

# Hotel Booking Analysis

## Capstone Project

By :

Shubham Papadkar

Shruti Jain

Renuka Yadav

Anuja shende

Suraj Singh

# Point of Discussion

1. Introduction
2. Problem statement
3. Data Collection and Data understanding
4. Data Summary
5. Data preprocessing and exploration
6. Data cleaning / null value implementation
7. Data Transformation and visualization
8. Conclusion

# Introduction



- To discuss the analysis of given hotel bookings data set from 2015 to 2017.
- 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 babies, and the number of available parking spaces etc.
- Hotel industry is a very volatile industry and the bookings 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.

# Introduction

- We'll be doing analysis of given data set in following ways :
  - Hotel wise analysis
  - Distribution Channel wise analysis
  - Booking cancellation analysis
  - Timewise analysis
- By doing this we'll try to find out key factors driving the hotel bookings trends.

# WHY ANALYZE THE HOTEL BOOKING DATA?



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. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business. We can use the patterns to predict the future bookings using time series or decision trees.

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

The overall goal of data analysis is to track patterns at properly and set yourself up to make accurate predictions.

In turn you can plan and strategy towards your ultimate desire: increased revenue.

# Problem statement

- We will be tackling this problem statement in three stages:
  - ❑ We will be analyzing some key metrics for hotel bookings like:
    - The number of cancellations
    - Number of bookings on weekday vs weekends
    - Most preferred meal types
    - Country wise bookings
    - Type of rooms preferred by customers
    - Booking types,
    - Hotels available for booking
  - ❑ We will be using various lenses to look through the data to analyze patterns associated with each segment such as:
    - The type of hotel
    - Day of week
    - Type of customers
    - Type of rooms

# Data Collection And Understanding

- After collecting data it is 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.

## ❖ Data Description:

- **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.

# Data Collection And Understanding



- **country** : Country of customer.
- **market\_segment** : Market segment designation. (TA/T0) .
- **distribution\_channel** : Booking distribution channel.(T/A/T0).
- **is\_repeated\_guest** : is 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 room type reserved.
- **assigned\_room\_type** : Code for the type of room assigned to the booking.
- **booking\_changes** : Number of changes made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation.
- **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 to the customer.



# Data Collection And Understanding

- **children** : Number of children.
- **babies** : Number of babies.
- **meal** : Type of meal booked.
- **customer\_type** : type of customer. Contract, Group, transient, Transient party.
- **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.

# Data Preprocessing and Exploration



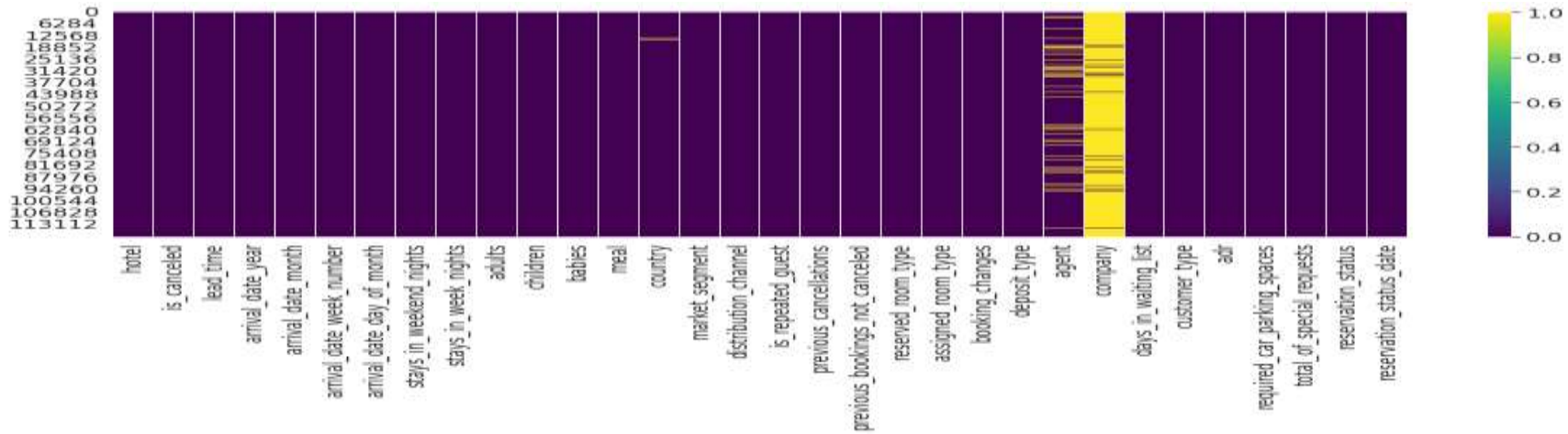
- ❖ **Data preprocessing is a process of preparing the raw data and making it suitable for a analysis. It increases the accuracy and efficiency of data.**
- ❖ **We are doing Data Preprocessing by following way:**

