

# Capstone Project-1

## Hotel Booking Analysis

### Team Members

Bandna Thakur  
Shivraj Saude  
Saurabh Waghmare  
Prerit Tyagi  
Saurav Sharma

# Contents :

1. Problem statement
2. About our dataset
3. Preparing the dataset
4. Data Visualisation
5. Hypothesis
6. Conclusion



# Importance of Hotel Booking Analysis ?

1. Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more.
2. Hospitality industry is big contributor to economic growth of any country.
3. This makes analyzing the patterns available in the past data more important to help the hotels plan better.
4. Using the historical data, hotels can perform various campaigns to boost the business.
5. We can use the patterns to predict the future bookings using time series or decision trees.
6. The growth in the hospitality sector and its contributions to the GDP will continue to be substantially increase.



# Dataset

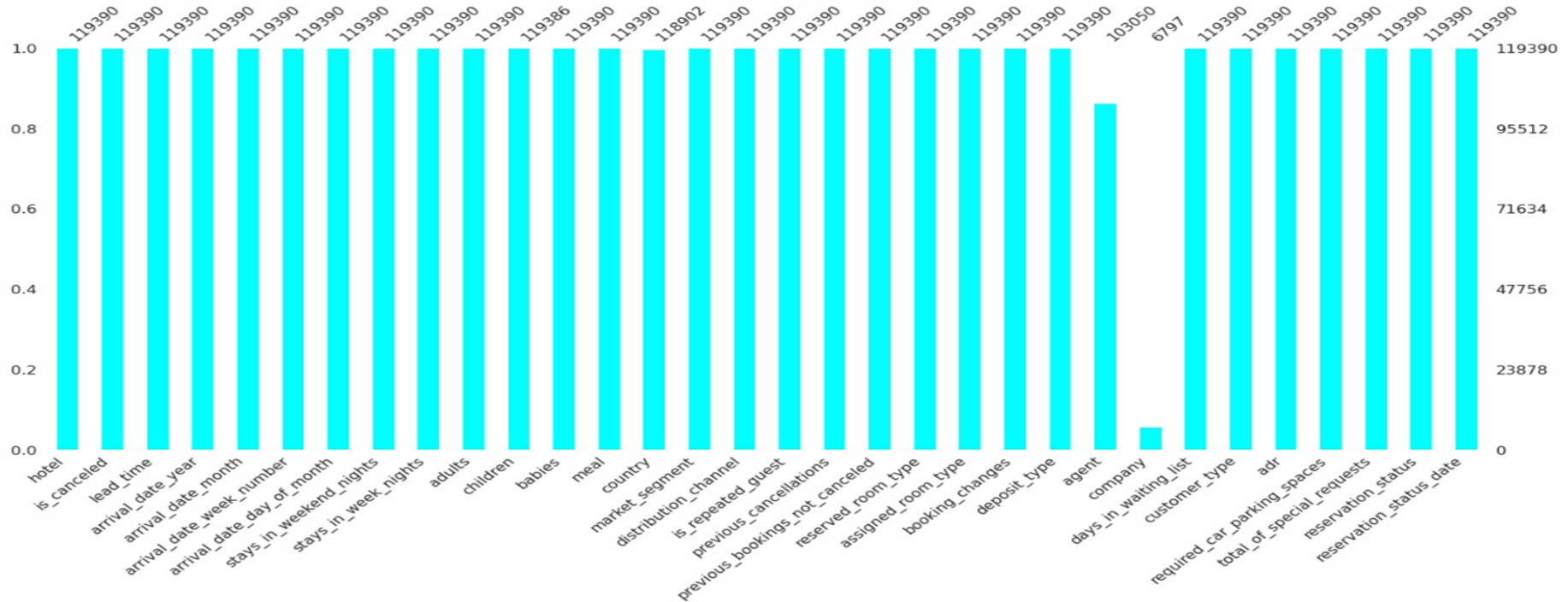
- **hotel:** Types of Hotels: Resort Hotel, City Hotel
- **is\_canceled:** Value indicates booking was canceled (1) or not (0)
- **lead\_time:** Number of days between the booking date and the arrival date
- **arrival\_date\_year:** Year of arrival date
- **arrival\_date\_month:** Month of arrival date
- **arrival\_date\_week\_number:** Week number of year for arrival date
- **stays\_in\_weekend\_nights:** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- **arrival\_date\_day\_of\_month :** Day of arrival date
- **stays\_in\_week\_nights:** Number of weeknights (Monday to Friday) 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. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch, and dinner)
- **country:** Country of origin.

- **market\_segment:** Market segment categories, “TA” : “Travel Agents” , “TO” : “Tour Operators”, Aviation, Complementary, Corporate, Direct, Groups.
- **distribution\_channel:** Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”, Corporate, Direct.
- **direct\_is\_repeated\_guest:** Value indicates if the booking name was from a repeated guest (1) or not (0).
- **previous\_cancellations:** Number of previous bookings that were canceled by the customer prior to the current booking.
- **previous\_bookings\_not\_canceled:** Number of previous bookings not canceled by the customer prior to the current booking.
- **reserved\_room\_type :** Code of room type reserved (A,B,C,D,E,F,G,H).
- **assigned\_room\_type :** Code for the type of room assigned to the booking (A,B,C,D,E,F,G,H,I,J,K).
- **booking\_changes:** Number of changes made to the booking from the moment the booking was entered until check-in or cancellation.
- **deposit\_type:** This variable has categories: No Deposit – no deposit was made; Non-Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of the stay.
- **agent:** ID of the travel agency that made the booking.

- **company:** ID of the company that made the booking or is responsible for paying the booking.
- **days\_in\_waiting\_list:** Number of days the booking was in the waiting list before it was confirmed.
- **customer\_type:** Types of customer four categories: Contract - when the booking has an allotment or other type of contract associated with it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract and is not associated to other transient bookings; Transient-party – when the booking is transient but is associated to at least other transient bookings
- **adr:** Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.
- **required\_car\_parking\_spaces:** Number of car parking spaces required by the customer.
- **total\_of\_special\_requests:** Number of special requests made by the customer.
- **reservation\_status:** Reservation status contains three categories: Canceled – booking was canceled by the customer; Check-Out – customer has checked in but already departed; No-Show – the customer did not check in and didn't inform the hotel its the reason.
- **reservation\_status\_date:** Date at which the last status was set.

# Preparing our dataset :

- Libraries used : NumPy, Pandas, Seaborn, Matplotlib.
- Shape : 119390 (rows), 32 (columns).
- Most of the null values were present in columns : company and agent.



