Capstone project

Hotel Booking Analysis

Table of Contents:

Problem Statements:

- **Analysis based on our data on hotels**
- **Analysis based on our data on customers**
- **Analysis based on special requests and parking spaces**
- **Analysis on cancellations**

Summarizing our data:

Name of our dataset:

Hotel Analysis Dataset consisting information on hotel bookings, customer info as well as data on cancellations for the years 2015, 2016 and 2017.

Shape of our dataframe:

- **Rows: 119390**
- **Columns: 32**

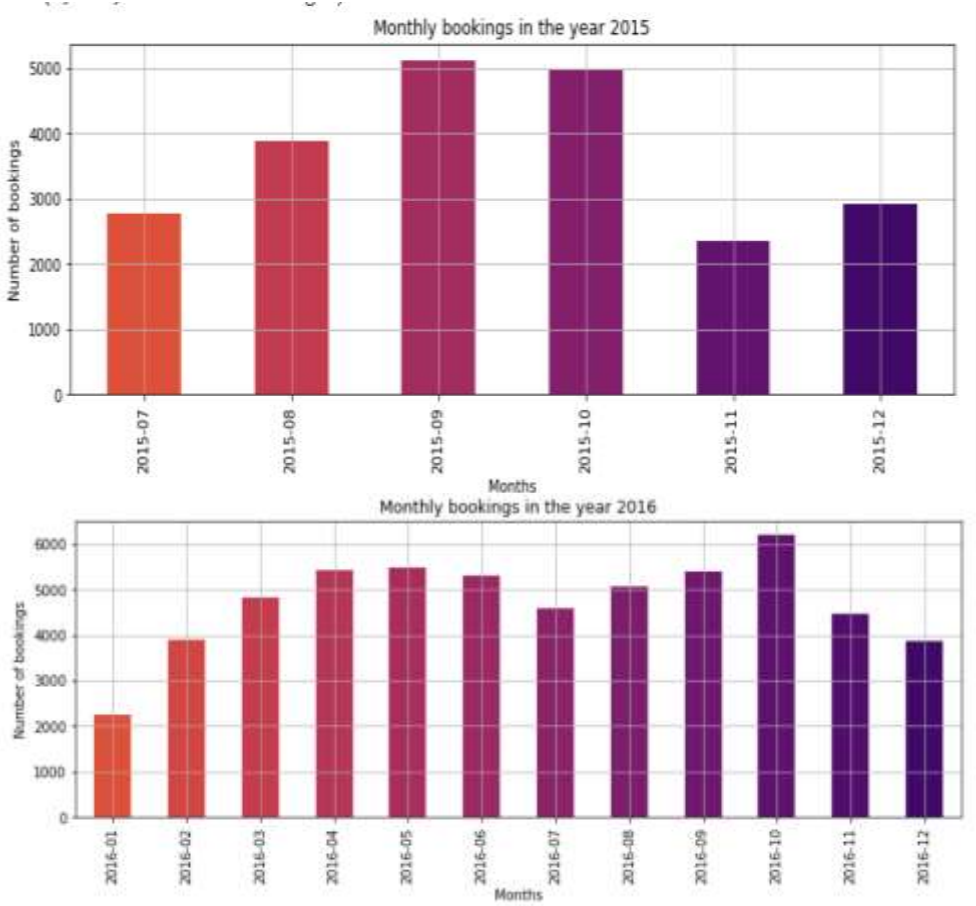
Most important columns we have worked on:

Hotel, is_canceled, arrival date, adr, repeated guest, previous cancellations

Columns we have not worked on:

