

CAPSTONE PROJECT

HOTEL BOOKING ANALYSIS

Team Members

Kirtesh Verma

Pravin Bejjo

Sahil Pardeshi

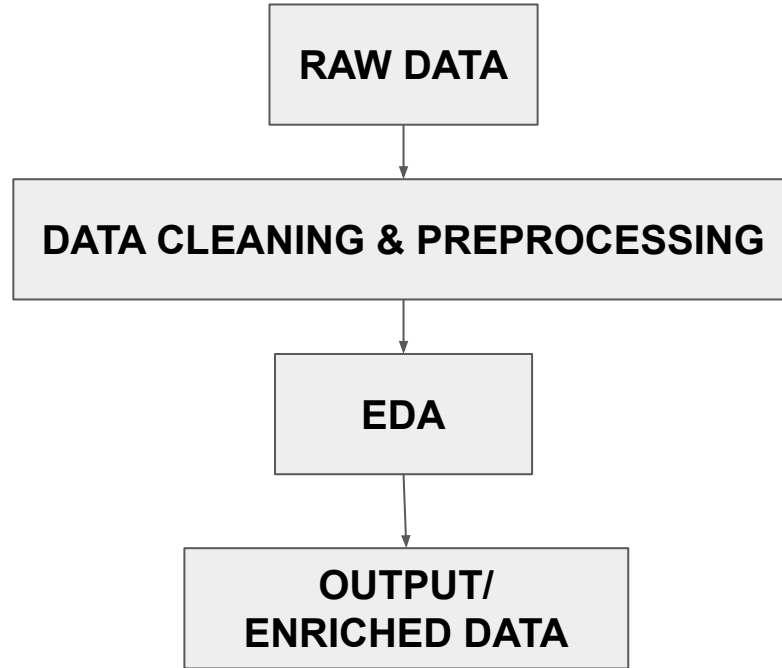
➤ STEPS

- **Introduction to EDA**
- **Data Understanding**
- **Data Preprocessing**
- **Problem statement**
- **Data Visualization**
- **Conclusion**

➤ INTRODUCTION TO EDA

- EDA stands for Exploratory Data Analysis
- It's an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.
- We will perform EDA on our data set of Hotel Booking and will try to understand features that governs the bookings.

➤ **PROCESS FLOW**



➤ DATA UNDERSTANDING

- The Hotel booking dataset is in .csv format.
- Hotel booking dataset contains record of hotel booking of a city and resort hotel from year 2015 - 2017.
- The dataset contains information like bookings, booking type, agent, etc.
- There are 119390 entries and 32 columns.
- This data has 3 type of data type i.e float(64), int64 and object.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   hotel                                119390 non-null object
1   is_canceled                          119390 non-null int64
2   lead_time                           119390 non-null int64
3   arrival_date_year                    119390 non-null int64
4   arrival_date_month                   119390 non-null object
5   arrival_date_week_number             119390 non-null int64
6   arrival_date_day_of_month            119390 non-null int64
7   stays_in_weekend_nights              119390 non-null int64
8   stays_in_week_nights                 119390 non-null int64
9   adults                               119390 non-null int64
10  children                             119386 non-null float64
11  babies                              119390 non-null int64
12  meal                                 119390 non-null object
13  country                              118902 non-null object
14  market_segment                       119390 non-null object
15  distribution_channel                  119390 non-null object
16  is_repeated_guest                    119390 non-null int64
17  previous_cancellations                119390 non-null int64
18  previous_bookings_not_canceled        119390 non-null int64
19  reserved_room_type                   119390 non-null object
20  assigned_room_type                   119390 non-null object
21  booking_changes                       119390 non-null int64
22  deposit_type                         119390 non-null object
23  agent                                103050 non-null float64
24  company                              6797 non-null float64
25  days_in_waiting_list                 119390 non-null int64
26  customer_type                         119390 non-null object
27  adr                                  119390 non-null float64
28  required_car_parking_spaces           119390 non-null int64
29  total_of_special_requests             119390 non-null int64
30  reservation_status                   119390 non-null object
31  reservation_status_date               119390 non-null object

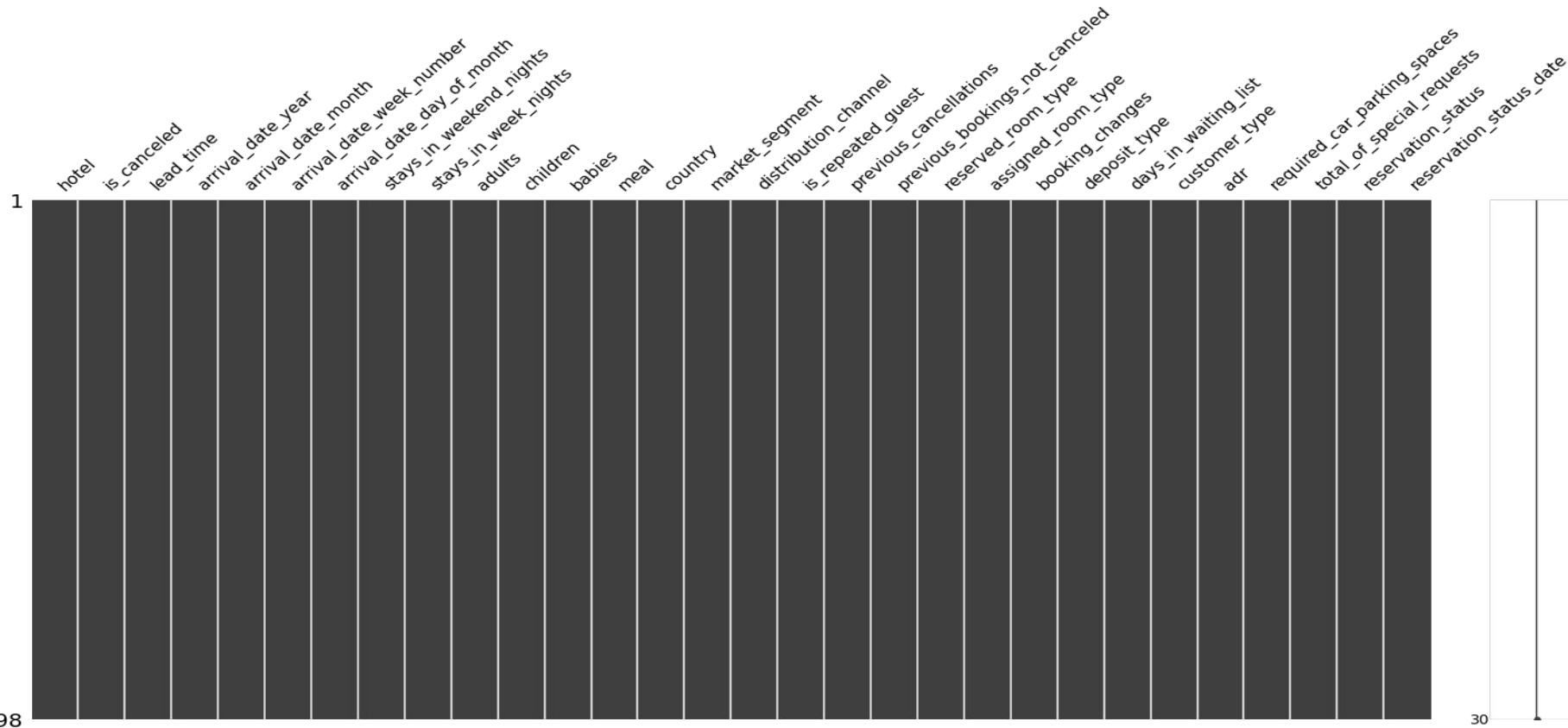
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

