# Capstone Project - 1

## Hotel Booking Analysis


**Team Members :**

1) Bhavesh Bhagwan Patil
2) Paresh Suresh Badgujar
3) Avinash Vasant Patil

# Index of Content

# Problem Statement

- For this project we will be analyzing Hotel Booking data.This dataset 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.
- Hotel industry is a very volatile industry and the booking depends on above factors and many more.
- The main objective behind this project is to explore and analyze 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

# Work Flow


**Exploratory Data Analysis (EDA)** is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

**Specific statistical functions and techniques you can perform with EDA tools include:**

- Univariate visualization of each field in the raw dataset, with summary statistics.
- Bivariate visualizations and summary statistics that allow you to assess the relationship between each variable in the dataset and the target variable you're looking at.
- Multivariate visualizations, for mapping and understanding interactions between different fields in the data.
- Clustering and dimension reduction techniques, which help create graphical displays of high-dimensional data containing many variables.
- Predictive models, such as linear regression, use statistics and data to predict outcomes.

# Data Collection and Understanding

After collecting data its very important to understand your data.So we had Hotel Booking Analysis data which had 119390 rows and 32 columns.So let's understand this 32 columns first.

**hotel :** Resort Hotel or City Hotel
**is_canceled** : Value indicating if the booking was canceled(1) or not(0)
**lead_time** : Number of days that elapsed between the entering date of the booking 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
**arrival_date_day_of_month** : Day of arrival date
**stays_in_weekend_nights** : Number of weekend nights
**stays_in_week_nights** : Number of week nights
**adults** : Number of adults
**children** : Number of children
**babies** : Number of babies
**meal** : Type of meal booked.
**country** : country of origin

**market_segment** : Market segment designation.(TA/TO)
**distribution_channel** : Booking distribution channel (TA/TO)
**is_repeated_guest** : is a repeated guest (1) or not (0)
**previous_cancellations** : Number of previous bookings that were cancelled by the customer
**previous_bookings_not_cancelled** : Number of previous bookings not cancelled by the customer 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 the changes made to the booking
**deposit_type** : No Deposit,Non Refund ,Refundable
**agent** : ID of the travel agency that made the booking
**company** : ID of the company/entity that made the booking
**days_in_waiting_list** : Number of days the booking was in the waiting list before it was confirmed
**customer_type** : type of customer (Contract/Group/Transient/Party)
**adr** : Average Daily Rate as defined by the 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 last status.

# Data Cleaning and Manipulation

#Checking for Null Values

```
df.isnull().sum().sort_values(ascending = False)[:4]
```

```
company     112593
agent        16340
country        488
children         4
dtype: int64
```

There were 4 columns **company**,**agent**,**country** and **children** with missing values

**'Company'** column has so many missing values so we drop the column from the dataset

Now we are replacing those missing values

```
df['agent'].fillna(value = 9.0, inplace = True)
df['country'].fillna('others', inplace = True)
df['children'].fillna(df['children'].mean(),inplace = True)
```

