

# Bookin Information for a Multi-service Business

## Data Cleaning and Exploratory Data Analysis (EDA) Report

### Project Overview

The dataset provided contains booking information for a multi-service business. It includes data for class bookings, subscriptions, facility rentals, and birthday party reservations. The dataset is a real-world export and may contain inconsistencies, missing values, and other imperfections that require cleaning and preprocessing.

---

### Dataset Columns

The dataset initially contained the following columns:

- **Booking ID**
  - **Customer ID**
  - **Customer Name**
  - **Booking Type**
  - **Booking Date**
  - **Status**
  - **Class Type**
  - **Instructor**
  - **Time Slot**
  - **Duration (mins)**
  - **Price**
  - **Facility**
  - **Theme**
  - **Subscription Type**
  - **Service Name**
  - **Service Type**
  - **Customer Email**
  - **Customer Phone**
- 

### Data Cleaning Process

## 1. Handling Missing Values

- Identified missing values in various columns.
- The column '**Subscription Type**' had no values, so it was dropped.
- The '**Time Slot**' column had missing values, which were filled using:
  1. **Mode Imputation**: Filled missing values with the most frequent time slot.
  2. **Forward Fill (ffill)**: Filled missing values with the previous row's value.
  3. **Backward Fill (bfill)**: Filled missing values with the next row's value.
- The '**Facility**' and '**Service Name**' columns contained the same values, so missing values in one column were filled using the other.
- After merging, the '**Service Name**' column was dropped.
- The '**Class Type**' column was also dropped after confirming it duplicated values present in other columns.

## 2. Handling Duplicate Rows

- Checked for duplicate rows in the dataset.
- Printed the number of duplicate rows found (if any).

## 3. Handling Categorical Data

- Replaced missing values in categorical columns with placeholders:
  - '**Instructor**' → Replaced missing values with 'No Instructor'.
  - '**Theme**' → Replaced missing values with 'Not Specified'.

## 4. Handling Numerical Data

- The '**Duration (mins)**' column had missing values, which were replaced with the median value of the column.

## 5. Handling Contact Details

- Missing values in contact details were replaced with placeholders:
  - '**Customer Email**' → 'No Email Provided'.
  - '**Customer Phone**' → 'No Phone Provided'.

## 6. Saving Cleaned Data

- The cleaned dataset was saved to a new Excel file:  
**File** : Cleaned\_Dataset\_for\_dataAnalysis.xlsx

---

## Exploratory Data Analysis (EDA)

After cleaning, we performed EDA to identify trends and insights:

### 1. Distribution of Booking Types

- Identified the most popular booking types (e.g., Class, Facility, Birthday).

### 2. Peak Booking Times

- Analyzed booking trends based on time slots.

### 3. Revenue Analysis

- Calculated revenue trends by price, duration, and service type.

### 4. Instructor-Based Analysis

- Determined which instructors had the most bookings.

### 5. Customer Engagement

- Identified frequent customers based on booking history.

This cleaned dataset is now ready for **Power BI visualization**.

---

## Power BI Dashboard Components

The dashboard contains multiple visualizations that provide insights into booking trends, revenue generation, and customer preferences.

### 1. Type of Services Booking (Pie Chart - Top Left)

What is the distribution of different types of services booked?

Which service category (Class, Facility, Birthday Party) has the highest bookings?

### 2. Total Revenue by Booking (Bar Chart - Bottom Left)

How much revenue is generated from each booking category?

Which booking type contributes the most to total revenue?

### 3. Total Revenue Generation by Facilities (Bar Chart - Center)

Which facility generates the highest revenue?

How do different facilities compare in terms of revenue contribution?

### 4. Monthly Revenue Generation (Line Chart - Top Right)

How does revenue change over different months?

What is the trend of confirmed vs. pending bookings over time?

In which months does the business generate the highest revenue?

### 5. Revenue Generated by Services (Pie Chart - Bottom Left)

What percentage of total revenue comes from each type of service?

Are all service categories contributing equally to revenue?

#### **6. Booking Time Slots (Line Chart - Bottom Center)**

What are the peak booking hours for different service types?

At what time slots do most customers book their services?

#### **7. Total Booking Facility (Vertical Bar Chart - Bottom Right)**

Which facility is booked the most?

How does the number of bookings vary across different facilities?

Which facility has the lowest number of bookings?

#### **8. Month Selection (Filter - Top Center)**

How do bookings and revenue change when selecting a specific month?

How do different months compare in terms of performance?

Helps in understanding the most profitable service type.

---

## **Business Insights and Decision-Making**

- **Identifying High-Demand Facilities** → Focus marketing efforts on the most popular facilities.
- **Revenue Optimization** → Track monthly revenue trends and adjust pricing strategies.
- **Customer Behavior Analysis** → Optimize staffing and resource allocation based on peak booking times.
- **Booking Type Analysis** → Invest in service types that generate higher revenue.