➤ DATA PREPROCESSING

- A null value indicates an unknown value.
- It is important to remove null values so that the chances of error during visualization is minimum.
- There are 4 columns that are present in our dataset that have null values. They are company, agent, country and children.
- Company column has the highest number of null values.
- As the number of null values are very small in country and children column they can be considered as negligible.

```
company          112593
agent            16340
country           488
children          4
lead_time         0
arrival_date_year 0
arrival_date_month 0
arrival_date_week_number 0
is_canceled       0
market_segment    0
arrival_date_day_of_month 0
stays_in_weekend_nights 0
stays_in_week_nights 0
adults            0
babies            0
dtype: int64
```

DATA PREPROCESSING(Cont..)



➤ **PROBLEM STATEMENT**

Our key objective is to find out the factors that affects the hotel bookings. For this we will explore and visualize the dataset of hotel booking using EDA techniques.

➔ **We will try to answer questions like:**

- **What is the booking ration between city and resort hotel?**
- **What is the cancellation rate between the two types of hotel?**
- **What are the types of guest?**
- **Which months are the busiest month for hotel booking?**
- **What is the most preferred meal type?**

PROBLEM STATEMENT(Cont..)

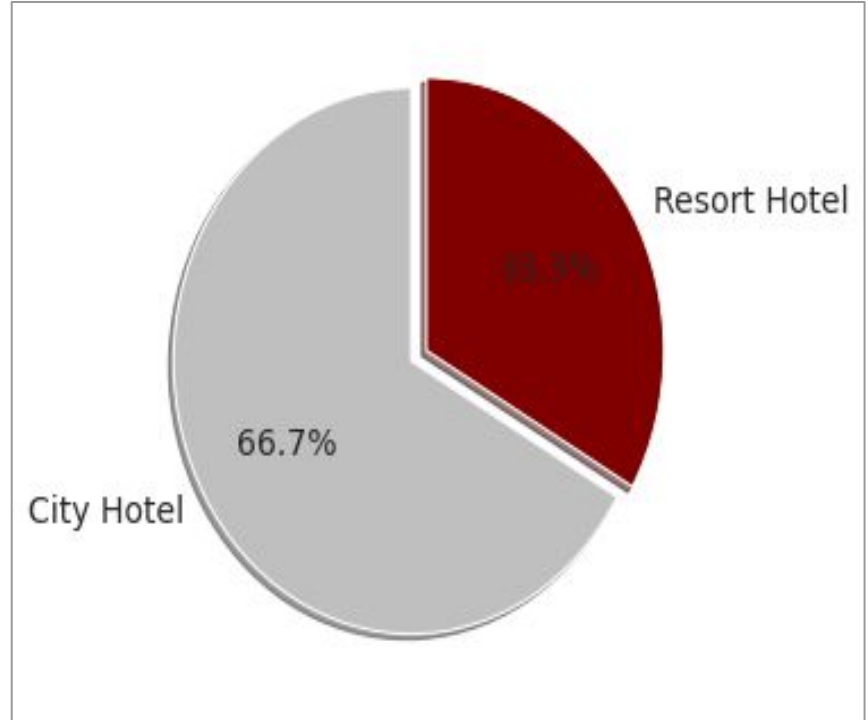
- What is the average stay of people on weekday and weekend night?
- What hotel type received the most number of special request?
- What is the average daily rate in both the hotel type?
- Was there any type of deposit before booking a hotel?
- What type of booking is most preferred by customers(online or offline)?
- From which country most guest comes?

➤ DATA VISUALIZATION



➤ HOTEL VS RESORT BOOKING

- City hotel are booked more than resort hotel.
- Almost 67% of bookings belong to city hotel where as resort hotel has only 33% of bookings.
- In general city hotels are cheaper than resort so people usually prefer city hotels.