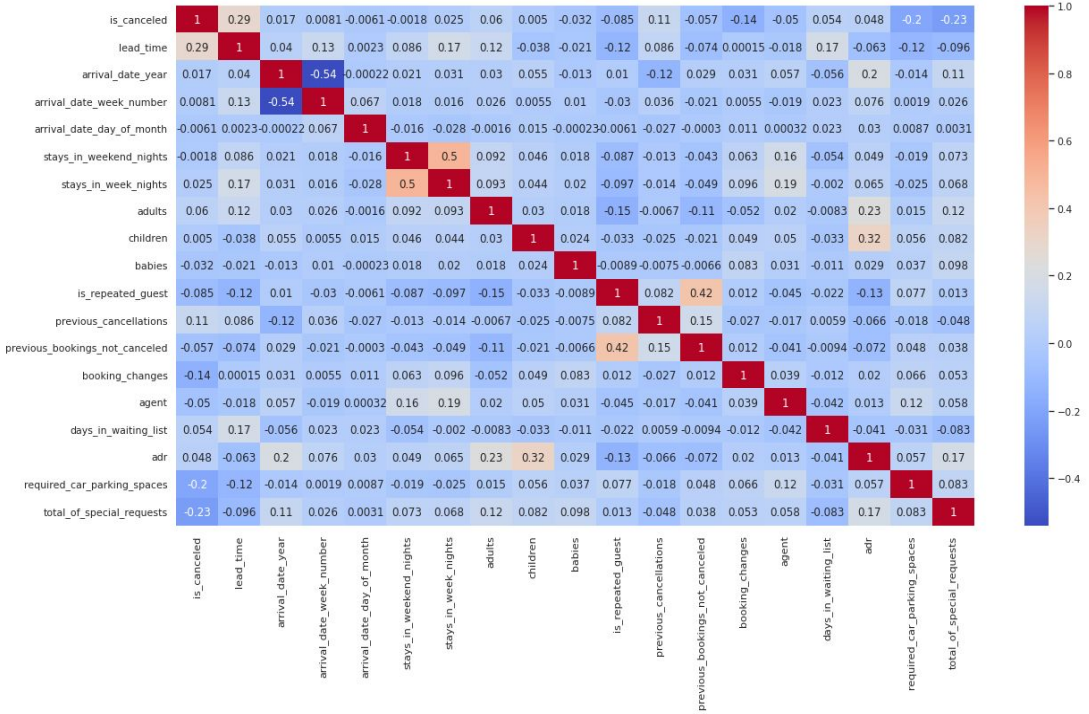
# Done with missing values

```
df.isnull().sum().sort_values(ascending = False)[:4]
```

```
hotel                        0
is_repeated_guest            0
reservation_status           0
total_of_special_requests    0
dtype: int64
```

Now , we have completely clean our data

# Exploratory Data Analysis

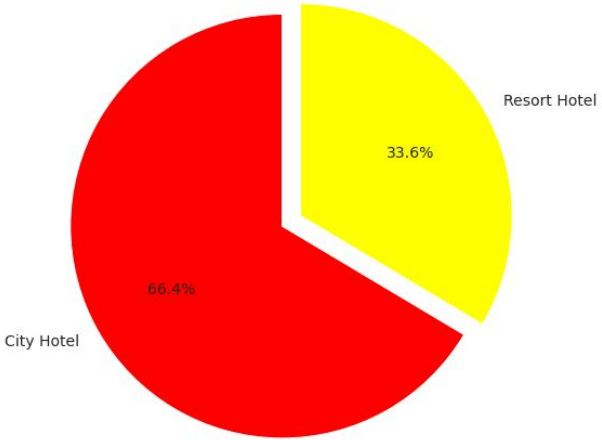



**Conclusion** :

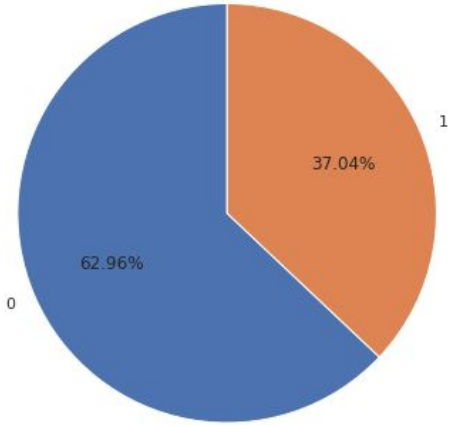
- is_canceled and total_of_special_request are negatively correlated. i.e there is no reason of booking cancellation when customer made the special request
- stays_in_weekend_nights and stays_in_week_nights are highly positively correlated with each other, it means that increase in weekend nights also increase in week nights
- Also the required_car_parking_space and is_canceled are negatively correlated with each other as there is less amount of car parking spaces which is required by the customer so high possibility that customer canceled the booking

