Capstone Project

Hotel Booking Analysis

Alma Better -Cohort Seattle

Presented By:-
Vaitul Sidhdhapara
Drashti Shah

Introduction

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Using the data available to analyze the factor affecting the hotel bookings. These factors can be used for reporting the trends and predict the future bookings.

01 Problem Statement

02 Data Summary

03 work Flow

04 Data Pre Processing

05 EDA

06 Conclusion

Problem Statements

Problem Statements in 3 stages.

1. Key metrics
   - Cancellation bookings
   - New Customer acquired
   - Type of rooms and hotel available for booking
   - Revenue of the hotel
2. Data to analyze patterns associated with each segment.
   - Type of hotel, rooms, customer
   - Day of week
3. Predict the future bookings.

Data Description

- **hotel :** Name of hotel (City or Resort).
- **is_canceled :** Whether the booking is cancelled or not (0 for no cancelled 1 for cancelled).
- **lead_time :** Time (in days) between booking transaction and actual arrival.
- **arrival_date_year :** Year of arrival.
- **arrival_date_month** : Month of arrival.
- **arrival_date_week_number :** Week number of arrival date.
- **arrival_date_day_of_month :** Day of arrival date.
- **stays_in_weekend_nights :** Number of weekend nights (Saturday or 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.
- **children :** Number of children.
- **babies :** Number of babies.
- **meal :** Type of meal. (**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.

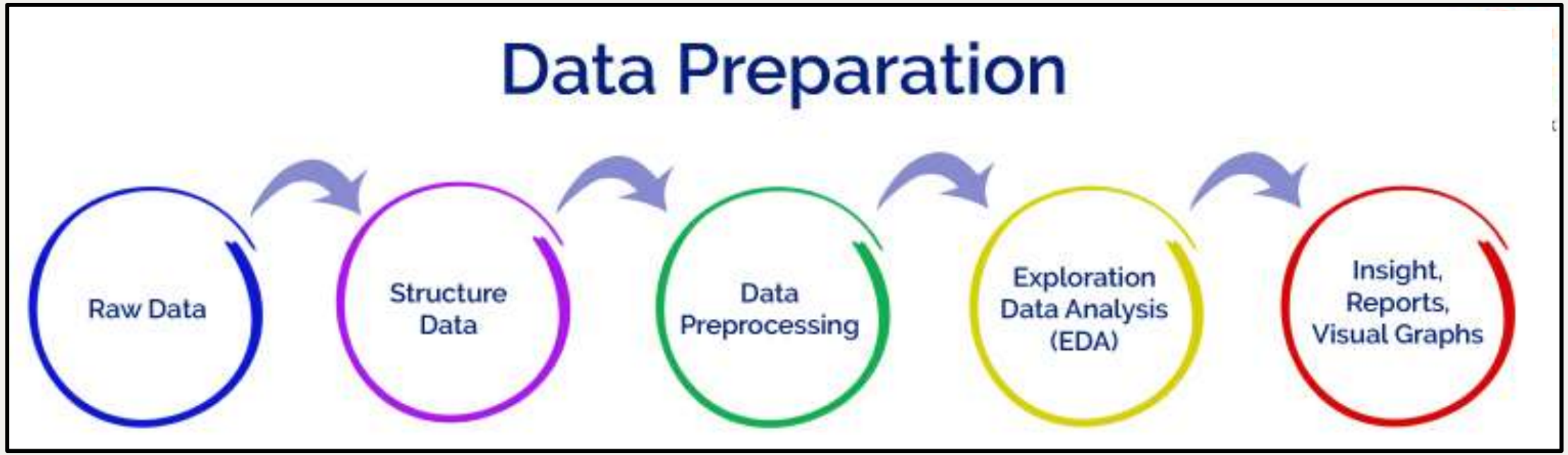
Data Description

- **market_segment :** Market segment designation. (**TA** =Travel Agents ,**TO**=Tour Operators).
- **distribution_channel :** Booking distribution channel.(**TA**=Travel Agents, **TO**=Tour Operators).
- **is_repeated_guest :** Value indicating if the booking name was from a repeated guest (1) or not (0).
- **previous_cancellations :** Number of previous bookings that were cancelled by the customer prior to the current booking.
- **previous_bookings_not_canceled :** Number of previous bookings not cancelled by the customer prior to the current booking.
- **reserved_room_type :** Code of reserved room. Code is presented instead of designation for anonymity reasons.
- **assigned_room_type :** Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request.
- **booking_changes :** Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation.
- **agent :** ID of the travel agency that made the booking.

Data Description

- **deposit_type :** Type of deposit made for booking(**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 stay).
- **customer_type :** Type of booking (**Contract** - when the booking has an allotment or other type of contract associated to it, **Groups** – 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 booking, **Transient-party** – when the booking is transient, but is associated to at least other transient booking).
- **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 (e.g. twin bed or high floor)
- **reservation_status :** Reservation last status (**Canceled** – booking was canceled by the customer, **Check-Out** – customer has checked in but already departed, **No-Show** – customer did not check-in and did inform the hotel of the reason why)
- **reservation_status_date :** Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when was the booking canceled or when did the customer checked-out of the hotel

Work Flow


Data Preparation
Raw Data
Structure Data
Data Preprocessing
Exploration Data Analysis (EDA)
Insight, Reports, Visual Graphs

01 Data Pre Processing

1. Cleaning the data

2. Manage the NaN(Missing) values

3. Changing datatype of necessary data

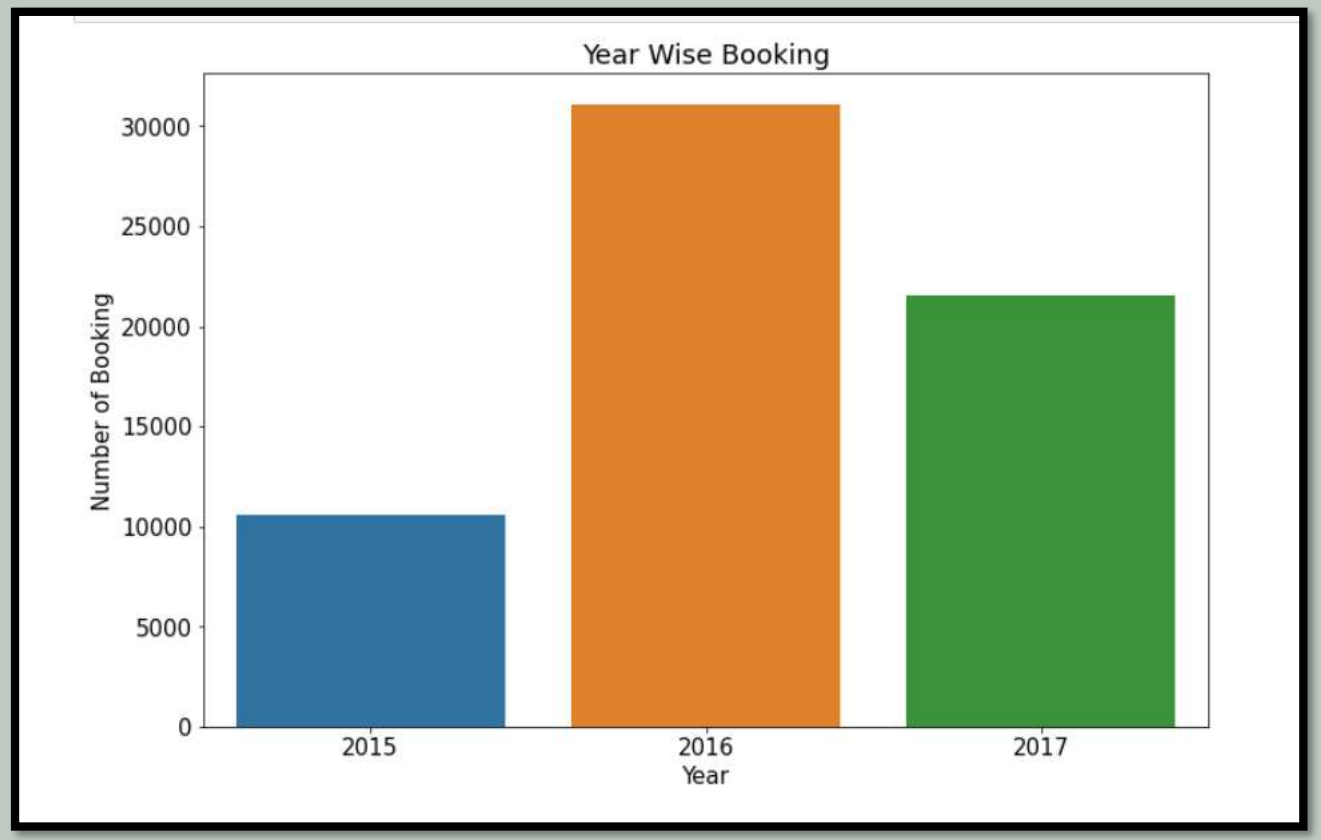
```
In [9]: # DataFrame of duplicate rows

        a[a.duplicated()].shape

Out[9]: (31994, 32)

In [10]: # Drop the duplicate rows from the dataset

         a.drop_duplicates(inplace=True)
```

```
In [22]: # Here datatype of "children","agent" and "company" columns is float
         # Convert float into int

         a[['children','agent','company']]=a[['children','agent','company']].astype('int')

In [23]: # Datatype of column 'reservation_status_date' is convert in to datetime format

         a['reservation_status_date'] = pd.to_datetime(a['reservation_status_date'], format = '%Y-%m-%d')
```

```
In [12]: # How many null values in the data

         a.isnull().sum().sort_values(ascending=False)[:5]

Out[12]: company      82137
         agent        12193
         country        452
         children         4
         lead_time        0
         dtype: int64

In [13]: # In the "company" and "agent" columns we can take a value is 0 inplace of NaN,
           #because sometimes customer not book the hotel through company and agent.

         a[["company","agent"]]=a[["company","agent"]].fillna(0)

In [14]: # For missing values of"children" column value, replace it with mean value of children.

         a["children"].fillna(round(a["children"].mean()),inplace=True)

In [15]: # For the missing values of "country" column, replace it with mode of this column. (value that is most repeated)

         # Find the mode
         a["country"].mode()

Out[15]: 0    PRT
         dtype: object

In [16]: # Replace  NaN  with  "PRT" which is mode

         a["country"].fillna("PRT",inplace=True)
```

