

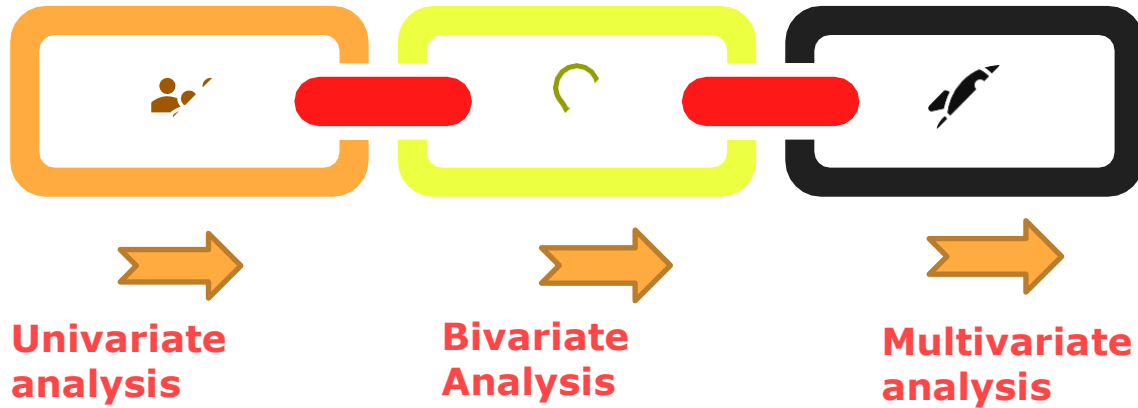
Project

Hotel Booking Analysis

Ankita Mishra

Agenda

We were given a 3 year hotel bookings dataset from 2015 to 2017, and we analyzed it as follows :



We had to try to figure out what's driving hotel booking trends this way.

Data description

The given dataset has crucial variables for hotel bookings. Some of them are:

Hotel :- Name of hotel (City or Resort)

• **Is_canceled** :- Whether the booking is canceled or not (0 for no canceled and 1 for canceled)

• **lead_time** :- time (in days) between booking transaction and actual arrival.

• **stays_in_weekend_nights** :- No. of weekend nights spent in a hotel

• **stays_in_week_nights** :- No. of weeknights spent in a hotel

• **Meal** :- Type of meal chosen

• **Country** :- Country of origin of customers (as mentioned by them)

• **market_segment** :- What segment via booking was made and for what purpose.

• **distribution_channel** :- Via which medium booking was made.

• **is_repeated_guest** :- Whether the customer has made any booking before(0 for No and 1 for Yes)

• **days_in_waiting_list** :- No. of days on waiting list.

• **customer_type** :- Type of customer(Transient, Group, etc.)

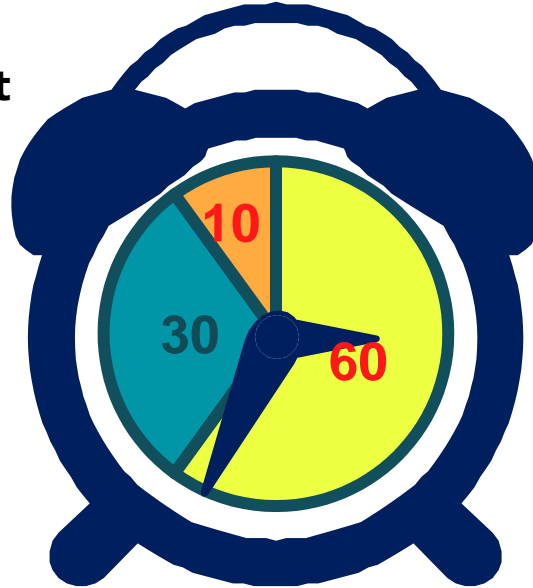
• **adr**:- Average Daily Rate.

• **total_of_special_requests** :- total no. of special request.

Dataset summary

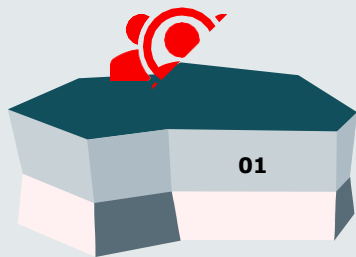
Basic summary of dataset

- dataset having 119390 observations and 32 columns
- There are 31994 duplicate values in this dataset.
- Null values for the features children, country, agent, and company are 4, 452, 12193, and 82137, respectively.



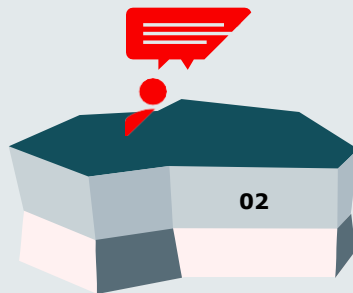
- lead_time and adr variables had outliers and were successfully removed.
- There are 87230 observations in the dataset after all null and duplicate values have been eliminated.

Dataset summary



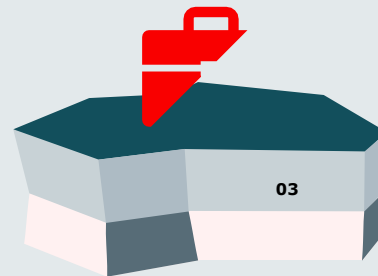
Binary

is_canceled,
is_repeated_guest



Categorical

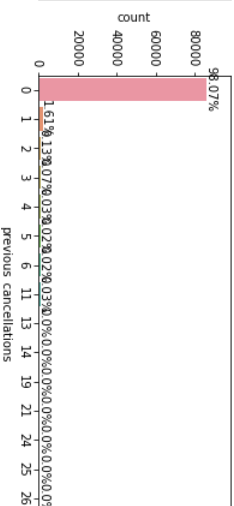
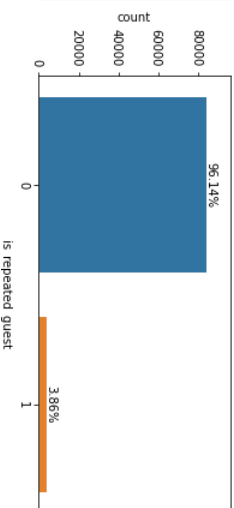
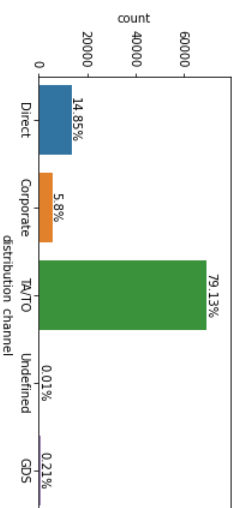
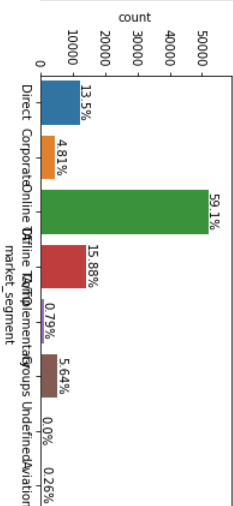
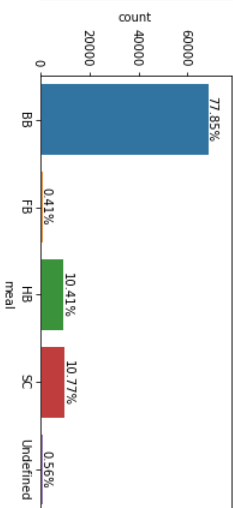
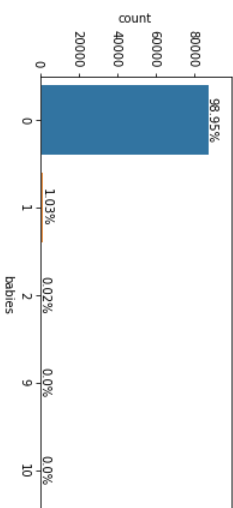
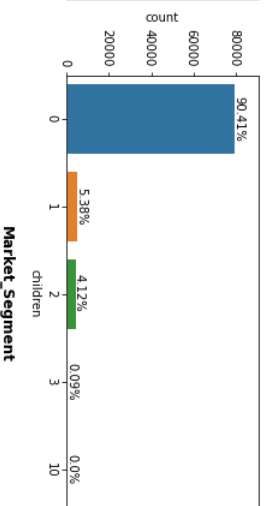
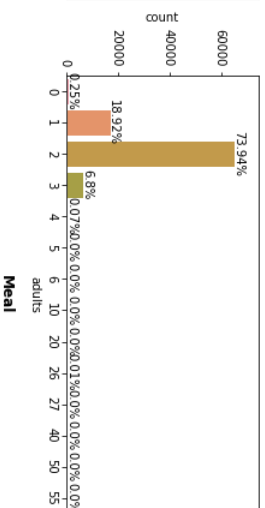
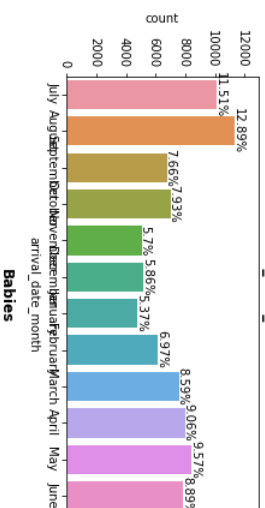
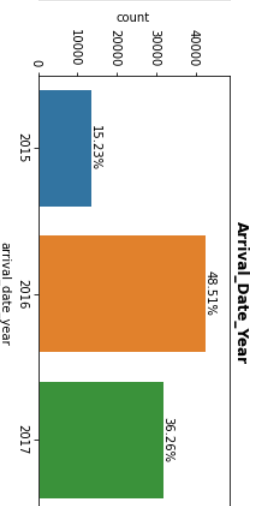
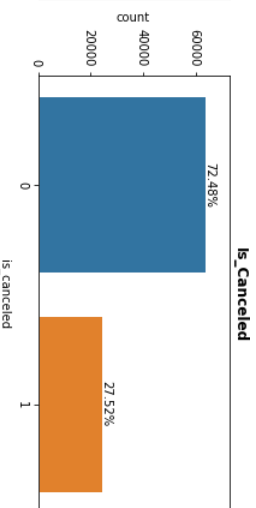
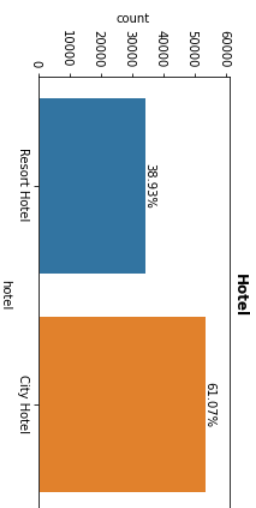
'hotel',
'arrival_date_month',
'meal', 'country',
'market_segment',
'distribution_channel',
'reserved_room_type',
'assigned_room_type',
'deposit_type',
'customer_type',
'reservation_status',
'reservation_status_date'



