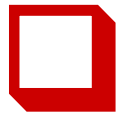


EDA Capstone Project

On
Hotel Booking Analysis
by
Sweta Seal



EDA(Exploratory Data Analysis)

□ Hotel wise Analysis

- Which hotel is mostly booked by customers?
- How long does people stay in the hotels?
- How does lead time effects cancelation of hotel ?
- What are the hotel preferences of distributors?

□ Booking Analysis

- Which is the most common customer_type ?
- Which country has the highest booking ?
- What are the effects of deposit on bookings ?
- Is assigned room type the causes for high cancelation?

□ Market Analysis

- Which segment of the market usually has the least number on the waiting list?
- Which segment of the Distribution_channel usually has the least number on the waiting list?
- Which hotel is mostly preferred by Distributors ?

□ Time Analysis

- How average daily rate (adr) changes with different months ?
- What is the busiest month of the year?
- Which is most preferred weekdays or weekends ?
- How average daily rate (adr) changes with Customer Type ?
- What is the chances of repeated guest cancelling the booking ?
- Which segment of distribution channel have highest bookings and cancellations?
- Which market_segment have highest bookings and cancellations?
- Which type of rooms are mostly reserved ?

Data Summary

This dataset contains 119390 observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and 31st of August 2017, including booking that effectively arrived and booking that were canceled.

The data has 32 columns :

1. Hotel : Resort hotel or city hotel
2. is_canceled : Value indicating if the booking was canceled (1) or not (0)
3. lead_time : Number of days that elapsed between the reservation and the arrival date
4. arrival_date_month : Month of arrival date
5. arrival_date_week_number: Week number of year for arrival date
6. stays_in_weekend_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
7. stays_in_week_nights : Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
8. adults :Number of adults
9. children :Number of children
10. babies :Number of babies
11. meal :Type of meals booked
12. Country : The different country of origin
13. Arrival_date_year :The year of arrival

- 15.booking_changes :Number of changes made to the booking
- 16.deposit_type :Indication on if the customer made a deposit to guarantee the booking. Three categories:
- a. No Deposit – no deposit was made;
 - b. Non Refund – a deposit was made in the value of the total stay cost;
 - c. Refundable – a deposit was made with a value under the total cost of stay
- 17.days_in_waiting_list :Number of days the booking was in the waiting list before it was confirmed to the customer
- 18.Adr :Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
- 19.required_car_parking_spaces :Number of car parking spaces required by the customer
- 20.total_of_special_requests :Number of special requests made by the customer (e.g. twin bed or high floor)
- 21.market segment : Market segment designation.
- a. Corporate
 - b. Direct
 - c. Complementary
 - f. Aviation
 - d. Offline TA/TO
 - e. Online TA

22.distribution_channel :Booking distribution channel.

- a. Corporate
- b. Direct
- c. GDS
- d. TA/TO

23..customer_type

- a. Contract - when the booking has an allotment or other type of contract associated to it;
- b. Group – when the booking is associated to a group;
- c. Transient – when the booking is not part of a group or contract, and is not associated to other transient booking;
- d. Transient-party – when the booking is transient, but is associated to at least other transient booking

24.Is repeated guest: when the guest has visited before this time.

25. previous cancellations : Number of previous bookings that were cancelled by the customer prior to the current booking

26.Previous bookings not canceled : Number of previous bookings not cancelled by the customer prior to the current booking

27.Reserved room type : Has different category of rooms

28.Assigned room type :The booked assigned for staying

29. booking changes : Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

30.agent : There are many agent

31.company :ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons

32.reservation status date:Date at which the last status was set. This variable can be used in conjunction with the reservation status to understand when was the booking cancelled or when did the customer checked-out of the hotel

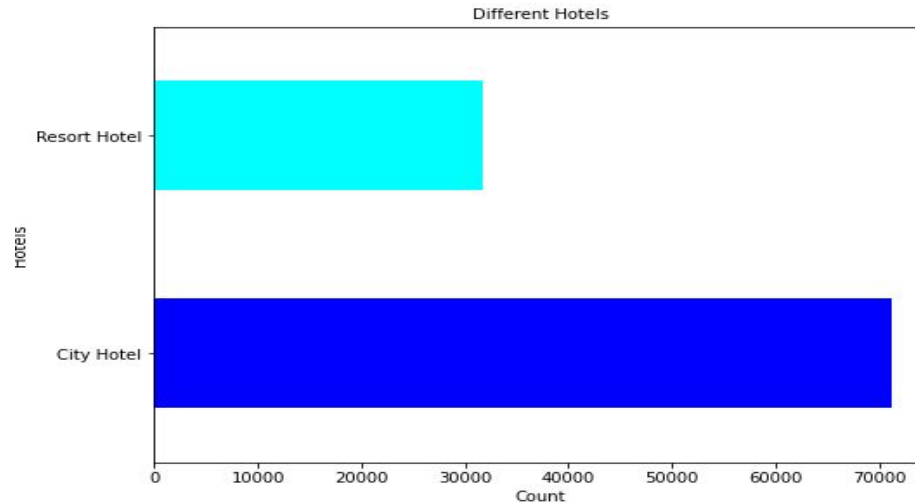
❑ Problem Statement

- Have you ever wondered when the best time of year to book a hotel room is? Or 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? This hotel booking dataset can help you explore those questions!
- This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.
- Explore and analyze the data to discover important factors that govern the bookings.

❑ Hotel booking Preferences

There are two types of hotel that is City hotel and Resort City .