➤ CANCELLATION RATE

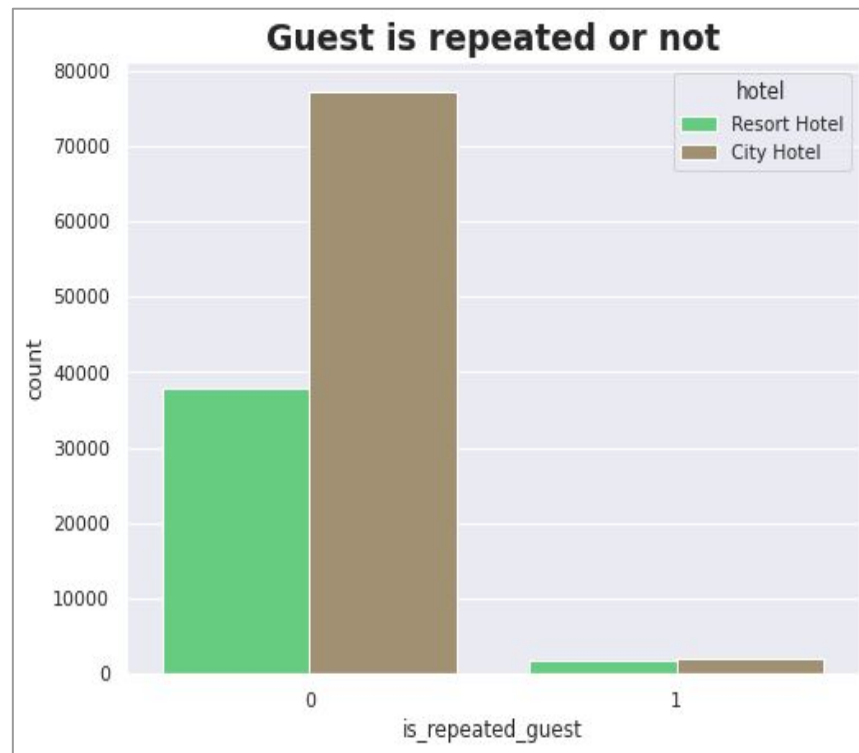
- City hotel has high cancellation rate as compared to resort hote.
- This may be the case due to the reason that city hotel bookings are quite high as as compared to resort hotel.
- Resort hotel falls under well planned trip so they are less likely to get canceled.



➤ GUEST TYPE

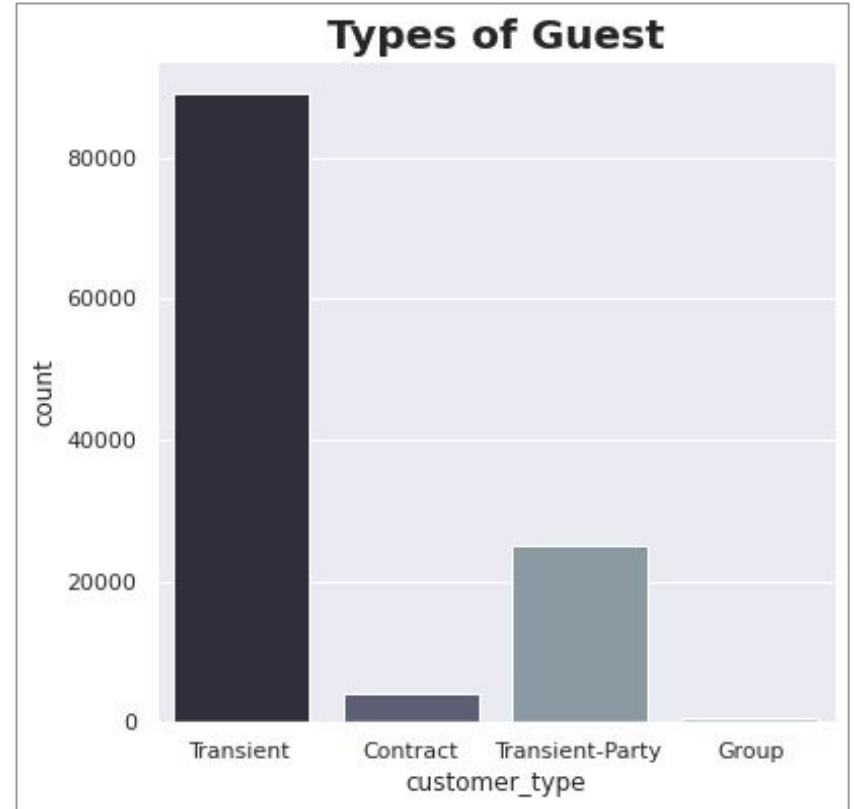
➔ Repeated or not repeated

- We can see that number of repeated guest is very low.
- Most of the customers are one timer and do not book for the second time.



→ TRANSIENT OR NOT

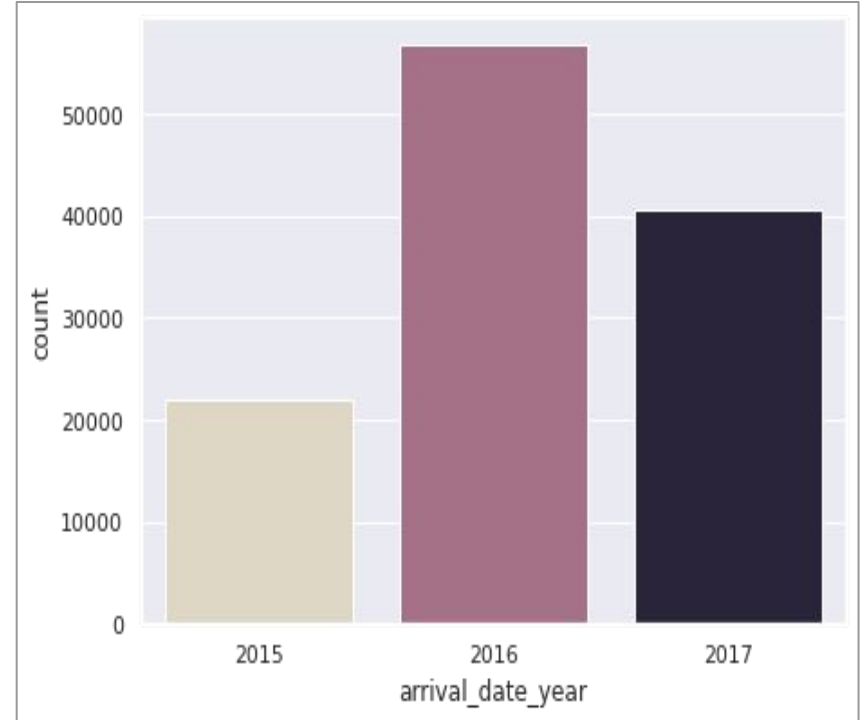
- Transient guest are guests making reservation on their own.
- Transient is the most common customer type.
- They represent 75% of total customers.



