Capstone Project
Hotel Booking Analysis

Koushik Das
M Anirudh Goud
Rupini Gandhaveeti
Chand Kamble

Points to Discuss:

- Agenda
- Data summary
- Univariate analysis
- Hotel wise analysis
- Distribution Channel wise analysis
- Timewise analysis
- Correlation heatmap
- Conclusion

Agenda

To discuss the analysis of given hotel bookings data set from 2015-2017.

We'll be doing analysis of given data set in following ways :

- Univariate analysis
- Hotel wise analysis
- Distribution Channel wise analysis
- Timewise analysis

By doing this we'll try to find out key factors driving the hotel bookings trends.

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.

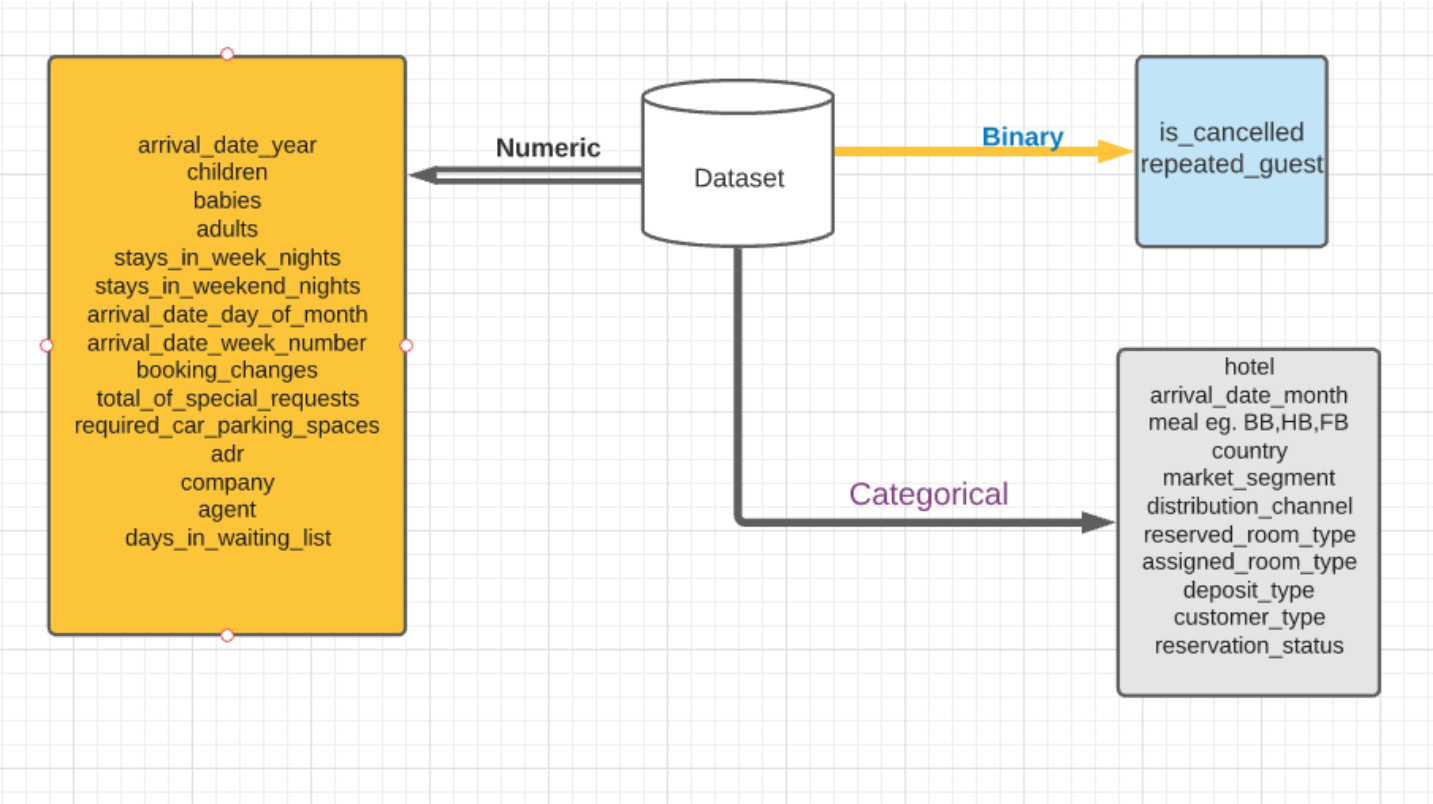
distribution_channel: The medium through booking was made.[Direct,Corporate,TA/TO,undefined,GDS.]

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 Summary



Univariate Analysis

EDA

❑Exploratory data analysis is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

❑EDA is primarily used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them.

In this project we will try to give answer of some real life problems by analyzing and visualization the given dataset.

Hotel wise Analysis

While doing hotel-wise analysis of given hotel booking dataset, we answered following questions:

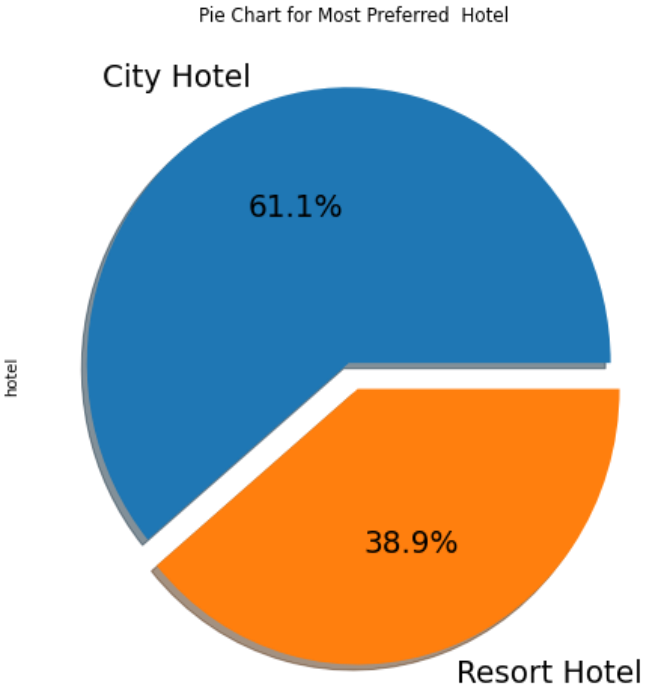
(1) Percentage of bookings in each hotels?

