

# **Capstone Project - 1**

## **Exploratory Data Analysis**

### **HOTEL BOOKING ANALYSIS**

**By - Rahul Shah**

# Summary

The hotel industry is solely concerned with the provision of guest accommodation and related services. It has a different set of booking arrangements with different distribution channels. I have worked on the Hotel Booking Dataset to get a comprehensive image of the hotel industry.

In this project, By understanding the dataset, I have framed few exploratory questions and performed exploratory data analysis on the same to draw out key insights about the hotel industry.

**Dataset Name :** Hotel Booking Analysis

**Shape :** Rows: 119390 , Columns:32

# Data Cleaning (Handling Missing Values)

In our dataset, we have four features with null values.

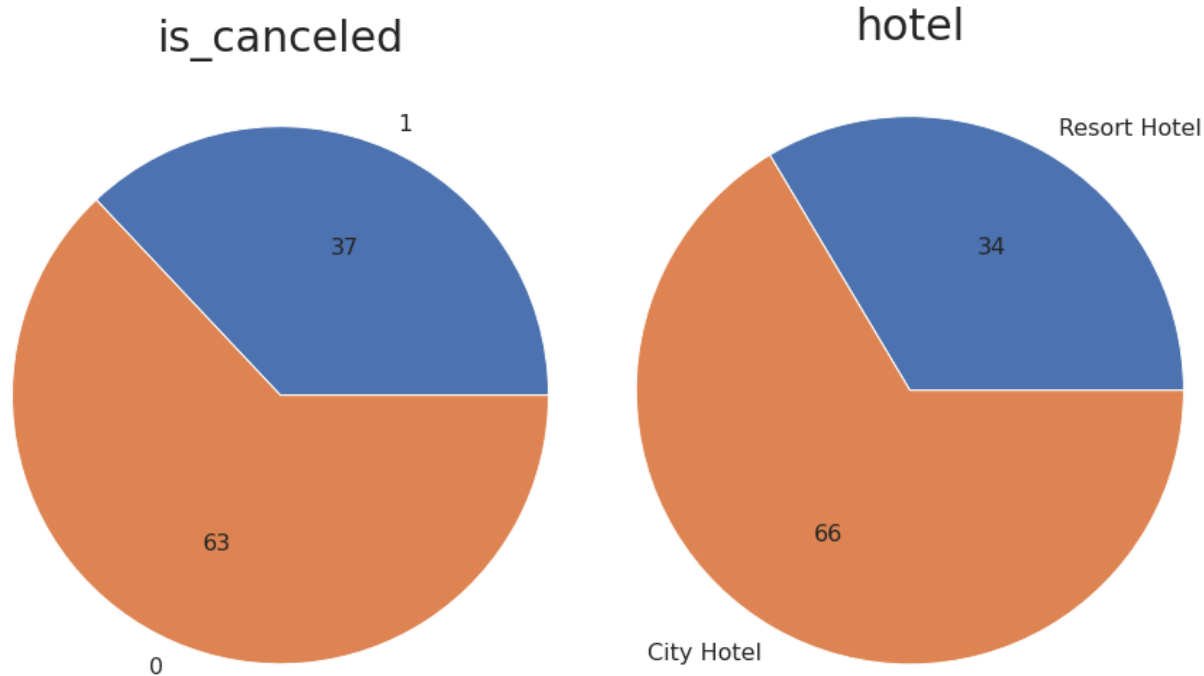
1. Children column contains the count of children, so we will replace all the missing values with 0.
2. Country column contains country codes representing different countries. It is a categorical feature, we replace missing values with global constant (Unknown).
3. Agent column contains the id number of the agent, so missing values are been filled with 0.
4. Company column contains 94% null values, so dropped this column.

	null_values	null_percent
children	4	0.003350
country	488	0.408744
agent	16340	13.686238
company	112593	94.306893

# Exploratory Data Analysis

# Basic Analysis

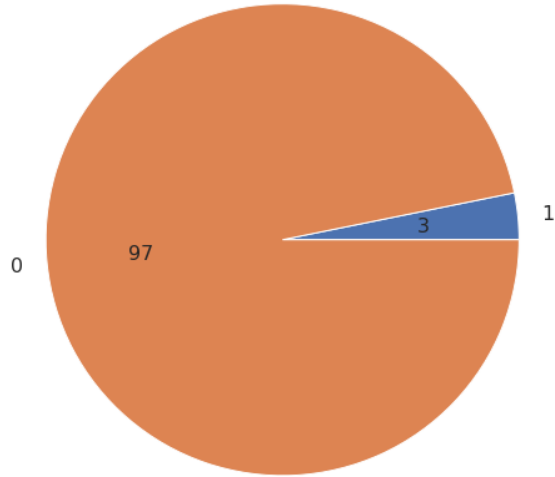
1) Understanding how data the revolves around the hotel bookings



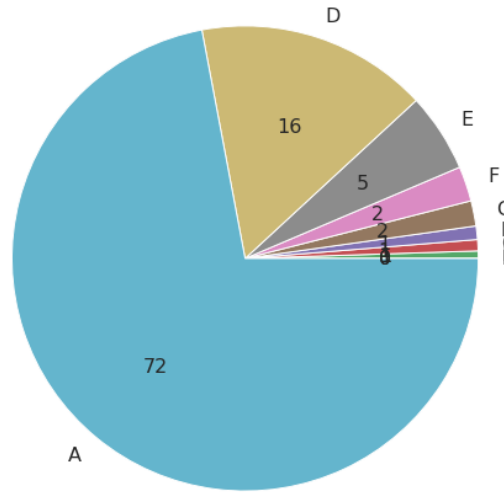
1) 63% bookings were made and 37% bookings were canceled

2) The Booking percentage of City Hotel (66%) is almost twice the percentage of bookings in Resort Hotel(34%)

