## **Problem Statement: Uncovering Patterns in Hotel Booking Data for Operational Efficiency and Revenue Growth**

### **Overview:**

The goal of this exploratory data analysis is to investigate a Resort Hotel’s booking dataset to identify key patterns, trends, and relationships that can support data-driven decision-making.

This includes analyzing booking behavior, customer demographics, pricing strategies, and operational factors such as room assignments and special requests. The study will involve cleaning and preprocessing the data, examining variable correlations, and validating key business assumptions through hypothesis testing.

### **Core Objectives:**

* Understand how customer attributes and booking behaviors impact revenue.
* Identify trends in lead time, stay duration, and booking channels.
* Detects inconsistencies or anomalies in room allocation and guest handling.
* Explore relationships between booking patterns and customer satisfaction indicators.
* Evaluate whether specific operational or customer variables significantly affect outcomes such as ADR or room upgrades.

### **Scope of Work:**

#### **1. Data Cleaning and Preprocessing**

Handle missing values in columns, Convert categorical variables into usable formats for analysis. Parse and standardize date columns, Create derived fields, Remove duplicate records if any, handle outliers if any.

#### **2. 📊 Exploratory Data Analysis**

Univariate and bivariate and multivariate analysis using relevant plots, Time-series analysis of booking trends. Guest demographics and distribution by country. Market segment share and ADR (Average Daily Rate) comparison. Booking lead time distribution across customer types

#### **3. Correlation Analysis**

Compute Pearson/Spearman correlation matrix for numeric feature. Visualize correlations using heatmaps to detect multicollinearity. Explore how strongly adr is influenced by lead time, special requests, and booking changes

#### **4. Hypothesis Testing**

Use statistical tests to validate business assumptions:

H0: There is no difference in ADR between bookings made through Online TA and Direct channels  
H0: Room upgrades are independent of lead time

H0: Average stay duration does not differ between customer types

#### **5. Key Business Questions**

* What influences ADR the most?
* Do guests who book earlier tend to request more changes?
* Are there pricing or booking differences across countries?
* Is there a pattern in room upgrades or reassignment?
* Are reserved room types consistently matched with assigned room types?
* What are the most common guest demographics (e.g., group size, nationality)?
* Are there patterns in guest types (e.g., transient vs. corporate) that influence booking behavior?
* How does booking lead time vary across customer types and countries?
* Are longer lead times associated with fewer booking changes or cancellations?
* What is the typical duration of stay, and how does it vary by customer type or segment?
* How often are guests upgraded or reassigned to a different room type?
* Are guests who make special requests more likely to experience booking changes or longer stays?
* Do certain market segments or distribution channels show higher booking consistency or revenue?
* What factors are most strongly associated with higher ADR?
* Are there customer types or segments consistently contributing to higher revenue?
* Do bookings with more lead time or from specific countries yield higher ADR?
* Are guests with higher ADR more likely to request special services or make booking modifications?
* Do guests from different countries behave differently in terms of booking timing or stay length?
* Are guests who make booking changes more likely to request additional services or cancel?

### **Deliverables**

* Exploratory Data Analysis Jupyter Notebook (.ipynb)
* Final Report (.docx)

## **Data Description**

The dataset contains detailed information about hotel bookings for a Resort Hotel (H1) and a City Hotel (H2). Each row represents a single booking.

| Column Name | Description |
| --- | --- |
| hotel | Type of hotel: "Resort Hotel" or "City Hotel" |
| is\_canceled | Indicates whether the booking was canceled (1) or not (0) |
| lead\_time | Number of days between booking and arrival date |
| arrival\_date\_year | Year of arrival |
| arrival\_date\_month | Month of arrival |
| arrival\_date\_week\_number | Week number of the year for arrival |
| arrival\_date\_day\_of\_month | Day of the month for arrival |
| stays\_in\_weekend\_nights | Number of weekend nights (Saturday or Sunday) stayed/booked |
| stays\_in\_week\_nights | Number of weekday nights (Monday to Friday) stayed/booked |
| adults | Number of adults included in the booking |
| children | Number of children included in the booking |
| babies | Number of babies included in the booking |
| meal | Meal plan booked (SC, BB, HB, FB, Undefined) |
| country | Country of origin (ISO 3166-1 alpha-3 format) |
| market\_segment | Market segment (e.g., Direct, Corporate, Online TA, Offline TA/TO) |
| distribution\_channel | Booking distribution channel (e.g., Direct, TA/TO) |
| is\_repeated\_guest | 1 if the guest has made previous bookings, 0 otherwise |
| previous\_cancellations | Number of previous bookings canceled by the customer |
| previous\_bookings\_not\_canceled | Number of previous bookings not canceled |
| reserved\_room\_type | Code of the room type initially reserved |
| assigned\_room\_type | Code of the room type actually assigned |
| booking\_changes | Number of changes made to the booking before check-in |
| 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 responsible for the booking/payment |
| days\_in\_waiting\_list | Days the booking spent on the waiting list before confirmation |
| customer\_type | Type of customer (Contract, Group, Transient, Transient-party) |
| adr | Average Daily Rate (lodging revenue / total nights stayed) |
| required\_car\_parking\_spaces | Number of parking spaces requested |
| total\_of\_special\_requests | Number of special requests made by the customer |
| reservation\_status | Final status of the reservation: Canceled, Check-Out, or No-Show |
| reservation\_status\_date | Date when the reservation status was last updated |