**TEAM 5   
  
Details :-**

REPAKA SOMASEKHARA VENKATA DURGA S PRAKASH (LEAD)

KADALI SATYA SRUTHI

LITHIKASRAYA C

JALLIGAMPALA VENKATARAJA

MOGALIKOLLA RAMALAKSHMI

GANDHAM LEELA VENKATA SATYANARA

**PROBLEM STATEMENT**

**CUSTOMER LIFETIME VALUE (CLV) PREDICTION**

**Project Overview**

**Introduction**

This project aims to predict the Customer Lifetime Value (CLV) for a retail business using historical transaction data. Accurate CLV predictions will enable the business to optimize marketing strategies and improve customer retention.

**Objectives**

* To develop a predictive model for estimating CLV.
* To identify key factors influencing CLV.
* To provide actionable insights for marketing and customer engagement.

**Scope**

The project will focus on analyzing transaction data from the past three years and will include customer segmentation based on predicted CLV.

**Background**

**Definition of CLV**

Customer Lifetime Value (CLV) is the total revenue a business can expect from a customer throughout their relationship. It is a critical metric for understanding customer profitability. **Importance of CLV Prediction**

Predicting CLV helps businesses:

* Allocate marketing resources effectively.
* Identify high-value customers.
* Enhance customer retention strategies.

**Data Requirements**

**Data Sources**

* Internal transaction database.
* Customer profile database.

**Data Attributes**

* **Transaction Data**: Transaction ID, Date, Amount, Customer ID, Product ID.
* **Customer Data**: Customer ID, Age, Gender, Location, Signup Date.

**Methodology**

**Data Preprocessing**

* Data cleaning: Handling missing values and duplicates.
* Data transformation: Normalizing transaction amounts.

**Feature Engineering**

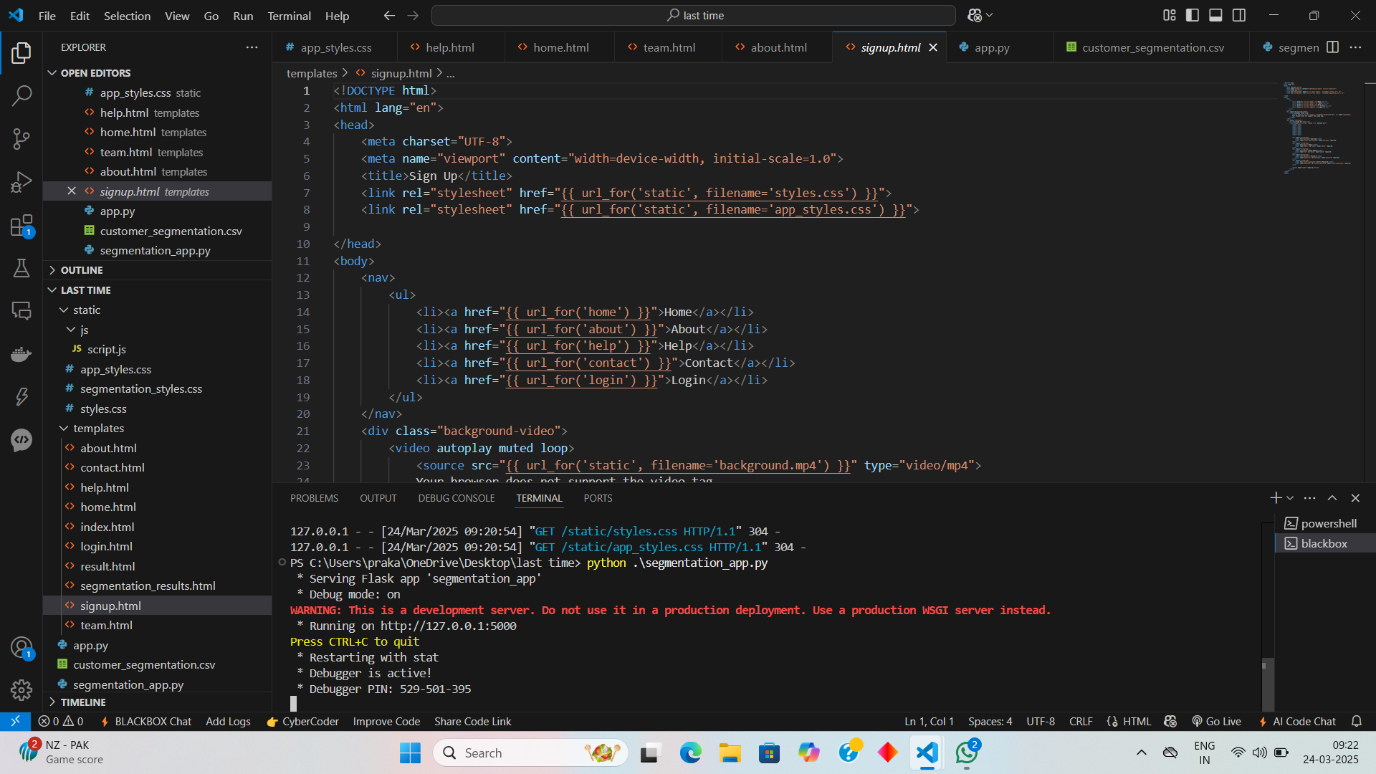
* Aggregate features: Total spend, purchase frequency, recency.
* Create new features: Customer tenure, average order value.

