HOTEL BOOKING ANALYSIS

Team Data Chronicle
First EDA

Team Members:

1. Kaushal Kumar Jha
2. Prabhat Pradip Manna
3. Shambhu Nath Jha
4. Asif PA
5. Priya Gupta

POINTS OF DISCUSSION

- Data Summary
- Busiest year among all three ye
- Maximum booking by month
- Most Preferred hotel
- Mostly reserved room
- Special request
- From which country most of the booking was made
- Percentage of Cancellation in Resort and City hotel
- What type of meal guest prefer most
- How many guests are repeated
- Deposit Analysis

POINTS OF DISCUSSION CONTINUE....

- Which hotel is generating more revenue
- Effect of required car parking space on average daily rates
- Which month having highest cancellation record
- Which type of guest made more cancellation
- Tourist preference in week nights v/s weekend nights
- monthly booking data in different years
- How lead time affects the cancellation with different type of customers
- Market segment analysis with waiting list
- Co-relation of the column

DATA SUMMARY

df : This data frame contains all the information related to hotel booking analysis given in a csv file and made it readable for analysis part.

The shape of our original data set was:(119390, 32)

1)Data Cleaning:

Data cleaning is required to get a correct analysis from a given dataset. In our data frame so many null values and duplicate values was present so we have done applied data cleaning process on our data set.

Before commencement of cleaning on our data set we have created a copy of original and name it as df1.

.

SUMMARY CONTINUE..

df1:On this data frame we have done cleaning and first of all dropped the duplicate value from this after that we have filled the null values with appropriate value. Also we have dropped the columns (company and agent) which contains maximum null values.

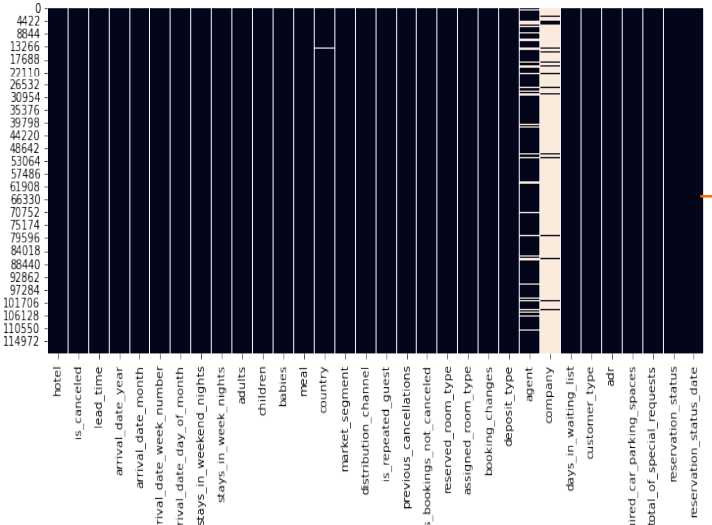
After cleaning the shape of our dataset was:(87396,30)

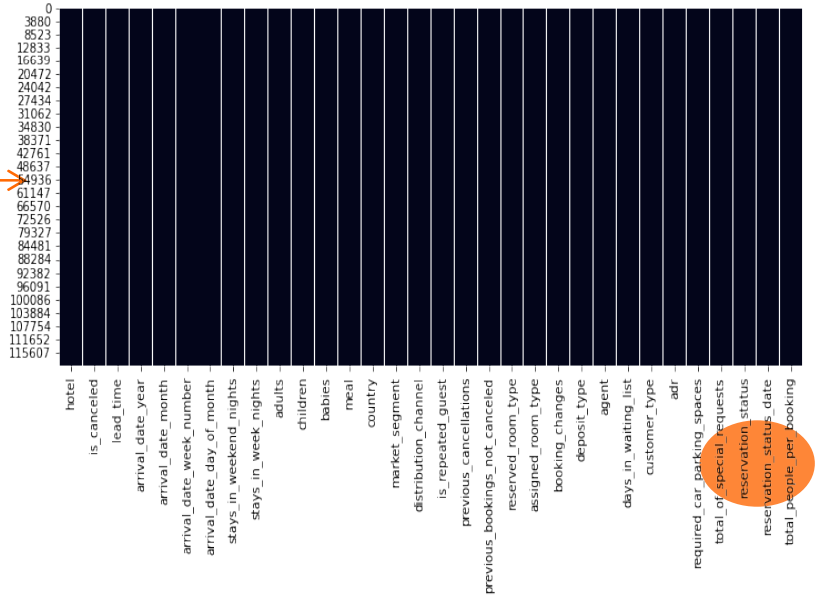
We have added one column total_people_per_booking by adding all adults, children and babies. And than we have check the total rows which contain total_people_per_booking equals to zero which means an invalid or booking was not made so we got a dataset for this condition with shape of 166 rows × 31 columns.

Df
(119390, 32)

Data Cleaning

Df1
(87230, 31)

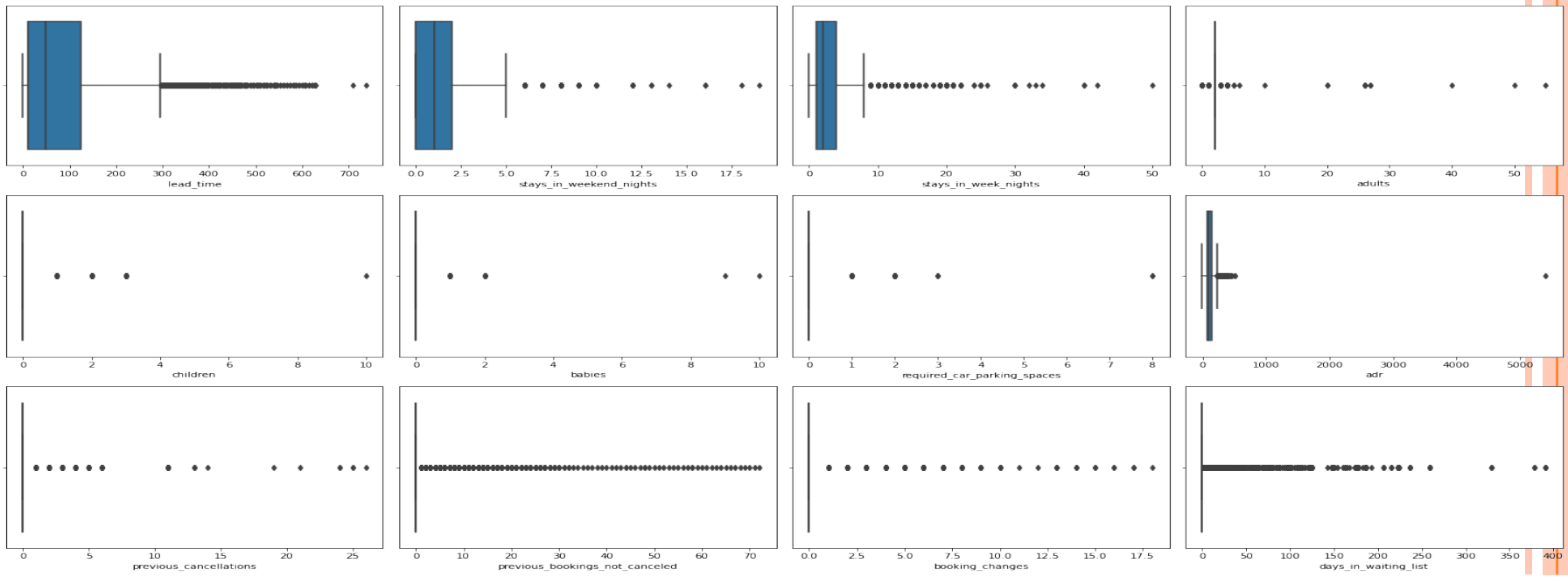




SUMMARY CONTINUE…

After finding all these data we have dropped them and finally we get a dataset with the shape of (87230, 31) .

