

Capstone Project-1

Exploratory Data Analysis

Team Members

1. Jitesh Kumar Srivastava

2. Aman Kumar Sharma

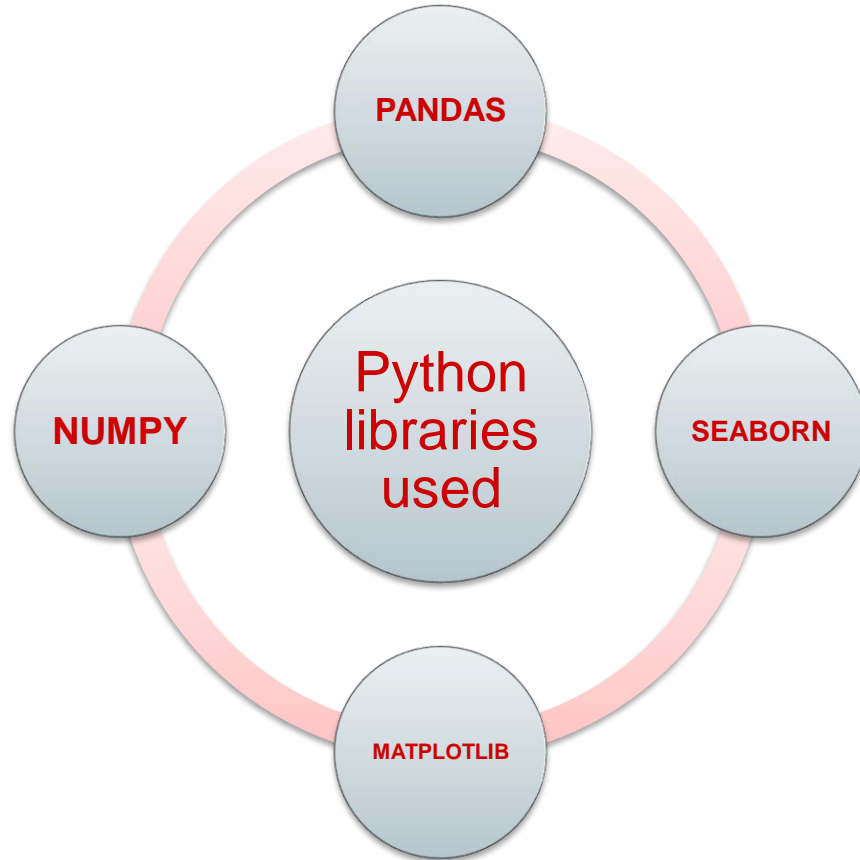
3. Pavani Billa

HOTEL BOOKING EDA

E- Exploratory
D- Data
A- Analysis



PYTHON LIBRARIES USED

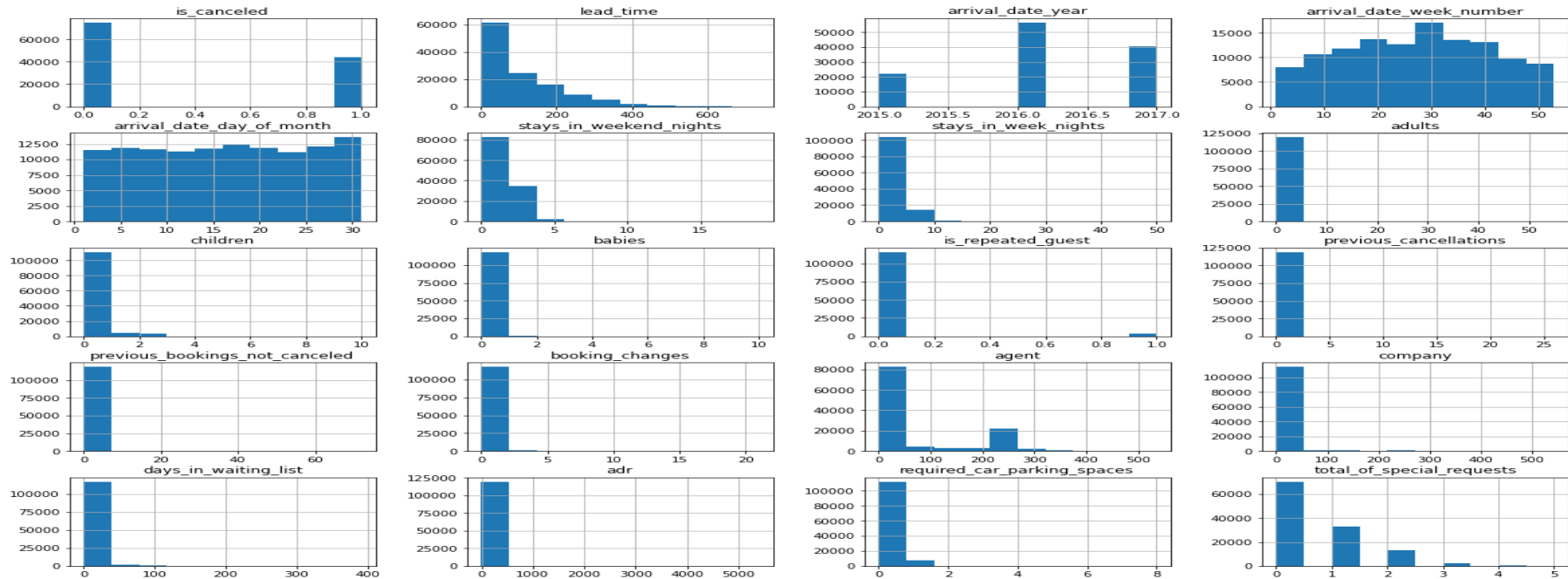


How we analysed relations in different dataset ?

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.