# Overview of type of the Hotel & Percentage of Cancelled Bookings

- Pie chart of Hotel preferred by the customers



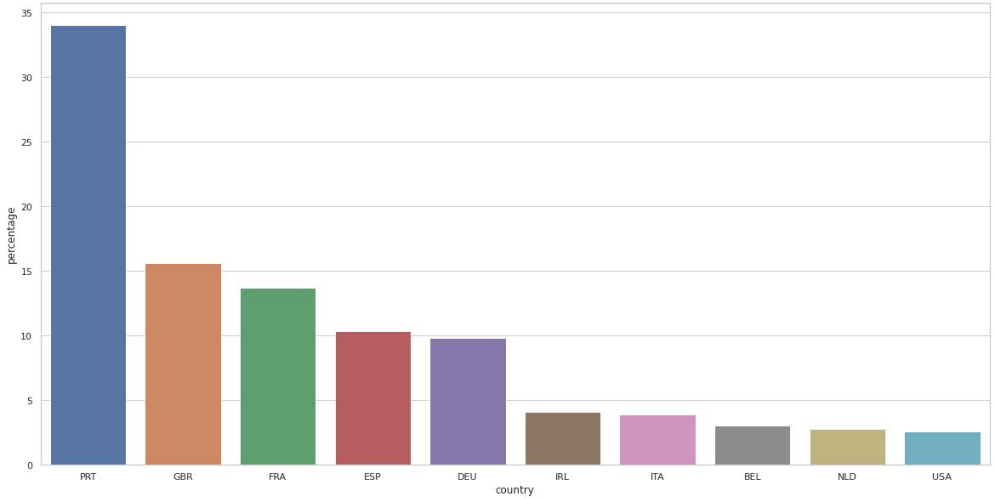

- Pie chart showing % of Booking canceled




**Conclusion** :

- From the 1st Pie chart it shows that 66% of customers choose City Hotel in comparison with the Resort Hotel.because of their facilities provided by the city hotels
- From the 2nd Pie chart it shows that around one-fourth of the total booking get canceled by the customer but 62% of the customer not canceled the bookings

# From which maximum number of guests are coming?

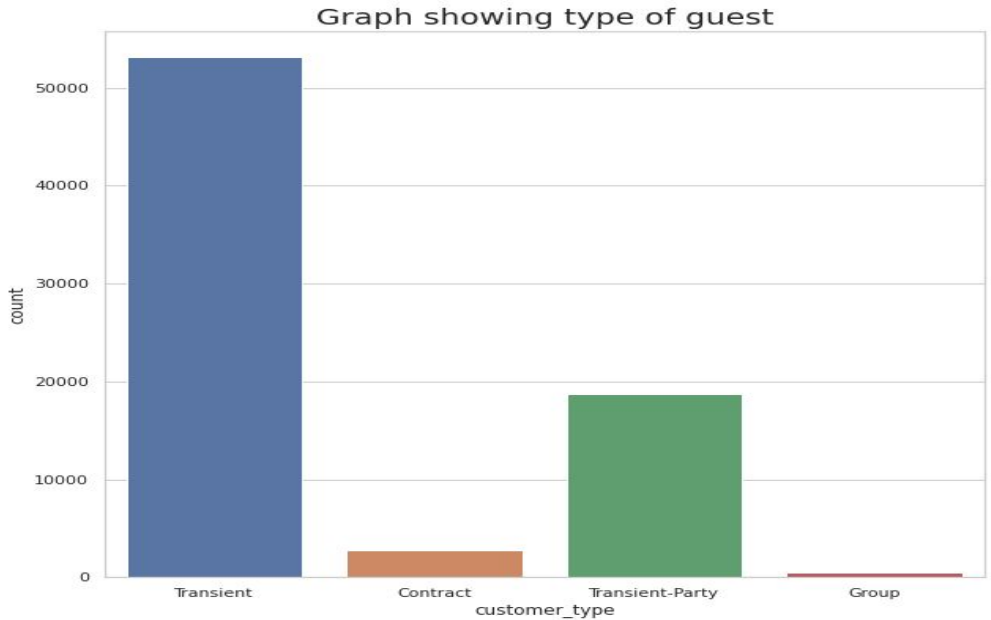




## Conclusion :

- Here are the Bar plots showing top-10 countries - it shows that more visitors are from Western Europe and minimum visitors are from United States.
- We have drawn the world map that shows the highest intensity of customers visited in the hotel The yellow region indicates about 20k customers from the PRT region i.e Western Europe and Portugal. Are visited to the hotels