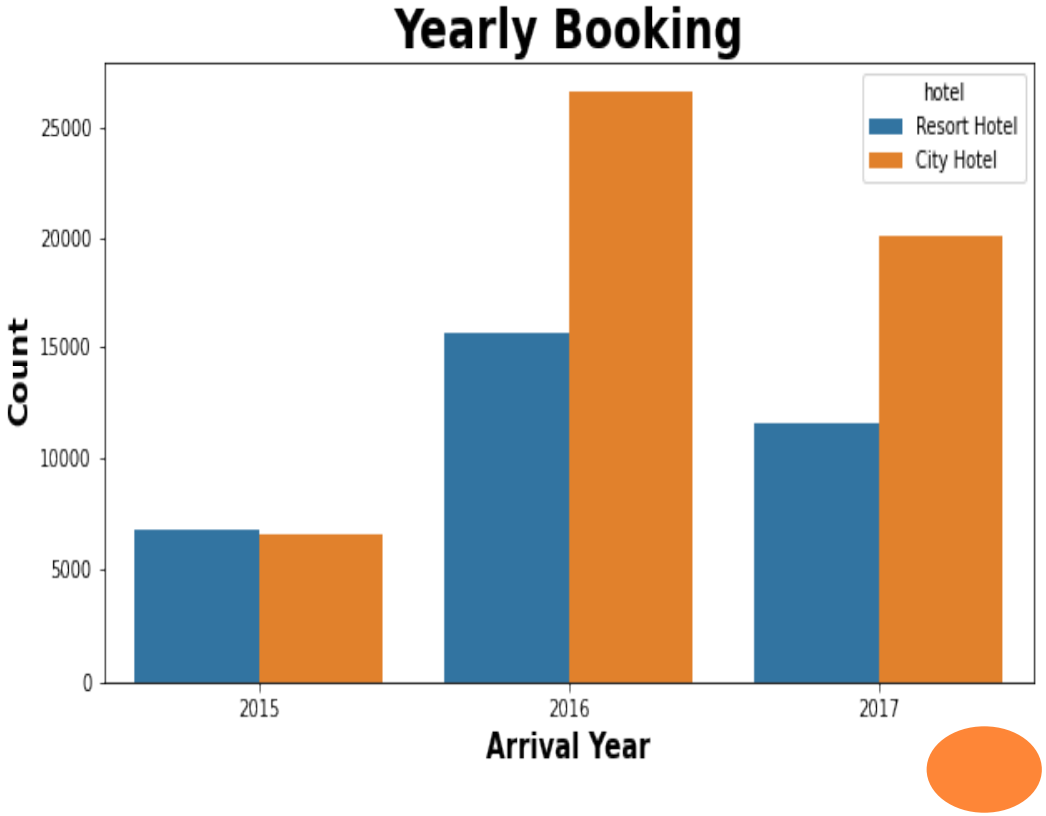
df1.describe:By using this we came to now that in the columns which contains numeric data having some outlier values so we have removed all the outliers data and updated those columns with new values.



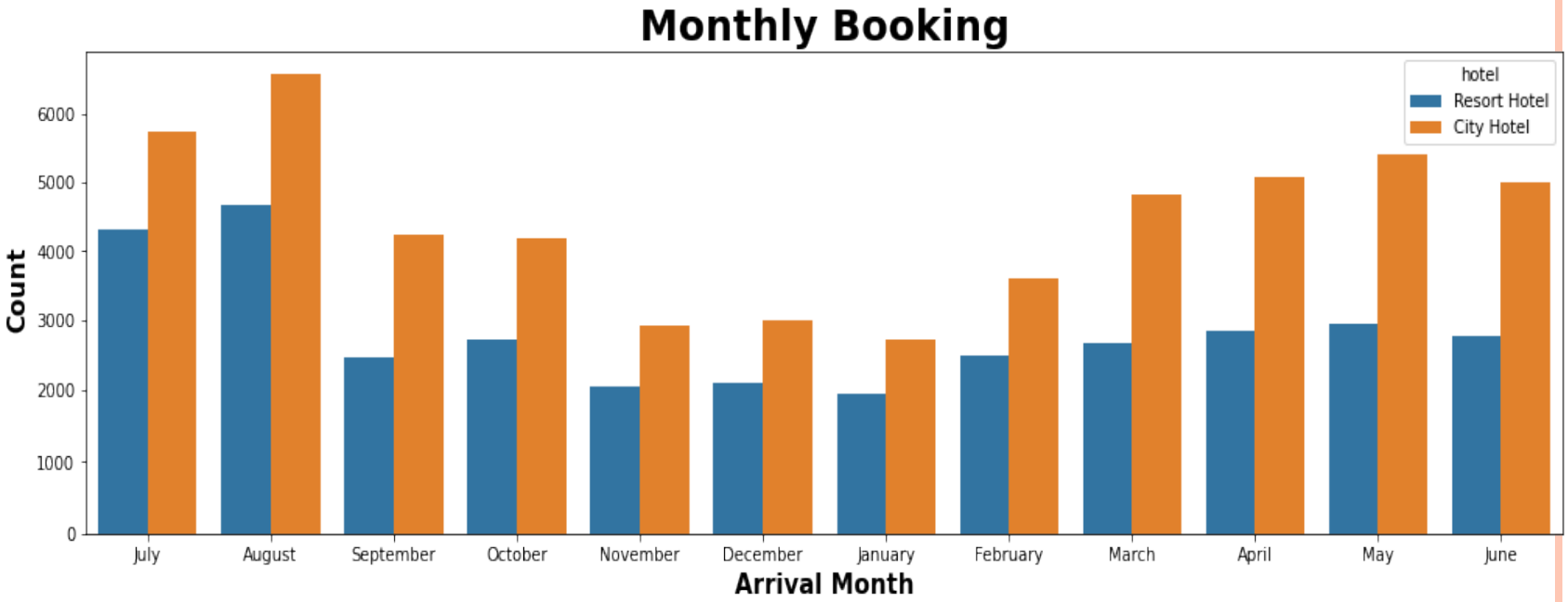
Let's Start EDA:

1.Busiest year among all three years.

From this count plot it is understood that busiest year for hotel booking was 2016 for both hotels type.



