Customer Lifetime Value (CLV) Prediction Project Report

---

1. Project Overview

Objective: To predict Customer Lifetime Value (CLV) using historical transaction data and provide business insights through RFM analysis and data visualizations.

Dataset Used: Online Retail Dataset (CSV format)

Tools & Technologies:

- Python (Jupyter Notebook)
- Pandas, NumPy, Scikit-learn, XGBoost
- Tableau
- Excel (for basic segment logic if needed)

---

2. Data Preprocessing

- Removed rows with missing `CustomerID` or `Description`
- Created `TotalPrice` column = Quantity * UnitPrice
- Converted `InvoiceDate` to datetime

---

3. RFM Analysis

RFM = Recency, Frequency, Monetary

- **Recency**: Days since last purchase
- **Frequency**: Number of purchases
- **Monetary**: Total amount spent

Used quantile-based scoring (1 to 4) for each metric

- RFM_Score = RecencyScore + FrequencyScore + MonetaryScore

---

4. Customer Segmentation

Created customer segments based on RFM_Score:

- **Best**: RFM_Score ≥ 9
- **Loyal**: $6 \leq$ RFM_Score < 9
- **Promising**: $4 \leq$ RFM_Score < 6

- **At Risk**: RFM_Score < 4

Saved as `Segment` column in dataset.

---

5. Predicting CLV Using Machine Learning

Target Variable: Monetary Value (used as proxy for CLV)

Features Used:

- Recency
- Frequency
- RFM_Score
- Segment

Model Used:

- XGBoost Regressor
- Train-test split (80/20)
- Evaluation metric: R^2 Score

Model Performance:

- R^2 Score (Baseline): \~0.58
- R^2 Score (XGBoost): \~0.74

---

6. Feature Importance

Used XGBoost's built-in feature importance plot:

- Frequency and RFM_Score were top predictors of CLV
- Recency had moderate importance
- Segment also contributed to CLV variation

---

7. Tableau Dashboard

Created a 4-part dashboard using `rfm_clv_tableau.csv`:

A. Bar Chart: Top 10 customers by predicted CLV\ **B. Scatter Plot:** Recency vs Frequency, sized & colored by CLV\ **C. Pie Chart:** Distribution of customers across segments\ **D. Histogram:** Distribution of Predicted CLV

Dashboard Title: **"Customer Lifetime Value (CLV) Insights & Segmentation"**

Exported dashboard as `.twbx` file.

8. Files in GitHub Repository

- `CLV_Project.ipynb`: Full ML pipeline and data preparation
- `rfm_clv_tableau.csv`: Cleaned dataset with predictions and segments
- `CLV_Dashboard.twbx`: Tableau dashboard
- `README.md`: Project summary and usage instructions
- `CLV_Project_Report.pdf`: This detailed report

---

9. Conclusion

This project demonstrates how to combine RFM-based segmentation and machine learning to predict Customer Lifetime Value. The resulting visual dashboard helps businesses identify top customers, understand spending behavior, and drive targeted marketing strategies.

Prepared by: Sagar kanojia Date: 17/7/2025