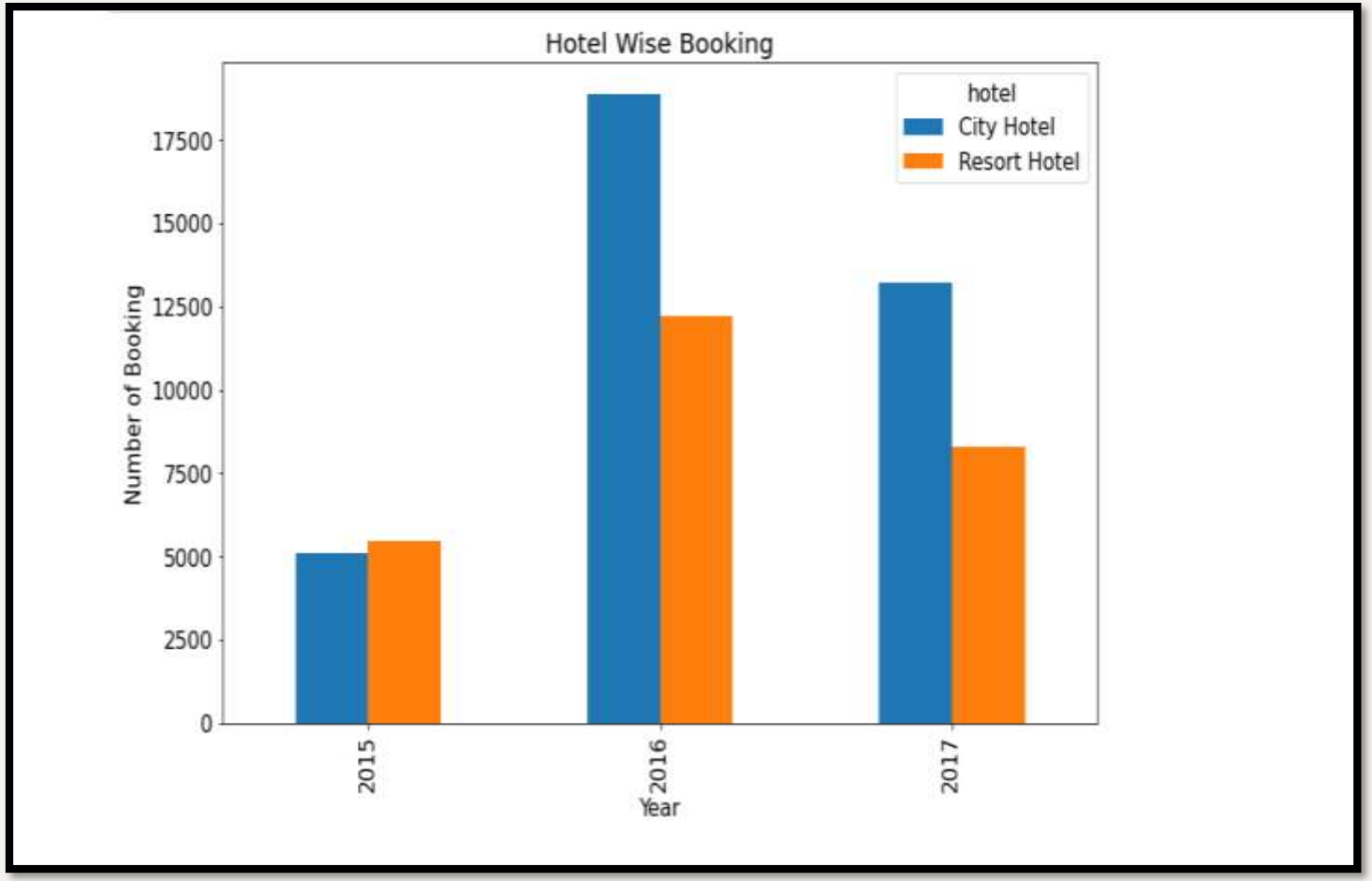
02 Exploratory Data Analysis (EDA)

Q1 Which year have a more number of bookings ?



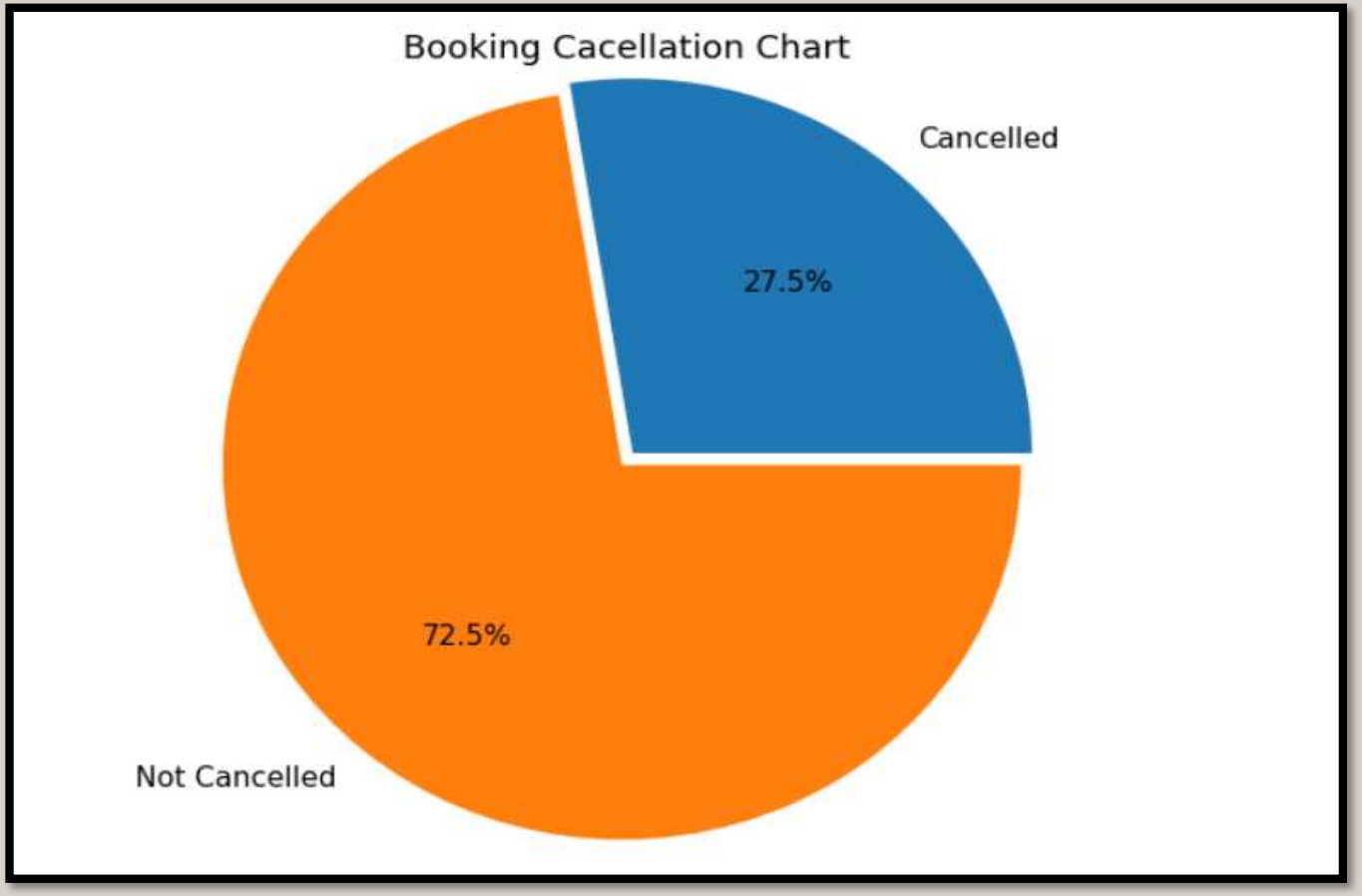
- 2016- Highest number of bookings
- 2015- Lowest number of bookings

Q2 How many bookings are done in every year according to the hotel ?



- City Hotel has more bookings than Resort Hotel.

Q3 What is the booking cancellation ratio ?

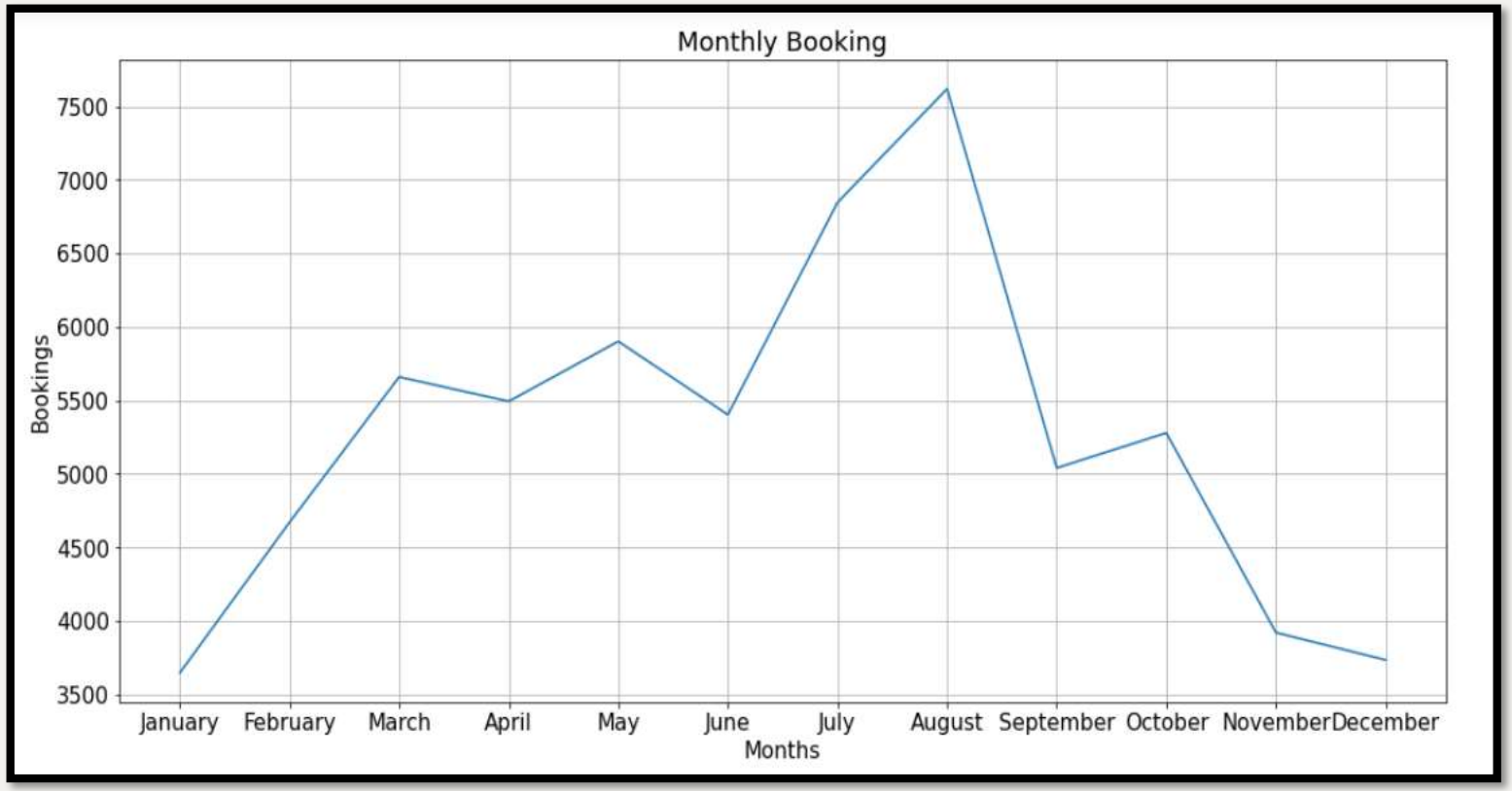


- 27.5% bookings are cancelled

Q4 Which hotel has higher bookings cancellation rate ?

