**HOTEL BOOKING DATA ANALYSIS**

**PROJECT DESCRIPTION**

1. **Aim of the Project:**

The aim of the Hotel Booking Analysis project is to uncover key insights into hotel booking behaviours and the factors that influence customer decisions. By analysing a comprehensive dataset of hotel reservations, the project seeks to identify trends in booking patterns, customer preferences, pricing strategies, and the reasons behind cancellations and modifications. Understanding these elements is crucial for hotel management to optimize their operations, enhance customer satisfaction, and improve revenue management.

The analysis focuses on comparing different types of hotels, such as City Hotels and Resort Hotels, to evaluate how factors like average daily rate (ADR), booking channels, and special requests influence customer choices. By examining these variables, the project aims to identify the drivers of customer behaviour and how they impact the booking outcomes.

Additionally, the project aims to assess the effectiveness of current marketing and sales strategies by analysing customer segmentation and the impact of various distribution channels on booking volumes. By leveraging predictive modelling techniques, the project seeks to forecast booking outcomes, such as the likelihood of cancellations or the expected revenue from a booking, allowing for more informed decision-making.

1. **Business Problem or Problem Statement:**

The hotel industry operates in a competitive and rapidly evolving landscape, where understanding and predicting customer booking behaviour is critical for success. Hotels face challenges in optimizing their operations, pricing, and marketing strategies due to the complexity of factors influencing booking decisions. The frequent occurrence of booking cancellations and modifications further complicates revenue management, making it difficult for hotels to forecast occupancy accurately and allocate resources efficiently. To stay competitive, hotels must fully leverage their data to gain insights into customer preferences, pricing dynamics, and the effectiveness of their marketing efforts.

 **Understanding Customer Behaviour**: Hotels need to comprehend the factors influencing customer booking decisions, such as booking timing, hotel type preference (City Hotel vs. Resort Hotel), distribution channels, and pricing sensitivity. Without this understanding, hotels risk misaligning their offerings with customer needs, leading to lost revenue and diminished customer satisfaction.

 **Optimizing Revenue Management**: The unpredictability of booking cancellations and modifications presents a significant challenge in managing revenue streams. Inaccurate occupancy forecasts can lead to inefficient resource allocation, such as overstaffing or understaffing, and missed opportunities to maximize revenue. Hotels need robust predictive tools and insights to anticipate cancellations and adjust strategies accordingly.

 **Leveraging Data for Competitive Advantage**: Despite collecting vast amounts of booking data, many hotels struggle to translate this data into actionable insights. This gap prevents them from fully optimizing pricing, marketing, and operational strategies, leading to missed opportunities for growth, customer retention, and overall competitiveness.

1. **Project Description:**

**Data Features**

* Hotel: Type of hotel (Resort Hotel or City Hotel).
* is\_canceled: Indicates if the booking was cancelled (1) or not (0).
* Lead-time: Number of days between the booking and the arrival date.
* arrival\_date\_year: Year of arrival date.
* arrival\_date\_month: Month of arrival date.
* arrival\_date\_week\_number: Week number of the arrival date.
* arrival\_date\_day\_of\_month: Day of the month of arrival.
* stays\_in\_weekend\_nights: Number of weekend nights stayed or booked to stay.
* stays\_in\_week\_nights: Number of week nights stayed.
* Adults: Number of adults.
* Children: Number of children.
* Babies: Number of babies.
* Meal: Meal plan booked (BB, FB, HB, SC).
* Country: Country of origin.
* market\_segment: Market segment (e.g., Travel Agents, Tour Operators).
* Distribution channel: Booking distribution channel (e.g., Travel Agents, Tour Operators).
* is\_repeated\_guest: Indicates if the guest is a repeat visitor (1) or not (0).
* Previous cancellations: Number of previous bookings cancelled.
* previous\_bookings\_not\_canceled: Number of previous bookings not cancelled.
* reserved\_room\_type: Room type reserved.
* assigned\_room\_type: Room type assigned.
* Booking changes: Number of changes made to the booking.
* Deposit type: Type of deposit made (No Deposit, Non Refund, Refundable).
* Agent: ID of the travel agency that made the booking.
* Company: ID of the company/entity responsible for the booking.
* days\_in\_waiting\_list: Number of days the booking was on the waiting list.
* Customer type: Type of customer (Group, Transient, Transient-party).
* Adr: Average Daily Rate (total lodging transactions divided by total staying nights).
* required\_car\_parking\_spaces: Number of car parking spaces required.
* total\_of\_special\_requests: Number of special requests made by the customer.
* Reservation status: Status of the reservation (Check-Out, No-Show).
* reservation\_status\_date: Date when the reservation status was last updated.
* Name: Name of the guest (anonymized).
* Email: Email of the guest (anonymized).
* Phone-number: Phone number of the guest (anonymized).