We see that the bookings for city hotel is comparatively higher which can be because factors like hospitality of the hotels, meals, room rates etc

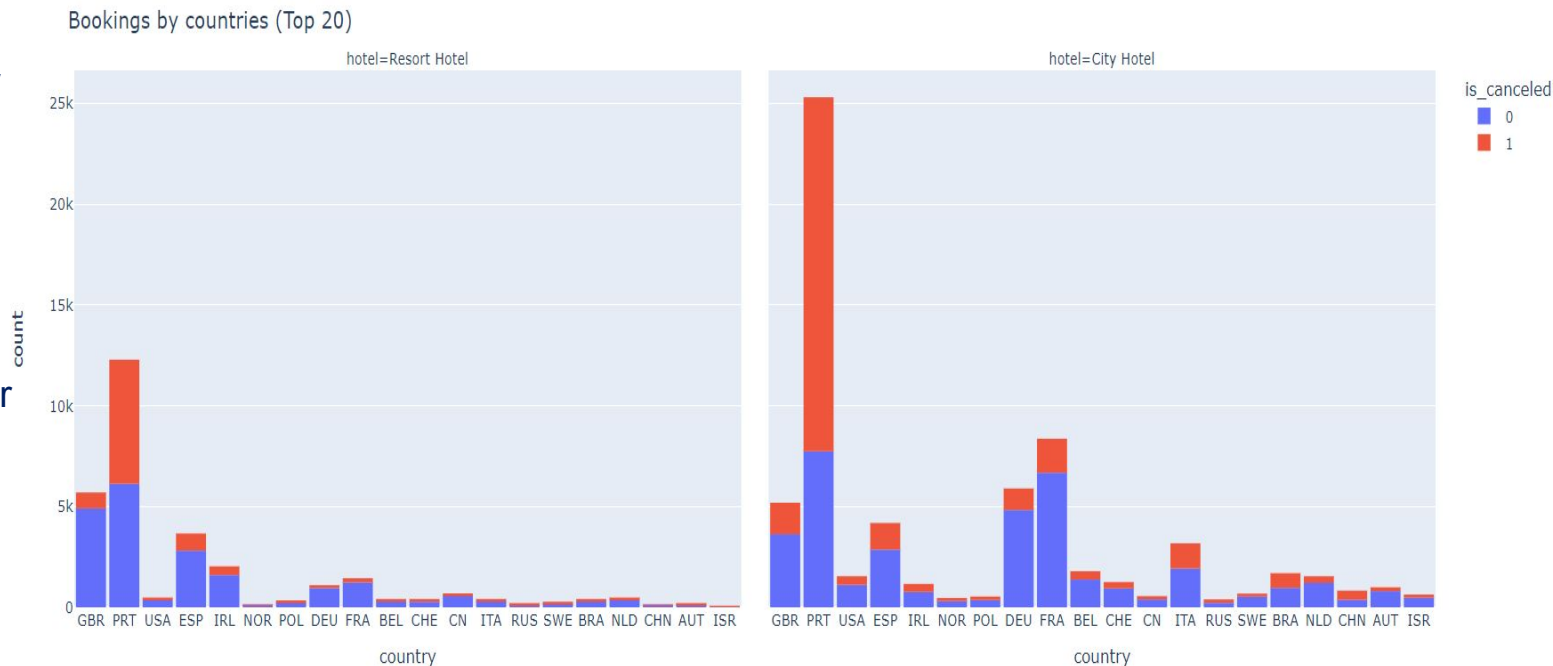


Top 20 countries:

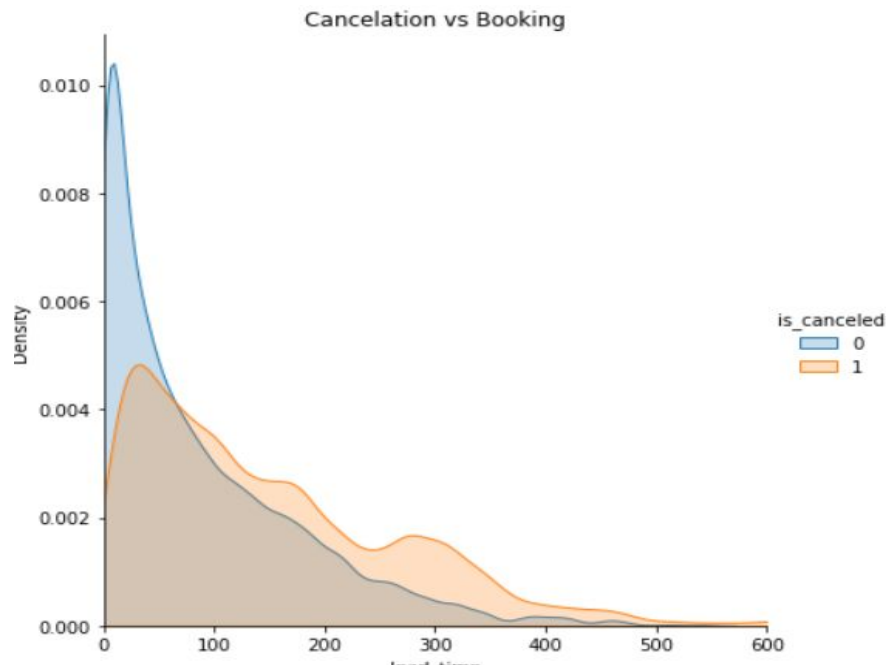
There are total 174 unique countries among we took top 20 countries to check the booking variation
The guests from Portugal has highest booking and cancelation ratio compared to all other country in both the hotels.

France has the next highest bookings for City hotel whereas ISR has next highest for city hotels.

Amongst total booking made for each country Portugal shows highest cancelation.



❑ Cancellation factors



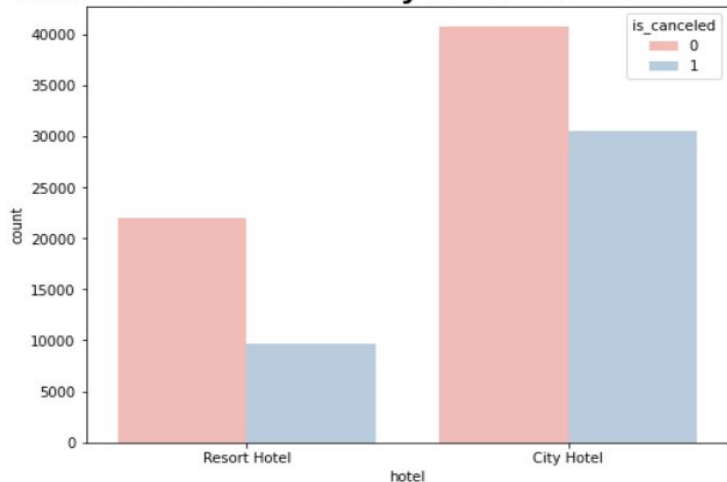
❑ Lead time has a positive correlation with cancellation

i.e. as lead time increases there is high chance of cancelling the booking.