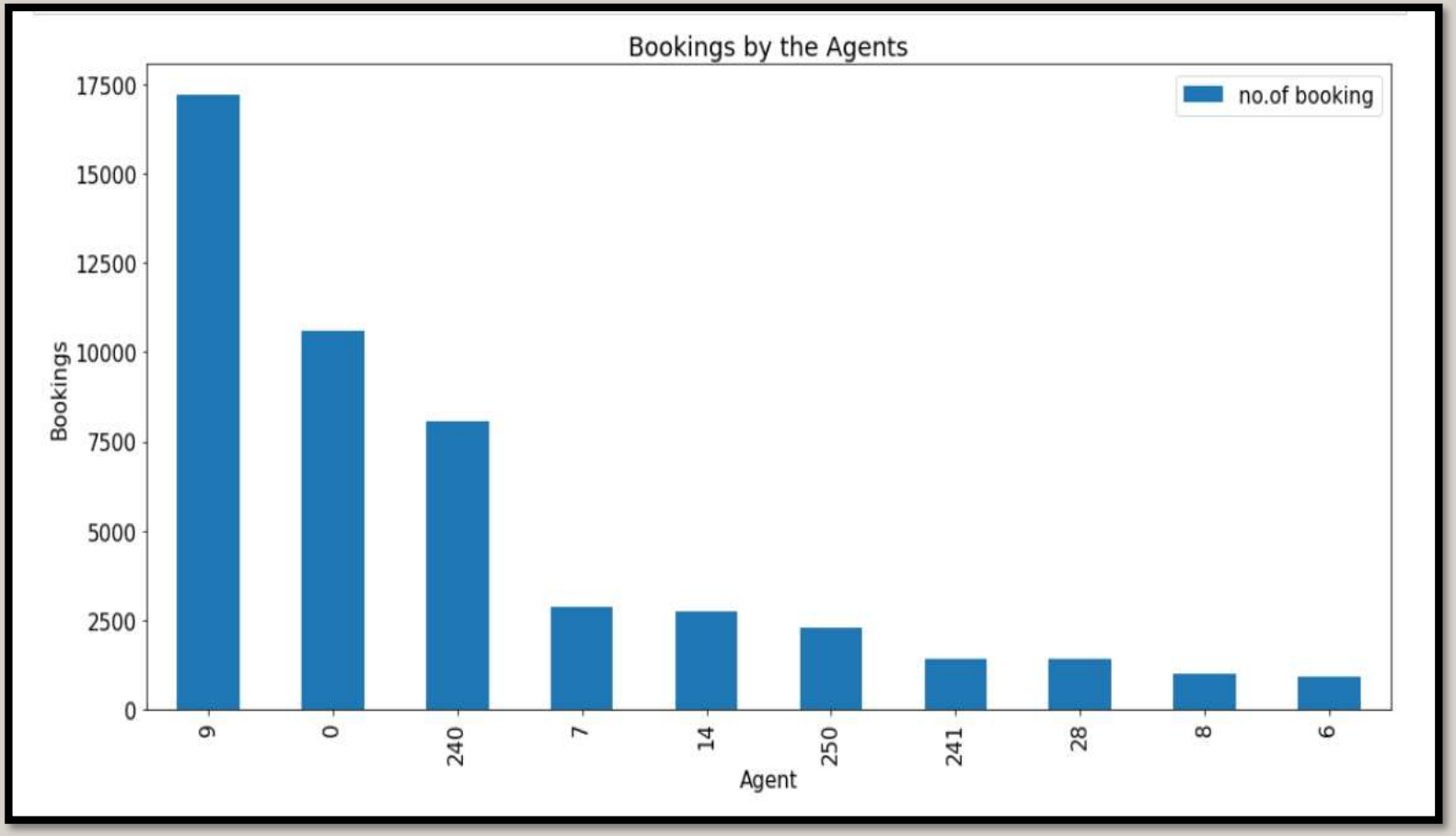
- City Hotel(30%) has higher cancellation percentage than Resort Hotel(23%).

Q5 Which month has maximum bookings ?



- Bookings increase from January to August while bookings decrease from August to December.
- August is the busiest month.

Q6 Which agent does the most bookings?

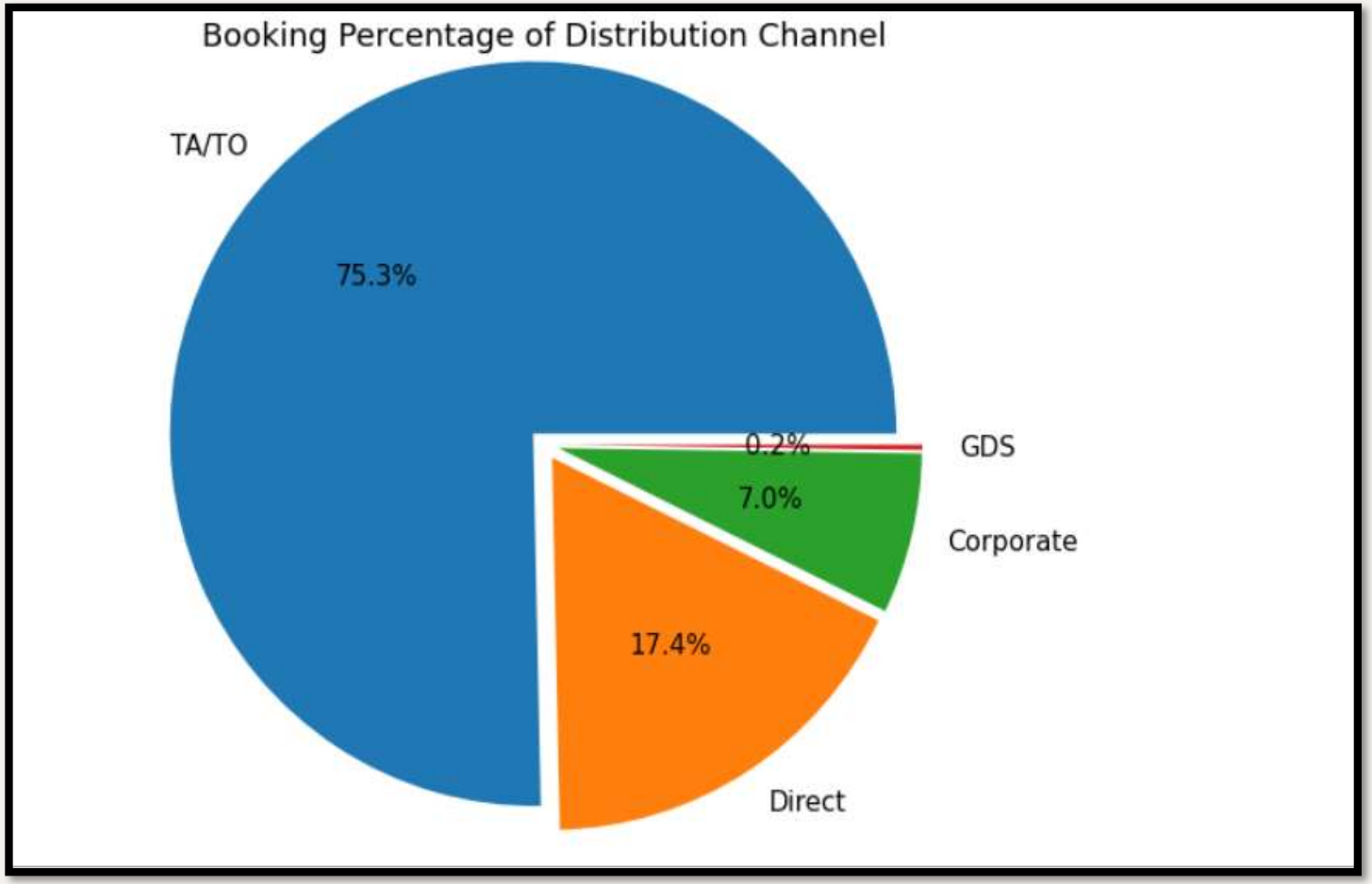


- Most bookings are done by Agent 9.

Q7 Which agent has the lowest booking cancellation ratio ?