(2) Which hotel makes more revenue?

(3) For which hotel, does people have to wait longer to get a booking confirmed?

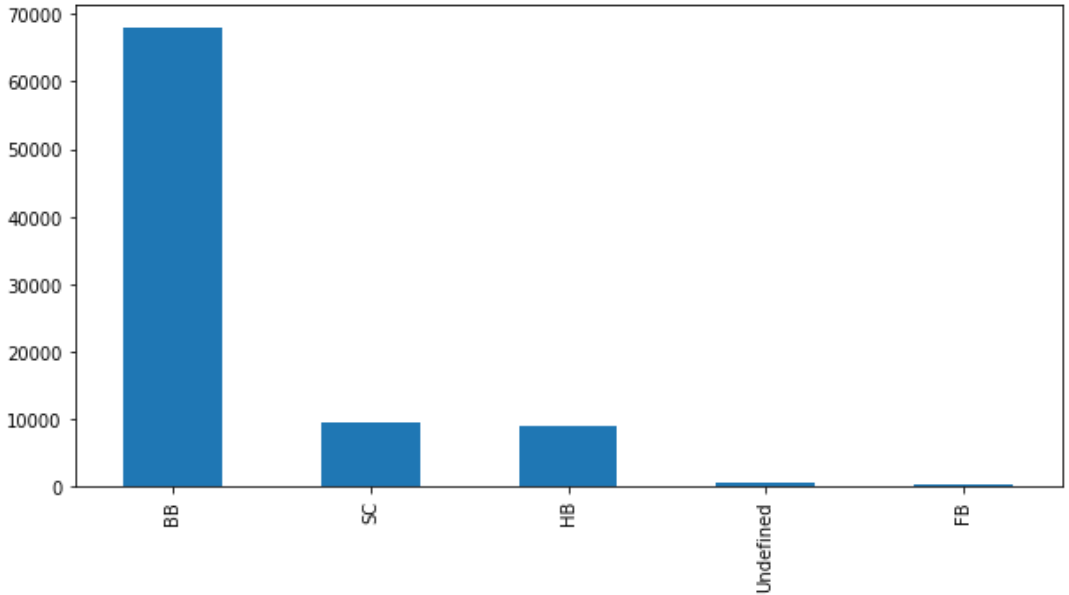
(4) Which hotel has higher booking cancellations rate?

Which type of hotel is mostly preferred?



- Around 60% bookings are for City hotel and 40% bookings are for Resort hotel.
- We can see that City Hotels are more preferred.

Which type of meal is mostly preferred?

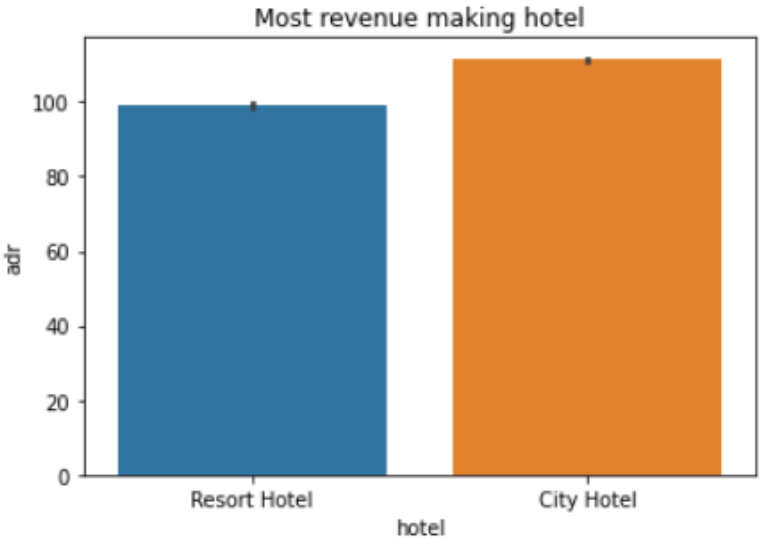


Types of meal in the hotels:

- ❑BB - (Bed and Breakfast)
- ❑HB- (Half Board)
- ❑FB- (Full Board)
- ❑SC- (Self Catering)

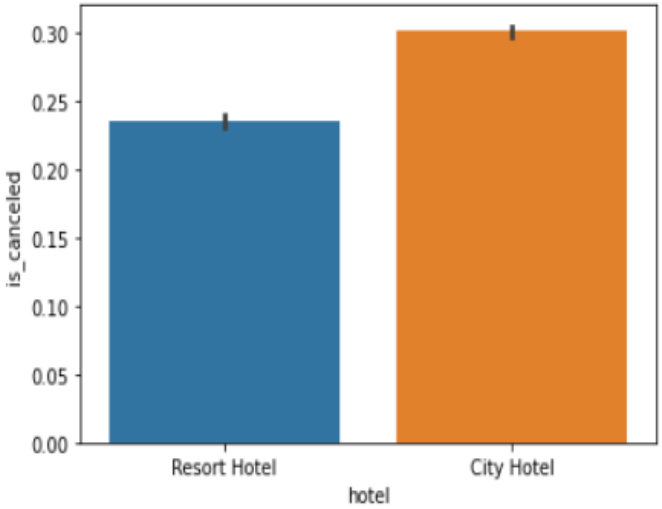
So the most preferred meal type by the guests is BB(Bed and Breakfast).

Which types of hotel makes more revenue?



❑ We can see City Hotel makes more revenue as compare to Resort Hotel.

