Hotel-bookings-EDA

The project is on Hotel booking dataset Here is the summary of the EDA project:

Introduction:

The hotel industry has high very competition. So, Hotels and resorts are always searching for new ways to attract the customers attention towards them. One of the most important aspects of this industry is managing bookings and reservations. Hotel owner need to be able to predict customer behavior, determine when to offer promotions and discounts, and optimize pricing strategies. This is where exploratory data analysis (EDA) can play an important role in understanding the factors that affect hotel bookings.

Objective:

The objective of this project is to analyze a hotel booking dataset and identify factors that affect hotel bookings. The dataset contains information on booking details for a city hotel and a resort hotel.

The project aims to answer several questions, such as:

- When is the best time of year to book a hotel room?

- What is the length of stay required to get the best daily rate?

- How does lead time affect bookings?

- Which type of hotel has high demand?

- What is the best time to offer discounts?

- How do the pricing trends vary for different months and hotel types?

- Which booking channels and customer types are most popular for different types of hotels?

- etc.

Data Source:

The dataset used in this project is a hotel_booking.csv. This dataset contains booking information for two types of hotels. The city hotel and resort hotel. The dataset has 119,390 rows and 32 columns. The data set has several columns from which the columns named 'agent','company' and 'children' have null values. There are some duplicate rows which needs to be removed.

Data Cleaning:

The first step in the project is to clean the dataset. This involves handling missing data, removing duplicate records, and correcting data types. I performed several data cleaning steps that are listed below:

- Removing duplicate records: There were some duplicate records were removed by code: "df.drop_duplicates(inplace = True)". This ensure that each row represented a unique booking.

- Handling missing data: There were some missing values in 'agent','company' and 'children' column. The missing values from 'agent' and 'company' was replaced by 0 and the missing values from 'children' was replaced by the mean value.

- Correcting data types: Some columns in the dataset had incorrect data types. These were corrected using appropriate data types. 'children', 'company', 'agent' columns were classified as float which was typecasted into int type. The 'reservation_status_date' column was typecasted into date format.

Exploratory Data Analysis:

The next step in the project was to perform exploratory data analysis (EDA) on the cleaned dataset. The EDA process involved analysing the data to identify patterns, trends, and relationships. Some of the insights found from the EDA process include:

- The majority of bookings were made between May and August, and the peak booking month was August.

- Most bookings were made for one or two nights.

- Approximately one-third of all bookings were cancelled, with resort hotel bookings having a higher cancellation rate than city hotel bookings.

- The most common room type was "A".

- The average daily rate for the resort hotel was higher than that of the city hotel.

- Both hotels had a higher average daily rate in the summer months.

- The average daily rate decreased with increase of lead time.

- The number of repeated guests were very less.

Conclusion:

The EDA process identified several factors that affect hotel bookings. It suggests that the best time to book a hotel room is during the summer months, and that the optimal length of stay is one or two nights. Most of the guests are new and non-repeated.

The insights from this project can be used by hotel owners to improve the booking process and optimize pricing strategies. For example, hotels can offer special promotions and discounts during the period of November and January.