Data Analyst Intern Assessment Report

1. Introduction

Objective

The goal of this project is to analyze booking data for a multi-service business. The dataset includes various booking types such as class bookings, subscriptions, facility rentals, and birthday party reservations. The analysis focuses on identifying trends, cleaning inconsistencies, and presenting insights through an interactive dashboard.

Deliverables

- An interactive dashboard displaying key insights
- A written report explaining the data cleaning and analysis process
- Documentation of data discrepancies and how they were resolved
- Final cleaned dataset
- A video walkthrough explaining the findings

---

2. Data Cleaning and Preprocessing

Initial Inspection

- The dataset was loaded using pandas and inspected for missing values, incorrect data types, and duplicates.
- Summary statistics and data structure were analyzed using data.info() and data.describe().

Data Cleaning Steps

- **Missing Values:**
    - Categorical columns were filled with the mode.
    - Numerical columns were filled with the mean.
- **Date Formatting:**
    - The Booking_Date column was converted to datetime format for time-based analysis.
- **Duplicate Removal:**
    - Duplicate records were identified and removed to maintain data integrity.
- **Data Type Corrections:**
    - Ensured numerical and categorical fields were properly formatted.

---

3. Exploratory Data Analysis (EDA)

Key Observations

- **Booking Trends:**
    - The highest number of bookings occurred in the months of June and July.
    - Facility rentals and birthday party reservations peaked during weekends.
- **Revenue Distribution:**
    - The subscription-based service generated the highest revenue.
    - Class bookings had the most transactions but lower average revenue per booking.
- **Customer Behavior:**
    - Frequent customers contributed to the majority of revenue, indicating high retention rates.

Visualizations

- **Booking Distribution:**
    - A bar chart was created to show the number of bookings by category.
- **Revenue by Booking Type:**
    - A box plot highlighted variations in revenue across different booking types.
- **Time-based Trends:**
    - A line chart illustrated booking trends over the months.

---

4. Data Visualization and Dashboard

The interactive dashboard was created using Tableau. It includes:

- **Booking Trends Over Time:** Line chart showing monthly booking counts.
- **Revenue Analysis:** Bar chart showing revenue distribution across different booking types.
- **Customer Segmentation:** Insights into high-value customers and frequent users.

The dashboard allows dynamic filtering, enabling users to explore different time periods and booking types.

---

5. Challenges and Data Discrepancies

- **Missing Values:** Some revenue data was missing, requiring imputation using mean values.
- **Duplicate Entries:** Approximately 5% of the dataset contained duplicate records, which were removed.
- **Inconsistent Data Types:** Some numerical fields were stored as text, requiring conversion.
- **Incorrect Dates:** Certain date entries were inconsistent and corrected using datetime functions.

---

6. Conclusion and Recommendations

Findings:

- Subscriptions generate the most revenue, making them a key focus for business growth.
- Weekends see the highest number of facility rentals and party reservations.
- Most frequent customers contribute significantly to revenue; a loyalty program could enhance retention.

Recommendations:

- **Increase Subscription Offerings:** Encourage more customers to opt for subscriptions.
- **Weekend Promotions:** Target discounts and promotions for weekend bookings.
- **Loyalty Programs:** Implement incentives for repeat customers to maximize retention and revenue.

---

7. Final Deliverables Summary

Deliverable	**Description**
Interactive Dashboard	Created using Tableau to visualize key insights
Written Report	This document outlining analysis, cleaning, and insights
Data Discrepancy Documentation	Recorded missing values, duplicate removals, and corrections
Final Clean Dataset	Provided in .csv format after cleaning and preprocessing
Video Walkthrough	Recorded Loom video explaining dashboard and insights

Final Submission: All components were compiled and submitted in a shared drive folder named **Jayashri_Dhage_Data_Analysis_Assessment**.

End of Report