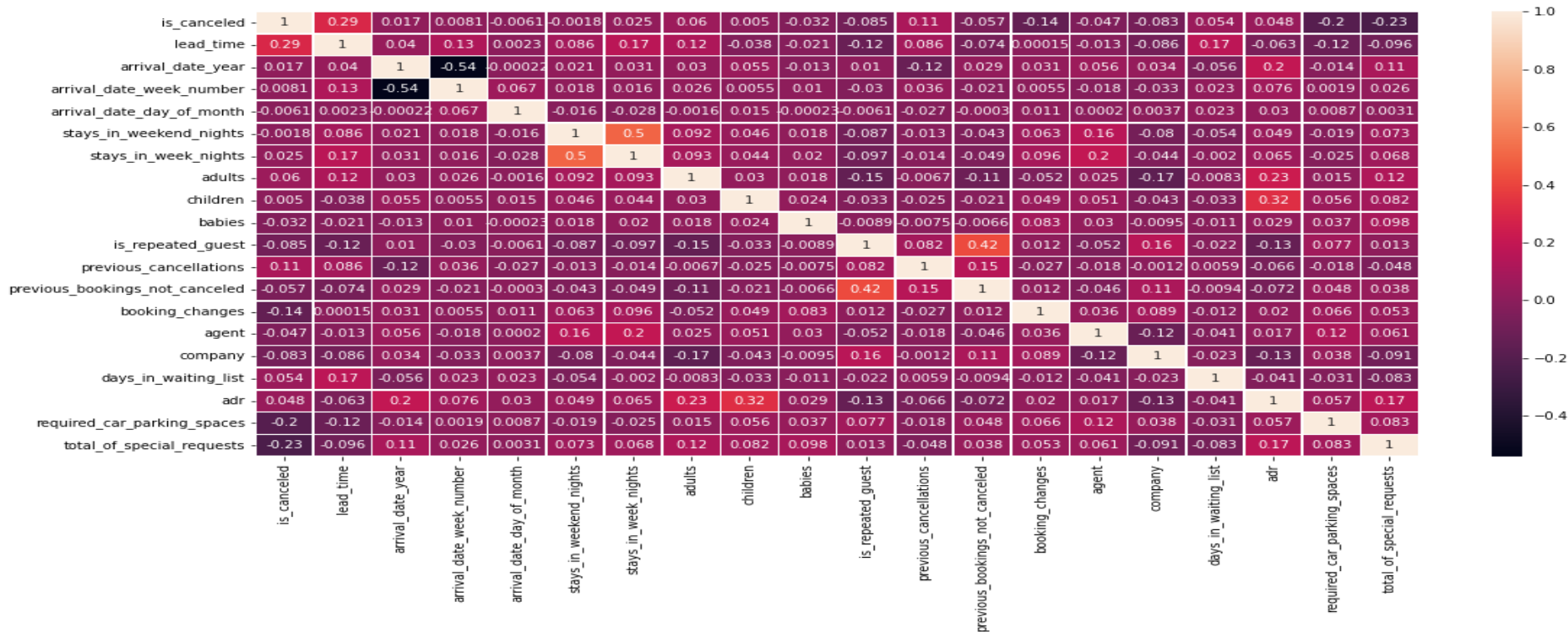
And we as a team have tried to do the same thing. The hotel booking dataset provided us with many things to lookout for and analyse them. The dataset was quite big and as a team we decided to divide the task that made our work a bit easier.

OUTLIERS



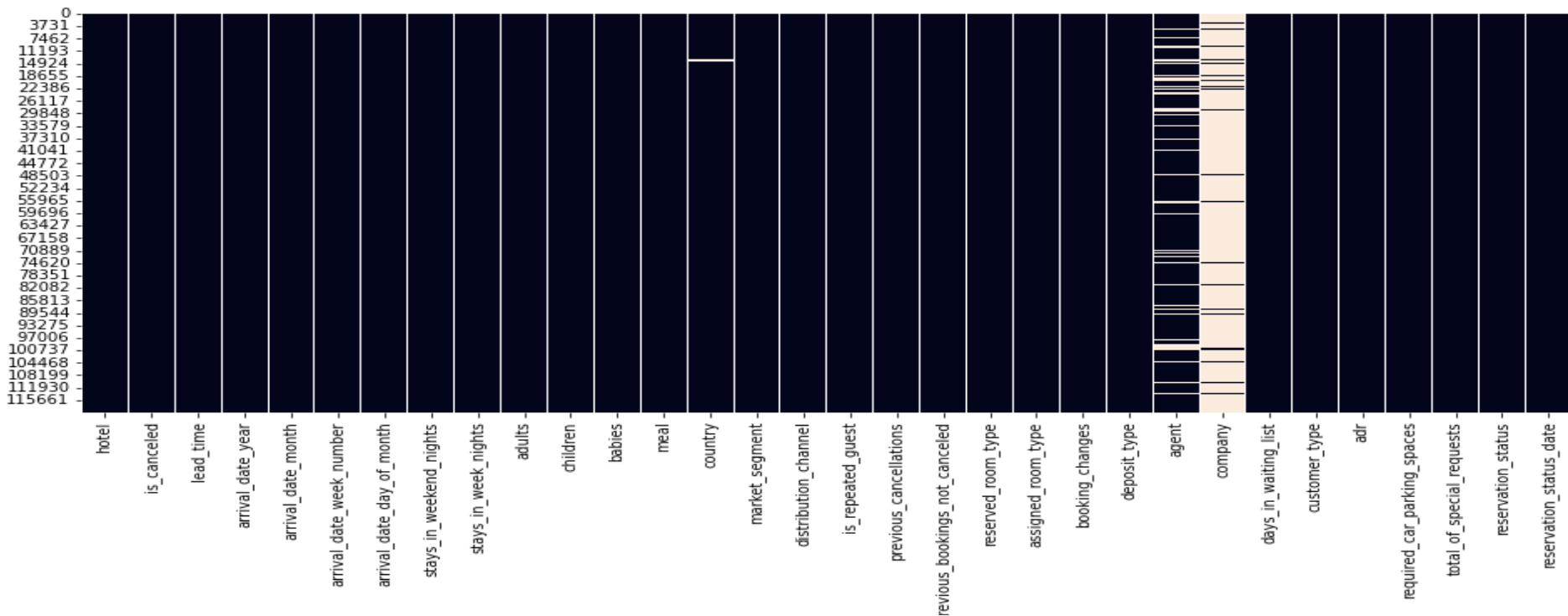
- We can see that the maximum guest came in the year 2016.
- Maximum arrival week number is 30.
- Maximum arrival happens in the last of the month.

CORRELATION HEATMAP



There is no correlation between the variables except diagonal element.
So we can deduce that variables are independent.