- **Step 1 : Importing Libraries and Dataset**

To perform data preprocessing using Python, we need to import some predefined Python libraries. Such as Numpy , Pandas , matplotlib, seaborn and etc. These libraries are used to perform some specific jobs. we need to import the datasets which we have collected for our project.

- **Step2 : Handling Null value**

- Finding Missing value in dataset.
- Finding Duplicated value in dataset.
- Removing Highly observations with Null values.
- Correct data type(INT,FLOAT,DATE).
- Extracting unique values.



- The Dataset contains 119390 rows and 32 features. There are missing values in some columns.
- Their count and percent in each are as follows:
  - company – 112593 (94.31%)
  - agent – 16340 (13.69%)
  - country – 488 (0.41%)
  - children – 4 (Negligible)

# Data Preprocessing and Exploration

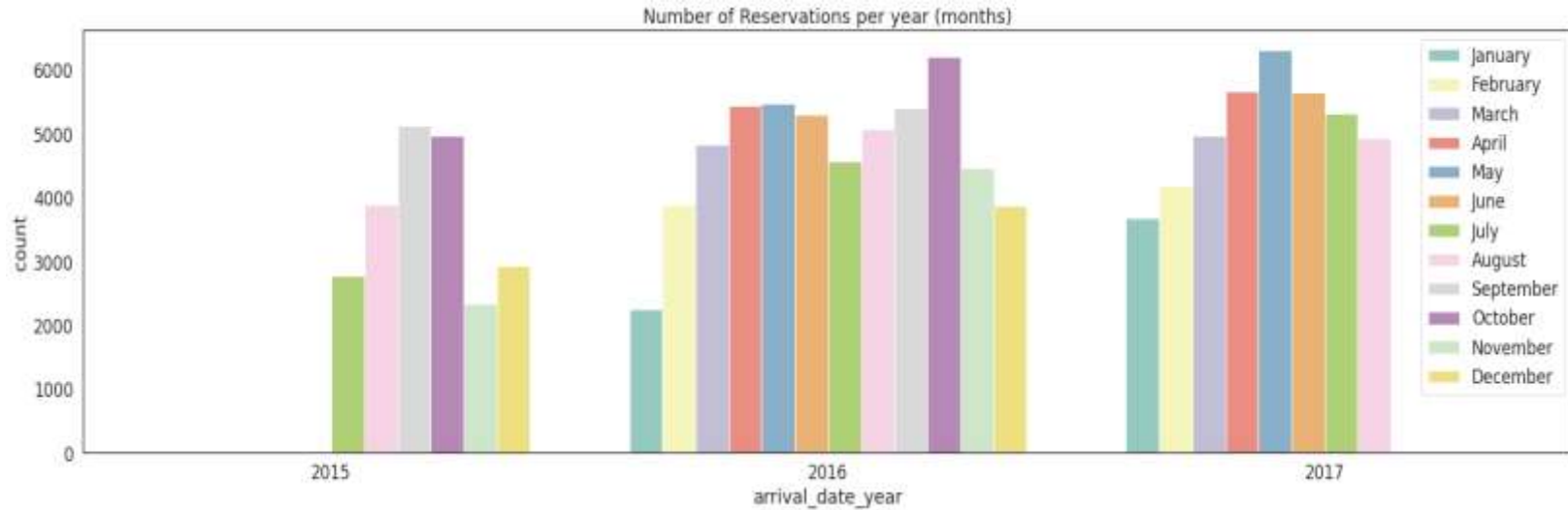
- **Step3 : Data Transformation** : This step is taken in order to transform the data in appropriate forms suitable.
- **Step4 : Analyze the data** : analysis of data is the practice of working with data to glean useful information, which can then be used to make informed decisions.
- **Step5 : Visualize the data** : the graphical representation of information and data. By using visual elements like charts, graphs data visualization tools provide an accessible way to see and understand trends, outliers, and patterns.