Lead time, babies, meal, agent, company

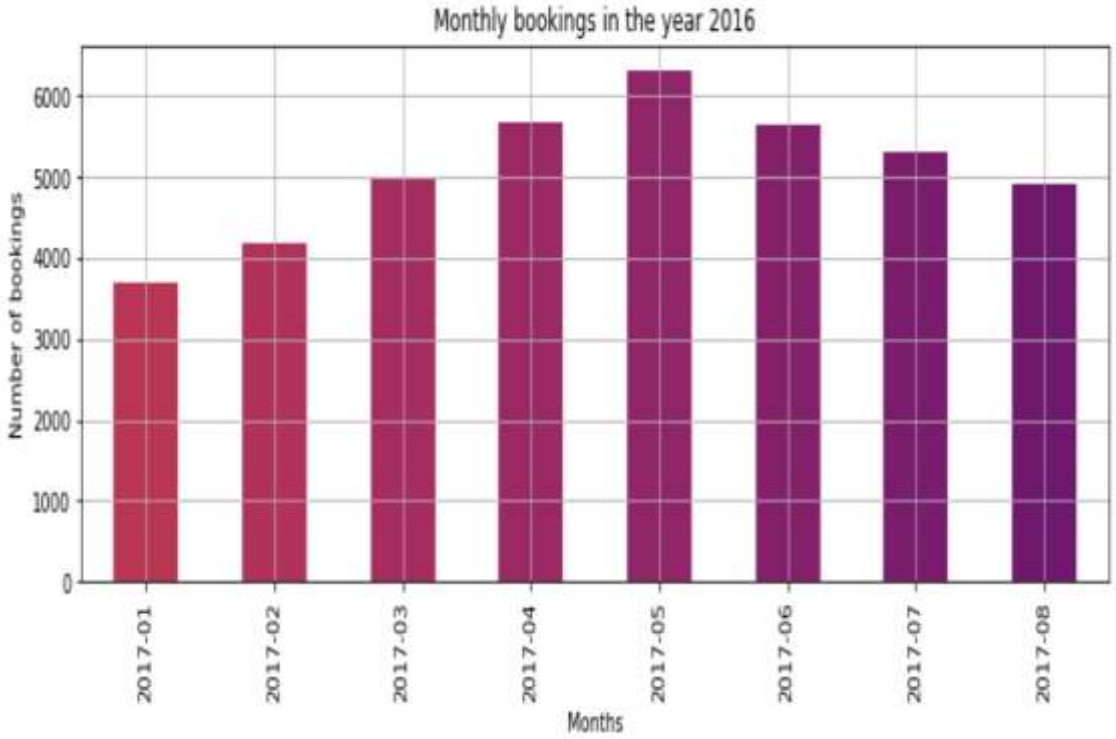
Analysis on yearly hotel booking data:



Bar plots on Number of bookings vs Months in years 2015 and 2016

We have plotted a bar graph between the number of bookings vs the month name, to compare the bookings between each and every month in all the three years. We noticed that the maximum bookings in the year 2015 was in the month of September with more than 5000 bookings and the least being in July. However, in the year 2016, we saw a boost in the number of bookings with an average booking of almost 4500 in the entire year, with the highest being in October and the least in January.

Analysis on yearly hotel booking data:(cont.)



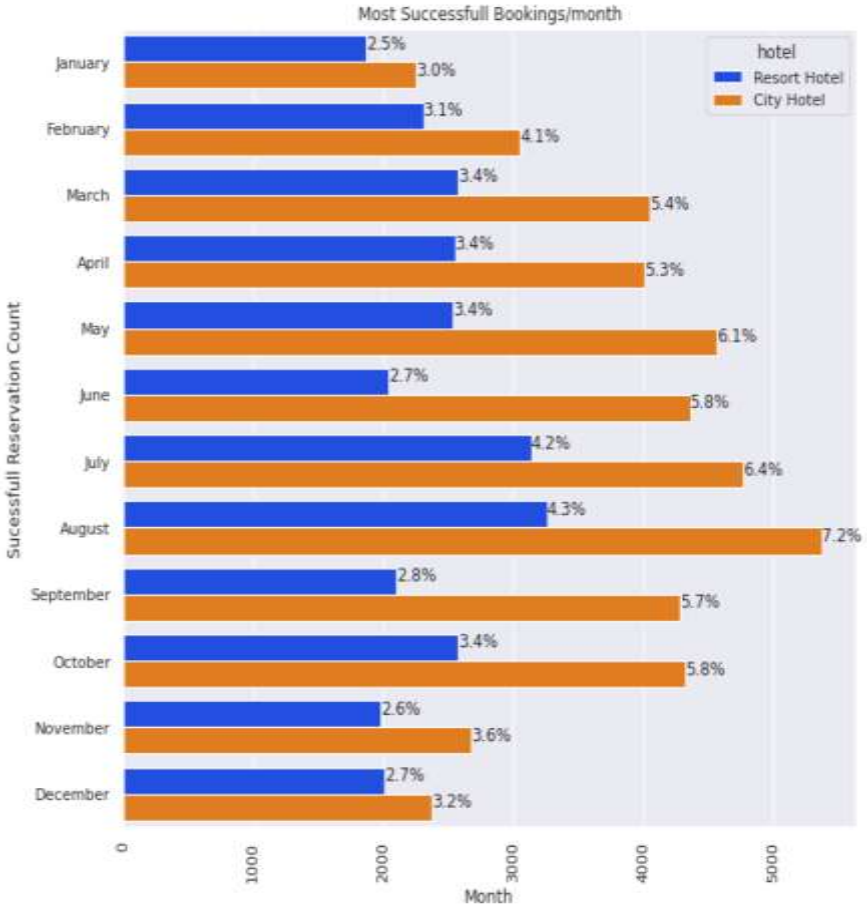
Bar plot on number of bookings vs the months in the year 2017

