CUSTOMER LIFETIME VALUE (CLV) FORECASTING FOR A SAAS SUBSCRIPTION BUSINESS

1. Objective

The aim was to build a data-driven model to predict Customer Lifetime Value (CLV) for an online subscription business. The goal was to help the company identify high-value customers, optimize customer acquisition spending, and improve retention strategies by understanding expected future revenue per customer.

2. Client and Use Case

Client Type: A B2C SaaS company offering monthly and annual subscriptions for productivity software in the U.K.

Use Case: The client wanted to replace their static, rule-based LTV estimates with a more dynamic, behavior-based forecasting model. The model needed to accommodate both monthly and yearly billing cycles, factor in churn likelihood, and forecast future revenue per user over a 12-month horizon.

3. Dataset Overview

Source: Internal customer database linked to Stripe subscription logs and website usage tracking

Time Period: Jan 2021 – Mar 2023

Volume: 88,000 customer records 300,000+ subscription events

Key Variables:

- Customer ID
- Subscription start and end date
- Billing cycle (monthly/yearly)
- MRR (Monthly Recurring Revenue)
- Number of support tickets
- Web app engagement score (scaled 0–100)
- Churn status (1 = churned, 0 = active)

4. Methodology

4.1 Data Preparation

- Used readtable and datetime functions in MATLAB to import and align transactional logs.
- Cleaned missing MRR data using conditional forward-fill logic.
- Segmented users by billing cycle and tenure band.

4.2 CLV Modeling Framework

1. Retention Probability Model (Logistic Regression):

- Modeled churn probability using logistic regression on variables like tenure, support interactions, and engagement score.
- Used MATLAB's fitglm with binomial distribution.

2. Revenue Forecasting (Time Series Forecasting):

- Applied ARIMA models on average monthly spend per segment using estimate and forecast functions.
- Incorporated customer-specific adjustments using weighted multipliers based on engagement deciles.

3. CLV Calculation:

$$CLV_i = \sum_{t=1}^{12} \left[P_{ret,i}(t) \times R_i(t)\right]$$

Where

- $P_{ret,i}(t)$: retention probability at time t
- $R_i(t)$: expected revenue at time t
- Discounting optional; ignored per client preference

5. Key Findings

- ~12% of customers accounted for ~52% of projected CLV
- Customers with >80 engagement score had 3x higher predicted CLV
- Yearly subscribers were 32% more likely to renew than monthly ones

- High-support-touch customers had lower retention probability

Model Metrics:

- Logistic regression AUC: 0.76
- Forecast MAE (next 3 months): £4.15
- Top-decile CLV accuracy: 91.2%

6. Visualizations and Tools

- **CLV Distribution Histogram** by segment
- **Retention Probability Curves** over time
- **Churn Heatmap** based on support tickets and tenure
- Time-series plots of ARIMA fit and forecast intervals

7. Deliverables

- Full CLV model in modular .m scripts
- Interactive MATLAB dashboard for retention team
- 25-page strategic insights report
- Customer-level prediction output (CSV + .mat formats)
- PowerPoint summary with charts for senior management

8. Strategic Outcomes

- Marketing was able to run targeted campaigns for top 20% CLV users
- Support team implemented proactive retention outreach for churn-prone segments
- Finance team improved CAC payback modeling using updated CLV values
- Customer segmentation strategy was revised based on CLV deciles

9. Tools and Libraries Used

- MATLAB R2022b
- Statistics and Machine Learning Toolbox

- Econometrics Toolbox
- App Designer for dashboard
- Integration with PostgreSQL via database toolbox