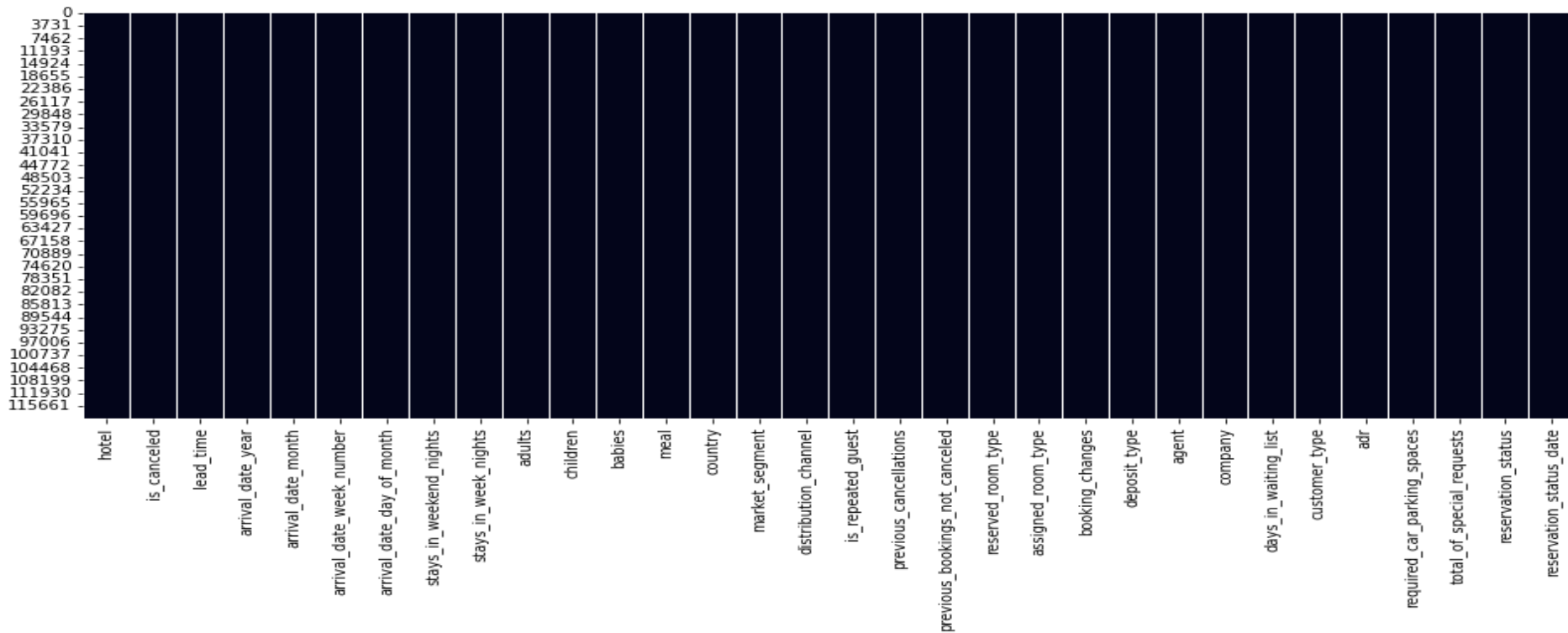
PREPARATION OF THE DATA



Columns company and agent are having most of the null rows.

It will not add any significant value if more than 50 percent data is missing.

REMOVAL OF NULL VALUES



We can see that now there is no null values in the dataset.

CONFIRMATION AND CANCELLATIONS BY DIFFERENT HOTEL



PERCENTAGE BOOKINGS OF RESORT HOTEL

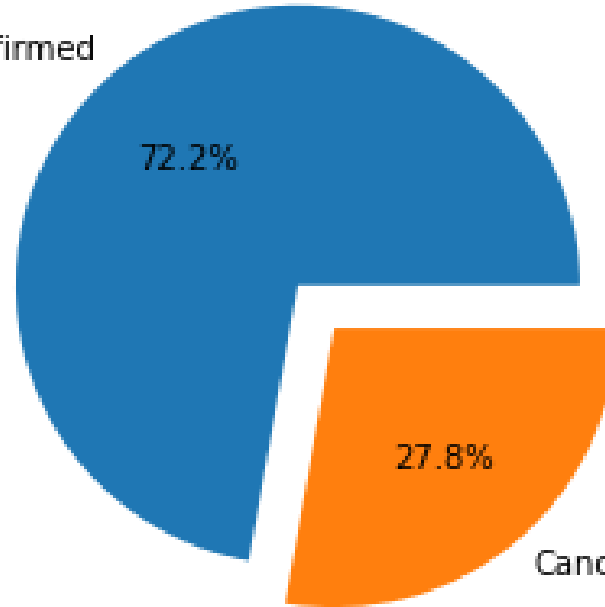
Resort Hotel Confirmed and Cancellations