This bar plot shows us the bookings data in the months of 2017. The highest bookings being in the month of May with over 6000 bookings and the least in the month of January.

Conclusion:

The highest number of bookings in the last two years are in the months from May to October in 2016 and from May to August in the next year. However, the least bookings in both these years are in the starting months of the year.

Analysis on successful bookings in each hotel(Hotel City and Hotel Resort)



Success rate comparison between Hotel City and Resort

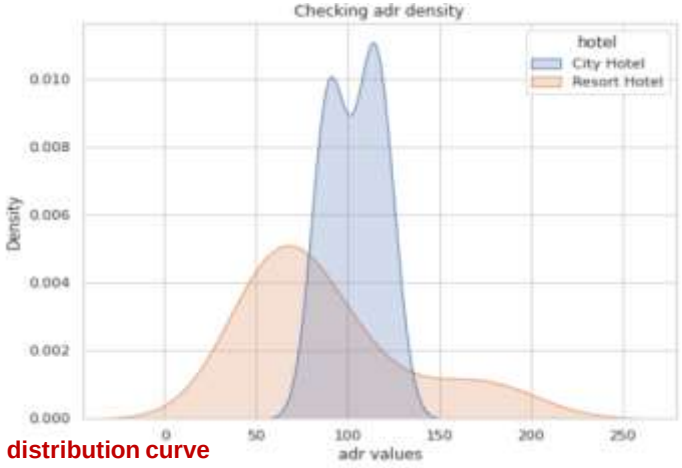
We have used a bar plot to compare the successful booking percentage in both the hotel and compared their data for the same months.

Conclusion:

City Hotel followed by Resort Hotel had the most successful bookings in July and August in all the three years. In July , City hotel had an increase in bookings to 6.4%, while Resort Hotel had an increase of bookings to 4.2% in the same month. During August, City hotel had an increase in bookings to 7.2% and Resort hotel had an increase of 4.3%.

Analyzing ADR values:

Plotting adr values for each month



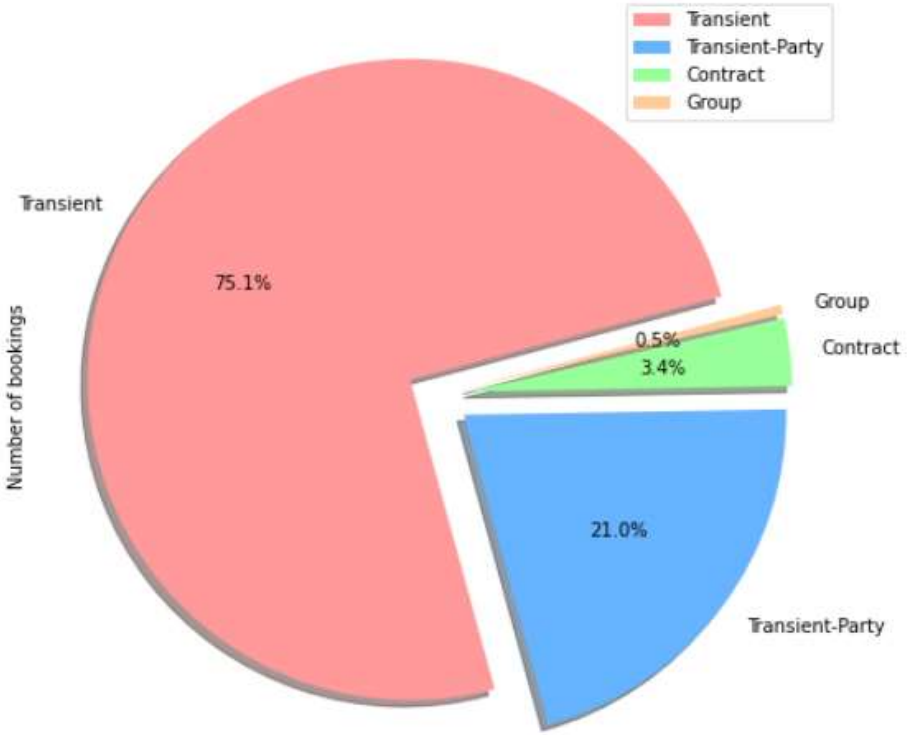
Adr distribution curve

Here, we have plotted a line curve for adr values versus all the months, and have also plotted a histogram curve to look at the adr density.

The conclusion we can draw from the line curve is this that the adr value seems almost constant over an entire year for Hotel City whereas there are some steep slopes in Hotel Resort from the month May to August.

From the histogram to our left, we can see the adr density for Hotel City is almost the same throughout the year with an average of about 110, but that of hotel Resort Hotel varies throughout the year with values as low as 0 to as high as 220.

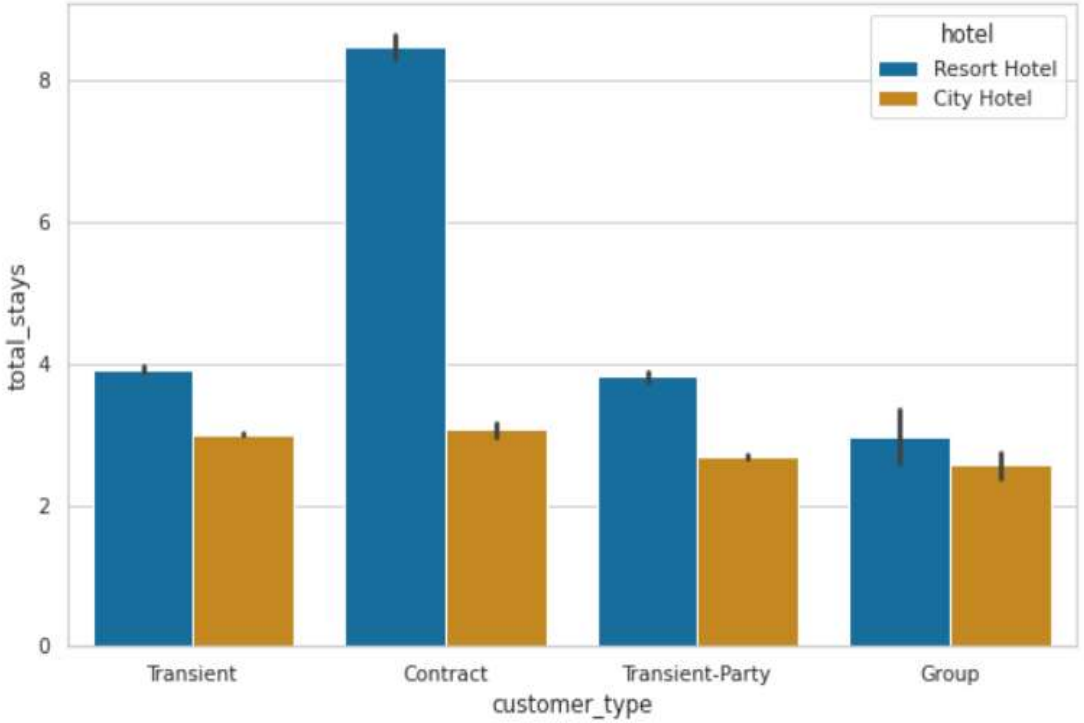
Analyzing customer's data:



We have analyzed the different types of customers who have booked these hotels in the years 2015-2017.