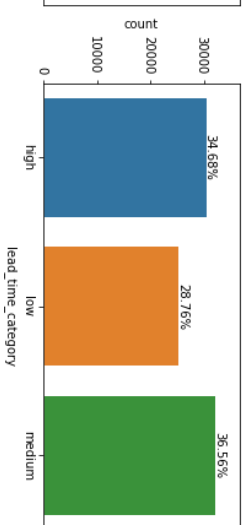
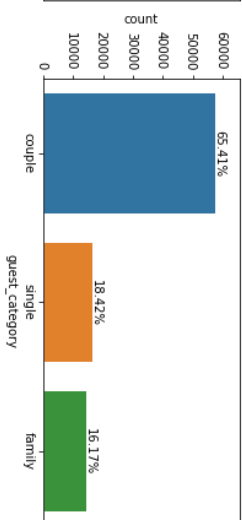
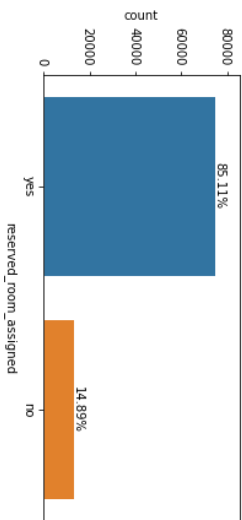
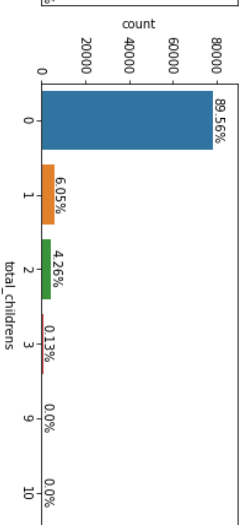
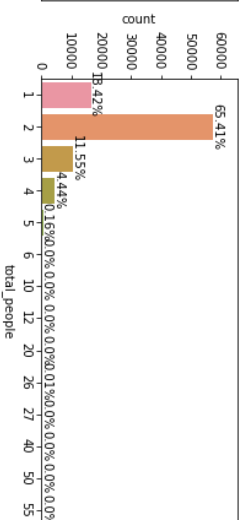
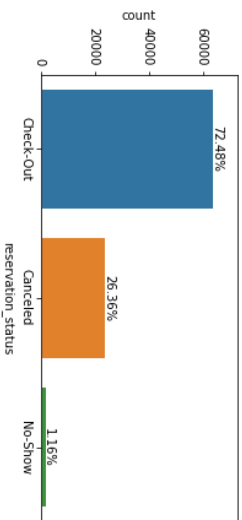
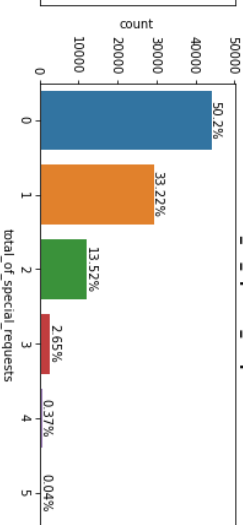
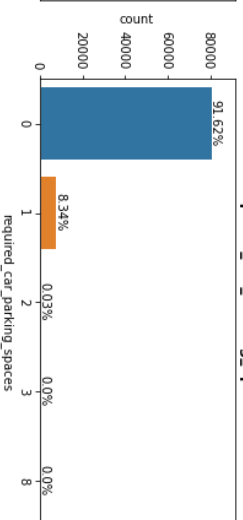
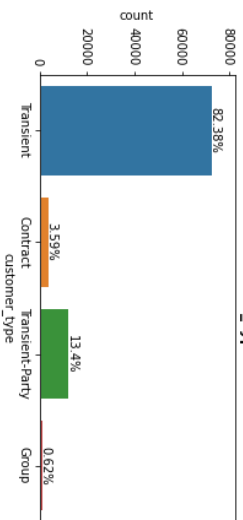
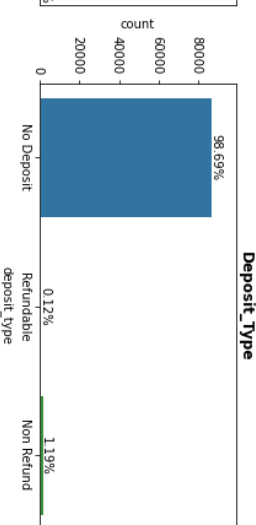
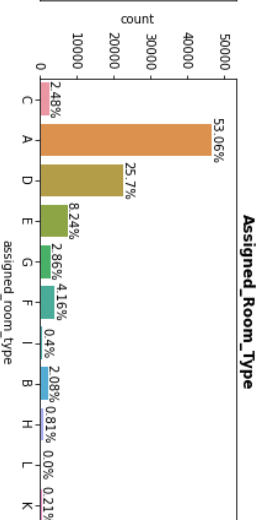
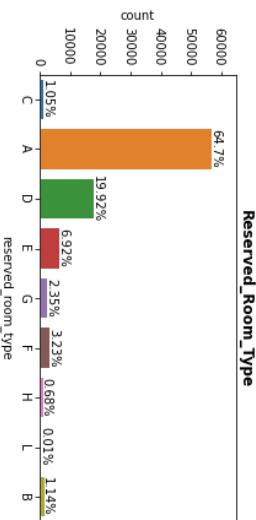
Numerical

'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',
'days_in_waiting_list', 'adr',
'required_car_parking_spaces',
'total_of_special_requests'

Univariate analysis



Univariate analysis (contd..)

