Capstone Project

Hotel Booking Analysis

KAMATAM HARSHITH
LAKSHMI KEERTHANA
TITO VARGHESE
ANMOL RAJ

Points to Discuss

- Agenda
- Data summary
- Data wrangling
- Hotel wise analysis
- Distribution Channel wise analysis
- Cancellation related analysis
- Time and Stay related analysis
- Conclusion

Agenda

To extract,observe and analyse the given hotel bookings data set from 2015-2017.

The analysis of given data set in following ways :

- Univariate analysis
- Hotel wise analysis
- Distribution Channel wise analysis
- Booking cancellation analysis
- Timewise analysis
- Conclusion

Data Summary

Given data set has different columns of variables crucial for hotel bookings:

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.

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 information

#	Column	Non-Null Count	Dtype
0	hotel	119390 non-null	object
1	is_canceled	119390 non-null	int64
2	lead_time	119390 non-null	int64
3	arrival_date_year	119390 non-null	int64
4	arrival_date_month	119390 non-null	object
5	arrival_date_week_number	119390 non-null	int64
6	arrival_date_day_of_month	119390 non-null	int64
7	stays_in_weekend_nights	119390 non-null	int64
8	stays_in_week_nights	119390 non-null	int64
9	adults	119390 non-null	int64
10	children	119386 non-null	float64
11	babies	119390 non-null	int64
12	meal	119390 non-null	object
13	country	118902 non-null	object
14	market_segment	119390 non-null	object
15	distribution_channel	119390 non-null	object
16	is_repeated_guest	119390 non-null	int64
17	previous_cancellations	119390 non-null	int64
18	previous_bookings_not_canceled	119390 non-null	int64
19	reserved_room_type	119390 non-null	object
20	assigned_room_type	119390 non-null	object
21	booking_changes	119390 non-null	int64
22	deposit_type	119390 non-null	object
23	agent	103050 non-null	float64
24	company	6797 non-null	float64
25	days_in_waiting_list	119390 non-null	int64
26	customer_type	119390 non-null	object
27	adr	119390 non-null	float64
28	required_car_parking_spaces	119390 non-null	int64
29	total_of_special_requests	119390 non-null	int64
30	reservation_status	119390 non-null	object
31	reservation_status_date	119390 non-null	object

Data Cleaning

Data Cleaning is a crucial step before EDA as it will remove the ambiguous data that can affect the outcome of EDA.

While cleaning data we will perform the following steps:

1) Remove duplicate rows (*df1[df1.duplicated()].shape)+df1.drop_duplicates(inplace = True)*

2) Handling missing values.

3) Convert columns to appropriate data types. (`df1[['children', 'company', 'agent']] = df1[['children', 'company', 'agent']].astype('int64'))`

Hotel wise analysis

- Hotel with higher bookings cancellation rate.
- Hotel with longest waiting time
- Hotel with most revenue.
- Chances of customer returning to hotel for another stay
- Factors Governing Booking
- Special requests by the guests

Distribution Channel wise analysis

- Most used Distribution Channel
- Which distribution channel brings better revenue generating deals for hotels?
- Market segments used by the guests
- Distribution Channel with highest cancellation

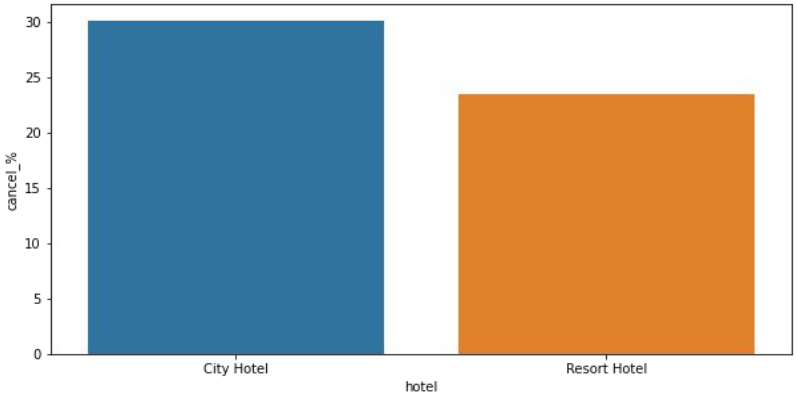
Cancellation related Analysis

- Waiting time(days)
- Lead Time
- Cancellation for not assigning same room
- Car parking space

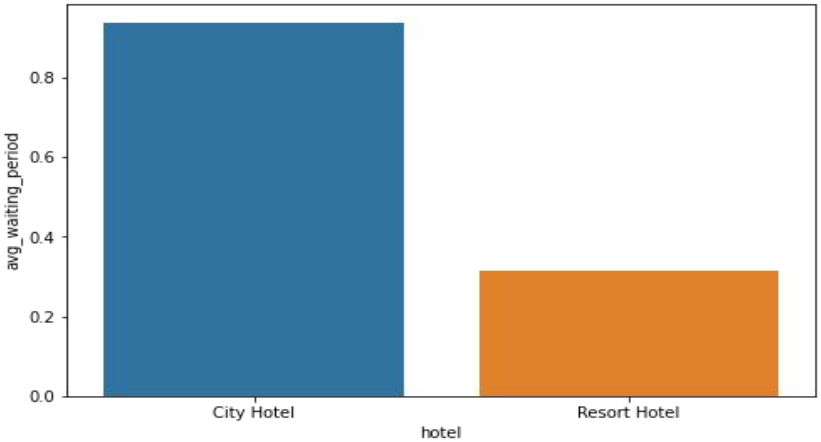
Time and Stay related Analysis

- Customer type with maximum Average Daily Rate
- Type of customers booking the most
- Heat correlation map Best time to book a hotel room
- Board Basis preferred
- Countries from which most customers are coming