Confirmed

72.2%

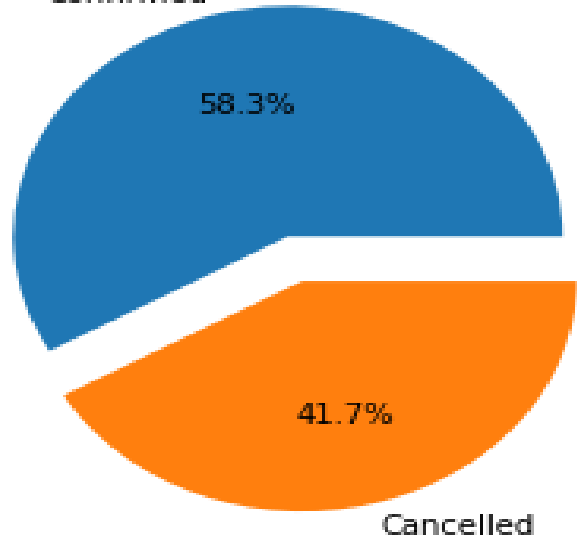
27.8%

Cancelled



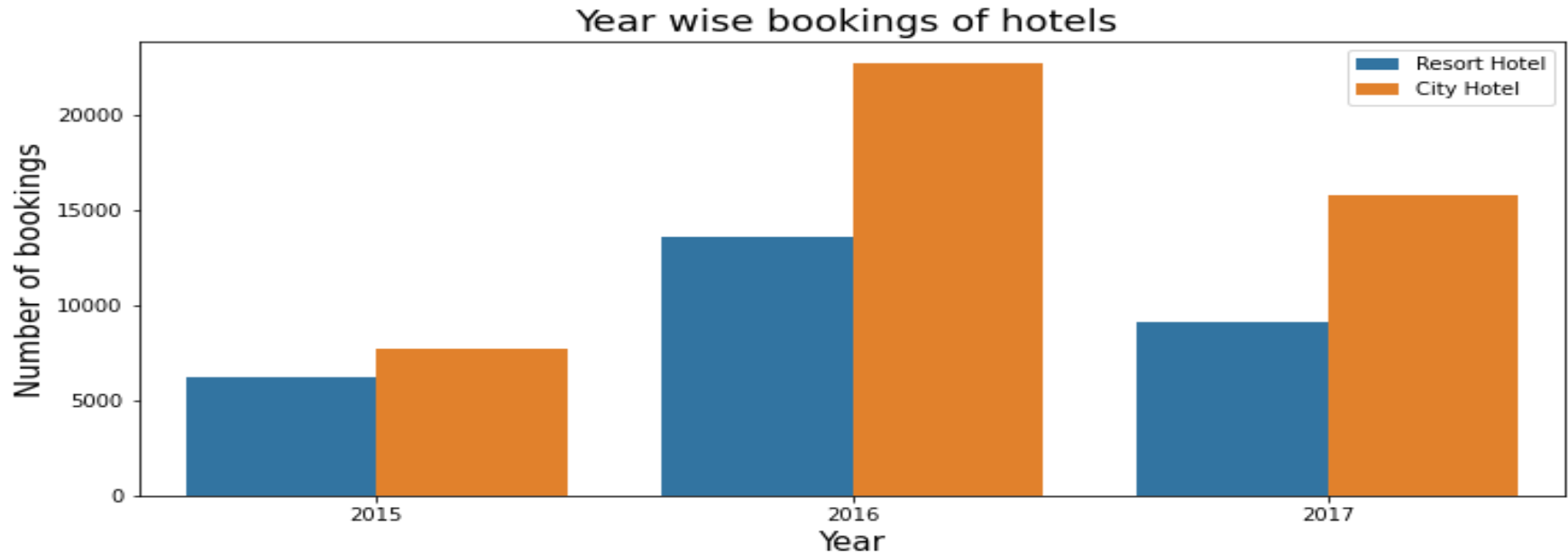
PERCENTAGE BOOKINGS OF CITY HOTEL

City Hotel Confirmed and Cancellations



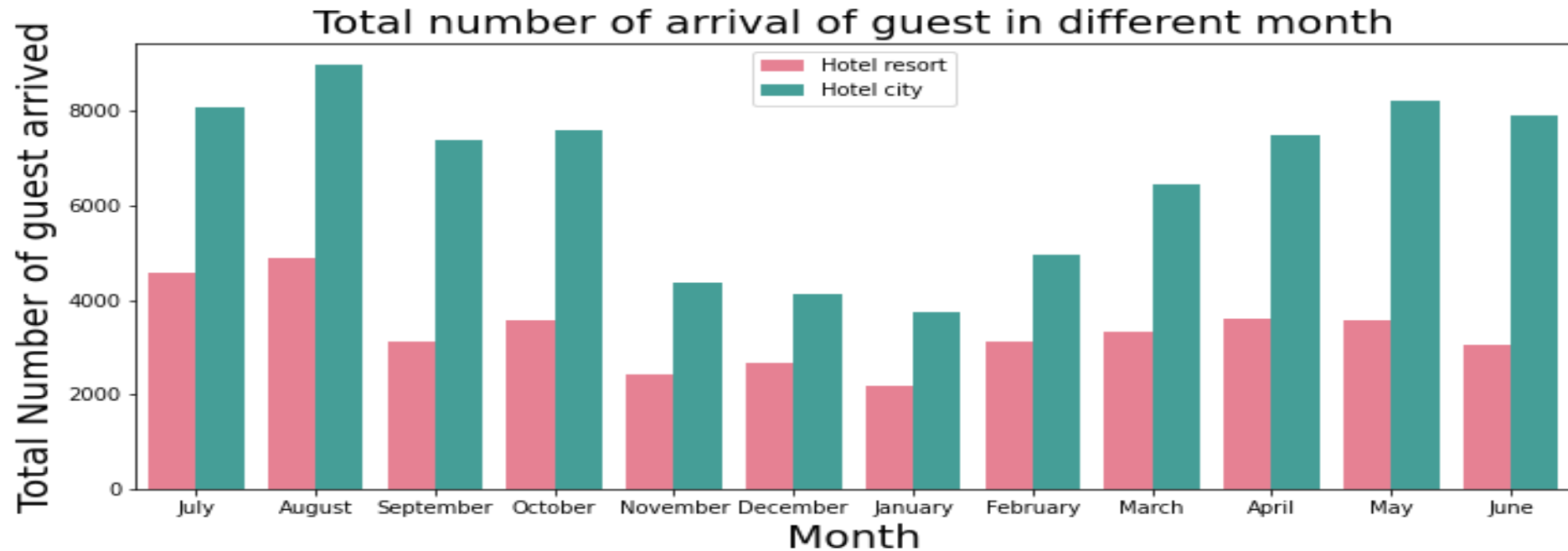
We can clearly deduce from the above graphs that the City hotel is having greater number of bookings as compared to Resort hotel. But , the cancellation percentage is high of the City hotel.

