Capstone Project

EXPLORATORY DATA ANALYSIS OF HOTEL BOOKING.

PRESENTED BY

SHANTANU CHOUDHARY

Agenda

❑ Introduction.

❑ Why analyse the hotel bookings ?

❑ Problem Statement

❑ About the Dataset

❑ Dataset Preparation

❑ Attributes in Hotel Booking Analysis

❑ Exploratory Data Analysis

❑ Conclusions

Why analyse Hotel Bookings?

Analyzing hotel bookings in the context of hotel booking analysis is crucial for hotel managers who want to make data-driven decisions to improve their business. By analyzing hotel bookings, managers can identify trends such as peak booking periods, popular room types, and booking channels. This information can help managers make better decisions on pricing, inventory management, and marketing strategies. Hotel booking analysis also helps managers understand the factors that influence booking behavior, such as seasonality, events, and competition, which can be used to optimize pricing strategies to maximize revenue. Additionally, hotel booking analysis can provide insights into customer preferences and behavior, helping managers personalize the customer experience, improve customer satisfaction, and build customer loyalty.

Problem Statement

The hotel industry is highly competitive, and hotels of all sizes are looking for ways to optimize their revenue management strategies. One key area of focus is understanding guest behavior and preferences when it comes to booking and canceling reservations. By analyzing hotel booking data, we can identify patterns and trends that can help hotels better anticipate cancellations, adjust their pricing strategies, and increase their occupancy rates. Specifically, this project aims to explore the factors that influence cancellations and occupancy rates in order to help hotels make data-driven decisions about their pricing and marketing strategies. By doing so, we hope to provide valuable insights that will help hotels improve their business performance and stay ahead in an increasingly crowded market.

About The Dataset

The dataset includes data on reservations made for two hotels, a resort and a city hotel, both of which are expected to open between July 1, 2015, and August 31, 2017. The identical data was gathered for both hotels: 32 variables were used to describe 40,060 observations for the resort and 79,330 observations for the city hotel. 119,390 hotel reservations, including those that were cancelled, are included in the dataset. All components that could identify hotels or clients were removed because this is genuine data.

Dataset Preparation

- **Load two datasets**: one containing hotel information and amenities, and the other containing customer reviews.

- **Import necessary libraries** like NumPy, Pandas, Seaborn, Matplotlib, Folium, and Plotly Express (pltx) to facilitate data analysis.

- **Data Cleaning & Feature Engineering:** During data cleaning and feature engineering, we remove null values, outliers, and duplicate data to ensure accuracy and consistency. Additionally, we create new features to improve model performance by extracting meaningful information from existing features or external data sources.

- **Perform exploratory data analysis** to gain insights into the data sets.

- Use **statistical graphics** and **data visualization techniques** such as scatter plots, bar charts, heat maps, and interactive maps created using Folium to summarize the main characteristics of the data sets.

Attributes In Hotel Booking Analysis.

- **hotel**: indicates the type of hotel i.e. (City or Resort Hotel)

- **lead_time**: the number of days between the booking date and the arrival date

- **arrival_date_year**: the year of arrival date

- **arrival_date_month**: the month of arrival date

- **arrival_date_week_number**: the week number of arrival date

- **arrival_date_day_of_month**: the day of the month of arrival date

- **stays_in_weekend_nights**: number of weekend nights (Saturday/Sunday) the guest stayed or booked to stay at the hotel

- **stays_in_week_nights**: number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

- **adults**: number of adults in the booking

- **children**: number of children in the booking

- **babies**: number of babies in the booking

- **meal**: type of meal booked (e.g. Full board, Half board, etc.)

- **country**: country of origin of the guest

- **market_segment**: market segment designation (e.g. Online Travel Agent, Offline Travel Agent, etc.)

- **distribution_channel**: distribution channel through which the booking was made (e.g. Travel Agents, Direct, etc.)

- **is_repeated_guest**: binary variable indicating if the guest is a repeated guest or not

- **previous_cancellations**: number of previous bookings that were cancelled by the same guest prior to the current booking

- **previous_bookings_not_canceled**: number of previous bookings not cancelled by the same guest prior to the current booking

- **reserved_room_type**: code of room type reserved

- **assigned_room_type**: code for the type of room assigned to the booking

- **booking_changes**: number of changes/amendments made to the booking before or during the stay

- **deposit_type**: type of deposit made to secure the booking agent

- **company**: ID of the company/entity that made the booking or responsible for paying the bill

- **days_in_waiting_list**: number of days the booking was on the waiting list before it was confirmed to the customer

- **customer_type**: type of booking (e.g. transient, contract, group, etc.)

- **adr**: average daily rate (total booking revenue divided by total number of nights)

- **required_car_parking_spaces**: number of car parking spaces required by the guest

- **total_of_special_requests**: number of special requests made by the guest (e.g. extra beds, late check-in, etc.)

- **reservation_status**: current status of the booking (e.g. cancelled, checked-in, no-show, etc.)

- **reservation_status_date**: date at which the reservation status was last updated

- **is_canceled**: Containing values 0 and 1 where 0 represents non cancelled bookings and 1 representing cancelled bookings.

- **agent**: represents the identifier of the travel agent or booking agent who made the reservation on behalf of the customer.

Overview Of Analysis.

DATA CLEANING & FEATURE ENGINEERING → DATA EXPLORATION → INSIGHTS & CONCLUSION

Eda will be divided into 3 analysis.

- **UNIVARIATE ANALYSIS -** Univariate analysis is the simplest of the three analyses where the data you are analyzing is only one variable
- **BIVARIATE ANALYSIS -** Bivariate analysis is where you are comparing two variables to study their relationships.
- **MULTIVARIATE ANALYSIS:** Multivariate analysis is similar to Bivariate analysis but you are comparing more than two variables

