# Capstone Project

# Hotel Booking Analysis

BY

## Suneel Nelaturi

**Data Science Trainee | AlmaBetter**

# OVERVIEW

- INTRODUCTION
- DATA SUMMARY
- DATA CLEANING
- DATA VISUALIZATION
- OBSERVATION
- CONCLUSION

# INTRODUCTION

❖ Hotel industry is a very volatile industry and the booking depends on many factors.

❖ Hotel booking provides both online and offline that helps to allowing the guests to staying in the hotel

❖ The main objective behind this project is to explore and analyse data to discover important factors that govern the bookings and give insights to hotel management, which can perform various campaigns to boost the business and performance.

❖ We are here to explore a hotel booking dataset to discover important factors that govern the bookings.

# PROBLEM STATEMENT

➔ Have you ever wondered when the best time of year to book a hotel room is?Or the optimal length of stay in order to get the best daily rate?

➔ What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

➔ This hotel booking data set can help you explore those questions!

# DATA SUMMARY

**Shape of The Data Set :**

★ Number of Rows in the given Dataset is : 119390

★ Number of columns in the given Dataset is : 32

# COLUMN DETAILS

**hotel**: The names of the hotel are City Hotel and Resort Hotel

· **is_canceled**: Cancellation type, if the booking was cancelled or not.

Which takes 2 values 0 and 1.

0 indicates not cancelled.

·         **lead_time**: Time between reservation and actual arrival.

·         **arrival_date_year**: Year of arrival date

·         **arrival_date_month**: Month name of arrival date.

·         **arrival_date_week_number**: Week number of arrival date

·         **arrival_date _day_of_month**: Day of the month of arrival date

# COLUMN DETAILS

**stays_in_week_nights**: Number of week nights the guest stayed or booked to stay at the hotel

**adults**: Number of adults

**children**: Number of children

**babies**: Number of babies

**meal**: Type of meal booked

**country**: Country of origin of customer

**market_segment**: Market segment designation.

In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators".

**distribution_channel**: The medium through booking was made

# COLUMN DETAILS

**is_repeated_guests**: Value indicating if the booking name was from a repeated guest (1) or not (0)

**previous_cancellations**: (0 or 1) Indicates whether or not the guest has previous cancellations

**previous_bookings_not_canceled**: (0 or 1) Indicates whether or not the guest has previous bookings which are not cancelled.

**reserved_room_type**: Code of room type reserved.

**assigned_room_type**: Code of room type assigned.

**booking_changes**: Number of changes made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

**deposit_type**: Whether refundable/non-refundable/no-deposit made

**agent**: ID of the travel agency that made the booking

# COLUMN DETAILS

**customer_type**: Type of customers (Transient, group, etc.)

**adr**: Average daily rate is the average revenue that a hotel receives for each occupied guest room per day

**required_car_parking_spaces**: Number of car parking spaces required

**total_of_special_requests**: Number of special requests made

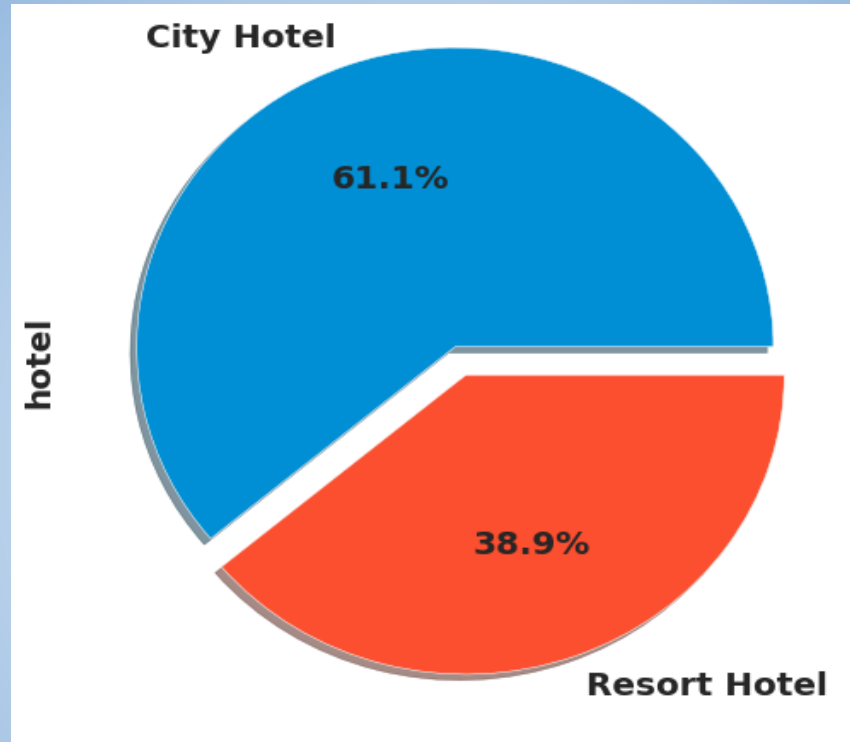**reservation_status**: Reservation last status

