Hotel Booking Analysis

Team Members

Affan Ahmad Abdul Wahid

Rithun Rajendran M.K

Prajwal Bharadwaj

Shreyash Kumar

Md Ashique Ali

**Introduction:**

Hotel industry is a very volatile industry and the bookings depend on a 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.

1. **Problem Statement**

* For this project we will be analyzing hotel booking data.
* This dataset contains booking information for resort and city hotels, and contains booking information muchas total stay, when the booking was made, the agent through booking was made, number of children, adults, babies and so on.
* The main objective behind the project is to explore and analyze data to discover important factors that govern booking and give insights to hotel management, which can provide various campaigns to boost the performance of the management

**2.Challenges faced**

1) There was a lot of duplicate data.

2) Choosing appropriate visualization techniques to use was difficult.

3) A lot of null values were there in the dataset.

**3.Modules used**

**. pandas**

**. numpy**

**. matplotlib**

**. seaborn**

**. barplot**

**. pie chart**

**4.Approaches Used**

**1. Data cleaning and preparations**

### 1) Removing Duplicate rows

All duplicate rows were dropped.

### (2) Handling null values

* Null values in columns  company and agent were replaced by 0.
* Null values in column country were replaced by 'others'.
* Null values in column children were replaced by the mean of the column.

### (3) Converting columns to appropriate data types

* Changed data type of children, company, agent to int type.
* Changed data type of reservation\_status\_date to date type.

# **2.EDA**(**Exploratory Data Analysis**

Exploratory data analysis (EDA) is a term for certain kinds of initial analysis and findings done with data sets.

Performed EDA and tried answering the following questions:

* What will be the overview of the arrival period?
* From which country most guests come?
* Which room type is most booked?
* Find the Distribution channel with a highcancelation\_percentage(%)?
* Which meal type is the most preferred meal of customers?
* Obtain the number of Adults reserving , and the number of Families with children reserving
* What is the percentage of booking in each hotel?
* Which agent makes the most no. of booking ?
* What is the percentage of the repeated customers?
* What is Preferred stay length in each hotel ?
* Which hotel has a high chance that its customer will return for another stay?

**5.Conclusion**

* The arrival of peoples in 2016 is double compared to 2015 but has gradually decreased in 2017.
* Portugal, United Kingdom, France , Spain and Germany are the top countries from which most guests arrived.
* ‘A’ type room is preferredby the majority of people.
* TA/TO has the highest cancellation percentage.
* Most common type meal isBB and least common is undefined
* From above two graph, in the first graph the adult with pair has most reservations and in the second graph adults with zero children is having most booking, so we can conclude that the adults without children has more booking
* The percentage of booking of a city hotel is 60% and the resort hotel is 36%.
* Agent ‘9.0’ has the most number of bookings above 30000 bookings and agent ‘8.0’ has the least number of bookings.
* 3.86% is the booking percentage of repeated guests and the percentage of new guests is 96.14%.
* Most common stay length is less than 4 days and generally people prefer city hotels for short stays, but for long stays resort hotels are preferred.
* Both hotels have a very small percentage that customer will repeat, but resort hotel has slightly higher repeat percentage than city hotel.