# Data Cleaning

## ❖ Understand the structure of the dataset and clean data before analysis

- The Dataset contains 119390 rows and 32 features. There are missing values in some columns.
- Their count and percent in each are as follows:
  - company – 112593 (94.31%)
  - agent – 16340 (13.69%)
  - country – 488 (0.41%)
  - children – 4 (Negligible)
- We replace the company null values with '0'. column because 94.3% of data was NA.
- If no. of children and agent is null we replaced it with mode of children and '0.0' respectively.
- For the NA values in the country column, we replaced it with 'others'.
- We find duplicate data and remove it for more accuracy.
- We also Extract unique values from the dataset.

# Booking Analysis



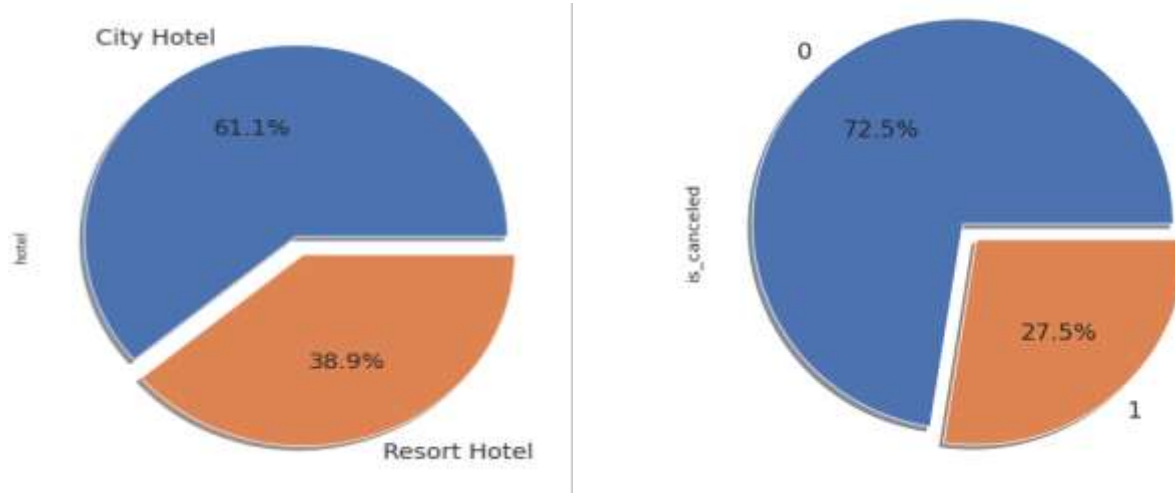
## ❖ Conclusion:

- Here we see that Reservation status is most in year 2016 and also most in canceled .
- August, September and October month has most of the booking during all three year.
- Here we are analyzed the booking in year 2015 starts from July to December.
- The booking in 2016 is throughout the year i.e. January to December.
- The booking in 2017 is from January to August only.
- so we have booking data of  $6+12+8 = 26$  months.