➤ OVERVIEW OF ARRIVAL PERIOD

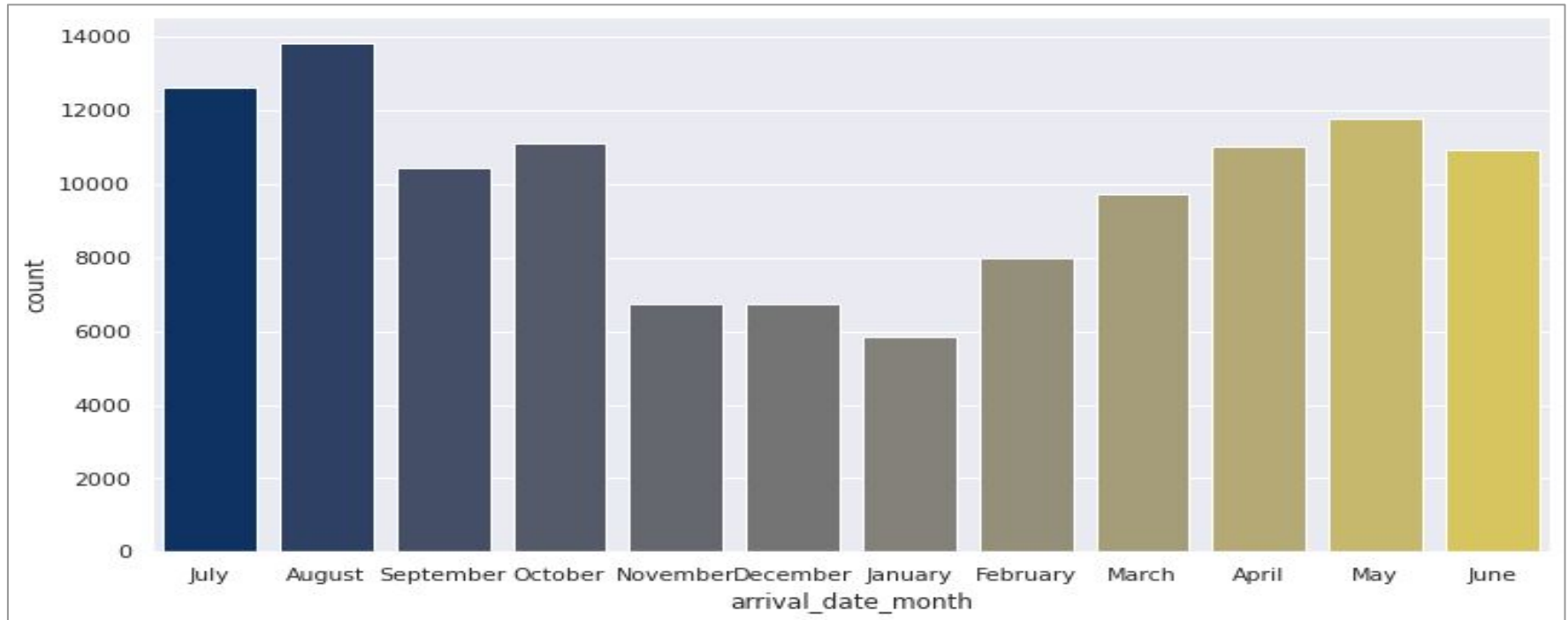
➔ YEAR WISE

- The number of Hotel booking is minimum in the year 2015 and highest in the year 2016.
- Hotel bookings got doubled in the year 2016 when compared to previous year.
- Bookings got decreased in 2017 but is still higher than year 2015.