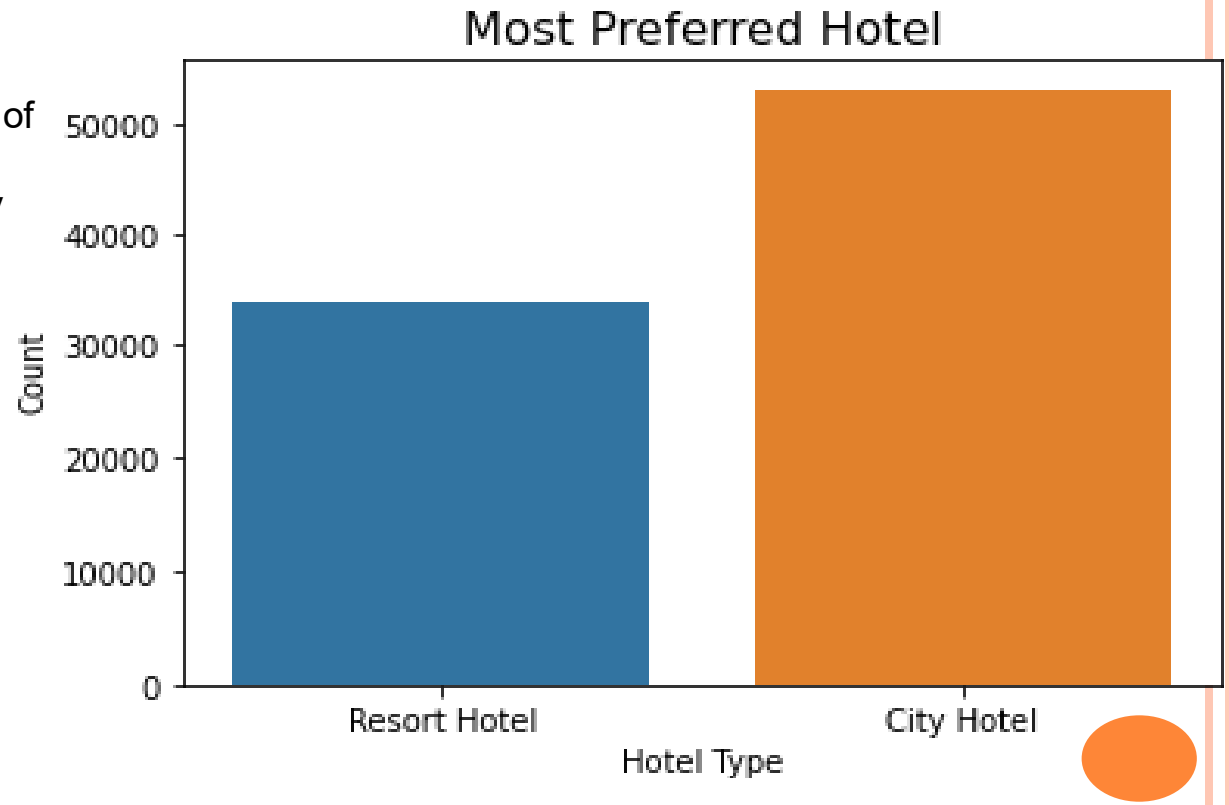
2.Maximum booking monthly wise



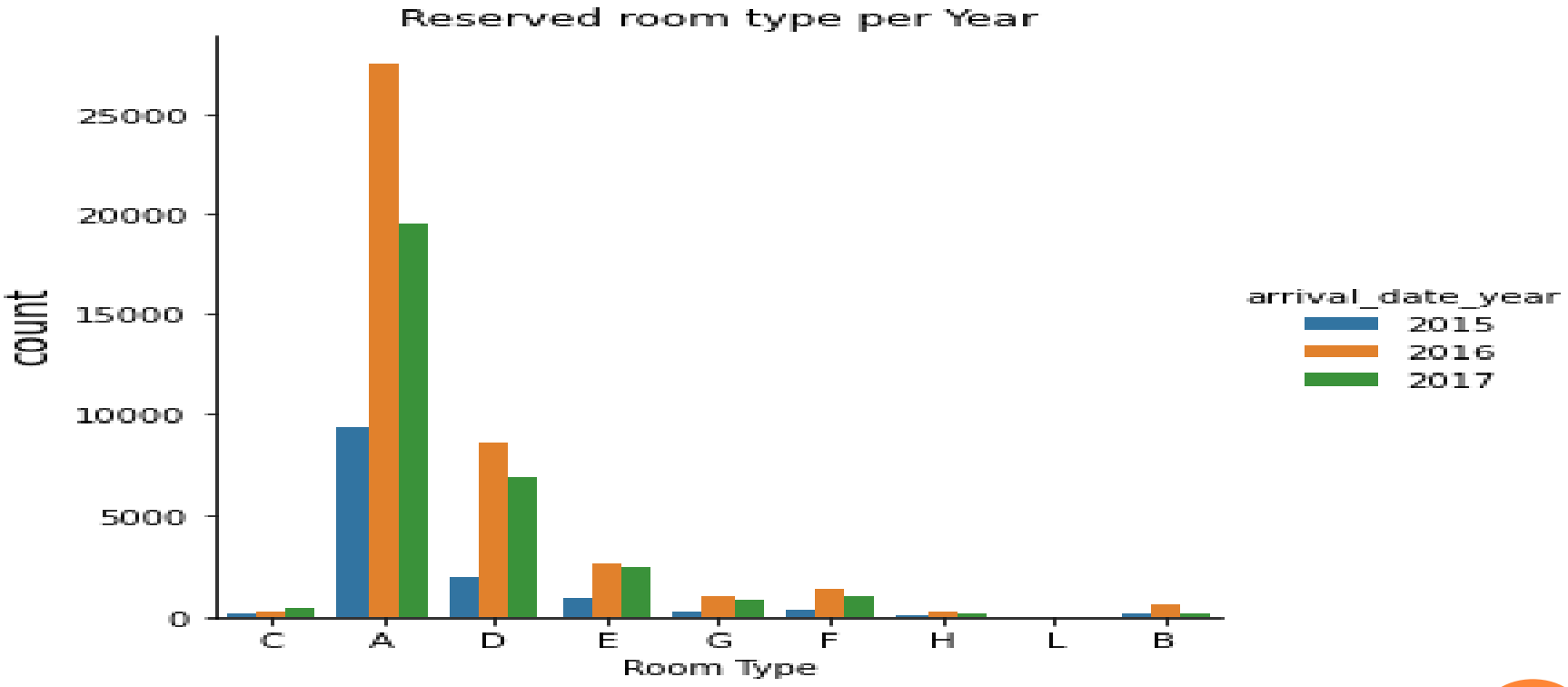
From above graph it is understood that most of the booking was done in the month of August for both type of hotels and lowest booking is in the month of January

3. Most Preferred hotel

From this count plot of both types of hotels we can understand that City hotel was mostly preferred by the guest.

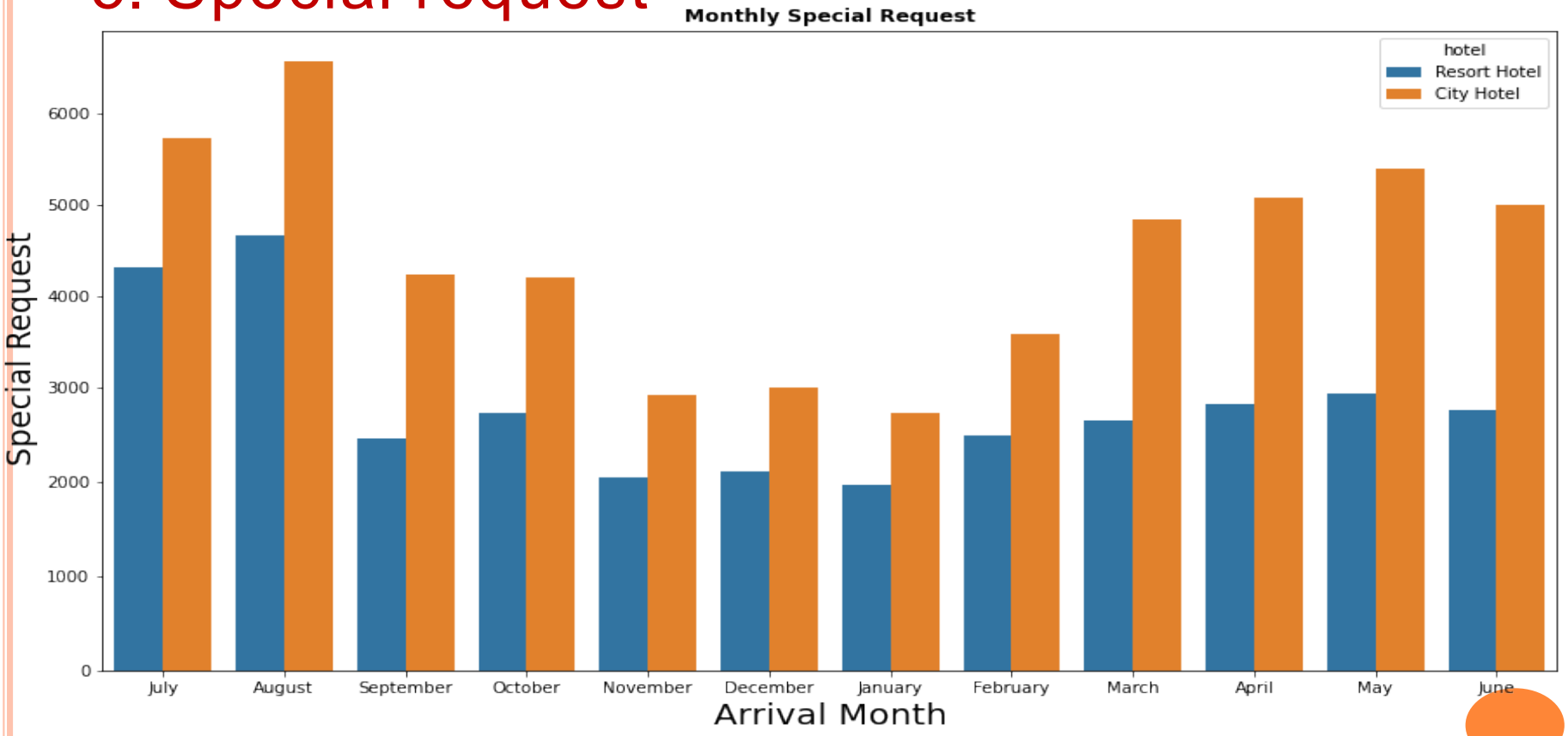


4. Mostly reserved room



From this graph it clearly shows the most reserved room are Type A followed by Type D