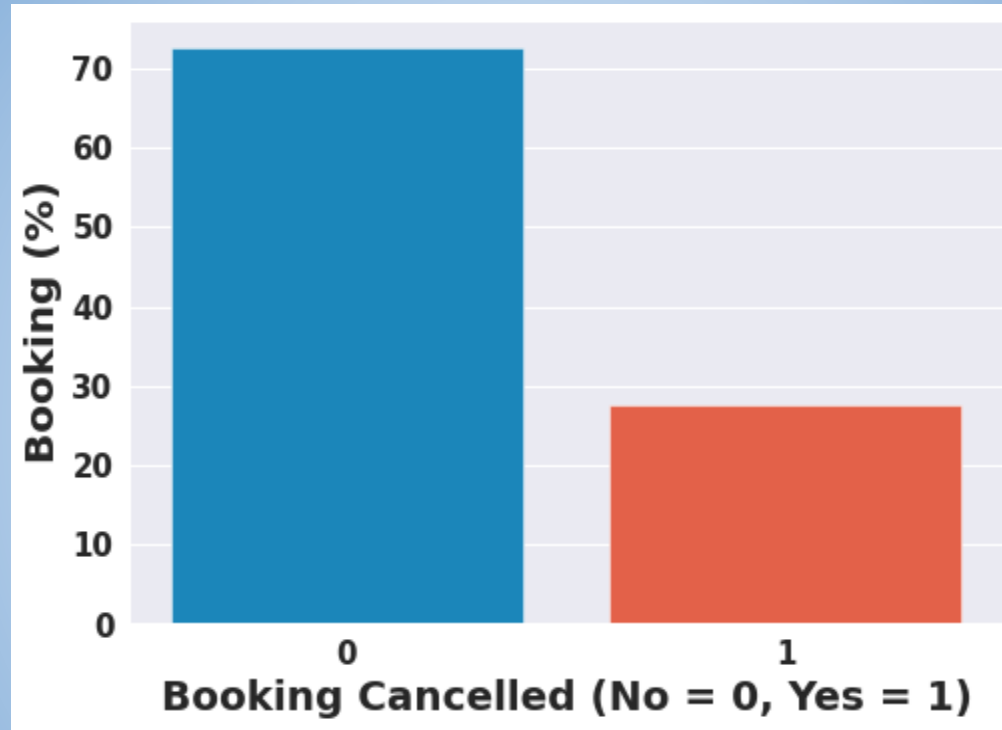**reservation_status_date**: Date of last reservation status

AI

# DATA VISUALIZATION

Here we are using Matplotlib and Seaborn libraries for the following graphs and plots had been used.

★ **Bar Plot**

★ **Histogram**

★ **Pie Chart**

★ **Line plot**

★ **Heatmap**

★ **Box Plot**

AI


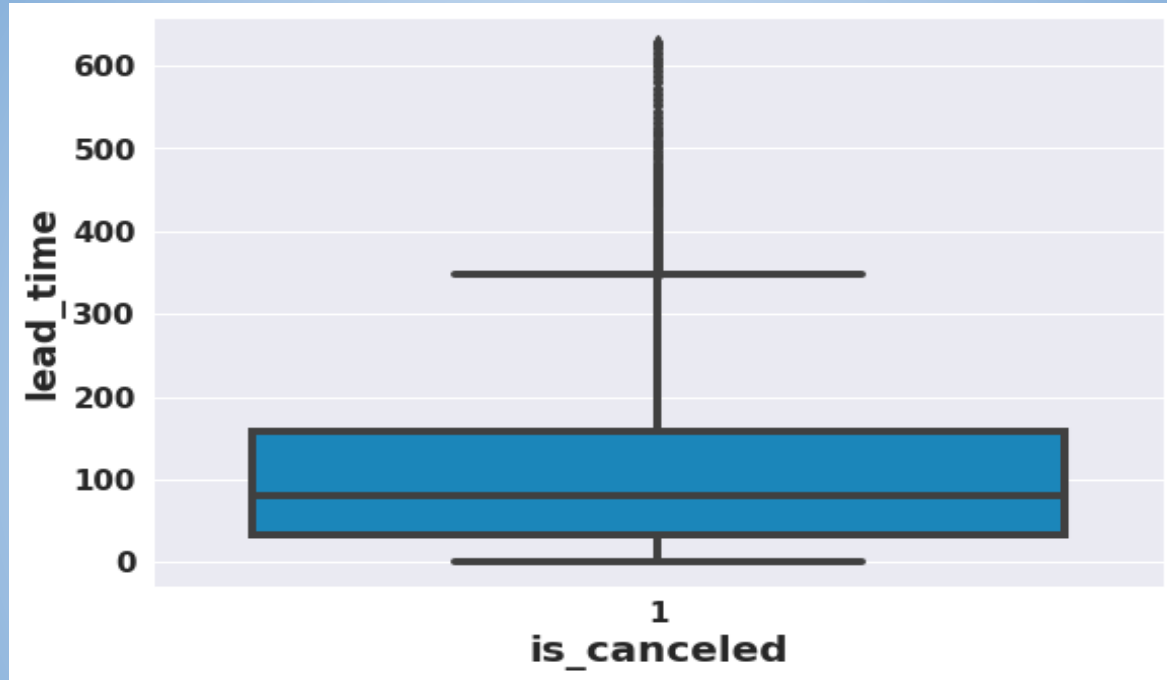
More than 60% of the bookings are of City Hotel.