Python Libraries Used

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

Import folium

from folium.plugins import HeatMap

import plotly.express as pltx

Python Libraries

Exploratory Data Analysis

- Which months over the course of 3 years had maximum arrivals for both hotels?
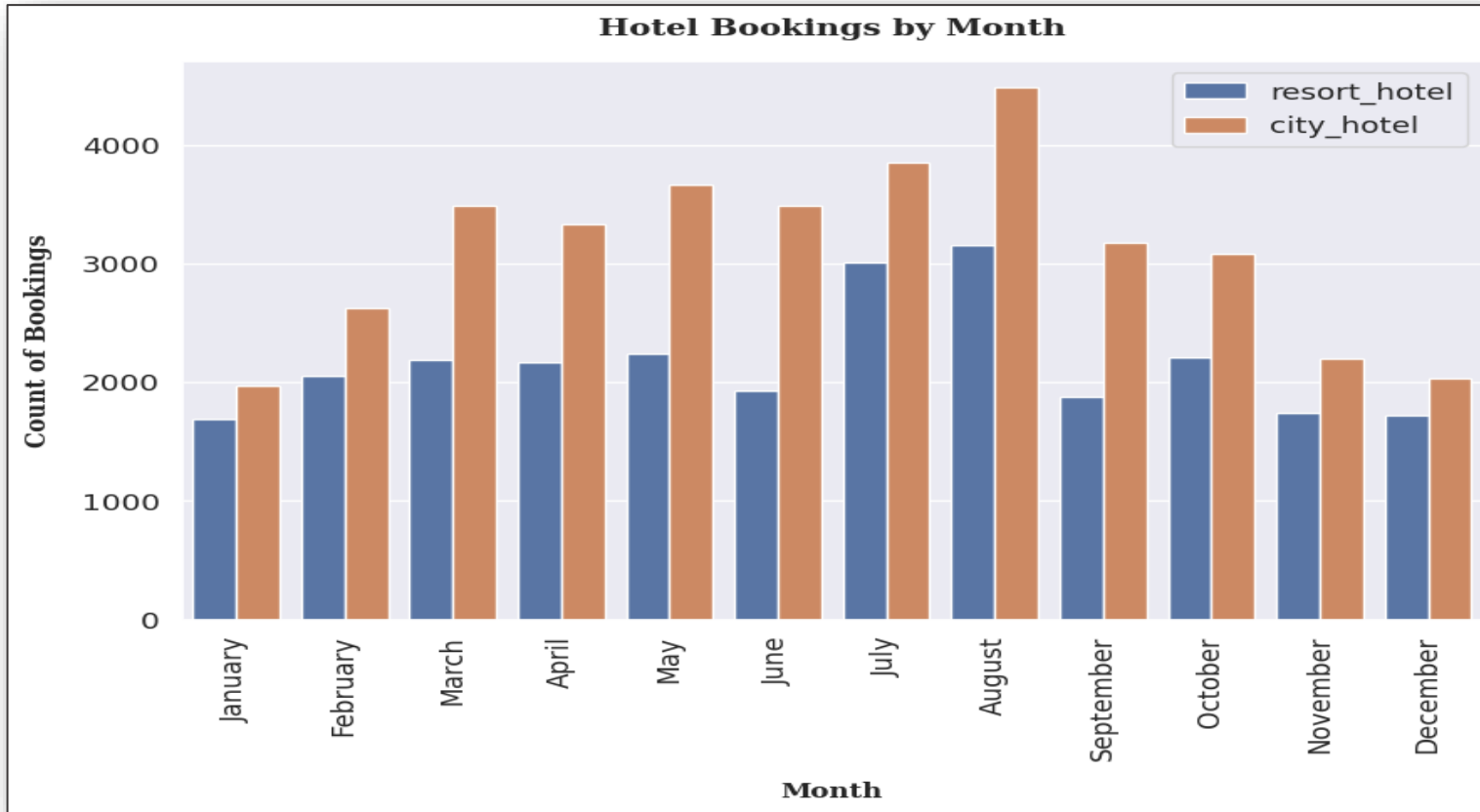
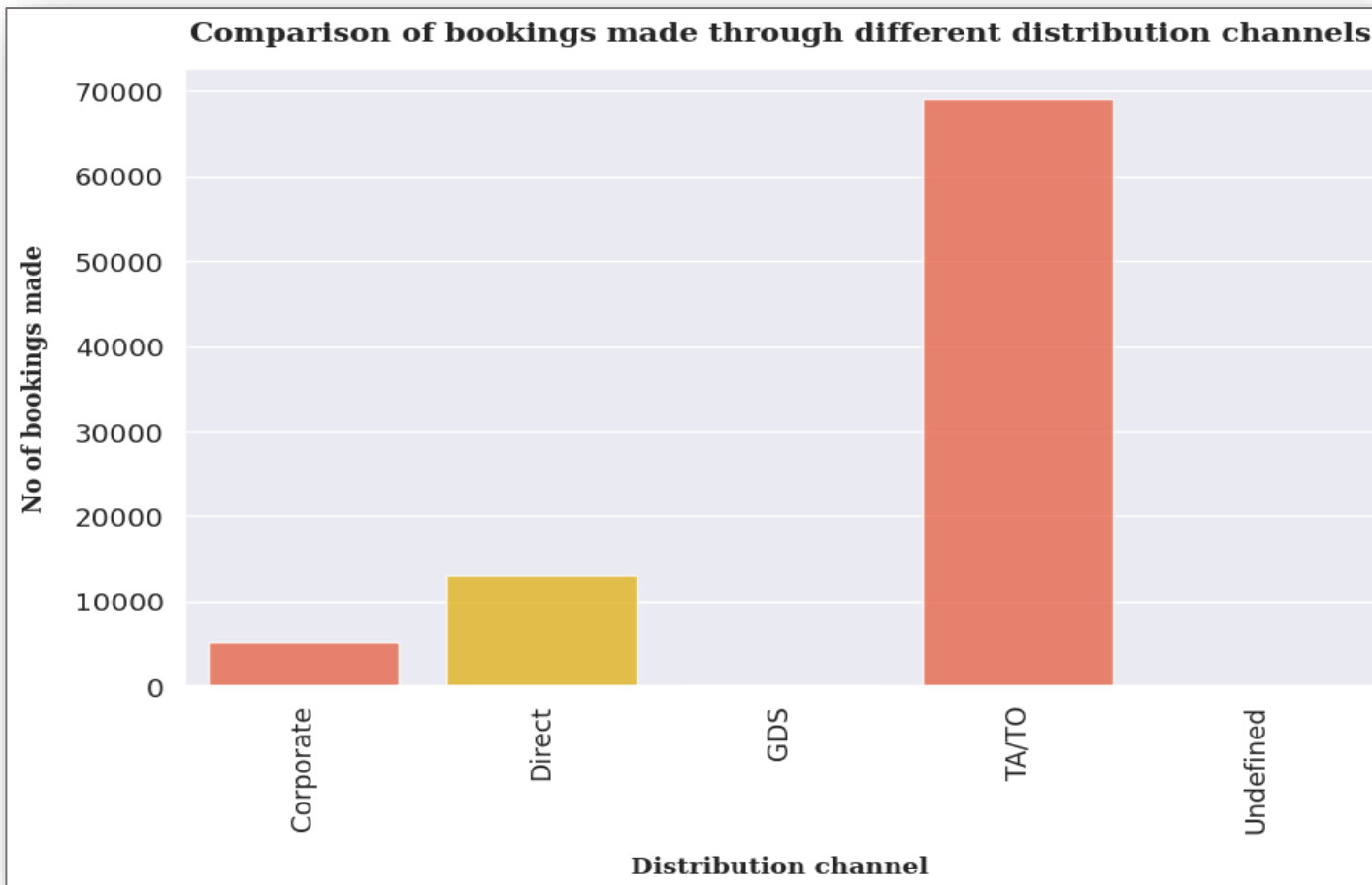For **Resort Hotel** – August was the month of maximum guest arrivals

For City Hotel – August was the month of maximum guest arrivals

- What are the bookings cancellation % of both hotels (City & Resort) ?

