Customer Lifetime Value (LTV) Prediction Project Report

Tools: Python (Pandas, Scikit-learn, XGBoost), Power BI, Excel

Author: SOMASEKHAR KADIYAM

Overview:

This project focuses on predicting the Customer Lifetime Value (LTV) for an e-commerce dataset. The objective is to identify high-value customers, understand the key behavioral drivers, and support targeted marketing strategies.

The project includes:

- End-to-end data cleaning & feature engineering
- Building an LTV prediction model using XGBoost
- Customer segmentation (High / Medium / Low LTV)
- A 2-page Power BI dashboard for business & model insights.
- Recommendations to improve customer retention and revenue.

Dataset Summary:

Files Used

- Customers.csv – Customer demographic + signup data
- Transactions.csv – Historical purchase data
- ltv_features_raw.csv – Engineered features (Frequency, Recency, Monetary)
- ltv_predictions.csv – Final model predictions
- feature_importance.csv – Model-derived feature ranking

Key Metrics Computed

- Frequency: Count of purchases
- Recency (days): Days since last purchase
- Monetary Avg: Average spend per order
- Orders_90d: Recent purchase behavior
- Predicted LTV: 12-month value forecast

Data Preparation & Feature Engineering

Data Cleaning

- Handled missing values
- Standardized date formats (YYYY-MM-DD)
- Removed duplicates
- Corrected inconsistent customer IDs

Merging & Transformation

- Merged customers & transactions using customer_id
- Snapshot date created for recency calculation
- Generated ltv_features_raw.csv

Feature Engineering (RFM Model)

- R (Recency): days since last order
- F (Frequency): number of orders
- M (Monetary): average order value

Output Generated:

- Clean feature dataset ready for ML modeling

Model Training & Evaluation

Model Used

- XGBoost Regressor for 12-month LTV prediction

Process

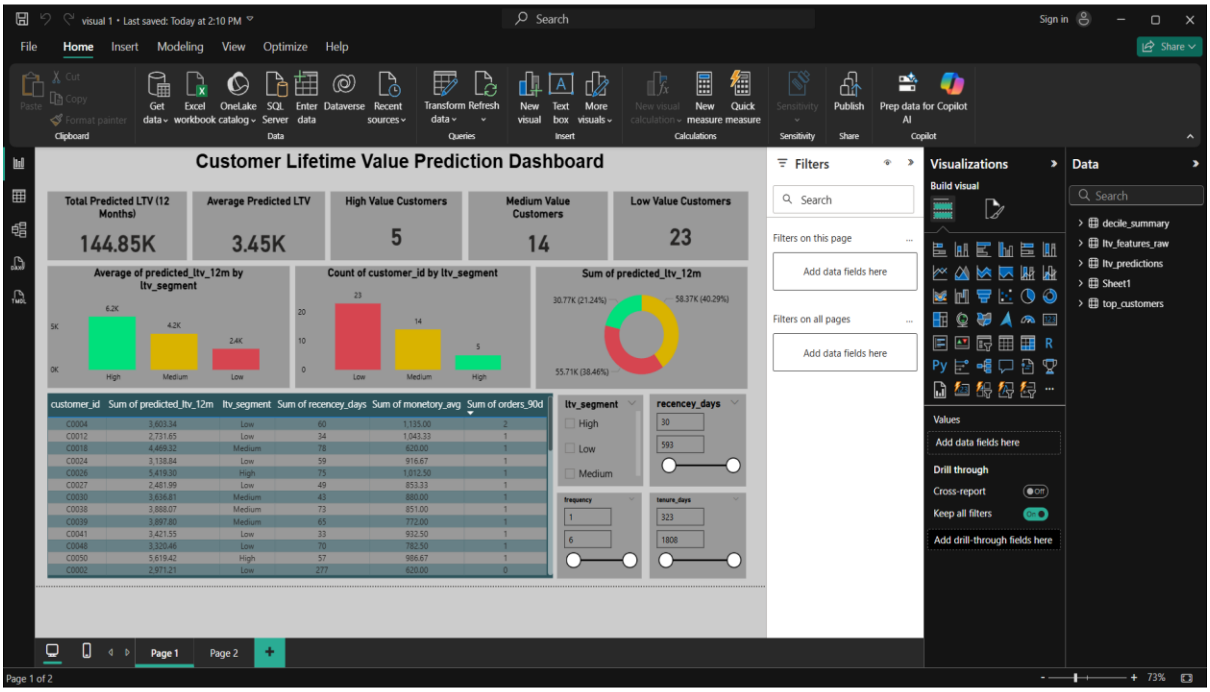
- Train/Test split
- Normalized numerical variables
- Grid search
- Model evaluation with MAE & RMSE

Results

- Model produced stable predictions for small dataset
- Generated ltv_predictions.csv
- Segmented customers into High / Medium / Low LTV

Why XGBoost?