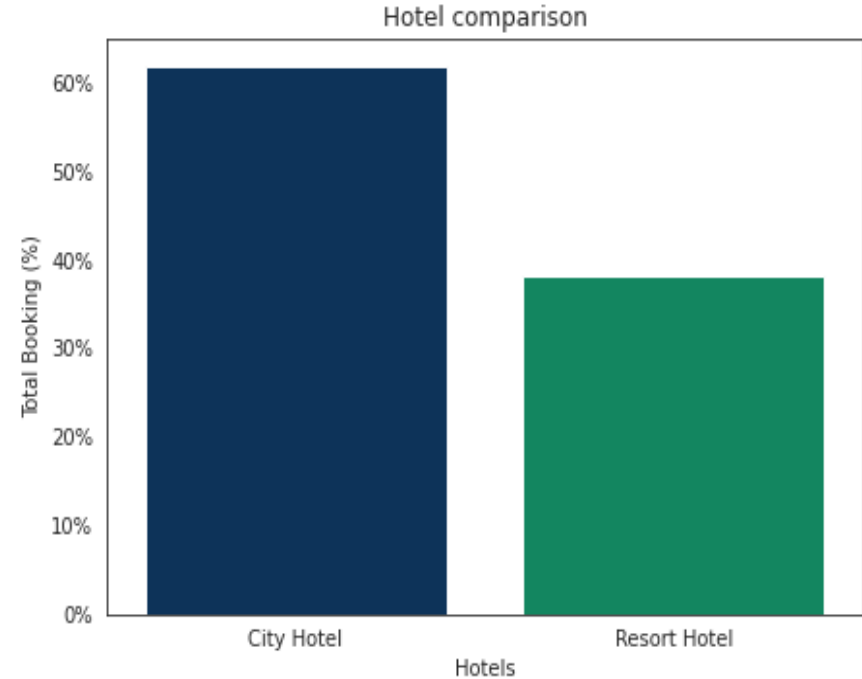
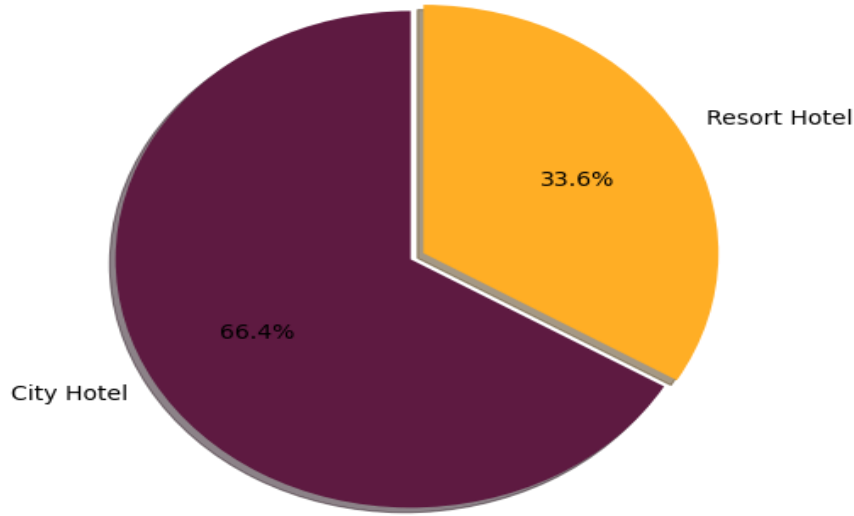
# **Exploratory Data Analysis**



# **VISUALIZATION**

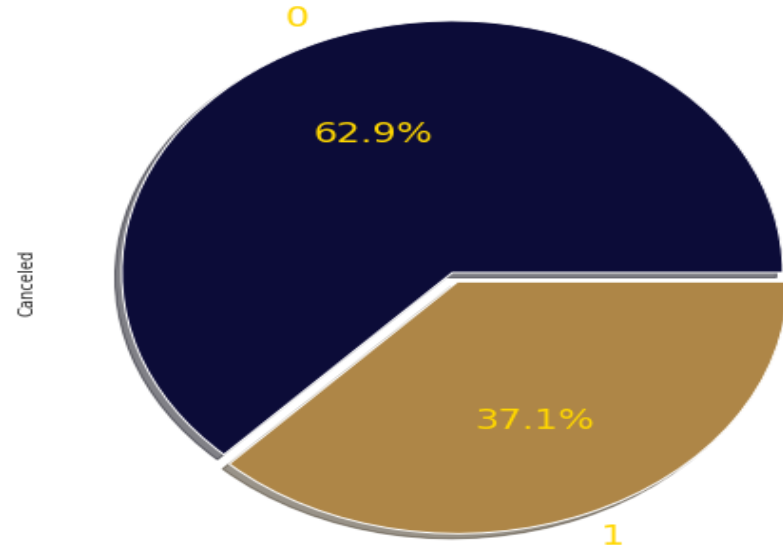
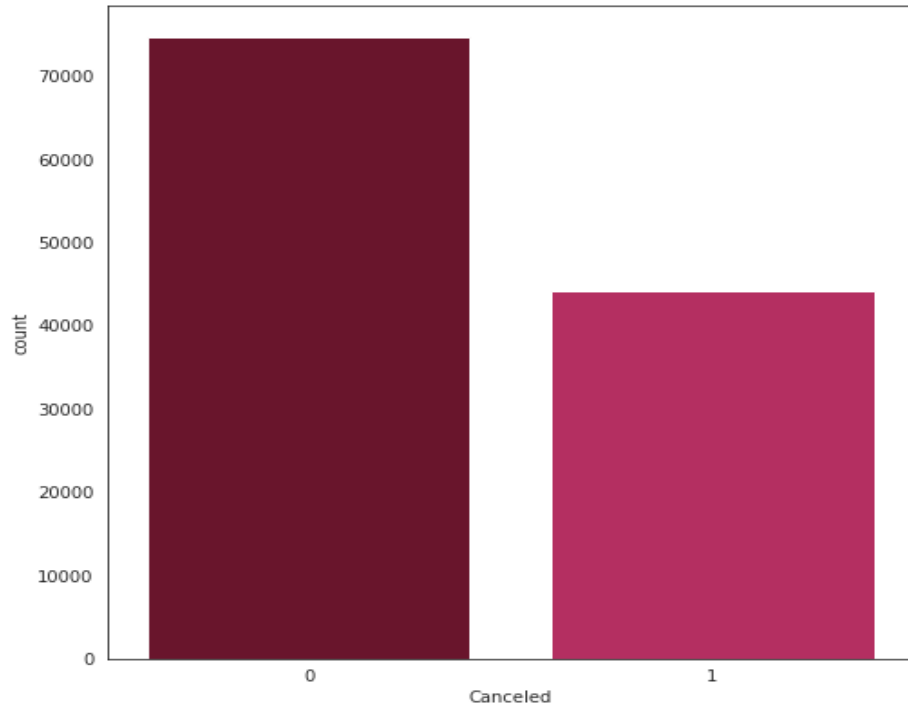