**Scope and Objectives:**

**Data Processing and Exploration**: Collect, clean, and pre-process hotel booking data to ensure accuracy and consistency. Perform exploratory data analysis (EDA) to identify key trends and relationships within the data.

**Booking Trends Analysis**: Analyse booking patterns over time, including seasonality, peak periods, and differences between hotel types (e.g., City Hotel vs. Resort Hotel).

**Customer Behaviour Analysis**: Examine factors that influence customer choices, such as room types, meal plans, and booking channels. Segment customers based on booking behaviour and preferences to identify distinct groups.

**Revenue Management**: Analyse pricing strategies, particularly the Average Daily Rate (ADR), to assess how pricing affects booking behaviour. Develop models to forecast occupancy rates and predict cancellations, aiding in revenue optimization.

**Marketing and Channel Performance**: Evaluate the effectiveness of different marketing strategies and distribution channels in driving bookings. Provide insights into the best channels for targeting specific customer segments.

**Predictive Modeling**: Build models to predict key outcomes such as booking cancellations, customer preferences, and revenue forecasts.

**Visualization and Reporting**: Develop interactive dashboards and visual reports to present findings, enabling hotel management to easily interpret data and make informed decisions. Use data visualization tools to effectively communicate complex insights and trends to stakeholders.

**Statistical Testing**: Conduct statistical tests (e.g., t-tests, chi-square tests) to validate hypotheses and determine the significance of relationships between variables. Use these tests to support data-driven decisions, such as identifying factors that significantly influence booking behaviour and pricing strategies.

**Technologies and Methodologies**

**Data Collection and Preparation:**

* **Data Sources**: Common formats for storing hotel booking data, often used for importing data into analysis environments using csv/excel formats files included with various columns.
* **Data Cleaning**: Pandas is essential for data cleaning, including handling missing values, filtering data, and converting data types. Techniques are applied to identify and address outliers, ensuring the dataset is clean and ready for analysis. Includes normalization, encoding categorical variables, and feature engineering to prepare data for analysis.

**Exploratory Data Analysis (EDA):**

* **Descriptive Statistics**: Used to calculate summary statistics such as mean, median, standard deviation, and range, providing an initial understanding of the data.
* **Visualization Techniques**:
* **Count plot**: Creates count plots to visualize the frequency of categorical data, helping to identify the distribution of bookings across different categories.
* **Bar graph**: Used to visualize the relationship between categorical variables and numerical values, such as the average daily rate (ADR) across different hotel types.
* **Pie chart**: Used to represent the proportion of bookings or cancellations across different categories, providing a clear visual summary of the data distribution.

**Statistical Testing:**

* **Independent 2 Sample t-test**: Used to compare the means of two independent groups (e.g., comparing ADR between City Hotels and Resort Hotels) to determine if there is a statistically significant difference between them.
* **Chi-Square Test**: Employed to assess the association between two categorical variables (e.g., booking status and customer type), testing whether distributions of categorical variables differ from one another.

