Report on Customer Lifetime Value Prediction Model — Streamlit Dashboard

Abstract

This project presents a Customer Spend Prediction Dashboard built using Streamlit, designed to transform raw transactional data into meaningful insights. The application enables businesses to clean and preprocess large datasets, extract RFM (Recency, Frequency, Monetary) features, and apply either heuristic rules or machine learning models to forecast customer spending behavior. The dashboard not only supports segmentation of customers based on predicted vs. actual spend but also includes rich visualizations to help interpret customer purchase patterns. Furthermore, it allows exporting of results into Excel-friendly formats, making integration with business workflows seamless. By combining analytical power with simplicity of use, this system equips marketing teams with actionable intelligence to identify high-value customers, minimize revenue leakage, and optimize campaign targeting.

Introduction

In today's competitive and data-driven business environment, understanding customer purchase behavior is a vital component of sustainable growth. Customers do not contribute equally to revenue, and predicting future spending can help organizations efficiently allocate resources, design retention strategies, and plan targeted marketing campaigns.

This project addresses those needs by providing a Streamlit-based interactive dashboard for customer spend prediction. The tool processes raw transaction data uploaded by the user, computes RFM metrics, and predicts the likelihood of future spending using either predefined rules or statistical machine learning models.

With its interactive interface, the dashboard allows users to filter customers dynamically, visualize predicted vs. actual performance, and download refined customer segments in a ready-to-use CSV format. This makes the solution equally useful for data analysts, marketing teams, and business decision-makers who want reliable insights without heavy technical effort.

Tools Used

1. **Programming Language: Python**
2. **Framework: Streamlit — for building the interactive dashboard.**
3. **Libraries:**

- pandas, numpy → Data cleaning, preprocessing, and feature engineering.
- matplotlib, plotly → Data visualization, scatter plots, calibration curves.
- scikit-learn → Logistic Regression, Linear Regression, calibration, and model evaluation.
- Excel/CSV integration → Smooth upload of raw files and easy export of customer segments.

4. Environment: Interactive web app with dark theme customization for better usability.

Steps Involved in Building the Project

1. Data Upload & Ingestion – Users can upload CSV/Excel files or use demo data.
2. Data Cleaning – Standardization of column names, removal of invalid records, and computation of missing values.
3. Feature Engineering (RFM Analysis) – Derivation of Recency, Frequency, Monetary values, and spend within 90/365 days.
4. Prediction Engine – Application of both rule-based heuristics and machine learning models to calculate propensity scores and predict spending.
5. Segmentation & Filters – Customers segmented according to predicted probability and shortfall threshold, enabling targeted analysis.
6. Visualization – Scatter plots for distribution, calibration curves for probability reliability, and Brier Score for model accuracy evaluation.

Conclusion

The Customer Spend Prediction Dashboard represents a practical combination of data preparation, feature engineering, and predictive analytics within a single, user-friendly platform. By leveraging RFM modeling and machine learning, it allows businesses to discover patterns in customer behavior, measure gaps between predicted and actual spend, and generate downloadable customer segments for campaign execution.

Its dual capability — offering both rule-based logic for simplicity and machine learning models for advanced accuracy — ensures that the tool remains effective even with limited datasets. This flexibility makes it suitable for organizations at different levels of analytics maturity.