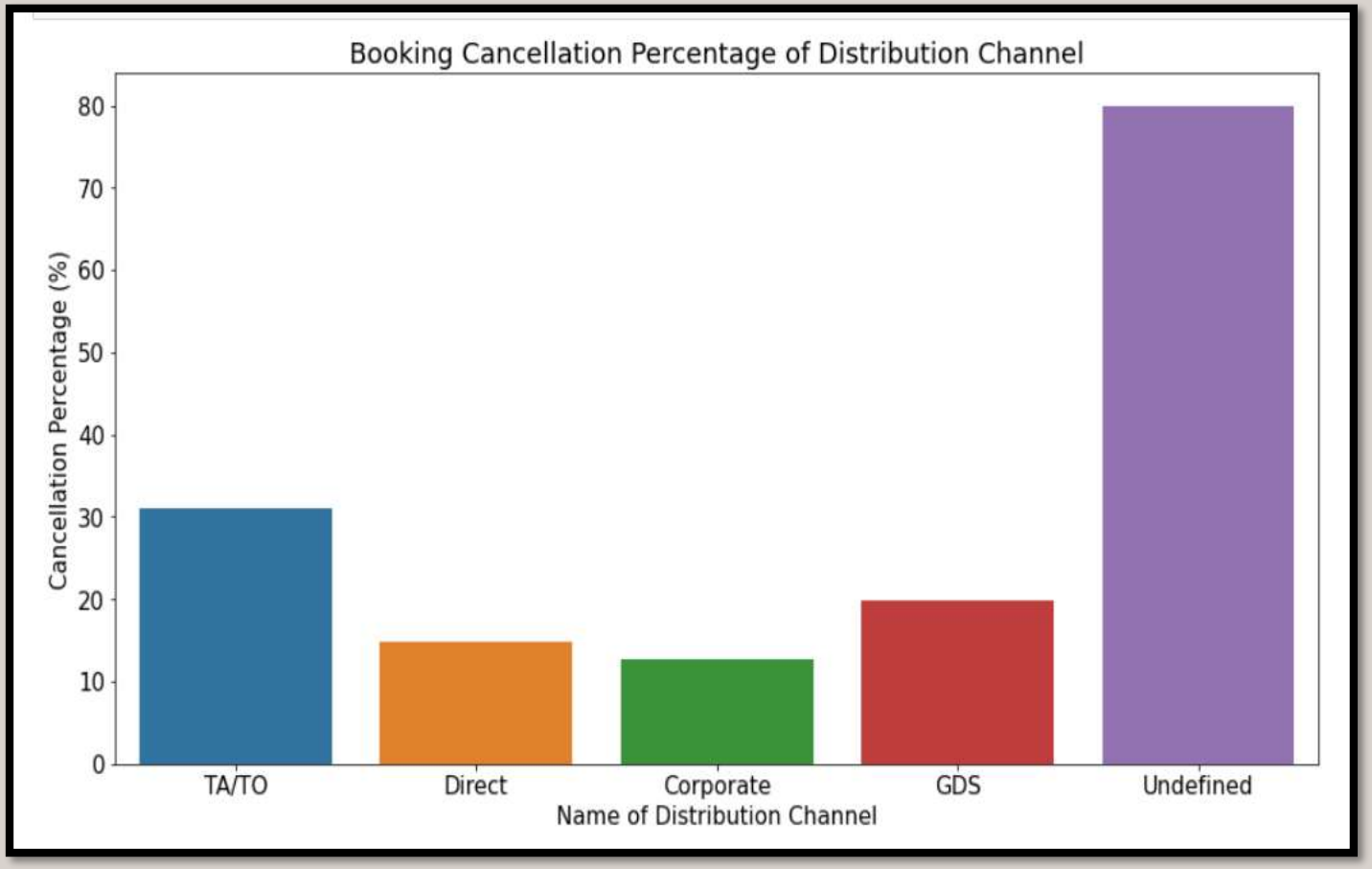
- Agent 464 has the lowest booking cancellation rate.

Q8 Which distribution channel has highest booking rate?



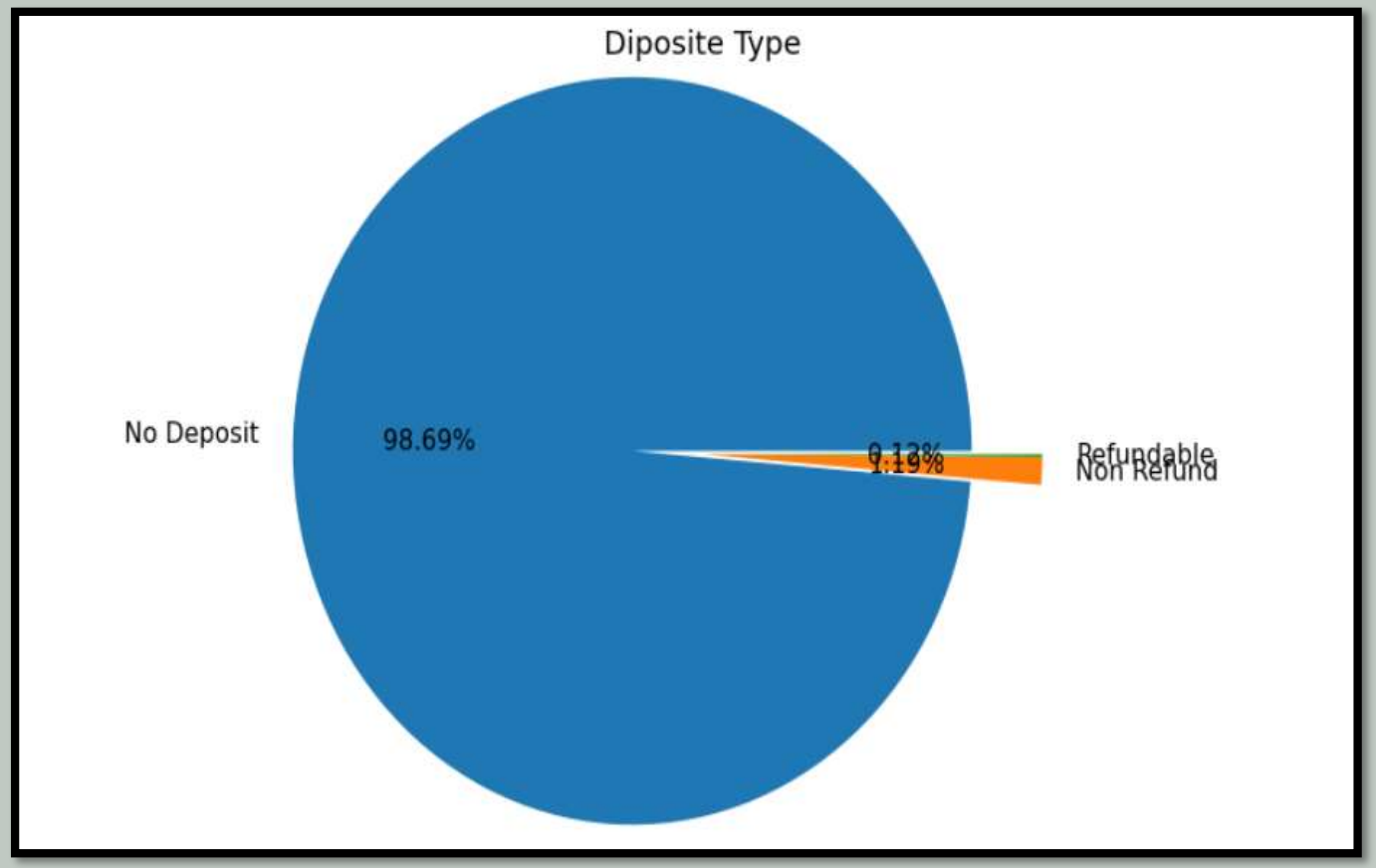
- TA/TO distribution channel does the highest bookings (75.30%).

Q9 Which distribution channel has highest cancellation percentage?



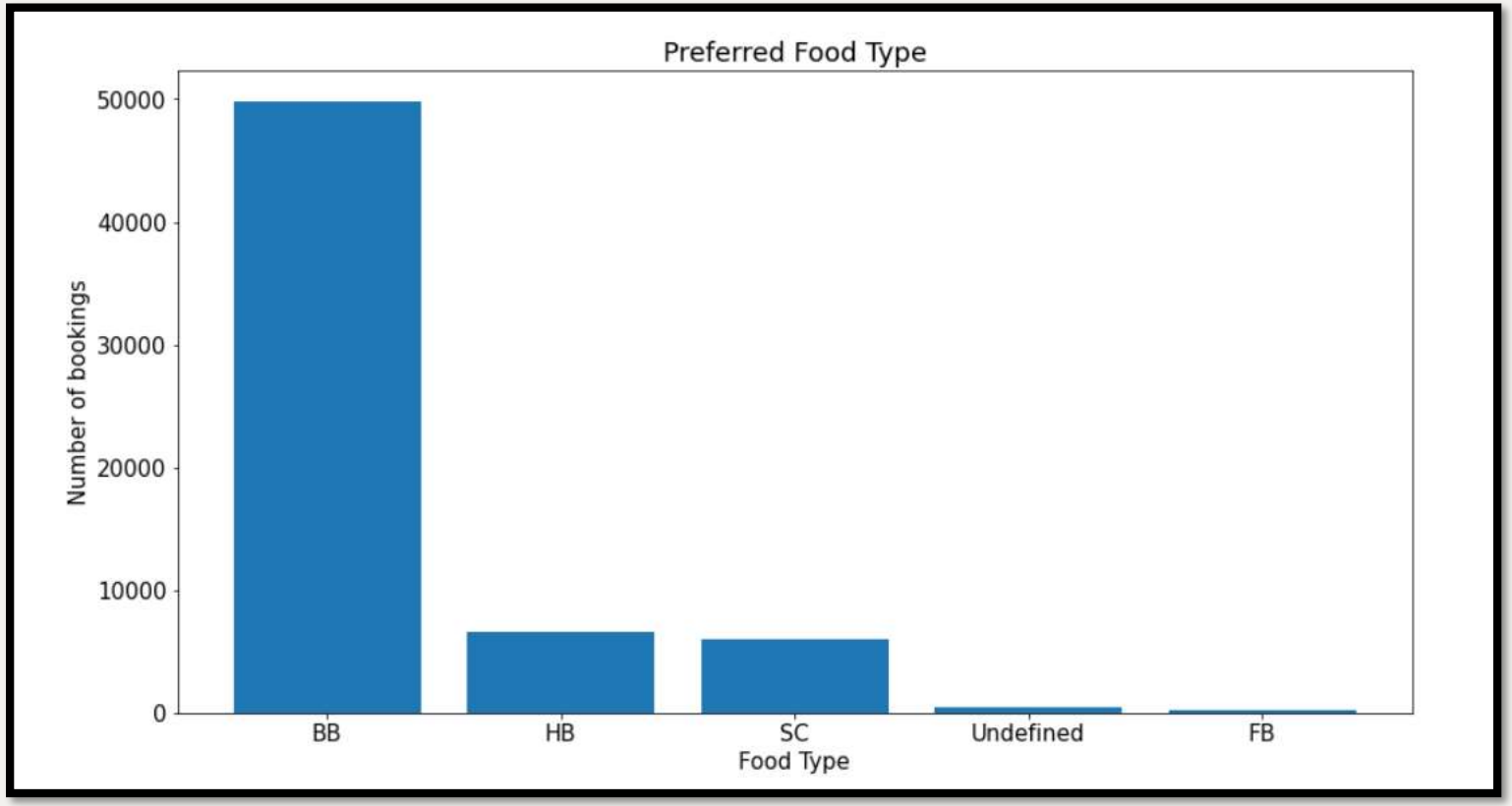
- Undefined (80%) and TA/TO (31%) has the highest cancellation percentage.

Q10 Which type of deposit is preferred by customers ?