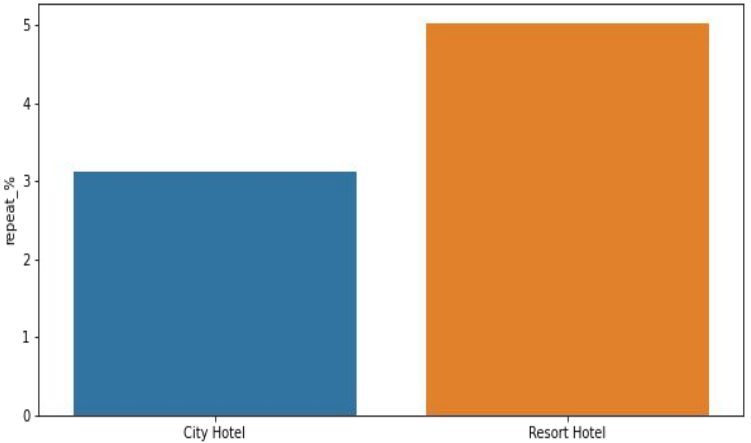
Hotel with higher bookings cancellation rate



Hotel with longest waiting time



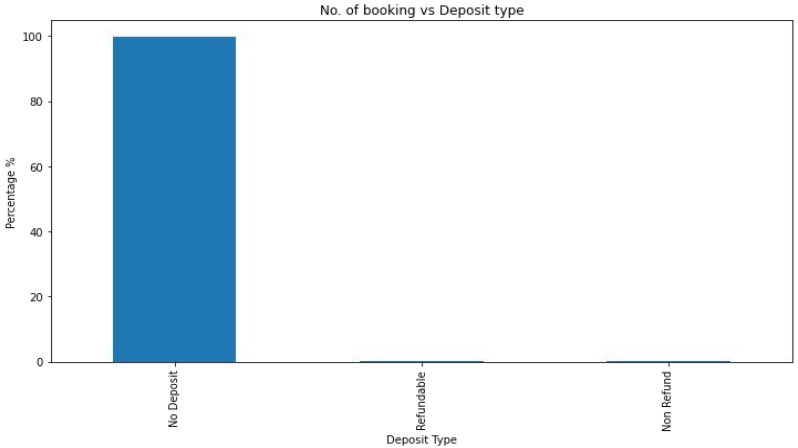
Hotel with most revenue



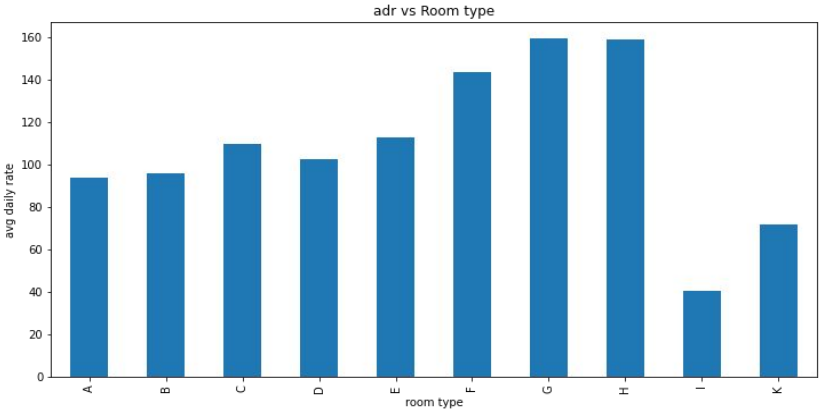
Chances of customer returning to hotel for another stay