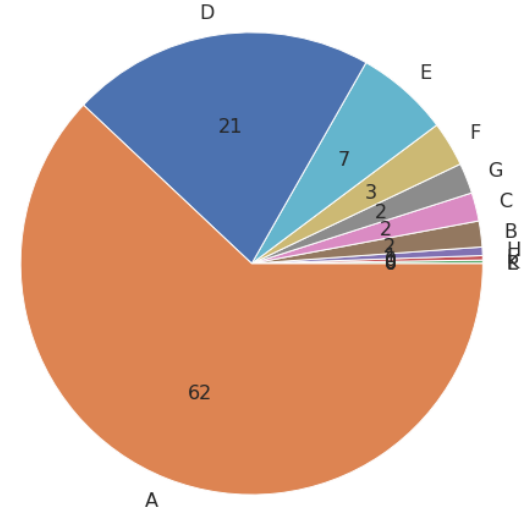
is\_repeated\_guest



reserved\_room\_type



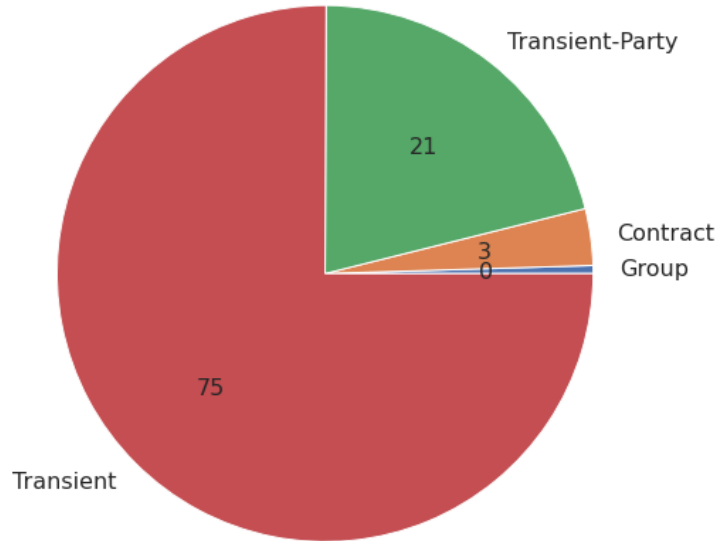
assigned\_room\_type



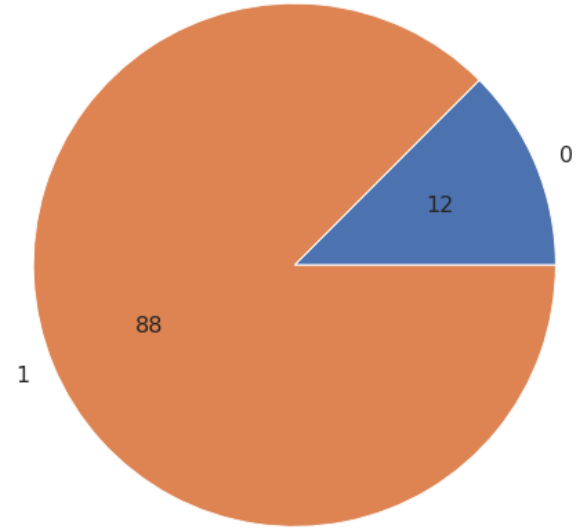
3) Only 3% of guests are repeated

4) 72% guests book room type A but only 62% guests were assigned the same room

customer\_type



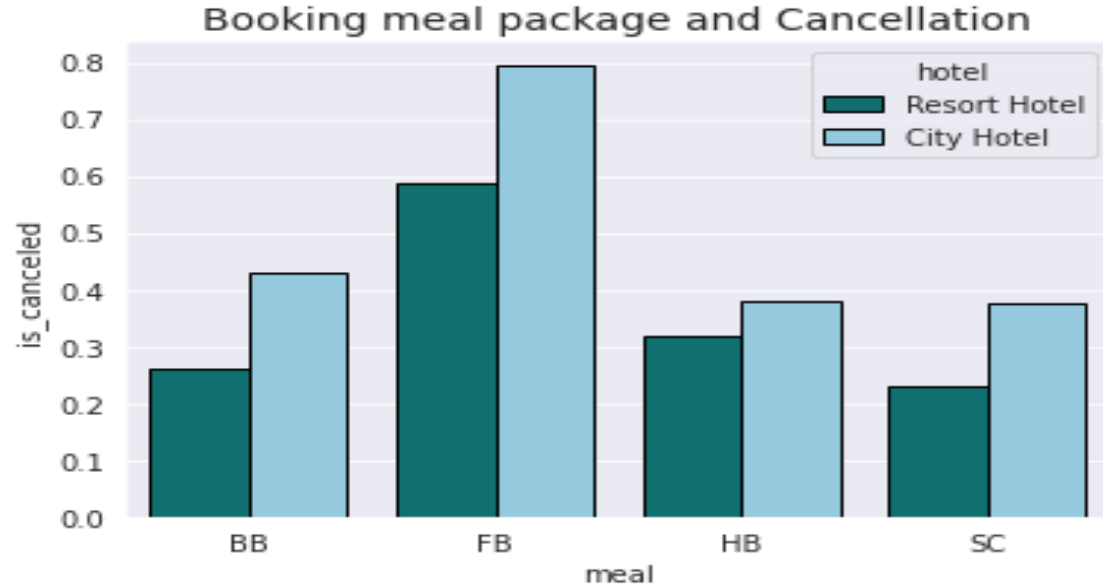
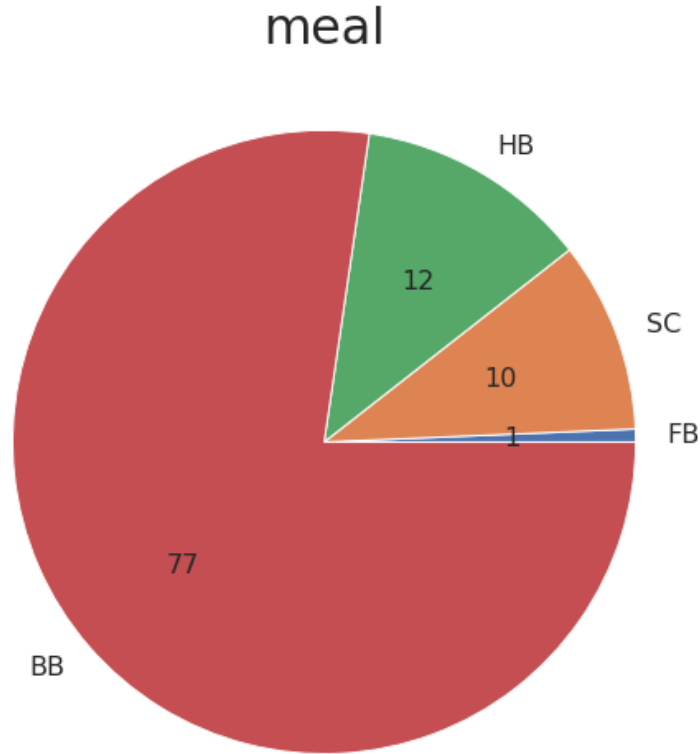
same\_room



5) 75% customers are transient. These are the temporary travelers.

6) 88% guests were assigned the same room.

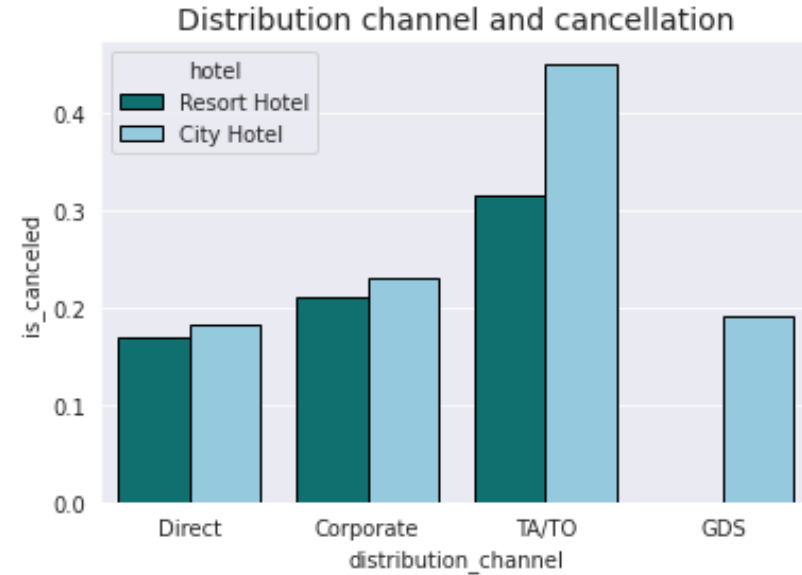
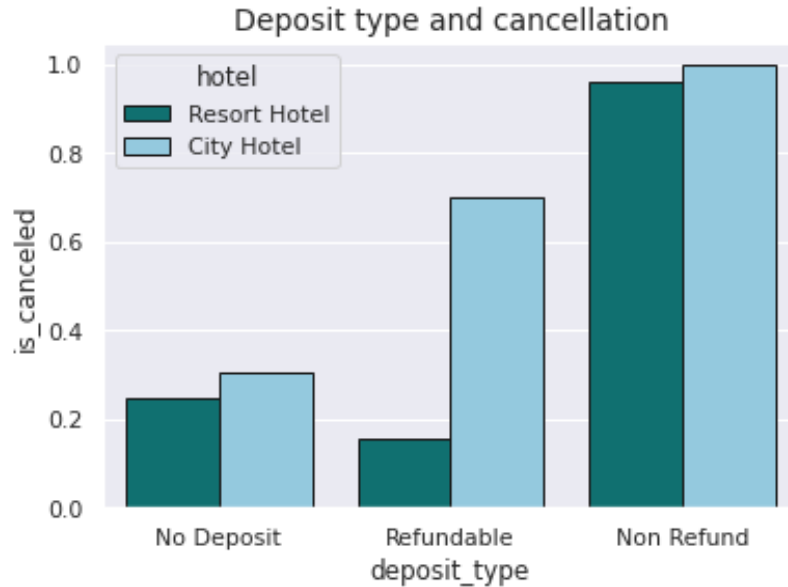
## 2) Booking meal package having highest cancellation rate



Bed & Breakfast is the most popular meal type in hotels.  
Resort hotel has highest confirmed meal bookings.  
Full Board meal package has the highest cancellation rate.