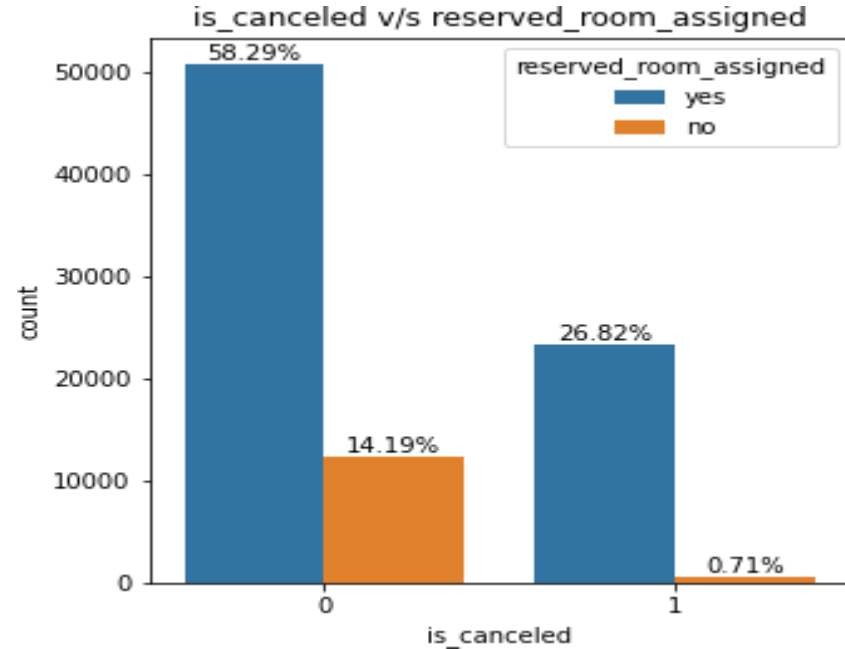


Obersvations :

- A city hotel was most preferred by 61.07 percent of customers over a resort.
- 72.48% of bookings are not cancelled. Almost one-third of all reservations are canceled.
- Bookings increased by 33.28% in 2016 compared to 2015, but fell by 12.25% in 2017.
- Customers make the most reservations in August, followed by July. Customers made the fewest reservations in November, December, and January. So we can make offers to customers in November, December, and January to maximise booking.
- BB is the most requested food.
- Most of the bookings are made through the online platform.
- The top distribution channel is TA/TO, which is used to make most of the bookings.
- The majority of hotel bookings are made by new customers. Very few customers (3.86%) visited again.
- The customer's top preference is for Room A to be reserved.
- Customers do not want to pay a pre-deposit for a reservation.
- Most customers (80%) preferred to book a hotel for a short stay.
- 90% of people do not require parking spaces for their vehicles.
- 70% chance that bookings will not be cancelled by customers.
- Reserved rooms were not assigned to 15% of customers. Ensure that customers receive the rooms they have reserved.
- Reservations were often made for two people. 10% or so of guests brought their families. Few bring their families with them. Offer family-friendly discounts to encourage reservations for family and business events.

Bivariate Analysis

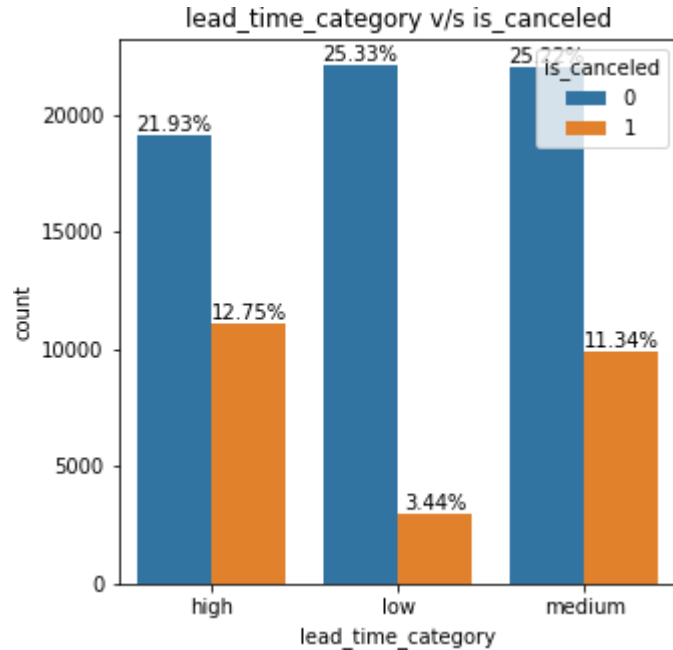
•**Question 1: Is not having a reserved room assigned a reason for booking cancellations?**



- The inability to assign a reserved room to a customer is not grounds for cancellation.
- Less than 1% of people who cancelled their booking when the reserved room was not assigned.

Bivariate Analysis (contd..)

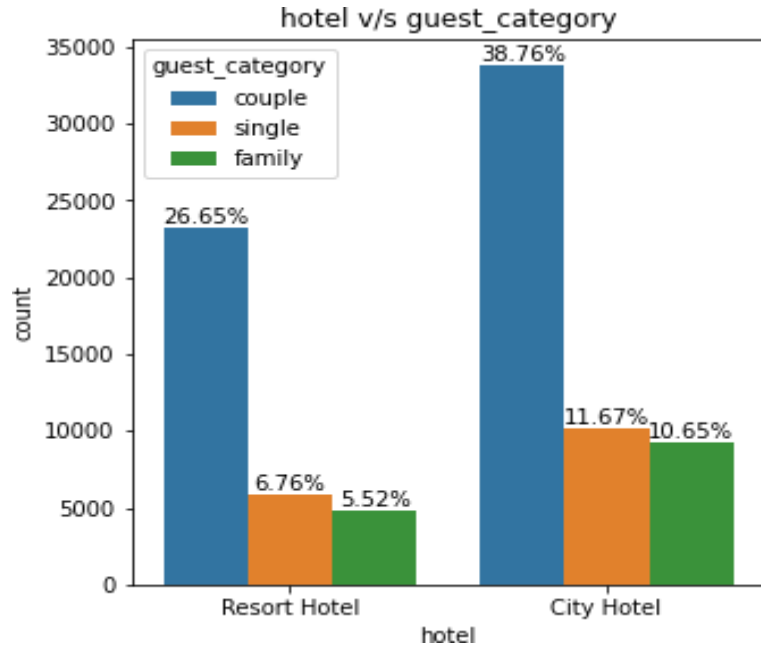
Question 2 : Is the high lead_time a reason for booking cancellations?



- Less lead time means fewer cancellations. Customers who book hotels more than 15 days in advance are more likely to cancel their booking.
- Booking cancellations are not caused by a longer Lead time.

Bivariate Analysis (contd..)

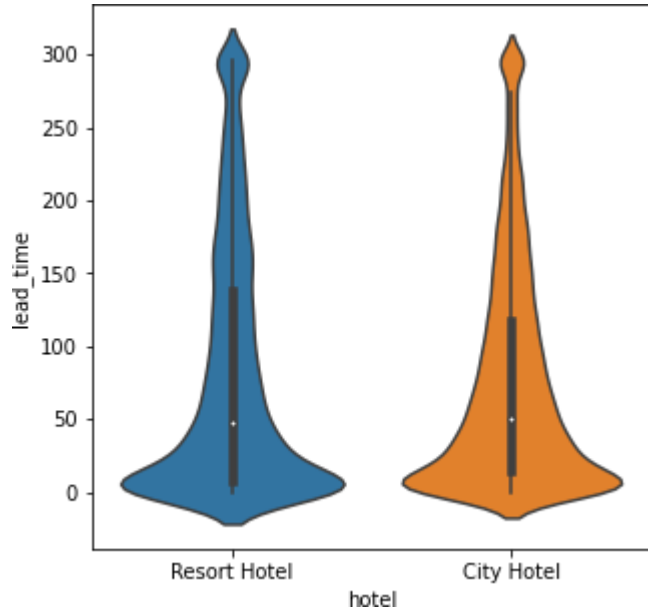
Question 3: How many people are reservations made for?



- Most customers book hotels for two people (couples). Customers prefer city hotels over resorts for family bookings.
- A city hotel is preferred when booking for a single person.

Bivariate Analysis (contd..)

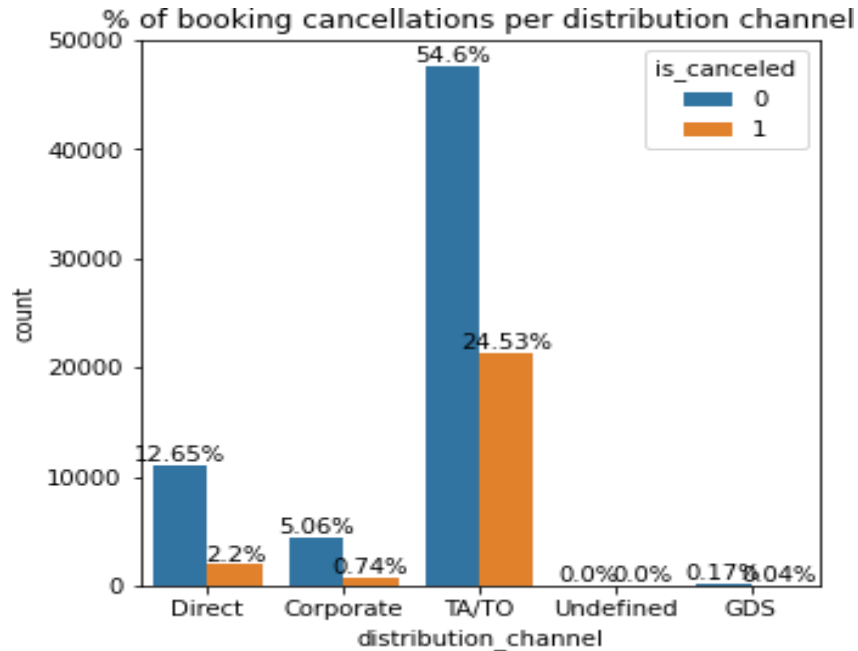
Question 4: Which hotel type has the most advanced reservations?