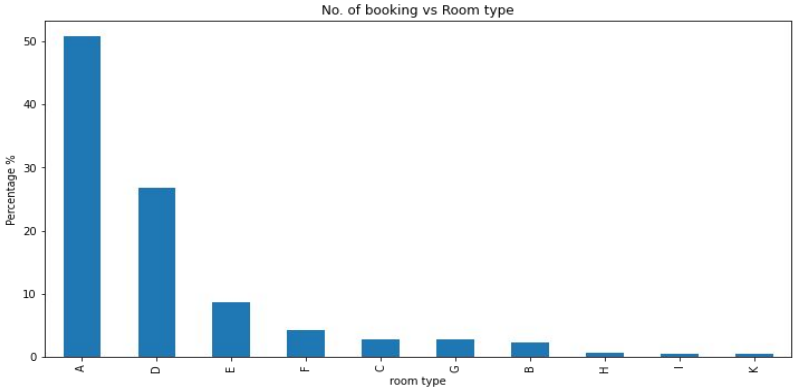
Factors governing booking

Deposit type

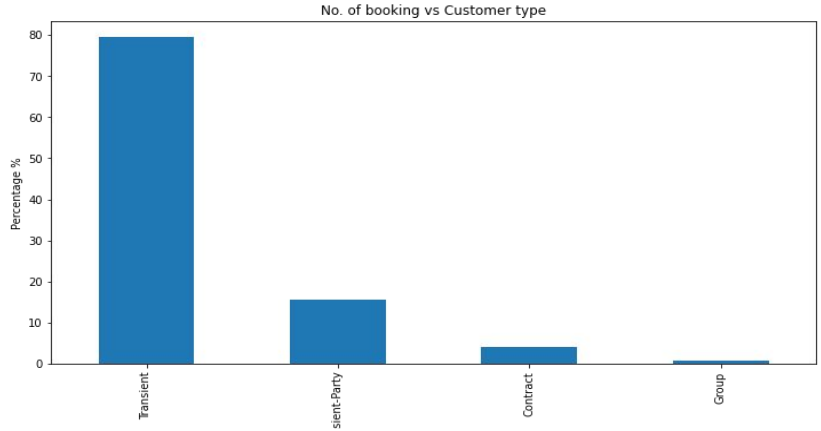


Room type assigned



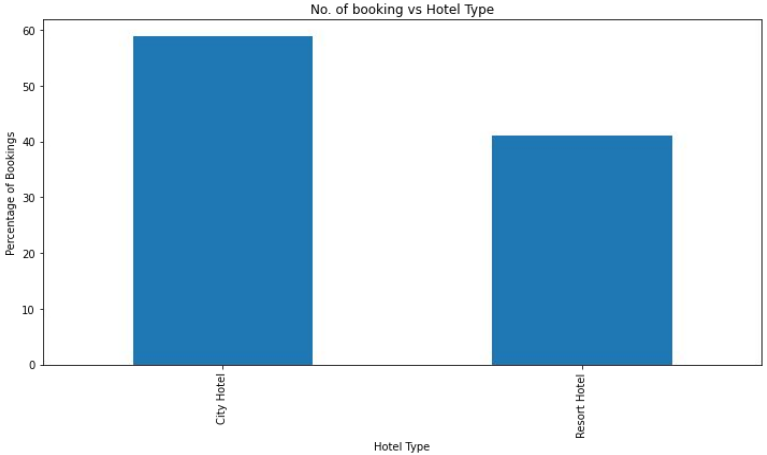


Room Type

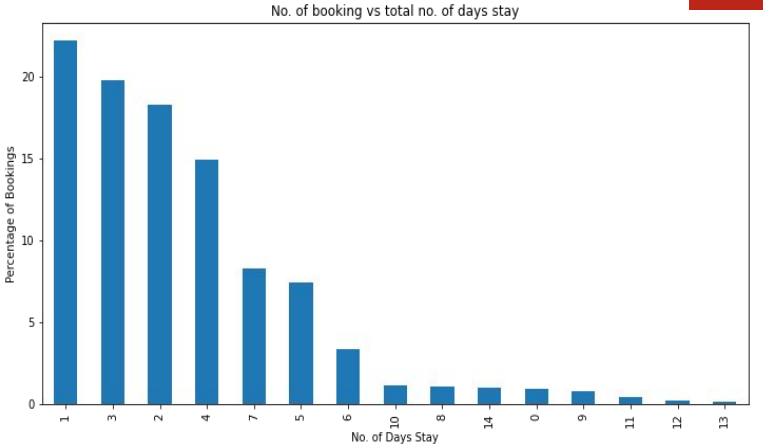


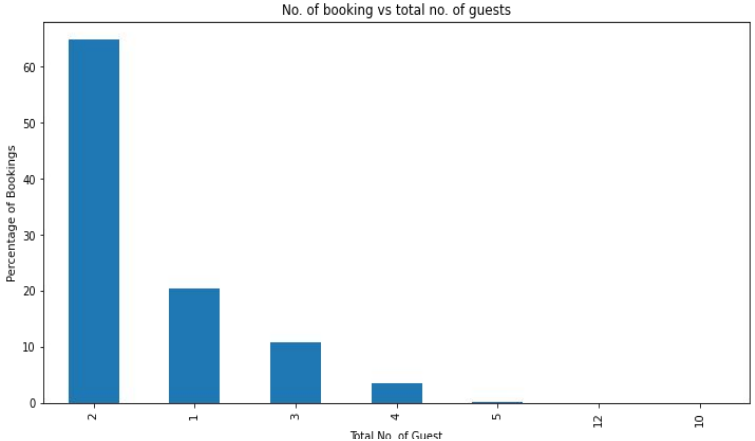
Customer Type

Hotel Type



Total number of days stays