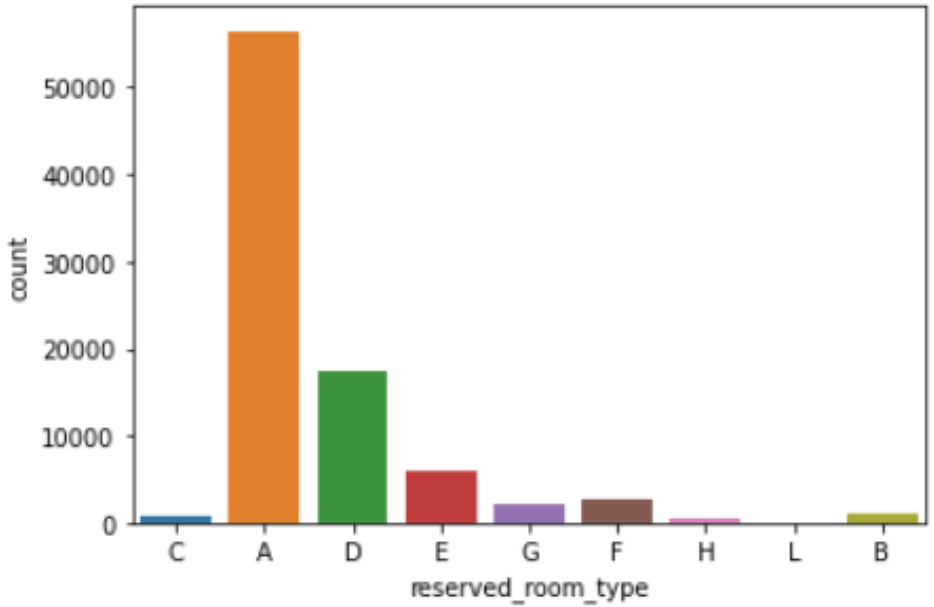
Which types of hotel has higher booking cancelation rate?



❑ So, City Hotel has the higher cancelation rate.

❑ Almost 30 % of City Hotel bookings and 25 % of Resort hotel bookings got canceled.

Which types of rooms are most reserved?



We can see ‘A’ type of rooms are most reserved.

Distribution channel wise Analysis

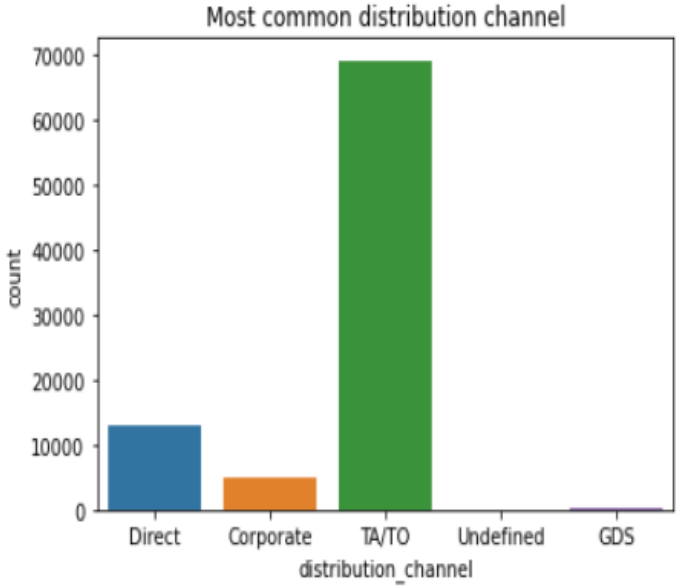
While doing Distribution channel wise analysis of given hotel booking dataset, we answered following questions:

(1) Which is the most common channel for booking hotels?

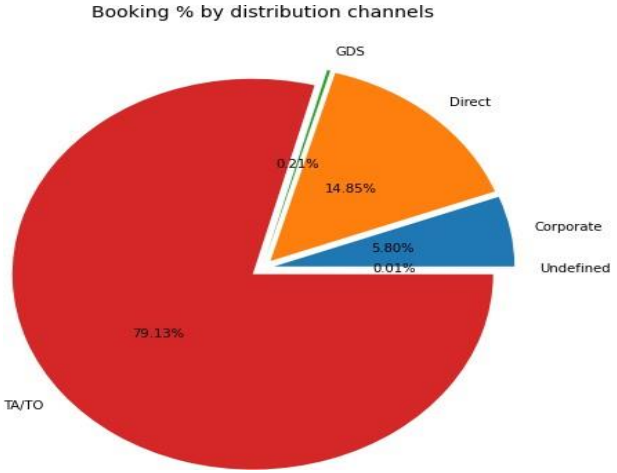
(2) Which type of market segment is mostly preferred by peoples?

(3) Most no of days in waiting list for different market segments.

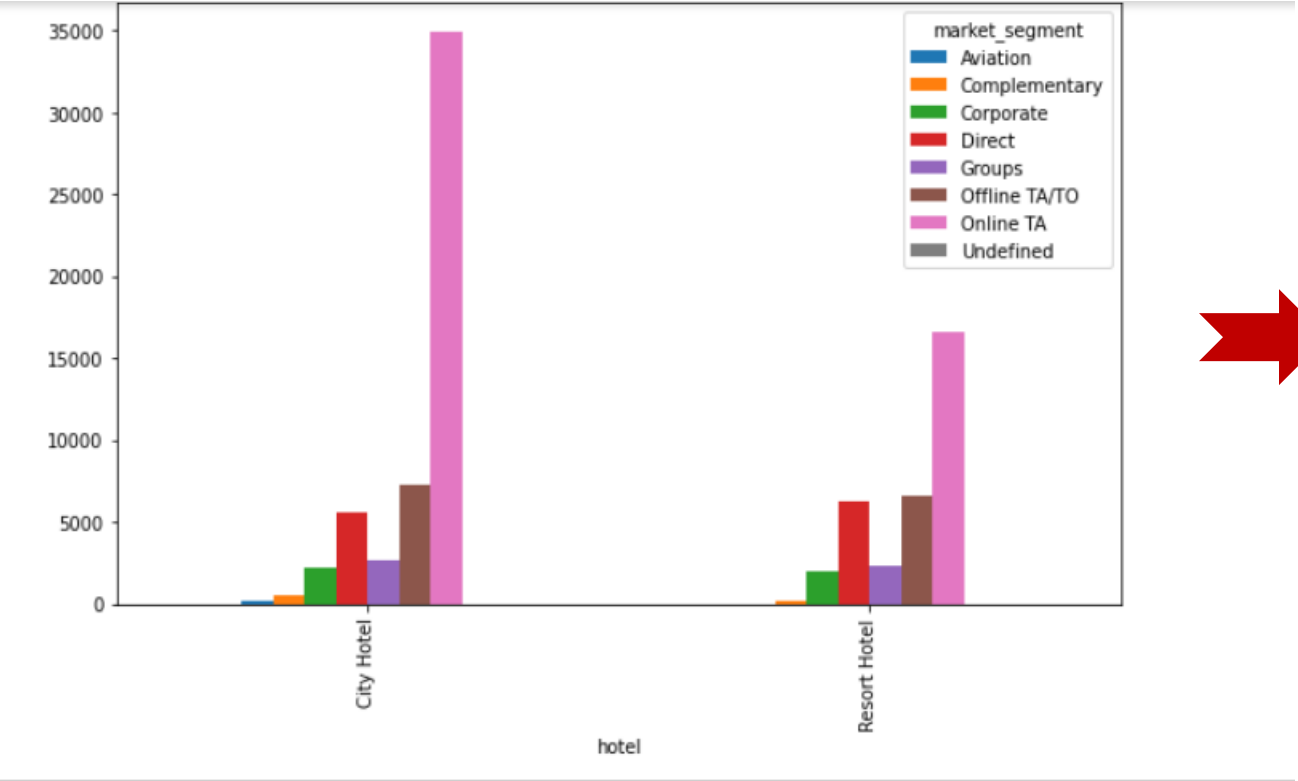
Which is the most common channel for booking hotels?