**Modeling Approaches**

* **Statistical Models**: Linear Regression, Logistic Regression.
* **Machine Learning Models**: Decision Trees, Random Forests, Gradient Boosting.
* **Probabilistic Models**: BG/NBD, Gamma-Gamma.

**Implementation**

**Tools and Technologies**

* Programming Language: Python
* Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn.
* Database: SQL for data extraction.
* 

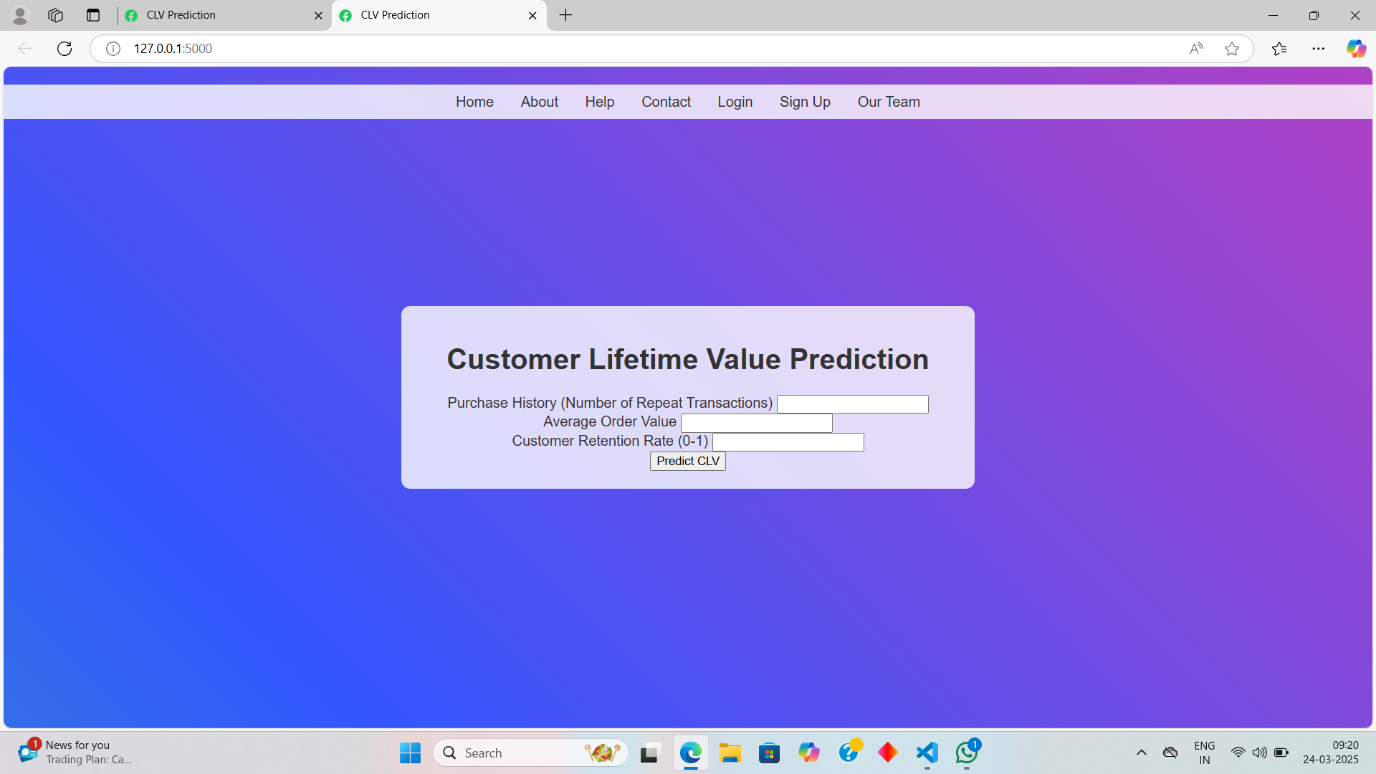
**Model Training and Evaluation**

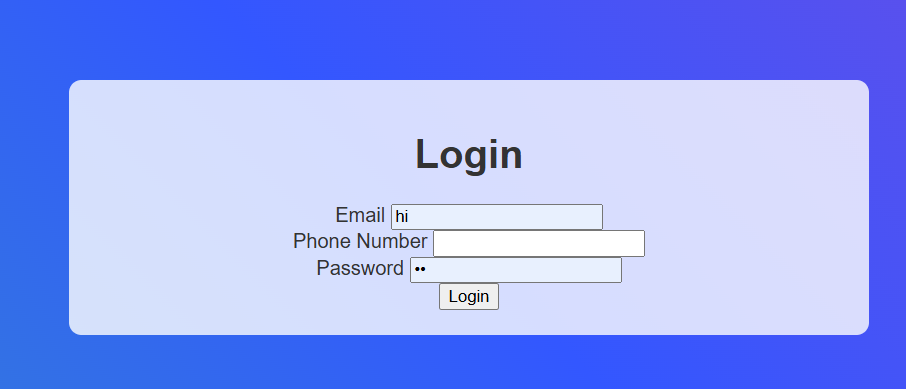
* Split data into training and testing sets (80/20).
* Train models and evaluate using metrics such as RMSE, MAE, and R-squared.

**Results**

**Model Performance Metrics**

* Summary of model performance (e.g., RMSE, R-squared values for each model).

