

# Customer Lifetime Value (LTV) Prediction Project

---

## 1. Objective

The goal of this project is to predict the Lifetime Value (LTV) of customers based on their purchase behavior. This will help the business target high-value customers more effectively through personalized marketing strategies.

## 2. Tools Used

- Python (Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, XGBoost, Joblib)
- Excel (for input datasets)

## 3. Dataset Description

Two datasets are used in this project:

1. transactions.csv – contains customer\_id, order\_date, and order\_value
2. customers.csv – contains customer\_id and registration\_date

## 4. Data Preprocessing

Dates were parsed into datetime format. Then, the transactions and customers datasets were merged on customer\_id. This created a unified dataset for feature extraction.

## 5. Feature Engineering

The following features were engineered for each customer:

- Frequency – Number of orders
- Recency – Days since last order
- AOV (Average Order Value) – Average of order\_value
- Total Value – Sum of order\_value (used as the target variable)

## 6. Modeling

The dataset was split into training and testing sets using an 80-20 split. An XGBoost Regressor was used to predict the total LTV of customers. Model performance was evaluated using MAE and RMSE.

## 7. Evaluation

- MAE (Mean Absolute Error)
- RMSE (Root Mean Squared Error)

## 8. Customer Segmentation

Based on predicted LTV values, customers were segmented into four categories:

- Low
- Medium
- High
- Very High

These segments help in devising targeted marketing strategies.

## 9. Deliverables

- Python Notebook for modeling
- Trained XGBoost model file (ltv\_xgb\_model.pkl)
- Final CSV with predicted LTV and customer segments (final\_ltv\_predictions.csv)
- Visualizations: Histogram of predicted LTV, Boxplot of LTV by segment