As the pie chart depicts, more than 75% of the customers who have booked rooms are in the Transient category and the other 25% belong to 'Transient Party's. People from the other two groups 'Contract' and 'Group' are almost negligible.

Let's find out the average number of days a customer from each customer type prefers to stay in Hotel City and Hotel Resort:



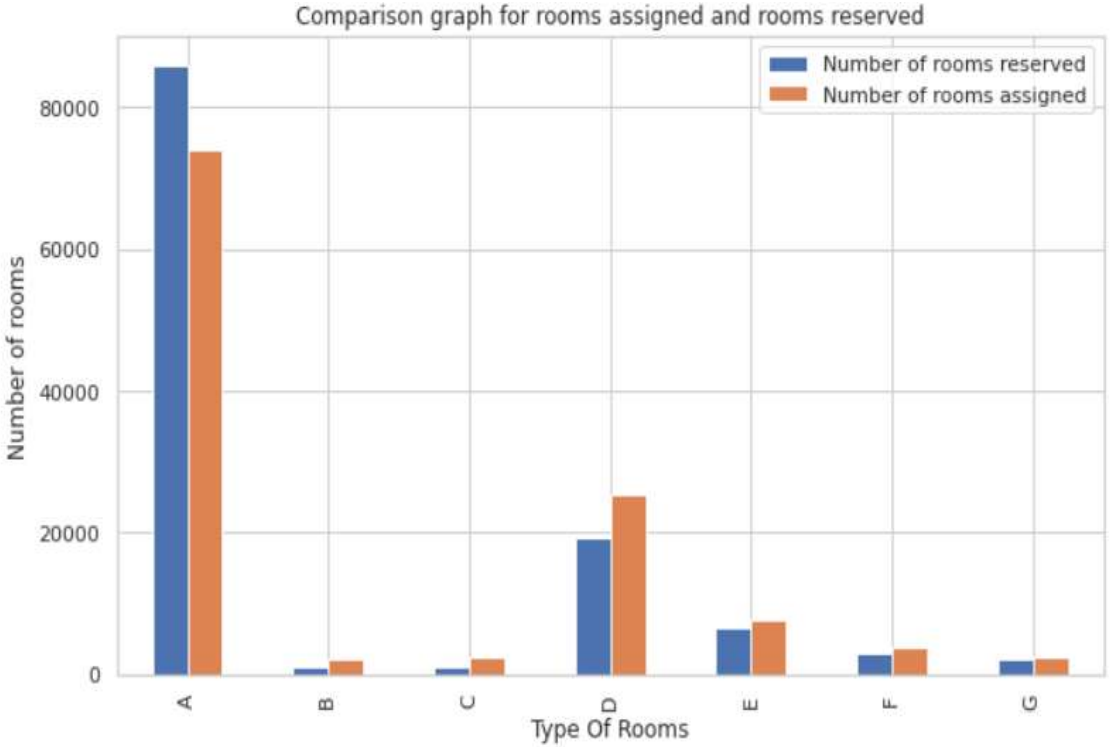
Plotting between customer types vs average number of stays

We have used a bar plot have compared the data between customer types and average number of stays in both the hotels.

Conclusion:

Guests love to stay at Resort hotel and most of them belong to Contract customer types(can be corporate contracts/any contracts made before reservation bet ween the hotel and the customer), stay ed for more than 8-nights on an average followed by Tran sient customer type with around 4-days and they keep relocating from on e place to another.

Analyzing room type customer prefers the most:



Plotting relation between Room types and Number of rooms booked and number of rooms assigned to customers

We have plotted to bar graph to get the details of room reserved by customers and assigned to them by hotel authorities.

We can see that room A has been booked the most number of times (almost 85000)times but not all of them have been assigned to the customers. However, in the other rooms, the scenario is quite opposite, assigned rooms are greater than reserved bookings.

