

Capstone Project

Hotel Booking Analysis
By
Prajakta Dangale

Outline

1. Problem Statement
2. Data Summary
3. Data Preparation
4. EDA
5. Challenges
6. Conclusion



Problem Statement

- ❖ Tourism is one of the world's most rapidly growing industries. Much of its growth is due to higher disposable incomes, increased leisure time and falling costs of travel.
- ❖ A hotel system manages information about rooms, reservations, customers, and customer billing.
- ❖ Hotel industry facing to analyse the problems like change in marketing trends and dynamics, housekeeping issues, customers' expectations, Data security, mode of bookings etc...
- ❖ We used the given data set to predict the future bookings using pandas data frame techniques.
- ❖ We will be using the data set available to analyze the factors affecting the hotel bookings.

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(contd..)

market_segment: This column show how reservation was made and what is the purpose of reservation. Eg, corporate means corporate trip, TA for travel agency.

distribution_channel: The medium through booking was made.

[Direct,Corporate,TA/TO,undefined,GDS.]

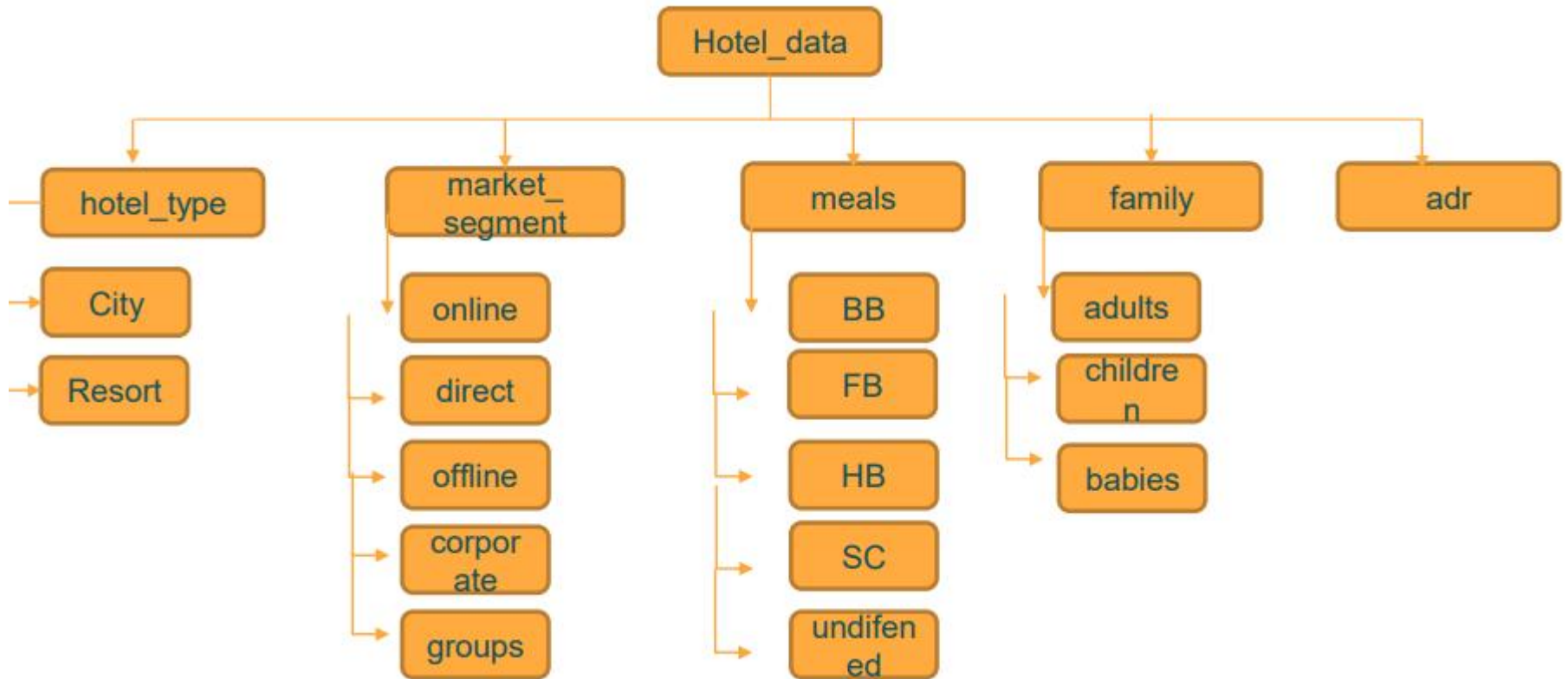
Is_repeated_guest: Shows if the guest is who has arrived earlier or not.

Values[0,1]-->0

indicates no and 1 indicated 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 Cleaning

Data cleaning is the very first important fundamental thing which every data scientist must know. It is the process of finding the inaccurate, incorrect and irrelevant or missing part of a data and then, modifying the data according to our necessity.

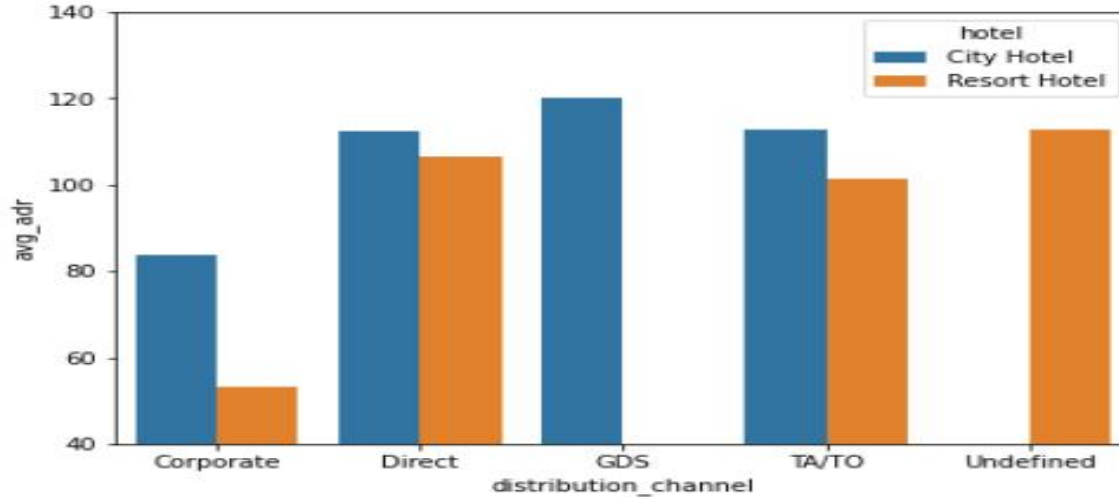
- Steps:
 1. Remove duplicate rows.
 2. Removing NULL values by replacing zero.
 3. Converting datatypes.
 4. Adding new columns if necessary.