Bookings Cancellation Percentages for Hotels

Resort Hotel

43.9%

56.1%
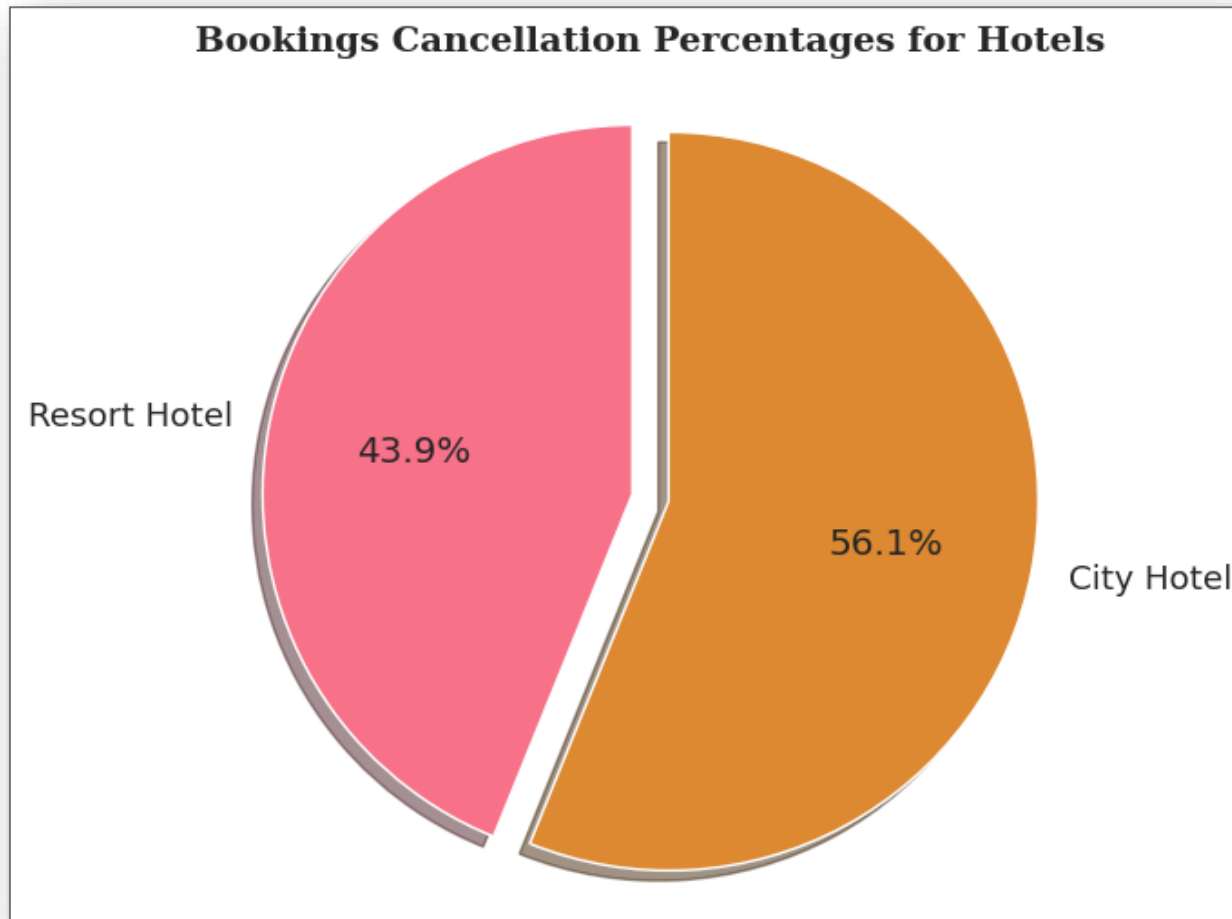
City Hotel

(**City Hotel**) has a higher cancellation rate as compared to Resort Hotel.

- Through which distribution channel was the maximum bookings made ?

Comparison of bookings made through different distribution channels

No of bookings made

70000
60000
50000
40000
30000
20000
10000
0

Corporate | Direct | GDS | TA/TO | Undefined

Distribution channel

(**Travel Agency/Travel Operator**) was the distribution channel through which the maximum bookings done followed by Direct,