- We see that TA/TO is the most common distribution channel.
- Than the second most used channel is direct.

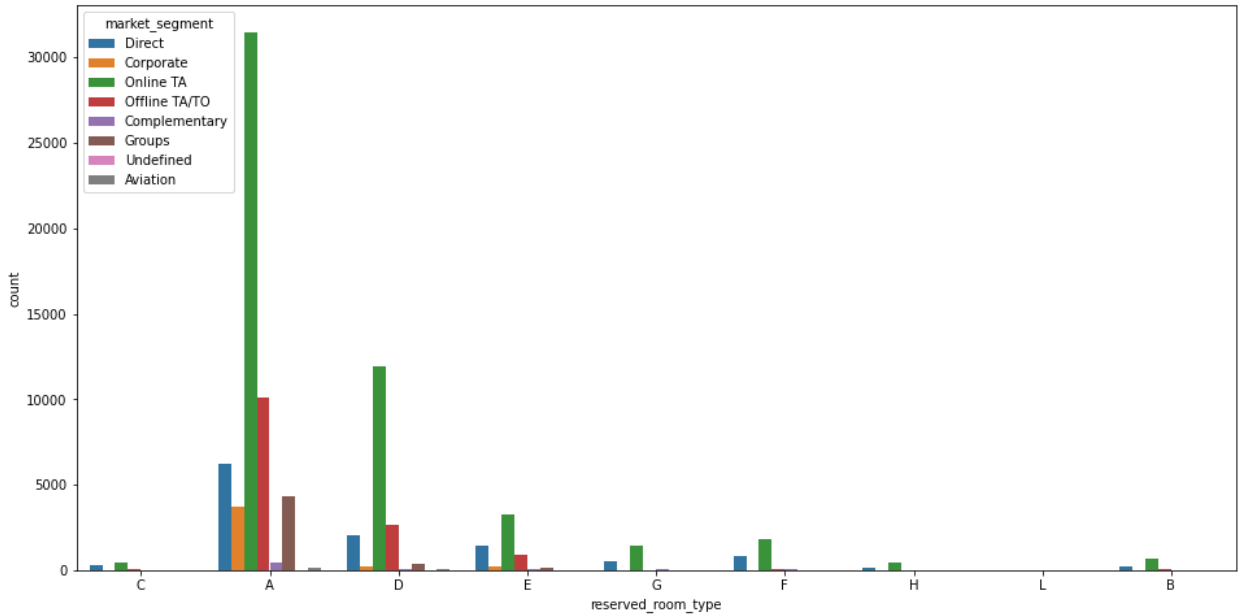


Market segment VS Hotel



❑ In each types of hotels, peoples preferred online TA.

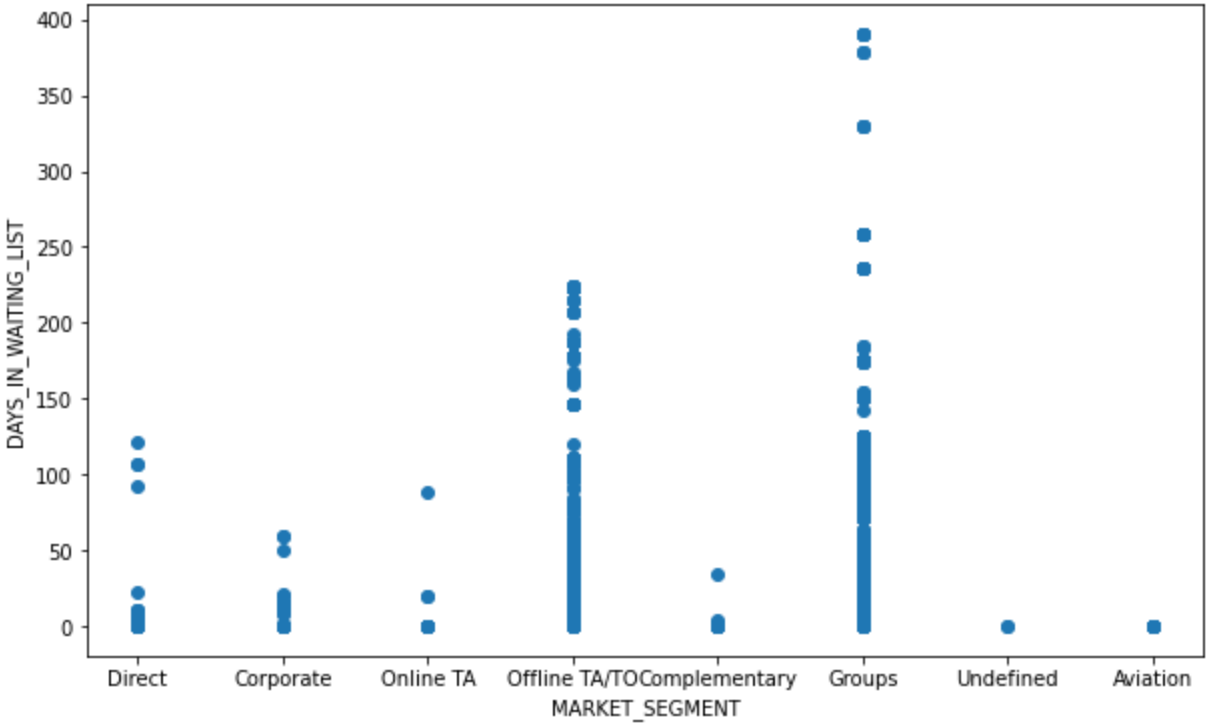
Market segment VS preferred room types



Majority of peoples reserved rooms via online TA from different market segments.

