Synopsis

Title: Hotel Booking Cancellation Analysis

Name: Aryan Verma

Student id: AF0417100

The dataset used for this analysis contains detailed information about hotel bookings, with a particular focus on cancellations. Each row represents an individual booking, and the columns provide various details, including:

1. **Hotel Type:** Type of hotel (e.g., city hotel, resort hotel).
2. **Lead Time**: Number of days between booking and check-in.
3. **Cancellation:** Whether the booking was canceled (Yes/No).
4. **Arrival Date:** The date of arrival for the booking.
5. **Length of Stay:** Duration of the stay in nights.
6. **Guests**: Number of adults, children, and babies in the booking.

7. Booking Channel: Platform through which the booking was made (e.g., online travel agency, direct).

8. Deposit Type: Type of deposit made (e.g., no deposit, refundable, non-refundable).

9. Special Requests: Number of special requests made by the guest.

10. Country: The country from which the booking was made.

11. ADR (Average Daily Rate): Average daily room rate charged.

12.Previous Cancellations: Number of past cancellations by the same guest.

Problem statement :

1. What is the overall cancellation rate?
2. How does the cancellation rate vary by lead time?
3. Which room types have the highest cancellation rates?
4. What impact does the booking source have on the cancellation rate?
5. How do cancellations differ across customer segments?
6. Does seasonality affect cancellation rates?
7. How does the cancellation rate vary by length of stay?

Data Preprocessing Steps:

1. Data Cleaning: Handle missing values and ensure consistency in the dataset.

2. Normalization: Normalize numeric features like ADR if necessary.

3. Encoding Categorical Variables: Convert categorical features (e.g., hotel type, booking channel) into numeric form using one-hot encoding or label encoding.

Implementation Process:

1. Data Ingestion: Load the dataset into a data analysis environment (e.g., Python, Power BI).

2. Preprocessing: Clean and prepare the data for analysis.

3. Exploratory Data Analysis (EDA): Perform EDA to uncover patterns in cancellations, such as the effect of lead time, booking channel, or deposit type.

4. Visualization: Use tools like Power BI, Matplotlib, and Seaborn to create visualizations (e.g., cancellation trends by hotel type, lead time vs. cancellation rate).

5. Reporting: Summarize findings in reports or dashboards, focusing on actionable insights to reduce cancellations.

Dataset:

- Hotel Booking Cancellations

Technologies:

- **Pandas:** For data manipulation.
- **Matplotlib & Seaborn:** For data visualization.
- **Power BI:** For interactive dashboards and reports.

Software Requirements:

- **Operating Systems:** Windows, Linux, macOS.
- **IDE:** Jupyter Notebook (for Python) or Power BI (for visualizations).

Hardware Requirements:

- **RAM:** Minimum 8GB (required for running Power BI smoothly), recommended 16GB.
- **Processor:** Minimum Intel i5, recommended Intel i7 for faster data processing.
- **Storage:** SSD recommended for quicker data handling, at least 256GB storage space for large datasets.
