

Omnify Data Analyst Intern Assessment Report

1. Data Cleaning Process

Objective

Prepare the raw dataset for analysis by handling missing values, fixing inconsistencies, and resolving discrepancies.

Steps

****Load Data****: Imported `bookings.csv` using pandas.

****Missing Values****:

`Instructor`: Filled with 'Unknown' (162 missing).

`Time Slot`: Filled with 'Not Specified' (326 missing).

`Duration (mins)`: Filled with median (90 mins, no missing).

`Price`: No 0.0 values; kept as is after verifying all are valid.

`Customer Email/Phone`: Filled with 'Not Provided' (174 and 173 missing, respectively).

`Class Type/Theme/Subscription Type`: Filled with 'N/A' (614, 644, 1000 missing).

`Booking Date`: Coerced to datetime, 472 NaT values remain.

****Inconsistencies****:

Converted `Booking Date` to datetime format.

Checked `Booking ID` duplicates: 0 found.

Standardized `Status` to capitalized form (e.g., 'Pending', 'Confirmed').

****Discrepancies****:

`Service Type` vs `Booking Type`: 0 mismatches (all aligned in this dataset).

Output

Saved cleaned data as `cleaned_bookings.csv`.

2. Analytical Approach

Tools

Python (Pandas for cleaning, Dash/Plotly for visualization).

Jupyter Notebook in Anaconda for execution.

Methodology

****Exploratory Data Analysis (EDA)**:** Examined distributions of `Booking Type`, `Status`, `Price`, and `Facility`.

****Visualization**:** Created an interactive dashboard to filter by `Booking Type` and display:

1. Status distribution.
2. Revenue by facility.
3. Booking trends over time.

3. Data Observations

****Booking Type Distribution**:** Birthday Party (44%), Facility (32%), Class (24%).

****Status**:** 51% Confirmed, 49% Pending.

****Revenue**:** Total ~\$162,000; Birthday Parties contribute ~\$80,000.

****Facility Usage**:** Play Area and Party Room each ~50% of bookings.

****Time Trends**:** Peak in May 2025 (372 bookings), likely seasonal demand.

4. Insights

****High Revenue Source**:** Birthday Parties are the most lucrative.

****Facility Demand**:** Equal usage suggests balanced capacity needs.

****Seasonality**:** May spike indicates summer planning opportunities.

5. Recommendations

****Data Improvement**:** Ensure `Booking Date` is consistently recorded.

****Business Strategy**:** Promote Classes in March, April to balance demand.

****Capacity Planning**:** Prepare Play Area/Party Room for May, June rush.