# What is the Total Number of Bookings of both hotel types?



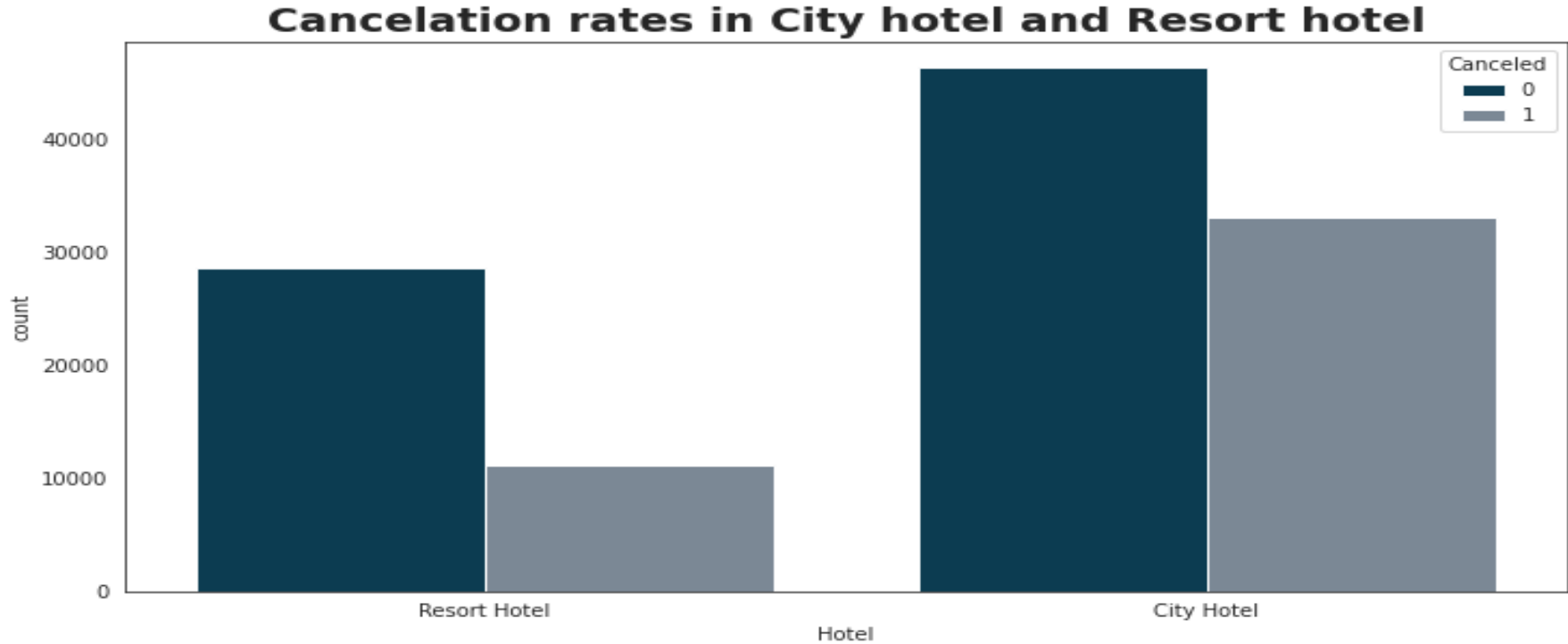
- We can see that the percentage of **City hotels** is **more** compared to **Resort hotels** in terms of bookings.

## Overall Cancellation Ratio:-



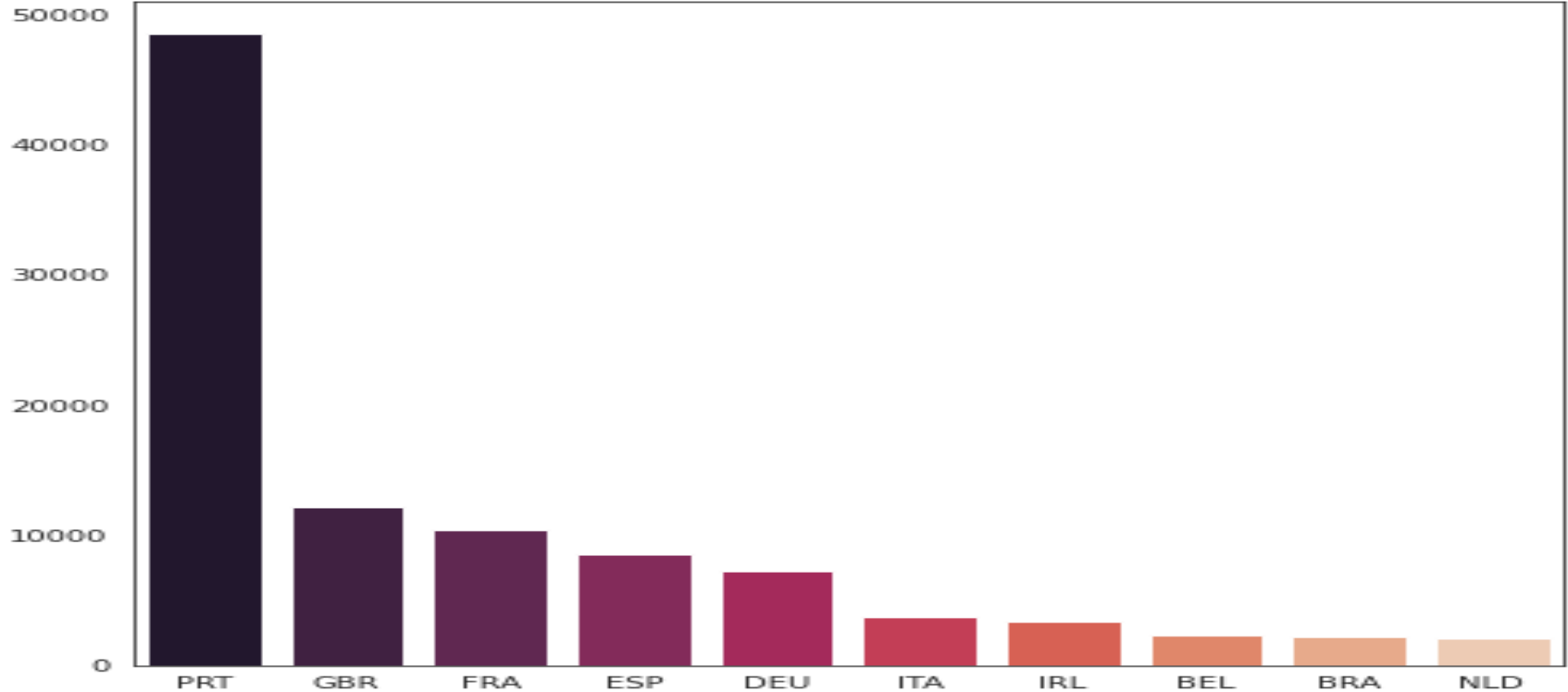
- *Majority of bookings were cancelled on the spot or at the hotels.*
- According to the pie chart, **63% (74745)** of bookings were not cancelled and **37% (44153)** of the bookings were cancelled at the Hotel.