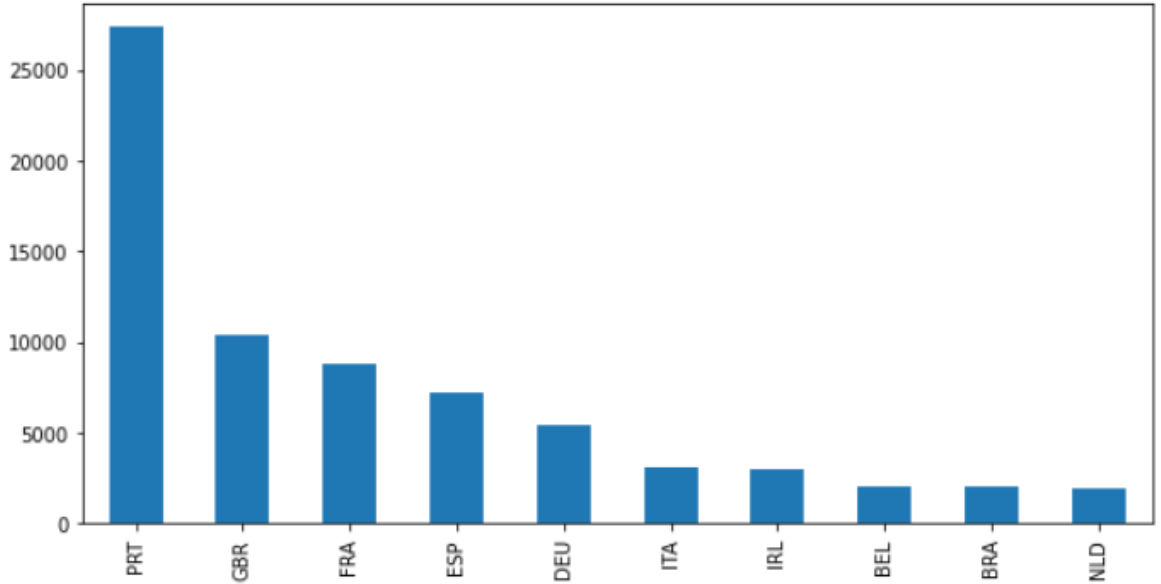
Also we can see majority of the people preferred room type A.

Most no of days in waiting list for different market segments.



Most no of days in waiting list are for the Groups market segments.

From which country most of the peoples are coming in the hotel ?



PRT	27355
GBR	10424
FRA	8823
ESP	7244
DEU	5385
ITA	3061
IRL	3015
BEL	2081
BRA	1993
NLD	1910

We see that most of the guests are come from Portugal, Great Britain, France and Spain.

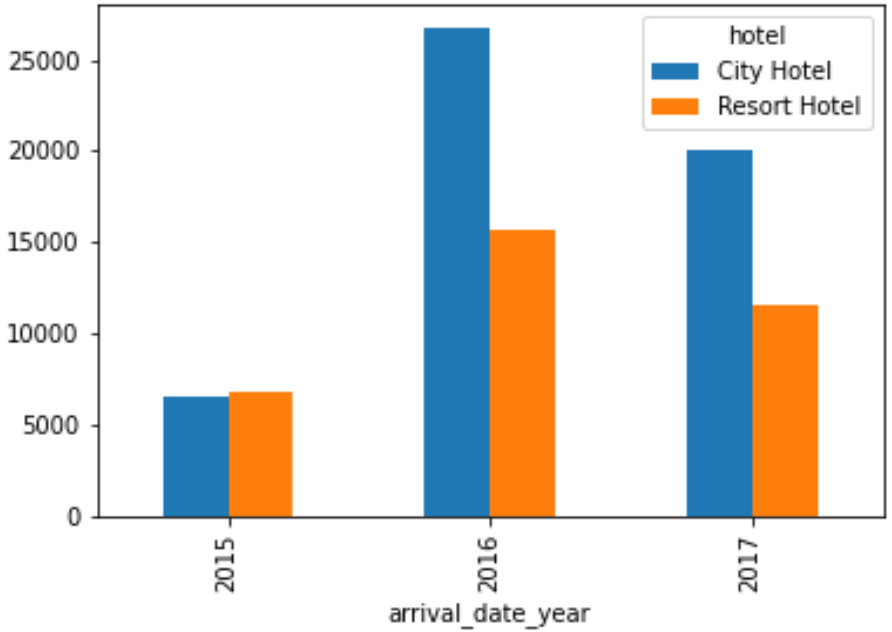
Time-wise Analysis

While doing time-wise analysis of given hotel booking dataset, we answered following questions:

(1) What are the most busy months for hotels?

(2) How does bookings varies along with the year?