# CANCELLATION OF BOOKING

**AI**

More than 70% of the people did not cancel the booking.

**AI**



The lead time and cancelation are co–related to each other
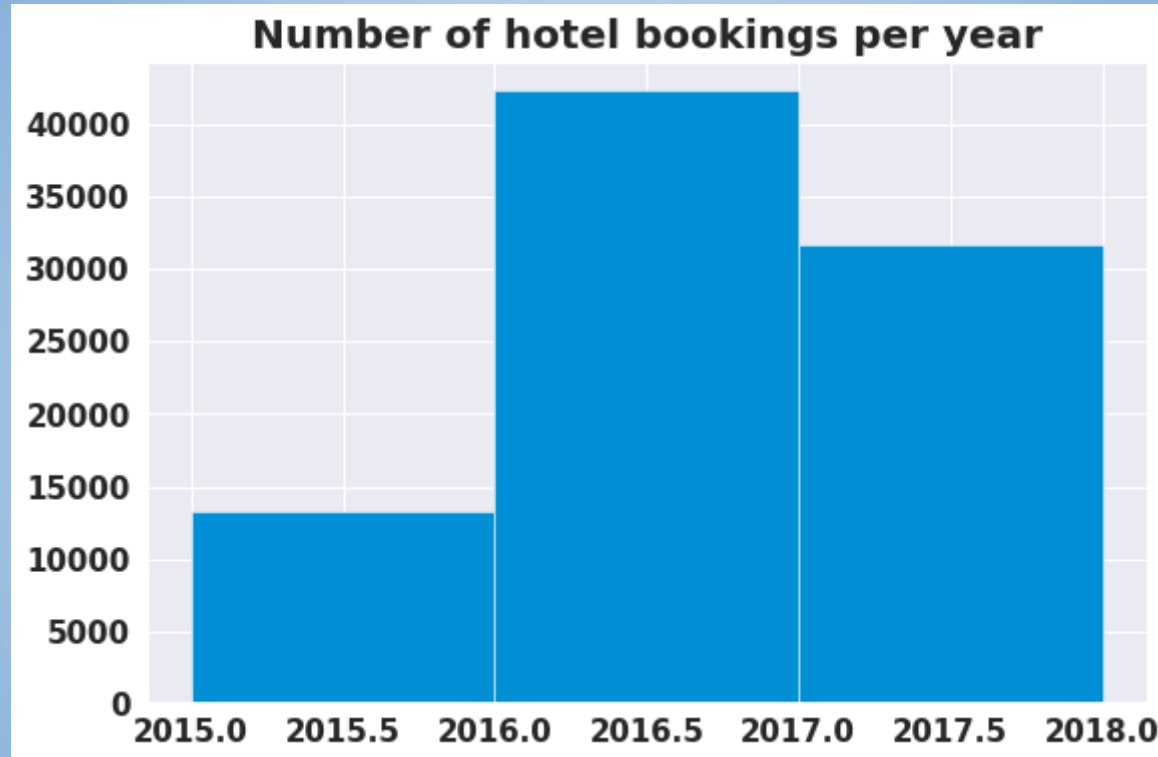When lead time increases, the chances for cancellation increases.