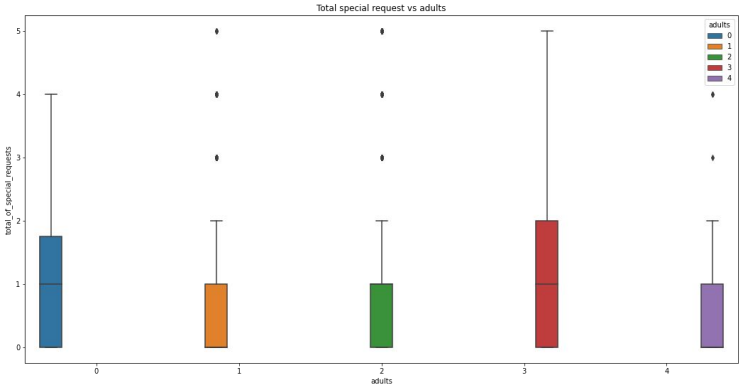




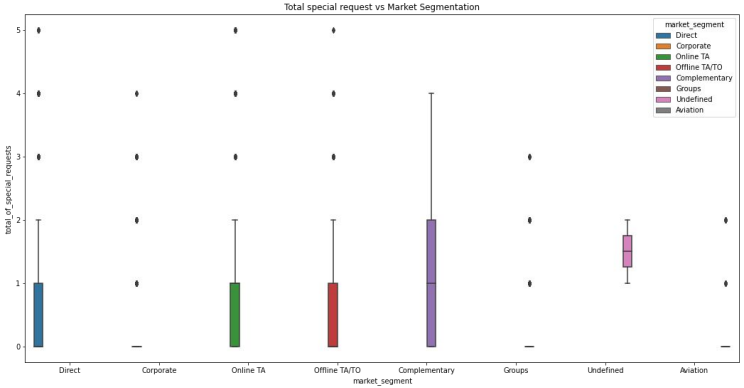
Total Number of Guest

Special requests by the guests

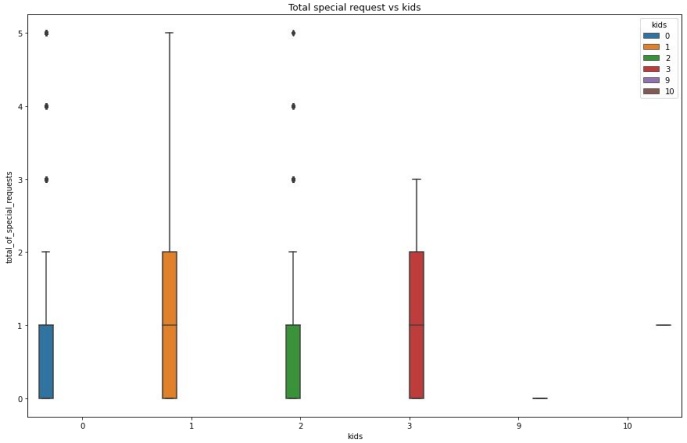
Special Requests According to Adults



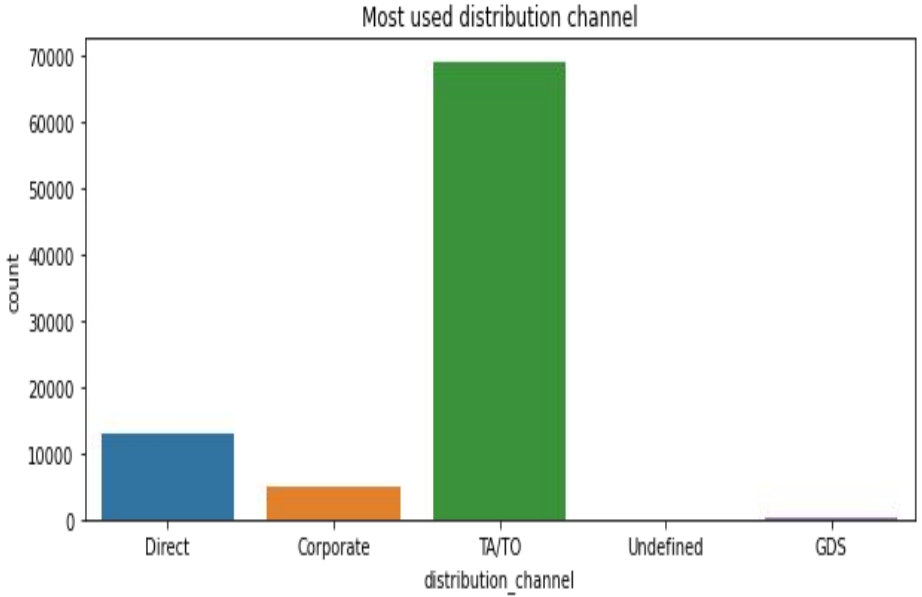
Special Requests According to Market Segmentation