Data Visualization

Mainly performed using Matplotlib and Seaborn library and the following graph and plots had been used:

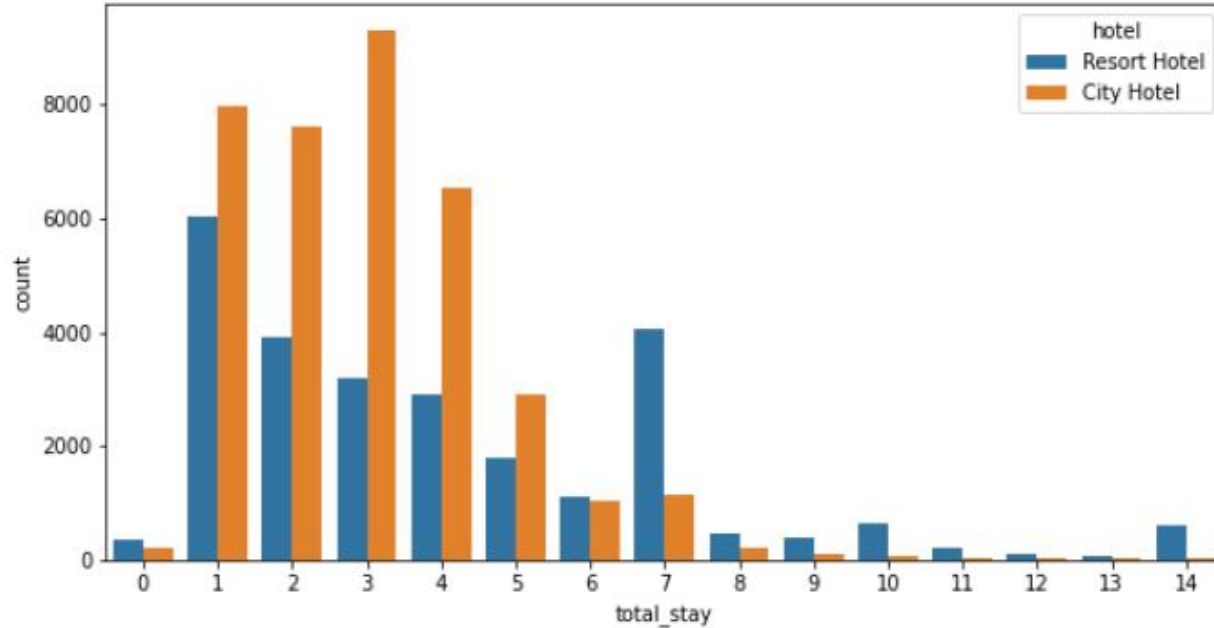
- 1) Bar Plot
- 2) Pie Chart
- 3) Heatmap
- 4) Line Plot
- 5) Histogram

EDA

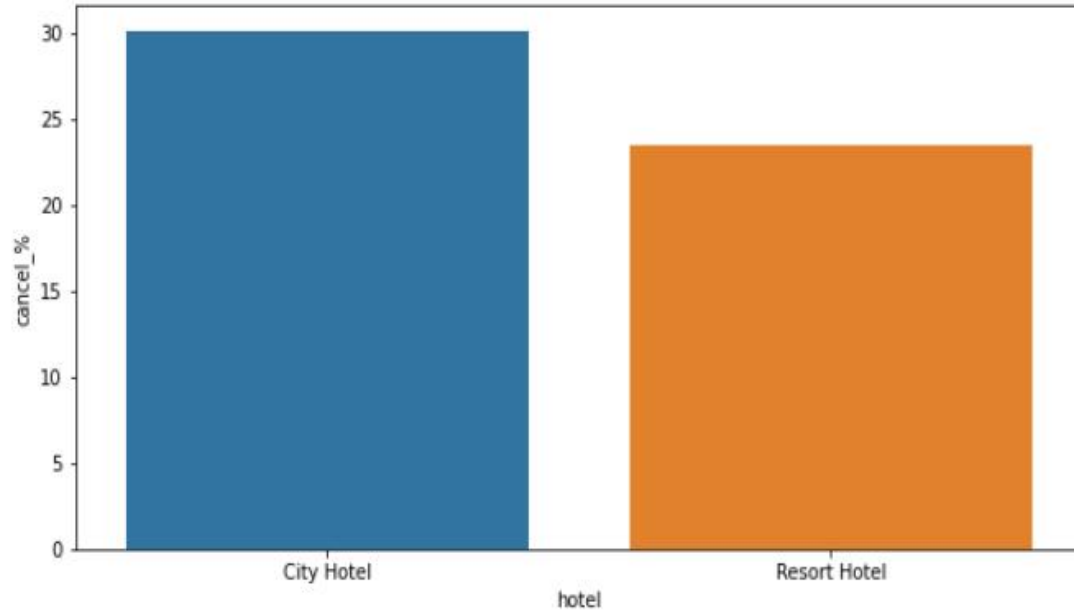


GDS channel brings higher revenue generating deals for City hotel, in contrast to that most bookings come via TA/TO. City Hotel can work to increase outreach on GDS channels to get more higher revenue generating deals.

