Booking Data Analysis & Dashboard

Executive Summary

This report presents a comprehensive analysis of booking data, including data preprocessing, exploratory data analysis, customer segmentation, time-series forecasting, and interactive dashboard development.

1. Data Analysis Methodology

1.1 Data Preprocessing Steps

1. **Libraries Imported**
    - Data manipulation: `pandas`, `numpy`
    - Visualization: `seaborn`, `matplotlib`
    - Machine Learning: `sklearn`
    - Time-Series Analysis: `statsmodels`
    - Dashboard: `dash`, `plotly`
2. **Data Loading**
    - Dataset: `DataAnalyst_Assesment_Dataset.xlsx`
    - Method: `pandas.read_excel()`
3. **Initial Data Exploration**
    - Examined data structure using:
        - `df.info()` : Column types and missing values
        - `df.head()` : First 5 rows preview
        - `df.isnull().sum()` : Missing value count
4. **Missing Value Handling**
    - Numerical columns (`Price`, `Duration`):
        - Replaced with median values
    - Categorical columns (`Status`, `Facility`):
        - Replaced with mode
    - Instructor names:
        - Assigned `"Unknown"` to missing entries

5. **Data Cleaning and Formatting**
    - Date columns converted to `datetime`
    - Text field standardization:
        - Trimmed spaces
        - Applied proper case to `Booking Type`
        - Converted `Status` to uppercase
        - Converted `Customer Email` to lowercase
        - Removed non-numeric characters from `Customer Phone`

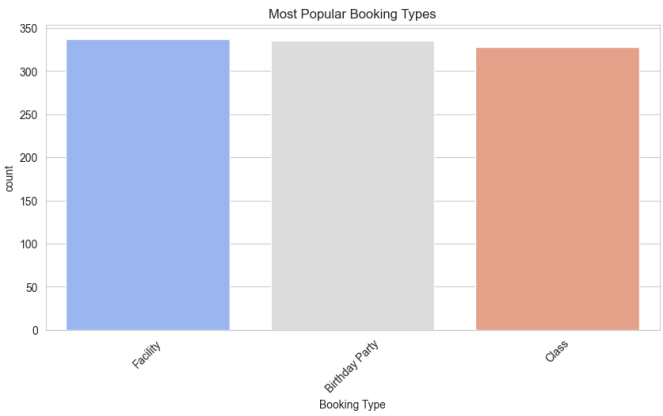
1.2 Feature Engineering

1. Extracted month from `Booking Date`
2. Converted `Time Slot` to `datetime.time`
3. Categorized Peak Hours (18:00 - 21:00)

2. Exploratory Data Analysis (EDA)

2.1 Visualization Insights

2.1.1 Most Popular Booking Type



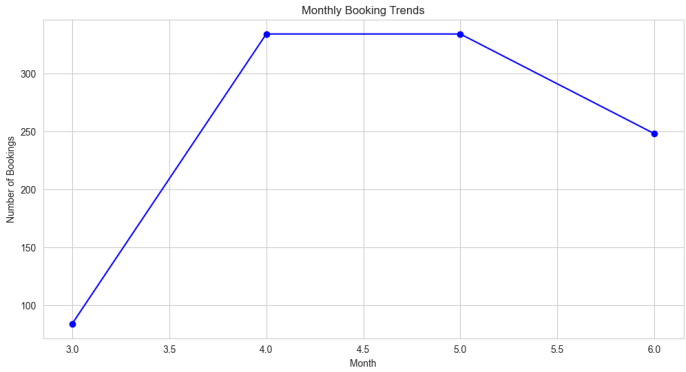
Key Observations:

- Facility Bookings: Highest number of bookings
- Birthday Party Bookings: Moderate booking volume
- Class Bookings: Lowest booking count

Recommendations:

- Investigate facility usage and popularity
- Enhance birthday party packages
- Develop strategies to increase class participation

2.1.2 Monthly Booking Trends



Key Observations:

- Sharp increase in bookings from month 3.0 to 4.0
- Peak booking period in months 4.0 and 5.0
- Decline in bookings after month 5.0

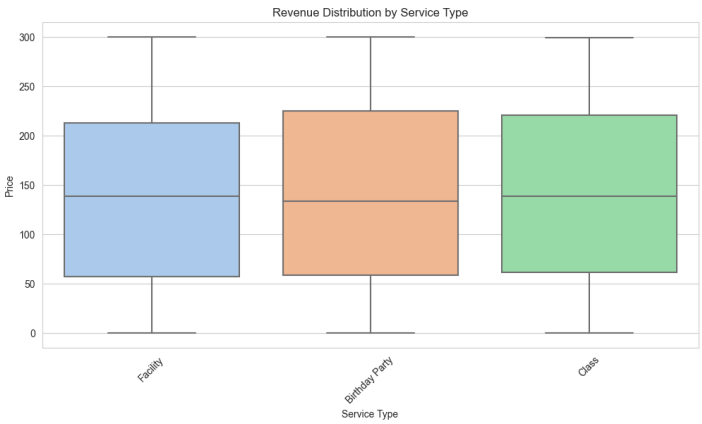
Recommendations:

- Investigate peak period drivers
- Develop off-season marketing strategies
- Optimize resource allocation

2.1.3 Revenue Distribution by Service Type

Key Observations:

- Consistent median price around 150 for all service types
- Significant price variability across service types
- Potential high-value outliers in facility bookings



Recommendations:

- Analyze price variability
- Consider tiered pricing packages
- Investigate high-revenue facility bookings

3. Advanced Analytics

3.1 Customer Segmentation

- Aggregated total spend and bookings per customer
- Applied K-Means clustering (k=3)

3.2 Time-Series Forecasting

- Aggregated monthly bookings
- Fitted ARIMA (2,1,2) model
- Generated 6-month booking forecast

4. Interactive Dashboard

4.1 Dash Web Application Features

- Dropdown for Service Type filtering
- Time-series revenue graph
- Revenue distribution histogram
- Revenue by Booking Type bar chart

- Dynamic updates using Dash callbacks

5. Outputs and Deliverables

- Cleaned dataset: `cleaned_booking_data.csv`
- Automated EDA report: `booking_report.html`
- Interactive dashboard application

6. Conclusion

The analysis provides comprehensive insights into booking patterns, revenue distribution, and customer behavior. Recommended strategies focus on optimizing service offerings, pricing, and marketing approaches.