



Customer Lifetime Value (LTV) Prediction Report

Introduction

In today's competitive business environment, customer retention is as important as customer acquisition. One of the key indicators used to measure a customer's potential is **Customer Lifetime Value (LTV)**. This project aims to predict the LTV of each customer based on their transaction history to aid in **targeted marketing and personalized engagement strategies**.

Abstract

The goal of this project is to develop a machine learning model to estimate Customer Lifetime Value using historical purchase data. By leveraging Recency, Frequency, and Monetary-based features, we train a regression model to predict future revenue a customer might generate. The customers are then segmented based on the predicted LTV, enabling businesses to optimize marketing efforts toward high-value customers. This model supports data-driven decision-making in customer relationship management (CRM).

Tools Used

- **Python (Jupyter Notebook / Google Colab)** – for data preprocessing and model building
 - **Pandas & NumPy** – for data manipulation and numerical operations
 - **Matplotlib & Seaborn** – for visualizations
 - **XGBoost** – for regression modeling
 - **Sklearn (scikit-learn)** – for data splitting and evaluation
 - **Excel** – for preliminary inspection (optional)
-

Steps Involved in Building the Project

1. **Dataset Upload & Cleaning**
 - Uploaded customer transaction data in CSV format
 - Removed rows with null customer IDs and negative quantities
 - Converted invoice date to datetime format
 - Created a new column `TotalPrice = Quantity × UnitPrice`
2. **Feature Engineering**
 - Aggregated data per `CustomerID` to compute:
 - **Recency**: Days since last transaction
 - **Frequency**: Number of unique transactions
 - **AOV (Average Order Value)**: Average of `TotalPrice`
 - **Monetary**: Total amount spent (used as LTV label)
3. **Model Training**

- Selected features: Recency, Frequency, AOV
 - Target: LTV (Total Monetary Value)
 - Used **XGBoost Regressor** to train the model
 - Evaluated performance using **MAE** and **RMSE**
 - 4. **Prediction and Segmentation**
 - Predicted LTV for each customer
 - Segmented customers into **Low**, **Medium**, and **High** LTV groups using quantile-based binning
 - 5. **Result Export & Visualization**
 - Saved the final output as `final_ltv_predictions.csv`
 - Created a boxplot to visualize the distribution of predicted LTV across segments
-

Conclusion

This project successfully demonstrates how to predict Customer Lifetime Value using transactional data. With a predictive LTV model, businesses can focus their resources more effectively—by retaining high-value customers, identifying at-risk ones, and tailoring promotions accordingly. The segmentation strategy enhances personalization in customer engagement, ultimately improving revenue and loyalty.