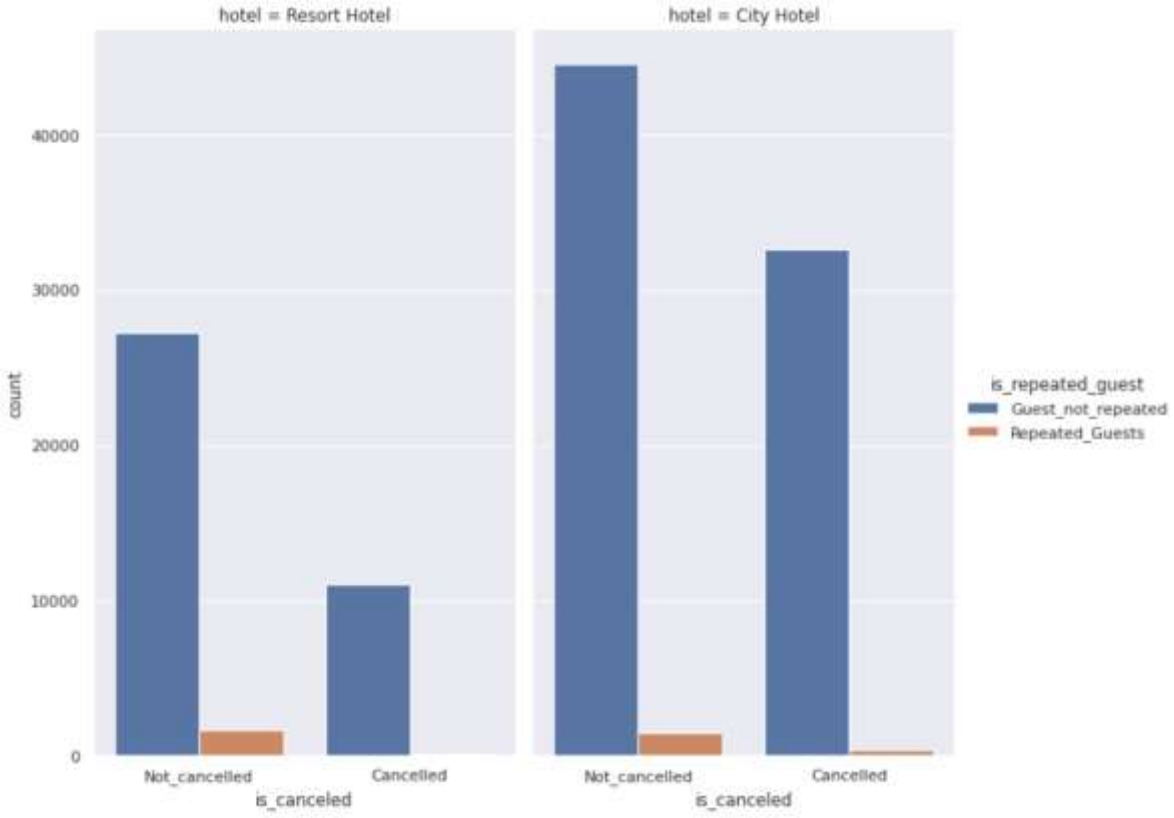
Analyzing on whether a booking gets cancelled or not based on deposit types of a hotel:

Relation between Deposit types vs cancellation

We have used two bar plots for each hotel to have a better comparison on deposit type vs cancellations.

In Resort Hotel, considering successful bookings, 70% of the time Resort hotel, successful bookings had refundable policy on deposits and 30% of the time after cancellation, the booking amount was refunded to customers.
In City Hotel, considering successful bookings, 60% of the time the hotels had non-refundable policy on deposit with having very less cancellations. However, 50% of the time, hotels had refundable policy on deposits and only 10% of the bookings got cancelled, and the amount was refunded back to the customers

Let's analyze on a booking getting cancelled or not based on the the nature of similar/different guests