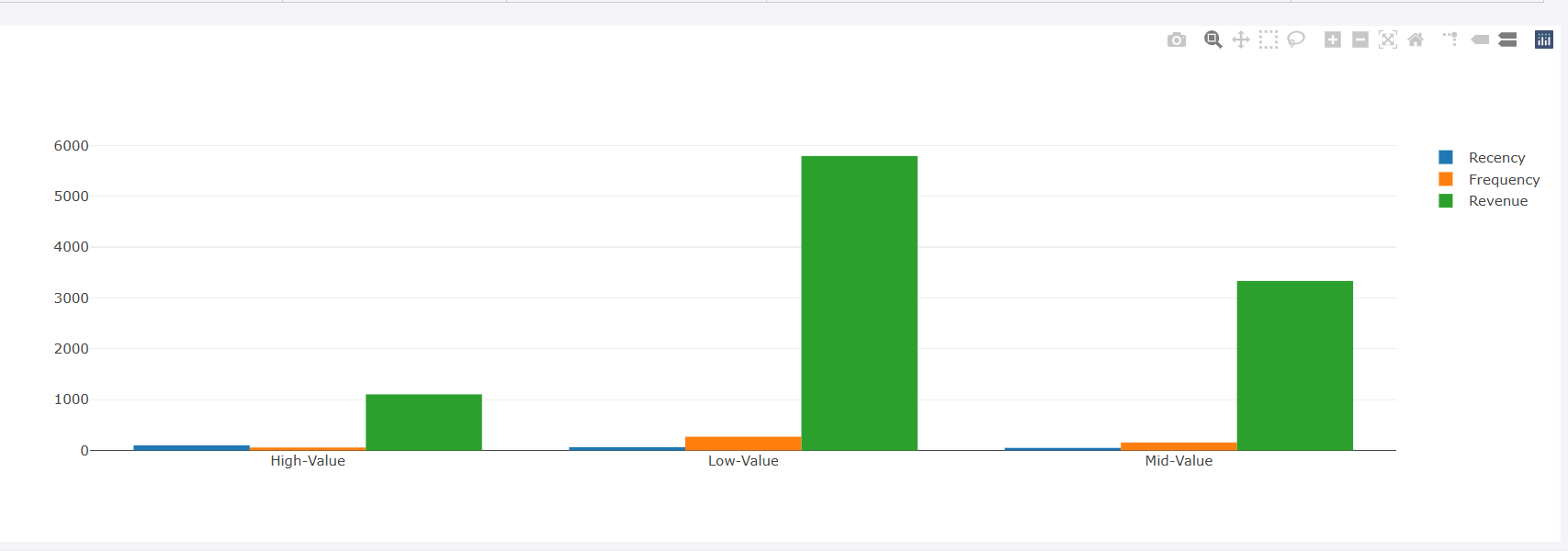


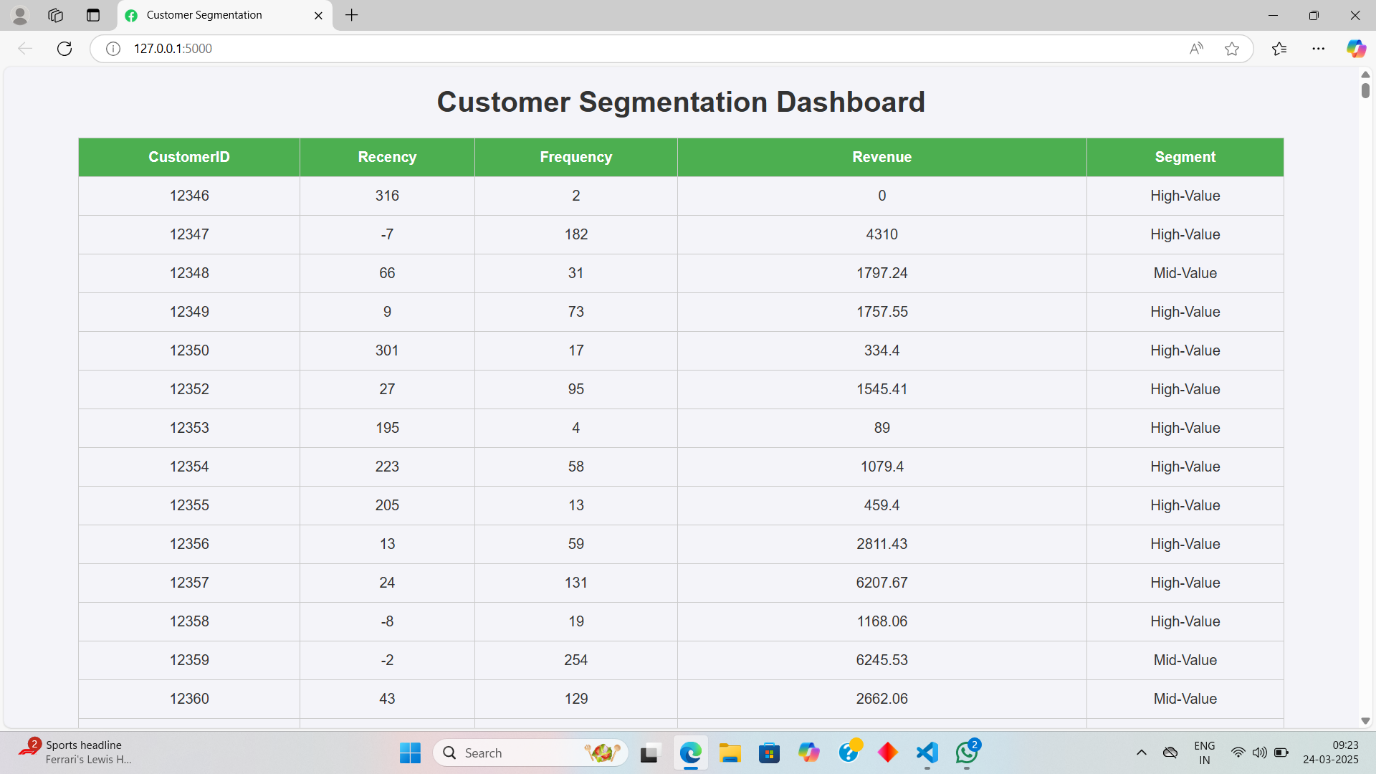
****

****

**Visualizations**

* Graphs comparing actual vs. predicted CLV.
* Feature importance plots.

****



**Insights and Recommendations**

* Key factors influencing CLV identified.
* Recommendations for targeted marketing strategies based on customer segments.

**Future Work**

* Explore advanced modelling techniques (e.g., deep learning).
* Implement real-time CLV prediction in marketing campaigns.

**Conclusion**

The Customer Lifetime Value (CLV) prediction project successfully developed a robust model to estimate the potential revenue from customers over their lifetime. By leveraging historical transaction data and employing various modelling techniques, we were able to identify key factors that influence customer value. The insights gained from this project can significantly enhance marketing strategies, allowing the business to focus on high-value customers and improve retention efforts. Future work will involve refining the model with new data and exploring advanced techniques to further enhance prediction accuracy.