**HOTEL BOOKING ANALYSIS**

**Alpashi Sadawarti**

**Shubham Masurkar**

**Data science trainees,**

**Almabetter, Bangalore**

**Abstract:**

When people search for a hotel, various factors such as price, location, availability, parking space, food, accommodation options, etc are taken into consideration. Prices of the Hotels vary according to the month of booking, the number of guests, days of stay, hotel locations, hotel ratings, etc.

The objective of this project is to deliver insights into the data set and draw meaningful conclusions so that it is helpful for further development of the hotel industry. This hotel booking dataset can help you explore those questions!

***Keywords: Exploratory data analysis, Hotel booking analysis***

**Problem Statement**

* 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.
* Explore and analyse the data to discover important factors that govern the bookings.

**Steps involved:**

* **Importing the dataset**

A directorial path for the Hotel Booking dataset was created using the Pandas read function for importing the data set.

## **Cleaning and Transforming**

Cleaning is the process of removing undesired features, values, or any suffix, prefix, or anything which can produce an exception whereas transforming is required to ensure the consistent data type of features. These two processes have specific sub processes as follows.

* Unwanted Data Removal-In this step we ensured to make a data type of feature consistent by removing characteristics from the values of features, to make them usable.
* Null values Treatment - The company and agent column has many Null values so it’s not feasible to fill that many null values so we drop this column.
* **Data analysis and Data visualisation**

### Here we do data analysis and data visualisation by using matplotlib and seaborn to the graph for better understanding of the insights.

**Exploratory Data analysis**

Following observations are concluded using Exploratory Data Analysis and visualisation:

**1. What types of hotels are in the market and how many numbers of hotels in each type?**

There are two types of hotels.

* City Hotels
* Resort Hotels

Out of 119390 Hotels, City hotels are 52108 whereas Resort bookings are 33265 in number.

**2. What is the ratio of reservation for each type of hotel?**

Out of 119390 Hotel bookings, City hotels have 61% bookings whereas Resort bookings make 39%.

**3. How Many bookings were cancelled for each type of hotel?**

Cancellation rates are high for city hotels as compared to resort hotels.

**4. Year wise comparison of booking of hotels.**

Year 2016 has the most numbers of guests for both types of hotels.

**5. Month wise comparison of booking of hotels.**

August is the busiest month for both the types of hotels.

**6. From which Countries highest numbers of guests are coming?**

We have fetched the data for the top 10 countries. Portugal (PRT) is at the top of the list, followed by Great Britain (GBR) and France (FRA).

**7. What are the number of Adults vs. Families with children preferring the reservations?**

Analysing the data based on reservations, we can see that reservation count is highest amongst the couple. While families with children are far less than the couples. Comparing families with children, no of families with one child is more than families with 2 children.

**8. What is the percentage of repeated guests?**

If we look at the pie chart for no of repeated guests, only 3% of the total guests are repeated.

**9. How long do people prefer to stay in the hotel?**

Most people prefer to stay for about 5 days in city hotels whereas in resort hotels people prefer to stay for 1-2 days.

**Conclusion:**

Starting with importing the data, followed by cleaning the dataset, we have done Exploratory Data Analysis. According to this analysis, following conclusions were drawn:

* City hotels have double the amount of reservation compared to resort hotels for the same time period.
* City hotels have high cancellation rates compared to Resort hotels.
* Reservations for both types of hotels peak in August and are lowest in the month of January.
* Bookings for city hotels are higher than resort hotels by considering every year’s data except the year 2015.
* In 2016 both the hotels had the highest number of guests.
* Both the hotels have good occupancy in the 2nd and 3rd quarter of the year with 'August' being the highest of all.
* Here we can conclude that the bookings in the month end are least. But that is not true because as every month doesn't have 31 days that's why we see the less number of arrivals in the month end.
* Majority of the reservation is by 2 adults followed by a few families with 1 or 2 children.
* Most people prefer to stay at hotels of <=5 days.
* Only 3% of the total guests are repeated. This shows that people do not prefer the same hotels for their stay.

**References-**

* Pandas user guide: <https://pandas.pydata.org/docs/user_guide/index.html>
* Matplotlib user guide: <https://matplotlib.org/3.3.1/users/index.html>
* Seaborn user guide & tutorial: <https://seaborn.pydata.org/tutorial.html>