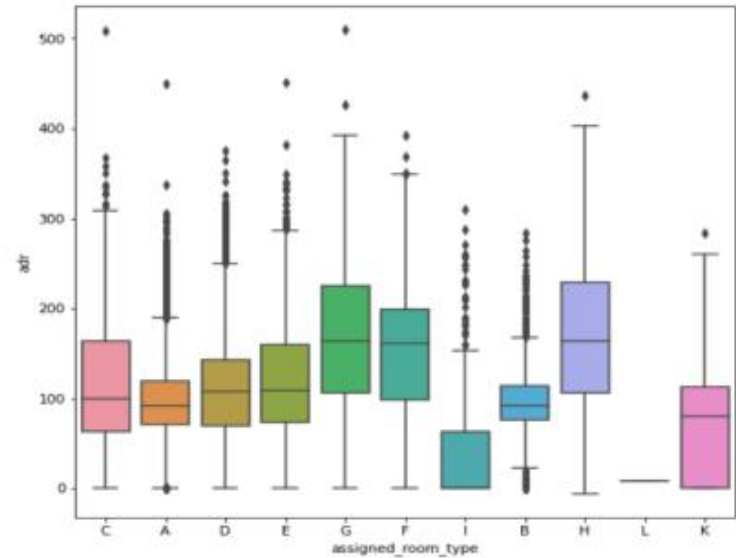
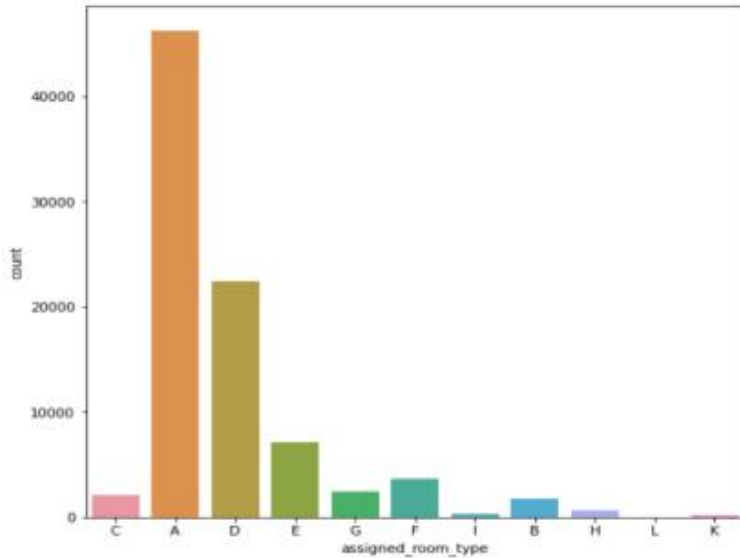
Resort hotel has more revenue generating deals by direct and TA/TO channel. Resort Hotel need to increase outreach on GDS channel to increase revenue.



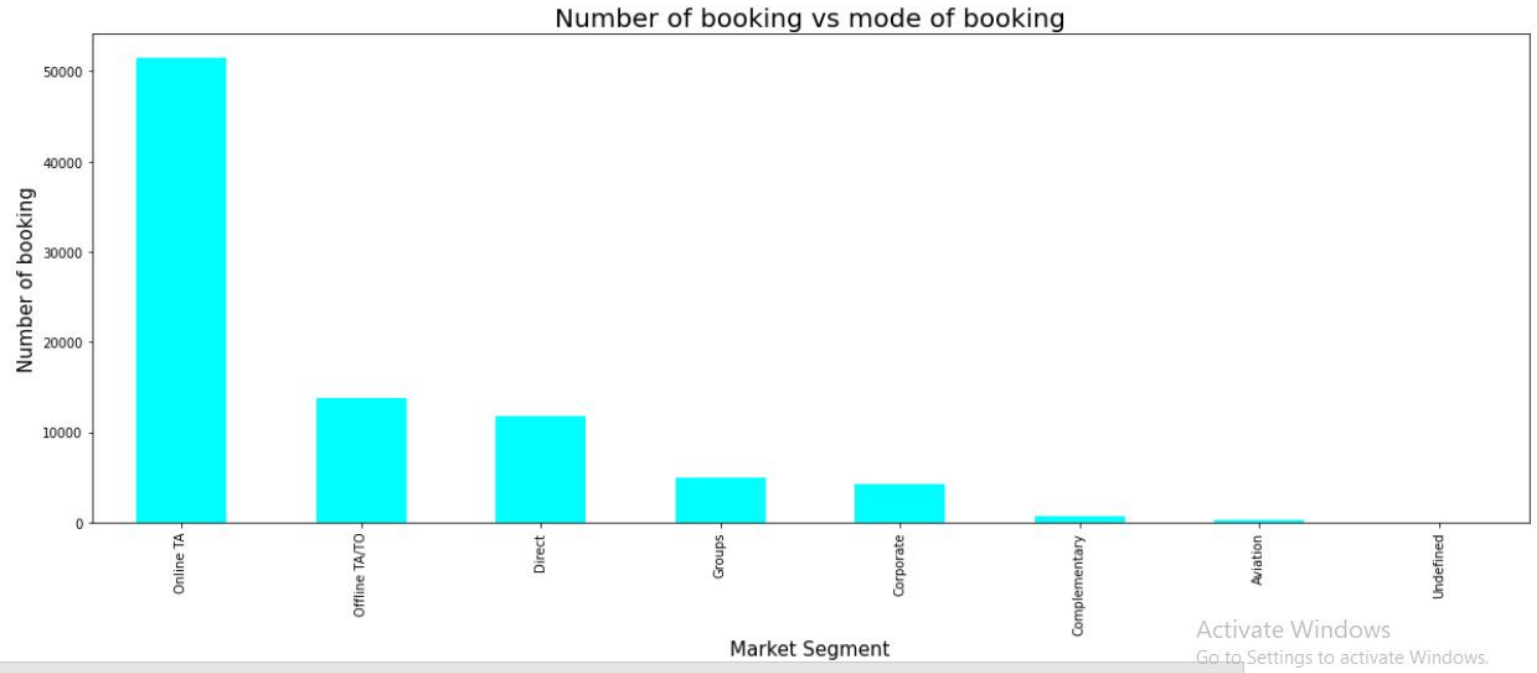
Most common stay length is less than 4 days and generally people prefer City hotel for short stay, but for long stays, Resort Hotel is preferred.