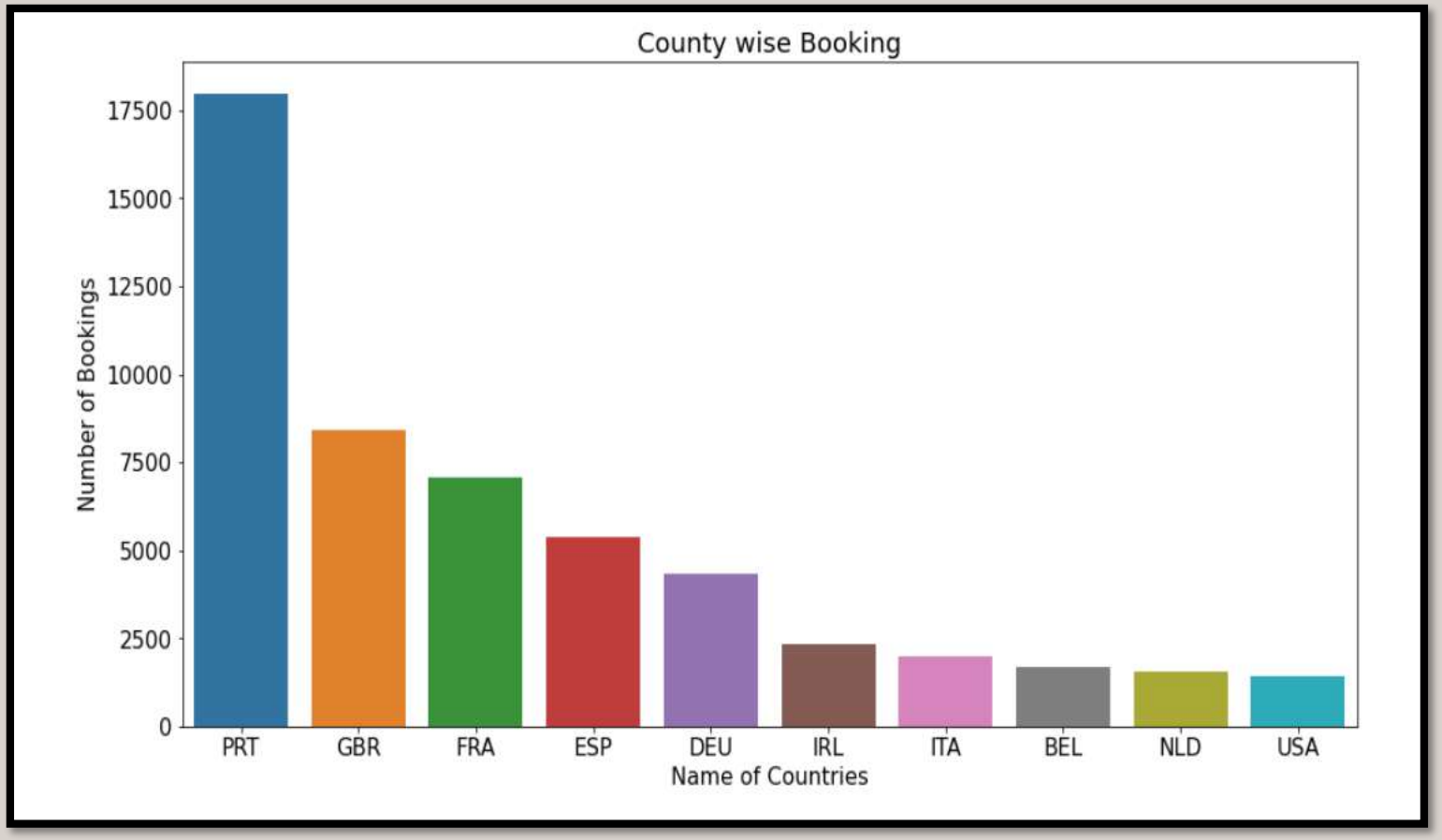
- The most preferred deposit type is the "NO DEPOSIT".
- Due to this, the Booking Cancellation Rate increases.

Q11 Which type of food(meal) is preferred by customers ?



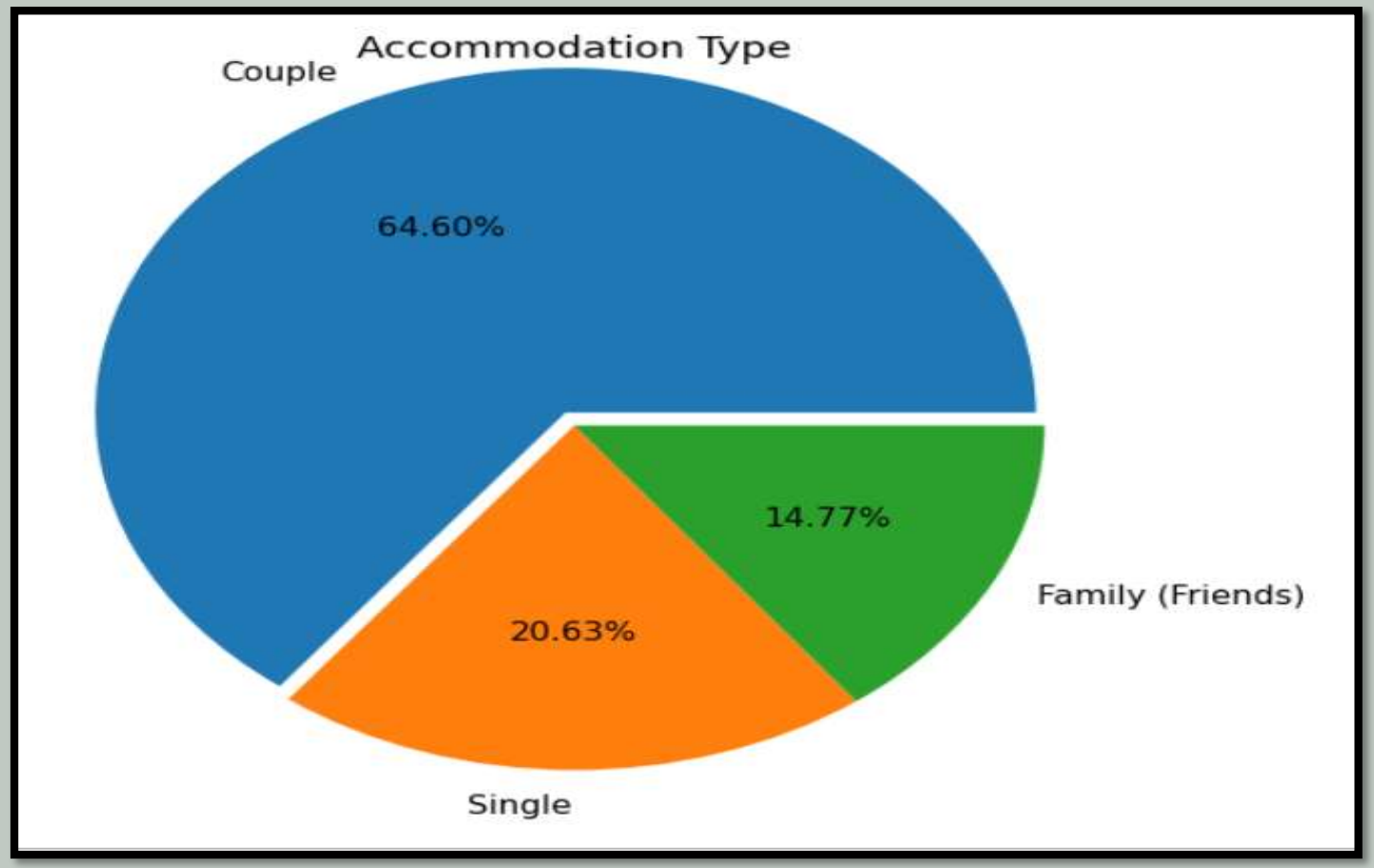
- BB is the most preferred Food Type.

Q12 Which country has the highest number of bookings by customers ?



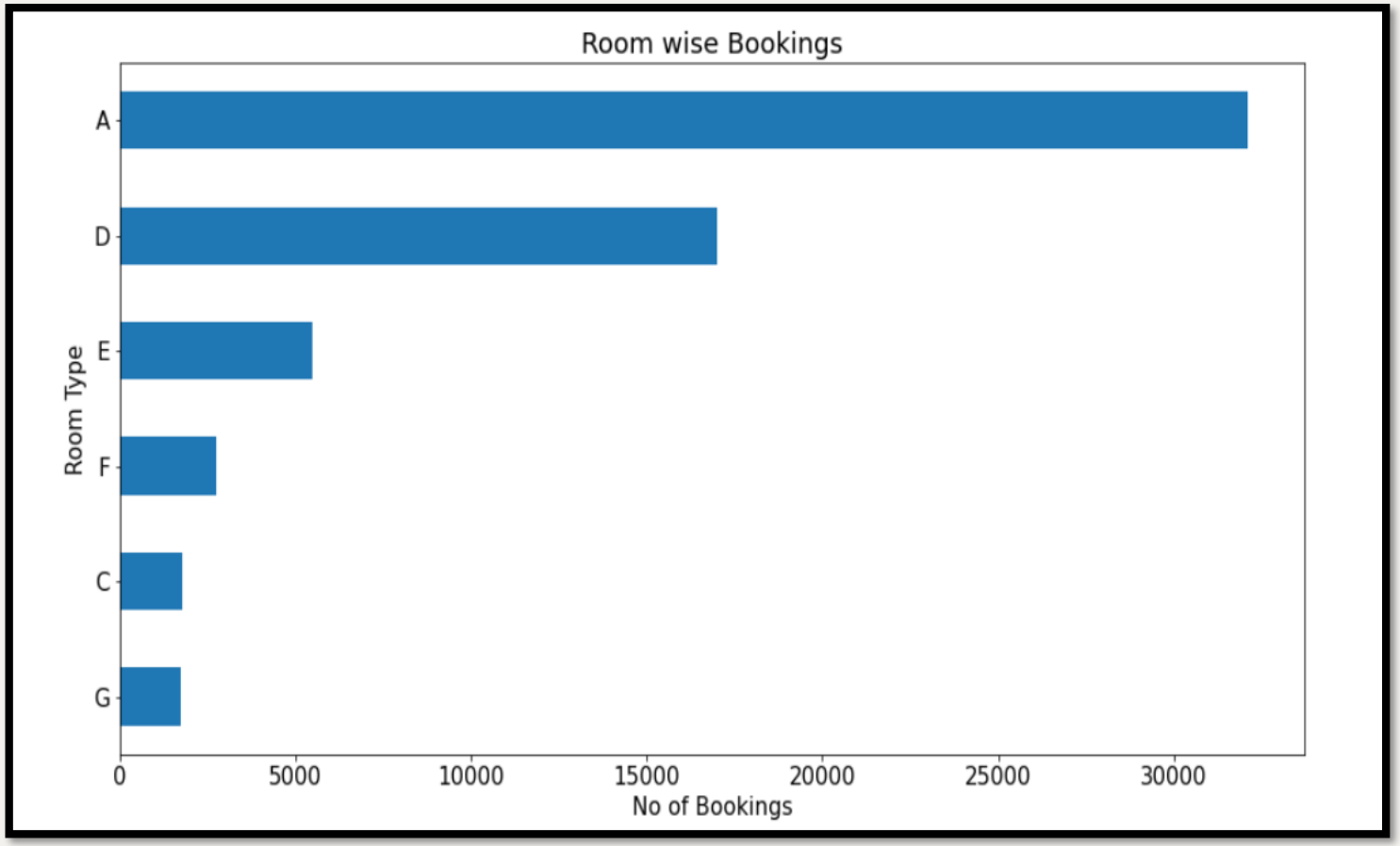
- Generally most of the customers are from European countries.
- Most bookings are done by PORTUGAL (PRT).