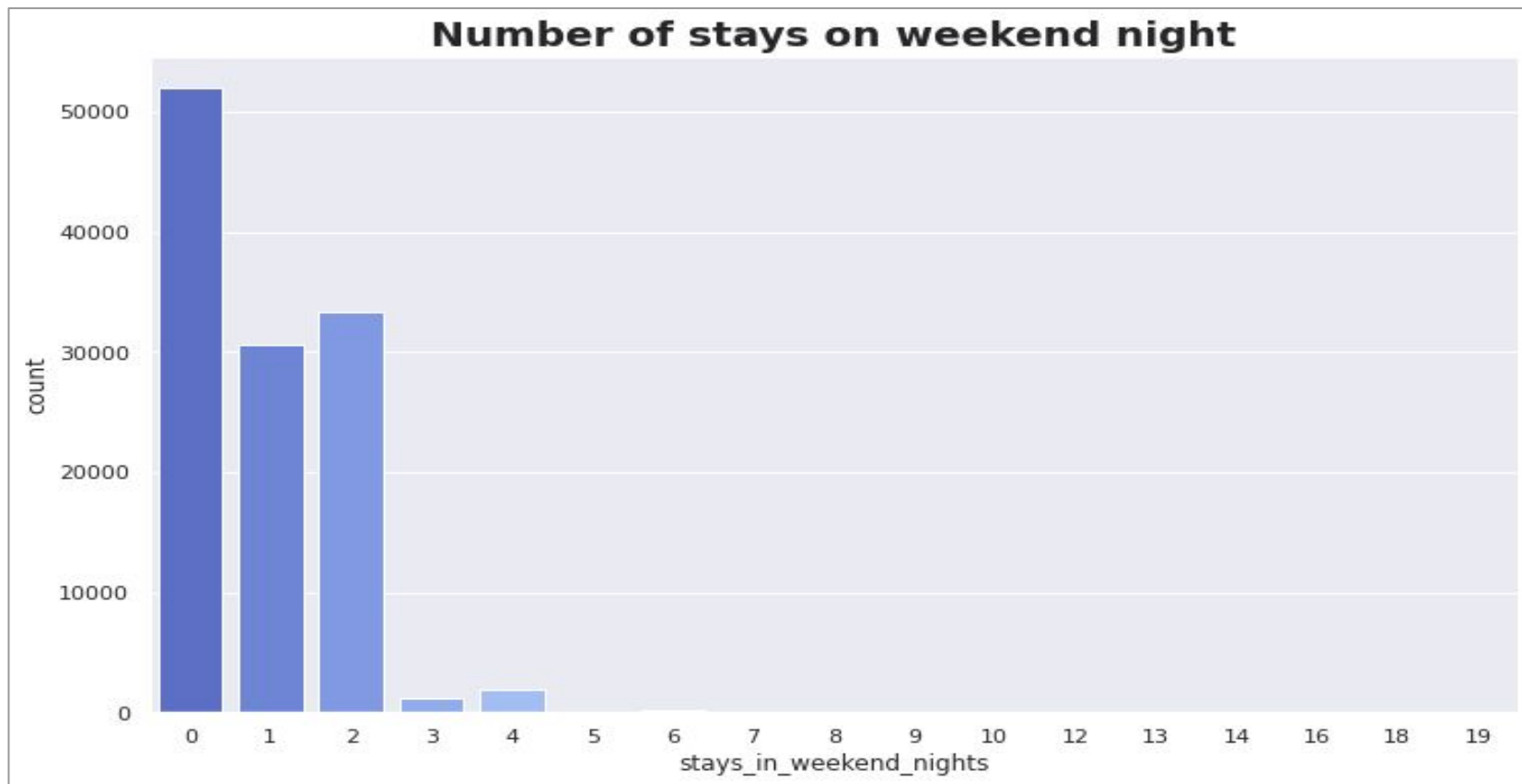


➔ MONTH WISE

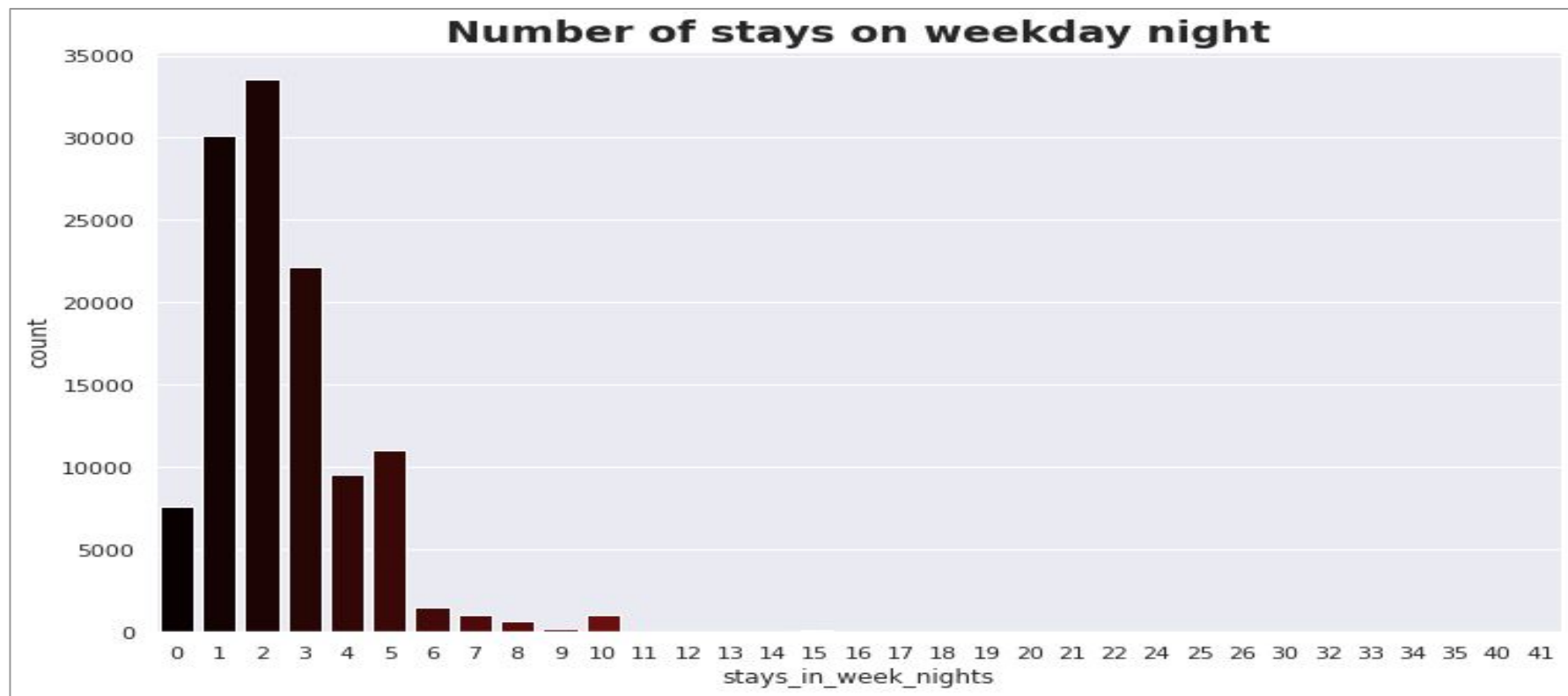
- Highest bookings are made in the month of August followed by July. This are the months of summer and people are more likely to plan trips during summer time.
- Lowest booking are made during winter time.
- January month recorded the lowest booking followed by November, and December