### 3) Deposit type and Distribution channel most likely to cancel

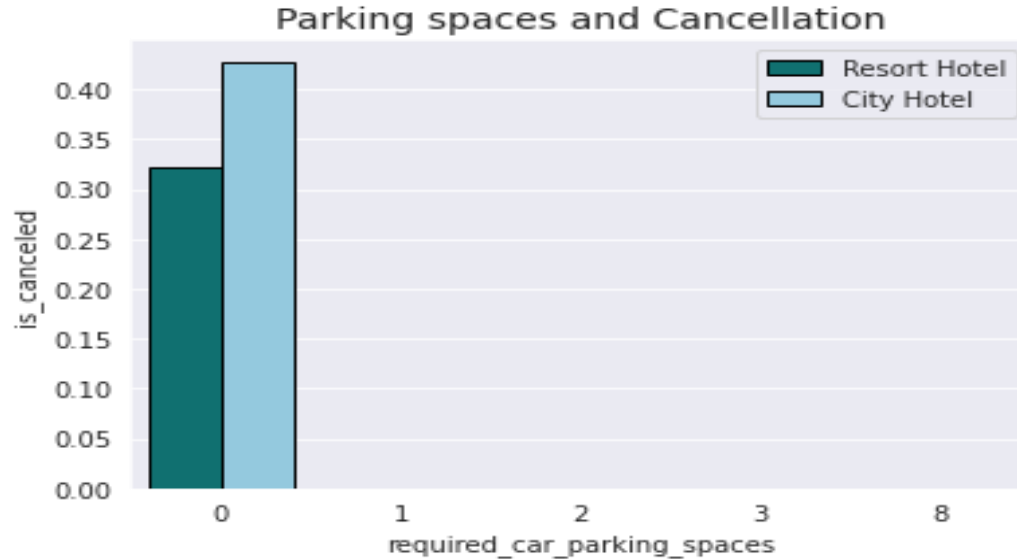


Non Refundable deposit has the highest cancellation rate.

Travel Agent / Tour Operators has the highest cancellation rate.

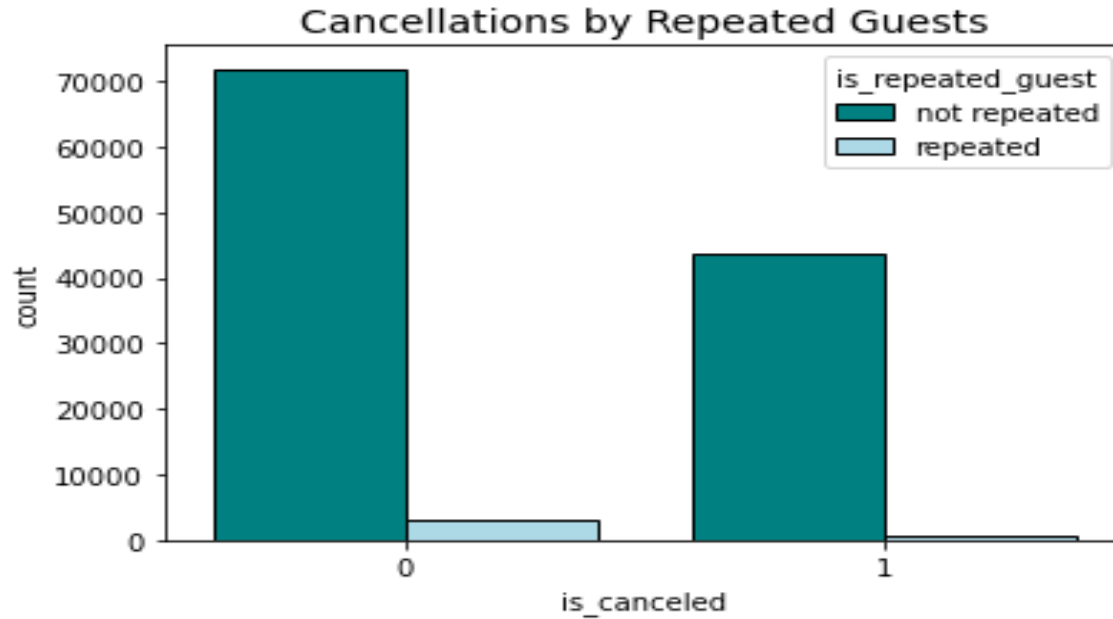
Direct has the lowest cancellation rate

## 4) Parking spaces and cancellation



Graph shows that the customers who needed a parking space for their vehicle never cancelled their hotel booking.

## 5) Repeated Guests and Cancellation



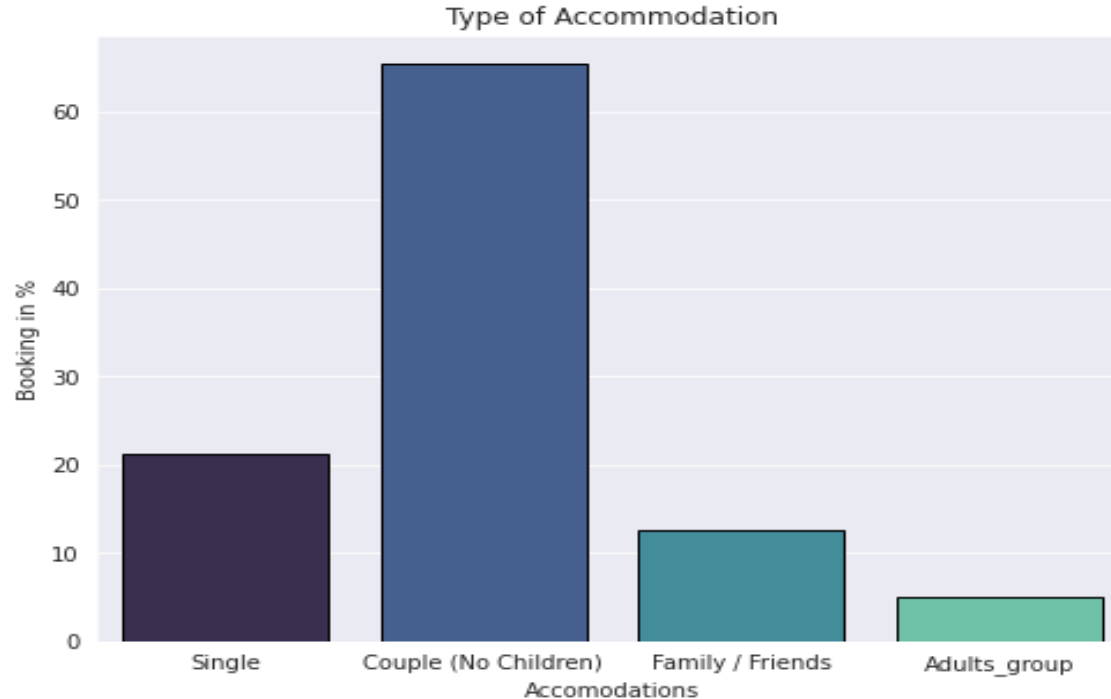
From the graph, it is clear that Repeated guests hardly cancel their bookings

