**Business Requirements Document**

**Project Title**: Customer Lifetime Value Prediction Across Active & Churned Customers  
**Folder**: data-science/clv-prediction/requirements/  
**Version**: v2 – Updated Scope

**1. Objective**

Develop a machine learning model to estimate Customer Lifetime Value (CLV) across both churned and active customers.  
The goal is to identify high-potential customers and their profile to help find where we likely have opportunity for upsells and high value new logo opportunities.

**2. Business Questions**

* Which features or behaviors are most predictive of high CLV?
* Can we create a profile of high-value customers to focus sales, marketing, and customer success efforts?
* How does CLV vary by region, product, and usage behavior?
* What levers might increase CLV in at-risk segments?

**3. Stakeholders**

| **Stakeholder** | **Role** |
| --- | --- |
| Customer Success | Proactive engagement |
| Sales | Account targeting & expansion |
| Marketing | Campaign targeting |
| FP&A | Revenue forecasting |

**4. Data Sources**

* Synthetic customer data (customer\_id, region, product\_family, etc.)
* Behavioral data (logins, feature usage, tenure)
* ACV and email engagement scores
* Historical churn status

**5. Modeling Approach**

* Supervised regression model (e.g., XGBoostRegressor or LightGBM)
* Target: clv (calculated for all customers based on tenure × ACV)
* Feature engineering: Encode categorical variables, scale numerics
* Model interpretation: SHAP for global/local driver insights
* Optional segmentation by region, product, etc.

**6. Deliverables**

* Trained model that estimates CLV for any customer
* Ranked list of customers by predicted CLV
* SHAP-based driver insights
* Visualizations: distribution of predicted CLV, SHAP plots, region/product breakdowns

**7. Success Criteria**

* High R² or low RMSE/MAE on test data
* Clear, interpretable drivers of CLV
* Actionable customer lists for GTM teams
* Reproducible notebook + clean GitHub repo