Customer Lifetime Value (LTV) Prediction Report

Objective

The objective of this project is to predict the Customer Lifetime Value (LTV) of customers based on their transaction history. The LTV predictions will help in identifying customer segments and enabling targeted marketing strategies.

Tools and Libraries

- **Python**: For data processing, machine learning, and visualizations.
- **Libraries**:
  - pandas for data manipulation.
  - numpy for numerical operations.
  - matplotlib and seaborn for visualizations.
  - scikit-learn for model training and evaluation.
  - xgboost for training the predictive model.
  - joblib for saving the trained model.
  - openpyxl or xlrd for reading Excel files.

Dataset Overview

1. Customer Data (Excel Sheet: customers)

The customer dataset contains information about each customer, including:

- customer_id: Unique identifier for each customer.
- name: Name of the customer.
- email: Email address of the customer.
- signup_date: Date when the customer signed up.

2. Transaction Data (Excel Sheet: transactions)

The transaction dataset contains details about each transaction, including:

- order_id: Unique identifier for each order.
- customer_id: Unique identifier for each customer (linked with the customers dataset).
- order_date: Date when the transaction occurred.
- order_amount: Monetary value of the transaction.

Methodology

Step 1: Data Preprocessing

- The transaction data was loaded from the Excel file.
- The order_date column was converted to a datetime format using pd.to_datetime().
- Any rows with invalid dates were removed.
- A **snapshot date** was set as the maximum order_date + 1 day to calculate **recency**.

Step 2: Feature Engineering

To predict LTV, the following features were created for each customer:

- **Recency**: The number of days since the customer's most recent transaction.
- **Tenure**: The number of days between the first and most recent transaction.
- **Frequency**: The total number of transactions made by the customer.
- **Monetary**: The total monetary value of all the customer's transactions.
- **Average Order Value (AOV)**: The average amount spent per order by the customer.

Step 3: Model Training and Evaluation

- **Model**: XGBoost Regressor was used to predict the **Monetary** value, which acts as the predicted **LTV**.
- The dataset was split into training and testing sets using train_test_split.
- The model was trained on features: recency, tenure, frequency, and AOV.
- Model evaluation metrics:
    - **Mean Absolute Error (MAE)**: Evaluates the average magnitude of the errors in a set of predictions.
    - **Root Mean Squared Error (RMSE)**: Measures the average magnitude of the error.

    After training, the model achieved the following performance:

    - **MAE**: X.XX (Replace with actual value)
    - **RMSE**: X.XX (Replace with actual value)

Step 4: Customer Segmentation

Customers were segmented into three groups based on their predicted LTV values:

- **Low**
- **Medium**
- **High**

This segmentation helps in tailoring marketing strategies based on customer value.

Deliverables

1. Python Notebook

- The Jupyter notebook, LTV_Prediction_Model.ipynb, contains all the code steps for data preprocessing, feature engineering, model training, evaluation, and visualizations.

2. Trained Model

- The trained model, ltv_model.pkl, is saved using joblib and can be used for future predictions.

3. LTV Prediction Output

- The final LTV predictions, along with customer segmentation, were saved in a CSV/Excel file (ltv_predictions.xlsx).

4. Visualizations

- **LTV Distribution**: A histogram showcasing the distribution of predicted LTV values.
- **Feature Importance**: A bar chart displaying the importance of each feature in predicting LTV.
- **Customer Segmentation**: A scatter plot of customer frequency vs. AOV, colored by their LTV segment.

Conclusion

The predictive model successfully estimated Customer Lifetime Value (LTV) using historical transaction data. With this model:

- **High-value customers** can be identified for targeted marketing.
- **Low-value customers** can be nurtured or analyzed for retention strategies.
- The model can be used in dynamic customer segmentation to optimize resource allocation and marketing campaigns.

The predictions and insights generated can help businesses improve customer acquisition, retention, and overall marketing effectiveness.