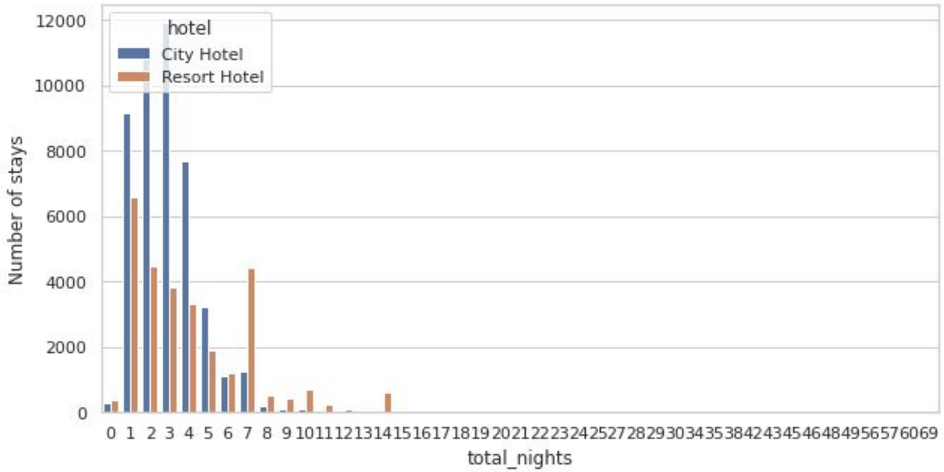
# Find Majority of booking done by which type of customers?

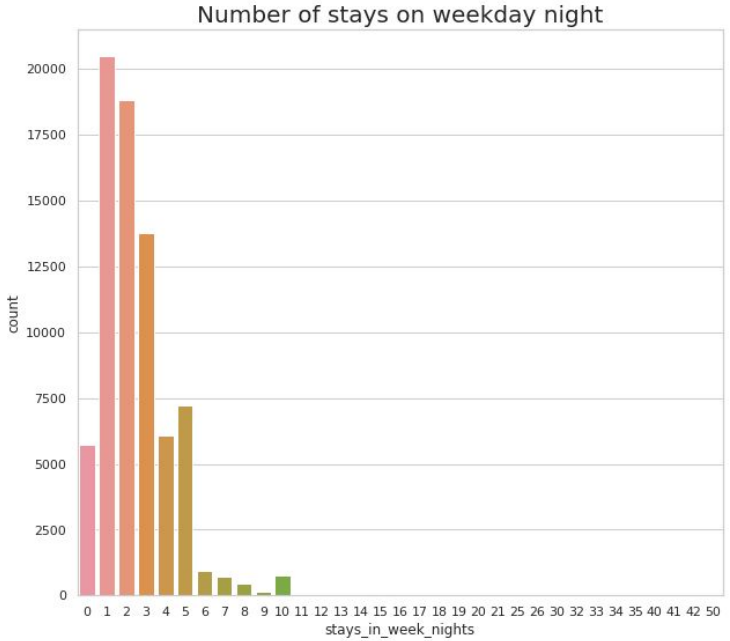



**Conclusion :**

- From the graph on the right side , it shows most customer who visited in the hotels are of Transient type.it means they are staying in the hotel for a short period of time.