## Q1. Which type of hotel is most likely to be cancelled?



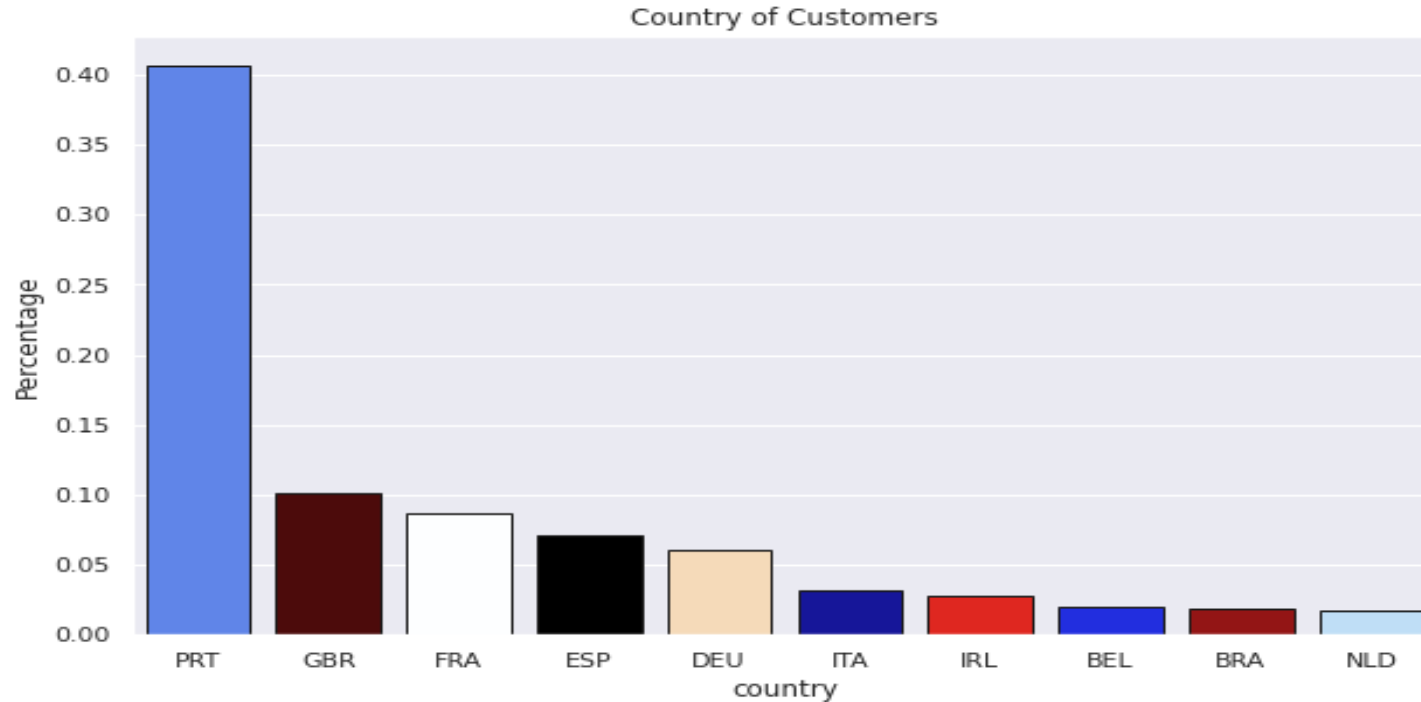
City hotels has higher cancellation rate than Resort hotel

## Q2. Which was the mostly booked accommodation type?



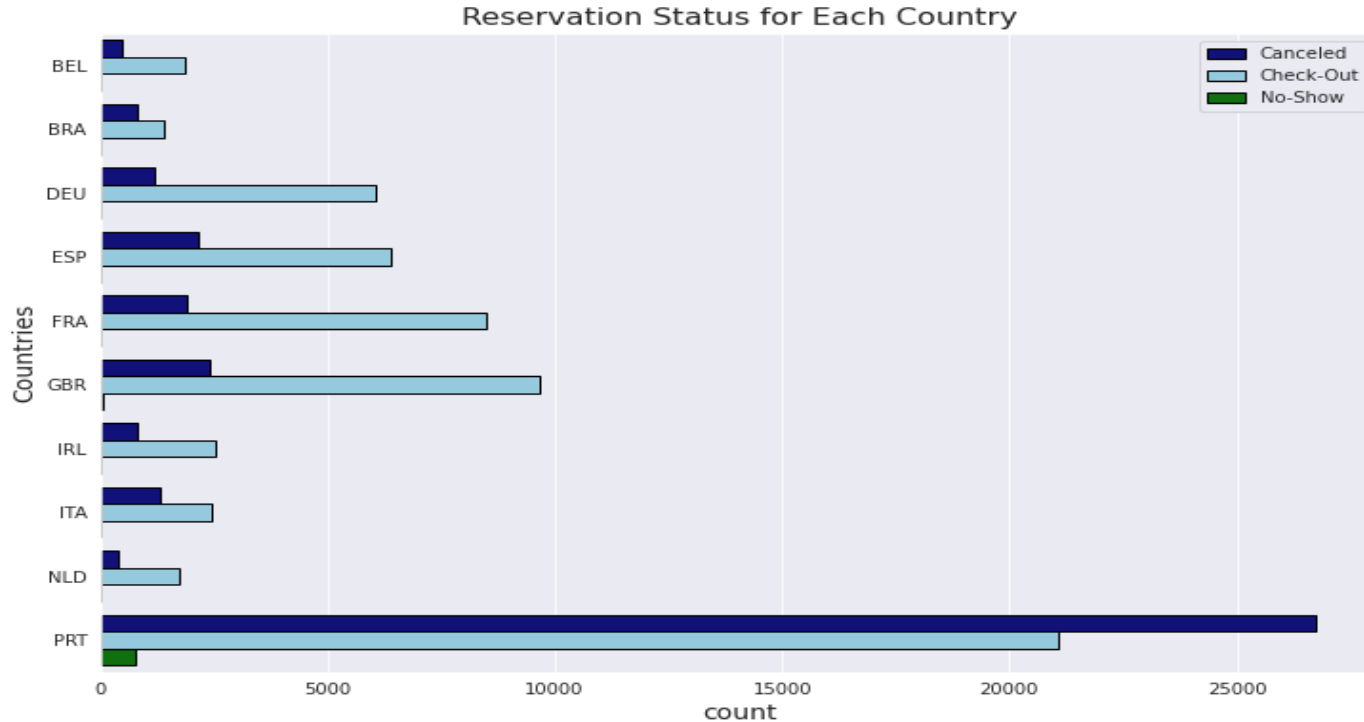
The graph shows that Couples visited the hotel the most

### Q3. Where do the customers come from?



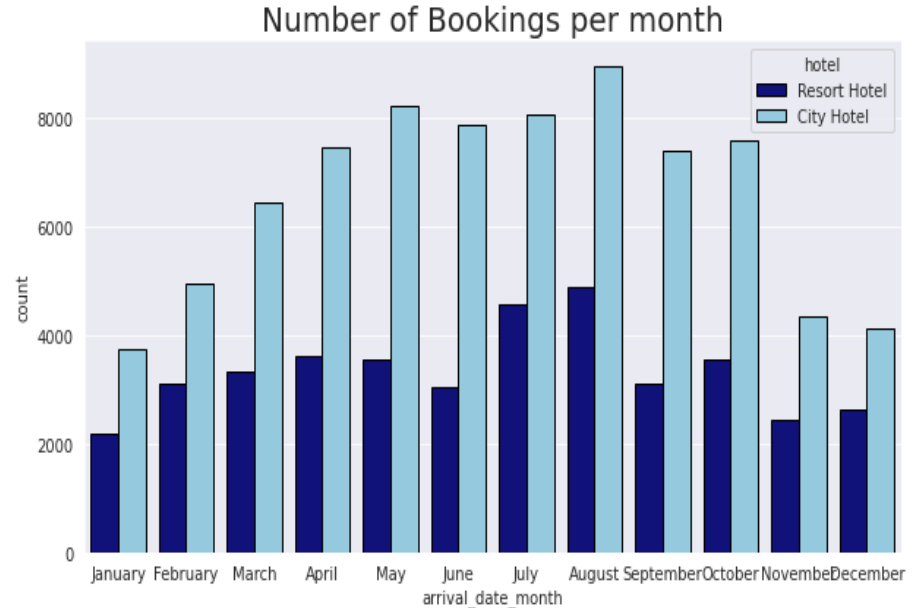
It is depicted from the graph that most of the customers are from Portugal