❑ Bookings made a few days before the arrival date are rarely canceled, whereas bookings made over one year in advance are canceled very often.

❑ Factors effecting bookings

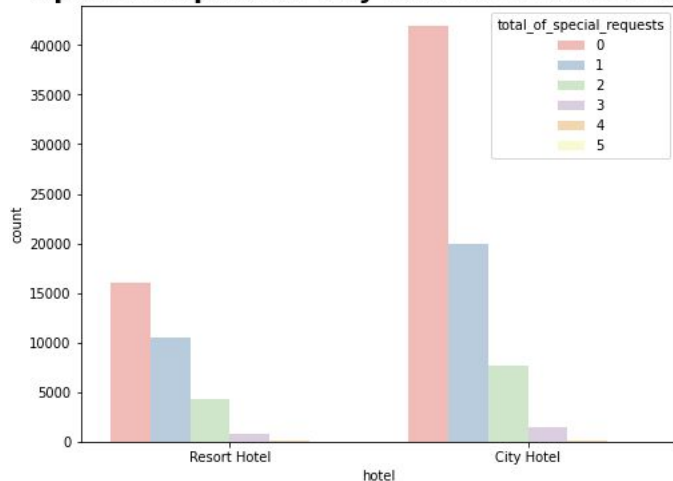
Cancellation rates in City hotel and Resort hotel



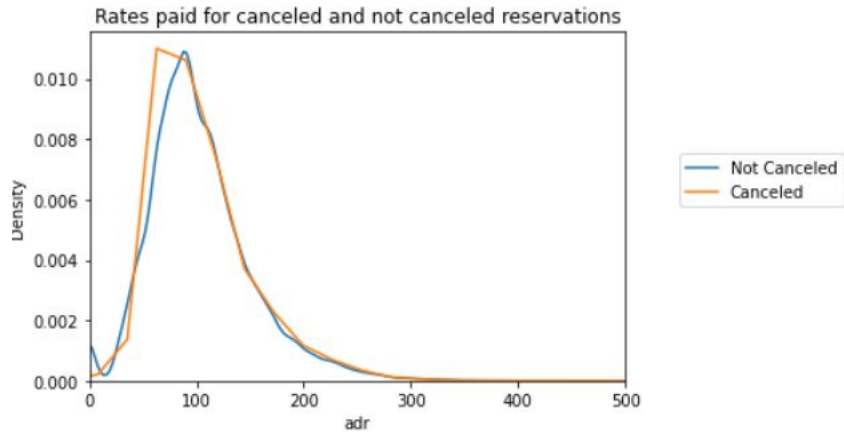
As we can see resort hotel total bookings were 31713 and city hotel total bookings were 71181. And the cancellations for resort hotel is 9684 and city hotel is 30477. i.e 30.53% and 42.81% of people are canceling from resort and city hotel. Instead of its high cancellation ratio city hotel have 54% more non canceled bookings.

As we can see that the city hotel have around 23000 special requests and resort hotel have around 10000.

Special request of City hotel and Resort hotel



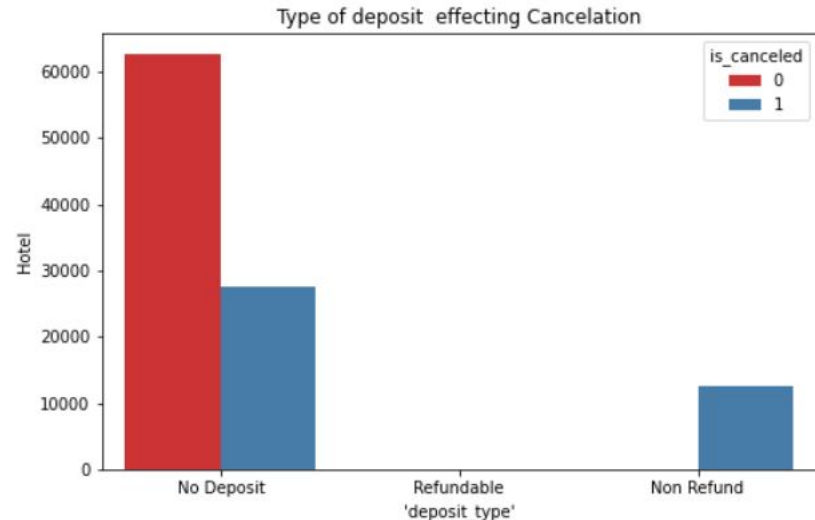
❑ Factors effecting bookings



It is interesting to note that non-refundable deposits had more cancellation than refundable deposits. Logically one would have assumed that refundable deposits have more cancellation as hotel rates are usually higher for refundable deposit type rooms and customers pay more in anticipation of cancellation.

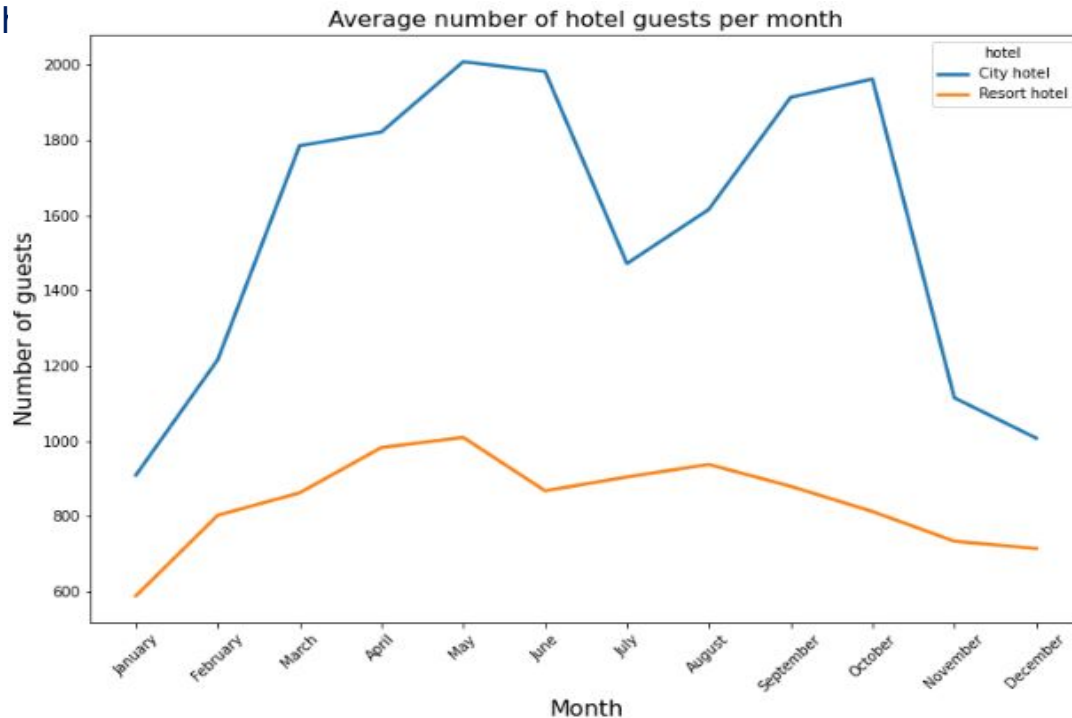
Average Daily Rate= Daily amount earned/Total booking on that day