YEAR WISE BOOKINGS OF HOTELS



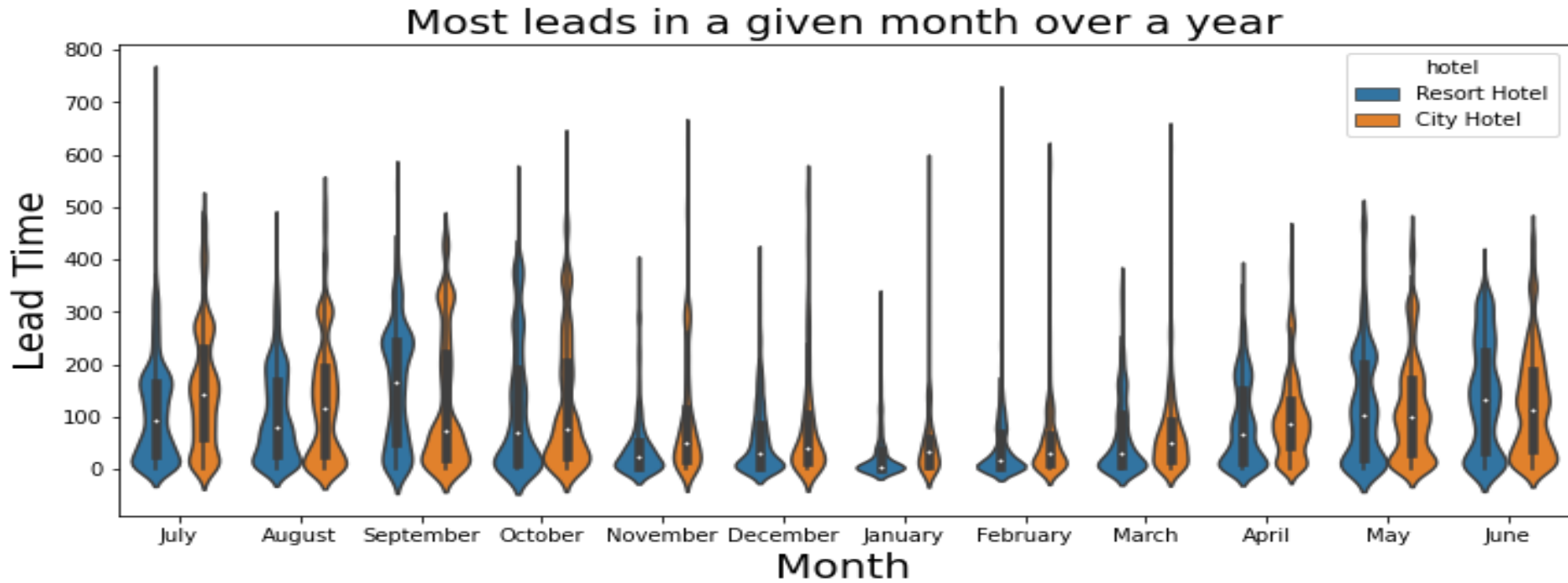
From the above graph it can be summarised that in the year 2016 both the hotel saw a massive increase in their bookings and by far the year 2016 is the year of the highest bookings of both hotel. In each year that is 2015, 2016 and 2017 the City hotel is having the highest number of bookings.

ARRIVAL OF GUESTS PER MONTH



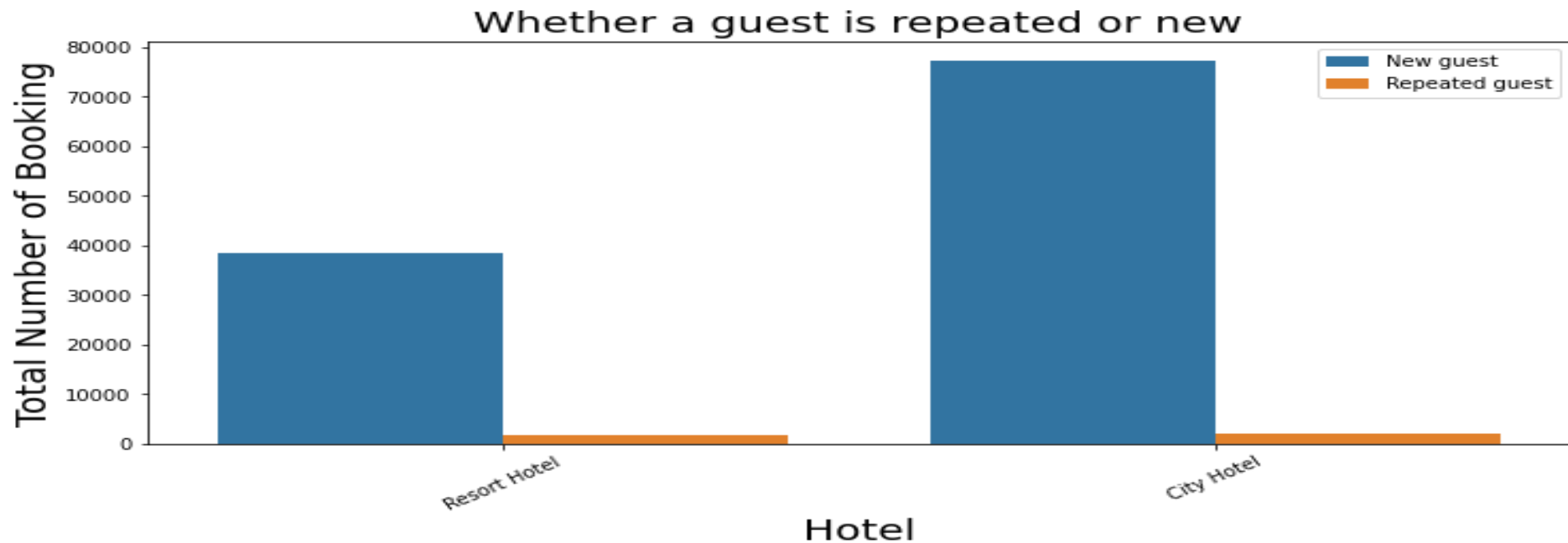
Most number of the guests arrived in the month of August and least was in the month of January.

LEAD TIME FOR VARIOUS SEGMENTS



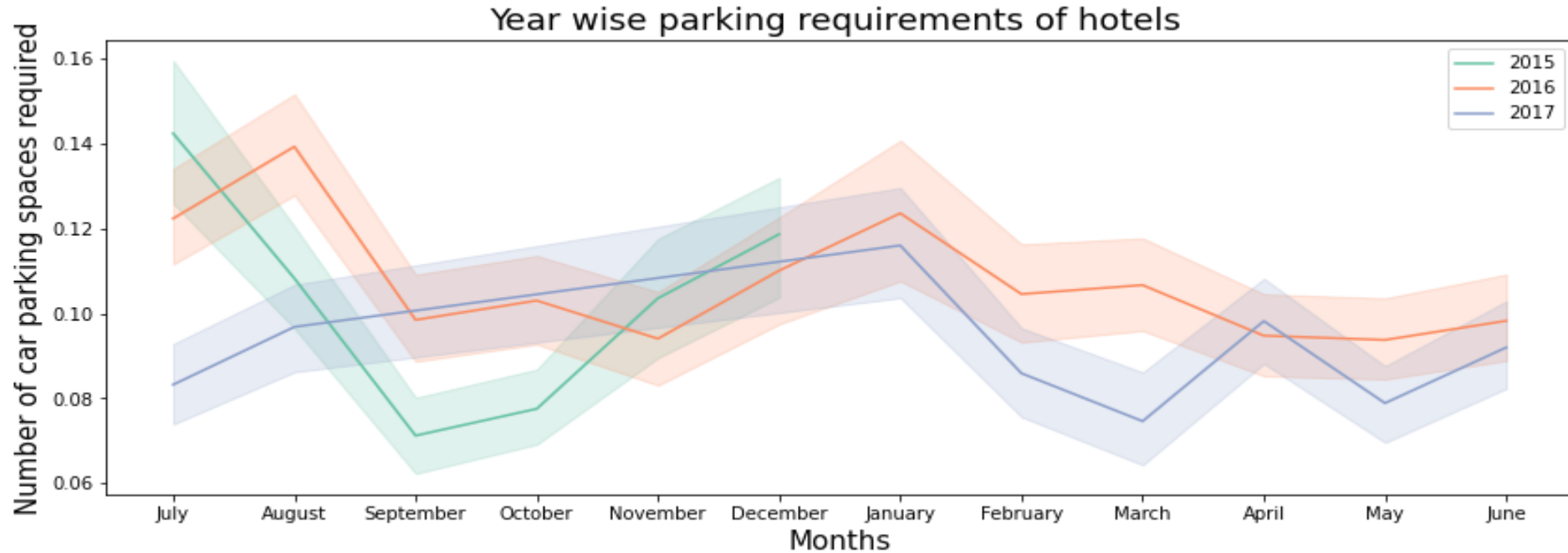
Most number of lead time was generated in the July month and least was in the January month.