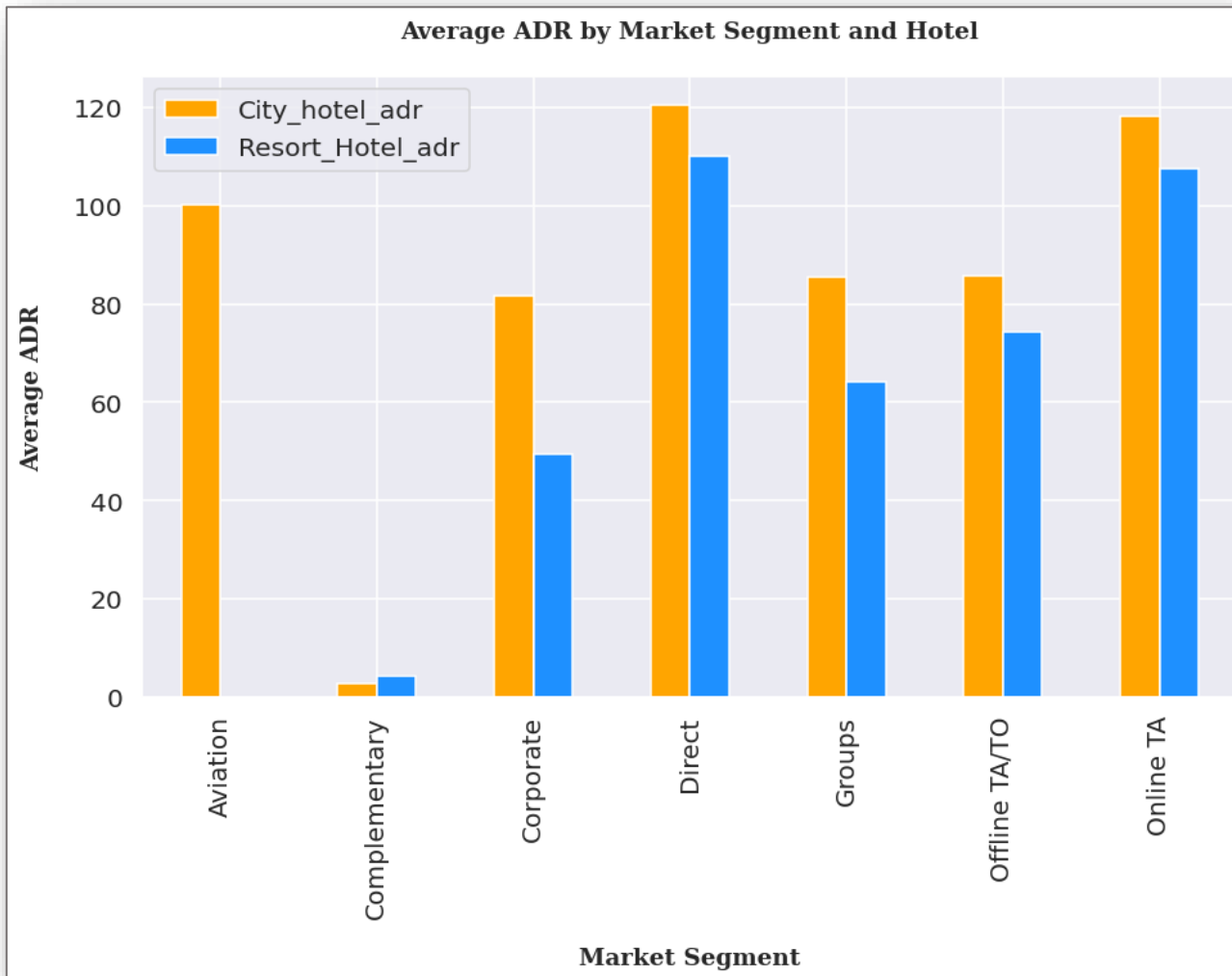- **Which category of rooms were assigned the most ?**
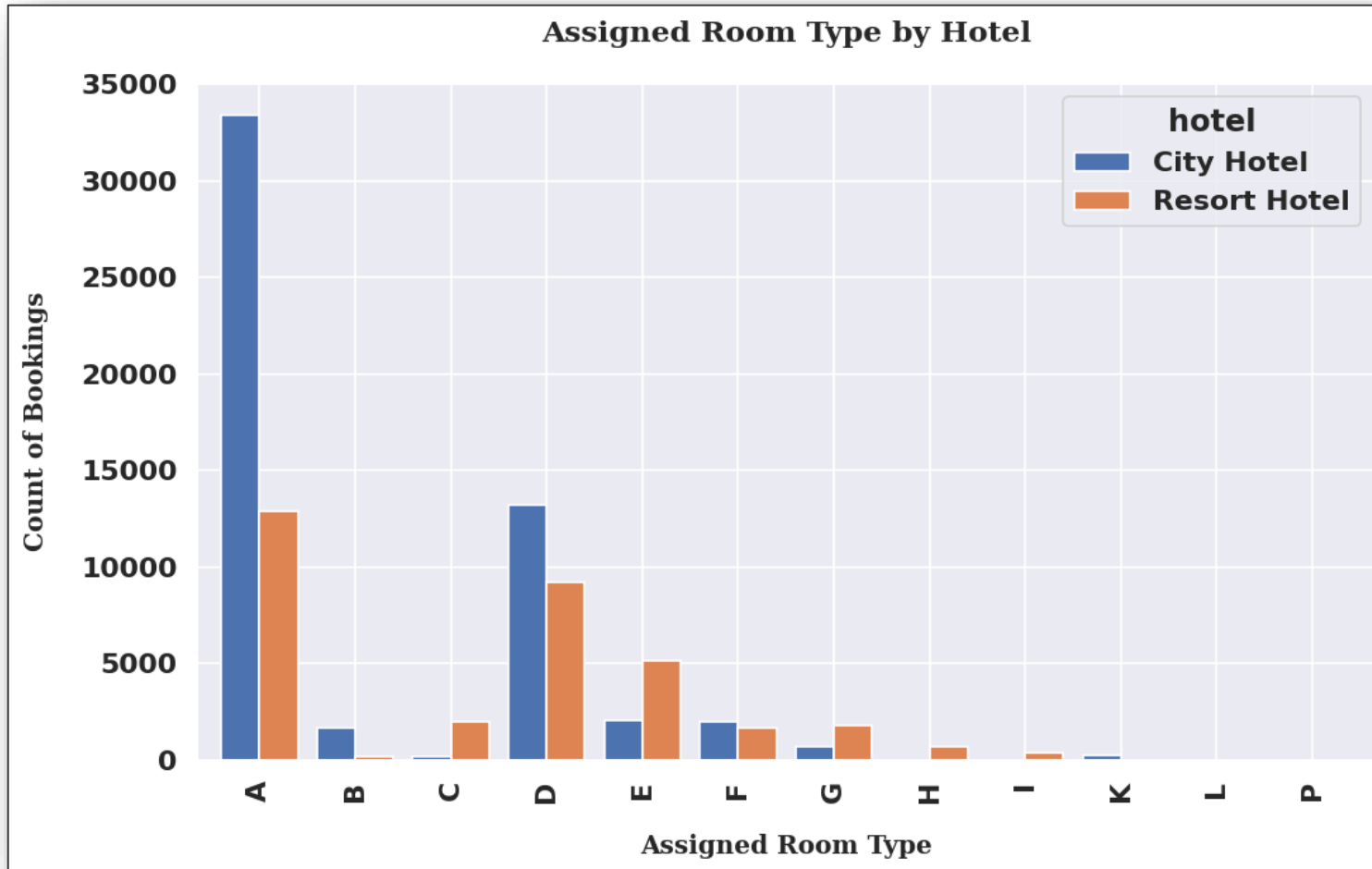
Category (A) rooms were assigned the most followed by (D) over the course of 3 years. By hotels.

- Which category rooms were reserved the most by the customers?
- Alternatively which category of rooms are least preferred by the customers ?


Reserved Room Type by Hotel
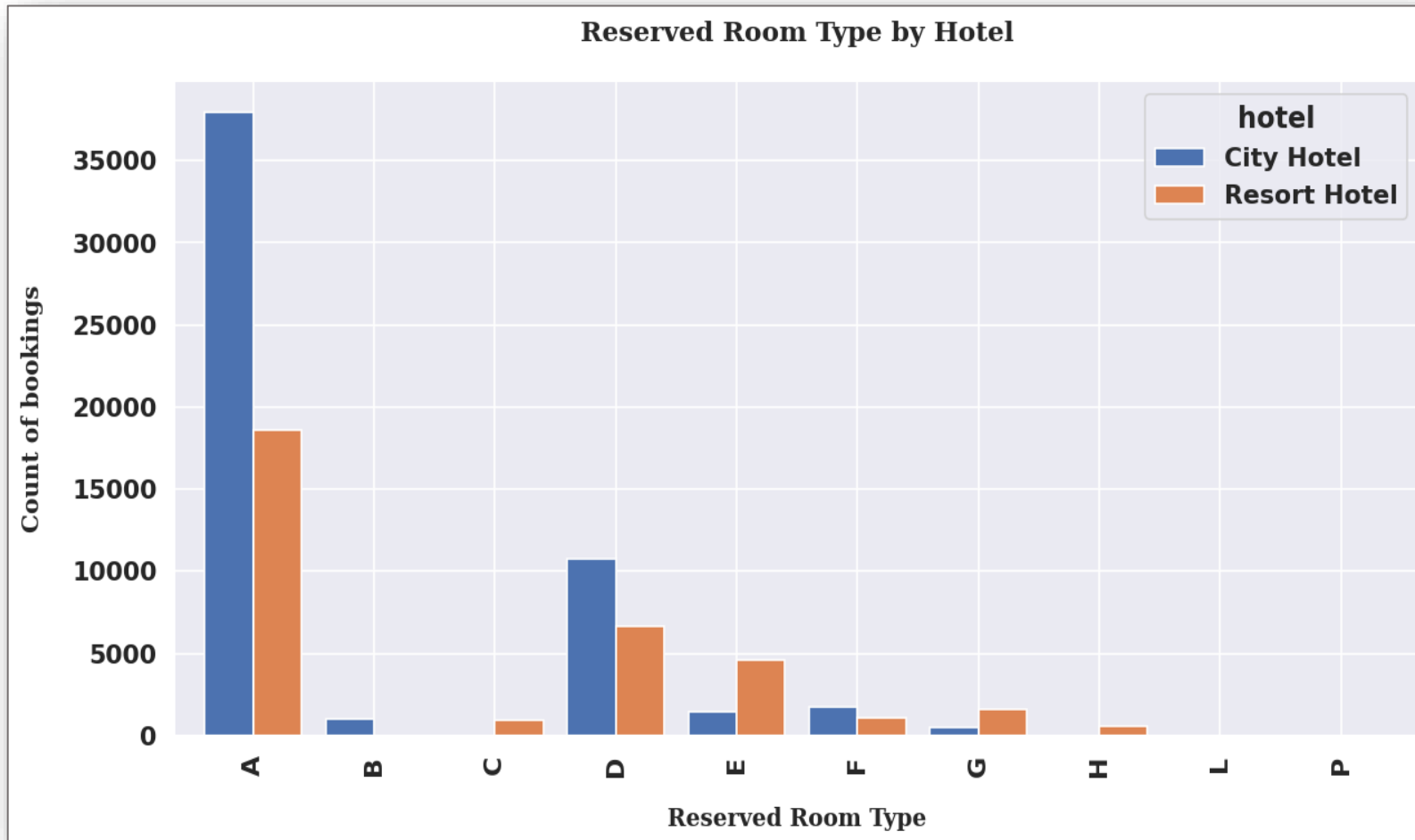
Room category (**A**) was reserved the most by both City and Resort Hotels.
Room(**P**) and (**L**) are least preferred rooms by the customers.