The average daily rate is not effected due to cancelation as average daily rate is not effecting the number of bookings



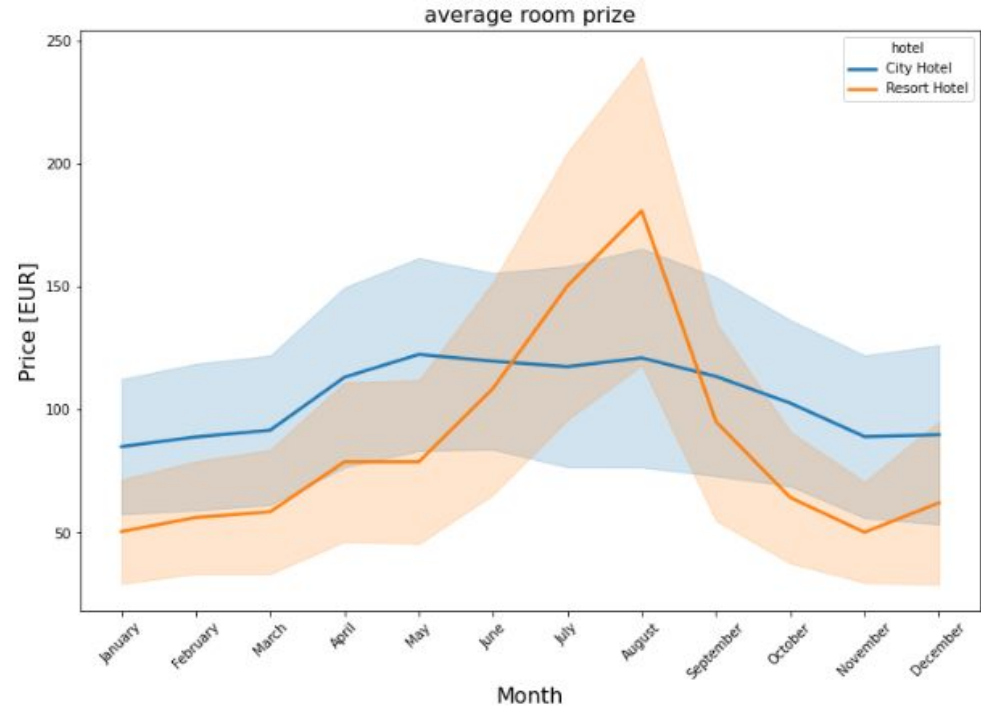
Booking trend of each month

- ✓ The city hotel and resort hotel has its booking in the month of May and October. The resort hotel have maximum bookings during March, April and May.
- ✓ City hotels have very few bookings during the winter (November and December) although at this time price were lowest.
- ✓ These graphs clearly depicts the number of hotel guests arriving in city hotel are way more than the guests arrived in resort hotel



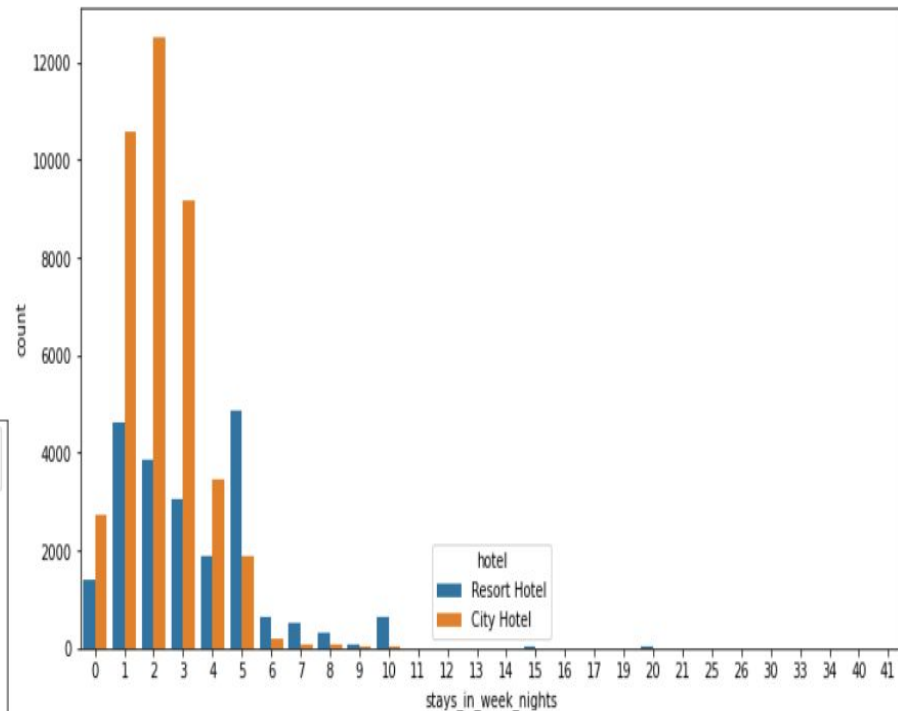
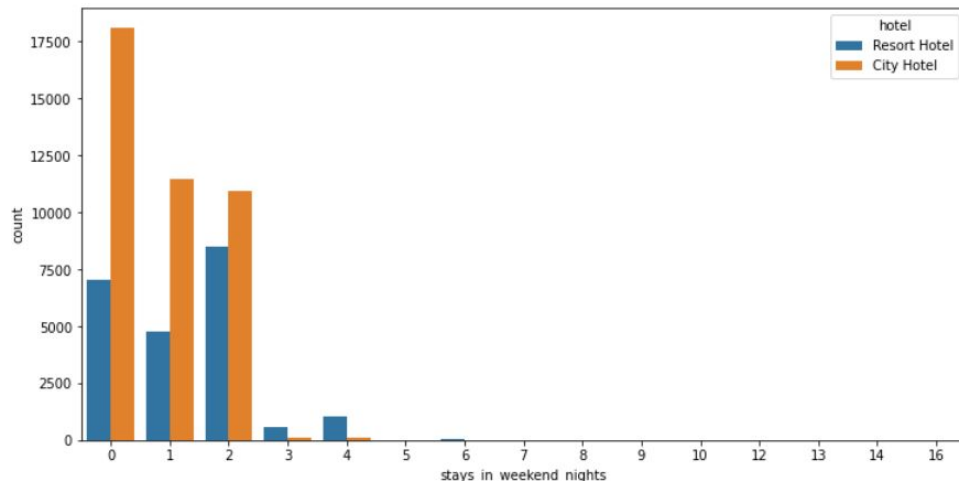
Monthly price Variation

