Project Title: Hotel Bookings Data Analysis

Overview:

The Hotel Bookings Data Analysis project utilizes Python to explore and analyze a dataset containing hotel bookings information. The dataset, stored in an Excel sheet ('hotel_bookings.csv'), is loaded into a Pandas DataFrame for manipulation and visualization.

Objectives:

Explore and clean the dataset, handling missing values and outliers.

Conduct data analysis to derive insights into hotel booking patterns.

Visualize key trends and patterns through various plots and charts.

Code Highlights:

Data Exploration and Cleaning:

Displaying dataset shape, columns, and information.

Converting the data type of the 'reservation_status_date' column to datetime.

Identifying unique values for categorical columns.

Handling null values by dropping columns with high null values and rows with remaining null values.

Removing outliers in the 'adr' column.

Data Analysis and Visualizations:

Calculating the percentage distribution of canceled and non-canceled reservations.

Creating a bar plot to visualize the distribution of canceled and non-canceled reservations.

Displaying a clustered column chart depicting reservation status in different hotels.

Analyzing the cancellation rate in resort and city hotels through pie charts.

Using scatter plots to show the average 'adr' for both types of hotels over time.

Adding a 'month' column and visualizing per-month reservation status.

Bar plot illustrating the average 'adr' per month for canceled reservations.

Analyzing the top 10 countries with the highest number of canceled reservations.

Pie chart showcasing the distribution of reservations market segment-wise.

Pie chart presenting the distribution of canceled reservations market segment-wise.

Conclusion:

The project provides valuable insights into hotel booking trends, cancellation rates, and market segment analysis. The visualizations enhance the understanding of the dataset, helping stakeholders make informed decisions in the hospitality industry.

Note:

Ensure that the necessary Python libraries (NumPy, Pandas, Matplotlib, Seaborn) are installed before running the code. Additionally, consider adapting the code for local file access if not using Google Colab.