

DATA ANALYST INTERN ASSESSMENT

Data Discrepancies & Inaccuracies Found

During the data cleaning and preprocessing phase, several data discrepancies and inaccuracies were identified and addressed to ensure the accuracy and reliability of the analysis. These issues are categorized as follows:

1. Missing Data

- **Price Column:** Some records had missing prices, which were imputed using the average price of the same service on the same day.
- **Time Slot Column:** Missing values were replaced with the most commonly booked time slot for each service.
- **Duration Column:** Missing duration values were filled using the mode of duration grouped by service name.
- **Customer Contact Details:** Missing email and phone numbers were labeled as "Not Provided."

2. Inconsistent Data Entries

- **Service Type & Status:** Some service types had inconsistent labels (e.g., "party room" vs. "Party Room"), which were standardized for uniformity.
- **Booking Status:** Entries labeled as "Pending" were excluded as they did not contribute to final revenue or confirmed bookings.
- **Facility & Theme:** Some records had different formats or missing values, which were assigned default labels ("Not Specified" and "Default Theme").

3. Incorrect Data Formats

- **Date Format Issues:** The booking date was initially stored in multiple formats. It was standardized to **YYYY-MM-DD** for consistency and ease of trend analysis.

4. Outliers & Data Integrity Issues

- **Revenue Anomalies:** Some bookings had an abnormally high or low price. These were verified, and any incorrect values were corrected based on similar bookings.
- **Booking Duplicates:** Duplicate records were identified and removed to prevent overestimation of total bookings and revenue.

5. Dropped Columns & Rows

- **Subscription Type Column:** This column was removed as it had null values across all records.
- **Bookings with Missing Essential Details:** Rows missing critical information like **time slot, duration, and price** were excluded to maintain data accuracy.

By addressing these discrepancies, the dataset was cleaned and structured for reliable analysis, ensuring that the insights derived from the Power BI dashboard are both **accurate and actionable**.

REPORT BY :- PRATHAM PATIDAR