- ✓ This plot clearly shows that prices in the Resort Hotel are much higher during June to August and with a steep fall after that.
- ✓ The prices of city hotel are more during January to April and September to December compared to 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.



Weekend stay vs Weekday stay

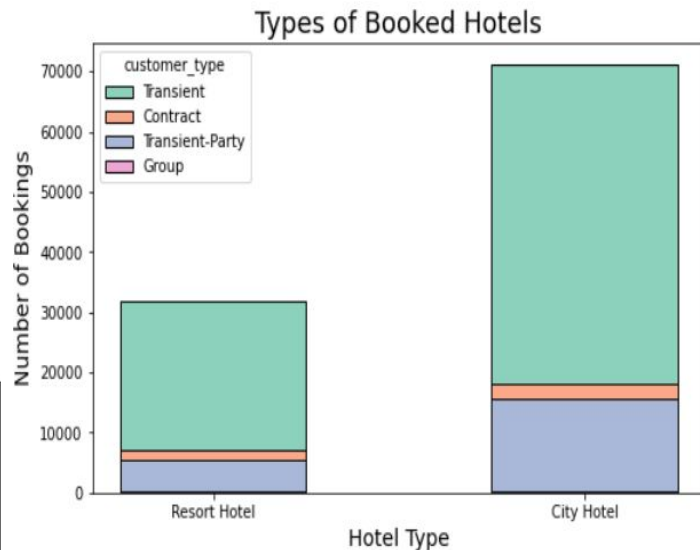
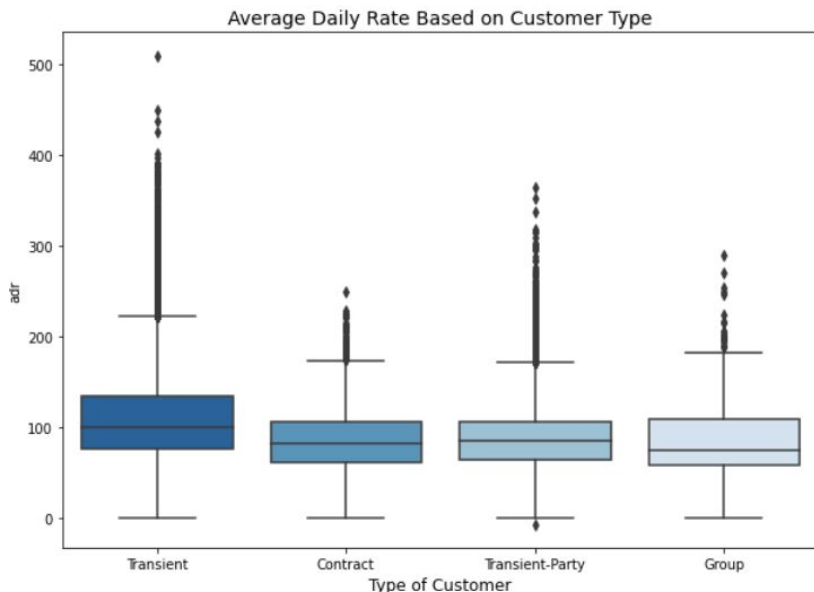
As we can see from the both the graphs majority of the people from city hotel tends to stay 0 week end nights followed by 1 and 2 weekend nights ,v
ery few people tends to stay longer than that, and
many people tend to stay 1 to 3 days on an avg.
In case of resort hotel people tends to stay at lea
st 2 weekends days,
and most people tend to stay atleast 4 days.