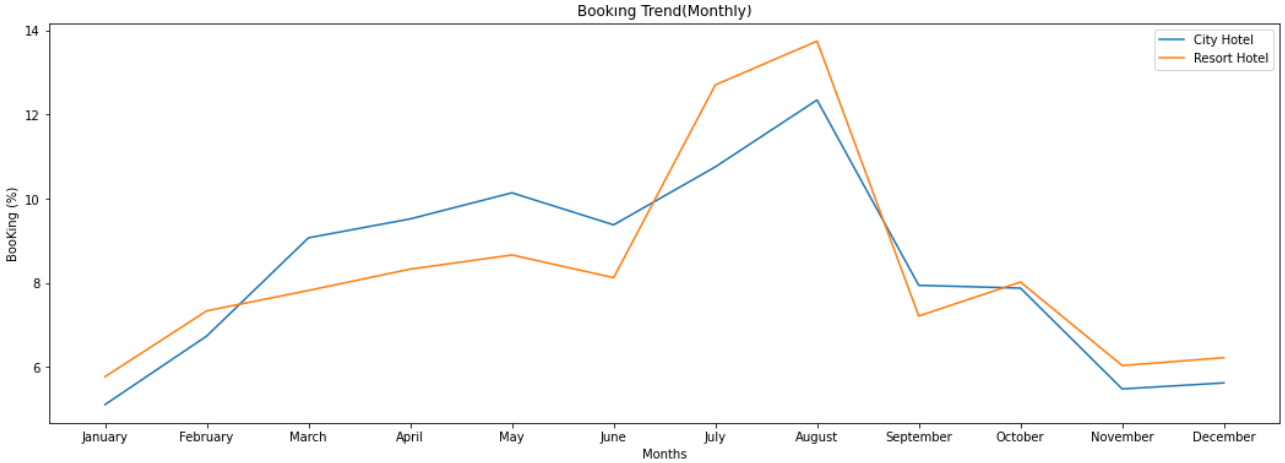
(3) In which month most number of peoples arrive?

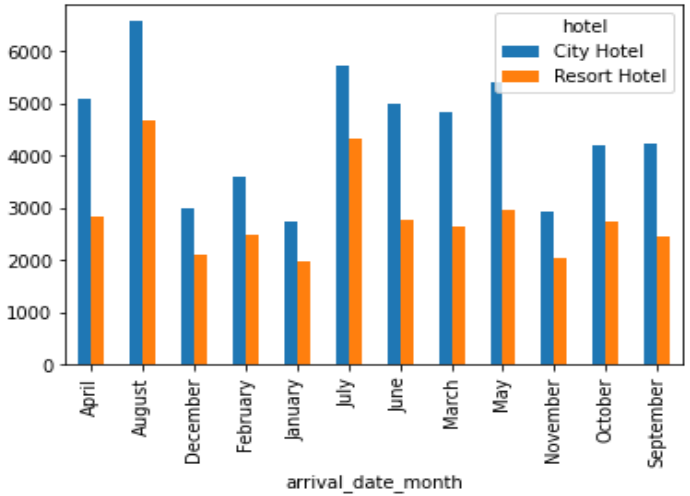


•2016 had the highest booking.

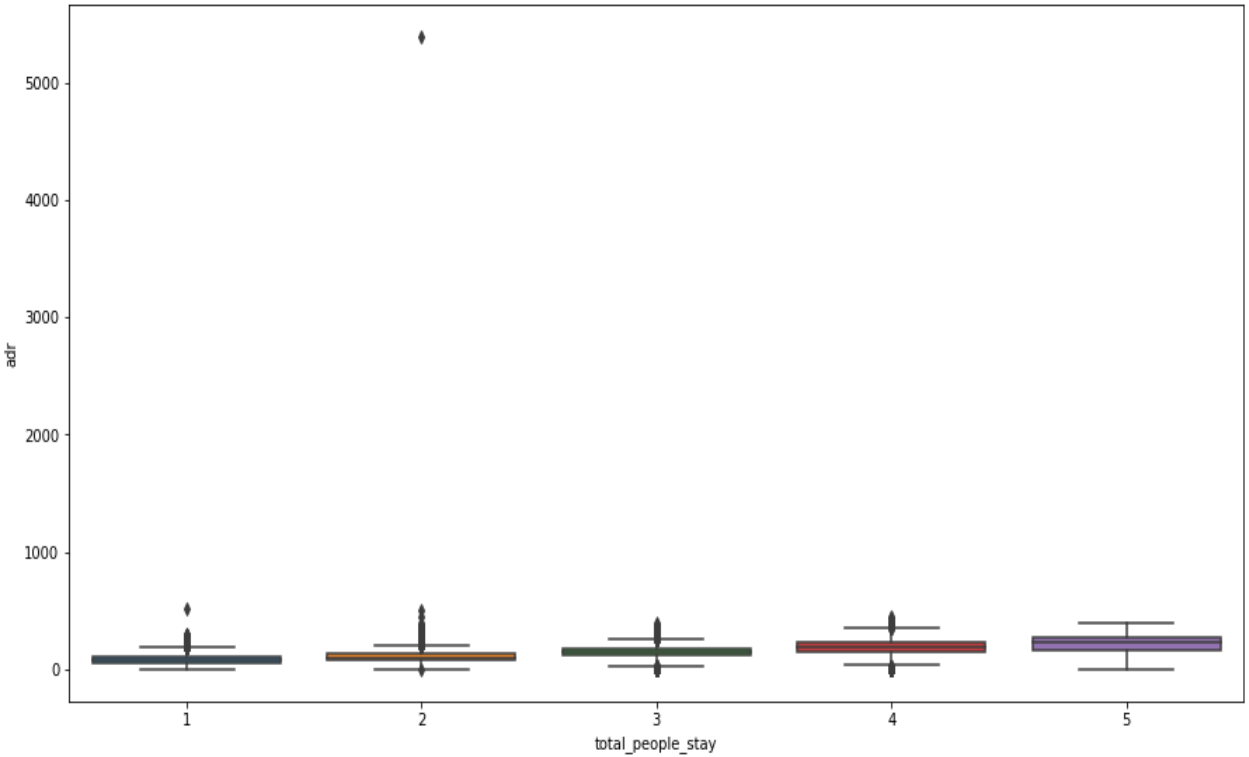
•2015 had lowest bookings.

•overall City hotels had the most of the bookings.