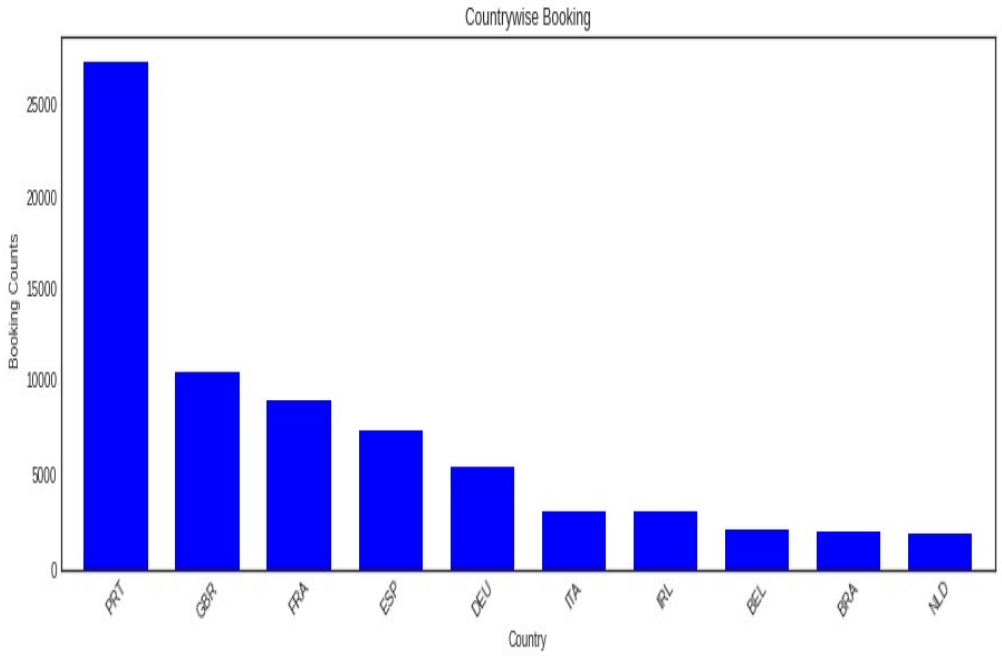
5. Special request



Most special request were from the month of August and least is in the month January for both type of hotel

6. From which country most of the booking was made

The guests was booking hotels from all over the world but from this graph it is clear that most of the guest was coming from Portugal followed by Germany and France



7. Percentage of Cancellation in Resort and City hotel

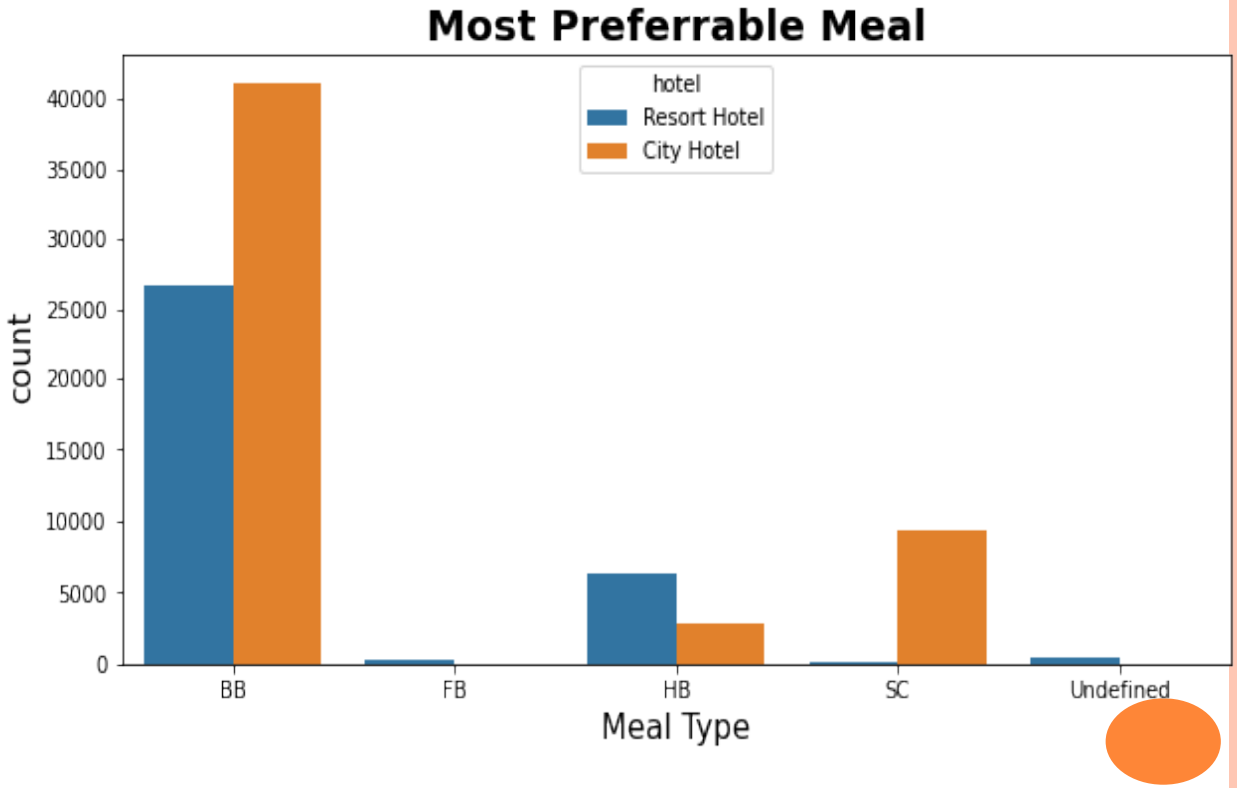
City hotel has 30.1% of cancelation.

Resort hotel has less cancelation chance than city hotel with only 23.5% of cancelation.

8. What type of meal guest prefer most

This data will help hotel management to understand customer needs and improve their services.