# Data Transformation and visualization

## Hotel wise Analysis

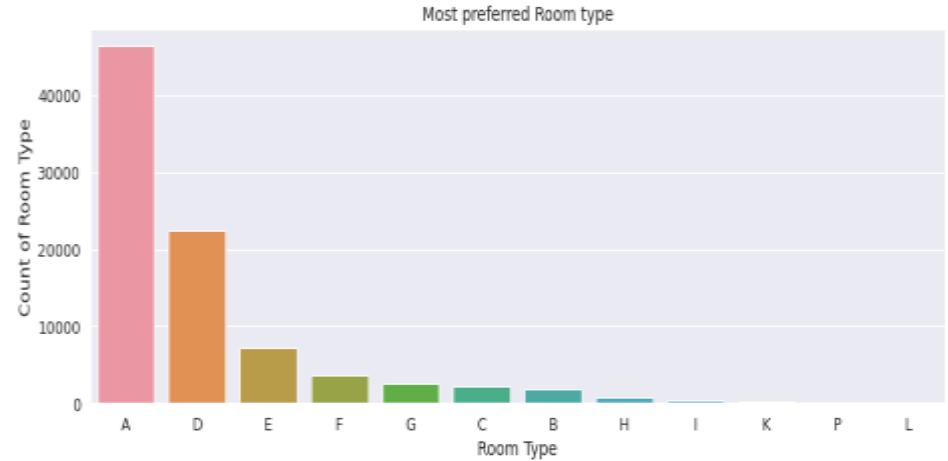
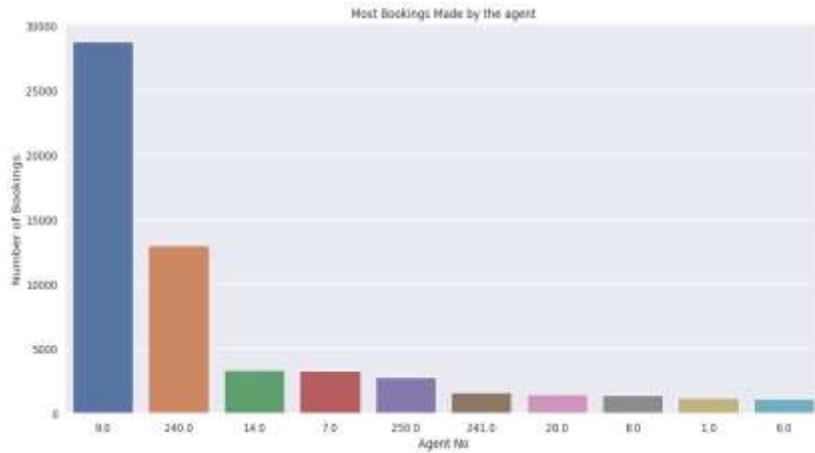
### ❖ Observation:



### ❖ Conclusion:

- City hotels is the most preferred hotel type by the guests .We can say City hotel is the busiest hotel.
- Nearly 27.5 % bookings were got cancelled out of all the bookings.

# Booking Analysis

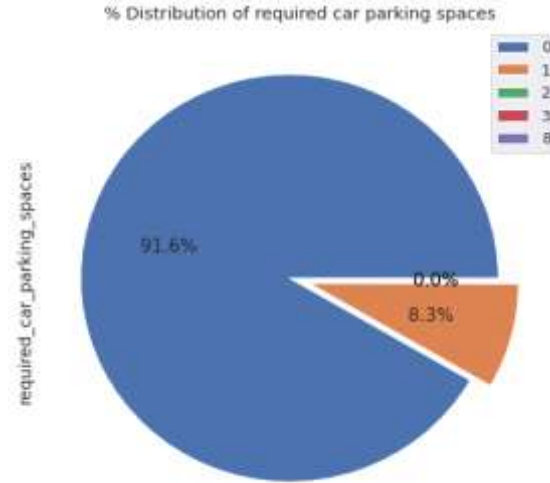
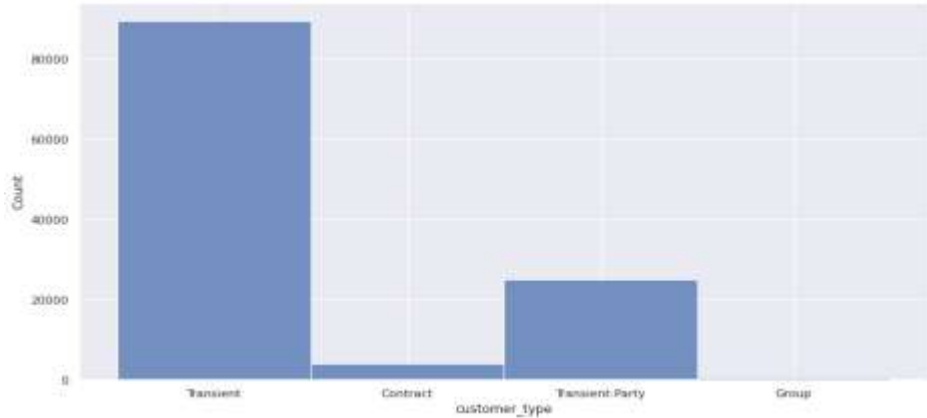