- In comparison to city hotels, guests book resort hotels a little bit in advance.

Bivariate Analysis (contd..)

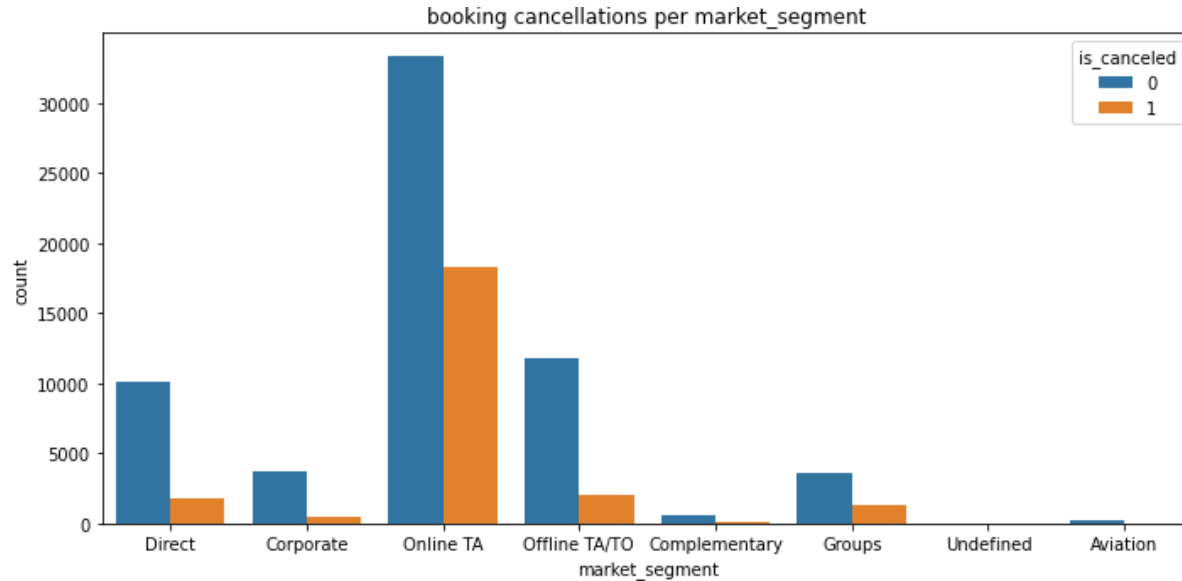
Question 5: Which distribution channels have the most cancellations of bookings?



- The majority of canceled bookings were made through the TA/TO distribution channel.
- Bookings made through the Direct, Corporate, and GDS distribution channels are extremely unlikely to be canceled.

Bivariate Analysis (contd..)

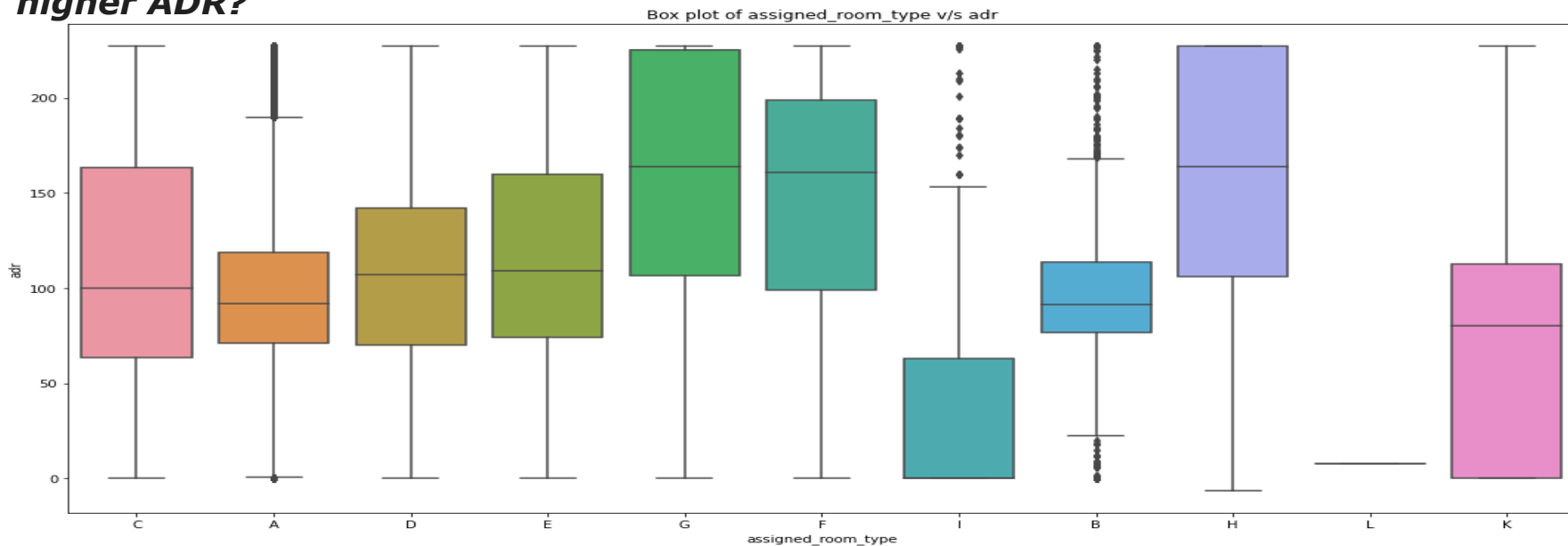
Question 6 : Which market segment is most used for booking hotels, and which market segment bookings are most canceled?



- The majority of hotel reservations are made online, as are the majority of cancellations of reservations made by customers who made their reservations online.

Bivariate Analysis (contd..)

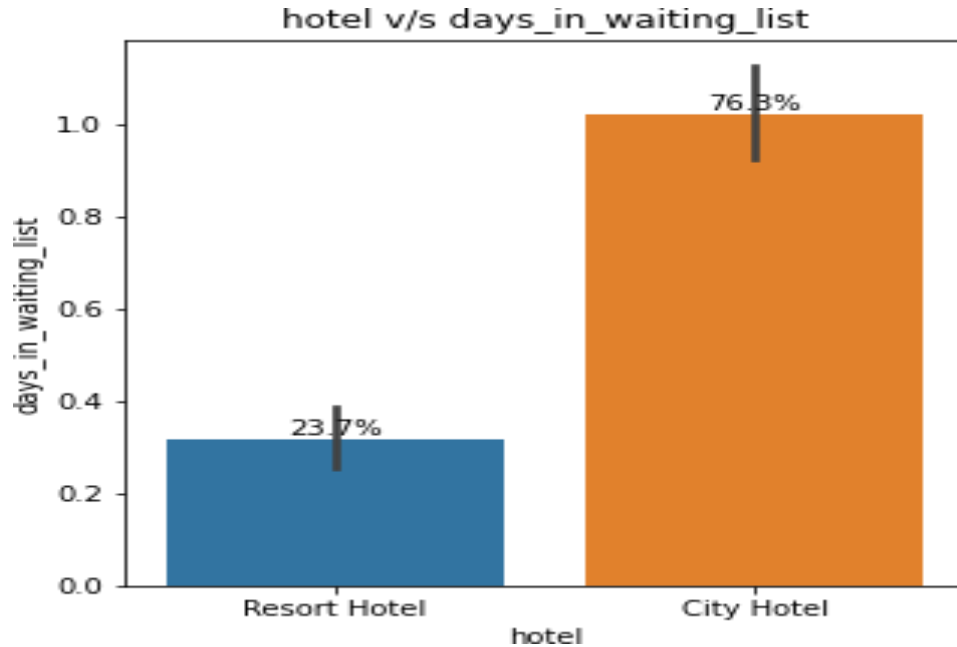
Question 7 : Which room generates a higher ADR?



- Room types G, followed by H, generate high ADR. Room I has a very low ADR. The most popular room is A, but it has a lower ADR than other room types that are less popular with customers for bookings.
- To maximize revenue, the hotel should increase the number of rooms in A, G and H.

Bivariate Analysis (contd..)