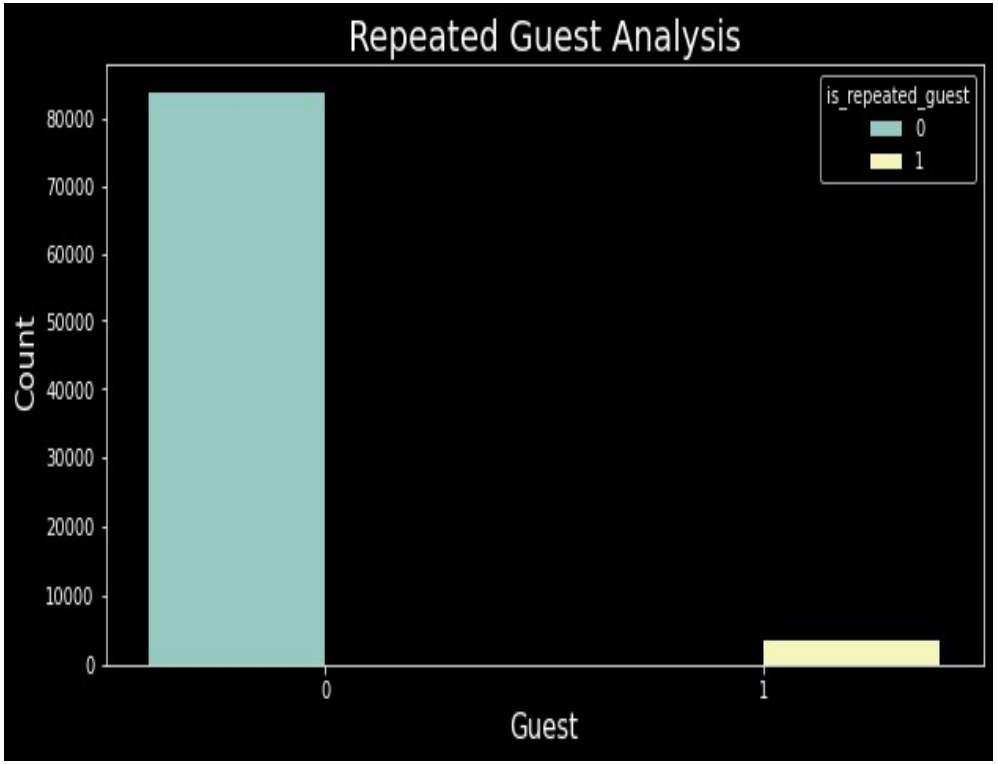
Mostly Bed and Breakfast type of meal was preferred in both hotels by guest.



9.How many guests are repeated

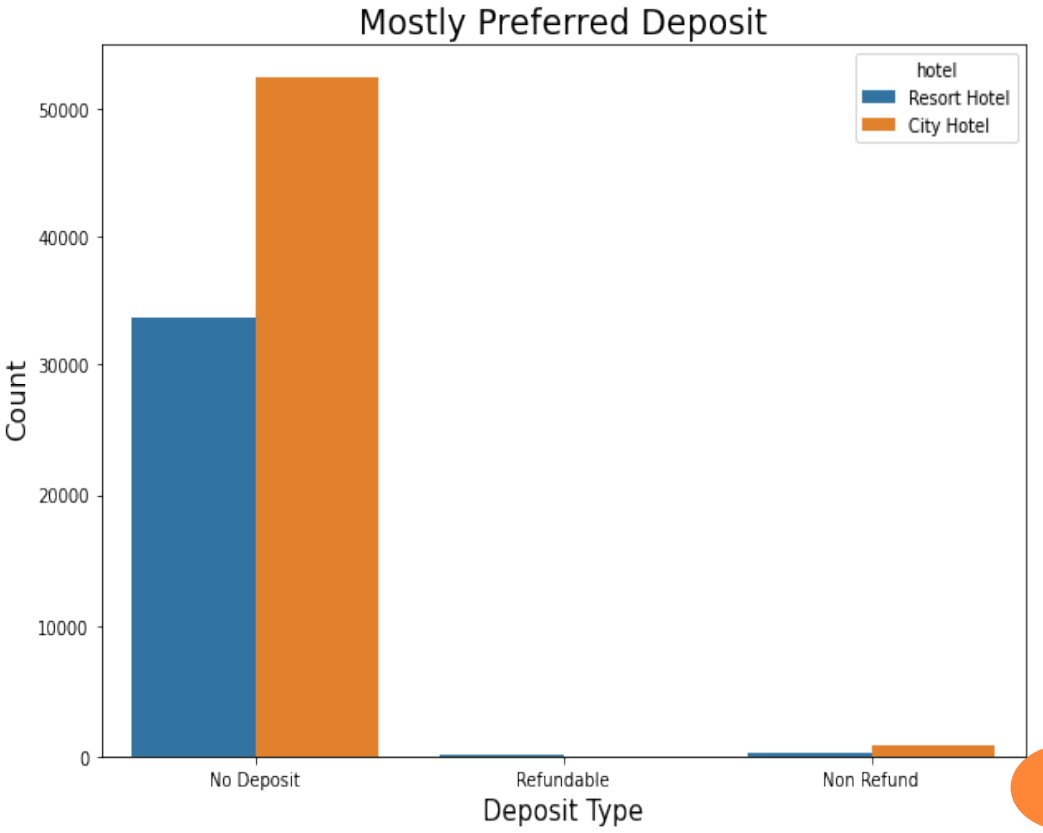
Most of the new guest is only doing booking so as a management we can learn from this graph that we have to make improvement in our services so that number of repeated guest will increase.

There might be a chance that hotels has not full filled guest requirement so that they are not repeating the booking in same hotel.

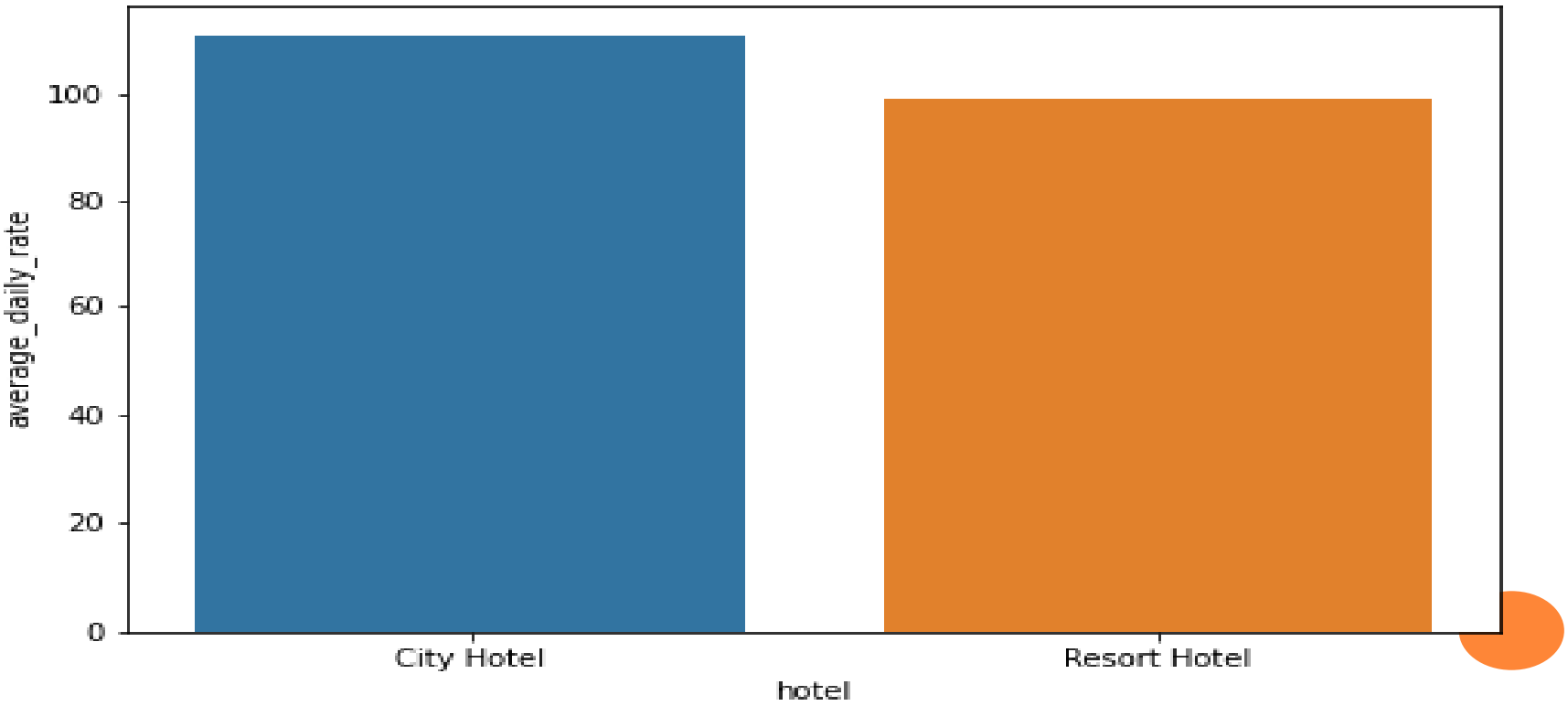


10. Deposit Analysis

From this graph we understand that guest are not making any kind of deposit before coming to the hotels.



