Project Report: Data Analysis and Price Prediction Dashboard Development

Project Title: Data Analysis and Price Prediction for Business Insights
Author: Vikram Singh
Date: 27 March 2025
Institution/Organization: Omnify

Table of Contents

1. Executive Summary

This report presents an analysis of the provided dataset, focusing on price prediction using machine learning and the development of an interactive dashboard. The report details the data cleaning process, analytical approach, key findings, and dashboard components. The interactive dashboard and machine learning model were designed to facilitate data-driven decision-making and improve pricing strategies.

2. Introduction

Purpose of the Project

The objective of this project was to explore the dataset, identify trends and patterns, and develop a machine learning model to predict prices for bookings. Additionally, an interactive dashboard was created to provide insights into pricing and demand patterns.

Scope & Objectives

- Perform thorough data cleaning and preprocessing.

- Conduct exploratory data analysis (EDA) to uncover insights.

- Develop a machine learning model to predict booking prices.

- Build an interactive dashboard using Streamlit and Plotly.

- Provide actionable insights for pricing optimization.

Dataset Overview

The dataset contained booking records, including booking dates, service types, instructors, prices, durations, time slots, and facility usage.

3. Data Cleaning & Preprocessing

Initial Data Exploration

- Identified missing or inconsistent values.

- Detected duplicate records.

- Analyzed data distribution and types.

Discrepancies & Solutions

- **Missing Instructor Assignments:** Replaced "Not Assigned" values with "Unassigned."

- **Date Formatting Issues:** Converted all dates to a uniform format.

- **Outliers in Pricing & Duration:** Applied filtering to remove anomalies.

- **Categorical Inconsistencies:** Standardized categorical labels.

- **Encoding Categorical Variables:** Used label encoding for categorical columns such as Booking Type, Class Type, Time Slot, Facility, Theme, Service Type, and Instructor.

- **Feature Scaling:** Applied StandardScaler to normalize numerical features.

4. Data Analysis & Insights

Key Observations

- **Booking Type Distribution:** Majority of bookings were for individual sessions, while group sessions were less frequent.

- **Revenue Trends:** Revenue fluctuated with peak periods aligning with seasonal demand.

- **Instructor Workload:** A few instructors handled most bookings, indicating potential workload imbalances.

- **Facility Utilization:** Certain facilities were underutilized, suggesting an opportunity for better scheduling.

- **Time Slot Preference:** Morning slots were the most popular, while late evening slots had lower engagement.

- **Price Prediction Model:** Machine learning model (XGBoost) was trained to predict prices based on features such as service type, instructor, duration, and time slot.

5. Dashboard Overview

Dashboard Components

The interactive dashboard was designed to provide a visual representation of key metrics. Components include:

1. **Booking Type Distribution (Pie Chart)** – Highlights the proportion of different booking types.

2. **Revenue Breakdown (Bar Chart)** – Displays revenue generated per service type.

3. **Peak Booking Dates (Area Chart)** – Shows booking trends over time.

4. **Instructor Workload (Bar Chart)** – Compares bookings handled by each instructor.

5. **Revenue Trends (Line Chart)** – Tracks revenue fluctuations over time.

6. **Customer Booking Patterns (Bar Chart)** – Displays most popular booking time slots.

7. **Facility Usage (Pie Chart)** – Represents the distribution of facility utilization.

8. **Duration vs. Price Correlation (Scatter Plot)** – Analysis pricing relative to session duration.

9. **Service Popularity by Month (Heatmap)** – Identifies seasonality in service bookings.

10. **Price Prediction Tool** – Allows users to input booking parameters and predict the expected price using the trained model.

6. Challenges & Solutions

Challenges Faced

- **Data Quality Issues:** Missing or inconsistent values affected analysis accuracy.

- **Visualization Performance:** Large dataset slowed down dashboard responsiveness.

- **Categorical Standardization:** Variations in category labels complicated grouping.

- **Machine Learning Model Optimization:** Training the model with imbalanced data affected prediction accuracy.

Solutions Implemented

- **Data Imputation & Cleaning:** Addressed missing values systematically.

- **Optimized Data Processing:** Used efficient Pandas operations to improve performance.

- **Standardized Labels:** Unified categorical naming conventions for consistency.

- **Feature Engineering:** Selected important features and applied transformations for better model performance.

- **Hyperparameter Tuning:** Optimized XGBoost model parameters to improve accuracy and reduce overfitting.

7. Conclusion & Recommendations

Summary of Insights

- Booking trends indicate strong demand for morning sessions.

- Certain facilities and service types are underutilized, suggesting room for optimization.

- Revenue follows a cyclical pattern, which could help in forecasting and strategic planning.

- Instructor workload distribution needs to be balanced to improve operational efficiency.

- Machine learning model successfully predicts booking prices, assisting in pricing strategies.

Recommendations

- Adjust scheduling to maximize utilization of underused facilities.

- Introduce promotions or discounts for low-demand time slots.

- Optimize instructor assignments for an even workload distribution.

- Use historical revenue trends for demand forecasting.

- Deploy the price prediction model for real-time pricing optimization.

8. Final Dataset & Appendices

Final Dataset Used for Analysis

- The cleaned and processed dataset has been saved and utilized in the dashboard and machine learning model.