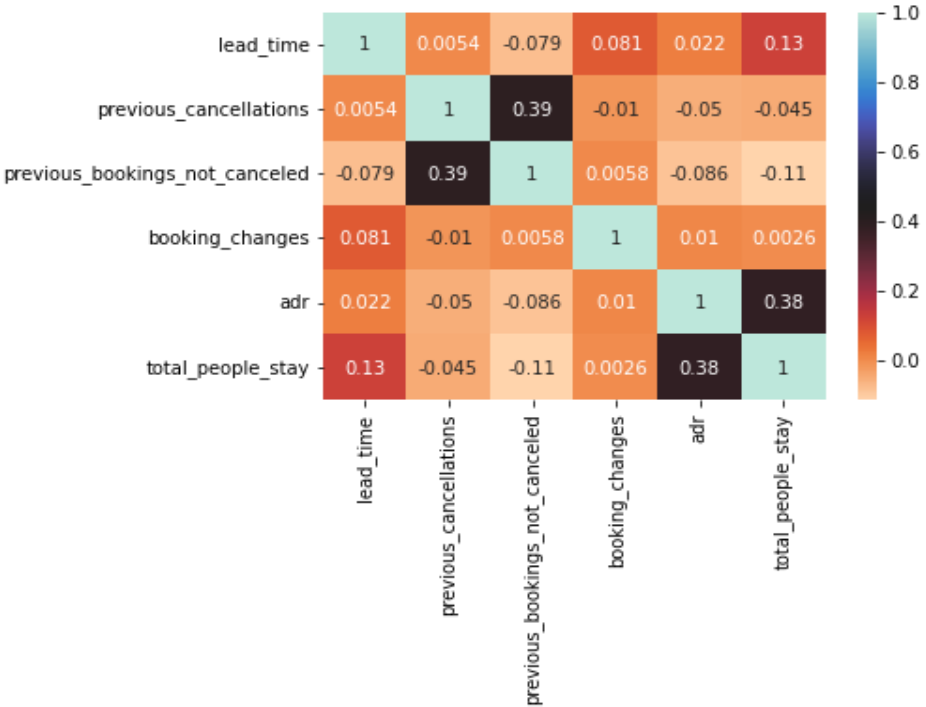


- July and August months had the most Bookings.



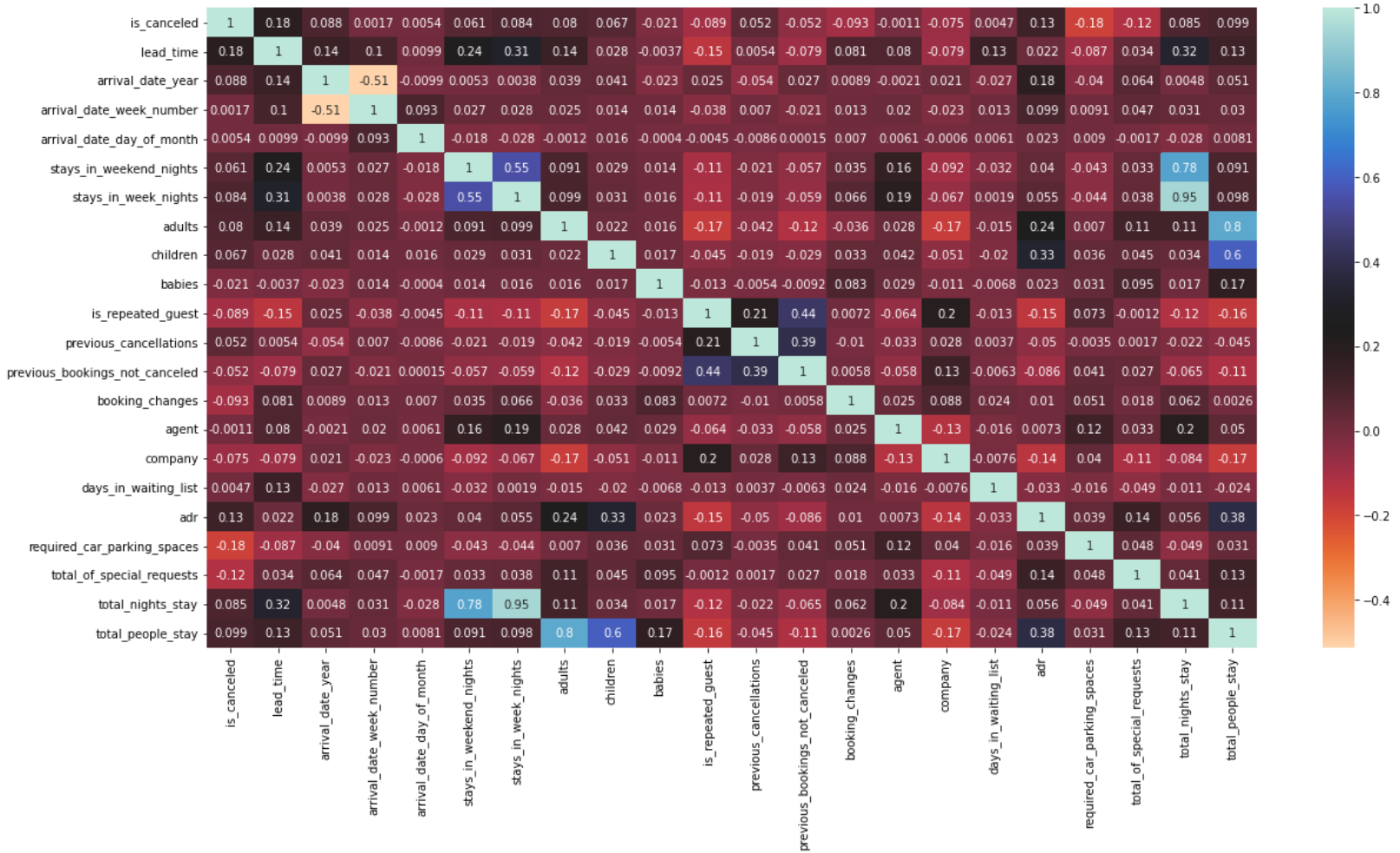
- From the Box plot we can observe that as the total number of people increases adr also increases.

Correlation Heatmap



❑Total no of peoples stay in the hotel and lead time have slightly positive correlation.