Q13 Which is the most booked accommodation type (Single, Couple, Family)?



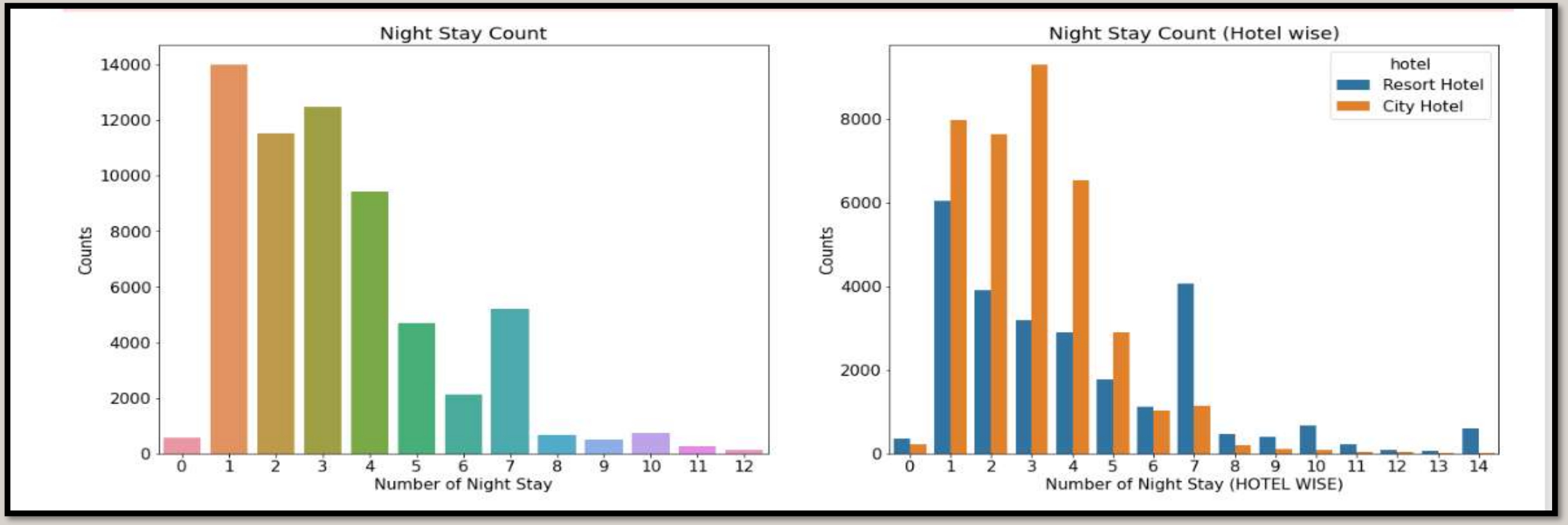
- Usually the highest bookings are done by couple.

Q14 Which type of room is most in demand ?



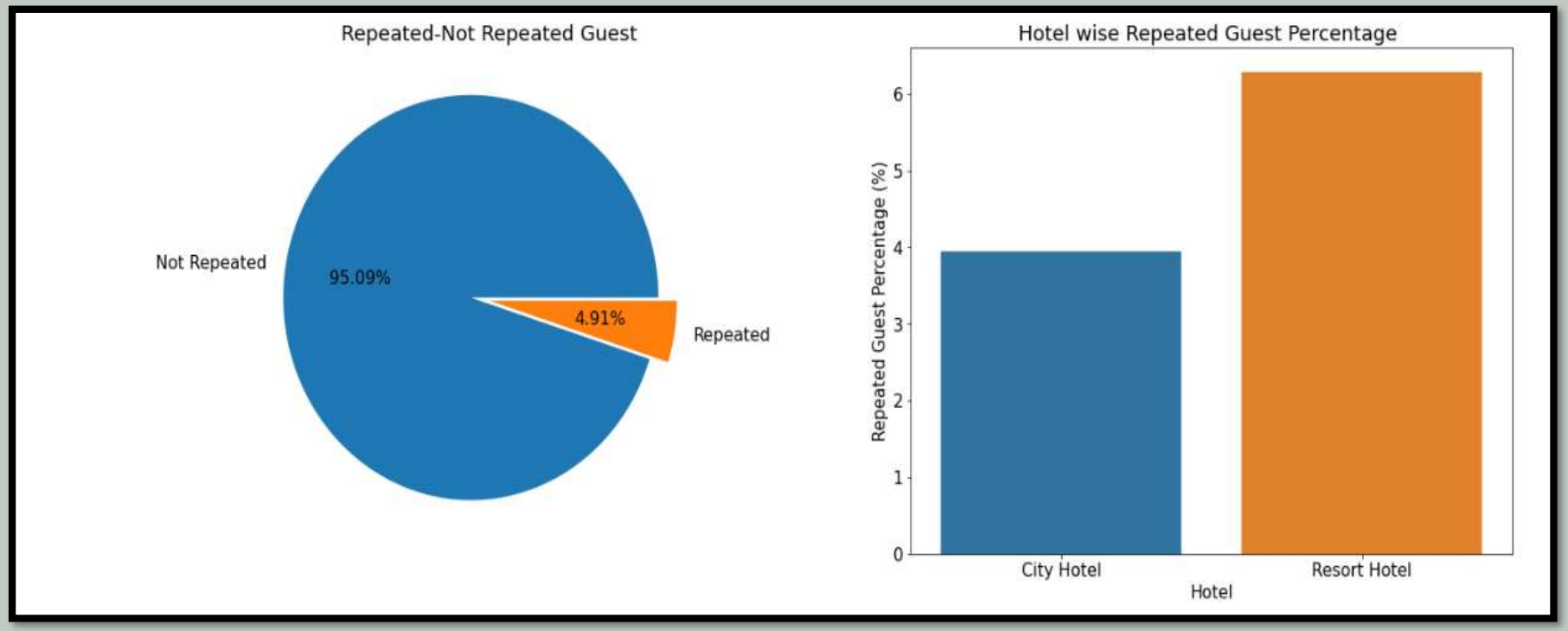
- "A" type of room is most in demand.

Q15 How many nights guests choose to stay the hotel ?



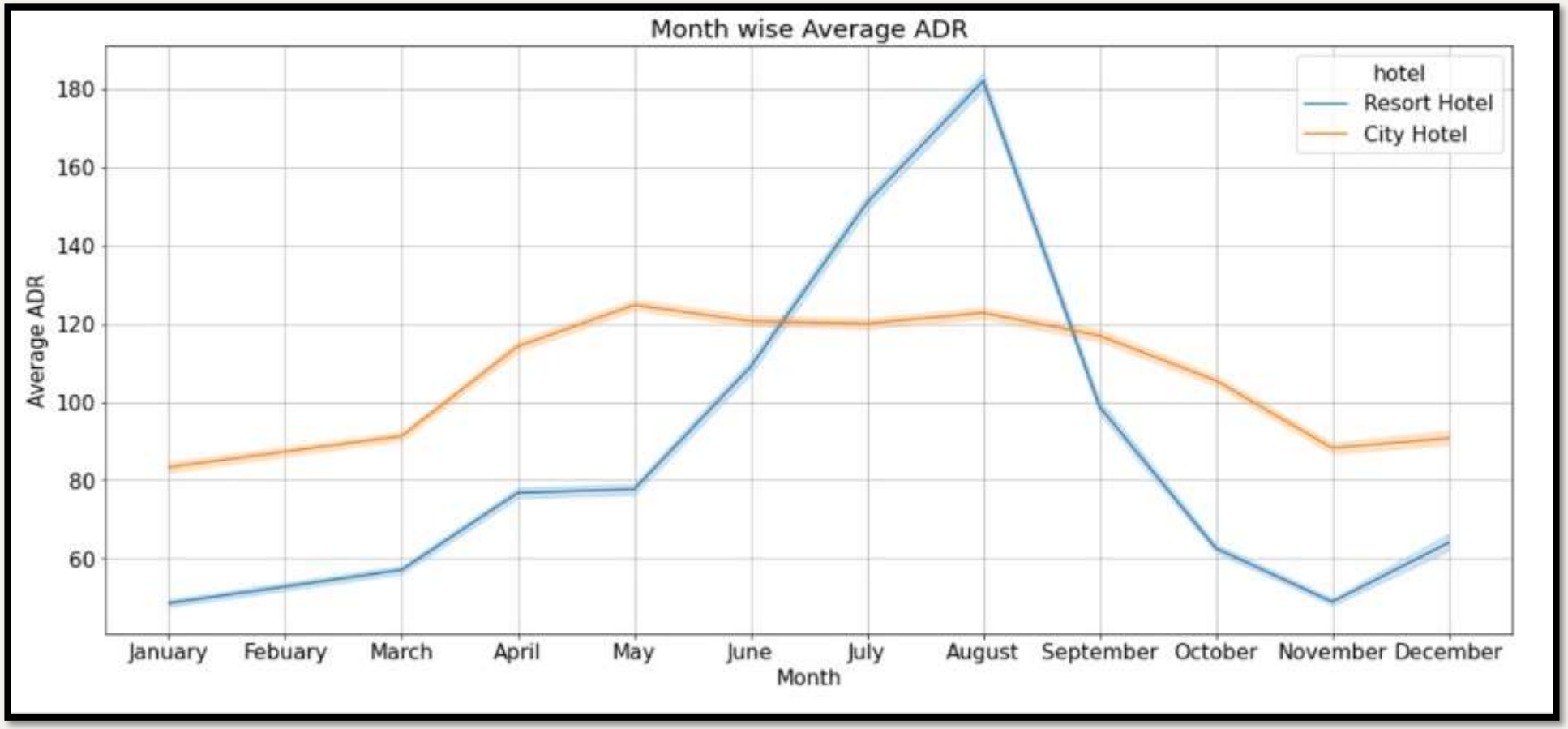
- From the 1st chart it can be said that generally people to stay 1 to 4 nights in hotel.
- From the 2nd chart it can be said that generally people prefer City hotel for short stay, but for long stays Resort Hotel is preferred.