- **Which Market Segment had the highest ADR and for which hotel ?**

Average ADR by Market Segment and Hotel

Legend:
- City_hotel_adr
- Resort_Hotel_adr

Y-axis: Average ADR
X-axis: Market Segment (Aviation, Complementary, Corporate, Direct, Groups, Offline TA/TO, Online TA)

- **Aviation**: Refers to the market segment of customers who are airline crew or staff.
- **Complimentary**: Refers to a hotel room or service that is provided to a guest for free, often as a gesture of goodwill or to enhance their overall experience.
- **Corporate**: Refers to the market segment of customers who are traveling for business purposes.
- **Direct**: Refers to a booking made directly with the hotel, rather than through a third-party booking platform or travel agent.
- **Group**: Refers to a market segment of customers who are traveling as part of a larger group, such as a conference or tour group.
- **Offline TA/TO**: Refers to travel agencies or tour operators who book hotel accommodations for customers
- **Online TA**: Refers to online travel agencies that allow customers to book hotel accommodations through their website or mobile app. Examples include Booking.com, Expedia, and Agoda.

Which customer category had the highest ADR and for which hotel ?

Average ADR by Customer Type and Hotel
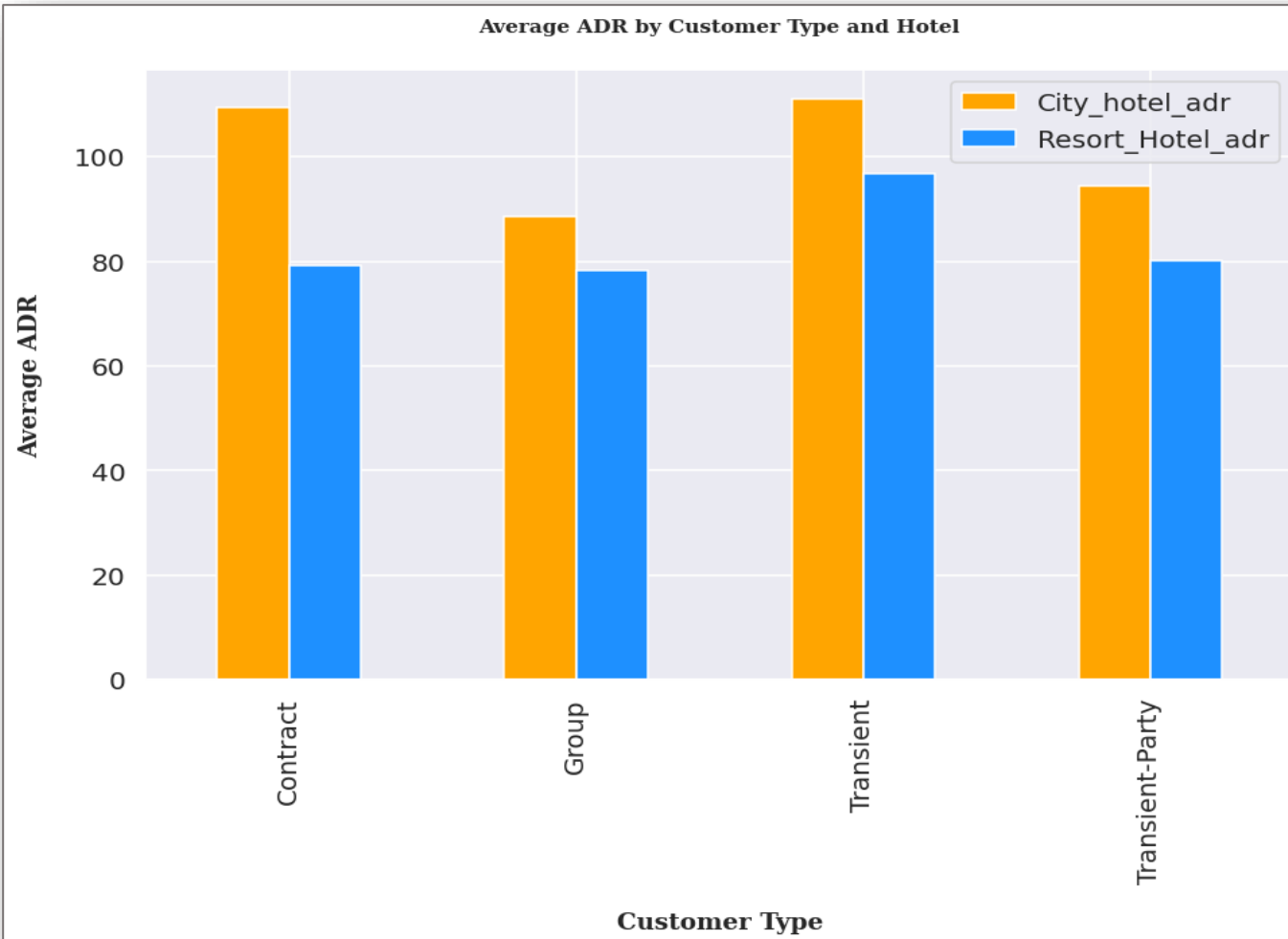
- **Transient**: An individual traveler or a small group of travelers who book a hotel room for a short stay and pay for their own accommodation.

- **Transient Party**: A group of individuals who are traveling together, but not as part of an organized group or event.

- **Contract**: An individual or group that has an agreement with the hotel for a specified period of time.

- **Group**: A large number of individuals who are traveling together as part of an organized event or tour.

(Transient) customer category had the highest average ADR for City Hotel.

- Which meal category was demanded the most according to different customer categories ?