➤ STAY ON WEEKEND NIGHT



➤ STAY ON WEEKDAY NIGHT

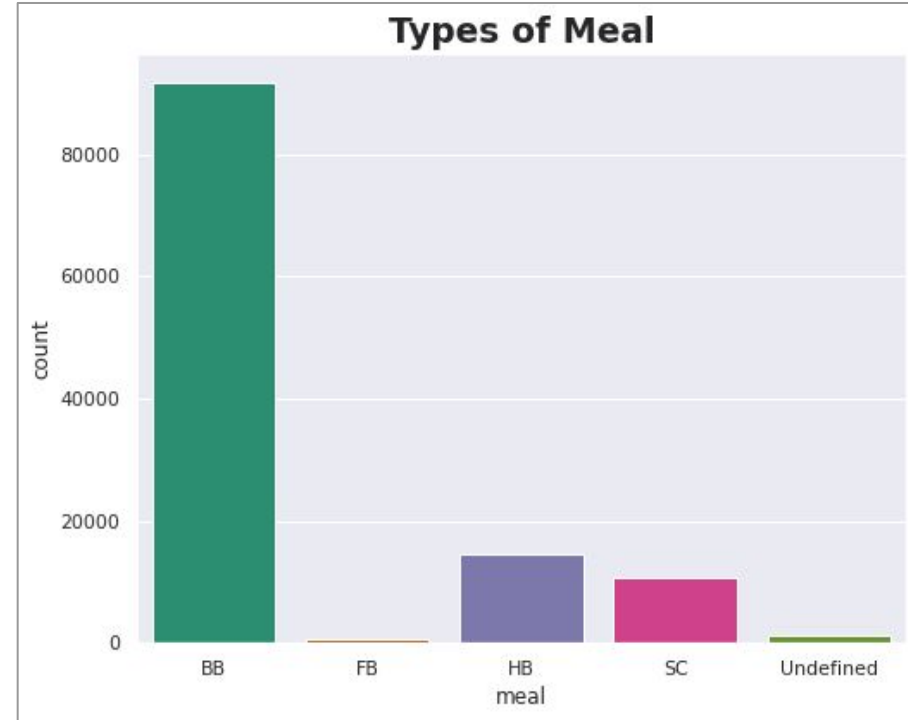


➤ MEAL PREFERENCE

- Guest particularly likes BB type of meal followed by HB and SC.
- More than 77% of time BB type of meal is preferred by the guest.

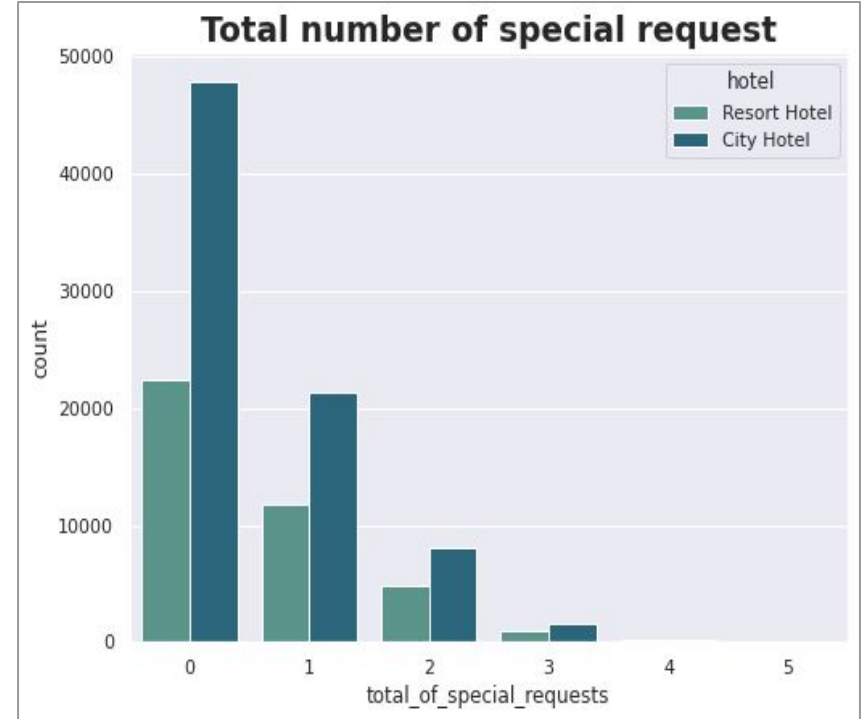
➔ Terminology

- Undefined/SC - no meal package
- Bed & Breakfast
- HB - Half Board(breakfast & one other meal type usually dinner)
- FB - Full Board(breakfast, lunch and dinner)