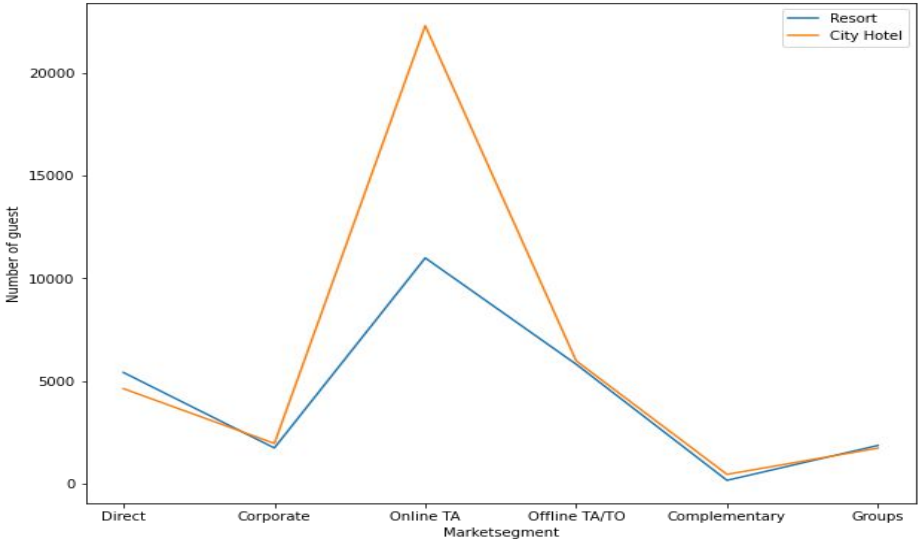
Special Requests According to Kids



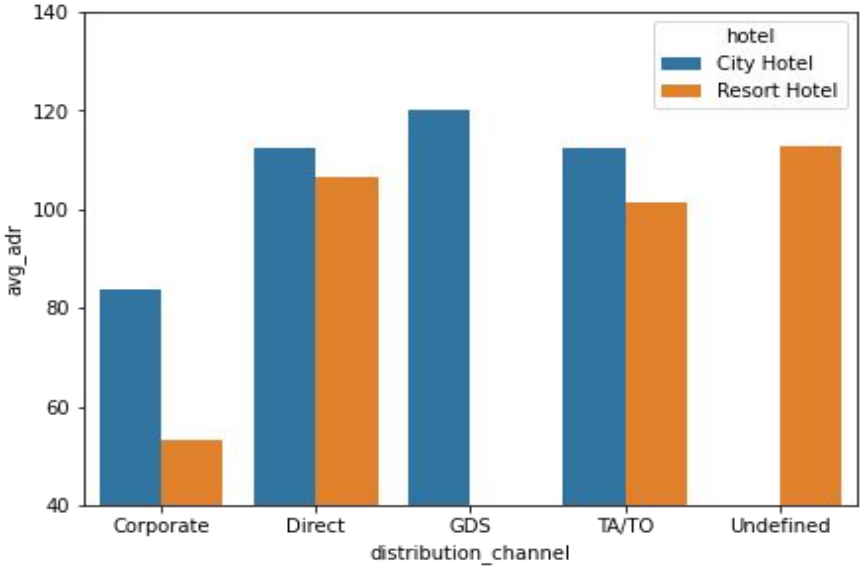
Most used distribution channel



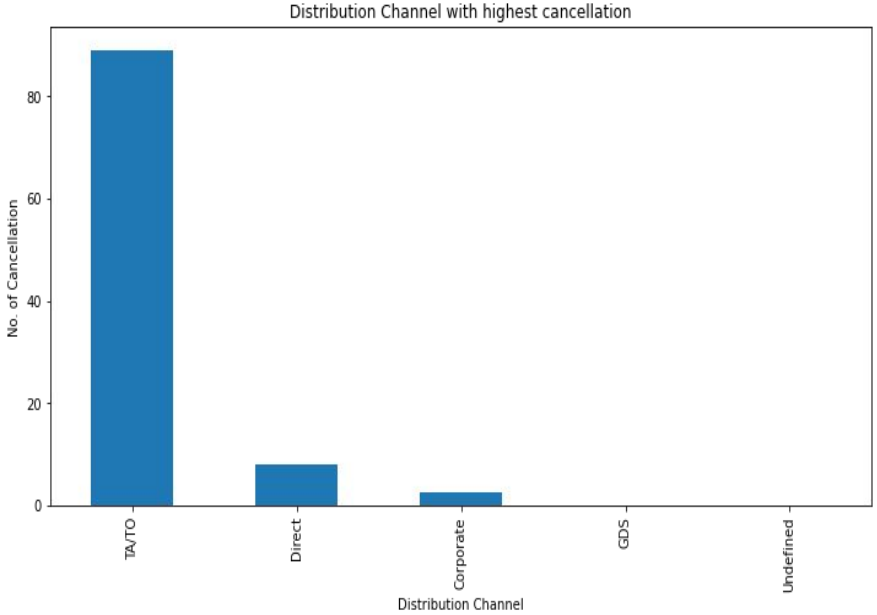
Most used market segment



Distribution channel bringing highest revenue

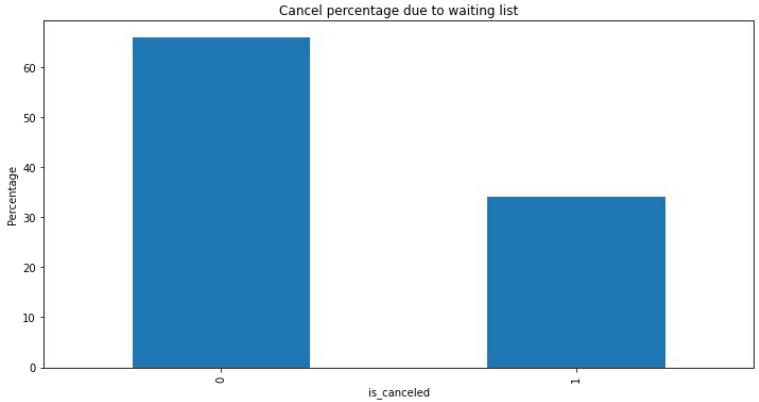


Distribution Channel with highest cancellation

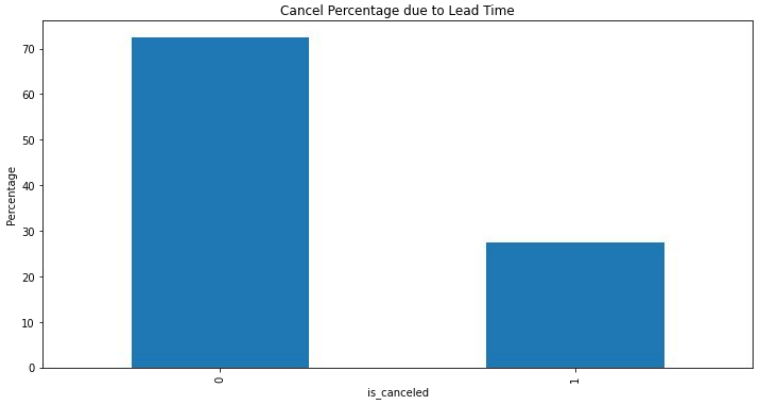


Cancellation related analysis

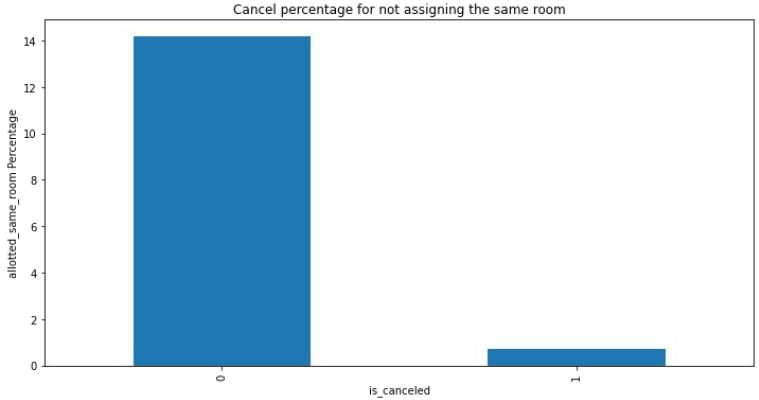
Waiting time(days)