Comparison of different meal categories for different customer categories
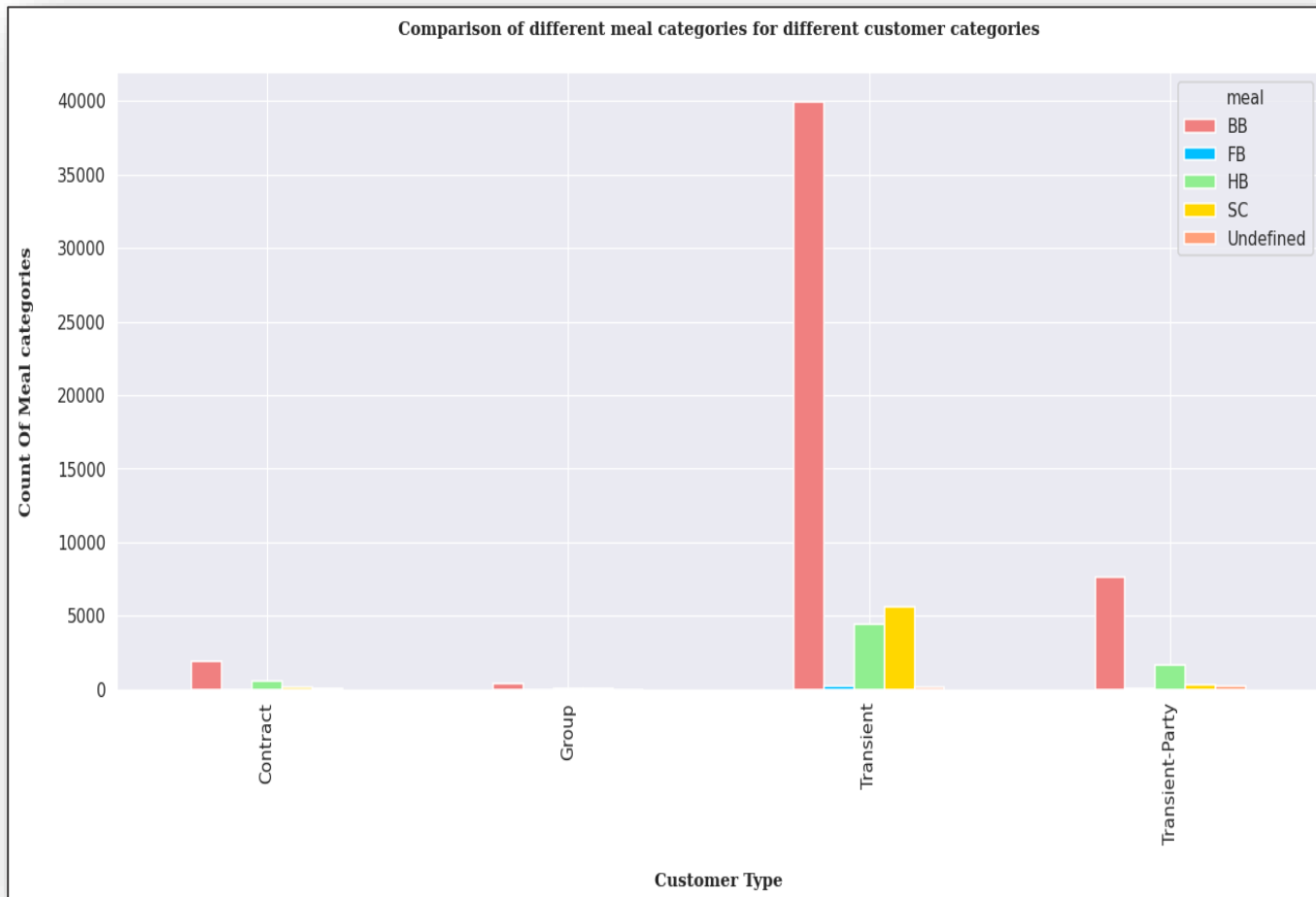
- **BB**: Bed and Breakfast

- **HB**: Half Board (Breakfast and one other meal, usually dinner)

- **FB**: Full Board (Breakfast, lunch, and dinner)

- **SC**: Self Catering (No meals included in the room rate)

- Undefined: Refers to reservations or bookings where the meal category is not specified or cannot be determined.

Consistently (**BB**) meal category was demanded the most across the customer categories.

- Which hotel out of the two earned the most revenue across the 3 year period ?