## O4. Which country make the highest number of reservations?



From the graph it is clear that Portugal has highest reservations. But unfortunately it also has more cancellations than check-outs

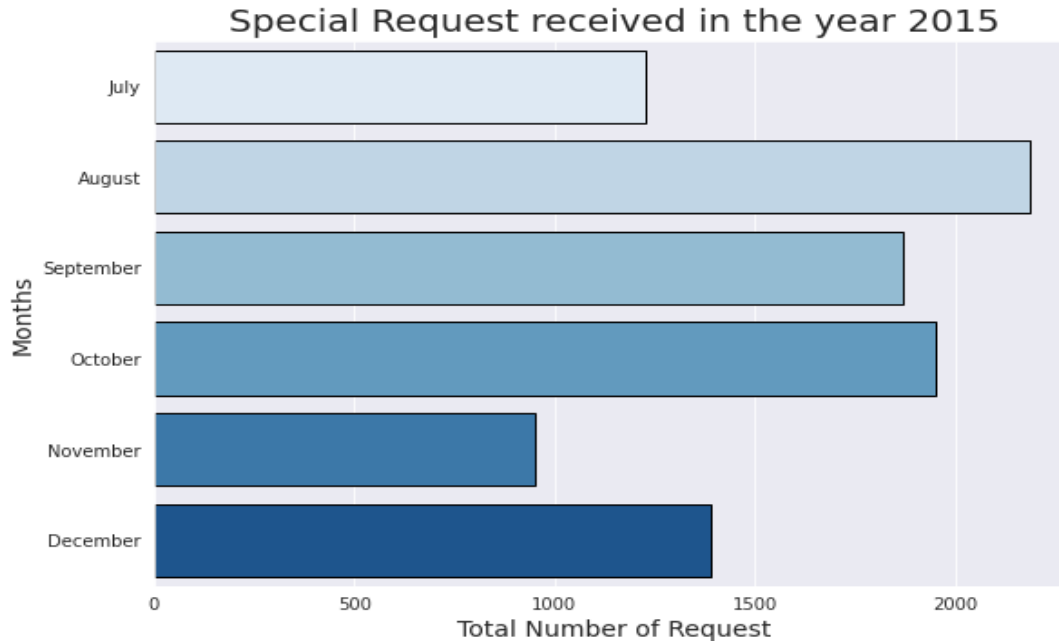
## Q5. Which year and month has the maximum number of bookings?



Most of the bookings were done in the year 2016. City Hotel has maximum bookings. August, May and July are the most occupied months for City Hotel.

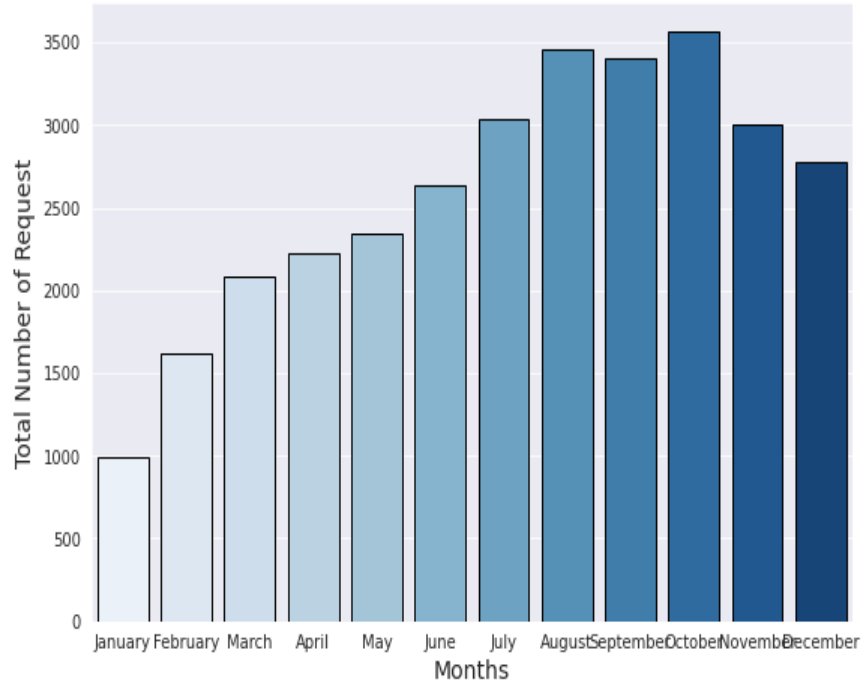


## O6. Which month has the maximum number of special requests per year?

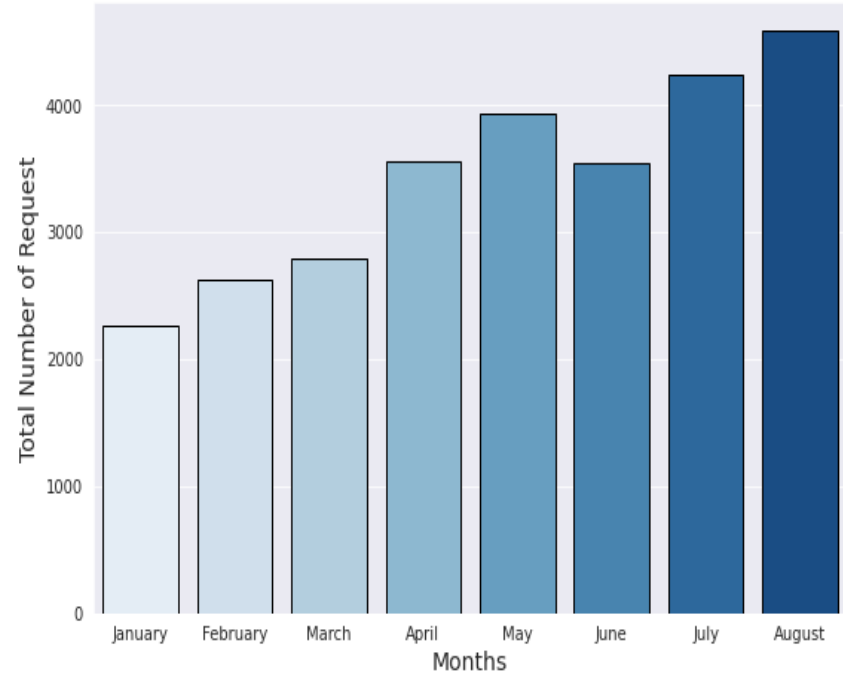


In the year 2015, hotels received maximum number of special requests in the month of August followed by October and September.