# In which Hotels cancellation rate is high?



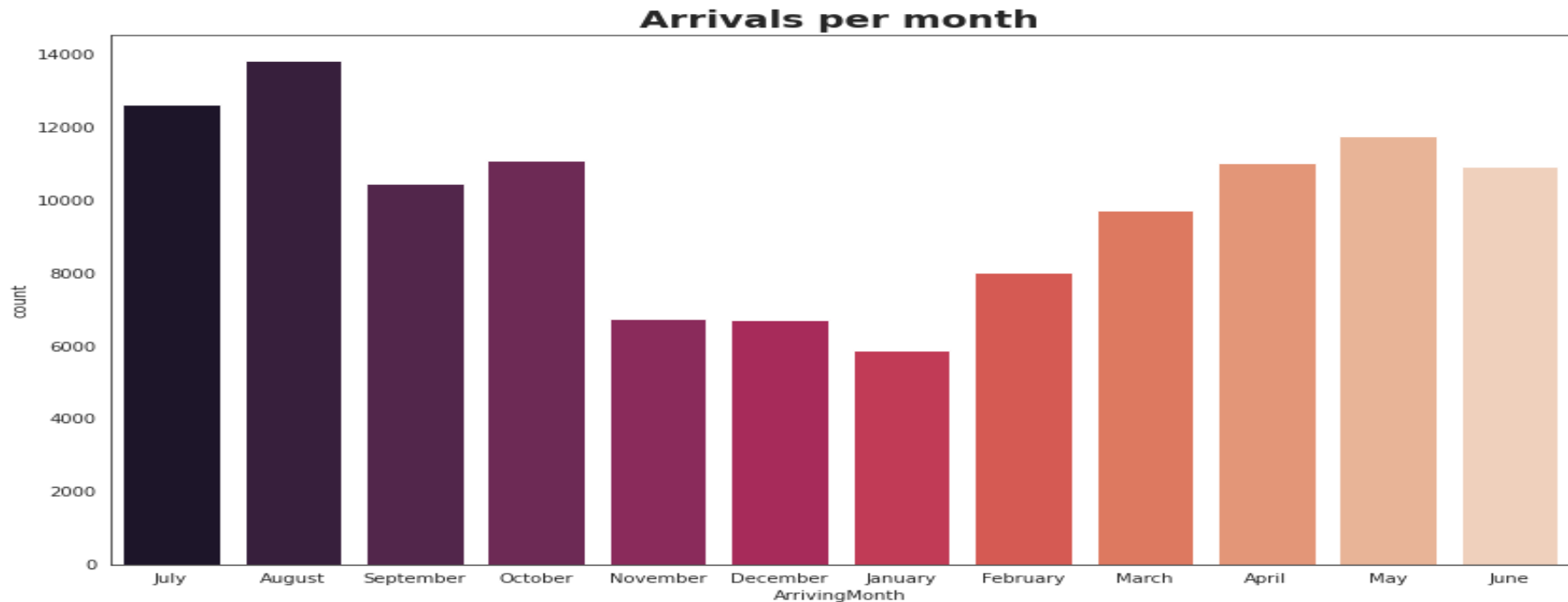
- Mostly the bookings were done from City hotels than Resort hotels.
- But the Cancellation rate is also **HIGH** in City hotels compared to Resort hotels.

## From which countries most of the guests came ?



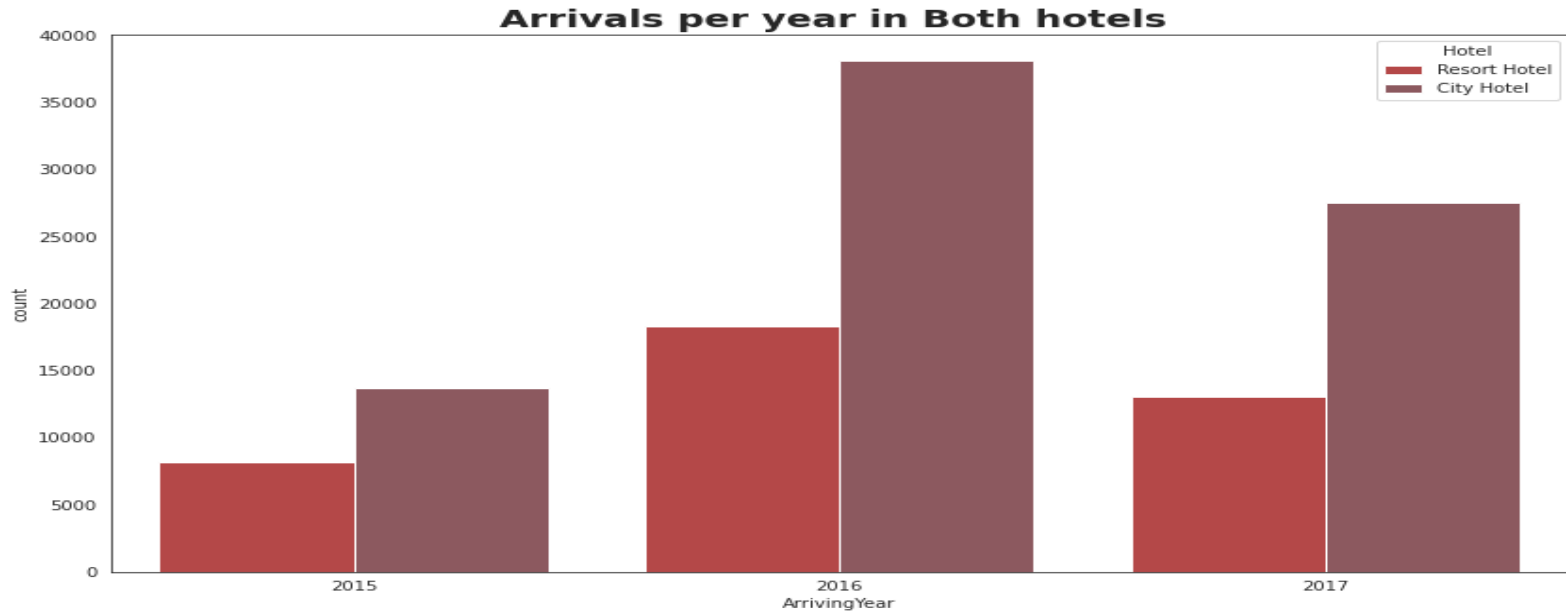
- More visitors are from western Europe, namely **Portugal** being the Highest, **Great Britain**(UK), **France** and so on...