11.Which hotel is generating more revenue



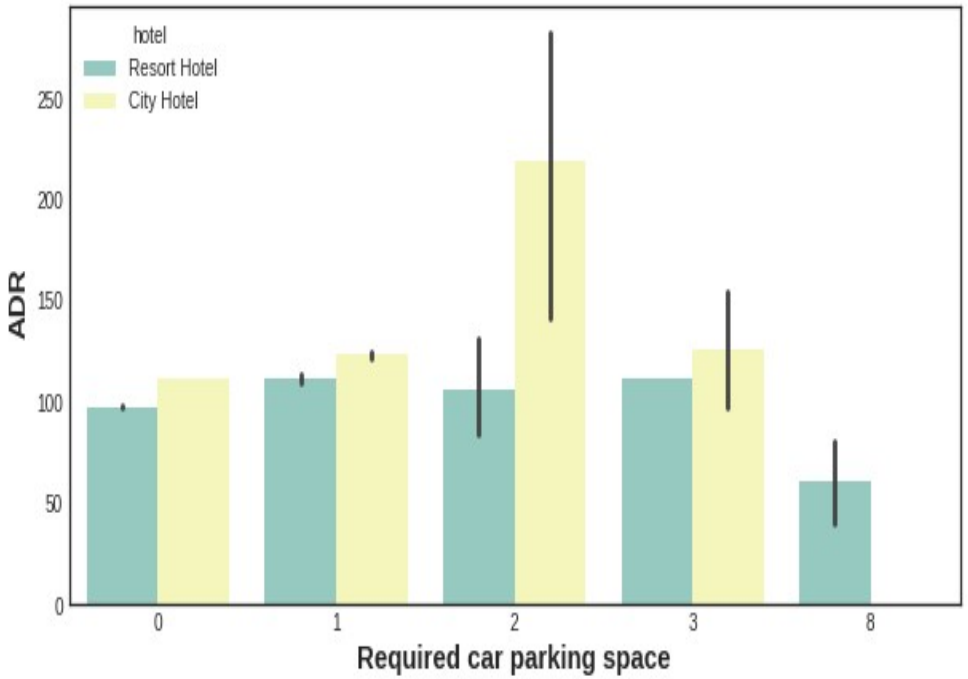
1. From above graph it is clear that City hotel has greater average daily rates as compare to Resort hotel.

2. And earlier we have seen that city hotel is mostly preferred by the guest so city hotels getting more booking than the Resort hotel.

3. So in terms of revenue City hotel will be generating more revenue than the Resort hotels

12. Effect of required car parking space on average daily rates

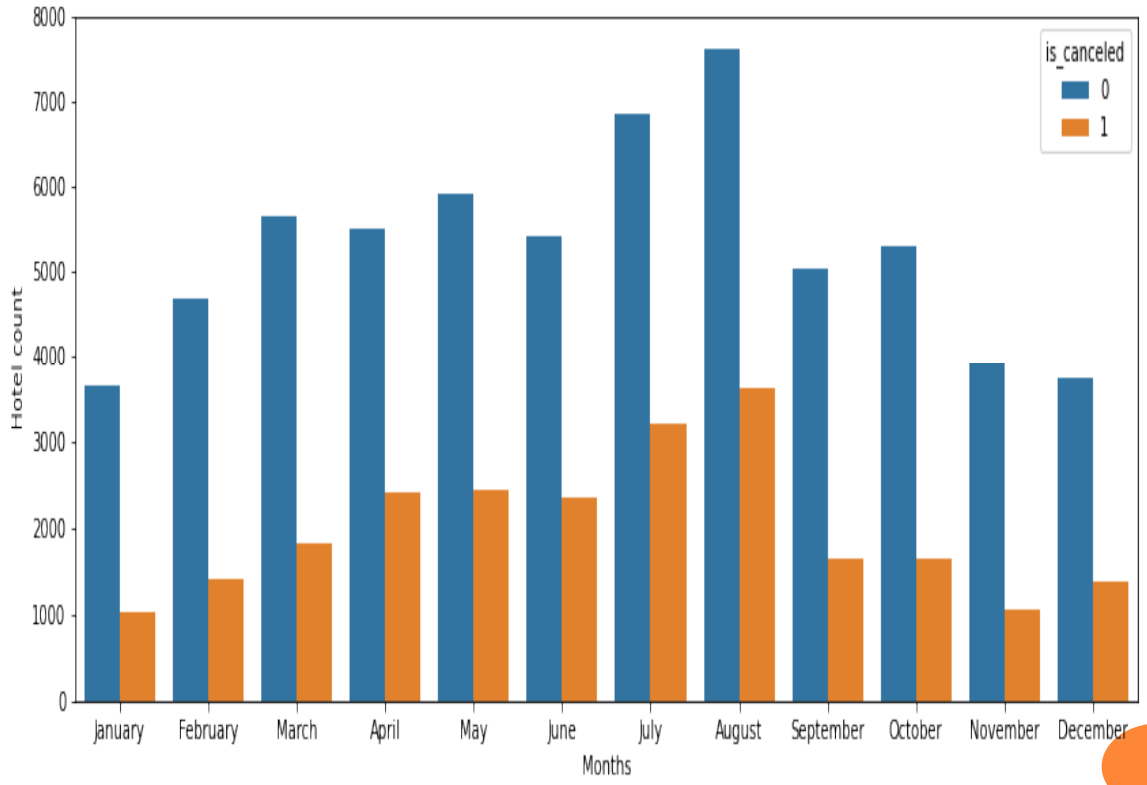
Average daily rate is high when the required car parking space is equals to two.



13.Which month having highest cancellation record

0=Not canceled,
1=canceled
From this graph we understand that in month of August maximum cancellation was done.

Also earlier we have seen that maximum booking was made in August .