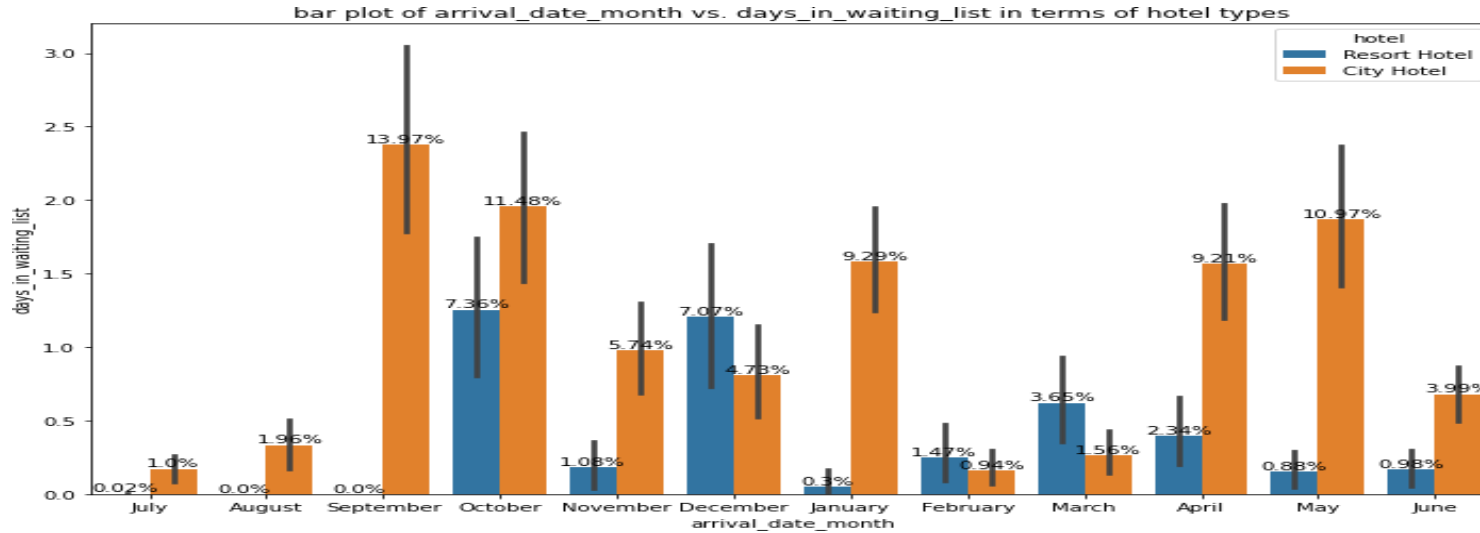
Question 8 : Which hotel type is the busiest?



➤ A city hotel is busier than a resort.

Bivariate Analysis (contd..)

Question 9: Which month is the busiest for hotels?

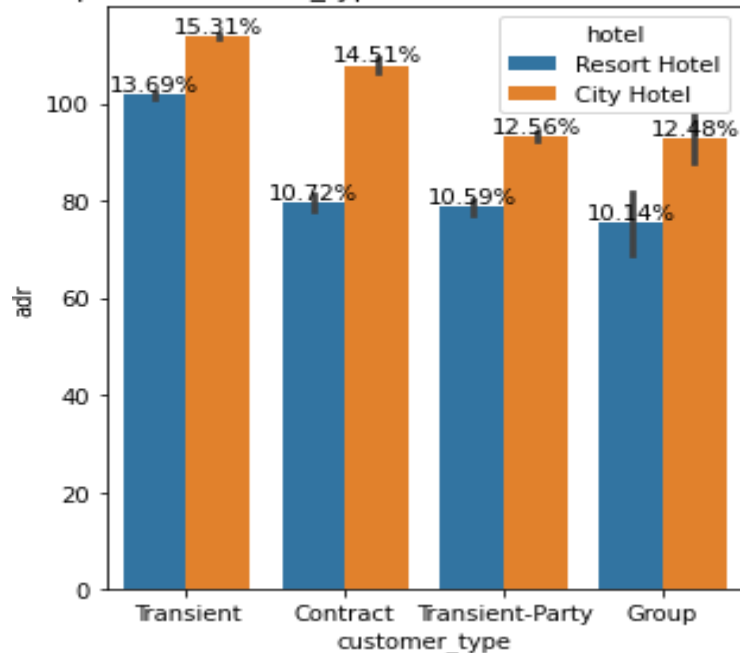


- Customers prefer city hotels over resort hotels for the New Year. The city hotel is busiest in October and September, which means that this month has the most bookings.
- In the fourth quarter of the year, in December, resorts are busier than city hotels. Resort hotels are significantly less busy in the third quarter of the year than they are throughout the year.
- In July there is no longer waiting period for bookings.

Bivariate Analysis (contd..)

Question 10 : Which customer type generates more revenue in terms of hotel types and customer types?

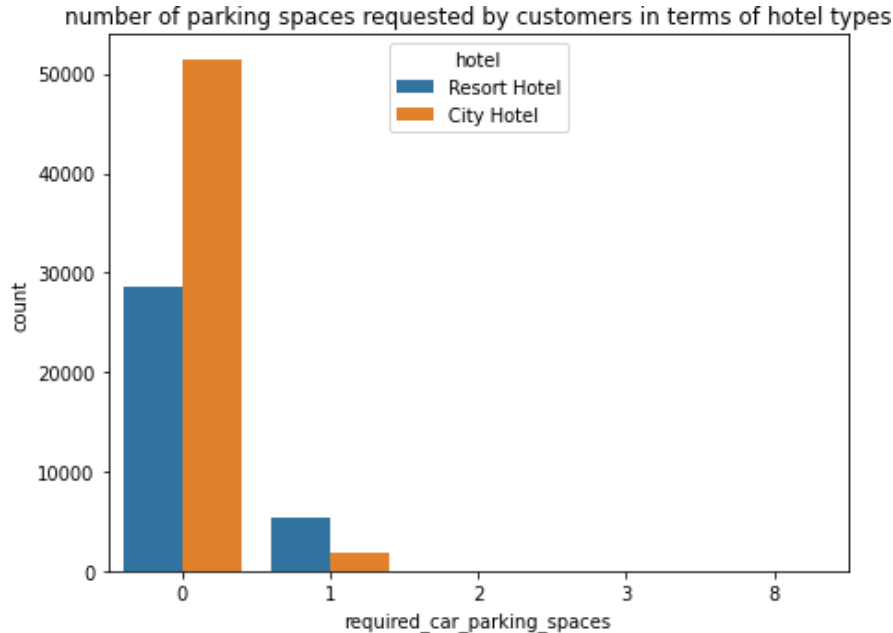
bar plot of customer_type vs. adr in terms of hotel types



- City hotels generate more revenue (54.86%) than resort hotels (45.14%).
- Transient customers who book rooms for a short period of time generate more revenue than other types of customers.

Bivariate Analysis (contd..)

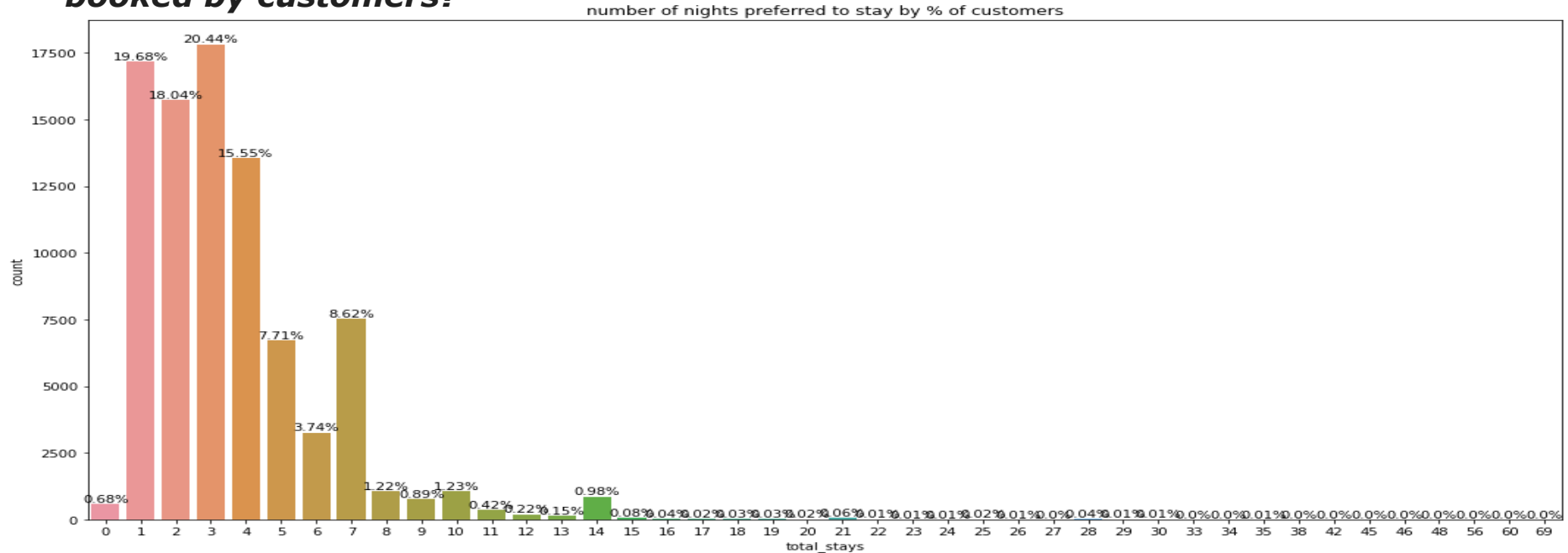
Question 11 : In terms of hotel types, how many parking spaces are most frequently requested by customers?



