**Final Report**

**Project Description**

This project focuses on Sales and Revenue Analytics and Forecasting for Strategic Business Insights. It utilizes machine learning models to analyze historical sales and revenue data, predict monthly units sold, forecast future monthly revenue, and analyze the relationship between unit price and revenue to identify optimal price points.

**Dataset Source**

The dataset used in this project is titled "Group 5-Dataset for Final Project.csv" and contains detailed information about sales transactions, customer demographics, product details, costs, and revenues. The specific source or origin of this dataset is not explicitly stated within the provided notebook.

**Summary of Methodology and Findings**

**Methodology:**

* **Data Acquisition & Preparation:** Loaded the dataset, performed initial checks for missing values and data types, handled duplicates and outliers, and imputed missing values.
* **Exploratory Data Analysis (EDA):** Conducted descriptive statistics and various visualizations (bar plots, histograms, scatter plots, line plots, heatmaps) to understand data distributions, relationships between variables, and trends in sales and revenue.
* **Predictive Modelling (Objective 1):** Prepared monthly aggregated data with lag features to predict monthly units sold. Evaluated Linear Regression, Random Forest, and Decision Tree models.
* **Time Series Forecasting (Objective 2):** Aggregated data to monthly and weekly levels for time series analysis. Fitted ARIMA models and explored Exponential Smoothing and Prophet for forecasting monthly and weekly revenue.
* **Unit Price and Revenue Analysis (Objective 3):** Analysed the relationship between unit price and revenue by product category using visualizations and price binning.

**Findings:**

* Data cleaning successfully addressed missing values, duplicates, and outliers.
* EDA revealed key insights into product popularity, revenue drivers (Bikes are highest revenue despite lower volume), geographical performance (US is dominant), customer demographics, and historical revenue trends (upward trend with recent sharp decline).
* Predictive models for monthly units sold did not meet the target R² of 0.85, primarily due to limited historical data length.
* Time series forecasts using ARIMA and Prophet were generated, but their reliability is limited by the short historical data. ARIMA forecasts showed a flat trend after the data decline, while Prophet showed a more optimistic increasing trend.
* Analysis of unit price and revenue suggested a positive correlation between higher unit prices and higher revenue within product categories, particularly for Bikes. However, identifying truly optimal price points requires further analysis of demand elasticity and profitability.
* Limited historical data was a significant constraint throughout the modelling and forecasting stages.