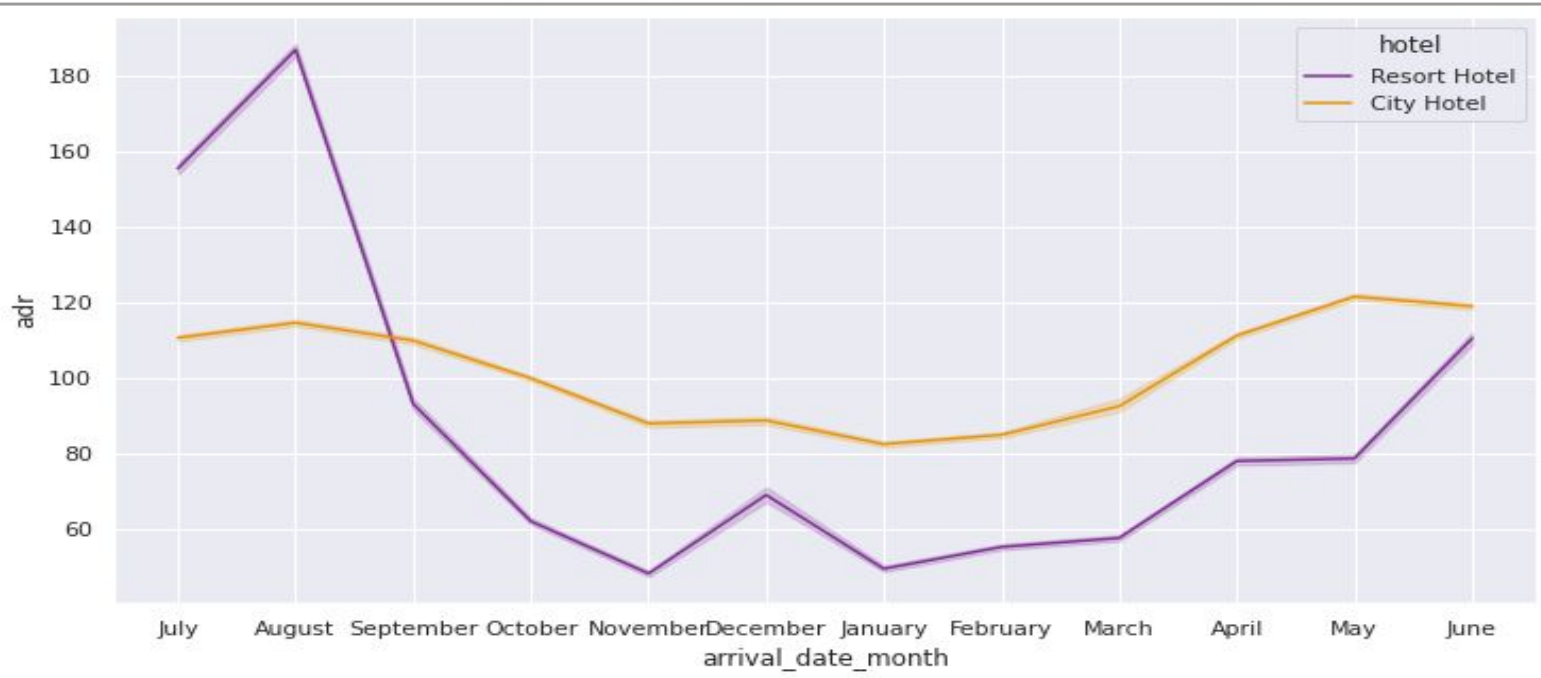
➤ SPECIAL REQUEST

- Special request are not very much high in both the city and resort hotel but between the two city hotels received more number of special request.
- If special request are met then people are more likely to book the same hotel type for the next time.



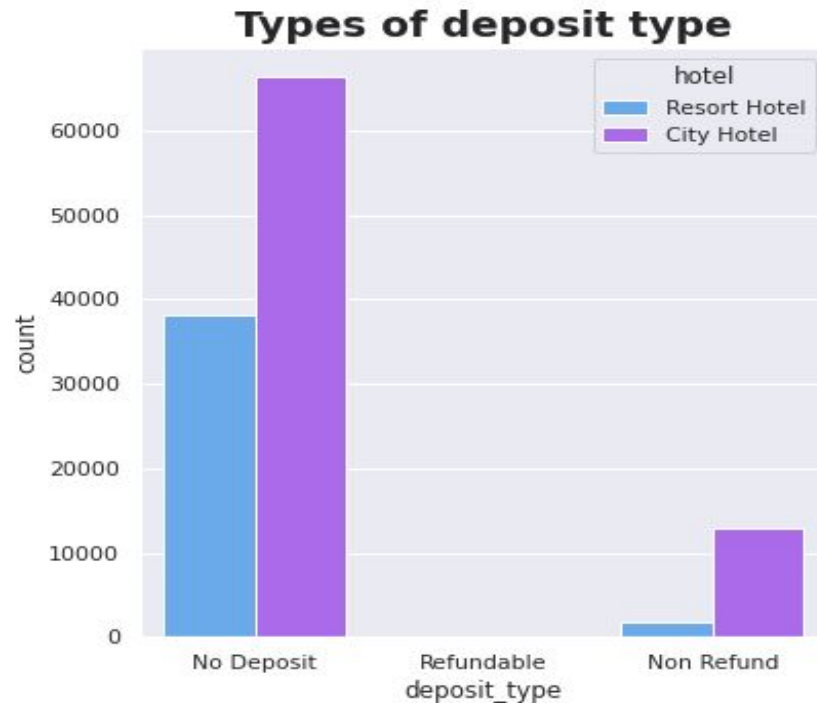
➤ AVERAGE DAILY RATE

- The rates are lowest during the winter time in both resort and city hotels.
- Since the booking are also lowest during winter the prices are also cheaper.
- Prices starts increasing during the summer time.
- Price are highest in the month of July and August.