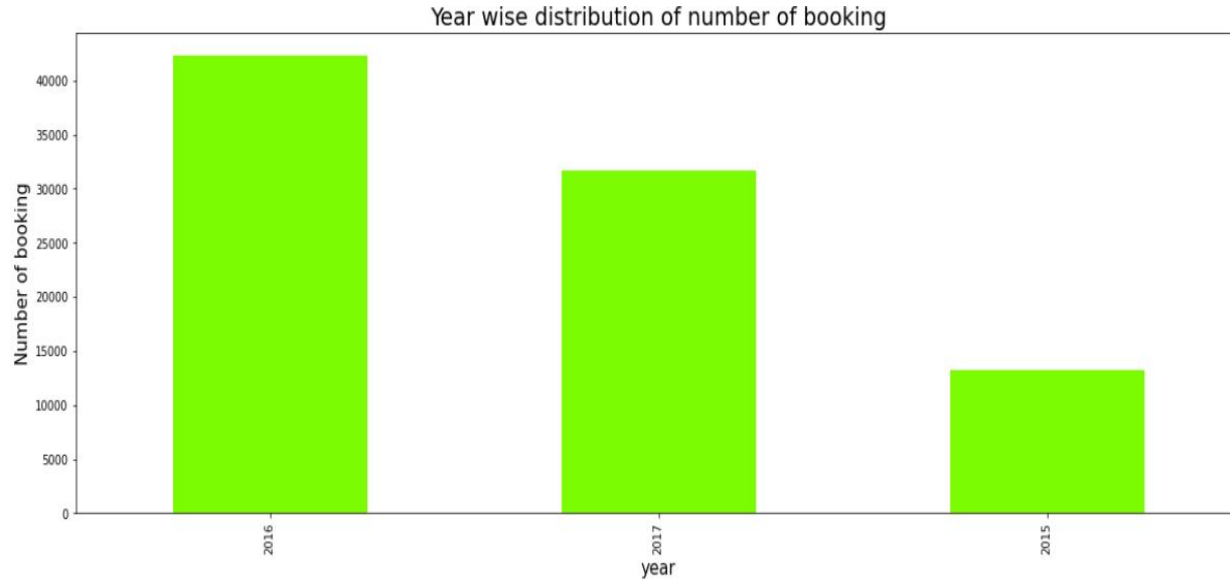
Almost 30 % of City Hotel bookings got canceled.



Most demanded room type is A, but better adr rooms are of type H, G and C also. Hotels should increase the no. of room types A and H to maximize revenue.



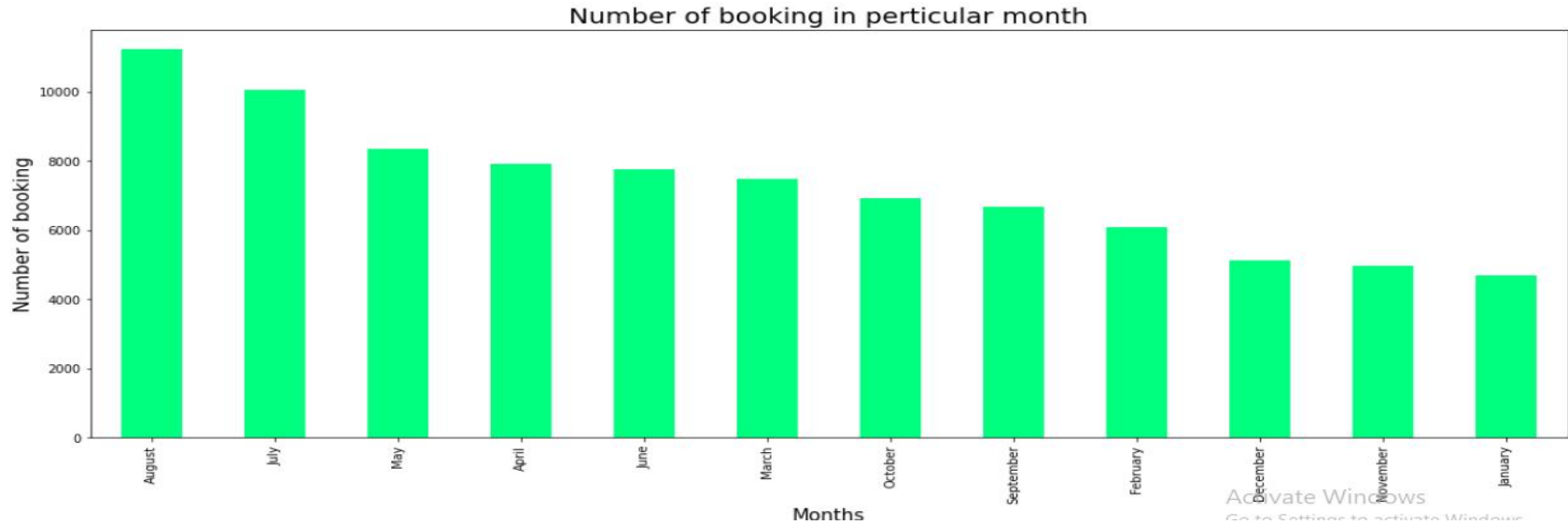
From above graph we can say that Online booking has almost same as sum of other mode of bookings.