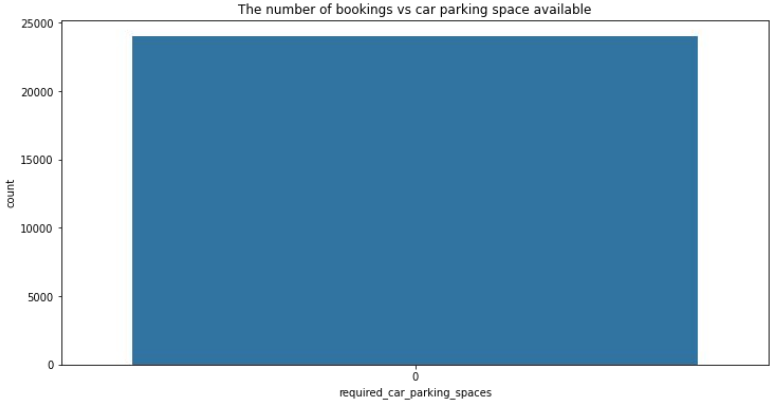
Lead Time



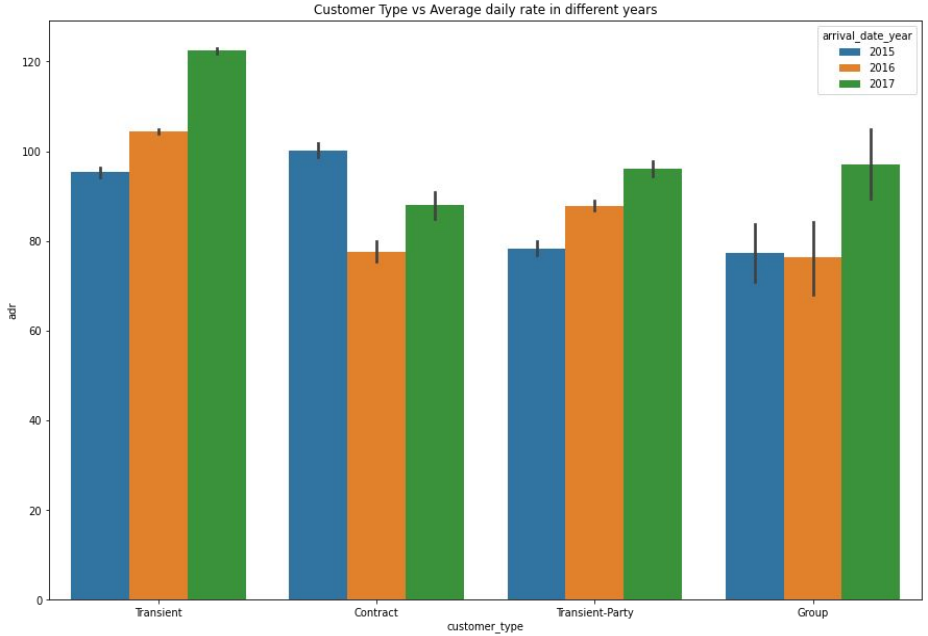
Cancellation for not assigning same room