**AI**



Number of canceled Bookings

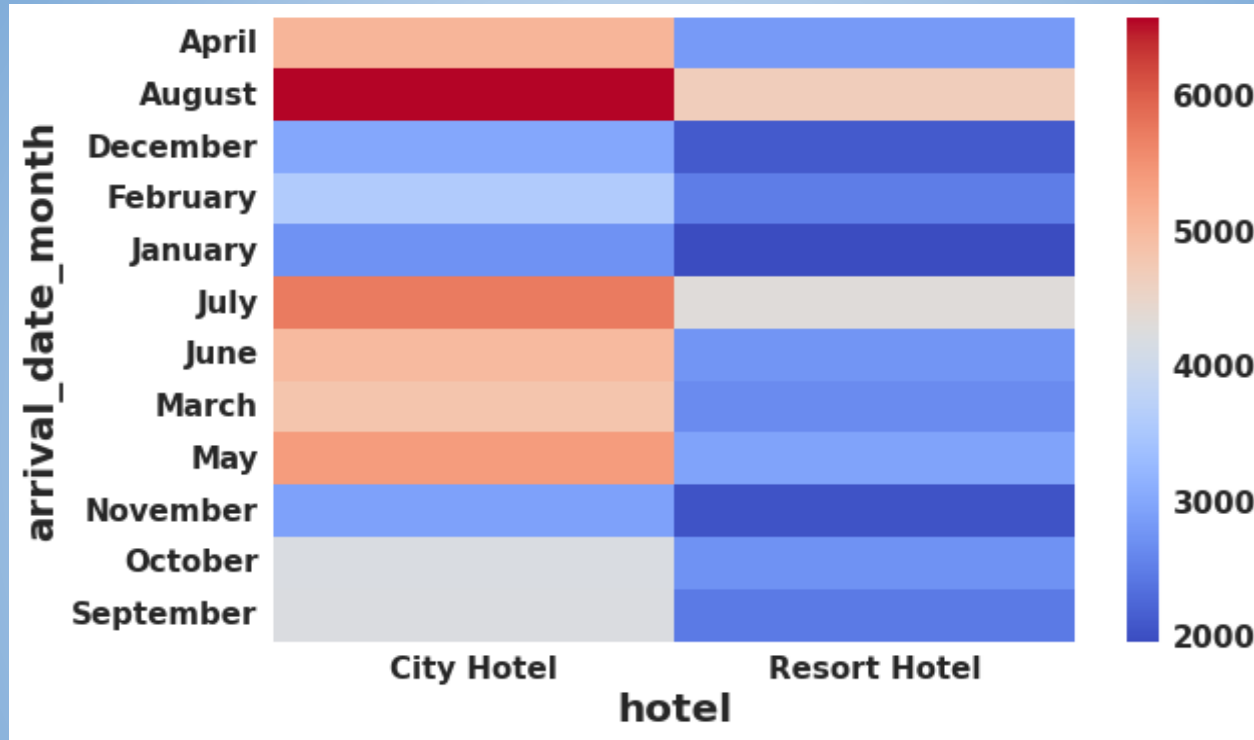City hotel has more number of cancelation compared to Resort hotel.

❖ **Most of the people prefer to reserve city hotel & most of the cancellations occur in city hotel.**

**Number of hotel bookings per year**

hiighest number of bookings are occurred in 2016

AI



August is the most busiest month in the  year.  Also the months after september is  the least demanded month in the year

# MOST BUSIEST MONTHS

Total no of guests per Months

Both city hotel and resort hotel have more guests in june,july and august months.
After august,  the number of guests are decreasing from September to December.

# PRICE VARIATION



Room price per night over the Months

From the above plot the price in the resort hotel is much higher during the summer and the price of city hotel varies less and is most expensive during summer and spring.

# TOP 10 COUNTRIES WITH MAX NUMBER OF PASSENGERS



country-wise comparison
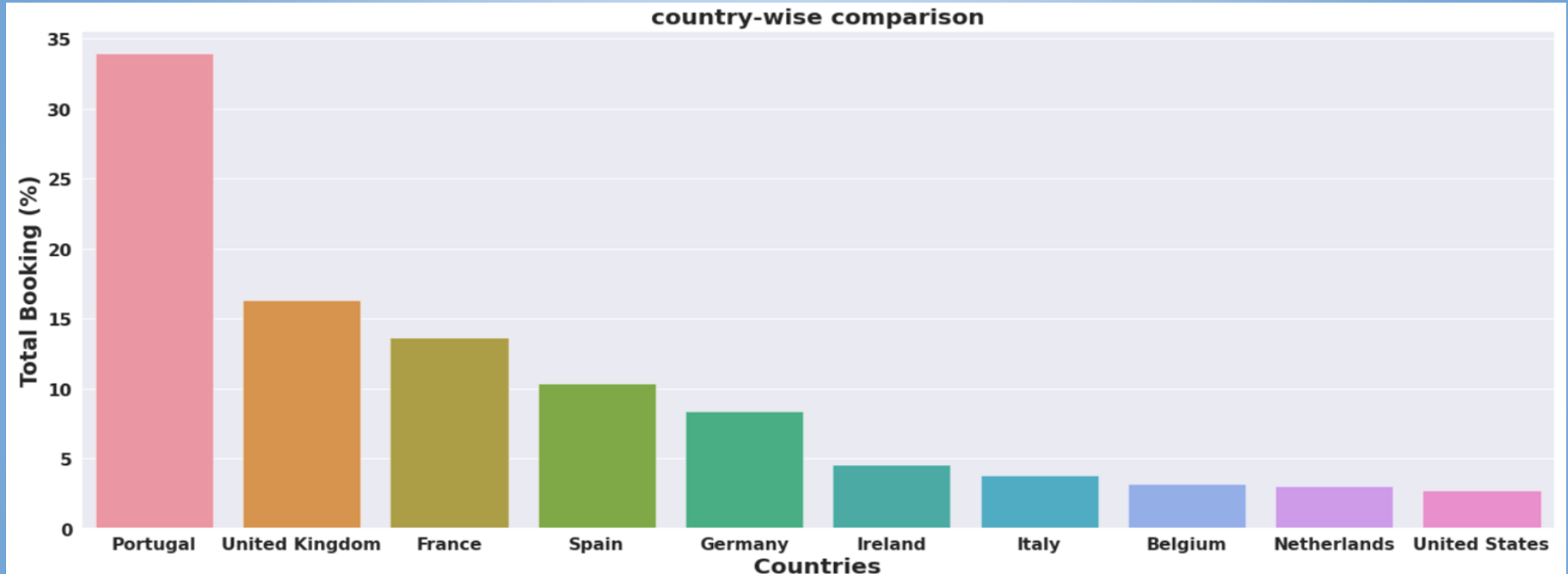
Most guests are from Portugal and other countries in Europe.