## ❖ Conclusions:

- Agent Id no -9 made the highest bookings i.e 28759.
- Room type 'A' is most preferred by the guests and second most preferred is 'D'.



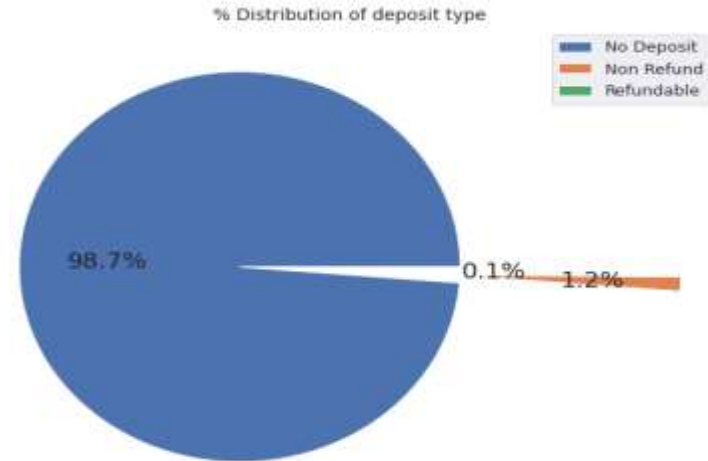
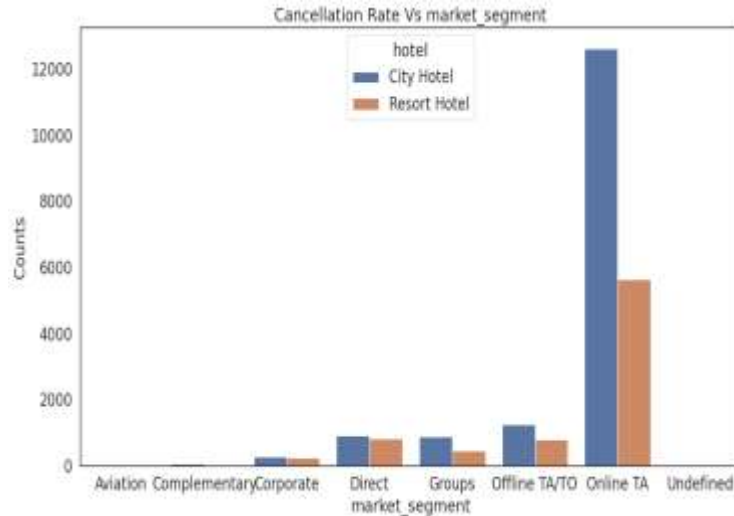
# Parking And Customer Type Analysis



## ❖ Conclusion:

- Most of the customers/guests were Transient type(82.4%). And transient party were 13.4% and 0.6 belongs to group. Remaining guests belongs to Contract type.
- Most of the customers(91.6%) do not require car parking spaces. Only 8.3 % people required only 1 car parking space.