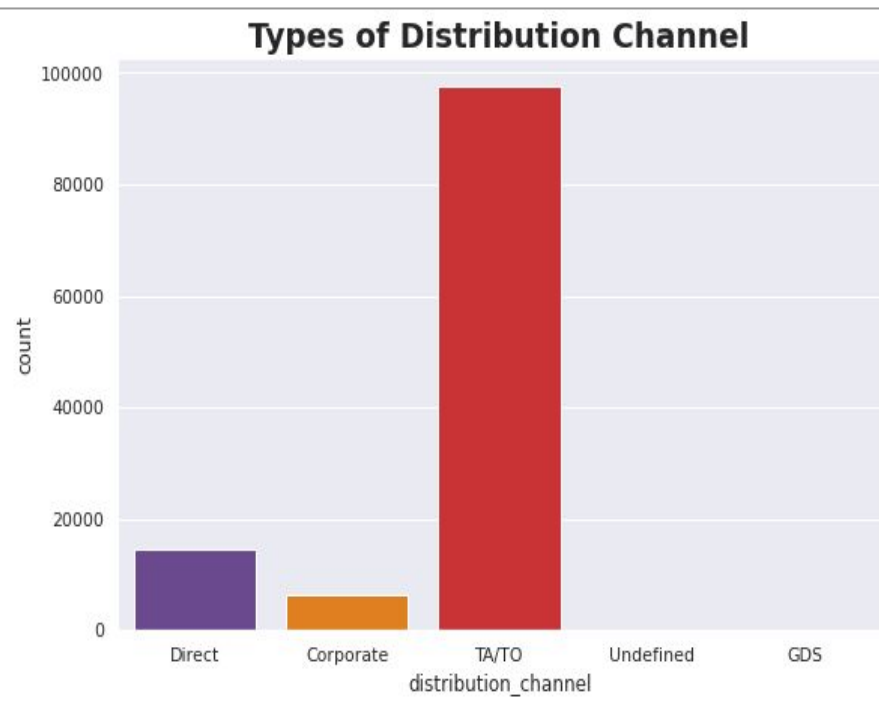
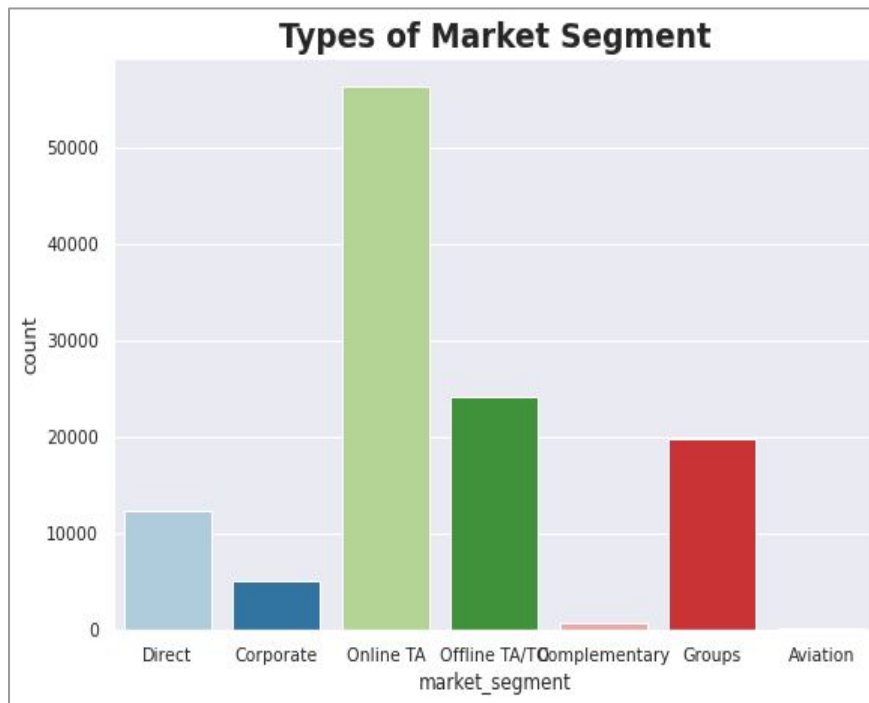
➤ DEPOSIT TYPE

- Both resort and city hotel doesn't require any deposit for booking.
- This could be one of the reason that leads to high cancellation.
- Cancellation of non refundable deposit type is higher than refundable which is quite surprising.



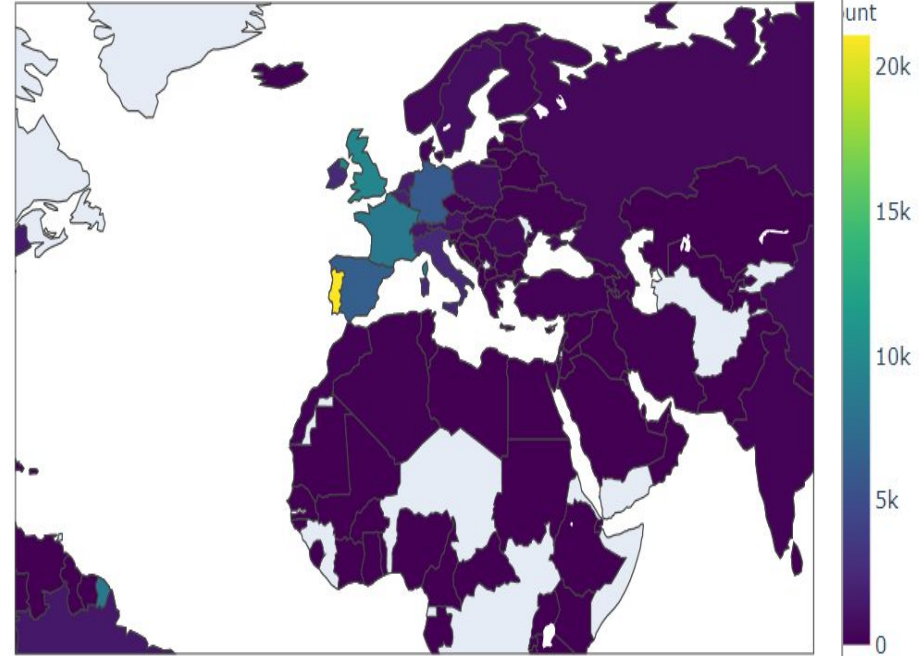
➤ MARKET SEGMENT AND DISTRIBUTION CHANNEL

- Majority of the market segment and distribution channel involves travel agencies(TA) online or offline.
- Bookings through online and offline travel agent is higher as compared to direct booking.



➤ VISITORS FROM DIFFERENT COUNTRIES

- Maximum numbers of visitors are from western Europe.
- Portugal, UK, France, Spain and Germany are the top countries where most of the guest comes.
- The highest number of customers comes from Portugal.



➤ CONCLUSION

- City hotels has higher bookings than resort hotel. 66% of bookings belongs to city hotel.
- City hotels have higher cancellation rate than resort hotel.
- Bookings are highest during summer and starts decreasing during winter.
- Average daily rates are lowest during winter and are highest during summer time.
- Number of stay on weekday night is higher than weekend nights.
- No prerequisite of deposit types leads to high cancellation rate.
- Number of repeated guest are very less.
- Bed and Breakfast(BB) is the most famous meal type.
- Number of special request are high in city hotel than resort hotel.
- Booking through online or offline travel agent is higher than any other type.
- Most of the guest are from western European countries like Portugal, UK, France and Spain.

THANK YOU!