Q16 What is the ratio of repeated guest ?



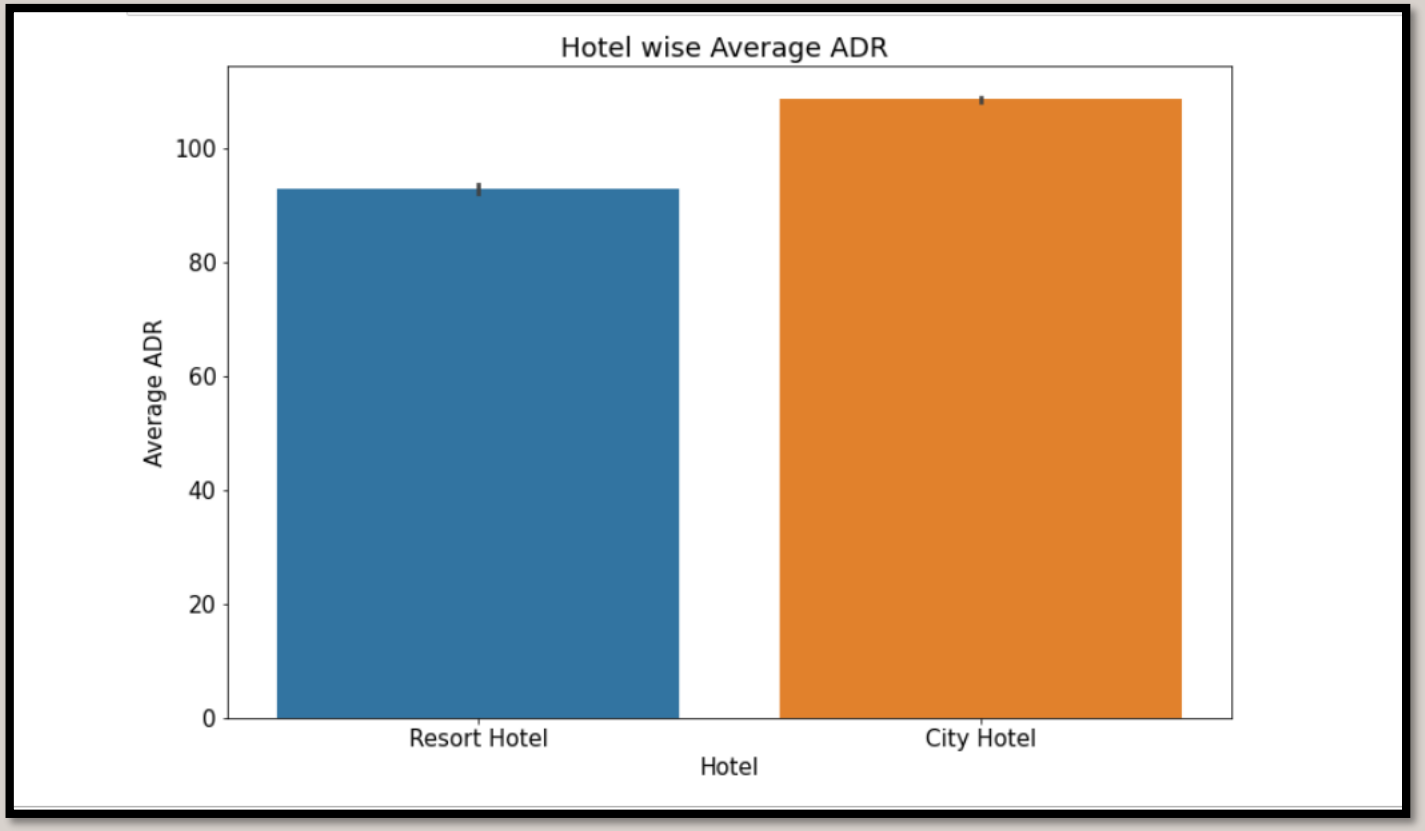
- From the first chart it can be said that the ratio of repeated guest is 4.91% , It is vey less.
- From the second chart it can be said that Resort Hotel has more number of repeated guest than City Hotel.

Q17 Which month has the highest ADR (Hotel wise) ?



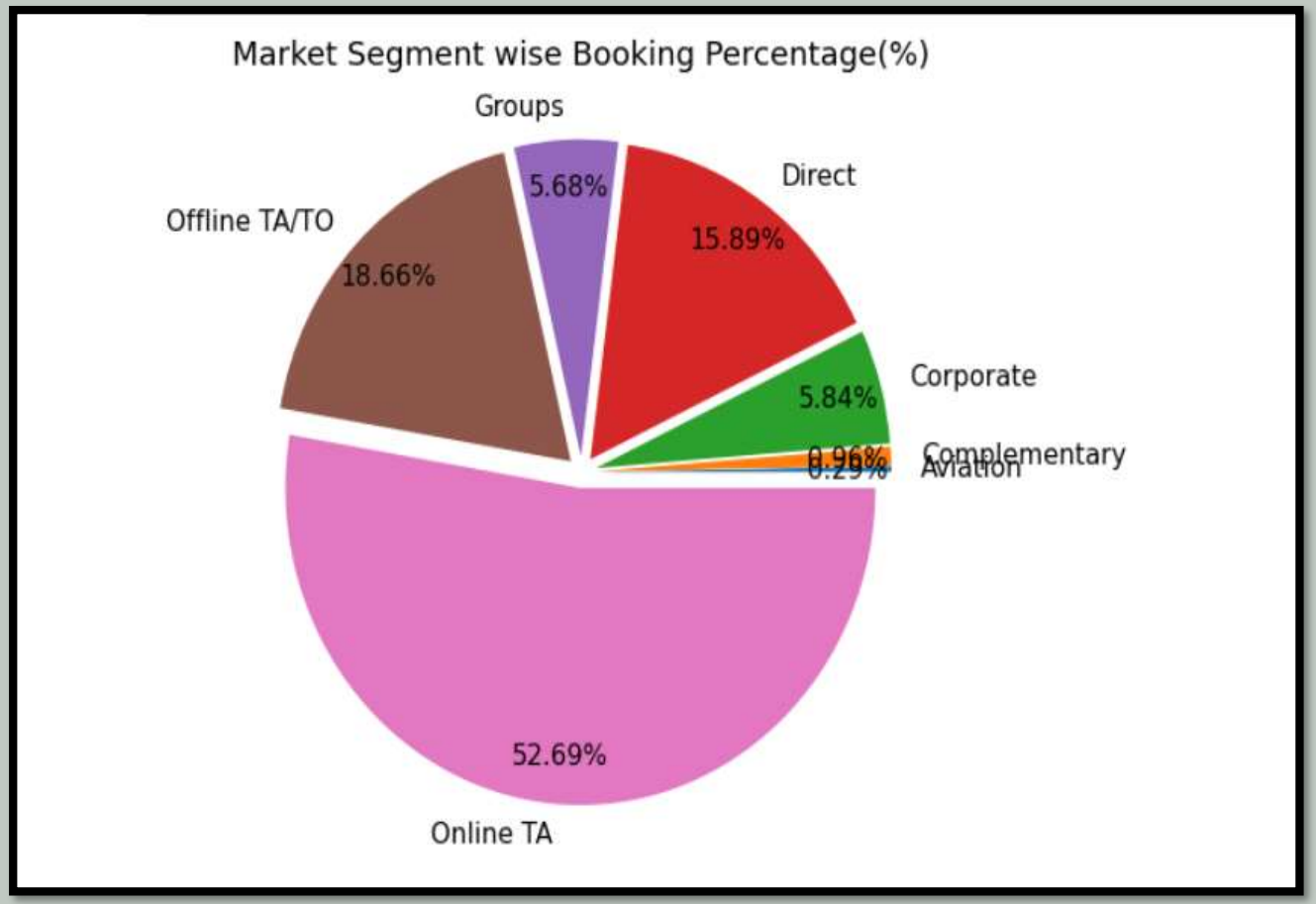
- From the chart it can be said that during the month of July to September the average ADR of the Resort Hotel is higher than City hotel while in rest of months average ADR of the City Hotel is higher than Resort Hotel.

Q18 What is the average ADR of hotel ?



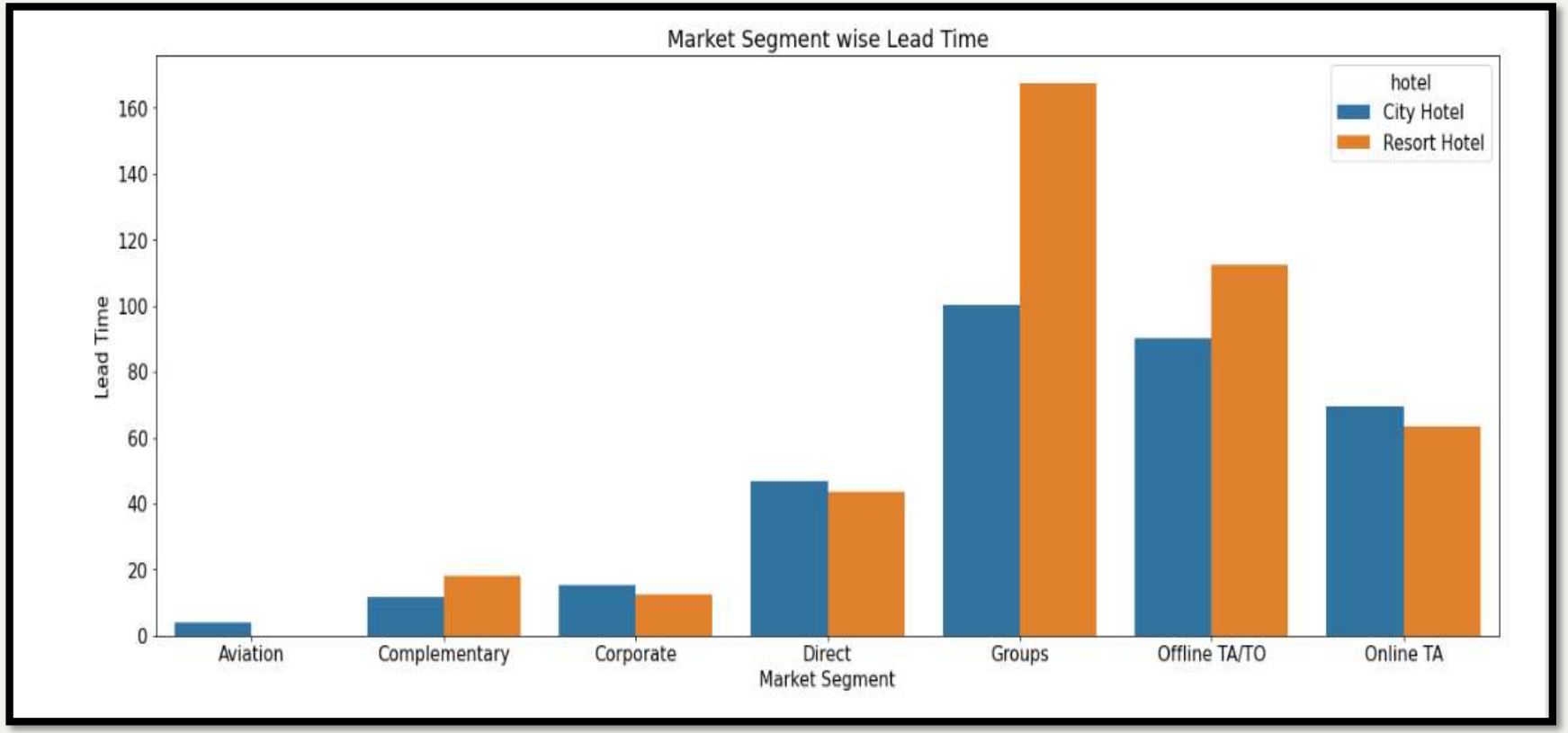
- Generally the average ADR of Resort Hotel is lower than City Hotel.