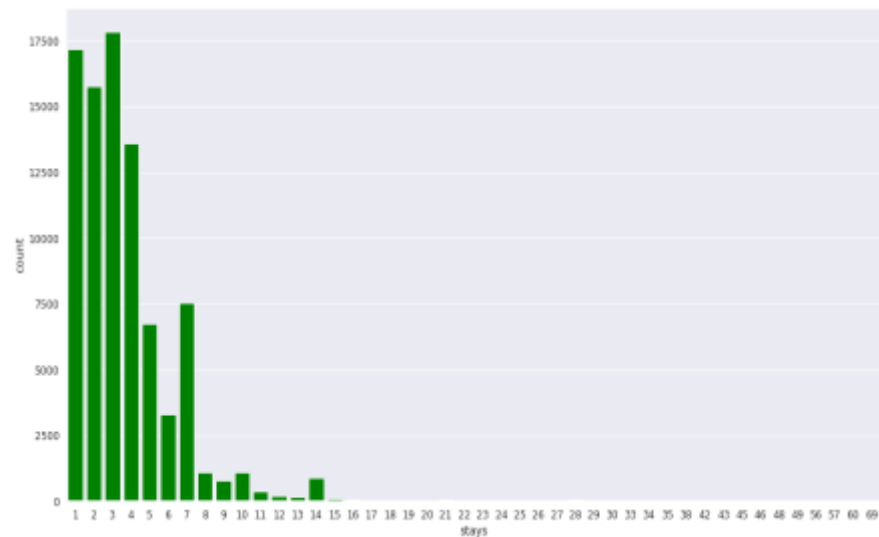
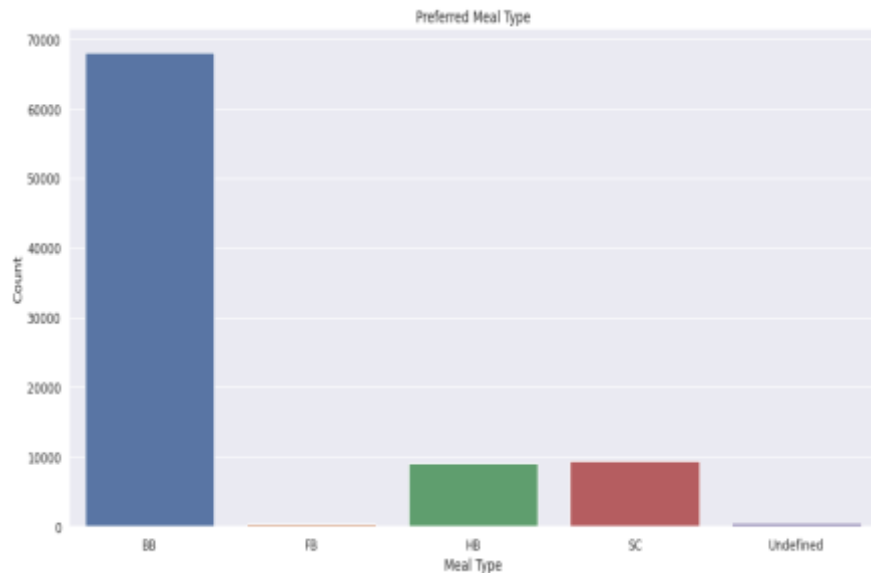
# Cancellation Rate Analysis



## Conclusion:

- 79.1 % bookings were made through TA/TO (Travel Agents/Tour Operators). Second most channel is direct booking.
- Almost 98.7% of the guests prefer 'No deposit' type of criterion while booking hotels.

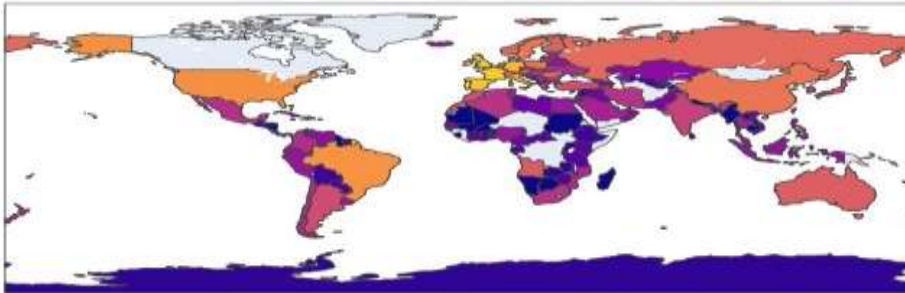
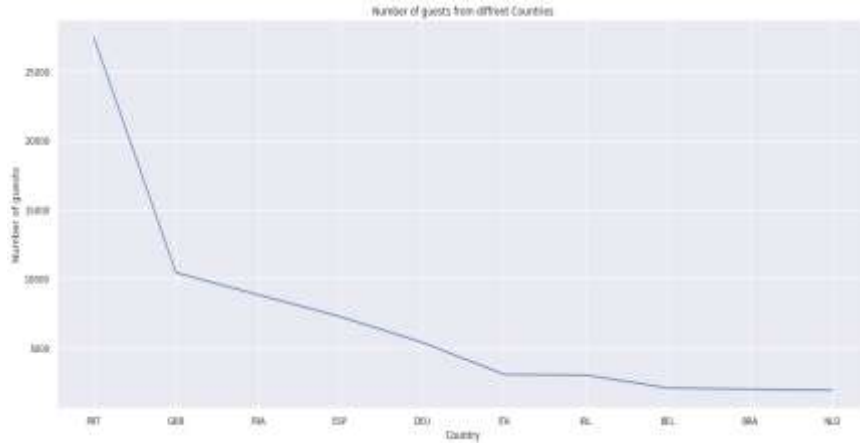
# Meals Pattern And Stay wise Analysis