Special Request received in the year 2016

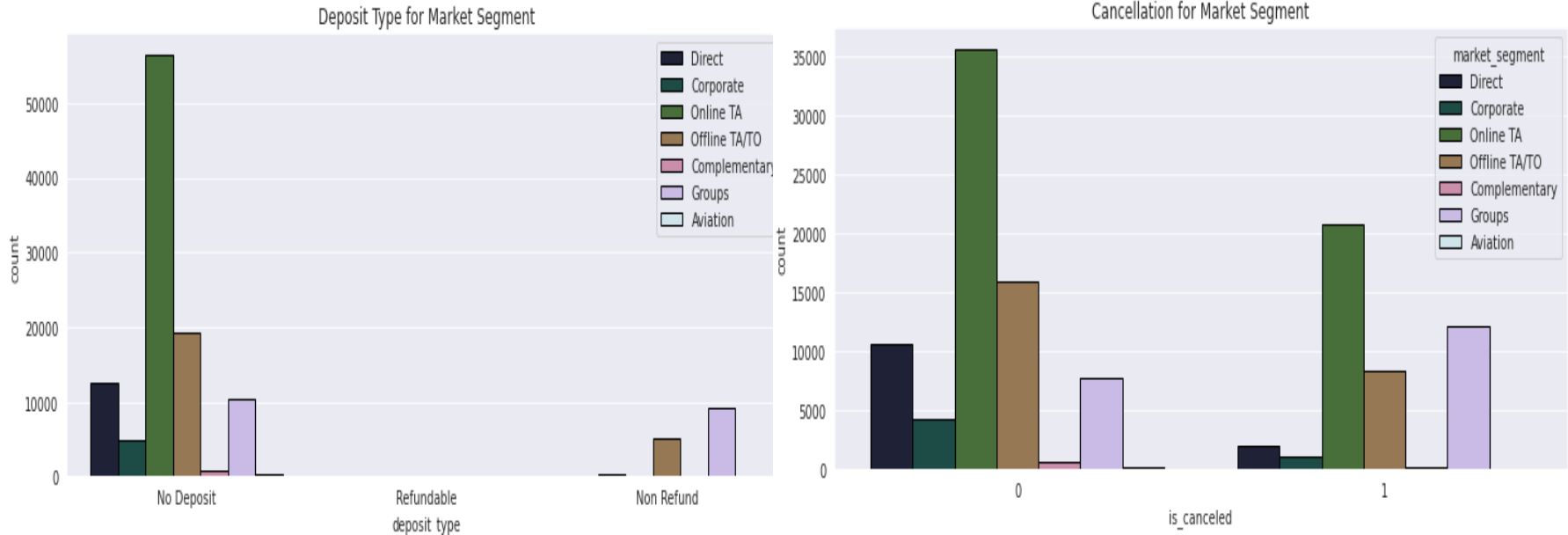


Special Request received in the year 2017



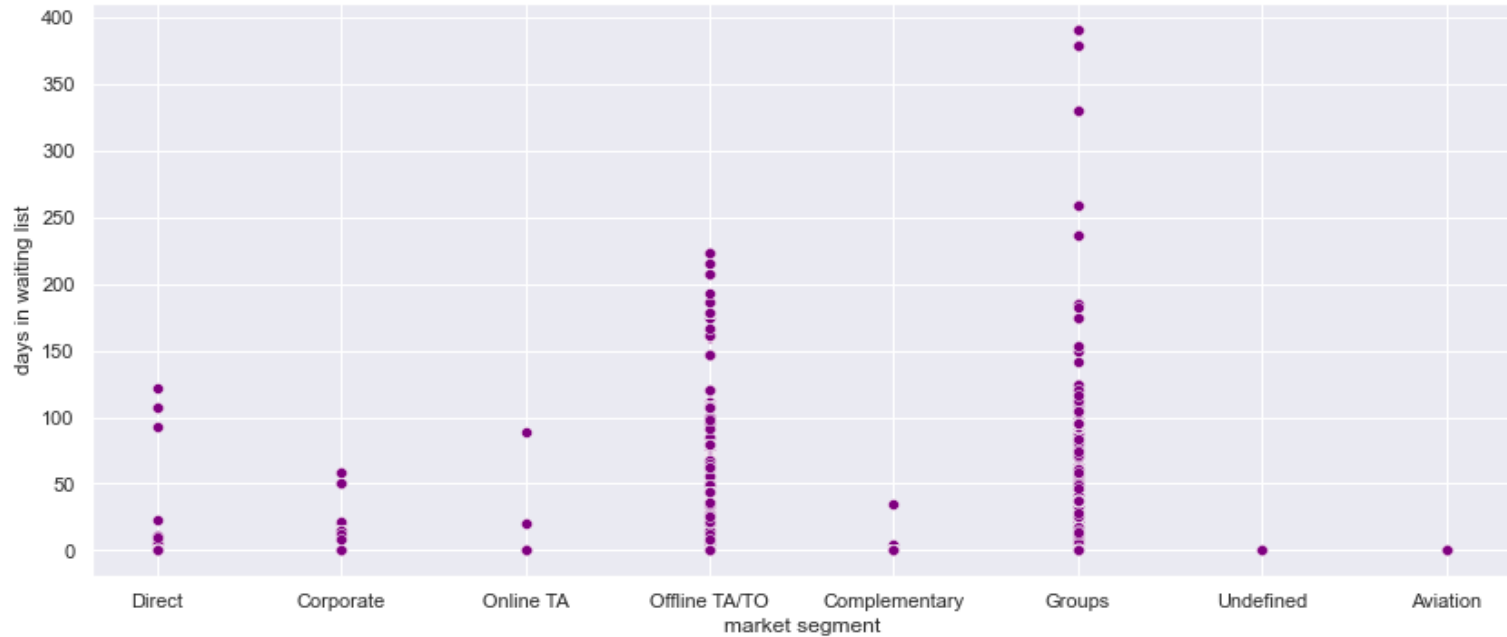
In the year 2016, October, August and September are the top 3 months with maximum special requests. Whereas In the year 2017, August, July and May are the top 3 months with maximum special requests.

## O7. What are the Deposits types and Cancellations for various types of Market Segments?



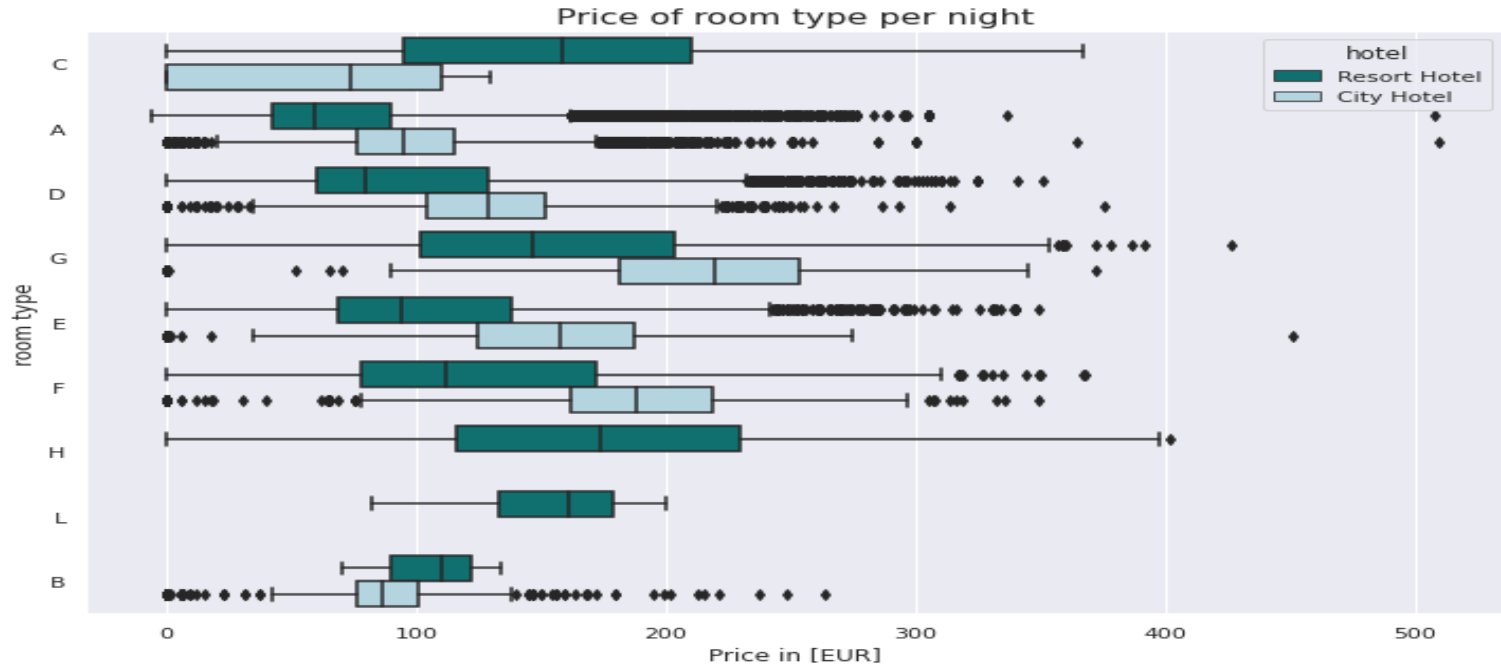
Online TA has the highest booking with No Deposit. Whereas Groups and Offline TA/TO has Non Refund booking. Online TA has the highest booking cancellation. Most of the Online TA provides the customers with facility to cancel their bookings without any charges.