# REPEATED GUESTS



**Percentage of Repeated Guests**
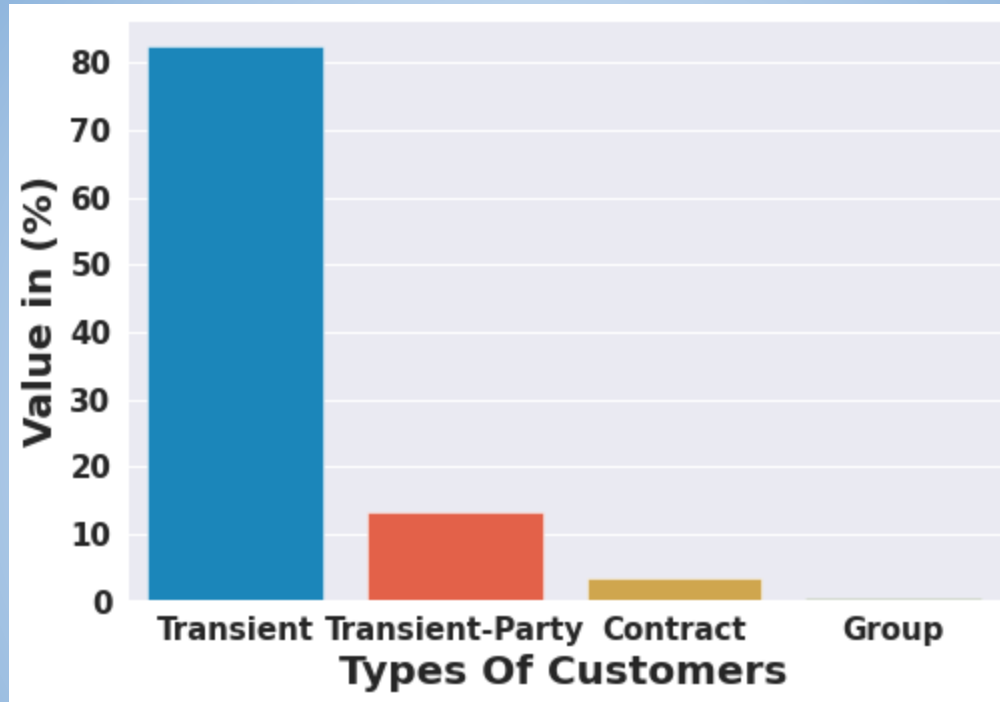
Guest Not Repeated — 96%
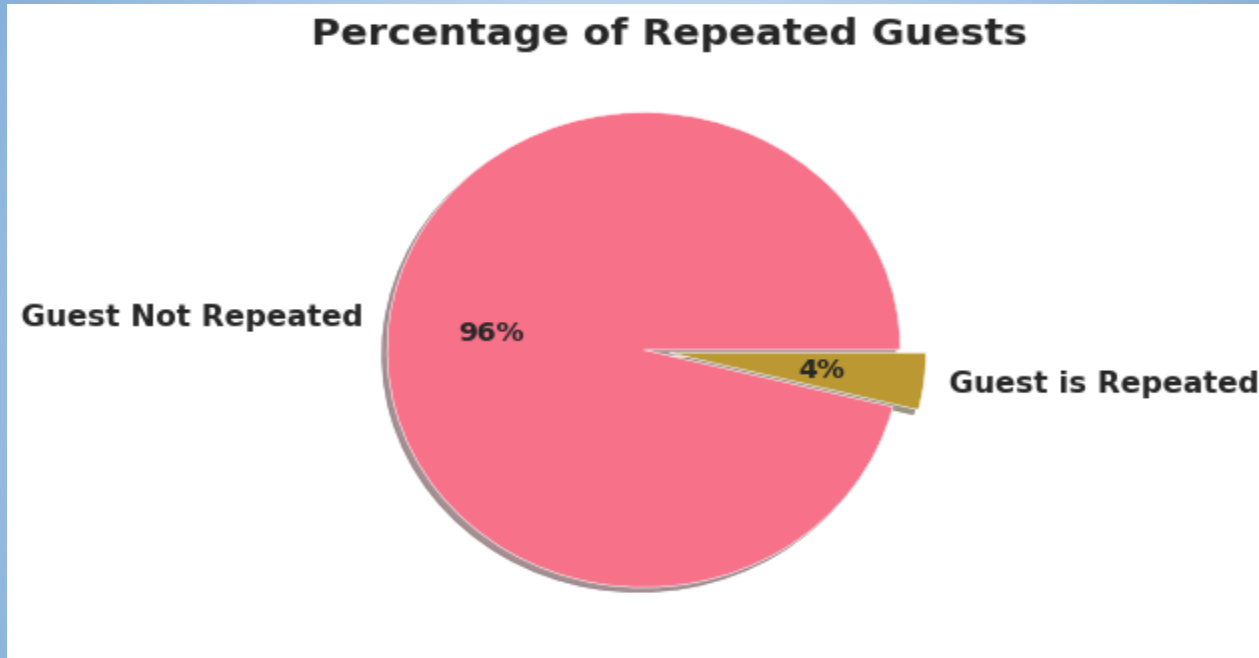
4% — Guest is Repeated