TYPES OF NEW AND REPEATED GUEST



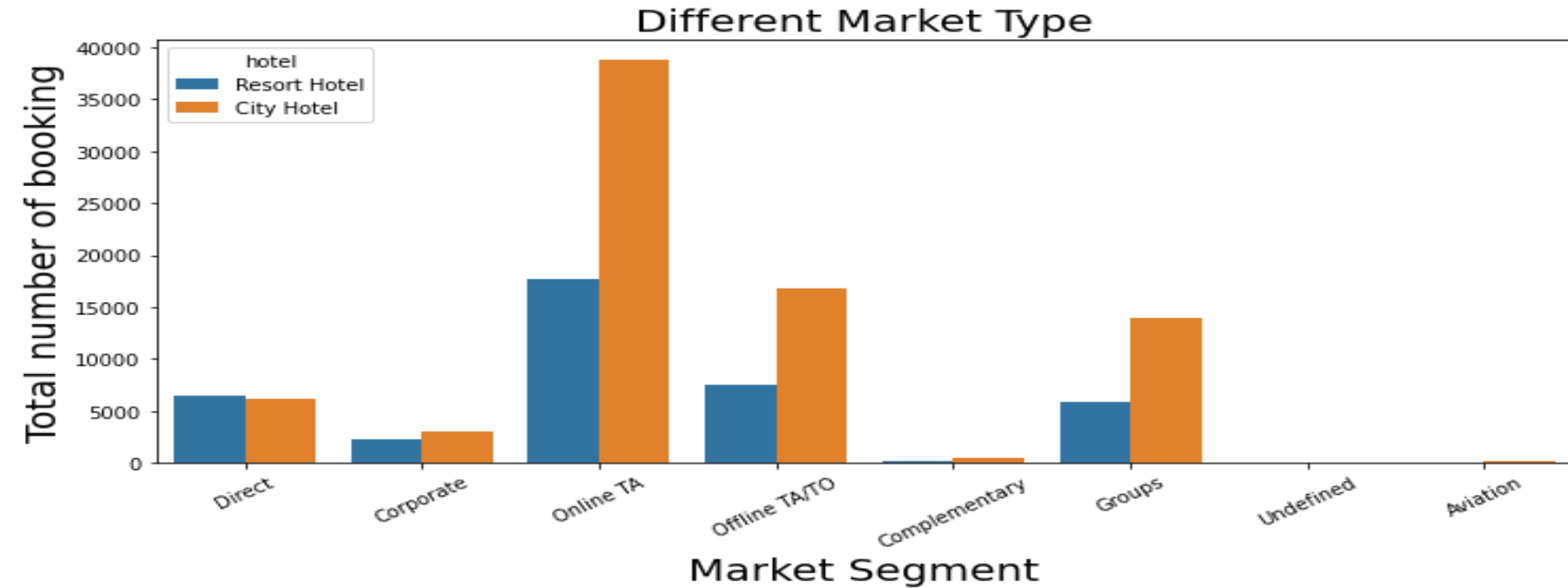
We can see that there is very low number of repeated guest in both the hotel. A lot of attention should be drawn to solve this problem because repeated guest know the services of hotel very well.

YEAR WISE CAR PARKING SPACES



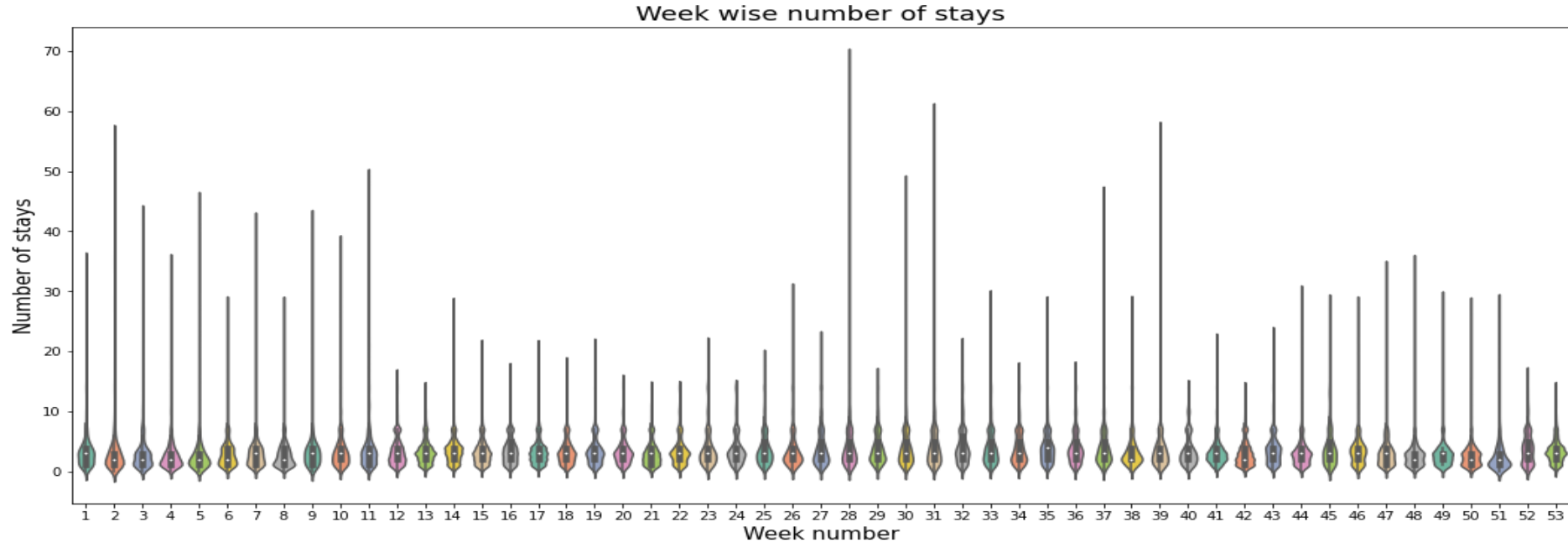
From the line chart it can be inferred that in the year 2015 from July the requirement of the car parking spaces was decreasing till September and then in the year 2016 in the month of August the requirement was on peak .where as in the year 2017 there was a steady requirement of parking spaces.