- ❑ adr is slightly correlated with total_no of People stay in the hotel, which makes sense as more no of people means more revenue, therefore more adr.


is_canceled 1 0.18 0.088 0.0017 0.0054 0.061 0.084 0.08 0.067 -0.021 -0.089 0.052 -0.052 -0.093 -0.0011 -0.075 0.0047 0.13 -0.18 -0.12 0.085 0.099
lead_time 0.18 1 0.14 0.1 0.0099 0.24 0.31 0.14 0.028 -0.0037 -0.15 0.0054 -0.079 0.081 0.08 -0.079 0.13 0.022 -0.087 0.034 0.32 0.13
arrival_date_year 0.088 0.14 1 -0.51 -0.0099 0.0053 0.0038 0.039 0.041 -0.023 0.025 -0.054 0.027 0.0089 -0.0021 0.021 -0.027 0.18 -0.04 0.064 0.0048 0.051
arrival_date_week_number 0.0017 0.1 -0.51 1 0.093 0.027 0.028 0.025 0.014 0.014 -0.038 0.007 -0.021 0.013 0.02 -0.023 0.013 0.099 0.0091 0.047 0.031 0.03
arrival_date_day_of_month 0.0054 0.0099 -0.0099 0.093 1 -0.018 -0.028 -0.0012 0.016 -0.0004 -0.0045 -0.0086 0.00015 0.007 0.0061 -0.0006 0.0061 0.023 0.009 -0.0017 -0.028 0.0081
stays_in_weekend_nights 0.061 0.24 0.0053 0.027 -0.018 1 0.55 0.091 0.029 0.014 -0.11 -0.021 -0.057 0.035 0.16 -0.092 -0.032 0.04 -0.043 0.033 0.78 0.091
stays_in_week_nights 0.084 0.31 0.0038 0.028 -0.028 0.55 1 0.099 0.031 0.016 -0.11 -0.019 -0.059 0.066 0.19 -0.067 0.0019 0.055 -0.044 0.038 0.95 0.098
adults 0.08 0.14 0.039 0.025 -0.0012 0.091 0.099 1 0.022 0.016 -0.17 -0.042 -0.12 -0.036 0.028 -0.17 -0.015 0.24 0.007 0.11 0.11 0.8
children 0.067 0.028 0.041 0.014 0.016 0.029 0.031 0.022 1 0.017 -0.045 -0.019 -0.029 0.033 0.042 -0.051 -0.02 0.33 0.036 0.045 0.034 0.6
babies -0.021 -0.0037 -0.023 0.014 -0.0004 0.014 0.016 0.016 0.017 1 -0.013 -0.0054 -0.0092 0.083 0.029 -0.011 -0.0068 0.023 0.031 0.095 0.017 0.17
is_repeated_guest -0.089 -0.15 0.025 -0.038 -0.0045 -0.11 -0.11 -0.17 -0.045 -0.013 1 0.21 0.44 0.0072 -0.064 0.2 -0.013 -0.15 0.073 -0.0012 -0.12 -0.16
previous_cancellations 0.052 0.0054 -0.054 0.007 -0.0086 -0.021 -0.019 -0.042 -0.019 -0.0054 0.21 1 0.39 -0.01 -0.033 0.028 0.0037 -0.05 -0.0035 0.0017 -0.022 -0.045
previous_bookings_not_canceled -0.052 -0.079 0.027 -0.021 0.00015 -0.057 -0.059 -0.12 -0.029 -0.0092 0.44 0.39 1 0.0058 -0.058 0.13 -0.0063 -0.086 0.041 0.027 -0.065 -0.11
booking_changes -0.093 0.081 0.0089 0.013 0.007 0.035 0.066 -0.036 0.033 0.083 0.0072 -0.01 0.0058 1 0.025 0.088 0.024 0.01 0.051 0.018 0.062 0.0026
agent -0.0011 0.08 -0.0021 0.02 0.0061 0.16 0.19 0.028 0.042 0.029 -0.064 -0.033 -0.058 0.025 1 -0.13 -0.016 0.0073 0.12 0.033 0.2 0.05
company -0.075 -0.079 0.021 -0.023 -0.0006 -0.092 -0.067 -0.17 -0.051 -0.011 0.2 0.028 0.13 0.088 -0.13 1 -0.0076 -0.14 0.04 -0.11 -0.084 -0.17
days_in_waiting_list 0.0047 0.13 -0.027 0.013 0.0061 -0.032 0.0019 -0.015 -0.02 -0.0068 -0.013 0.0037 -0.0063 0.024 -0.016 -0.0076 1 -0.033 -0.016 -0.049 -0.011 -0.024
adr 0.13 0.022 0.18 0.099 0.023 0.04 0.055 0.24 0.33 0.023 -0.15 -0.05 -0.086 0.01 0.0073 -0.14 -0.033 1 0.039 0.14 0.056 0.38
required_car_parking_spaces -0.18 -0.087 -0.04 0.0091 0.009 -0.043 -0.044 0.007 0.036 0.031 0.073 -0.0035 0.041 0.051 0.12 0.04 -0.016 0.039 1 0.048 -0.049 0.031
total_of_special_requests -0.12 0.034 0.064 0.047 -0.0017 0.033 0.038 0.11 0.045 0.095 -0.0012 0.0017 0.027 0.018 0.033 -0.11 -0.049 0.14 0.048 1 0.041 0.13
total_nights_stay 0.085 0.32 0.0048 0.031 -0.028 0.78 0.95 0.11 0.034 0.017 -0.12 -0.022 -0.065 0.062 0.2 -0.084 -0.011 0.056 -0.049 0.041 1 0.11
total_people_stay 0.099 0.13 0.051 0.03 0.0081 0.091 0.098 0.8 0.6 0.17 -0.16 -0.045 -0.11 0.0026 0.05 -0.17 -0.024 0.38 0.031 0.13 0.11 1
is_canceled
lead_time
arrival_date_year
arrival_date_week_number
arrival_date_day_of_month
stays_in_weekend_nights
stays_in_week_nights
adults
children
babies
is_repeated_guest
previous_cancellations
previous_bookings_not_canceled
booking_changes
agent
company
days_in_waiting_list
adr
required_car_parking_spaces
total_of_special_requests
total_nights_stay
total_people_stay
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Conclusion

- Around 60% bookings are for City hotel and 40% bookings are for Resort hotel, therefore City Hotel is busier
  than Resort hotel. Also the overall adr of City hotel is slightly higher than Resort hotel.
  BB types meal is most preferred by the guest.
- City hotels have significantly higher booking cancellation rates
- Most of the guests came from european countries, with most no. of guest coming from Portugal.
- Guests use different channels for making bookings out of which most preferred way is TA/TO.
- July- August are the most busier and profitable months for both of hotels.
- 2016 had the highest bookings.
- adr and the total number of peoples are positively correlated.

Thank You