Graph representing Customer Preference

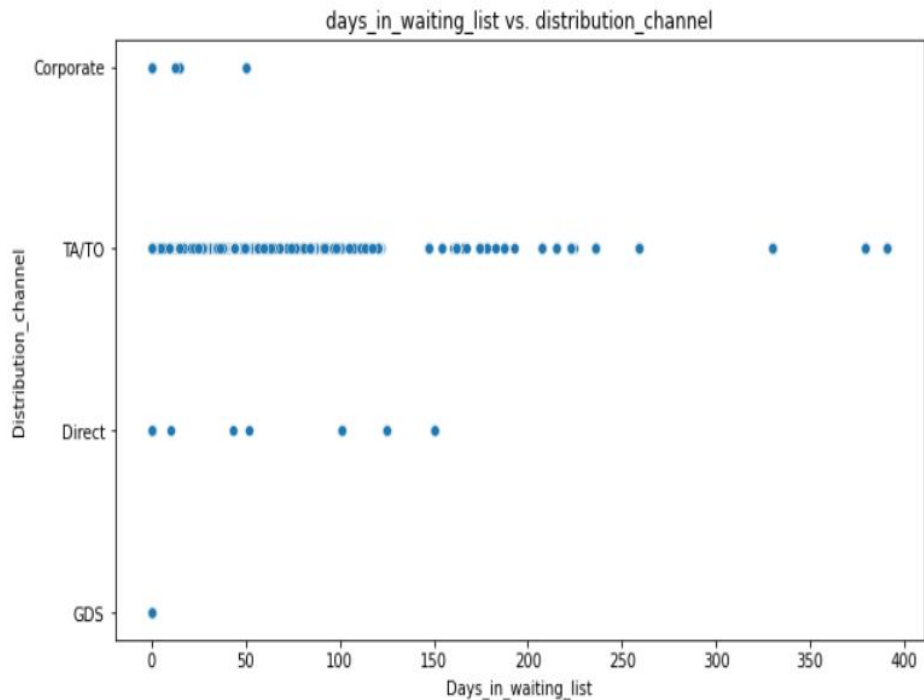
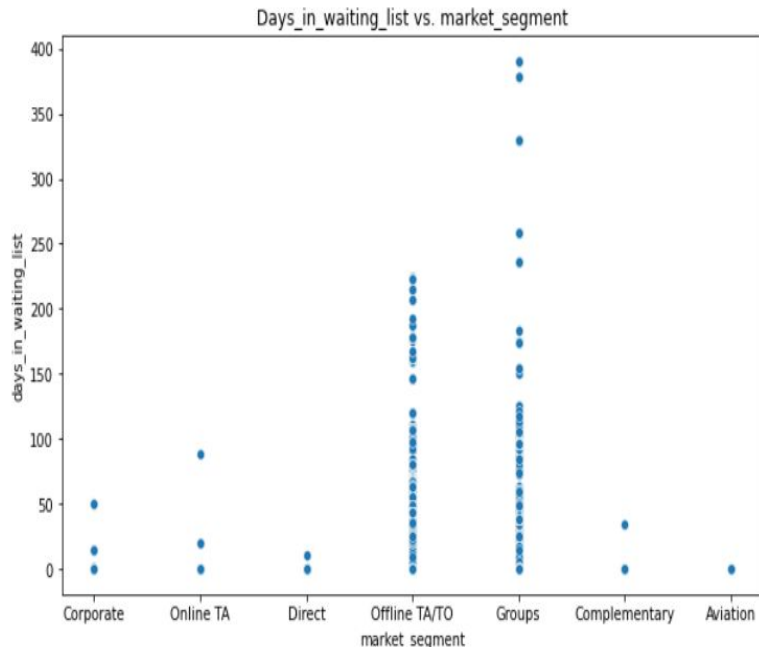
- The most common type of customer are Transient type followed by Transient-Party.
- And the less common type is Group type followed by Contract type.

As we can see that contract customer type has the least average daily rate(adr) followed by group and transient -Party and the highest average daily rate(adr) is for transient customer type.



Days in Waiting list

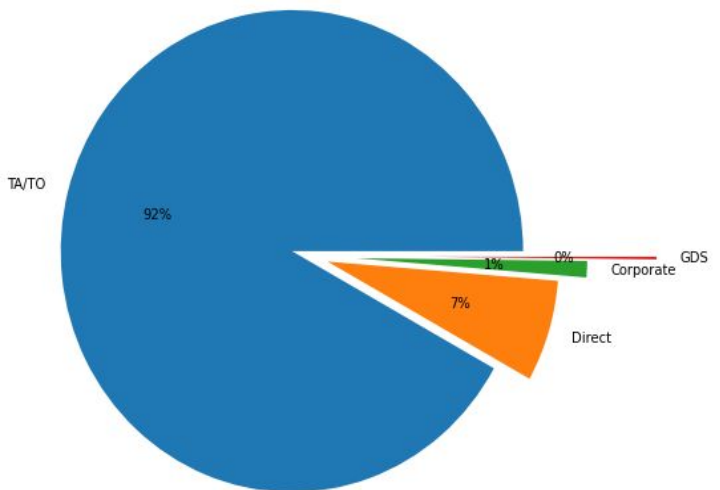
- TA/TO have the maximum booking in the waiting list
- And aviation has the least waiting list days, as there will be people who are pilots and cabin crew followed by remaining segments.



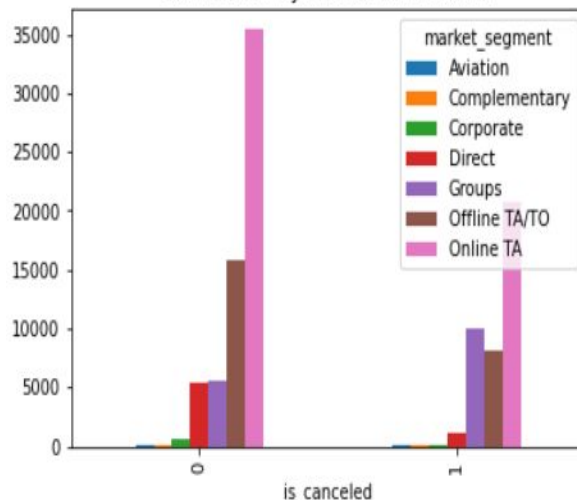
Bookings by Distribution Channel

TA/TO has highest booking of 92% and cancellation of 7% followed by direct distribution channel having highest booking and least cancellation. Direct bookings are having highest contribution to revenue of hotels.