NUMBER OF BOOKING FOR MARKET SEGMENT



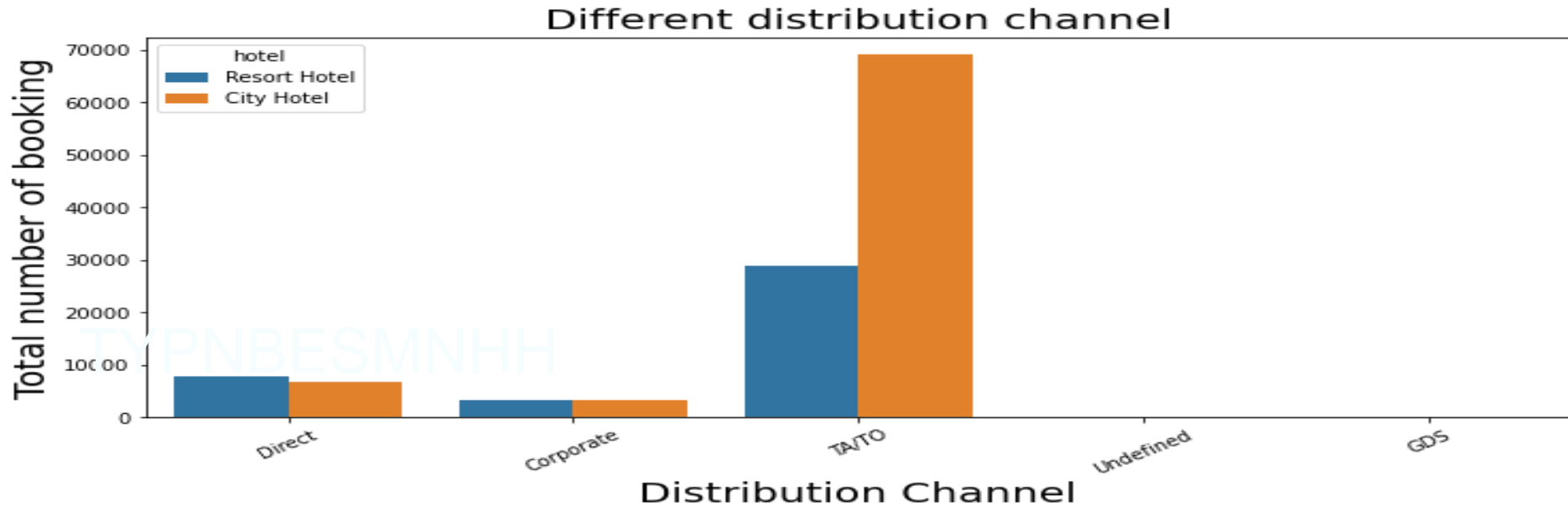
Clearly, We can see that Online TA has the maximum booking followed by offline TA/TO in maximum booking and Aviation has the least booking.

DURATION OF STAY



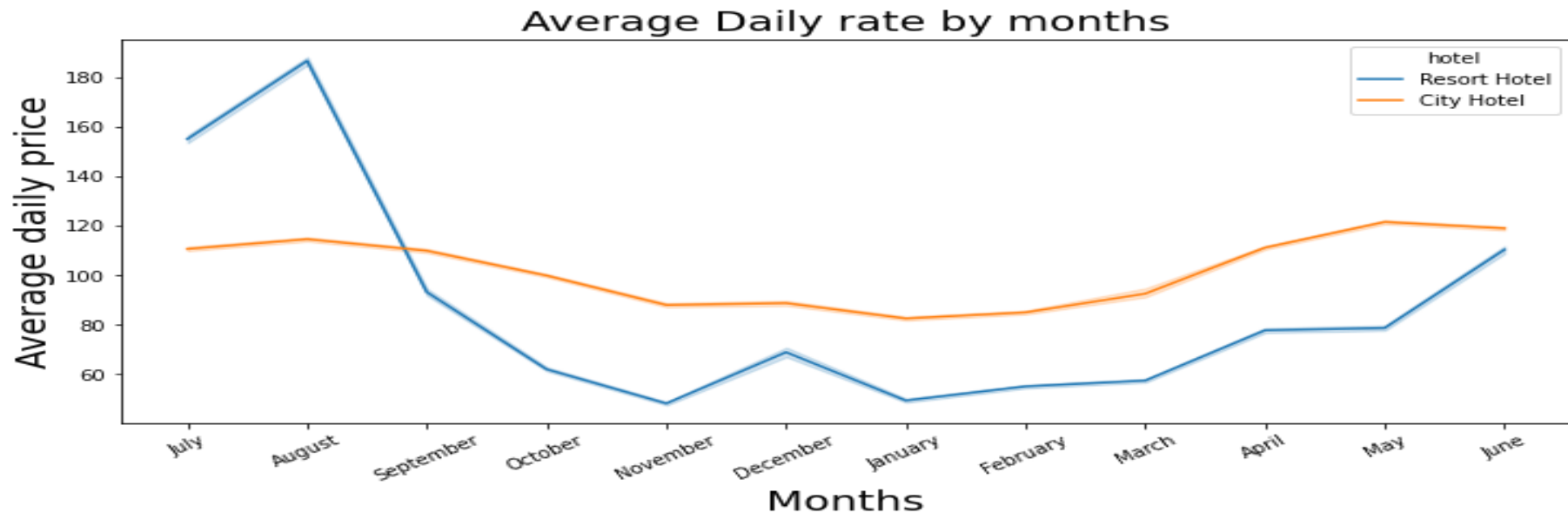
The above violin chart it can be noticed that from the week 28 to 31 has shown the highest days of stay where as from the week 1 to 11 has shown a very steady trend in the number of stays and also the week 18 to 22 has shown the least number of stays by the visitors in aggregate of all 3 years 2015, 2016 and 2017.