- Handles non-linear relationships
- Works well with small-to-medium datasets
- Excellent feature importance interpretability

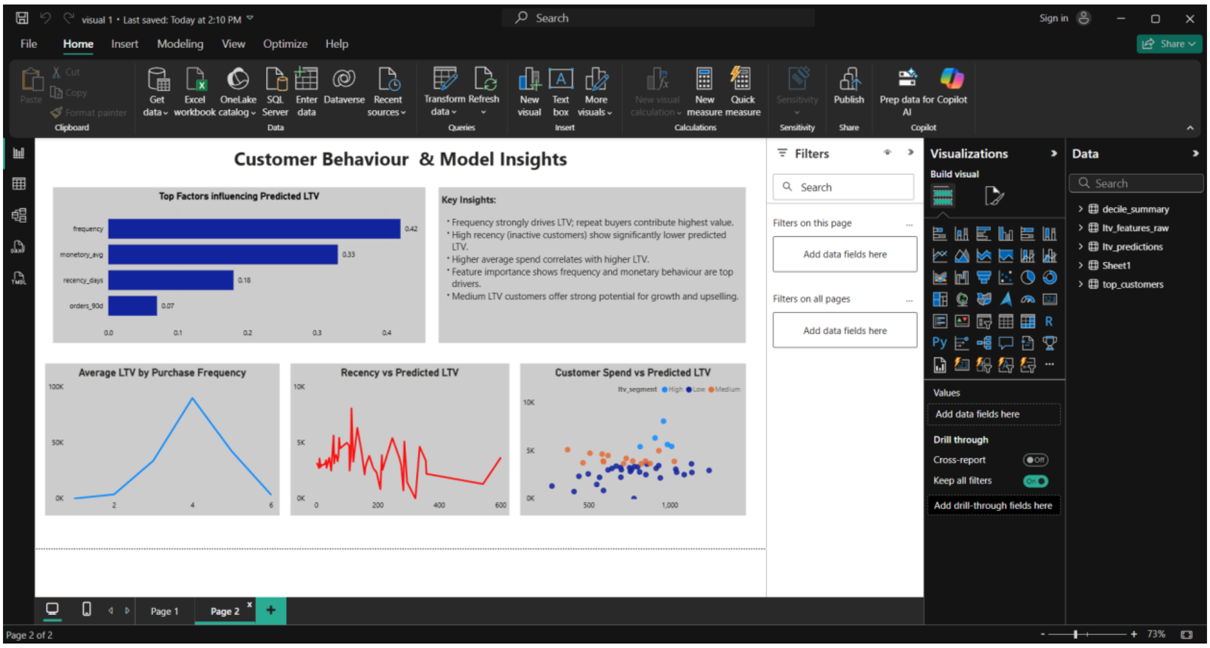


Page 1 : Business Overview

Key Insights

- Total predicted LTV
- Breakdown of High / Medium / Low value customers
- Segment contribution via donu chart
- Top predicted LTV customers
- Slicers to filter by behavior

This page provides a business -friendly view of customer value distribution.



Page 2: Behavior & Model Insights

Insights from Page 2

- Feature importance : Frequency and monetary value drive LTV the most
- LTV vs Frequency : Most frequent buyers have significantly higher LTV
- LTV vs Recency : Inactive customers show lower LTV
- LTV vs Monetary : High spenders contribute most to long-term value

This page explains WHY the model predicts certain customers as high or low value.

Final Business Recommendations

Based on the model and behavioral analysis:

1. Focus on High-LTV Customers
   - Retention campaigns
   - Loyalty rewards
   - Personalized offers

2. Upsell to Medium-LTV Customers
   - Combo offers
   - Product bundling
   - Targeted email campaigns
3. Re-engage Low-Recency Customers
   - Win-back discounts
   - SMS remainders
   - Time-limited promotions
4. Increase Average Order Value
   - Add-on recommendations
   - Free shipping above threshold
   - Multi-buy discounts
5. Improve Customer Retention Strategy
   - Track RFM metrics monthly
   - Automate churn flags
   - Develop custom campaigns by LTV segment