Percentage of Revenue by Hotel Type

City Hotel
Resort Hotel

City Hotel — 52.9%
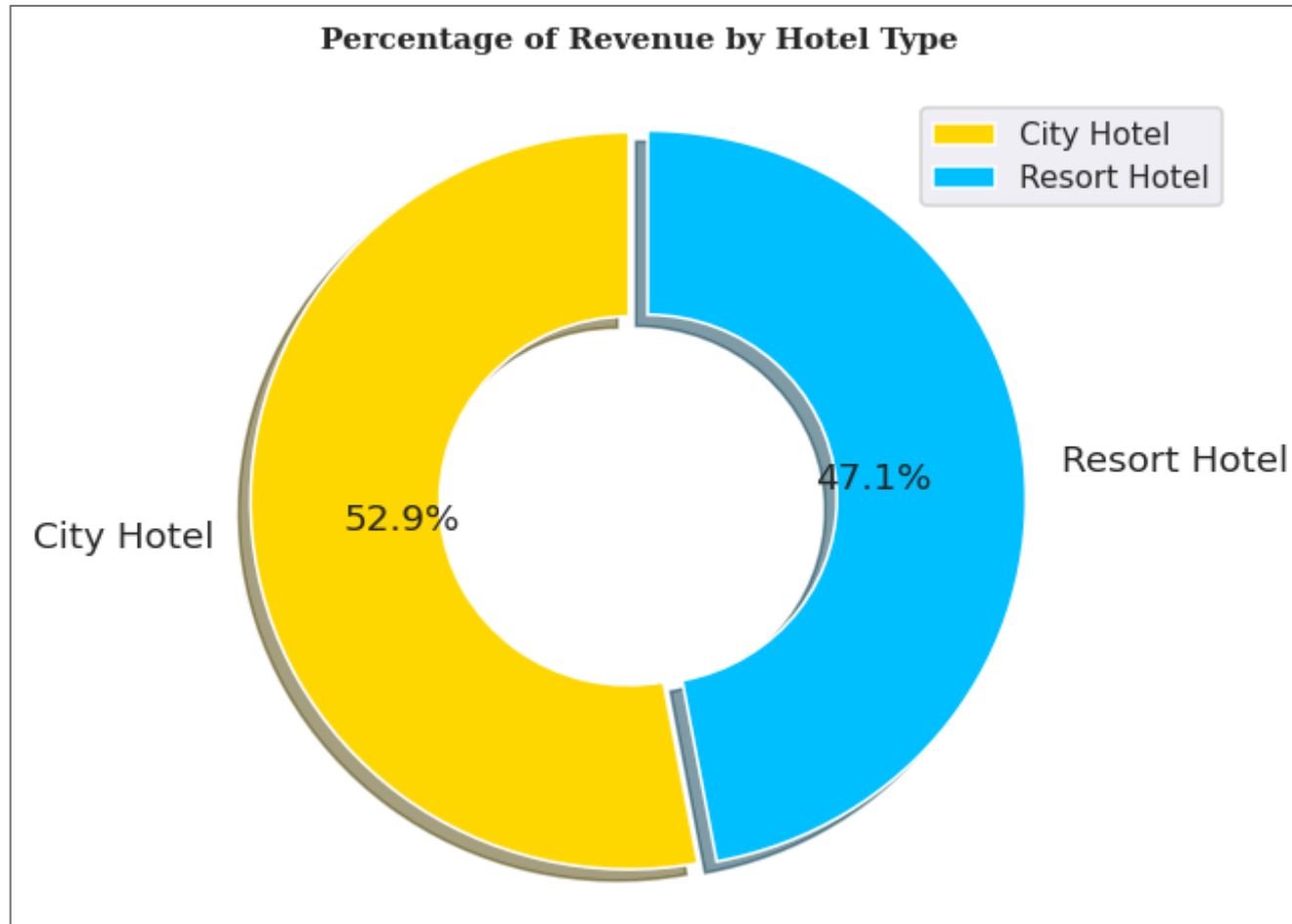
Resort Hotel — 47.1%

(City Hotel) earned more revenue than the (Resort Hotel) by almost 6 %.

- Show the relationship between Total Stay and Waiting Days.
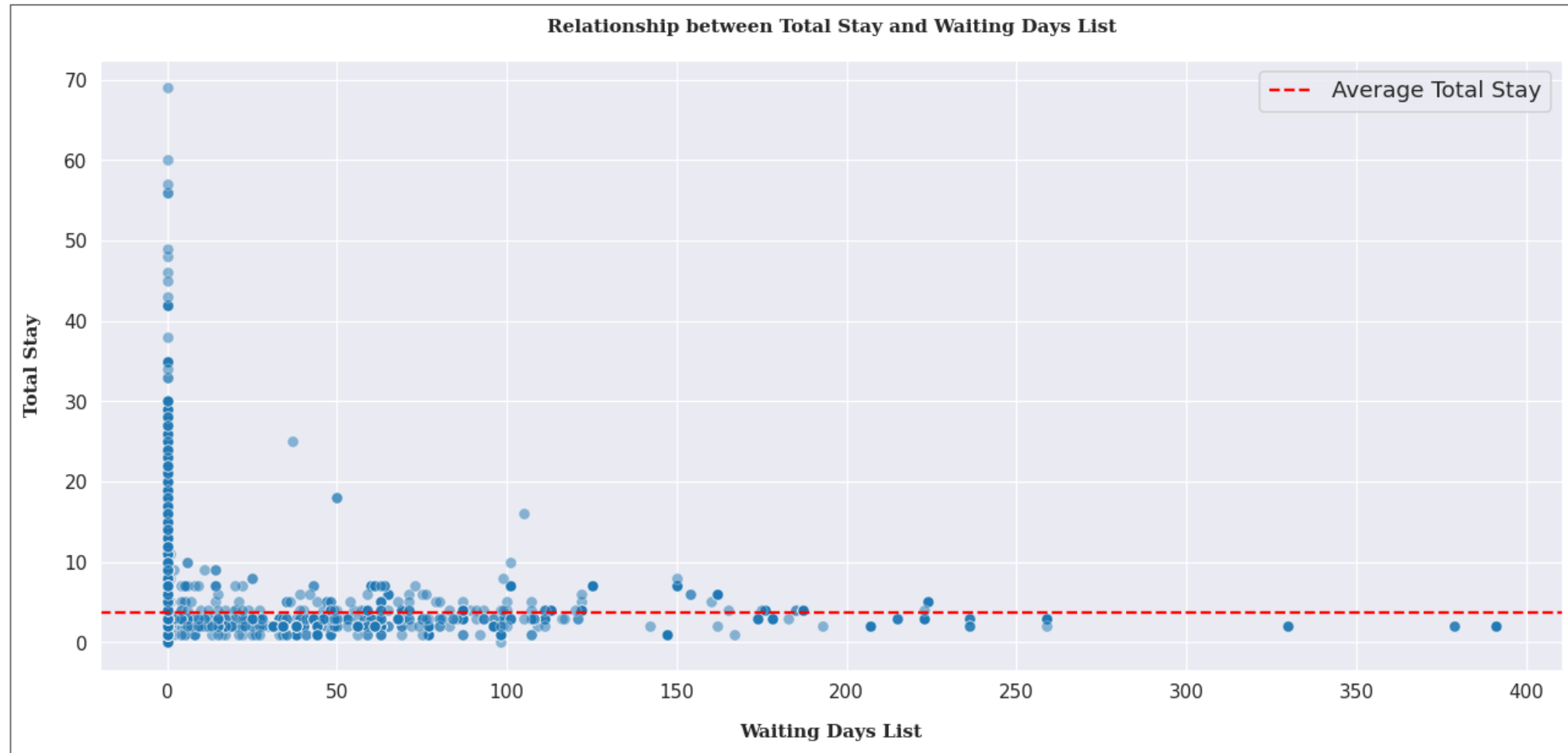
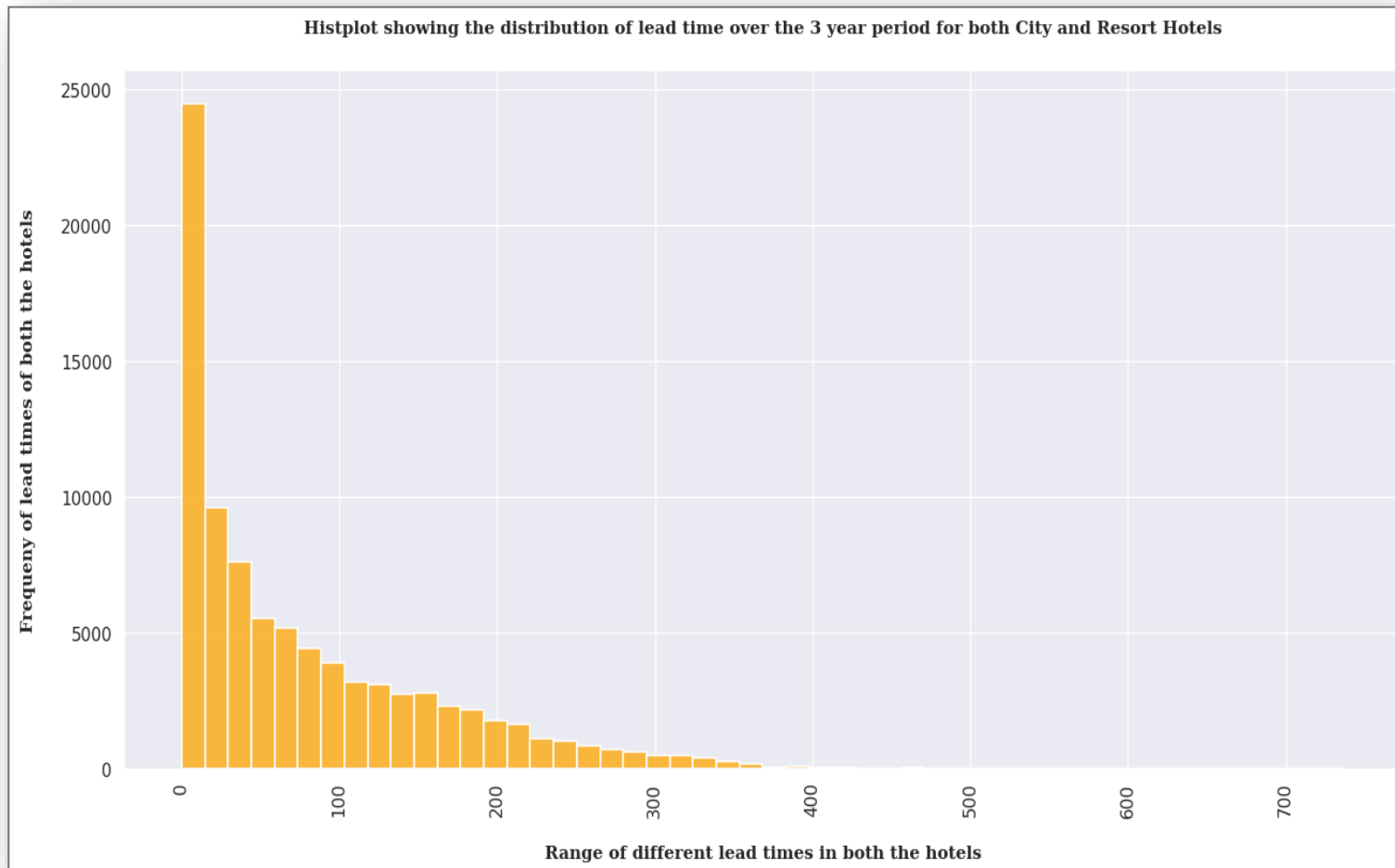**Which countries guest helped in generating the maximum revenue across the 3 year period ?**