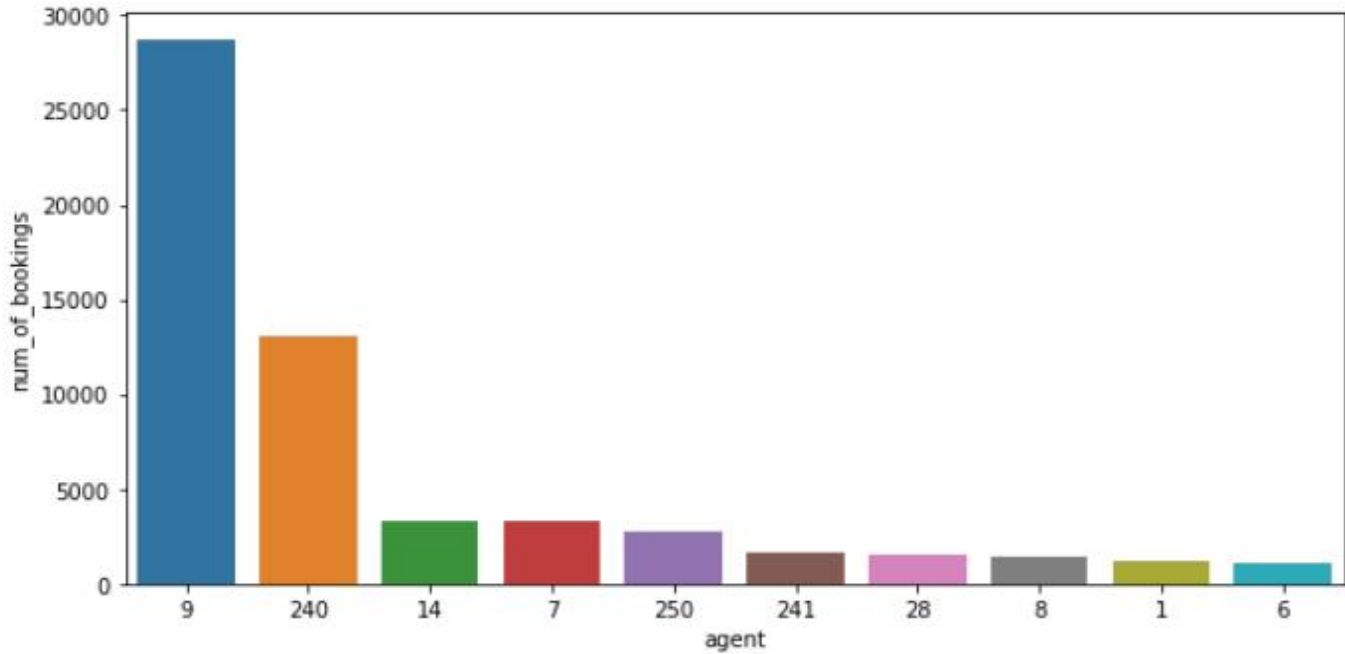
2016 has higher bookings than 2015 and 2017.

From 2015 number of bookings increases in 2016 but booking decreases in 2017.

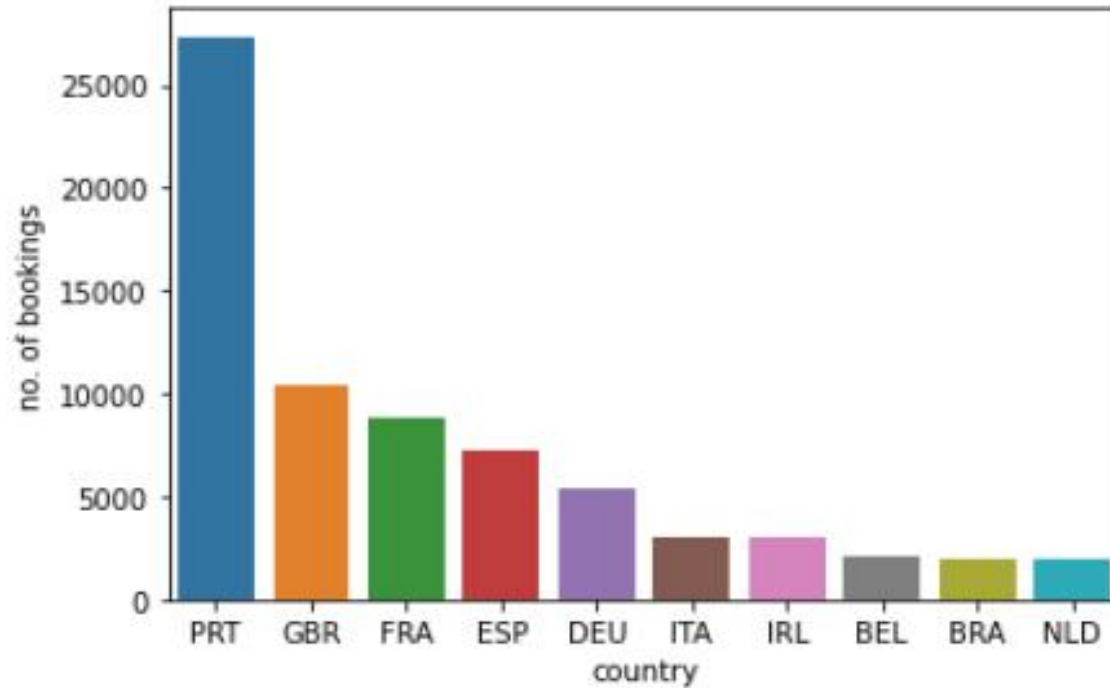
2016 has higher bookings than 2015 and 2017.