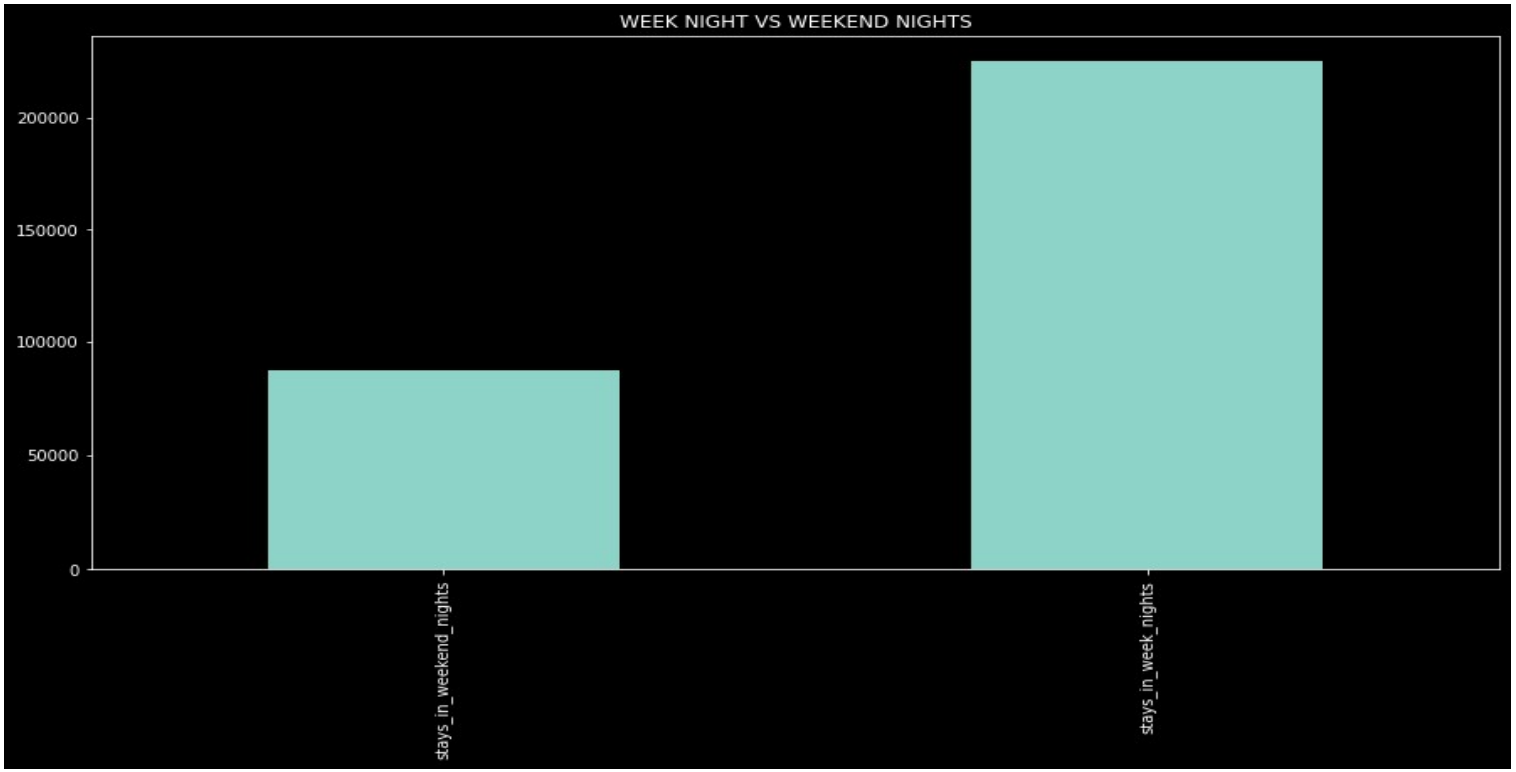


14. Which type of guest made more cancellation

As we can see that most of the cancellation was made by Transient type of customers.

This graph will help the management to look on these type of customers and make their booking confirmed by providing all the requirements asked by the customers.

15.Tourist preference in week nights v/s weekend nights

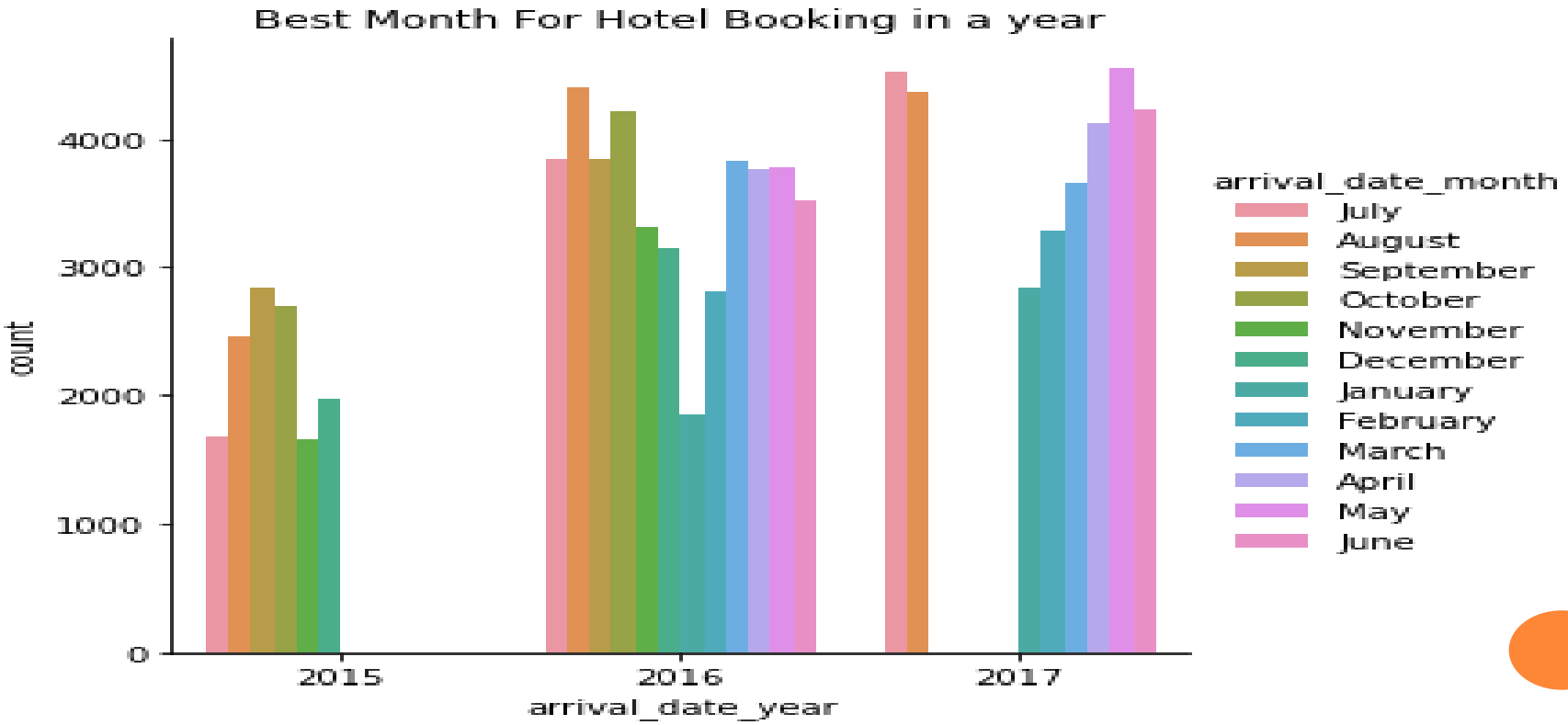


The above chart helps us visualize the variation of stays in weekend nights and stays in week nights.

We can see that the majority stays by the guests is during week nights.

So a myth has been cleared as we have always thinks that weekend night will be getting most of the booking as compare to week nights but now everyone's doubt has been cleared from above graph.

16.MONTHLY BOOKING DATA IN DIFFERENT YEARS



From above graph we got the following conclusions:

1.In 2015:

I. Most of the booking was made in month of September
II. Least booking was made in month of November.

2.In 2016:

I. Most of the booking was made in month of August
II. Least booking was made in month of January.

3. In 2017:

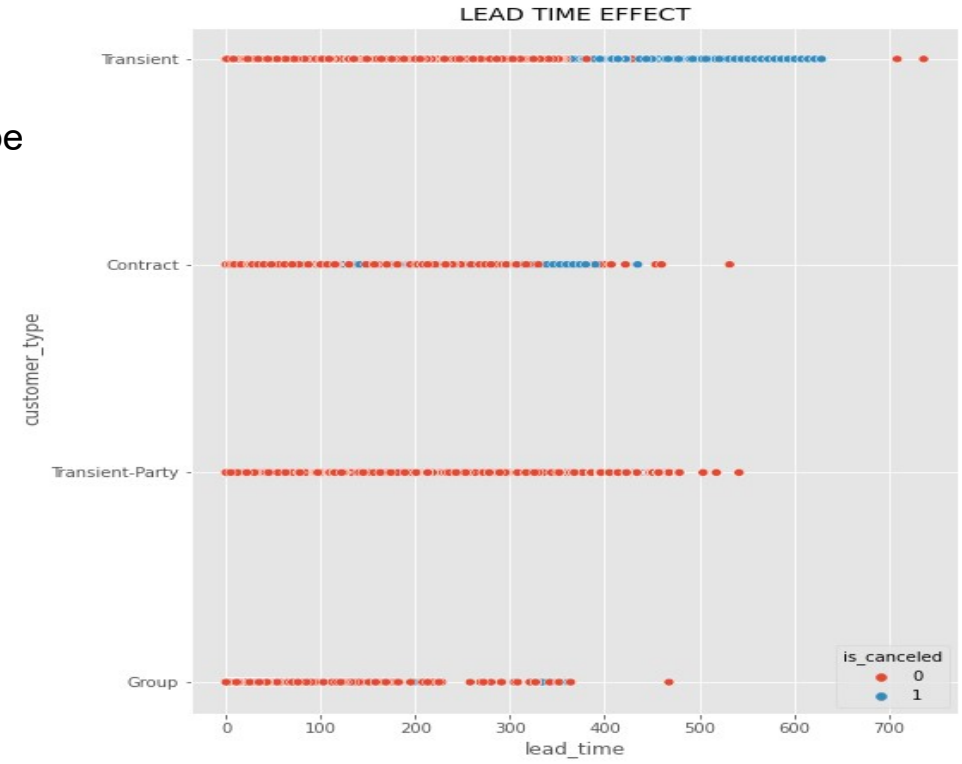
I. In 2017 the most of the booking was made in month of May