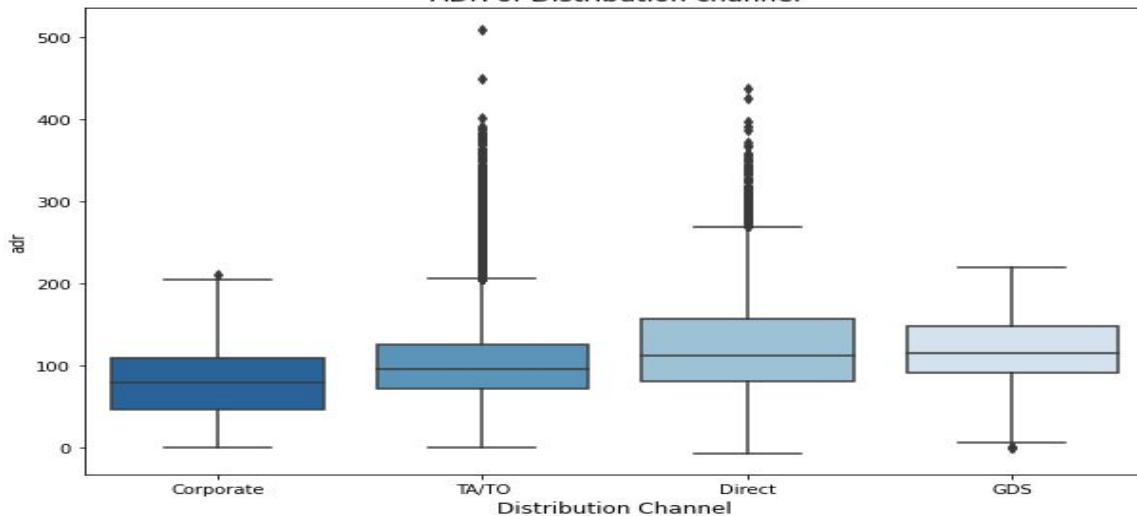
Bookings by distribution channel



Cancellation by distribution channel



ADR of Distribution channel

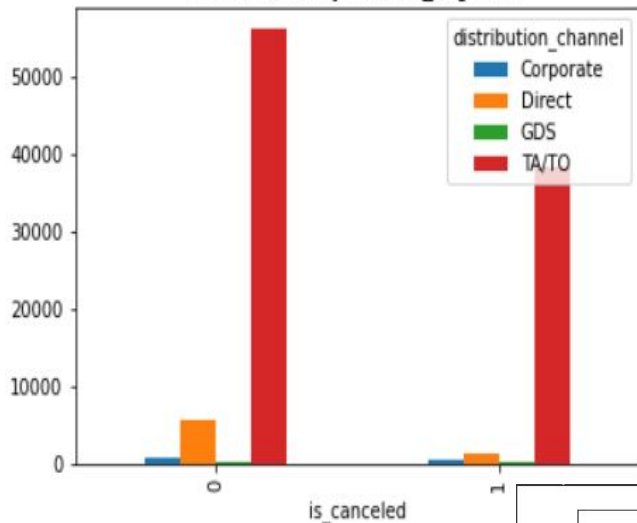




Bookings by Market Segment

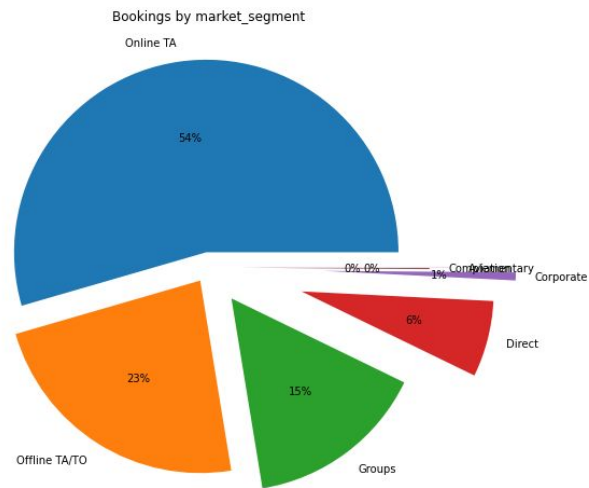
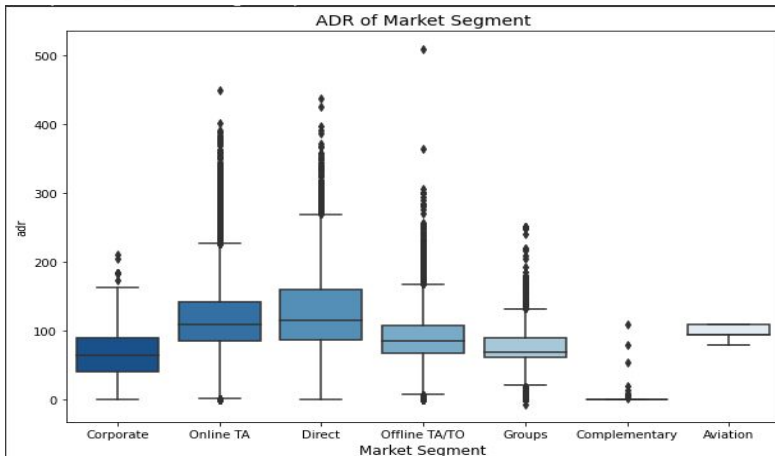
AI

Cancellation by market_segment



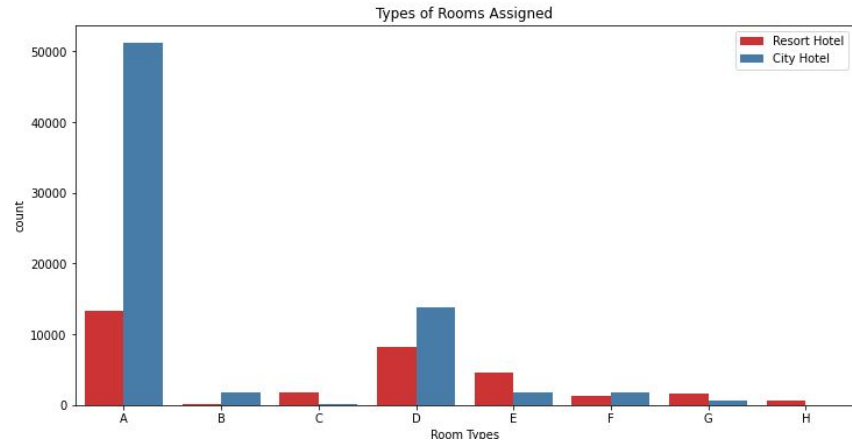
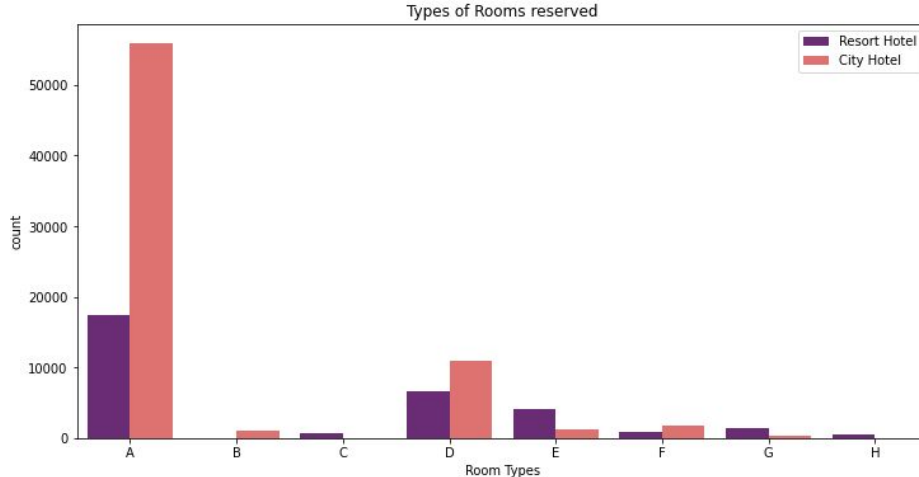
Offline TA/TO (Travel Agents/Tour Operators) and Online TA has booking rates of 23% and 54%, and cancellation rate of 33.8% and 36.9%.