## Conclusions:

- BB( Bed & Breakfast) is the most preferred type of meal by the guests.
- Full Board i.e. FB is least preferred. HB (Half Board) and SC(Self Catering) are equally preferred.
- Optimal stay in both the type hotel is less than 7 days. Usually people stays for a week.
- For stay more than 7 days people likes to stay in Resort hotels

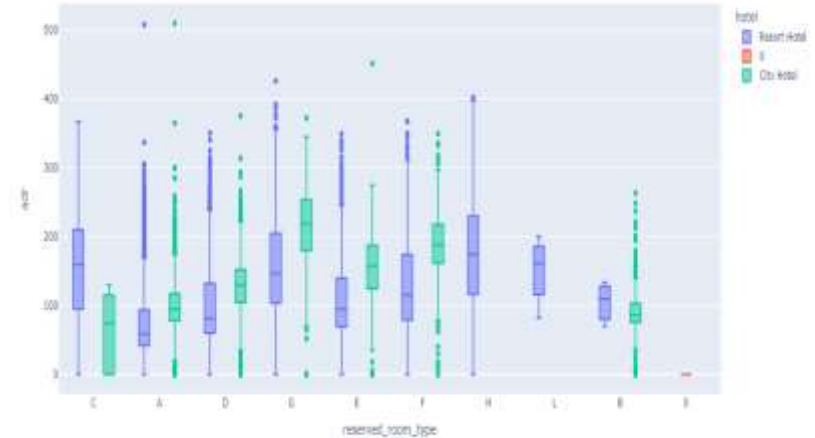
# Timewise Analysis



## ❖ Conclusions:

- Resort hotels had the highest ADR in June, July and August than the City hotels. But in other months ADR of Resort hotel was less than the City hotels.
- Thus we can say that, the January, February, March, April, November and December are the good months for customers to get good ADR.
- Maximum number of guests were from Portugal. i.e. more than 25000 guests.
- After Portugal, GBR(Great Brittan),France and Spain are the countries from where most of the guests came.

# Average Daily Rate Analysis



## ❖ Conclusion:

- For Resort hotel is ADR is high in the months June, July, August as compared to City Hotels. May be Customers/People wants to spend their Summer vacation in Resorts Hotels.
- The best time for guests to visit Resort or City hotels is January, February, March, April, October, November and December, as the average daily rate in this month is very low.

# Conclusion

- Most of the booking made in Hotel - City Hotel.
- Max\_ preferred meal - Bread and Breakfast.
- Best performer Agent - Agent No. 9.
- Most preferred and assigned room type - A
- One out of every three bookings are cancelled.
- People prefer to tour more in August.
- Average days in advance booking : 77 days and prefer 'No deposit' type of criterion while booking hotels.
- Most preferred meal is BB(Bread and Breakfast).
- Online marketing is the best way to attract customers.
- People do not want to pre-deposit the money for booking.
- Only 10% of people require parking space . In these 8.3% customer only required parking space for 1 car.
- Resort hotel is preferred mostly for longer stay, day time stays and when the parking space is needed.
- Best time to book a hotel is in January. Because the adr is less in JAN

Thank you!