# Which are the busiest months for hotel ?



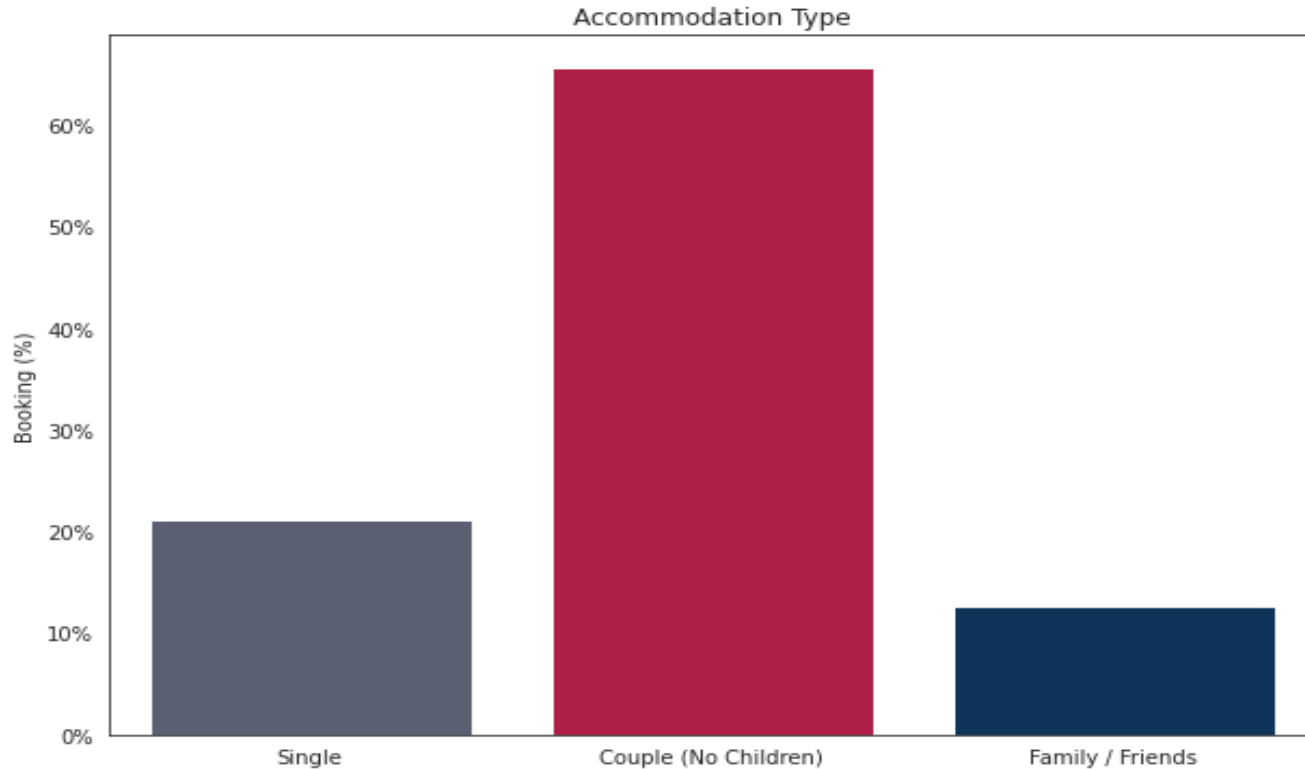
- Most bookings were done in the month of **August, July, May, June**, and so on...
- **August** is the most **occupied (busiest)** month with **11.62%** bookings and **January** is the most **unoccupied** month with **4.96%** bookings.

# In which years maximum bookings are done for both hotels?



- **47%** of bookings were done in the **year 2016**, then **34%** in **2017** and **18%** in **2015**.
- We can see the increasing tendency in bookings year wisely.
- We can see most of the bookings were in the year 2016 and **highest bookings** were from **City hotels**.

# Which was the most booked accommodation (Single,Couple,Family)?

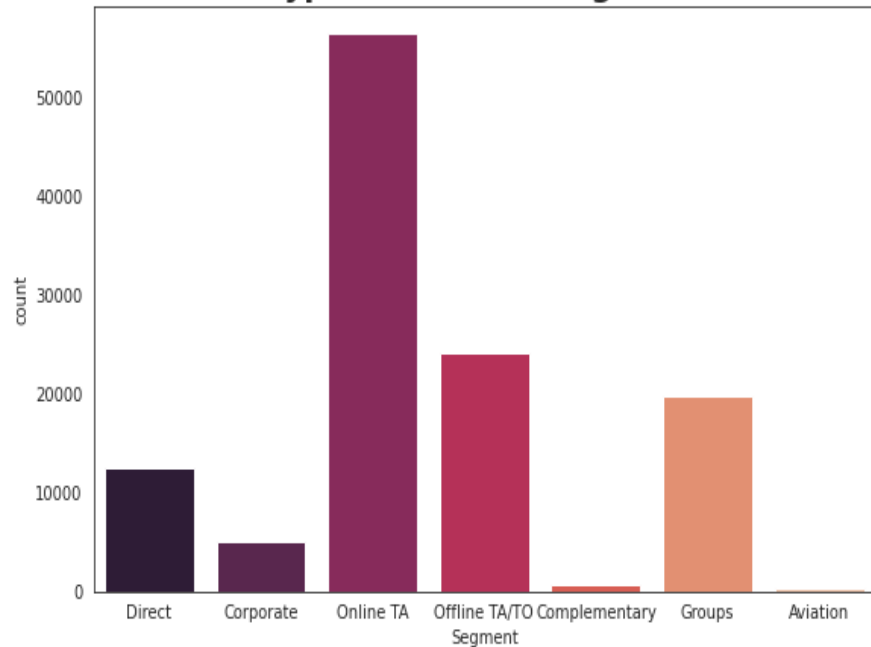


- The **highly** booked Accommodation was for **Couples**, then Single and Family/Friends.

