Customer Lifetime Value Prediction Project

Introduction:

Customer Lifetime Value (CLV) estimates the total revenue a business can expect from a customer throughout their relationship. Predicting CLV allows businesses to identify high-value customers, personalize marketing, and increase profitability by focusing retention strategies on the most valuable segments.

Abstract:

In this project, transactional data from an online retail dataset was analyzed to predict customer lifetime value. Key RFM (Recency, Frequency, Monetary) features were engineered, and a Random Forest regression model was trained to estimate CLV for each customer. Customers were then segmented into four groups based on their predicted CLV, enabling targeted marketing strategies. The final results were saved as a CSV file containing each customer's predicted CLV and assigned segment.

Tools Used:

- Python Libraries: Pandas, NumPy, Scikit-learn, Matplotlib
- Environment: Jupyter Notebook
- Data Source: Online Retail dataset (Excel format)
- Deliverables: Python notebook, final predictions CSV, project report PDF

Steps Involved in Building the Project:

* Loaded and explored transactional data from Excel.

* Cleaned data by removing missing customer IDs and calculating total purchase amounts.

* Engineered RFM features:

- Recency: Days since last purchase.
- Frequency: Number of transactions.
- Monetary/AOV: Average Order Value.

* Built and trained a Random Forest regression model to predict CLV.

* Evaluated model performance using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

* Predicted CLV for each customer and segmented them into Low, Mid-Low, Mid-High, and High categories using quantile-based binning.

* Saved final predictions and segments as final_clv_predictions.csv.

Conclusion:

This project successfully demonstrated how to engineer RFM features and build a predictive model for customer lifetime value. By segmenting customers into value-based groups, businesses can focus marketing and retention efforts on the most profitable customers. The approach provides a practical framework for data-driven customer relationship management and can be extended to any retail dataset.