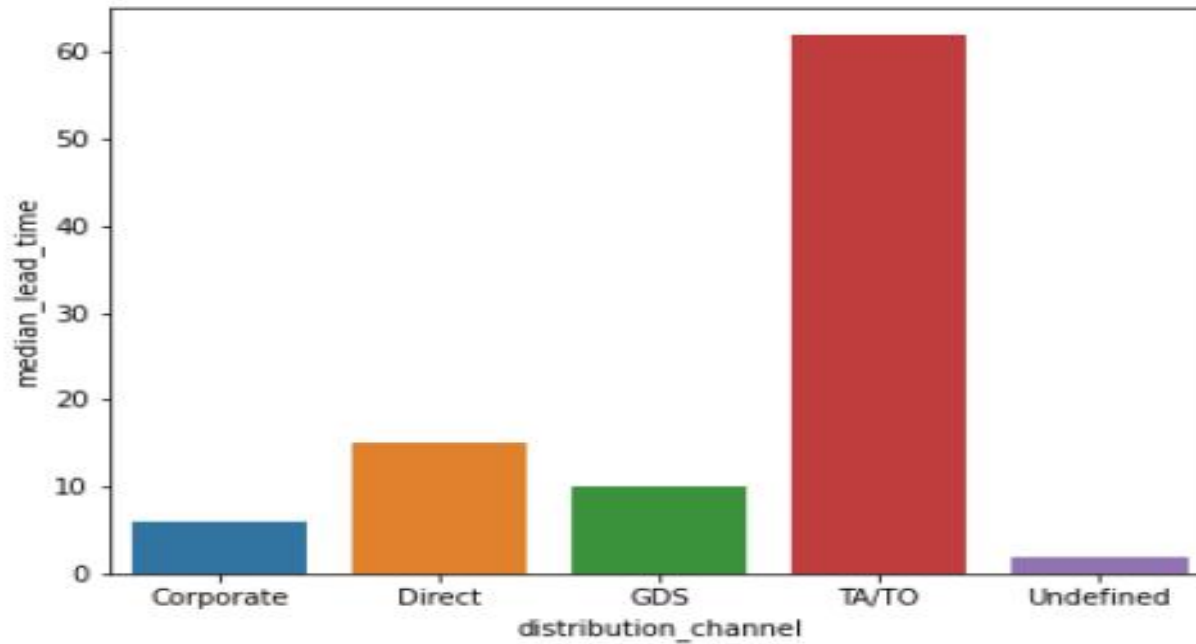
August month has highest number of booking and january month has lowest number of booking.



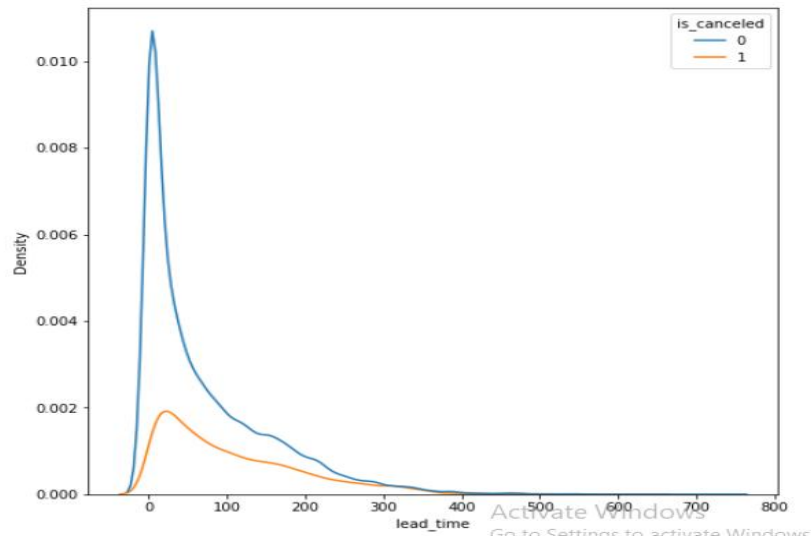
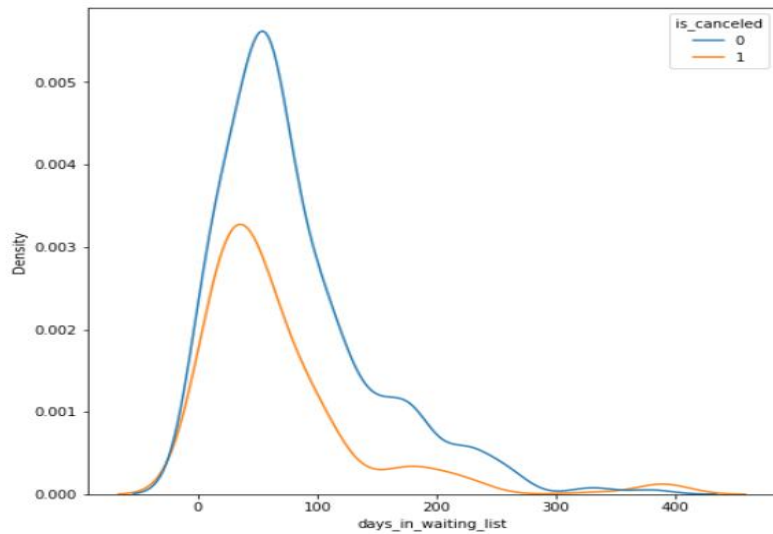
Agent no. 9 has made most no. of bookings



Most of the customers come from Portugal, Great Britain, France and Spain.

