**Hotel Booking Analysis**

**Suraj Kumar**

**Data Science Trainee**

**AlmaBetter, Bangalore**

**Abstract:**

The data collection includes information on when the reservation was made, how many nights they stayed, the numbers of adults, kids or babies and the number of parking spaces, among other things, for both a city hotel and a resort hotel.

Exploratory Data Analysis (EDA) allows us to explore, visualize and understand the specific variables in our datasets that have the most impact on the outcome of the Hotel Booking analysis.

1. **Problem Statement:**

There are many online hotels booking websites that allow travelers to book hotel rooms online. However, there is a problem with an online hotel booking. The problem is that travelers often do not receive the hotel room that they booked. This is because the hotel room that the traveler booked may not be available when the traveler arrives at the hotel. This can be a problem for travelers because they may not be able to stay in the hotel room that they booked.

There are several possible solutions to this problem. One solution is for the traveler to contact the hotel directly to book the hotel room. This way, the traveler can be sure that the hotel room is available. Another solution is for the traveler to use a hotel booking website that guarantees that the hotel room will be available when the traveler arrives at the hotel.

In order to understand the customer's journey, we first looked at the data that the client had provided. This data included information on customer bookings, as well as customer feedback. We then used this data to create a customer journey map, which allowed us to see the steps that customers take when booking a resort or hotel. Finally, we analyzed this data to identify any potential areas that may need improvement.

1. **INTRODUCTION:**

The hotel booking data collection consists of binary, categorical, and numeric data. We were able to gain significant insights from the data set with the columns like the hotel type, is\_canceled, arrival\_date\_year, arrival\_date\_month, stays\_in\_weekend\_nights, nation, market\_segment, distribution\_channel, etc. Here, our goal is to comprehend the critical elements that influence hotel reservations.

I will be analyzing data from a resort or hotel booking dataset. The dataset includes information on the resort or hotel, the guests, and the booking itself. I will be looking at various aspects of the data to see what insights can be gleaned from it.

The main goal of this project is to conduct an exploratory information analysis and draw conclusions about all the important factors that influence hotel reservations.

In order to help hotels accept bookings and receive payments online, the Hotel Booking System is an online booking system that enables customers to make safe online reservations through hotel websites.

In order to assist hotels in automating daily hotel operations and boosting bookings, the Hotel Booking System is a complete hotel estimate booking system that includes the major roles of Hotel XML IN, Hotel XML Out, Hotel Channel Manager, Hotel Extranet, and Own Contracting.

## **Types of Hotels:**

## Resort Hotel

## City Hotels

## **How hotel booking works?**

When a customer wants to book a hotel, they usually go online to a hotel booking website and enter in their travel dates. The website will then search for hotels that have availability for those dates and display a list of options. The customer can then browse through the list of hotels and read reviews from other guests before making their decision. Once they have decided on a hotel, they can book their stay and pay for it online. The following distribution channels provide bookings:

* Direct
* Corporate
* TA/TO
* GDS

The hotel assigns a room based on supply and demand. A different room is assigned if the specified room is not available.

There is no non-refundable or refundable deposit given to the hotel in exchange for the reservations.

## **STEPS INVOLVED:**

## **Importing Libraries:**

## Our main goal in this step was to import all the libraries we would need to analyse the issue statement and do EDA to draw conclusions from the data.

## **Understand the data set:**

## The data set was checked after that. In order to solve the issue statement, we need to know how many rows and columns are available, as well as which columns could be critical.

* **Manipulating datasets:**

In this section, we worked on manipulating datasets refers to making changes to data in order to better understand it, to clean it up, or to prepare it for analysis. This might involve tasks such as sorting data, calculating new values, or reorganizing the data into a different format.

* **Null values Treatment:**

There are a few ways to deal with null values:

1. Drop them: This is probably the most common way to deal with null values. Simply drop any rows or columns that contain null values. This is easy to do but may not be the best solution, as you may be losing valuable data.

2. Replace them: Another common way to deal with null values is to replace them with another value, such as the mean or median of the column. This is easy to do but may not be the best solution, as you may be introducing bias.

* **Exploratory Data Analysis:**

After importing datasets, I explored, deleted several columns, and handled with different columns null values. After then, in order to better understand and make conclusions from each attribute, we compared them both separately and together.

* **Uni-variate Analysis:**

There is only one reliable variable in a univariate analysis because uni means one and variate means variable. Univariate analysis aims to derive the data, characterise and summarise it, and examine any patterns that may be there. It explores every variable in a dataset separately. It is possible for two kinds of variables- Categorical and Numerical.

* **Histogram**: A histogram is a graphical representation of the distribution of data. It is a type of bar chart that shows the frequency of occurrence of a certain event. The histogram can be used to show the distribution of data in a variety of ways, such as the distribution of ages in a population, the distribution of test scores, or the distribution of income levels.
* **Pie Chart:** Pie charts can be used to visualise population percentages of various factors, such as Type variable, to determine how many percentages of the overall population of applications are free or paid.
* **Count Plot**: A count plot is a graph that shows the number of occurrences of each unique value in a given column. This type of plot is useful for visualizing categorical data.
* **Distplot**: A distplot plots a univariate distribution of observations. The matplotlib hist function, the seaborn kdeplot(), and rugplot() functions are all combined in the distplot() function. Related subject: Examples from Matplotlib and a video tutorial. Example Distplot illustration An easy distribution is displayed in the plot below. With random, it creates random values. randn().
* **Bivariate Analysis**:

Bivariate analysis is the statistical method used to determine the relationship between two variables. The variables can be any two things that can be compared or related to each other, such as height and weight, age and income, or gender and voting behavior. There are three main types of bivariate analysis:

1. Correlation: A measure of the strength of the relationship between two variables.

2. Regression: A statistical method used to predict the value of one variable based on the value of another variable.

3. Chi-square: A statistical test used to determine whether there is a significant relationship between two variables.

* **Bar plot**: Here, we've analysed categorical data in relation to numerical data with the primary goal of obtaining categorical characteristics for specific numerical features. This graph helps us comprehend the relationship between two variables.
* **Joint plot:** Our primary reason for selecting this graph is because it more clearly contrasts one numerical property from others. We can quickly comprehend the variation and get more understanding by looking at it in isolation.
* **Heatmap**: A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors. Heatmaps are used in a variety of fields, including biology and finance, as they allow for easy visualization of complex data sets.

## **4. Conclusion:**

At the conclusion of our exercise, that is where we are at. We started by importing our datasets, engaged in some data manipulation, imputed null values, and then conducted exploratory data analysis in two stages: univariate analysis and bivariate analysis, where we made use of various graph types to gain a better understanding and come at a better conclusion.

## The data has been loaded, null values in category columns have been handled, and the main factors that affect hotel reservations have been discovered along with actions to improve them.

* Most reservations were made through internet travel agencies.
* In comparison to resort hotels, city hotels receive the most visitors. However, city hotels have a higher cancellation rate.
* Hotels in cities receive more regular visitors. The ratio of returning guests is higher for resort hotels, however, as city hotels receive more visitors than resort hotels offer.
* The majority of visitors prefer bed and breakfast meals.

## **References**:

* Geeks for Geeks
* Analytics Vidya
* Research Gate

### 