- Most of the customers (91.63%) did not request any parking spaces. Only a few customers (8.31%) requested parking.
- One parking space is most desirable to customers.
- The hotel can eliminate parking spaces 3 and 8 and concentrate on parking spaces 1 and 2, which will reduce the cost of providing parking spaces.

Bivariate Analysis (contd..)

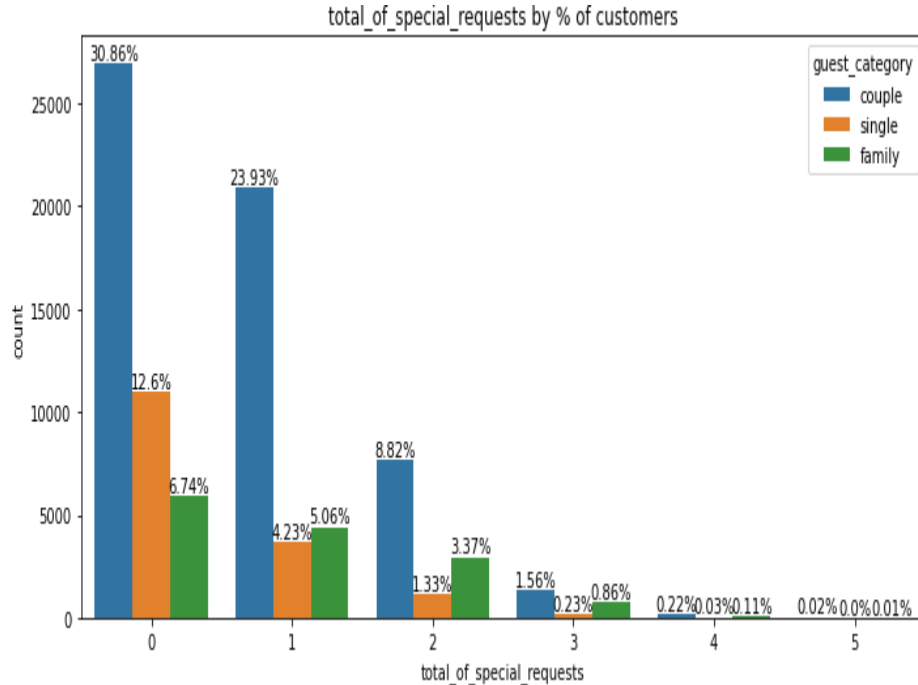
Question 12 : What is the most common number of nights booked by customers?



- The majority of the guests are staying at the hotel for three nights. Customers who book hotels for more than one week are very rare. Most hotel bookings are for less than 7 nights.
- The hotel should pay more attention to keeping reservations available for less than 7 nights to not lose customers and to generate more revenue.

Bivariate Analysis (contd..)

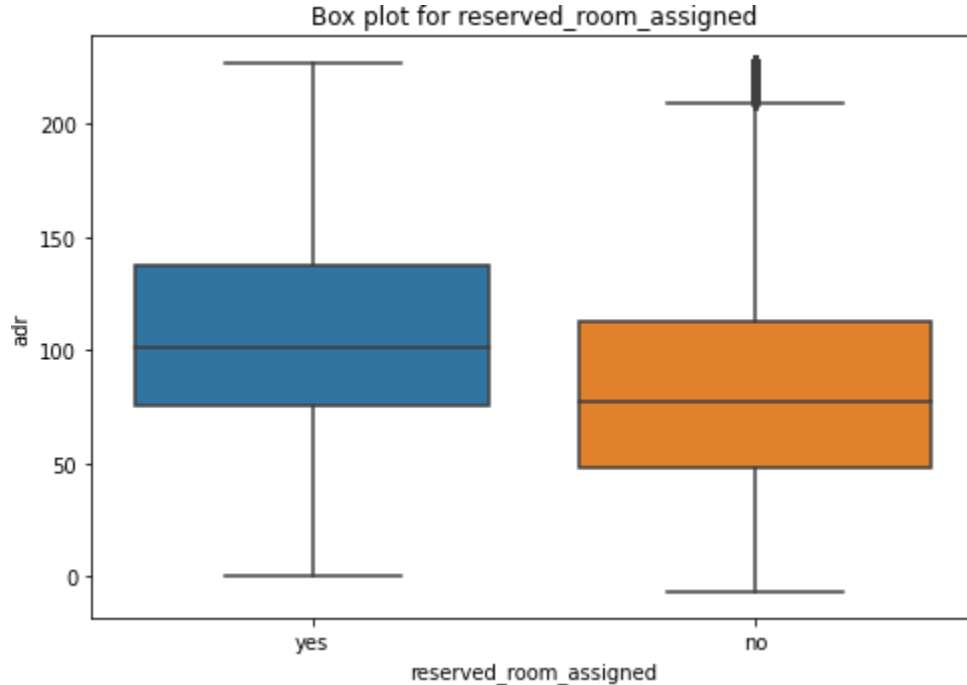
Question 13 : What is the most common number of special requests made by customers, and what kind of customer are they?



- Approximately half of all customers do not make special requests. Customers frequently make one special request.
- Couples make the majority of special requests.
- Bookings made for single people are not more requested for special requests than those made for couples, followed by family and business events.

Bivariate Analysis (contd..)

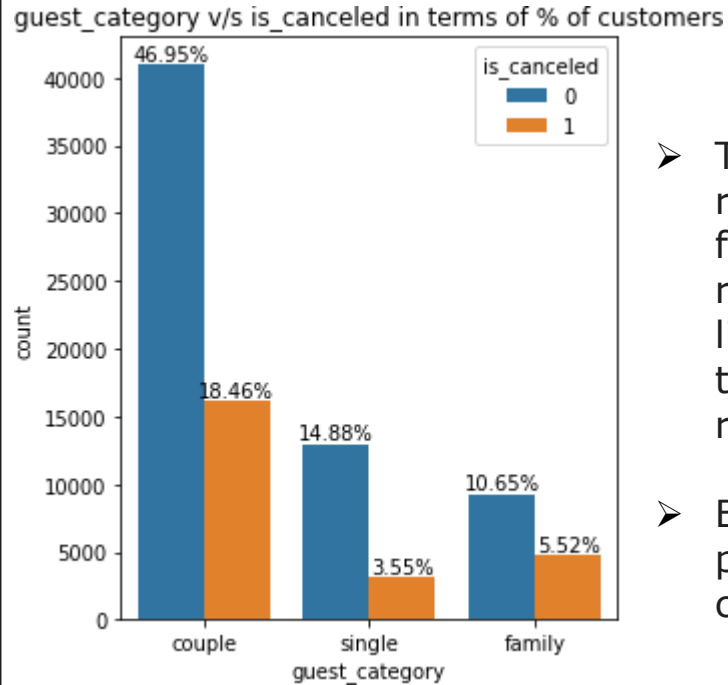
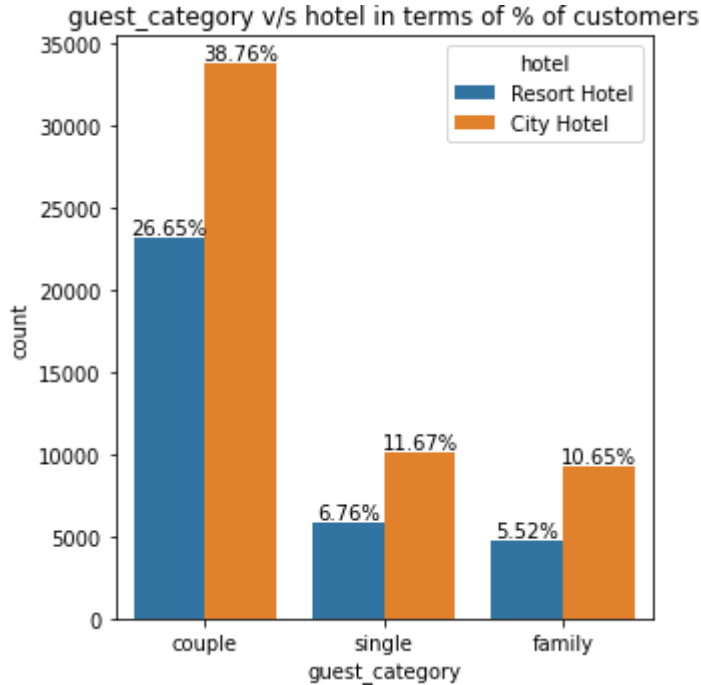
Question 14 : Is the ADR affected by the hotel not giving a reserved room?