**Advanced Analytics**:

* **Correlation Analysis**: Analyses the correlation between variables, visualized through heat maps to understand relationships and dependencies in the data.
* **Predictive Modelling**: Used for developing models to predict outcomes like booking cancellations, based on features such as lead time and customer type.
* **Bar plot**: A bar plot is utilized to compare different groups, such as the average revenue generated by different customer segments, aiding in identifying profitable customer groups.

**Data Visualization Tools:**

* **Matplotlib**: A foundational Python library used for creating static, animated, and interactive visualizations. Matplotlib is versatile and widely used for simple plots like line graphs, bar charts, and histograms.
* **Seaborn**: Built on top of Matplotlib, Seaborn provides a high-level interface for drawing attractive and informative statistical graphics, including heatmaps, count plots, and violin plots.
* **Jupyter Notebooks**: An interactive environment that combines code, visualizations, and narrative text, allowing for dynamic analysis and easy sharing of results. Jupyter Notebooks are ideal for iterative data analysis and reporting.

1. **Code Implementation**

* **Data Collection and Preparation:**
* **CSV/Excel Files:** Hotel booking data often comes in CSV or Excel formats. You can load these files into Python using libraries like Pandas for further analysis.
* **Handling Missing Values:** Data cleaning involves filling or dropping missing values to maintain the integrity of your dataset. For example, missing values in columns like 'agent' or 'company' could be replaced with 0, while numerical columns like 'children' could be filled with the median value.
* **Data Transformation**: This includes converting categorical data into numerical format for analysis. For instance, converting the 'arrival\_date\_month' from text (e.g., "January") to a numerical value (e.g., 1 for January).
* **Visualization Techniques:**
* **Count plot:** Seaborn can be used to create count plots that show the frequency of categories within a variable, such as the number of bookings per market segment.
* **Bar graph:** Bar graphs can be created using Matplotlib or Seaborn to compare numerical values across different categories, like the average daily rate (ADR) by hotel type.
* **Pie chart:** Pie charts in Matplotlib are useful for visualizing the proportion of different categories, such as the percentage of bookings from different customer types.
* **Statistical Testing:**
* **T-Test:** Using Scipy, you can conduct an independent two-sample t-test to compare the means of two groups. For example, comparing the ADR between City Hotels and Resort Hotels to see if there is a significant difference.
* **Chi-Square Test:** This test can be used to determine if there is an association between two categorical variables. For instance, you might want to test if booking status (cancelled or not) is related to the market segment.
* **Correlation Analysis:** Using Pandas and Seaborn, you can calculate and visualize the correlation between different variables, helping you identify relationships that might influence booking behaviour.

1. **Conclusion:**

The Hotel Booking Data Analysis provided comprehensive insights into the booking behaviours and preferences of customers for City Hotels and Resort Hotels. We identified key trends and factors that significantly influence hotel bookings.

Our analysis began with rigorous data cleaning to ensure accuracy and reliability. This included removing irrelevant variables, handling missing values, and ensuring consistency across the dataset. Through Exploratory Data Analysis (EDA), we examined booking patterns over time, revealing that City Hotels tend to attract more bookings during business periods, while Resort Hotels see higher demand during vacation seasons.

Statistical testing played a crucial role in our analysis. We employed the independent two-sample t-test to compare the Average Daily Rate (ADR) between City Hotels and Resort Hotels, uncovering that Resort Hotels generally command higher ADRs, particularly during peak seasons. Additionally, the Chi-Square test was used to explore the relationships between categorical variables, such as market segments and distribution channels, providing insights into how these factors influence booking decisions.

Advanced analytics allowed us to further investigate the key determinants of booking behaviour. Factors such as the type of meal plan, room type, and deposit requirements emerged as significant predictors of whether a booking would be made at a City Hotel or a Resort Hotel. These findings are critical for informing strategic decisions in pricing, marketing, and operational management.

The use of data visualization tools, including Matplotlib and Seaborn, enabled us to present these insights clearly and effectively, making the analysis accessible to stakeholders. Overall, this analysis offers valuable guidance for optimizing hotel operations, enhancing customer satisfaction, and driving profitability through data-driven decision-making.