TYPES OF DISTRIBUTION CHANNEL



TA/TO of the hotels have maximum contribution to the booking followed by Direct . Both the hotels need to expand their distribution channel which in results will expand their businesses.

AVERAGE PRICE BY MONTHS

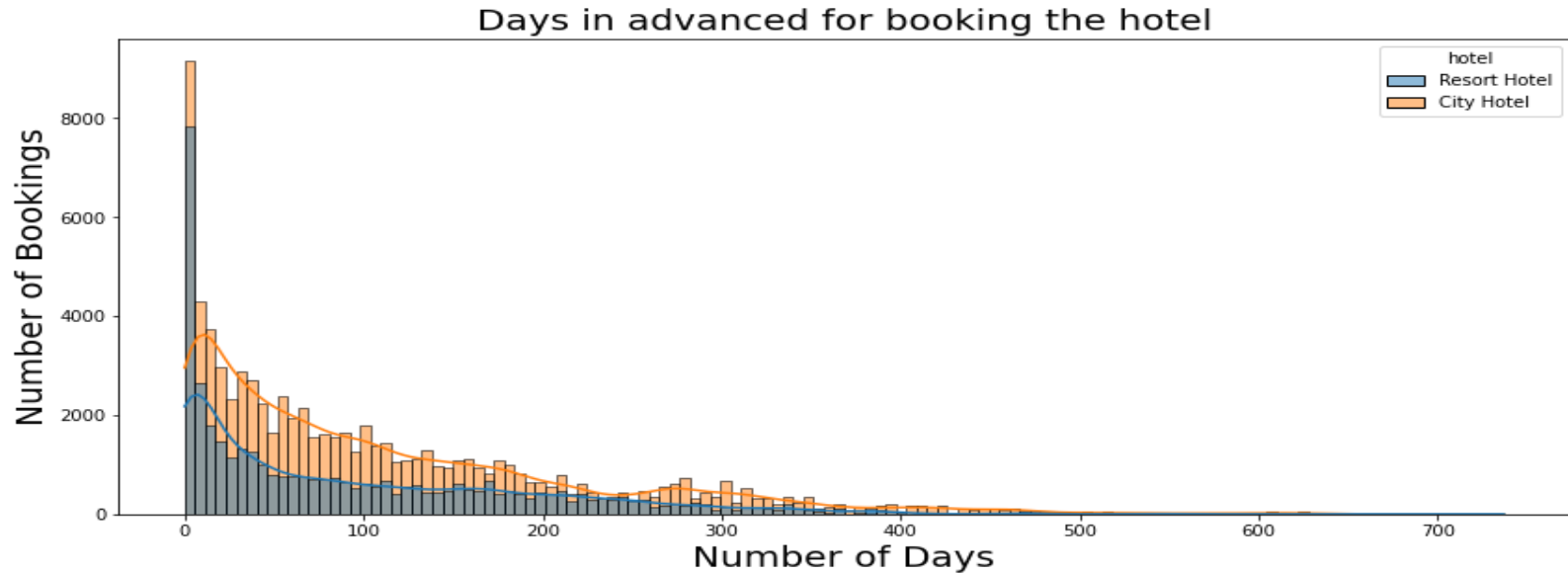


By looking closely at the line chart, we can say that for Resort hotel the highest average daily rates is in the month of July and City hotel the highest average daily rates is in the month of May. For Resort hotel the least average daily rate is in the month of November and for City hotel it is in month of January.

We can also say that average price (adr) of City hotel is somewhere between 85 to 125 and for Resort hotel the average price (adr) is between 55 to 185 thus we can say that the average price of hotel Resort is volatile.

The graph depicts that for the month of July to mid August stay at hotel Resort is costlier and from month of September to June stay at hotel City is costlier.

BOOKINGS FOR NUMBER OF DAYS



From the histogram we can say that most of the booking of both the hotels are done within 100 days of check-in date to hotel and maximum booking are done on the same check-in day.

CONCLUSION

Now , we are down to the last part of this project that is concluding, so far we have analysed the various variables and on the basis of that we have visualised the data.

The findings were as follows:

1. The City hotel is having greater number of bookings as compared to Resort hotel . But, the cancellation percentage is high of the City hotel.
- 2 . In the year 2016 both the hotel saw a massive increase in their bookings and by far the year 2016 is the year of the highest bookings of hotel . In each year that is 2015,2016 and 2017 the City hotel is having the highest number of bookings.
- 3 . Most number of the guests arrived in the month of August and least was in the month of January.

- 4 . Most number of lead time was generated in the July month and least was in January month.
- 5 . We can see that there is very low number of repeated guest in both the hotel. A lot of attention should be drawn to solve this problem because repeated guest know the services of hotel very well.
- 6 . Most of the visitors do not carry children with them every year. Very less number of children been carried by them around 1 and 2 only every year.
- 7 . In the year 2015 from July the requirement of the car parking spaces was decreasing till September and then in the year 2016 in the month of August the requirement was on peak. Where as in the year 2017 there was a steady requirement of parking spaces.
- 8 . Online TA has the maximum booking followed by offline TA/TO in maximum booking and Aviation has the least booking.
9. Portugal is the country from where maximum visitors came followed by Great Britain, France and Spain. We have taken only the booking which has been completed.

10 . From the week 28 to 31 has shown the highest days of stays whereas from the week 1 to 11 has shown a very steady trend in the number of stays and also the week 18 to 22 has shown the least number of stays by the visitors in aggregate of all 3 years 2015, 2016 and 2017.

11. TA/TO of the hotels have maximum contribution to the booking followed by direct. Both the hotels need to expand their distribution channel which in results will expand their businesses.

12. Resort hotel the highest average daily rates is in the month of July City hotel the highest average daily rates is in the month of May . For Resort hotel the least average daily rate is in month of November and for City hotel it is in month of January. We can also say that average price(adr) of City hotel is somewhere between 85 to 125 and for Resort hotel the average price(adr) is between 55 to 185 thus we can say that the average price of hotel Resort is volatile. The graph depicts that for the month of July to mid August stay at hotel Resort is costlier and from month of September to June stay at hotel City is costlier.

13. Most of the bookings of both hotels are done within 100 days of check-in date to hotel and maximum booking are done on the same check-in day.

We treated the various null values accordingly and did the correlation analysis. We took help from the various python documentation present and the various modules to analyze and visualized the data.

Hope you have enjoyed the journey till here,

THANK YOU !!