# **📄 Project Report: Customer Lifetime Value (CLTV) Prediction Model**

## **1. Project Title**

**Customer Lifetime Value (CLTV) Prediction using Machine Learning**

## **2. Objective**

To build a machine learning model that predicts the **Customer Lifetime Value (CLTV)** using past purchase behavior such as **Recency**, **Frequency**, **Average Order Value (AOV)**, and **Tenure**.  
 This enables targeted marketing strategies and customer segmentation for retention campaigns.

## **3. Dataset Details**

* **Dataset Name**: UK Online Retail Dataset (Kaggle)
* **File Name**: data.csv
* **Size**: ~500,000 transactions
* **Fields Used**:  
  + InvoiceNo (Order ID)
  + CustomerID
  + InvoiceDate
  + Quantity
  + UnitPrice
  + Country

## **4. Tools & Technologies**

* Python (Google Colab)
* Libraries: pandas, NumPy, scikit-learn, XGBoost, matplotlib, seaborn
* CSV File Handling & Data Preprocessing
* Machine Learning Regression Modeling

## **5. Methodology**

### **Step 1: Data Cleaning**

* Removed rows with missing CustomerID, InvoiceDate, Quantity, or UnitPrice.
* Filtered out cancelled orders (negative quantities or amounts).

### **Step 2: Feature Engineering**

Created customer-level metrics:

| **Feature** | **Description** |
| --- | --- |
| Recency | Days since the customer's last purchase |
| Frequency | Total number of unique purchases (Invoice count) |
| Tenure | Days between first and last purchase |
| AOV | Average spend per order |
| LTV (Target) | Total money spent by the customer (sum of orders) |

### **Step 3: Model Building**

* Split the dataset into training (80%) and testing (20%).
* Used **XGBoost Regressor** with:  
  + n\_estimators = 100
  + learning\_rate = 0.1
  + max\_depth = 4

### **Step 4: Model Evaluation**

* **MAE (Mean Absolute Error)**: ±£143.XX
* **RMSE (Root Mean Squared Error)**: ±£YYY.XX  
   *(Values to be filled based on your model output)*

### **Step 5: Customer Segmentation**

* Predicted LTV scores for all customers.
* Segmented customers into 4 categories using quantiles:  
  + Low
  + Medium
  + High
  + Very High

## **6. Visualizations Created**

| **Visualization** | **Purpose** |
| --- | --- |
| LTV Distribution Histogram | Understand spread of actual LTVs |
| Actual vs. Predicted LTV Scatter Plot | Evaluate model prediction accuracy |
| Feature Importance (XGBoost) | Identify key behavioral drivers of LTV |
| Customer Segments Pie Chart | Visualize customer distribution by segment |

## **7. Key Outcomes**

* Built a predictive model with **MAE of ±£143** using XGBoost.
* Identified **Frequency** and **AOV** as top predictors of LTV.
* Successfully segmented customers into actionable tiers for marketing.
* Exported a CSV file with **CustomerID, Predicted LTV, and Segment**.
* Visuals created for easy reporting and presentation.

## **8. Conclusion**

This project demonstrates how machine learning can effectively predict Customer Lifetime Value based on transactional data.  
 Such models empower businesses to focus on high-value customers and improve retention strategies through targeted offers and personalized marketing campaigns.

## **9. Future Scope**

* Add **demographic features** if available.
* Incorporate **time-based LTV prediction (e.g., next 6 months)**.
* Build a **dashboard (Tableau/Power BI)** for live customer insights.
* Automate model retraining with new data.

## **10. Files Delivered**

* Python Notebook (Google Colab)
* CLTV Predictions CSV
* Visualizations (PNG images)
* Model File (xgb\_ltv\_model.pkl) *(Optional)*