Relation between guests and cancellation

We used a bar plot to give a relation between cancelation and guest nature

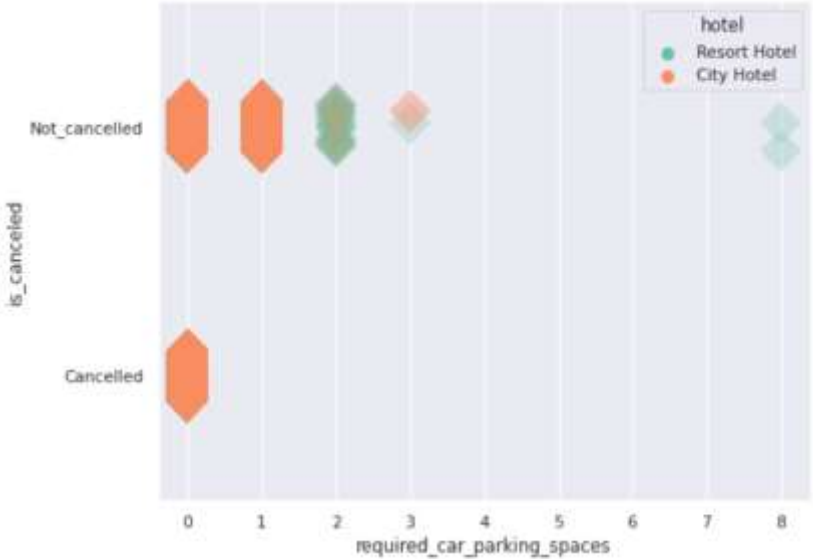
Conclusion:

Around 94% of the bookings which didn't have previous cancellations were booked by same guests in City Hotel and 6%of the bookings without previous cancellations had no show.

While in Resort hotel 65% of the previous bookings which didn't have previous cancellations were booked by same guests. 35% of the bookings without any previous cancellations had no repeated guests who booked again.

Analyzing on parking spaces:

Relation between parking spaces and cancellation

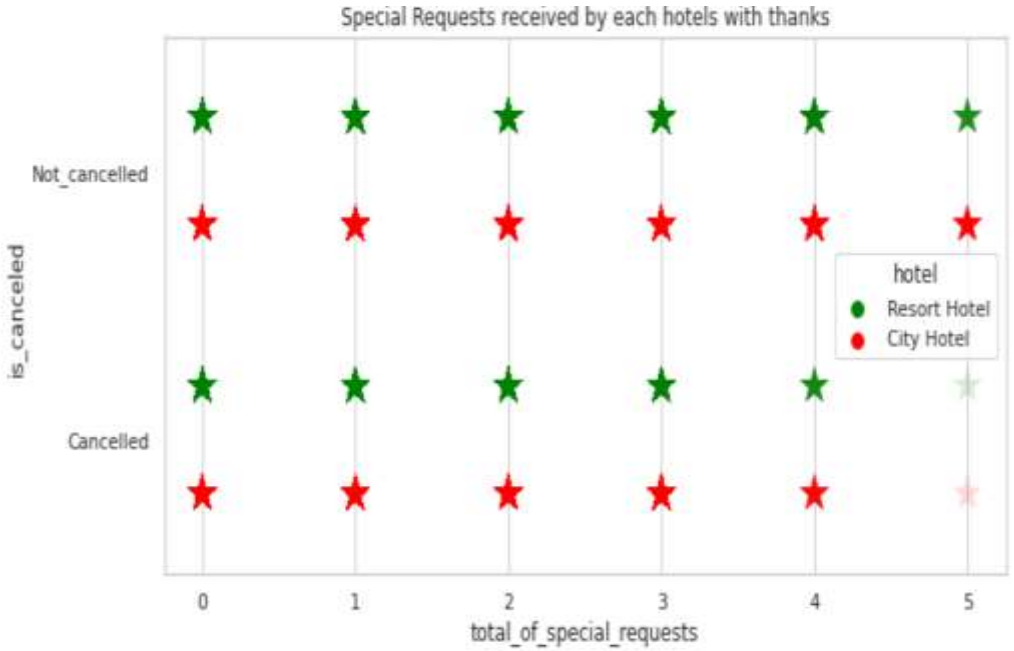


Minimum number of parking spaces for customers in each hotel

Conclusion:

Guests who didn't have any cancellations on bookings, opted for parking spaces.
Few of them opted for three parking spaces
from City Hotel while at max 8 parking spaces were opted by a few guests at Resort Hotel

Analyzing on special requests:

