

# Customer Lifetime Value Prediction

---

## Introduction

Customer Lifetime Value (LTV) prediction is a key strategy in modern retail for identifying and retaining high-value customers. By analyzing transactional data, businesses can optimize marketing efforts and enhance profitability. This project leverages machine learning to predict LTV and segment customers based on their potential value.

---

## Abstract

We analyzed historical e-commerce data to develop a model that predicts LTV using behavioral features—**Recency**, **Frequency**, and **Average Order Value (AOV)**. A Random Forest Regressor was trained to estimate LTV, followed by customer segmentation into **Low**, **Mid-Low**, **Mid-High**, and **High** categories. Results were visualized, and the model was saved for future use.

---

## Tools Used

- **Python, Pandas, NumPy** – Data handling
  - **Matplotlib, Seaborn** – Visualization
  - **Scikit-learn** – Machine learning
  - **Joblib** – Model saving
  - **Jupyter/Colab** – Development environment
- 

## Datasets Used

### 1. Original Dataset – Online Retail .xlsx

- Raw transactional data including:
  - InvoiceNo, CustomerID, InvoiceDate, Quantity, UnitPrice
- Used for preprocessing and feature extraction

### 2. Processed Dataset – Predicted\_LTV\_Segments .csv

- Contains customer-level data with:
    - Recency, Frequency, Monetary, AOV
    - LTV (total spend), Predicted\_LTV
    - Segment (based on LTV quartiles)
- 

## Project Workflow

## 1. Preprocessing

- Removed null CustomerID and canceled orders (InvoiceNo starting with "C")
- Created TotalAmount as Quantity × UnitPrice
- Converted InvoiceDate to datetime format

## 2. Feature Engineering

- Calculated:
  - Recency: Days since last purchase
  - Frequency: Purchase count
  - Monetary: Total spending
  - AOV: Monetary / Frequency
- Defined LTV as Monetary

## 3. Modeling

- Features: Recency, Frequency, AOV
- Model: **Random Forest Regressor**
- Evaluation: MAE, RMSE

## 4. Prediction & Segmentation

- Predicted LTV for all customers
- Segmented into:
  - **Low, Mid-Low, Mid-High, High** (using pd.qcut)

## 5. Visualization

- Histogram of predicted LTV
- Bar chart of customer segments

## 6. Export

- Saved predictions and segments to CSV
- Serialized model with joblib

---

## Conclusion

This project demonstrates how behavioral data can effectively predict Customer Lifetime Value. By segmenting customers based on predicted LTV, businesses can tailor strategies to maximize retention and revenue. Future work could include integrating time-based models, campaign response data, or customer demographics for enhanced prediction accuracy.