The chart shows only 4% guest is repeated in hotel.It means very low guests repeated to visit same hotel, i.e 3358 guests out of 83981 is repeated.

Type of Accommodation( Single vs Couple vs Family)
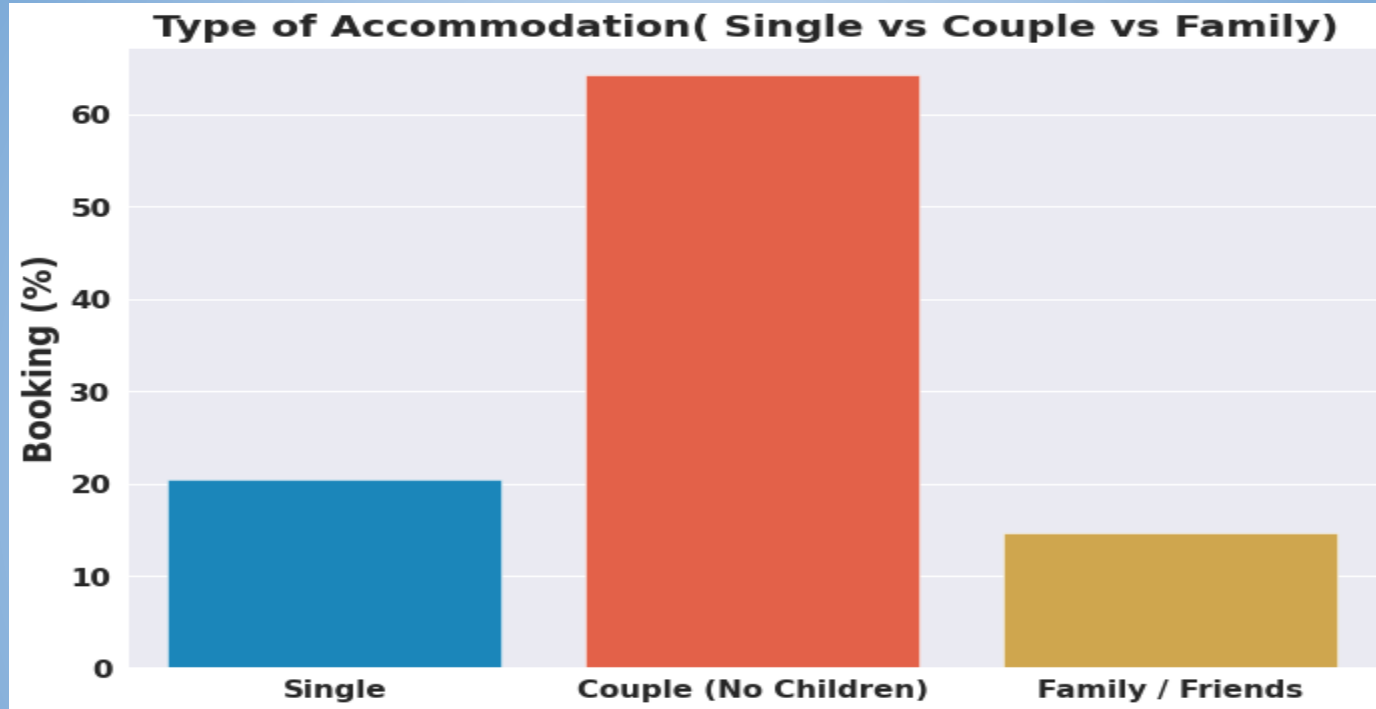
Around 65% of the people visit with their family. While most of the visitors are couples.
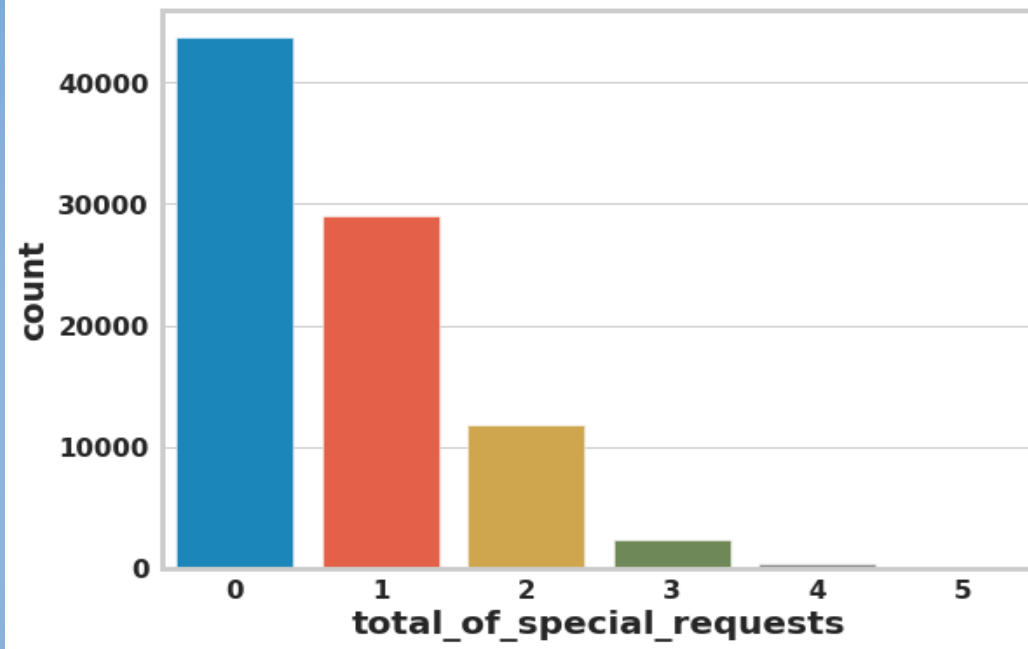
**AI**