# How long do people stay at the hotels?

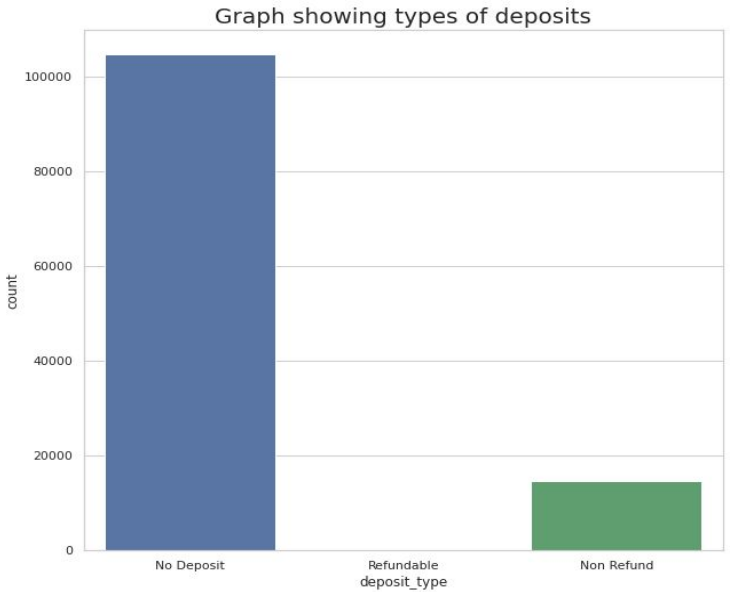


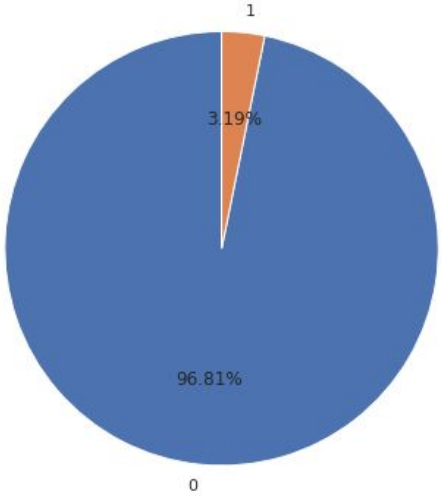




## Conclusion :

- From the above graph , it gives the information about customer stays in the hotel.for how long they will stays in the hotel. From the 1st graph it shows most of the peoples prefer to stay in the hotel are less than 5 days especially in the city hotels
- From the 2nd graph , it also shows that less than 5 days most of the guests are likely to stay in the hotel for the weekend nights.

# Looking into Deposit Type



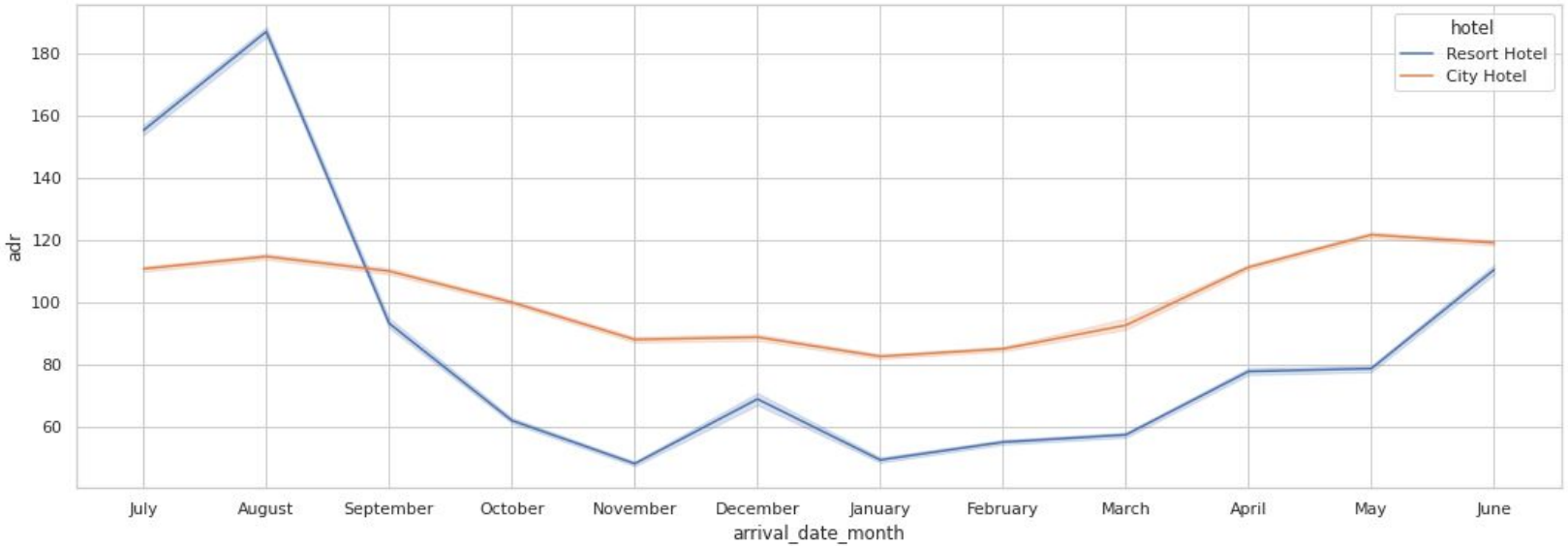




- Pie chart of No.of repeated guest

**Conclusion :**

- ➢ Majority of the booking does not require deposit.that could explain why explanation rate was actually 50% of Non-cancellation rate.thats why people are so comfortable in those hotels.
- ➢ From the Pie chart , high percentage of number of customers shows the guests are not repeated again again.