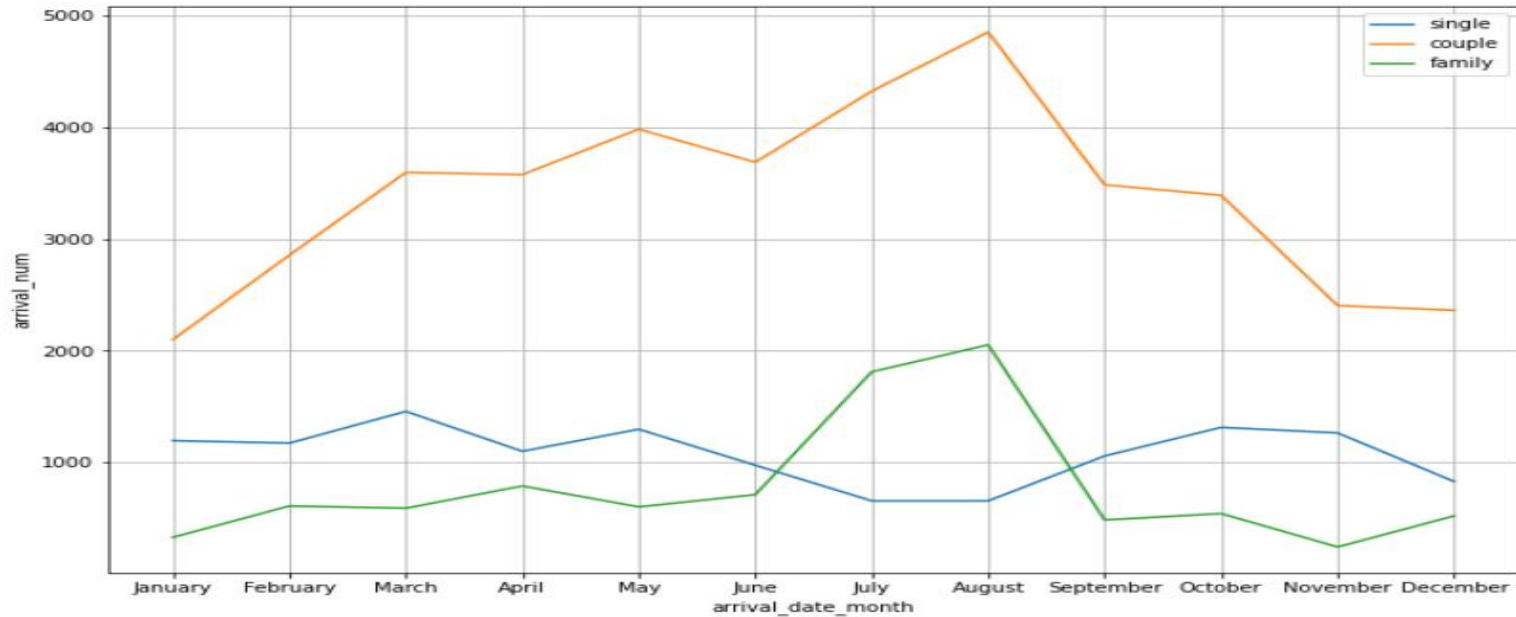


TA/TO is mostly used for planning Hotel visits ahead of time.
But for sudden visits other mediums are most preferred.



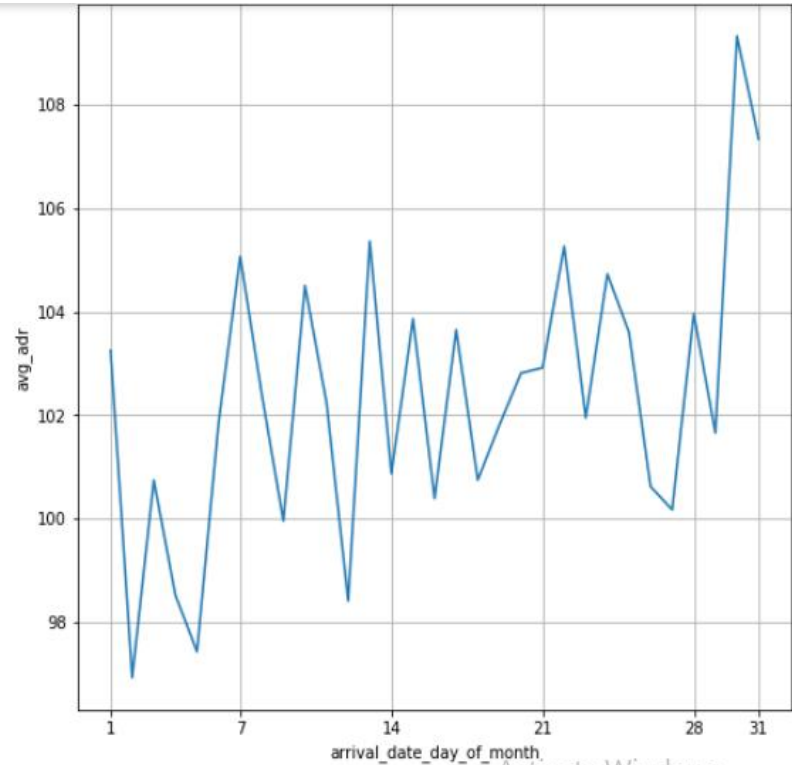
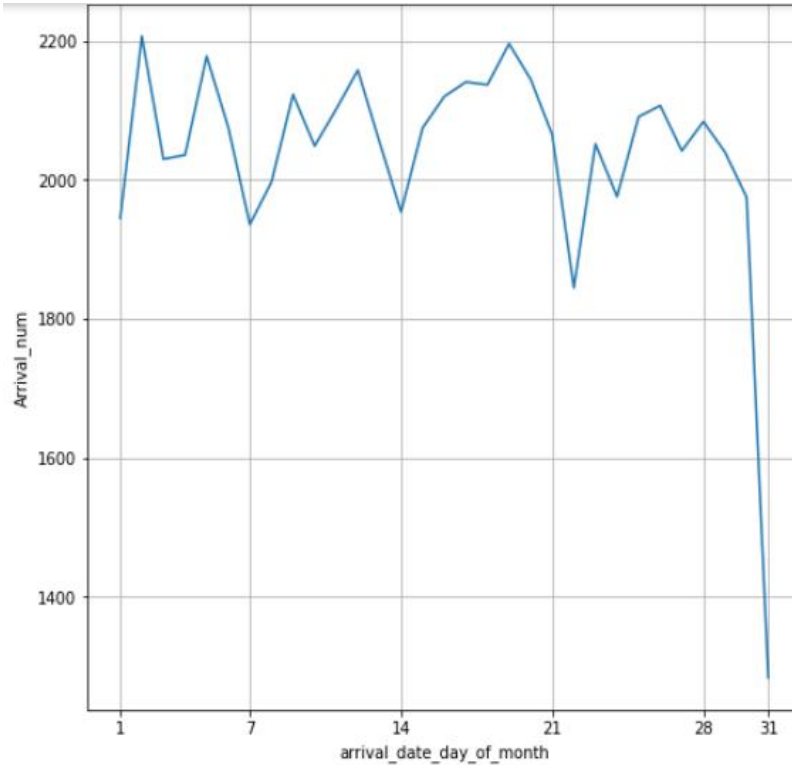
Most of the bookings that are cancelled have waiting period of less 150 days but also most of bookings that are not cancelled also have waiting period of less than 150 days. Hence this shows that waiting period has no effect on cancellation of bookings.