Most customers are of type Transient.

Hotel wise night stay duration (Top 10)

In Resort hotel people like to stay 1 day and in City hotel people like to stay 1-4 days. As number of days increases, people prefer Resort hotel the most.
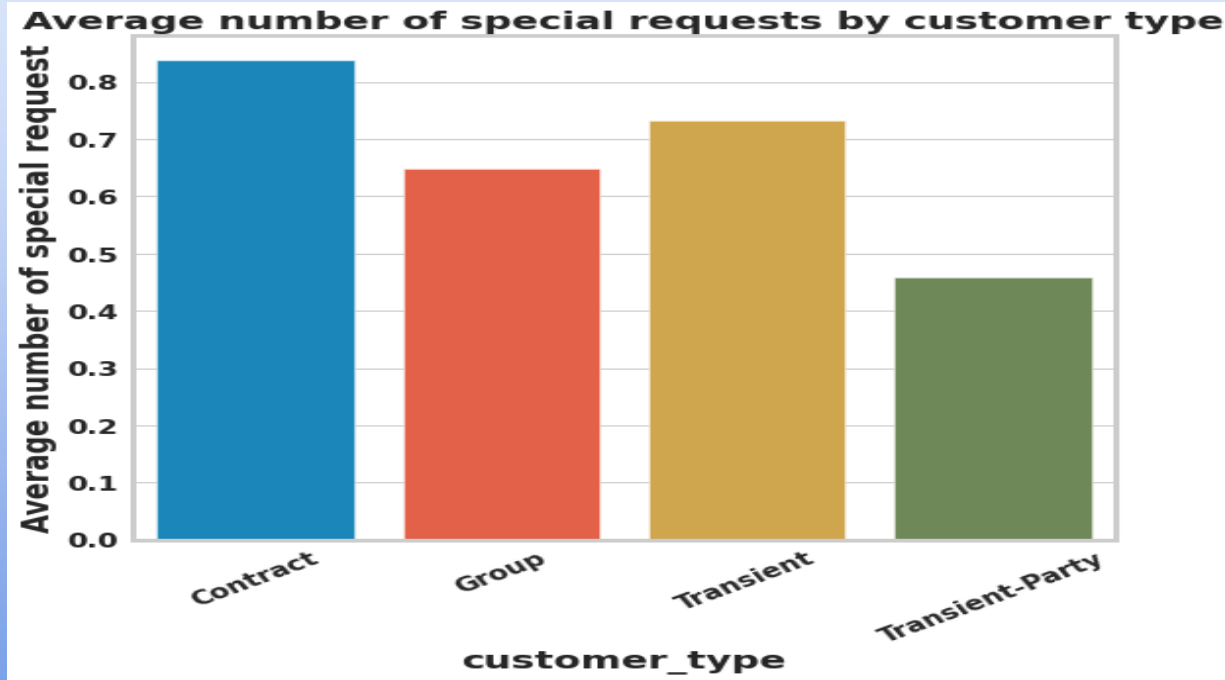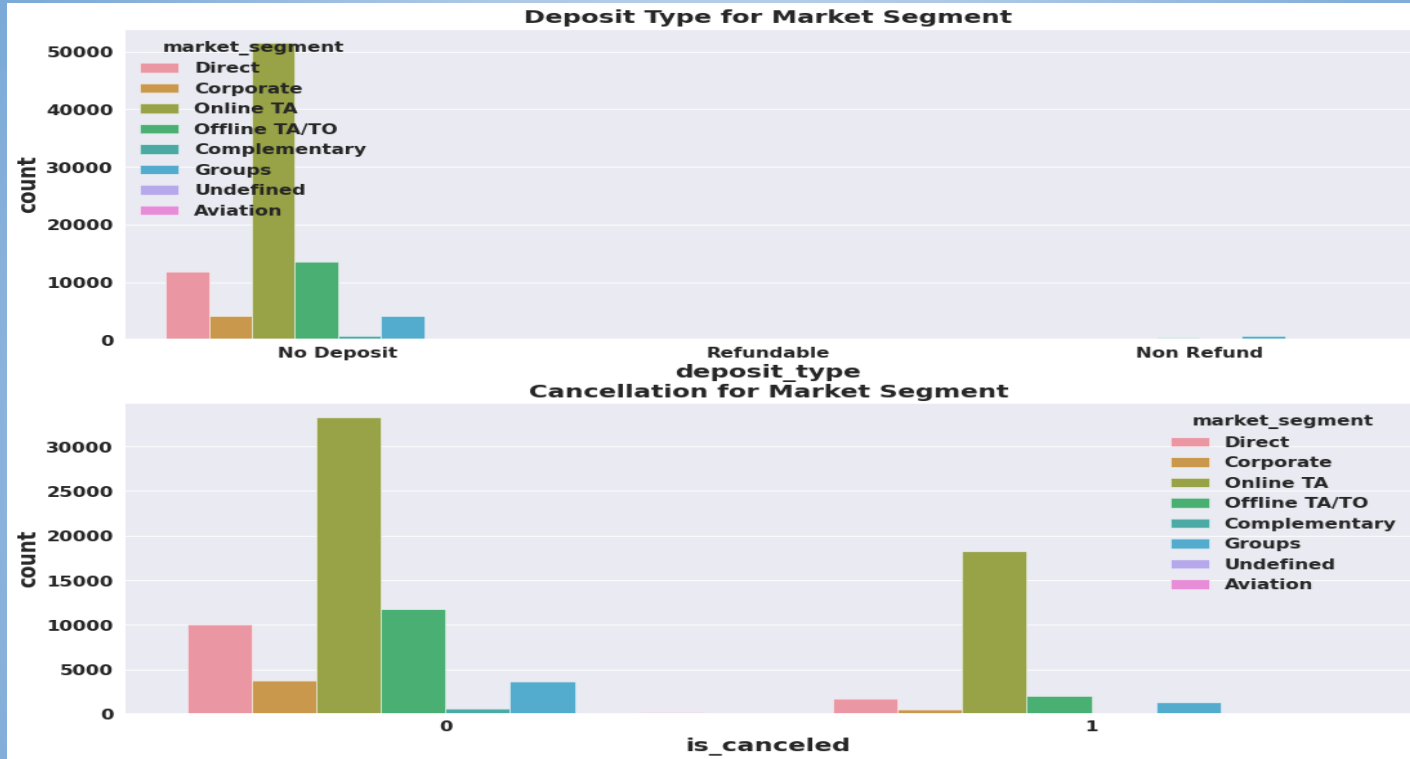
Here the number of 0 special requests is the highest. After the number of 1 special requests for booking

SPECIAL REQUESTS

Average number of special requests by customer type

They are not entertaining much special requests. Contract people have more special requests.

Most of the bookings are done through online TA segment and most cancelation is done through online TA segment only.People don't want to pre-deposit  the money

# DISTRIBUTION CHANNEL



The most preferred distribution channel is TA/TO.

## CONCLUSION

- ➜ We found out there is huge demand for City Hotel and Resort Hotel in months of June, July and August. In other side, we found out that the prices in the above months are higher. So the best time to book the hotel will be between Nov-Mar, where the demand is low and prices too. The user can get better price and room easily.
- ➜ In a business perspective, we should focus more on summer season due to high demand.
- ➜ If people stay for less than 5 days, they will get the best price.
- ➜ They are not entertaining much special requests.
- ➜ Most of the bookings are coming through online.
- ➜ Most of the bookings are of 2 adults.

# CONCLUSION

→ When users book the hotel much earlier, they are more likely to cancel the booking due to various reasons.So we have to reduce lead time to decrease cancellation rate.

→ Majority of guests are from Western Europe.

THANK YOU!