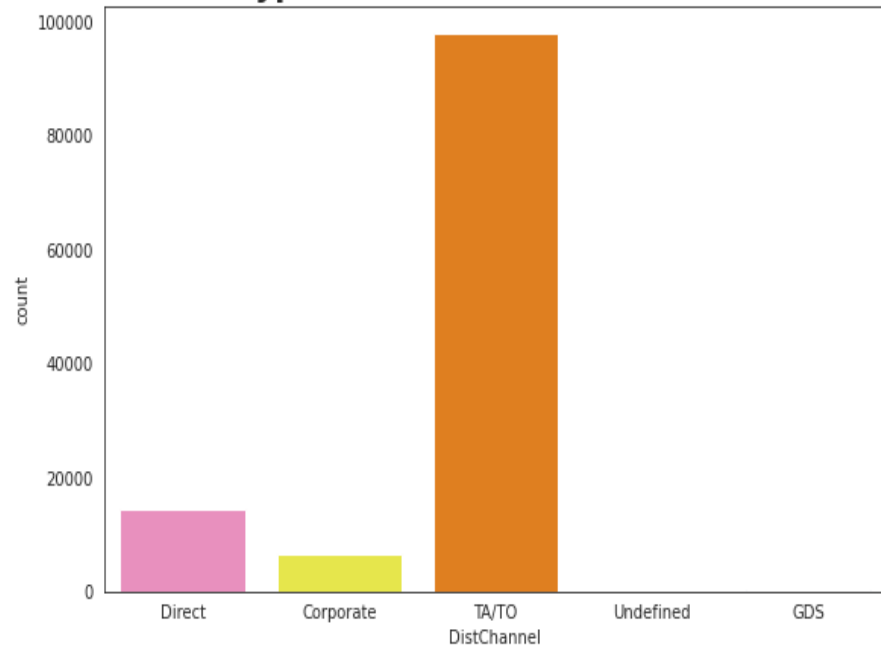


# What are the Market Segment of both hotels?

Types of market segment



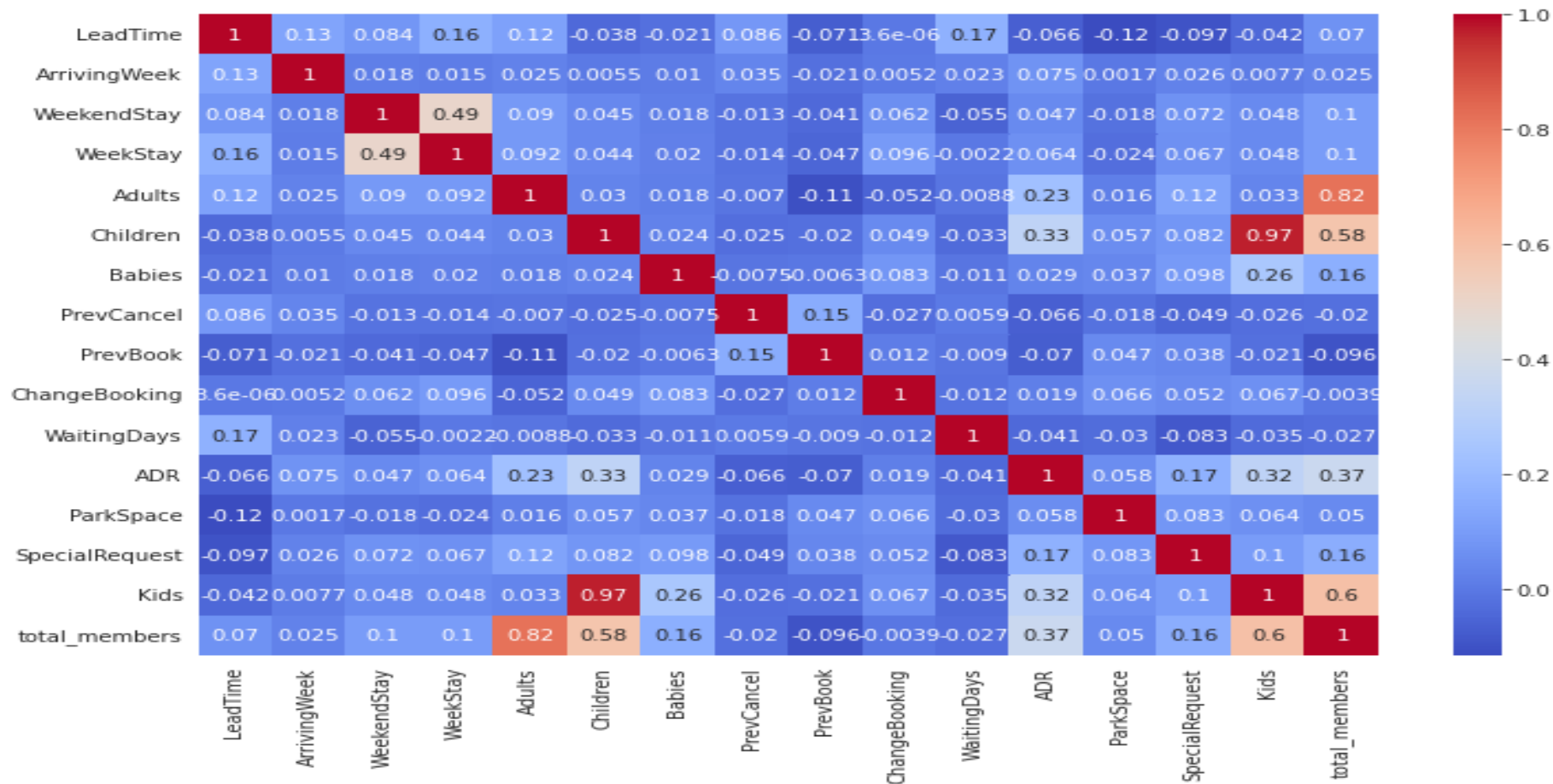
Types of distribution channels



➤ Majority of Market segments and Distribution channels is **Travel agencies(offline/online)**.

