* **Abstract :**

The data set included customer booking information. As part of the analysis, descriptive statistics were calculated for each variable, and visualizations were created to explore the relationships between various variables. To get insight from the dataset, we built a variety of charts, including a count plot, bar plot, kdeplot, heatmap, pairplot, violin plot, and boxplot.

To prepare the dataset for analysis, we go through several steps. We removed null values, outliers, and duplicated values from the dataset. We convert some variables to the appropriate data type. We done feature engineering to gain a better understanding of datasets. Then we performed statistical analysis by visualization and concluded our results. Some of the most significant outcomes include not assigning a reserved room and a longer lead time, neither of which are grounds for booking cancellation.

* **Table of contents :**
* **Abstract**
* **List of terms**
* **Acknowledgment**
* Problem Statement
* **Introduction**
* **Background**
* **Materials and apparatus**
* **Procedure**
* **Conclusion**
* **List of terms :**

~ Descriptive statistics :-Descriptive statistics are used to describe or summarize data in ways that are meaningful and useful.

~ kdeplot :- A kernel density estimate (KDE) plot is a method for visualizing the distribution of observations in a dataset, analogous to a histogram.

~ Feature engineering :- the pre-processing step of machine learning, which extracts features from raw data.

~ Univariate analysis :- Univariate analysis explores each variable in a data set, separately

~ Bivariate analysis:- The bivariate analysis is will measure the correlations between the two variables.

~ Multivariate analysis:- is a Statistical procedure for analysis of data involving more than two type of measurement or observation.

* **Acknowledgments :**

The Edunet team has been extremely helpful in this exploratory data analysis of hotel bookings. Thank you for helping me analyze the hotel booking dataset. I appreciate the opportunity to explore the data and uncover useful insights. We hope the analysis will be useful to you in making future decisions.

* **Problem Statement :**

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

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. All personally identifying information has been removed from the data.

#### Explore and analyze the data to discover important factors that govern the bookings.

* **Introduction :**

For this hotel booking analysis, the goal was to explore the customer data of a hotel and identify any potential trends or correlations.

The purpose of this exploratory data analysis (EDA) was to explore the hotel booking data set and identify potential relationships between key variables.

* **Background :**

Hotel booking analysis is the process of studying customer booking and reservation patterns in order to better understand customer preferences, identify areas for improvement, and maximize revenue. It is a critical aspect of hotel management and helps hotels to understand their customers and make informed decisions that lead to improved customer satisfaction and increased revenue.

The analysis can include data such as room rate, length of stay, room type, booking source, and more. By analyzing this data, hotel managers can better understand customer behavior, identify potential issues, and develop strategies to improve customer experience.

* **Materials and apparatus:**

Materials and apparatus used for Exploratory Data Analysis include :

- Computers and programming languages such as Python for data manipulation and visualization.

- Jupyterlab software for descriptive statistics, correlation testing, regression, and other modeling techniques.

- Text mining library Pandas for extracting useful information from dataset.

* **Procedure :**

*[1] Data Inspection :*

We examine the dataset's shape and columns after importing it. We check variables and their datatypes for null values using the info() function. We know the basic description of each variable by using the describe() function, such as the mean, median, count, and so on. The variable description provided assists us in understanding the variable's meaning.

This aided us in comprehending datasets.

*[2] Handling duplicated values :*

In the dataset, duplicated values make no sense. All duplicate values in our dataset are removed using the drop duplicates function (.drop\_duplicates()).

*[3] Handling null values*

Null values create errors in datasets and mislead results, so we gave appropriate treatment to null values. By mode, null values were replaced with the .fillna() function. If there are more than half of the null values, it's a better technique to remove that variable.

*[4] Handling outliers:*

Outliers distort results and should be avoided. We trimmed the dataset and removed outliers by using the interquartile range method.

*[5] Feature engineering and data wrangling:*

For ease of comprehension and convenience while performing analysis, we performed feature engineering and developed new features from ones that already existed.

The astype() function is used to change some variables into the correct datatype because they are not in the correct datatype by default.

*[6] Exploratory data analysis:*

We ran our first univariate analysis to investigate data patterns and gain insight. Then, bivariate and multivariate analysis used to examine the correlation of each variable with other variables and to uncover new insights.

* **Conclusion :**

We went through a number of stages before arriving at a resolution. Following EDA, we discovered some key insights that will enhance hotel reservations and reveal data patterns.

The top country with the most bookings is PRT, and the number one agent is 9. Customers favored city hotels more than resort hotels by a margin of 61.07%. The vast majority of hotel bookings are made by new guests. Almost no consumers (3.86%) returned.

Booking cancellations are not caused by a longer Lead time. Only 10% of visitors require space to park their cars. Not assigning a reserved room does not affect ADR. There isn't a lengthy wait for reservations in July.