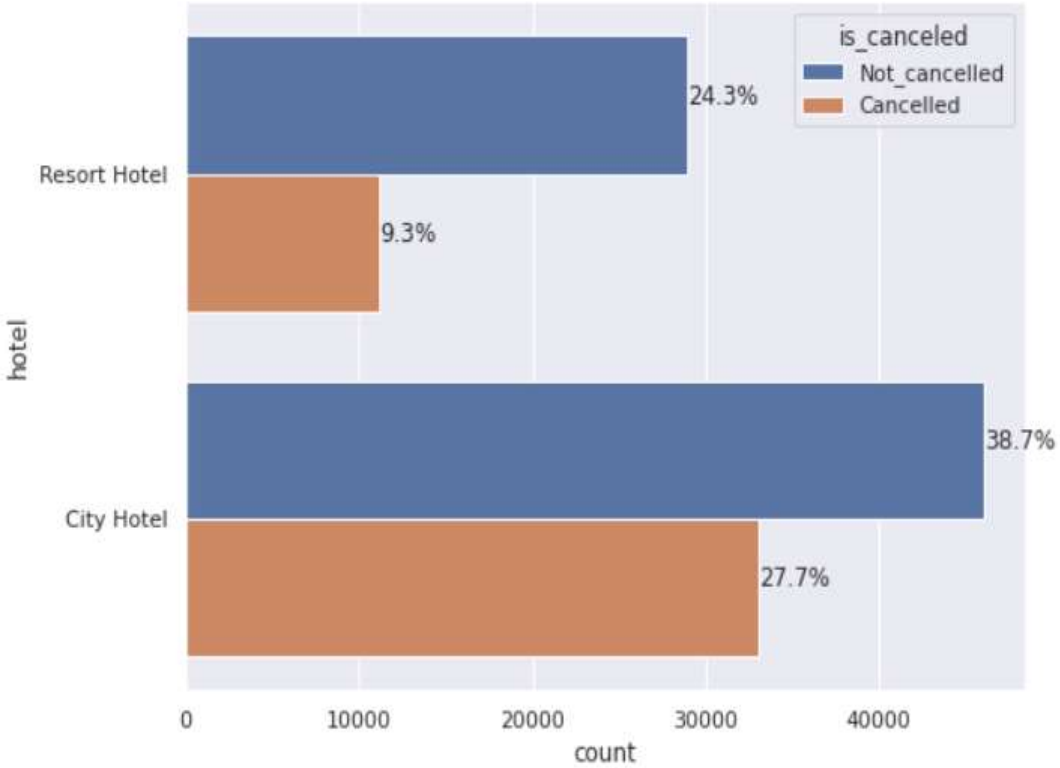


Relation between the number of special requests and cancellation

We have used a strip-plot to have a better visualization and to have a better understanding of the relation between the number of special requests and cancellations

The conclusion that can be drawn from the above plot is this that both the hotels get special requests upto 5 where the bookings did not get cancelled. While for cancelled bookings, hotels receive special requests upto 4.

Analysis on cancellation on both the hotels:



Relation showing the number of cancellations in each hotel over the three years

We have used two horizontal bar plots, one for each hotel and compared their cancellation percentages.

38.7% of the total bookings in City hotel were without cancellations, whereas 24.3% of the bookings in Resort were not cancelled. However, the cancellation percent of City and Resort hotels are 27.7% and 9.35 respectively.

Challenges we faced:

- **Difficult to recognize some of the columns and how to operate on those columns, such as babies, adults, meals, agent.**

- **Ample time to load due to large amounts of data**

Conclusion:

- After analyzing every single relevant variable in the dataset, we can observe that although bookings numbers are much more in Hotel City, people love to spend their time more in Hotel Resort than in Hotel City.

- Even if the booking numbers are higher in Hotel City, the overall cancellation percentage in City Hotel is also much higher than that in Hotel Resort.

- Percentage of people coming to hotels increased from 2015 to 2016, whereas it decreased slightly in the next year.

- People who had successful reservations in any of the hotels booked the hotel again in their next visit and cancellation percentages of repeated customers were much less than normal customers.

- The ADR values of Hotel City are mostly constant over a year which denote that both the hotels earned higher profits as in price per night in months July & August. Also, both the hotels had lowest price/night(adr) during starting months(Jan-March) every year. Resort hotel makes good profit during year-end as well(November & December)s.