17.How lead time effect cancellation

From this scatter plot we can Understand how lead time effect the cancellation with different type of customers

As the lead time increases the chances of cancellation also increases

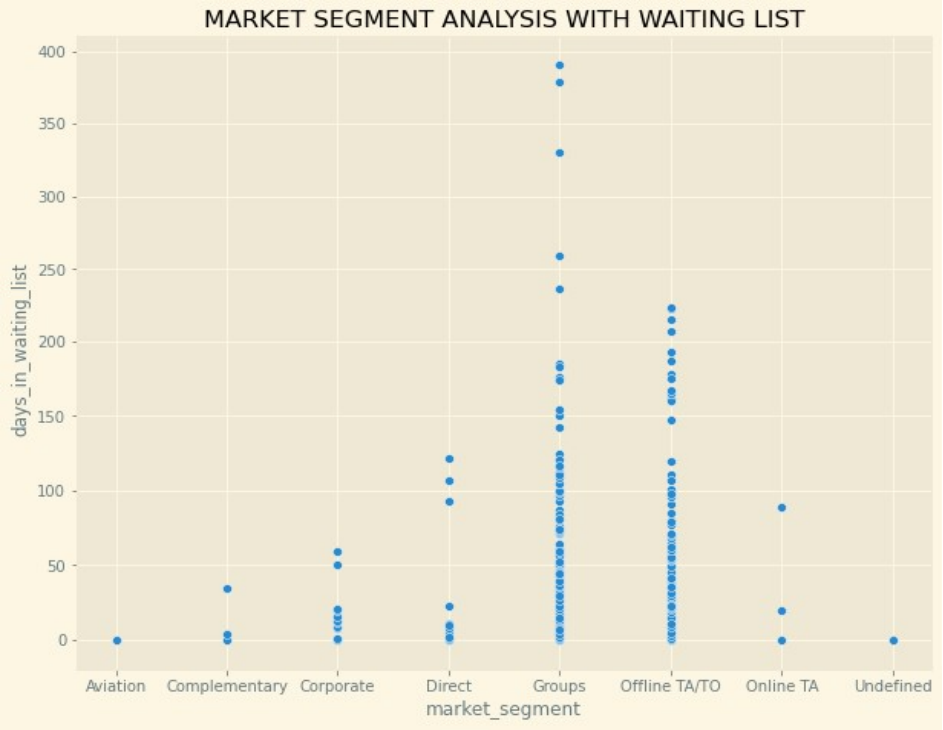
Transient type of customers cancel most bookings as lead time increases



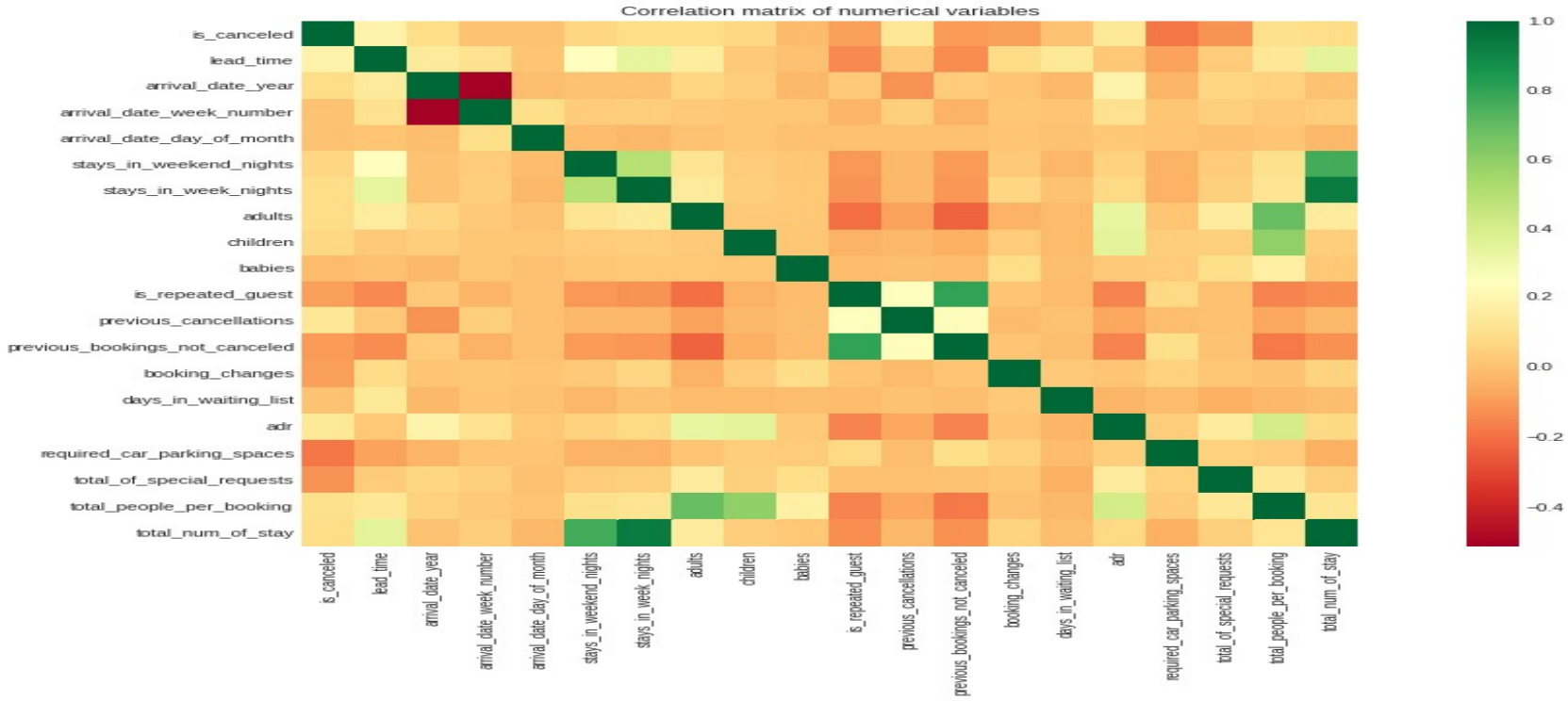
18. Market segment analysis with waiting list

From this scatter plot we can understand how different market segment varies with waiting list time

It clear understand that waiting list time is high for Group market segment



19:Co-relation of the column



Co-relation is helpful to understand the relationship between different parameters in the dataset, how they affect each other and the overall business. Understanding correlation is essential to get a brief overview of how the different parameters simultaneously vary with each other. This gives us a brief idea of the overall dataset and how the values of each table vary throughout.

Co-relation shows us the relationship between the columns like tthey are positively co-related to each other or negatively related.

Using the heatmap, we have drawn the above co-relation map.

Thank You!