Project Overview

This project applies survival analysis and customer lifetime value (CLV) modeling to a telecom churn dataset. The primary goals are:

- To understand the factors influencing customer churn
- To identify high-value customer segments
- To compute Customer Lifetime Value using survival probabilities
- To detect at-risk subscribers for retention planning

All analysis is conducted using Accelerated Failure Time (AFT) models with the **flexsurv** package in R.

Data Description

The dataset contains **1,000 subscribers** with variables including:

- **Customer attributes:** age, income, address (years at current address)
- **Demographics:** marital status, education, retirement status, gender
- **Service features:** voice plan, internet subscription, call forwarding
- **Customer category:** Basic, E-service, Plus, Total
- **Churn information:** tenure (months until churn or censoring), churn flag

The survival time is **tenure**, and the event indicator is 1 if the customer churned.

Methods Used

Survival Modeling

I estimated multiple AFT models:

- Exponential
- Weibull
- Lognormal
- Loglogistic

Model comparison using **AIC/BIC** selected the **lognormal AFT model** as the best fit.

Survival Curves

I generated survival curves for a “typical” subscriber (average demographics and most common categorical levels) to visualize retention differences across model assumptions.

Customer Lifetime Value (CLV)

CLV is computed using discounted survival probabilities:

$$CLV = \sum_{t=1}^{24} MM \cdot S(t) \cdot \left(1 + \frac{r}{12}\right)^{-(t-1)}$$

Where:

- Horizon = **24 months**
- Monthly Margin (MM) = **$50**
- Annual Discount Rate = **12%**

CLV is computed for each subscriber based on predicted individual survival curves from the final model.

Segmentation

CLV is summarized by:

- Customer Category
- Marital Status
- Gender

Boxplots and summary tables highlight differences in long-term value.

At-Risk Subscriber Detection

For each customer:

$$Risk_{12} = 1 - S(12)$$

The top high-risk subscribers are flagged based on lowest survival probability at 12 months.

Key Outputs

- Final survival model: **Lognormal AFT**
- Survival curve comparison across distributions
- Individual-level CLV estimates
- Segment-level CLV ranking

- Predicted churn risk at 12 months
- Identification of at-risk customers
- Recommended retention budget and strategies