## O8. Which Market Segment has the lowest number of days on waiting list?



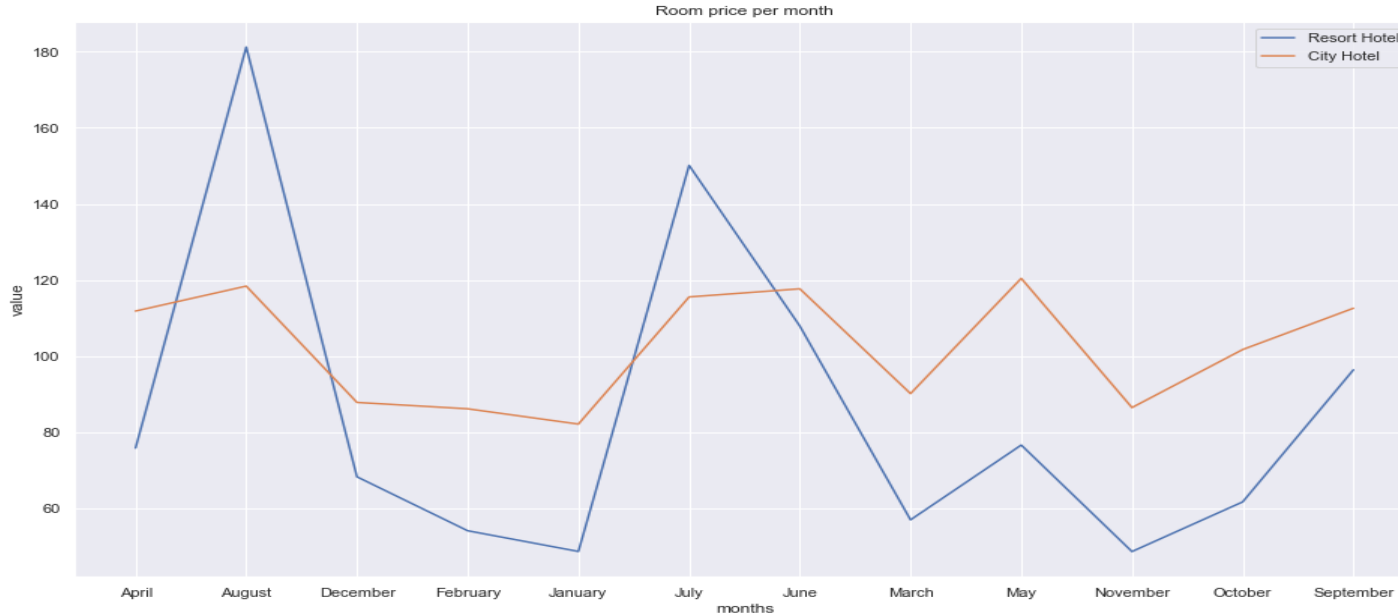
Aviation industry has minimum number of days on waiting list. Groups and Offline TA has highest number of days on waiting list.

## Q9. How much do customers pay for a room per night?



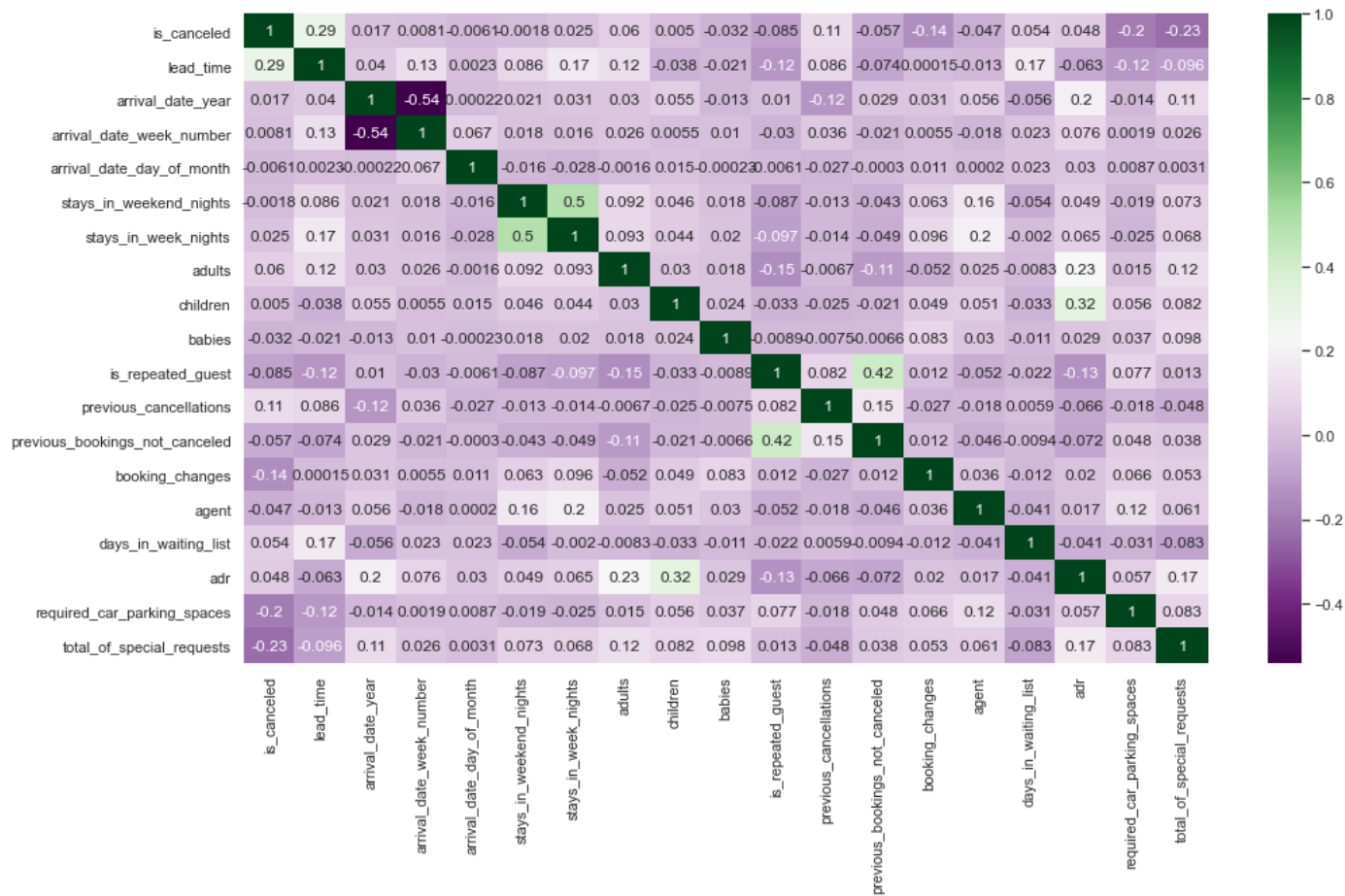
The graph shows that the room type A has highest price that is 500 euro. City Hotel prices are high for room type G.