Also, lead time has no effect on cancellation of bookings, as both curves of cancellation and not cancellation are similar for lead time too.

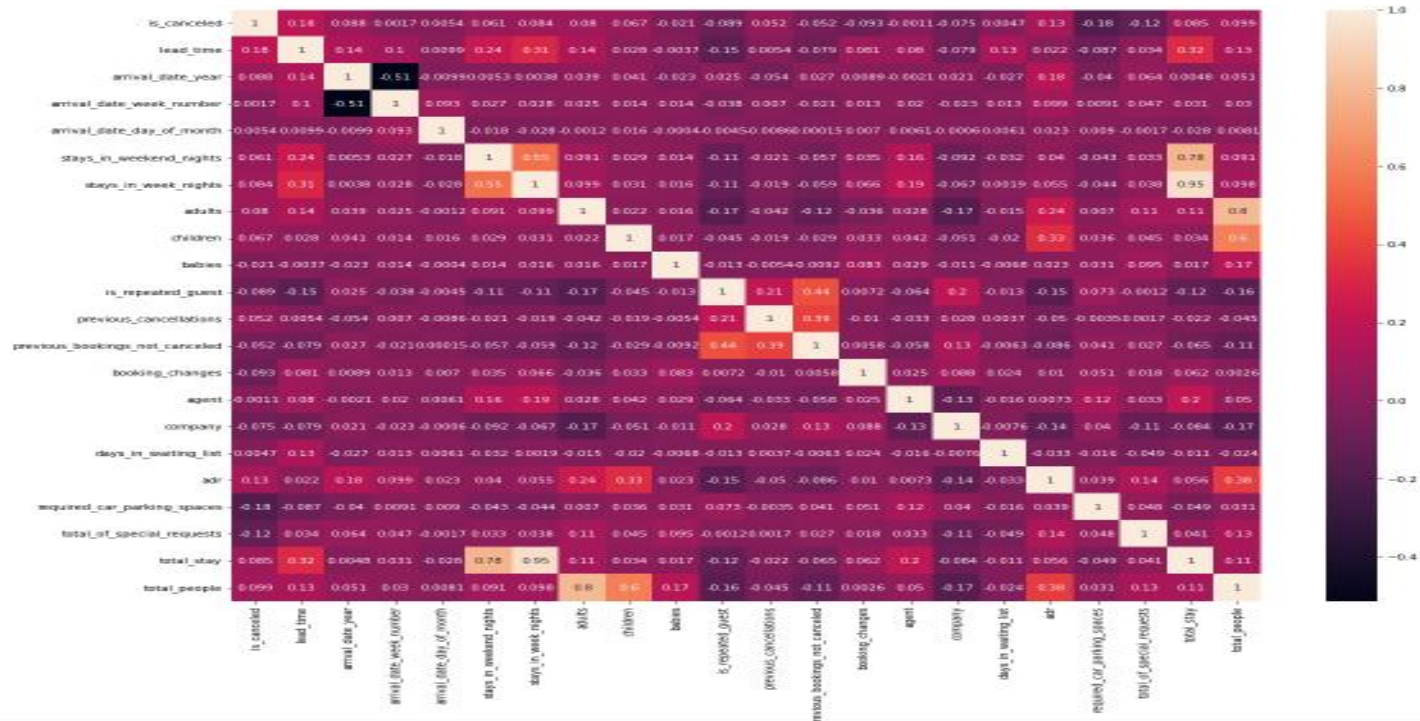


Mostly bookings are done by couples.

It is clear from graph that there is a sudden surge in arrival num of couples and family in months of July and August. So better plans can be planned accordingly at that time for these type of customers.



We can see that graph Arrival_num has small peaks at regular interval of days. This can be due to increase in arrival weekend. Also the avg adr tends to go up as month ends. Therefore charge more at the end of month.



It can be observed arrival_date_week_number and arrival_date_year are 54% negative correlated Company and agent are positive correlated by 35% adr and children are positive correlated by 32%

Challenges

- Data cleaning is the difficult thing to resolve in a short time.
- It was hard to find which graph technique to use.
- Removing null values.
- Analyzing the data and visualizing box plot and scatter plot made confused.

Conclusion

- 1. August month has highest number of booking and january month has lowest number of booking.
- 2. 2016 has higher bookings than 2015 and 2017.
- 3. From 2015 number of bookings increases in 2016 but booking decreases in 2017.
- 4. From above graph we can say that Online booking has almost same as sum of other mode of bookings.
- 5. Most demanded room type is A, but better adr rooms are of type H, G and C also. Hotels should increase the no. of room types A and H to maximise revenue.

Conclusion

- 5. Resort hotel has maximum cancellation than city hotel.
- 6. Agent no. 9 has made most no. of bookings.
- 7. Most of the customers come from Portugal, Great Britain, France and Spain.
- 8. Most common stay length is less than 4 days and generally people prefer City hotel for short stay, but for long stays, Resort Hotel is preferred.





THANK

YOU !

