

# Customer Lifetime Value Prediction

---

## Introduction:

Customer Lifetime Value (CLTV) is a critical metric used to estimate the total monetary value a customer is expected to contribute to a business over their relationship period. This project focuses on predicting CLTV using transactional purchase history to help businesses identify and retain high-value customers through targeted strategies.

## Abstract:

Customer Lifetime Value (CLTV) is a critical metric used to estimate the total monetary value a customer is expected to contribute to a business over their relationship period. This project focuses on predicting CLTV using transactional purchase history to help businesses identify and retain high-value customers through targeted strategies.

## Tools Used:

- **Python:** Primary programming language
- **Pandas & NumPy:** Data wrangling and feature engineering
- **Scikit-learn (Random Forest):** Model training and evaluation
- **Joblib:** Model saving and loading
- **Matplotlib & Seaborn:** Data visualization
- **Jupyter Notebook:** Development environment

## Steps Involved in Building the Project

### 1. Data Cleaning & Preparation

The dataset contained 541,909 rows with 8 columns.

Removed missing values, refunds, and transactions with zero or negative amounts.

## 2. Feature Engineering

**Recency:** Days since last transaction.

**Frequency:** Number of purchases.

**AOV:** Total spent divided by frequency.

**Monetary:** Total amount spent (used as the LTV target).

## 3. Model Building

Used `RandomForestRegressor(random_state=42)`.

Split dataset into training and testing sets (80-20).

Target variable: **Monetary** value (proxy for LTV).

## 4. Model Evaluation

**Mean Absolute Error (MAE):** 191.60

**Root Mean Squared Error (RMSE):** 2554.26

The model generalizes well for the skewed nature of customer value distribution.

## 5. Segmentation & Export

Customers were segmented into **Low**, **Mid-Low**, **Mid-High**, and **High** based on predicted LTV.

Results saved as CSV and model saved as `customer_ltv_model.pkl`.

## Conclusion:

This project demonstrates a scalable pipeline to predict customer lifetime value from transactional data. It enables businesses to:

- Identify top customers early.
- Prioritize high-LTV customers for retention.
- Personalize marketing campaigns based on predicted customer value.

The methodology is extendable to include churn prediction, time-based LTV, and advanced clustering for more sophisticated insights.