## Q10. What are the prices per month for both the types of hotel?

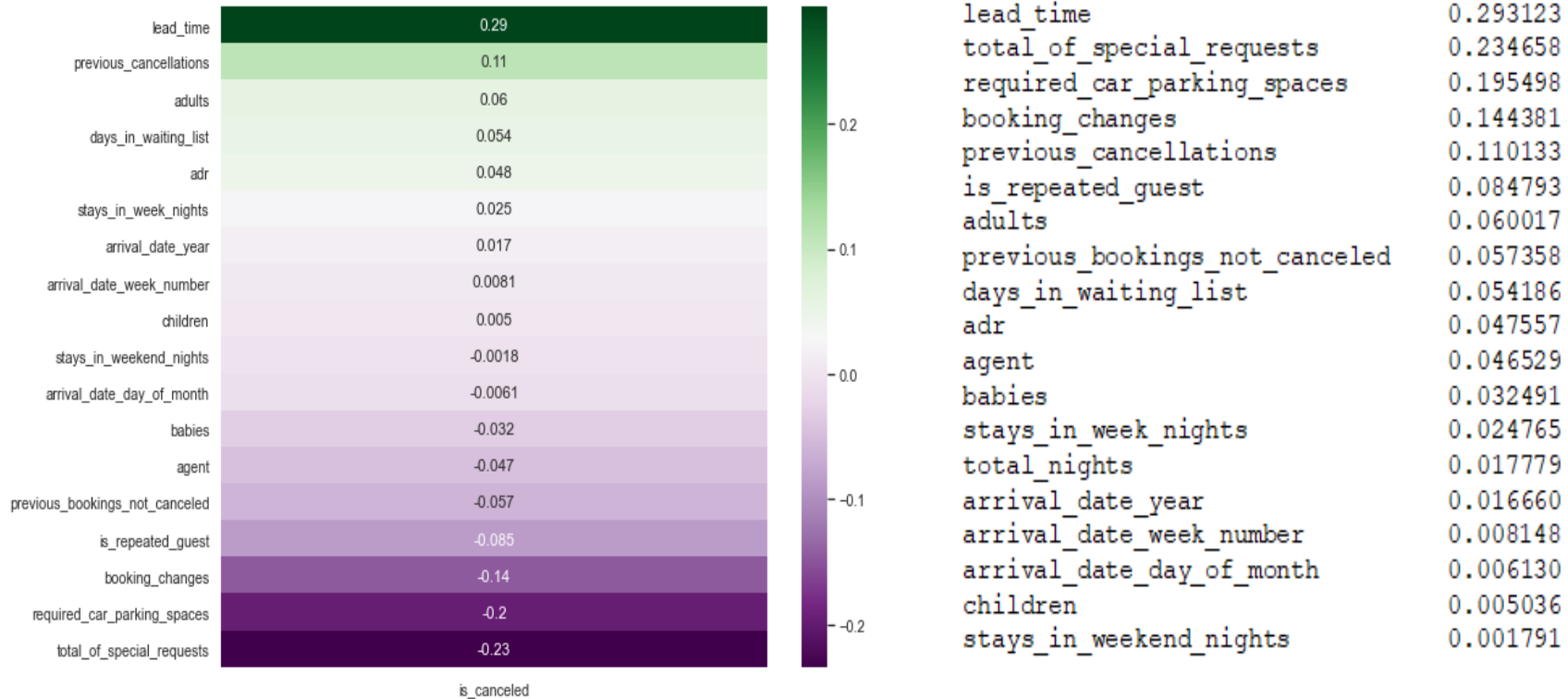


The graph shows that prices in the Resort Hotel are much higher during the months of August and July. Whereas City Hotels are most expensive during August, May and June.

# Correlations



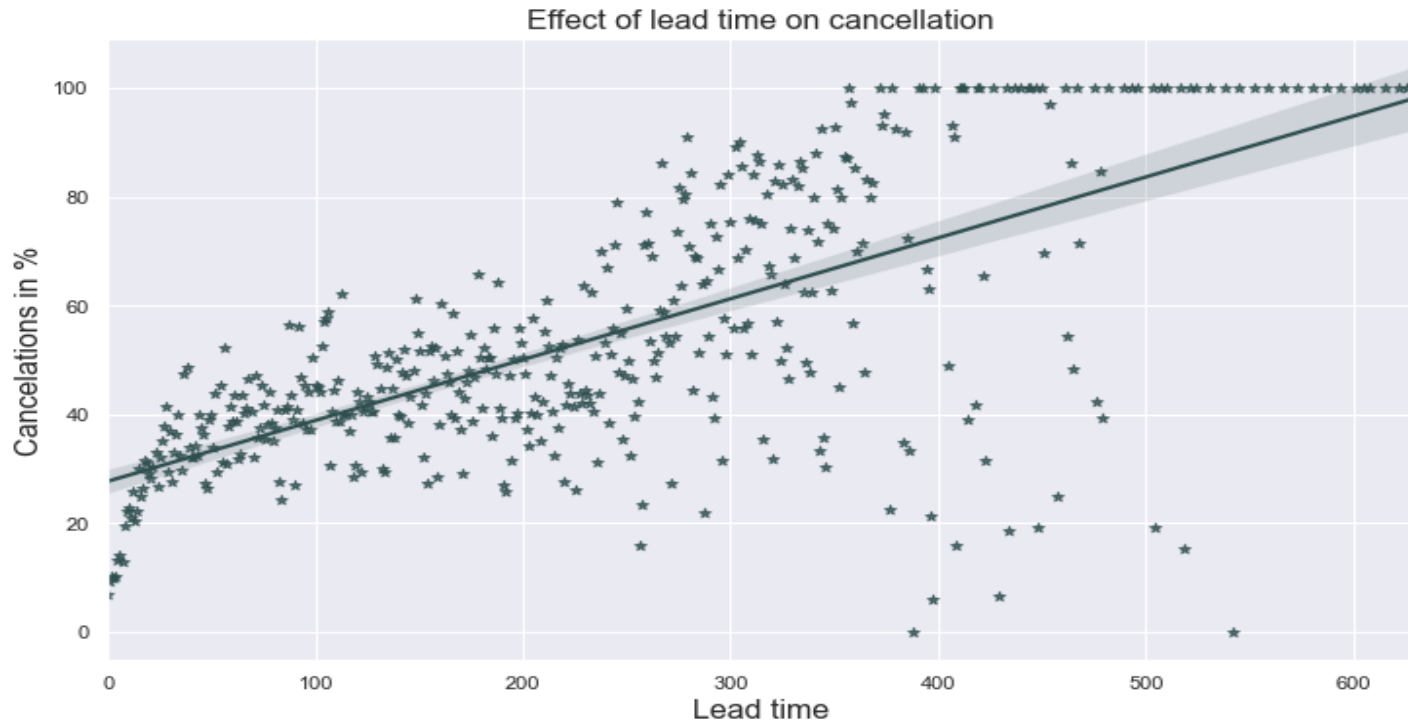
# Most Important factors for cancellation of bookings:





# Lead time and Cancellation rate:

What is the relationship between lead time and cancellation rate?



# Conclusion

After analysing the data here are some important conclusions:

1. Majority of the hotels booked are city hotel. Also City hotels are mostly cancelled than Resort hotels. It does not matter if parking space is available. Most customer do not consider it as a preference while booking hotels.
2. Majority of the reservations are made by couples.
3. Majority of the guests are from Western Europe. Highest number of guest from Southern Europe i.e. Portugal.
4. Most of the visitors aren't repeated guests.
5. The highest booking is using Online TA with no deposit. Also the highest booking cancellation is using Online TA.
6. Portugal has the highest number of reservations. But it also has high cancellation rate than check-outs.
7. Aviation industry has the minimum number of days on waiting list.
8. The prices varies with peak seasons, prices are higher for peak seasons and lower for off seasons

# Strategy

## Suggestions based upon insights to build a sound marketing strategy.

1. Cancellation rates are high for OnlineTA & Offline TA/TO channels across all the months. Collecting deposits, Setting non refund rates and Implementing more strict policies for booking cancellations will help hotel prevent loss of revenue.
2. Attracting more customers by implementing schemes such as Free Parking, Free Charging for Electric Vehicles. Hosting a new car launch event will attract potential customers.
3. Number of Cancellations can be reduced by adding non refund option in Festivals and Holiday seasons when there are chances of high bookings.
4. Since there are very less number of repeated guests, these customers should be targeted through advertisements.
5. Now a days almost all of the customers read hotel reviews and also considers them extremely important before making final reservation. Hence, improving hotel online reputation will help increase the booking rate.

**Thank You**