Top 10 revenue generating countries with respect to guest arrivals
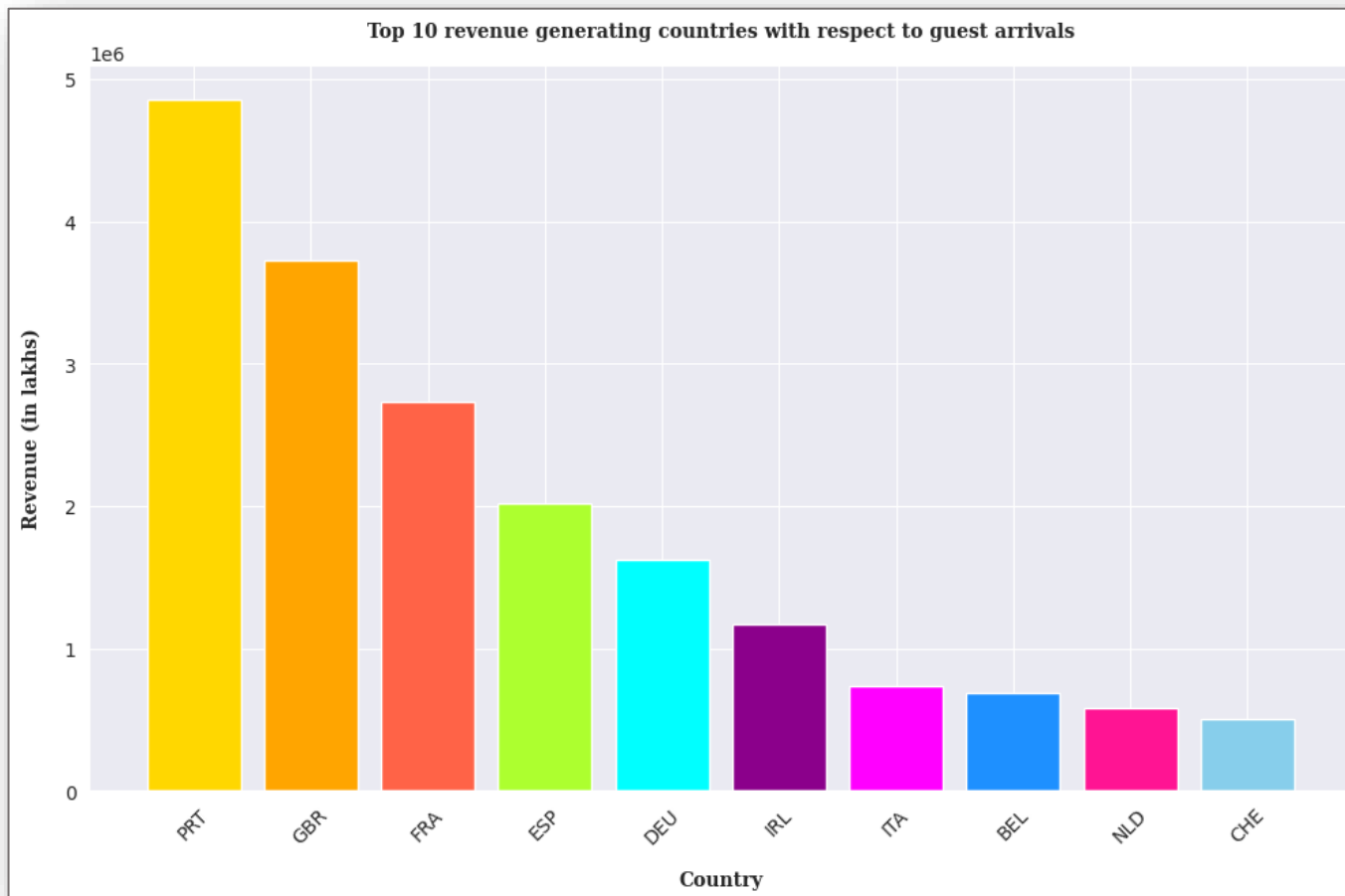
(PRT) country guest helped in generating the maximum revenue followed by (GBR) and (FRA).

- Show the frequency of Lead Time across the 3 year period for both hotels.

Majority of lead time lies in the range between 0-100 days for both the hotels.

Conclusion

- It is undeniable that city hotels attract the majority of reservations and thus, they necessitate the most precise allocation of expenses.

- Hotels need to work towards decreasing their cancellation rates, as a 15% rate is considered to be notably high. To attain this objective, hotels should concentrate on elevating their guests' experience and guaranteeing that their amenities fulfill their guests' anticipations.

- Out of the total number of guests, only 3415 are repeat customers, which is relatively low considering the three-year timeframe and the two hotels. This presents an opportunity for the hotels to focus on improving guest loyalty and increasing the number of repeat customers. It is essential to investigate the reasons why guests are not returning and to take steps to address any issues that may be contributing to this trend. By doing so, the hotels can work towards building a loyal customer base and increasing their revenue over time.

- Summer months (June to September) see the highest hotel bookings for both City Hotel and Resort Hotel likely due to vacations and warmer weather especially for beach or resort destinations indicating seasonal trends.