- There is no significant difference in ADR between reserved rooms assigned and not reserved rooms assigned.
- Not assigning a reserved room does not affect ADR.

Bivariate Analysis (contd..)

Question 15: The majority of bookings were made for how many people, and the majority of cancellations of bookings were made for how many people?

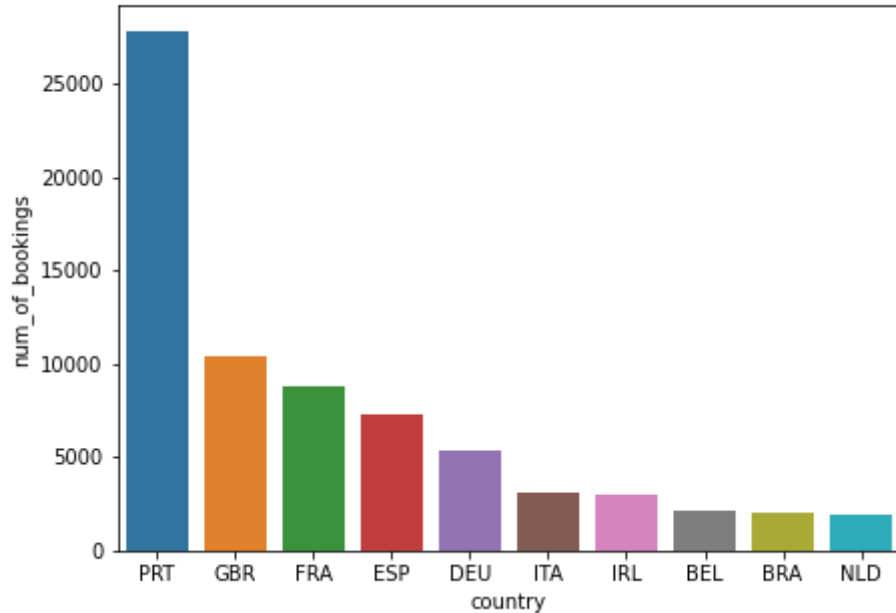


- The majority of hotel reservations are made for couples. Couples' reservations are more likely to be canceled than singles' or families' reservations.
- Bookings for single people are rarely canceled

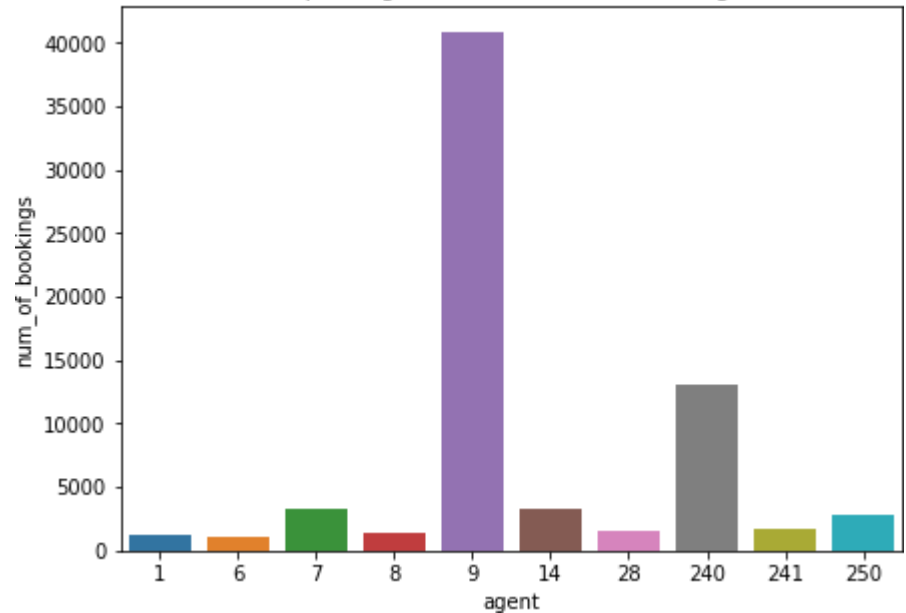
Bivariate Analysis (contd..)

Question 16 : Which country makes the most reservations, and which agent makes the most bookings?

Top 10 countries with number of bookings



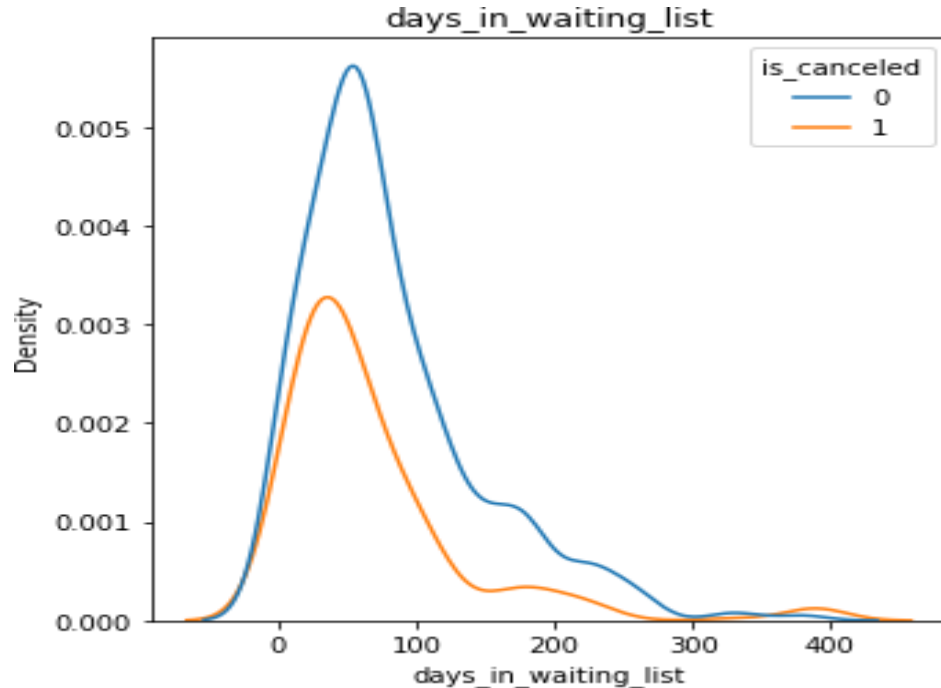
Top 10 agent with number of bookings



- The majority of reservations are made through country PRT. Customers make the most bookings in the following top 5 countries: PRT, GBR, FRA, ESP, and DEU.
- Agent number 9 made most number of bookings. 9, 240, 7, 14 and 250 are the top 5 agents by number of bookings made.

Bivariate Analysis (contd..)

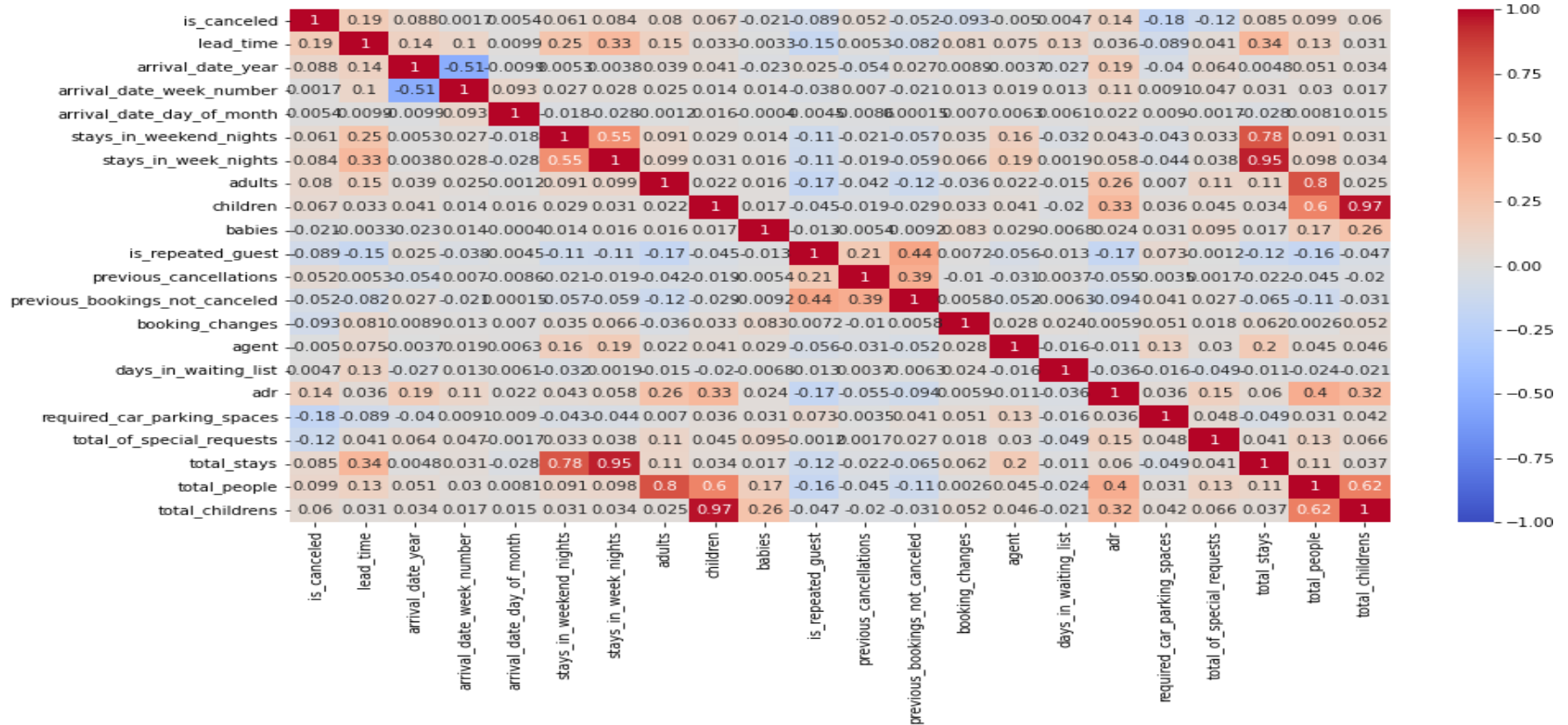
Question 17 : Does a longer waiting period cause the cancellation of bookings?



