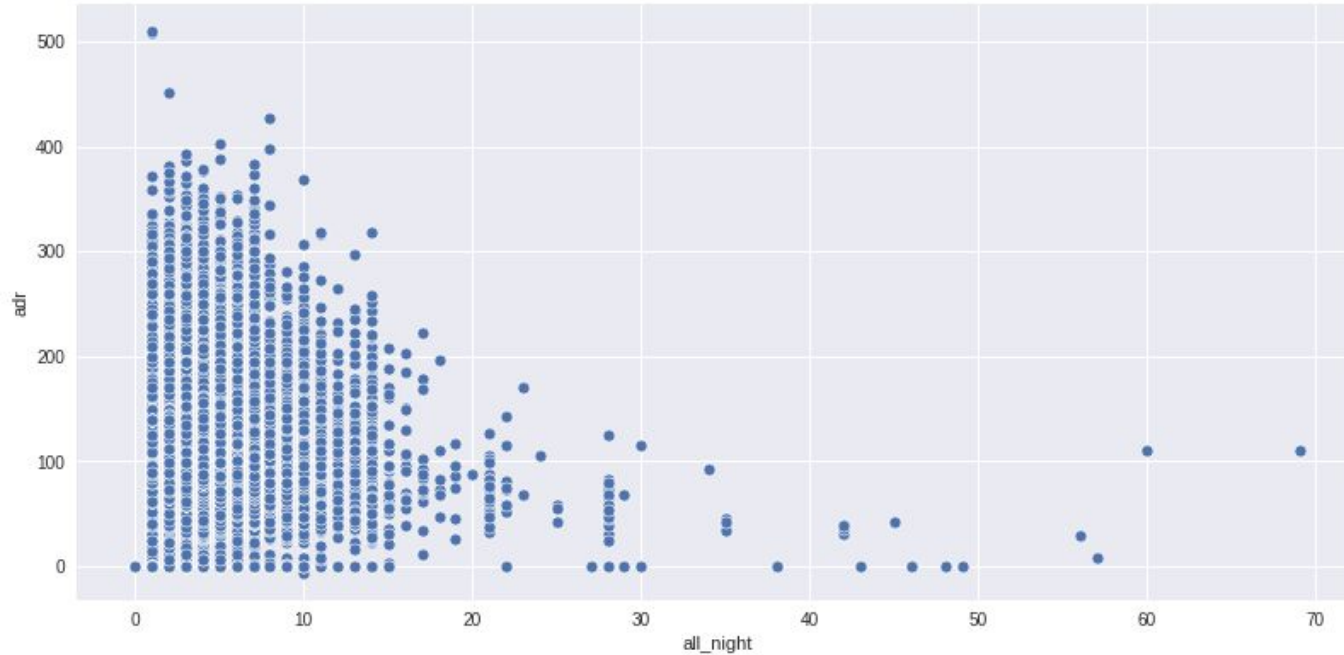
- When the best time of year to book a hotel room is?
- The optimal length of stay in order to get the best daily rate?
- What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

When the best time of year to book a hotel room is?



- As the demand is high in City hotels in the month of July and August so the ADR is also high for those particular months.
- ADR is higher in May, June in case of Resort hotels. Customers are preferring summer times in Resort hotels.
- The optimal time for booking the hotels is January and December to get the best daily rates

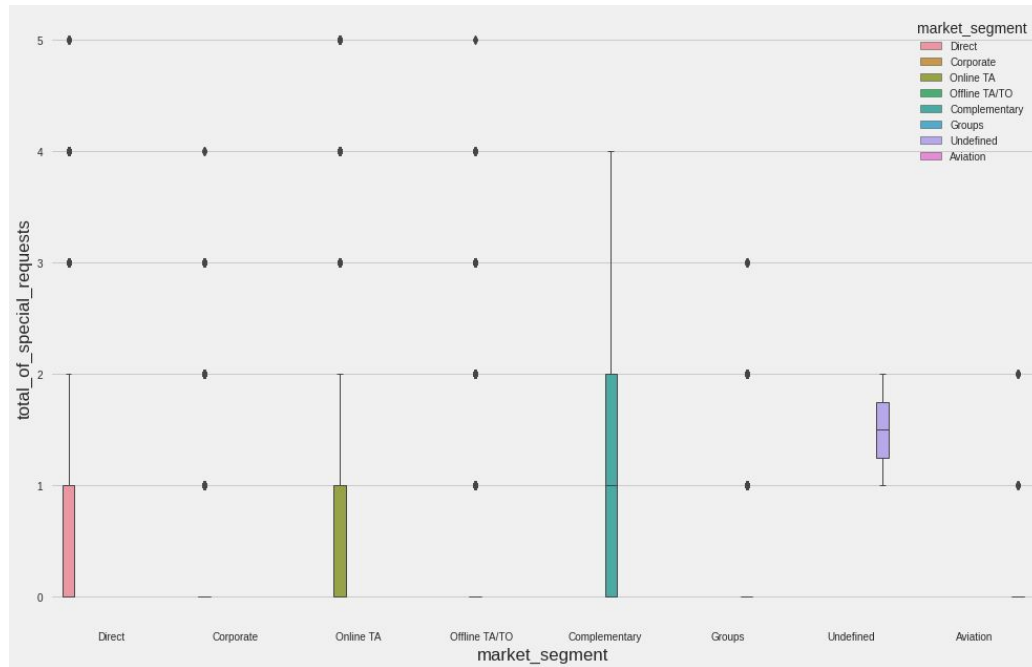
The optimal length of stay in order to get the best daily rate?



Customers who are staying in the hotels more than 20 days can get the best daily rate.

What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

- We can conclude from this chart that from every market segment the hotels are getting disproportionately high number of special requests.



Conclusion

1. Around 61.5% bookings are for City hotels and 38.5% bookings are for Resort hotels, therefore City Hotels are busier than Resort hotels. Also the overall ADR of City hotels is slightly higher than Resort hotels.
2. Mostly guests stay for less than 5 days in hotel and for longer stays Resort hotel is preferred.
3. Most of the customers came from European countries, with most of them coming from Portugal.
4. Customers use different channels for making bookings out of which most preferred way is TA/TO.
5. For hotels higher ADR deals come via GDS channel, so hotels should increase their popularity on this channel.
6. Nearly 30% of bookings via TA/TO are cancelled.
7. More number of people in guests results in more number of special requests.
8. Most of the customers prefer Type A rooms.
9. Longer lead time and waiting time do not affect cancellation of bookings.
10. July-August are the most busier and profitable months for both of hotels.
11. Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.
12. Singles prefers summer and winter holidays over other months. Hotels should provide better deals for singles at that time.
13. For customers, generally the longer stays (more than 20 days) can result in better deals in terms of low ADR.
14. City hotels should provide more two spaced parking and Resort hotels should provide more one and three spaced parking in order to get more bookings.