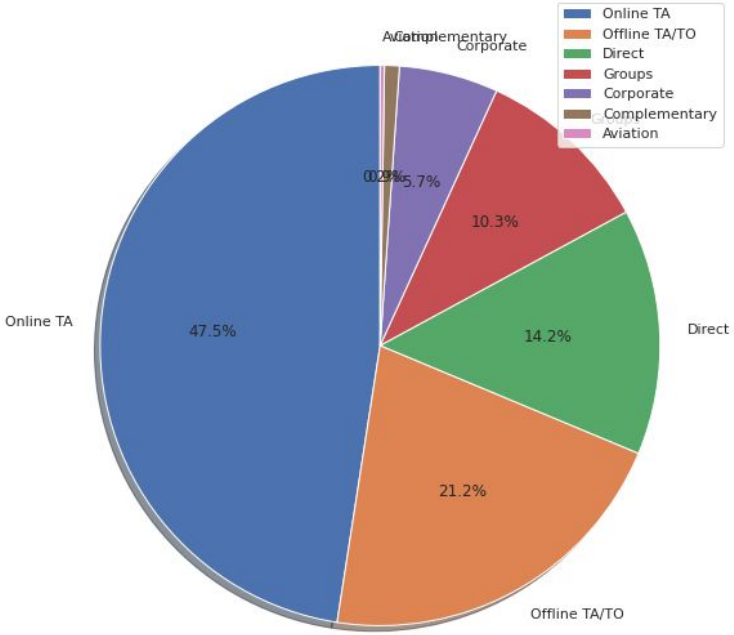
# Which month have cheaper booking rates?




**Conclusion :**

- From the month of November to January , there are average cheaper daily rates for the both types of hotel - Resort Hotel and City Hotel.

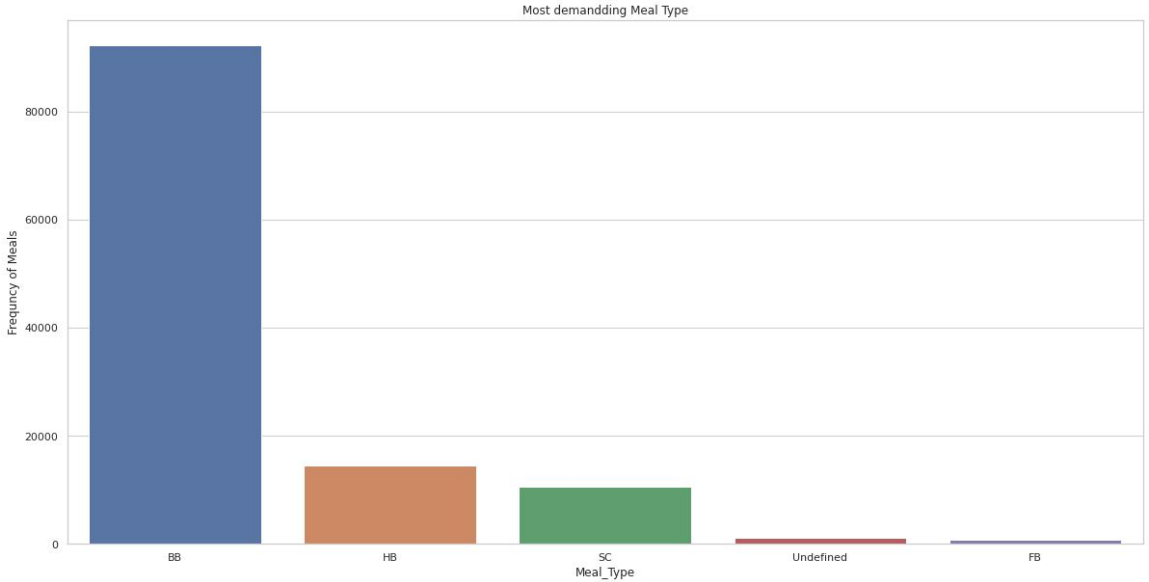
# Number of bookings made by different market segments.?




**Conclusion :**

- Here we shows that the maximum number of bookings were made by online TA(Travel Agencies) which was also followed by offline TA/TO(Travel Operators) and Direct

# Which is the most demanding Meal Type?




**Conclusion :**

- From the above bar chart it shows that , guest coming to the hotel are prefered BB means bread and breakfast meal type in the breakfast.