- The majority of canceled bookings have a waiting period of less than 150 days, but those that are not canceled bookings by customers have a waiting period of less than 150 days, which has a higher density than the canceled bookings.
- So a longer waiting period is not a reason for booking cancellation.

Multivariate Analysis

1] Correlation Heatmap



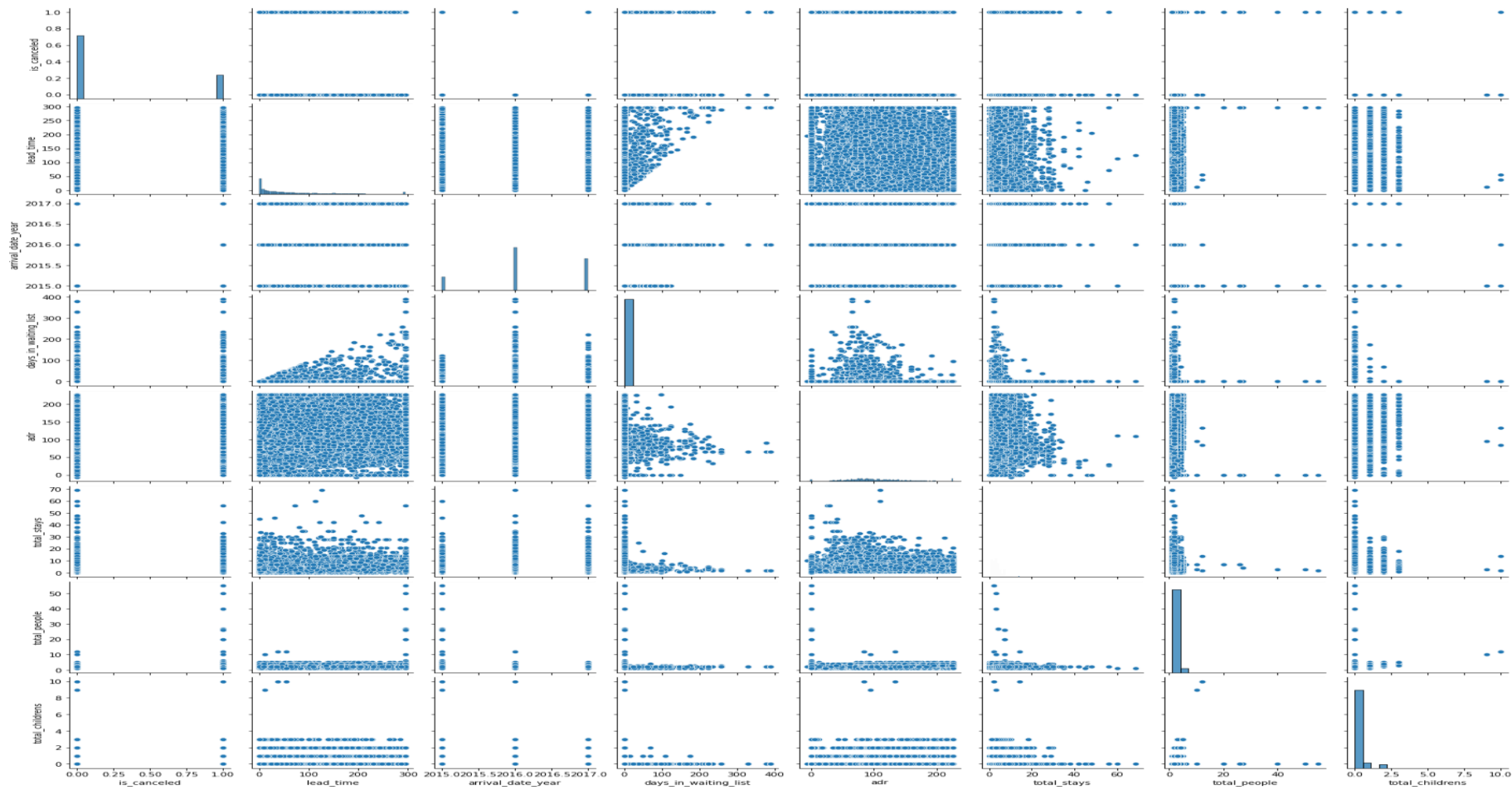
Multivariate Analysis (contd..)

1] Correlation Heatmap (contd..)

- In the heatmap, we see some high correlation between a few variables because we created new variables total_stays, total_people, and total_children from existing variables and did not drop old variables.
- The variables lead_time and is_canceled have weak relationships. The most likely reason for cancellation is a longer lead time.

Multivariate Analysis (contd..)

2] Pair Plot



Multivariate Analysis (contd..)

2] Pair Plot (contd..)

- We can conclude from the relationship between lead_time and is_canceled that a longer lead time does not result in cancellations.
- We can conclude from the lead_time and arrival_date_year graphs that people were consistently interested in booking rooms in advance in 2015, 2016, and 2017.
- According to the graph of arrival_date_year and days_in_waiting_list, 2016 had the longest waiting period for room bookings.
- We can conclude from the graph of ADR and days_in_waiting_list that a short minimum waiting period for bookings results in a high ADR.
- From the graph of "is_canceled" and "days_in_waiting_list," we conclude that waiting for bookings is not a reason for booking cancellation.

Solution to Business Objective

- The GDS distribution channel can be eliminated by hoteliers because bookings made through it are extremely low.
- Hotels can cut costs by eliminating market segments such as complementary and aviation, as bookings through these segments are very low.
- Only 3.86% of customers return to the hotel, so hotels can increase repeat bookings by offering the right repeat booking incentives.
- Offer packages and promotions to promote bookings for the resort hotel.
- Bookings for 3 and 8 parking spaces were rarely requested by customers, hotels can only keep bookings for 1 and 2 parking spaces to save money.
- Almost 25% of customers cancelled their bookings. Hotels should implement a cancellation policy, discount on confirmed bookings, and send booking reminders to guests to reduce booking cancellations.
- People typically book rooms for two people, so encourage family and group bookings to maximize revenue.

Conclusion

- The top country with the most number of bookings is PRT, and the number one agent with the most number of bookings is 9.
- Customers favored city hotels more than resort hotels by a margin of 61.07 percent.
- One of the four reservations is canceled.
- The most popular food is BB.
- The Online (internet) platform is used to make the majority of bookings.
- The majority of the bookings are made using TA/TO, the leading distribution channel.
- The vast majority of hotel bookings are made by new guests. Almost no consumers (3.86%) returned.
- The customer wants Room A to be reserved the most.
- Customers do not wish to make a bookings with a pre-deposit.
- Customers (80%) favored making a hotel reservation for a short visit.

Conclusion (contd..)

- Only 10% of people require space to park their cars.
- Most visitors are couples.
- The inability to assign a reserved room to a customer is not grounds for cancellation.
- Booking cancellations are not caused by a longer Lead time.
- A city hotel is busier than a resort.
- The busiest months for hotels are October and September. There isn't a lengthy wait for reservations in July.
- Not assigning a reserved room does not affect ADR.

Challenges

- The data contained a large number of duplicates.
- The improper data type format was used for the data.
- It was challenging to select the best visualization techniques.
- The dataset contained a large number of null values.

Thank You !