It is surprising that the cancellation rate in these segments is high despite the application of a deposit.

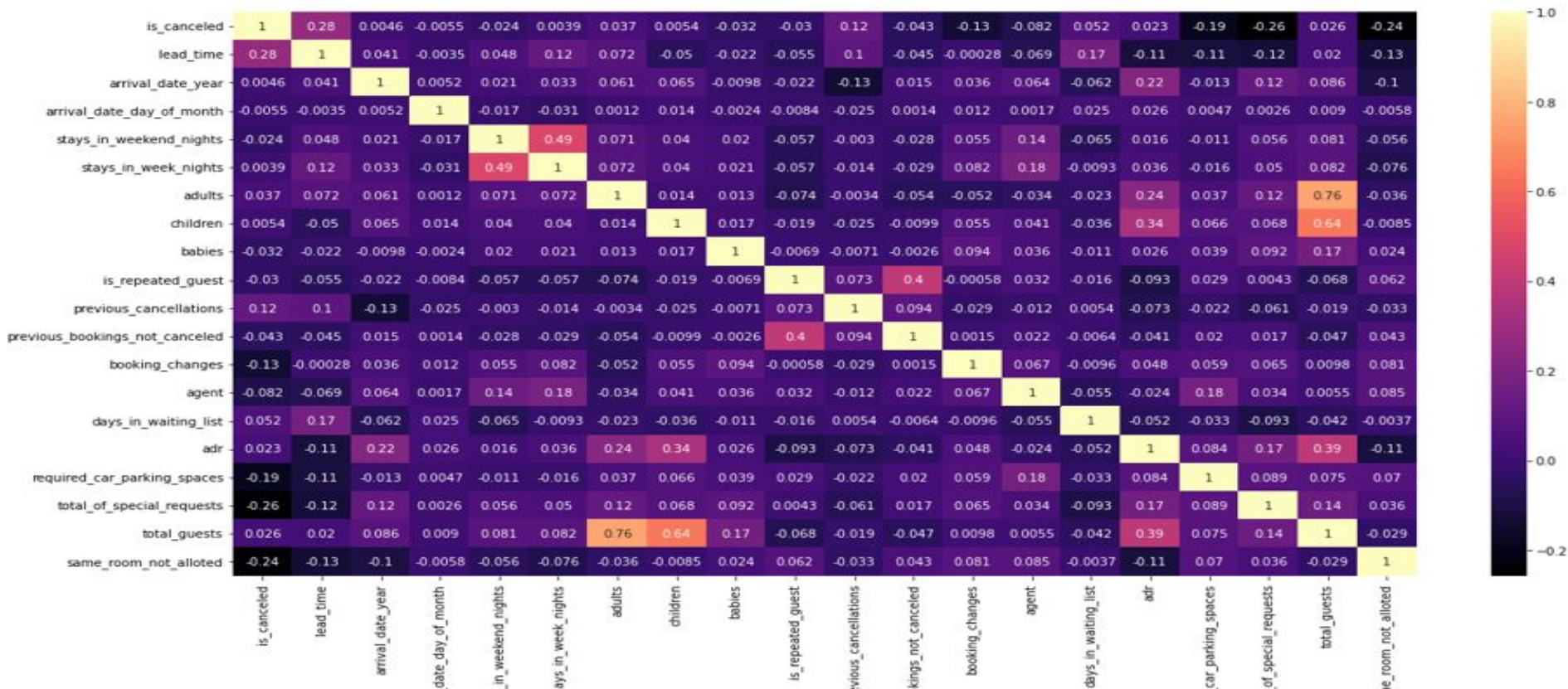


Highly assigned and reserved room

We see that the mostly A categories rooms are booked and of resort hotel and mostly allotted rooms are same as reserved so this indicates that the hotel gives more preference to the customers choices .



Factors influencing the booking



Factors influencing the booking:

- From the above heatmap we found that few column like same_room_not_allotted, no_of_special_requests,car_parking spaces,booking_changes are negatively corelated like these column will not effect the bookings.
- Whereas lead_time and previous_cancelation,stay in weekend is highly correlated,which means that this columns effect my bookings .
- If we dug deep we will see that adults and children are positively corelated with stay in weekend hence we can conclude that people staying in wekekdn are mostly family.



Conclusions

- The city hotel is the busiest as it has maximum booking. The avg adr is also high for city hotels making it more profitable.
- The City hotel has more guests during spring and autumn, when the prices are also highest. In July and August there are less visitors, although prices are lower. Thus, customers can get a good deal on bookings in July and August in city hotel.
- The number of guests for Resort Hotel go down slightly from June to September, which is also when the prices are highest. Thus, these months should be avoided for bookings. Broadly, May is the peak season for city hotel for bookings whereas for resort hotels March, April and May have the highest bookings.
- Higher lead time has higher chance of cancellation. Also, history of previous cancellations increases chances of cancellation.
- The prices in the resort hotel are higher during the summer and prices of city hotel varies but is most expensive during spring and autumn. Both hotels have few bookings in Winter.
- Transient customers, direct customers are contributing maximum revenue to the hotel business.
- No deposit cancellations are high compared to other categories, but these should not be discouraged as bookings in this category are also very high as compared to nonrefundable type bookings.
- Cancellations are high when done through agents compared to direct bookings. Hotels need to do marketing and give special incentives for direct bookings as these may establish personal one to one relationships promoting customer loyalty.

THANK YOU