

Data Analyst Internship Assessment – Report

By Sivaramakrishna Bandla

1. Introduction

Overview of the Dataset: The dataset contains booking records for a multi-service business, including classes, facility rentals, subscriptions, and birthday party reservations. The data may include inconsistencies, missing values, and duplicate records that require cleaning and preprocessing.

Key Objectives:

- Clean and preprocess the dataset to ensure accuracy.
- Perform exploratory data analysis (EDA) to extract meaningful insights.
- Execute SQL queries to derive key metrics.
- Create an interactive dashboard in Power BI/Tableau.
- Provide business recommendations based on findings.

2. Data Cleaning

Issues Found:

- Missing values in customer phone numbers, time slots, and durations.
- Inconsistent data in Booking Type and Service Type columns.
- Duplicates in booking records.
- Formatting inconsistencies in customer names and emails.

Resolutions:

- Missing phone numbers were imputed with 'Not Provided' or retrieved from other entries.
- Missing time slots were replaced with the most frequent slot per service type.
- Standardized Booking Type and Service Type by grouping similar categories.
- Duplicates were removed using a unique identifier (Booking ID).
- Email addresses were formatted correctly by removing extra spaces and ensuring consistency.

3. Exploratory Data Analysis (EDA)

Summary Statistics:

- Total Bookings: 1001
- Total Revenue: 139K
- Average Booking Price: 139
- Most Popular Service: Facility Booking (34%)

Key Trends:

- Peak booking times occur between 5 PM - 8 PM.
- Birthday Party reservations generate the highest revenue per booking.
- Pending bookings (510) are slightly higher than confirmed bookings (490), indicating a need for follow-up.

4. SQL Queries & Insights

Total Revenue by Month:

```
SELECT DATE_FORMAT(Booking_Date, '%Y-%m') AS Month, SUM(Price) AS Total_Revenue  
FROM bookings  
GROUP BY Month  
ORDER BY Month;
```

Insight: Revenue trends over time help identify seasonal booking patterns.

Top 10 Customers by Total Spend:

```
SELECT Customer_Name, SUM(Price) AS Total_Spent  
FROM bookings  
GROUP BY Customer_Name  
ORDER BY Total_Spent DESC  
LIMIT 10;
```

Insight: Identifies high-value customers for loyalty programs.

Peak Booking Hours:

```
SELECT Time_Slot, COUNT(*) AS Total_Bookings  
FROM bookings  
GROUP BY Time_Slot  
ORDER BY Total_Bookings DESC;
```

Insight: Determines the most popular booking times to optimize staffing.

Average Revenue Per Customer:

```
SELECT Customer_Name, AVG(Price) AS Avg_Spending  
FROM bookings  
GROUP BY Customer_Name  
ORDER BY Avg_Spending DESC;
```

Insight: Helps understand customer spending behavior.

Booking Trends by Facility Usage:

```
SELECT Facility_Name, COUNT(*) AS Total_Bookings  
FROM bookings  
GROUP BY Facility_Name  
ORDER BY Total_Bookings DESC;
```

Insight: Identifies the most utilized facilities.

5. Dashboard Explanation

The Power BI/Tableau dashboard consists of:

- **KPIs & Overview:** Total Bookings, Revenue, Average Price, Most Popular Services.
- **Booking Trends:** Line chart for monthly bookings, heatmap for peak hours.
- **Customer Insights:** Top 10 customers by spend, total bookings.
- **Revenue Analysis:** Revenue breakdown by service type and facility.

Key Insights from Visuals:

- Facility bookings drive the highest number of transactions.
- Birthday Party bookings contribute the highest revenue share.
- Significant potential in converting pending bookings into confirmed ones.

6. Business Recommendations

- **Optimize Booking Slots:** Encourage customers to book in off-peak hours with discounts.
- **Loyalty & Rewards Program:** Reward high-spending customers with special offers.
- **Follow-up Strategy:** Implement automated reminders for pending bookings.
- **Service Expansion:** Consider adding more facilities based on high-demand services.