Q19 By which market segment are the most bookings done ?



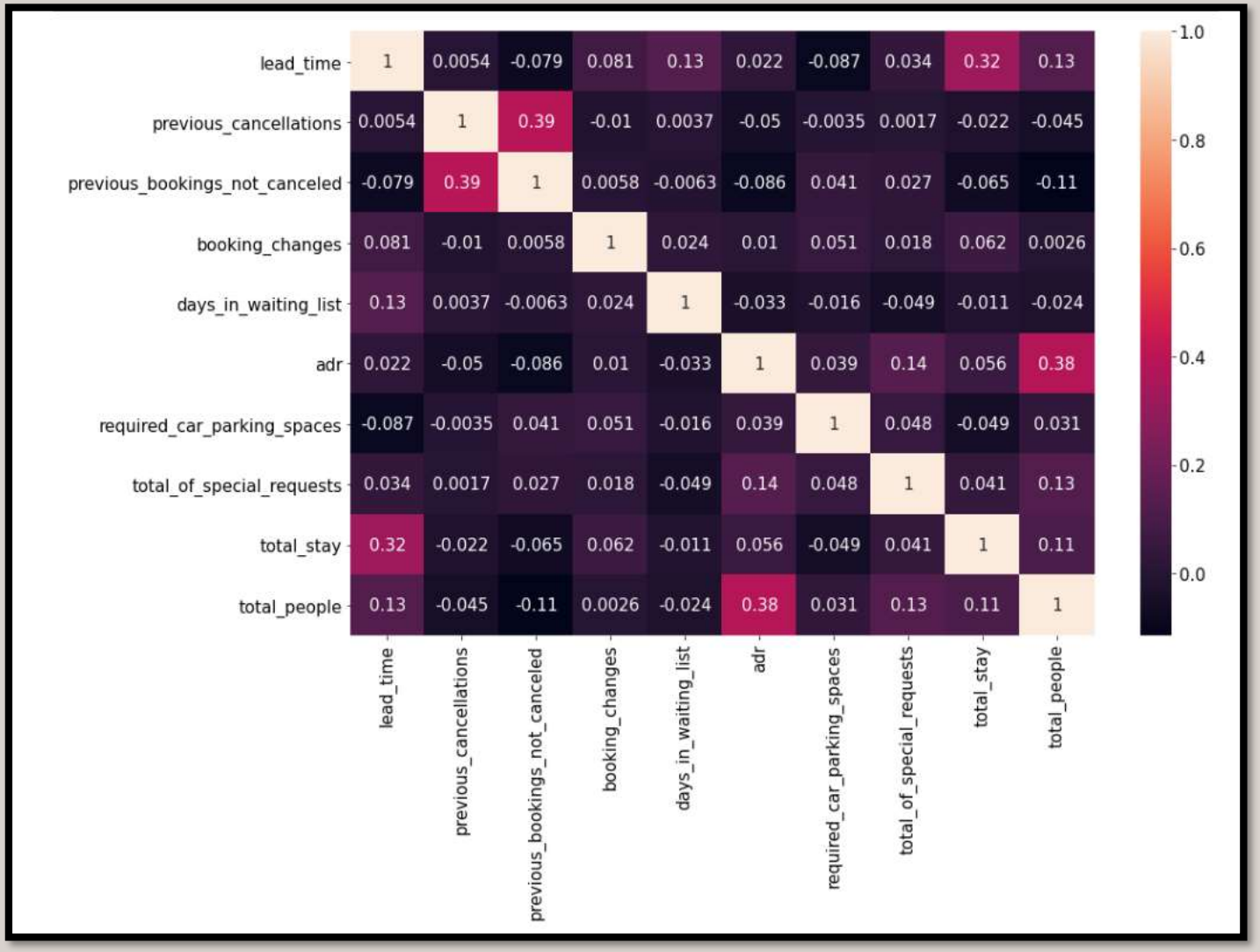
- The highest bookings is done by "Online TA".

Q20 Which market segment has the longest lead time ?



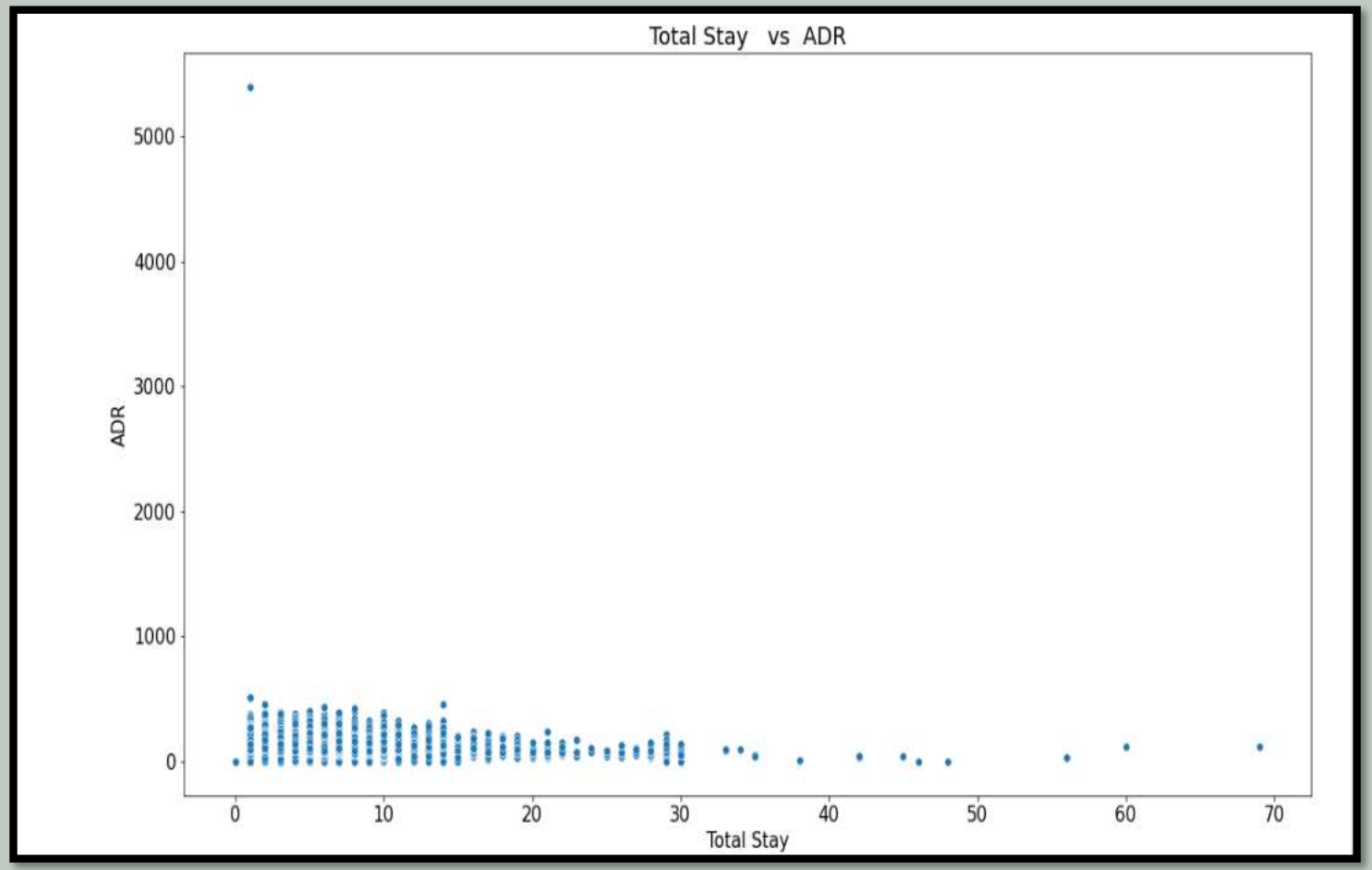
- "Groups" type Market Segment has the longest lead time for both Hotels. so we can say that early booking is essential for "Groups" type Market Segment

Q21 Correlation between two parameters.



- "Total stay length" and "lead time" have slight correlation, that means for longer hotel stays people generally plan little before the actual arrival.
- "ADR" is slightly correlated with "total_people", which makes sense as more number of people means more revenue, therefore more ADR.

Q22 Relationship between number of Total Stay and ADR.



- From the chart it can be said that length of Total Stay increases the ADR decreases. That means longer stay is beneficial for customer.

Conclusion

- Overall adr of City hotel is slightly higher than Resort hotel.
- Mostly guests stay for less than 4 nights in hotel and for longer stays they preferred Resort hotel.
- 4% of visitors return for another booking at the City hotel, and 6% of visitors return to the Resort hotel for another stay.
- Europeans and Portuguese are the majority of guests.
- Mostly preferred channel for booking is TA/TO
- Booking cancellation ratio in TA/TO is 30%

Conclusion

- July- August are the most busier and profitable months for both of hotels.
- Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.
- Low adr for longer stays (more than 15 days)
- Booking cancellation ratio- 27.5%.
- Agent 9 is done more number of bookings while Agent 464 has lowest cancellation ratio.
- When it comes to longer stays in hotel, guest do plan much in advance.

Thank
You