**Hotel Booking Analysis Project**

**Step 1: Setup and Libraries**

Ensure you have the following libraries installed:

- Pandas for data manipulation.

- NumPy for numerical operations.

- Matplotlib for visualizations.

**Step 2: Load and Explore the Dataset**

1. Load the dataset into a Pandas DataFrame.

2. Inspect the first few rows of the dataset using `.head()`.

3. Check for missing values using `.isnull().sum()` to identify columns that may require cleaning.

4. Data types: Ensure that columns like dates, categories, or numerical values are correctly typed.

**Step 3: Data Cleaning**

1. Handle missing values:

- Decide whether to drop rows/columns or fill missing values based on the importance and the percentage of missing data.

2. Convert data types if needed. For example, if dates are stored as strings, convert them to datetime objects.

3. Remove duplicates if there are any, using `.drop\_duplicates()`.

**Step 4: Exploratory Data Analysis (EDA)**

4.1 Descriptive Statistics

- Use Pandas `.describe()` to get a statistical summary of the dataset, including mean, median, and standard deviation for numerical columns.

- Calculate specific statistics using NumPy, like the average lead time, median number of special requests, etc.

4.2 Correlation Analysis

- Use `.corr()` to generate a correlation matrix that shows how variables are related to each other.

- Visualize the correlation matrix using a heatmap (you might use Matplotlib or another library like Seaborn if allowed).

4.3 Distribution of Bookings by Market Segment

- Use a bar chart to visualize the distribution of bookings across different market segments.

- Discuss which market segment is most prevalent and what that might imply for the hotel’s marketing strategy.

4.4 Lead Time Analysis

- Plot the distribution of lead times with a histogram.

- Analyze the skewness of the distribution—long tail, short lead times, etc.

- Discuss how lead time might affect booking behavior.

4.5 Customer Type and ADR (Average Daily Rate)

- Group the data by customer type and calculate the average ADR.

- Create a bar chart to visualize the differences in ADR across customer types.

- Discuss which customer type is more valuable in terms of revenue and how this could influence pricing strategies.

4.6 Booking Cancellations

- Analyse the cancellation rates by different customer segments, market segments, or room types.

- Create a bar chart or pie chart to visualize the cancellation distribution.

- Discuss potential factors leading to higher cancellation rates and suggest strategies to mitigate them.

4.7 Time Series Analysis

- Analyse trends over time:

- Monthly Trends: Number of bookings per month, ADR trends, etc.

- Yearly Trends: How has the number of bookings changed over the years?

- Use line charts to visualize these trends.

- Discuss any seasonal trends or noticeable shifts in booking behavior over time.

4.8 Correlation Between Variables

- Investigate relationships such as:

- The correlation between lead time and cancellation rates.

- The relationship between number of special requests and ADR.

- How booking lead time affects ADR.

- Visualize correlations using scatter plots or pair plots.

**Step 5: Advanced Analysis**

5.1 Segmented Analysis

- Perform segmented analysis for different categories such as:

- Weekday vs. Weekend Bookings: Analyze how booking patterns differ between weekdays and weekends.

- Family vs. Solo Travellers: Compare booking behaviors between family groups and solo travellers.

5.2 Predictive Analysis (Optional)

- If you're familiar with machine learning, you might consider building a predictive model to estimate the likelihood of a booking being canceled, based on features like lead time, market segment, etc.

**Step 6: Visualisation**

- Ensure that all visualizations are clear, labeled, and include titles, axes labels, and legends where necessary.

- Use different chart types to communicate different insights:

- Bar Charts for categorical comparisons.

- Line Charts for trends over time.

- Histograms for distribution analysis.

- Scatter Plots for correlation analysis.

**Step 7: Conclusions and Recommendations**

- Summarise key findings from your analysis:

- Which market segments are the most valuable?

- What trends are evident over time?

- Are there any red flags like high cancellation rates in specific segments?

- Provide actionable recommendations based on your findings:

- Suggest marketing strategies to target high-value segments.

- Recommend pricing adjustments for certain customer types or booking lead times.

- Offer suggestions to reduce cancellations.

**Step 8: Saving and Sharing Results**

- Save the cleaned dataset if necessary, for future analysis or sharing.

- Document your process: If using a Jupyter Notebook, include markdown cells to explain each step.

- Share visualisations: Export graphs and charts as images to a PowerPoint presentations.

- Project presentation: Build a PowerPoint presentation for this project.