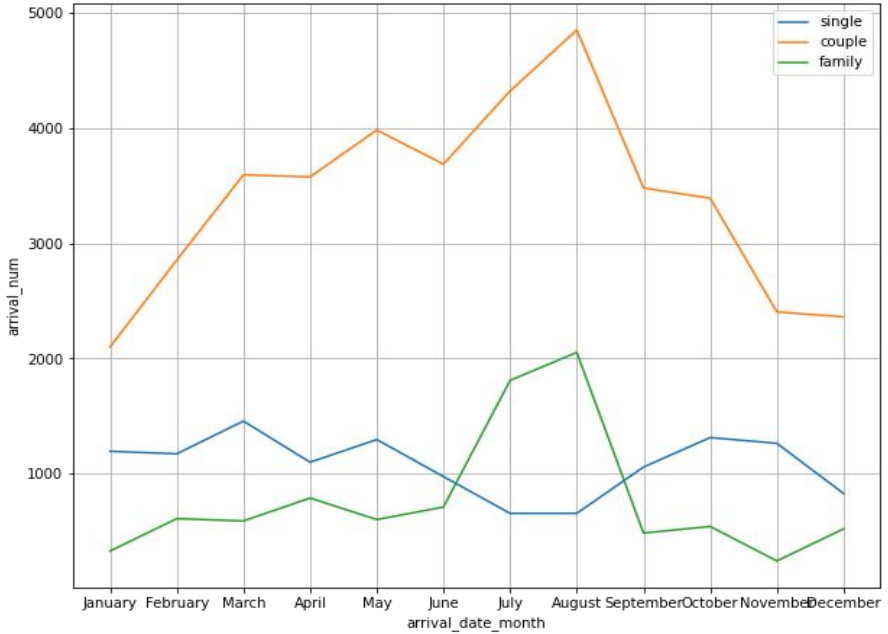
Car parking space



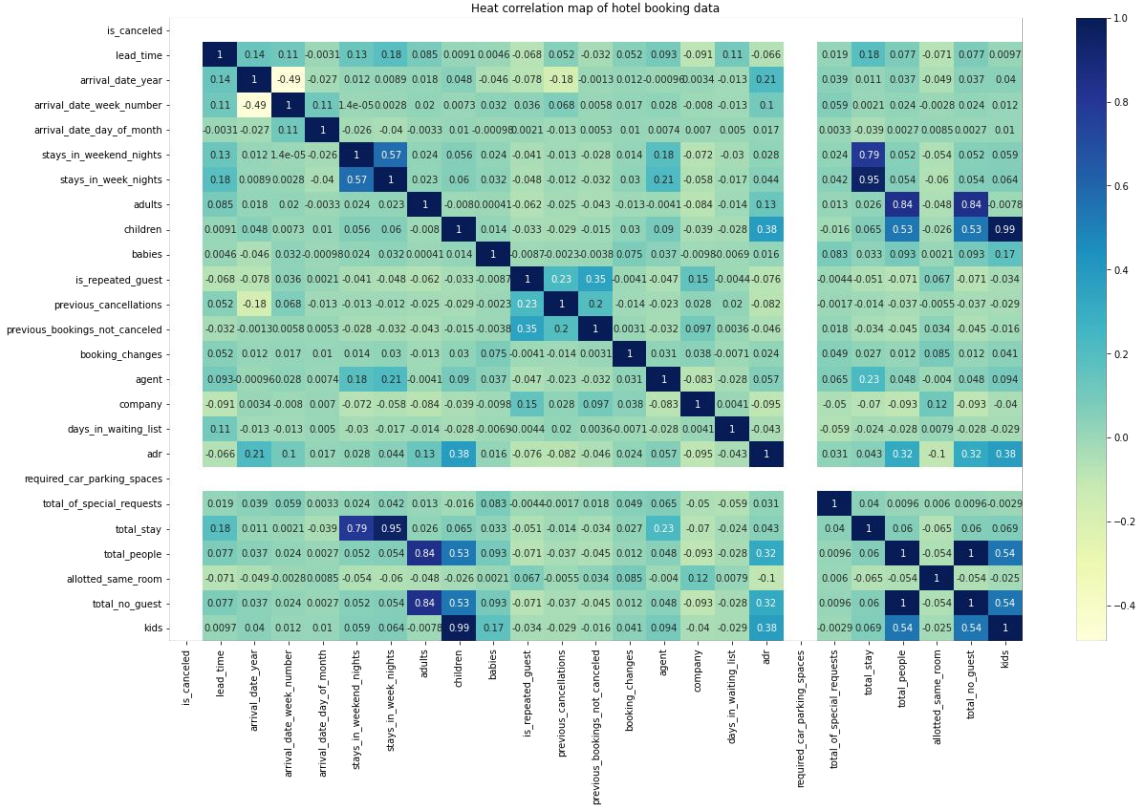
Customer type with maximum Average Daily Rate



