Customer Lifetime Value (LTV) Prediction Project Report

1. Objective

The primary objective of this project was to predict the Lifetime Value (LTV) of customers based on historical purchasing behavior. Accurate LTV prediction enables businesses to:

- Segment customers based on their value.

- Allocate marketing budgets efficiently.

- Personalize retention strategies for high-value customers.

2. Problem Statement

Understanding customer value is crucial for sustainable business growth. However, manually identifying high-value customers can be time-consuming and inaccurate. Thus, we aim to automate LTV prediction using machine learning models.

3. Dataset Overview

We used a transactional dataset containing:

- Invoice ID

- Customer ID

- Invoice Date

- Quantity

- Unit Price

The dataset included thousands of transactions across different customers over time.

---

4. Tools & Technologies Used

- Python 3 (Pandas, Scikit-learn, XGBoost)

- Google Colab (cloud-based coding)

- Excel/CSV for output file management

- FPDF (for report generation)

---

5. Data Preprocessing Steps

- Removed transactions with missing Customer IDs.

- Filtered out canceled transactions (negative quantities or prices).

- Created a `TotalPrice = Quantity × Unit Price` field.

- Converted `InvoiceDate` to datetime format.

- Set a reference date to calculate recency for each customer.

---

6. Feature Engineering

We generated customer-level features:

- **Recency**: Number of days since the customer's last purchase.

- **Frequency**: Total number of unique purchases made.

- **Average Order Value (AOV)**: Average amount spent per order.

- **Total Revenue**: Sum of all purchases made by the customer.

- **Target Variable**: Customer Lifetime Value (CLV).

---

7. Modeling Approach

- Chose **XGBoost Regressor** for predicting LTV based on Recency, Frequency, and AOV.

- Split the dataset into a training set and a testing set.

- Evaluated model performance using:

  - Mean Absolute Error (MAE)

  - Root Mean Squared Error (RMSE)

- After prediction, segmented customers into 4 groups based on LTV:

  - Low Value

  - Mid Value

  - High Value

  - VIP

---

8. Results & Observations

- LTV predictions were successfully generated.

- Clear customer segments emerged from the predicted LTV values.

- The final LTV predictions were saved in the file `ltv_predictions.csv`.

9. Business Value

The LTV prediction model allows marketing and sales teams to:

- Focus promotions on high-value customers.

- Design loyalty programs for top customers.

- Identify low-value customers and strategize reactivation offers.

- Reduce churn by targeting at-risk customers proactively.

10. Future Improvements

- Include demographic/customer profile data for better predictions.

- Explore other machine learning models like LightGBM and CatBoost.

- Implement real-time LTV predictions integrated into CRM systems.

- Build time-series forecasting models to predict dynamic LTV over time.