# Extracting the correlation using Heatmap.

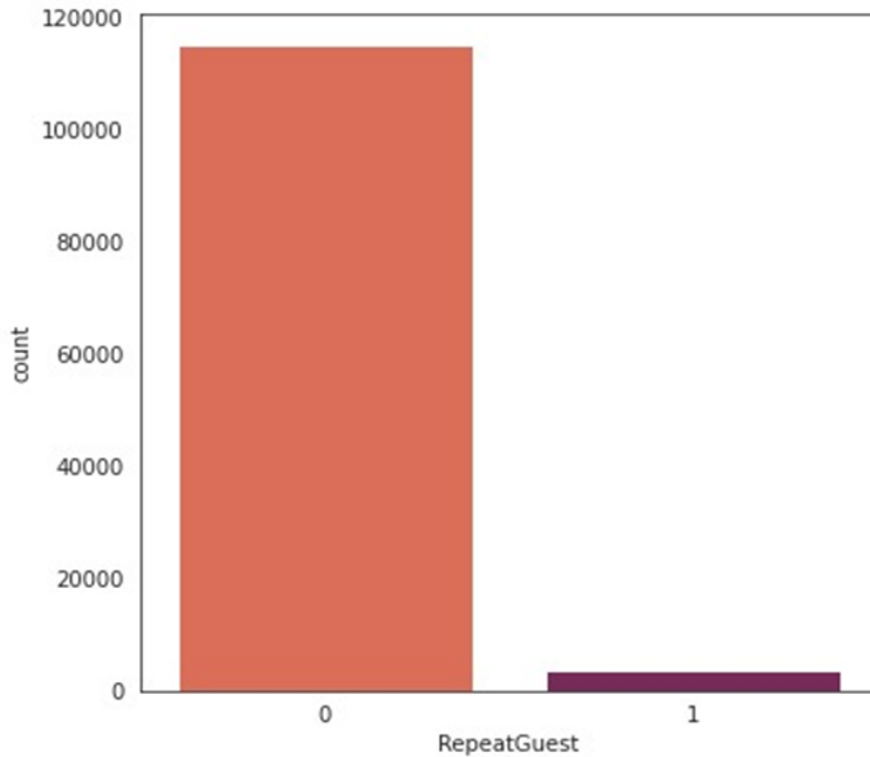
AI



# **HYPOTHESIS**

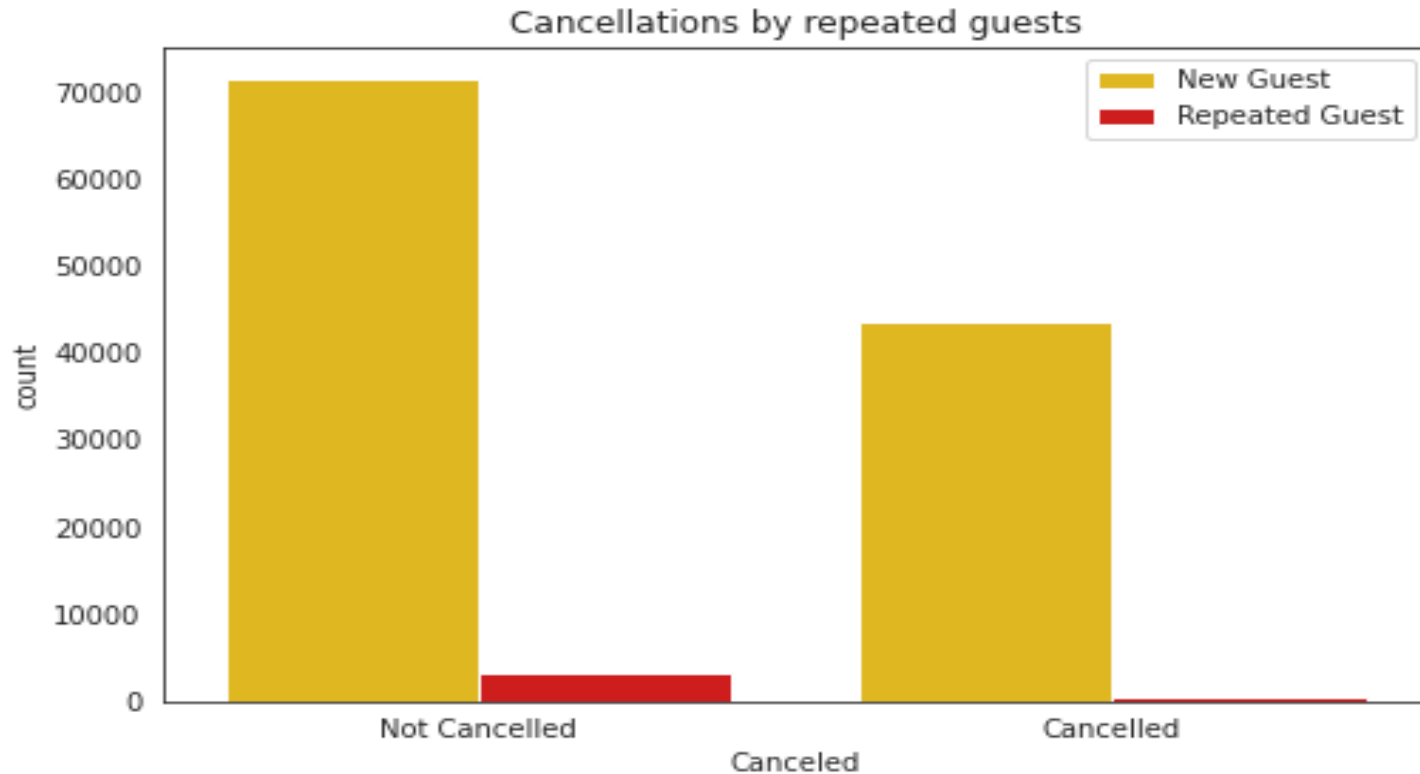


# New Guest & Repeated Guests Count:



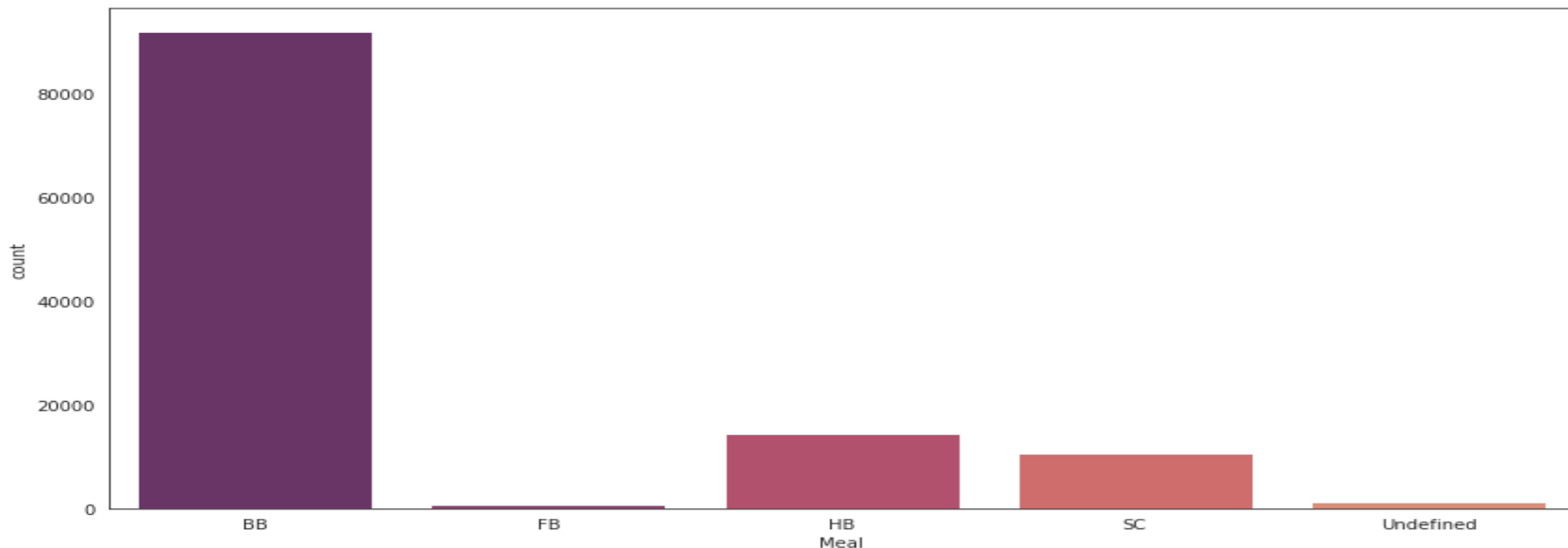
- Repeated guests(1) are **very less** comparative to New guests(0)
- The **ratio** of New guests is **high** then Repeated guests.

## Cancellations by type of Guests (Repeated/NonRepeated) :-



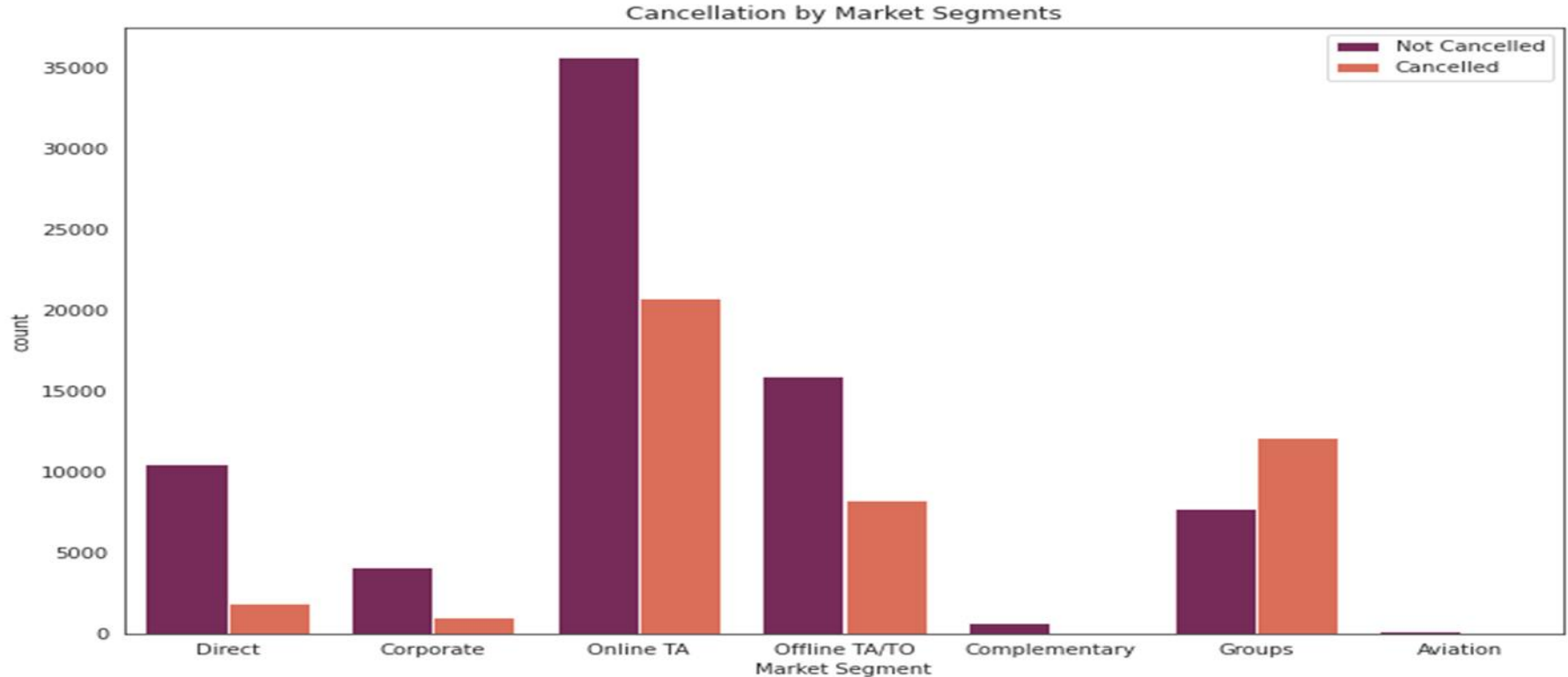
- We observed that **New Guests CANCELS** their bookings **more** than the Repeated Guests.

# Most popular meal type?



- Type of meal booked. Categories are presented in standard meal packages:-
- Undefined/SC — no meal package.
- BB — Bed & Breakfast.
- HB — Half board (breakfast and one other meal — usually dinner).
- FB — Full board (breakfast, lunch, and dinner).
- **BB Meal plan is most popular meal plan followed by HB.**

## Cancellation by Market Segments:-



- Almost 95%, cancellations are from online TA , offline TA and group segments.
- Online TA has the **highest** cancellations of around **50%**, followed by groups and offline TA.

# Conclusions





1. City hotel was booked in the majority. So it's good to spend the most targeting fund on these hotels.
2. Cancellation ratio of a city hotel is high than resort hotel though city hotels bookings are also high compared to resort hotels.
3. Many guests come from Western European countries. So hotels should spend a significant amount of budget in these areas.
4. May to August is the busiest month so the hotels should target more customers in this period and try to do more business for profit.
5. Most customers are couples and bringing kids along with them is rare so the hotels should advertise in such a way that it attracts the couples more.
6. In the Market segment many of the hotel bookings are done from Travel agencies (online and offline).
7. Encourage Direct bookings by offering special discounts.
8. Hotel has low repeated guests, Hotel should depend on online marketing companies for advertising to increase the repeated guests.
9. New Guests cancel the most than the repeated guests.
10. In food most popular meal plan was BB(Bed & Breakfast) preferred by guests.

**Thank you**