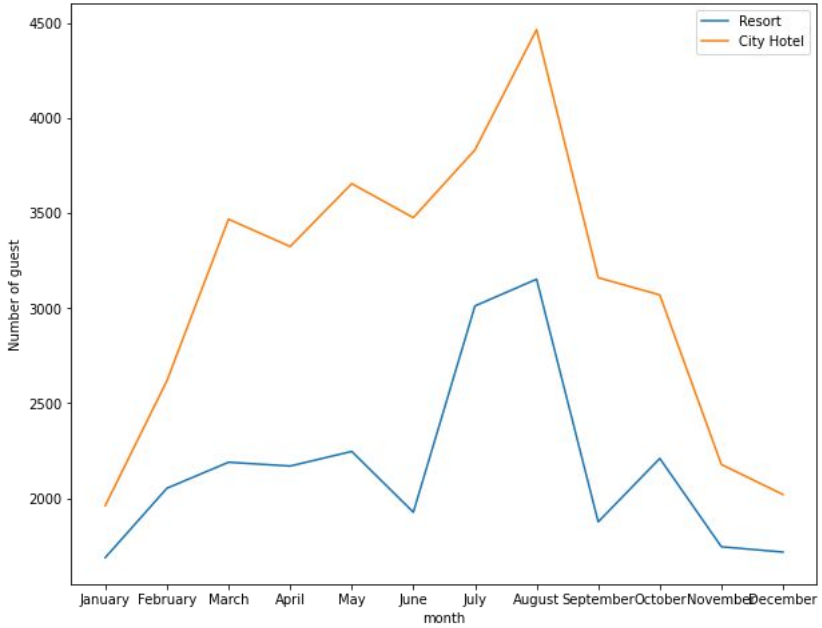
Type of customers booking the most



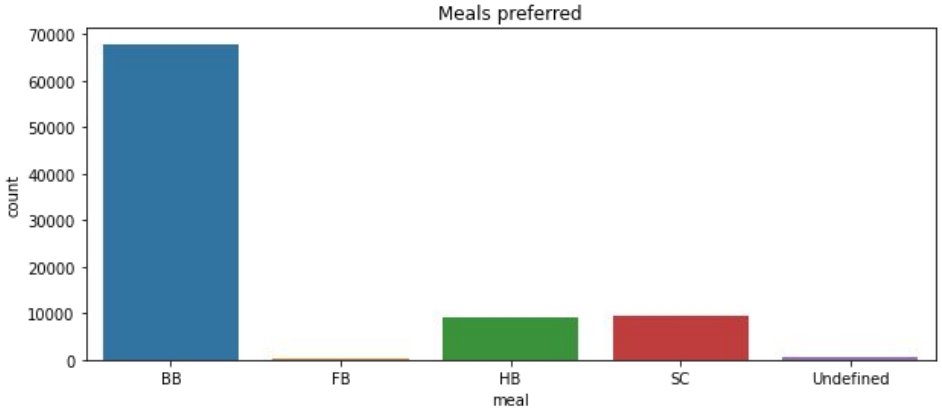
Heat Correlation Map



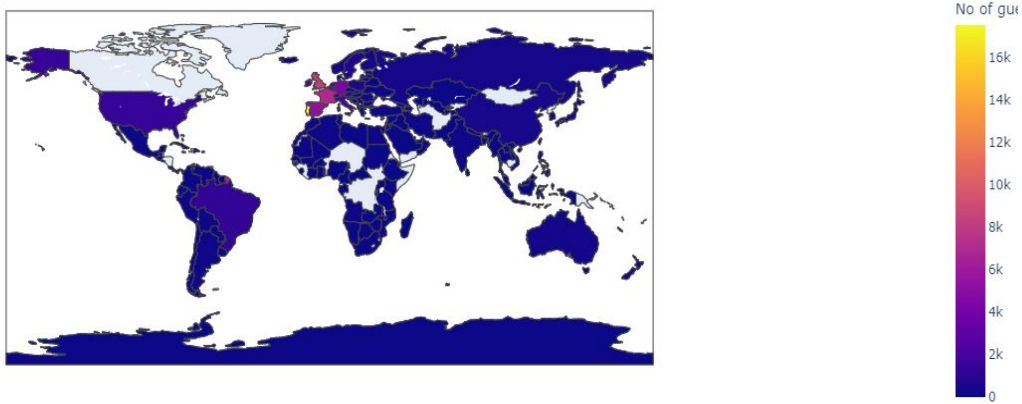
Best time to book a hotel room

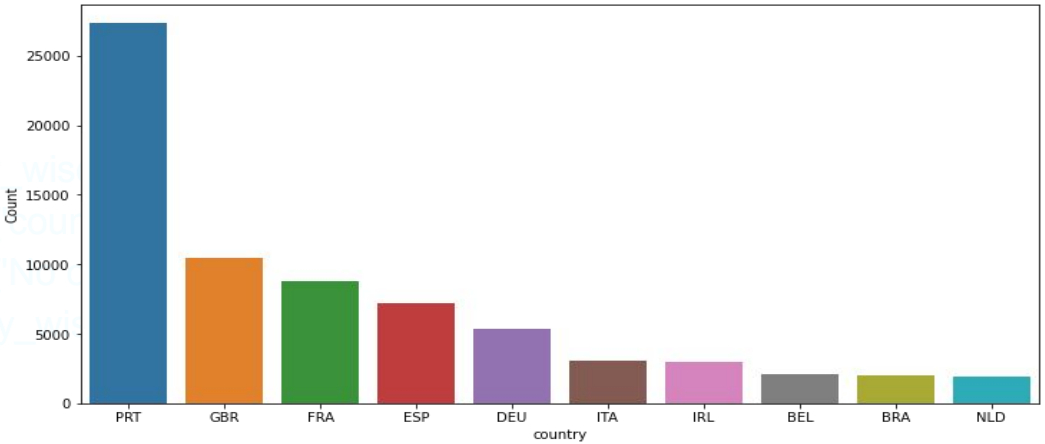


Board Basis preferred



Countries from which most customers are coming





Conclusion

Hotel wise analysis

- City hotel generates most revenue.
- 30% of customers of City Hotel have cancelled their booking.Whereas 5% of customers have cancelled their booking in Resort Hotel.
- City Hotel has a higher waiting time than that of Resort Hotel
- Room type A has the highest number of bookings compared to the other room types. The most number of bookings was by Transient Customer Type and the least was by Group customer type. Most number of customers used No Deposit option. The number of days stay was mostly 1. Most preferred hotel was City Hotel. Most number of bookings was done by couples.
- Both the customers have less chances of its customer returning for the stay.
- The most number of special request will be asked from Complementary market segment. And the cases where the number of adults is more than 3 and when there are 1 to 3 number of kids, we can expect for more special requests.

Distribution channel analysis

- Most number of customers have used TA/TO(Travel Agency/Travel Operator) distribution channel for hotel booking.
- GDS channel brings higher revenue for City hotel. Whereas for Resort hotel gets more revenue by direct and TA/TO channel.
- Mostly used market segment by the guests was Online TA to book City hotel and Resort hotel.

Cancellation related analysis

- Main reason for cancellation has been because of no car parking space. Waiting period, lead time or assigning of different room has not been a reason for cancellation.

Time and Stay related analysis

- We have used seaborn barplot to analyse the best customer type. We can conclude that, transient customer type generates maximum adr. The adr of transient-party has been increasing with the year. The group customer type does not show much of a progression.
- Board Basis preferred: In this analysis, we have concluded that the most preferred meal type by the customer is Bed and Breakfast(BB).
- As per the plot graph, we can conclude that most of the customers visit in the month of August.
- According to the graph, it has been mostly couples or families that has been visiting during the month of July and August.
- Most of the guests were from Portugal with 25k customers. Second by Great Britain with 10k customers. Followed by France, Spain and Germany respectively.