

## Customer Lifetime Value Analysis Case

Targeted marketing and cost sensitiveness by companies has changed the way companies view customers. Lifetime value – the net present value of customers calculated profit over a certain number of months is an important aspect in promotions. Customer life time value is a very powerful measure that analyses customer attrition and profitability risk at individual customer level. Customer life time value can help companies to develop customer loyalty and treatment strategies to increase the customer value whereas for newly acquired customers, customer lifetime value can help companies develop strategies to grow the right customers.

To calculate CLTV, we should first get the cost of Acquiring customers and then use the Survival Analysis - how long the customer would survive in the system along with the profitability of each customer to calculate the Lifetime Value of each customer.

We assume that Acquisition Cost is available and use Survival Analysis to estimate the customer Lifetime Value.

As for Survival Analysis, it is the study of the time between entry of a customer in the system and a subsequent event such as churn.

Conventional statistical methods (e.g. logistic regression, decision tree etc.) are very successful in predicting customer Churn but these methods are not used to get survival curve based on historical data.

In this exercise, we categorize customers who are yet to experience churn are called censored cases, while those customers who already churned are called observed cases (*or failed in our analysis*).

Finally, we calculate the total profitability of the customer for the entire duration with us and then deduct Acquisition cost from this to derive Lifetime Value of each customer.

### **1. CLTV OBJECTIVE:**

The objective of this study is to develop the concept of customer lifetime value. In this case, we present an overview of the data and analytics required to setup a CLTV framework.

### **2. DATA REQUIREMENTS:**

To develop a CLTV framework, we must understand the different data elements that should be available for the analysis. Here is a list of data that are generally needed for this type of exercise.

1. Customer Acquisition Cost
2. Customer Profitability and Survival Analysis
  - a. Age

- b. Demographics/lifestyle
- c. Type and intensity of product ownership
- d. Satisfaction level

One of the biggest challenges to implementing a CLTV framework is to ensure that these data elements are available for analysis.

Let us next discuss the analytics framework that is needed for putting together a CLTV measure. As we mentioned above, one of the biggest challenges in the process is that we have to collect a lot of data from very different sources. Therefore, it is very important to evaluate the data quality carefully. In our experience, putting together the data from different sources can be a big challenge.

Next, we need to develop strategies for combining the different data elements and treating the data for anomalies like outliers and missing values. This is not very different from what we did in the previous class.

Once we get past the data preparation stage we need to consider the modelling that will be required for the process. Usually three types of algorithms need to be developed

1. Develop algorithm to understand how long the customer will remain a customer (any form of survival modelling)
2. Develop algorithms to understand the purchase behaviour of the customer.
3. Develop algorithms to segment the customers. This might be needed because we might find that the predictors that drive the retention or purchase behaviour are significantly different across groups and cannot be captured by a single model.

### **3. Data Extraction for analysis**

Of all the variables we need for this project, some of them are readily available and can be sourced from the Data Libraries but the remaining variables are derived variables which have to be derived from the directly available variables.

### **4. MODELING:**

Through the series of steps mentioned below, we can calculate the Customer's current life time value.

#### **Step 1:**

Creating master dataset with all the required variables

#### **Step 2:**

Calculate total transactions of each customer in first 6,12,18,24 months

#### **Step 3:**

Calculate total transaction amount for each period

#### **Step 4:**

Consolidate all the variables i.e. derived & available in one dataset

### **Step 5:**

Calculate the probability of survival of each customer

We take data from 2008 to 2012 timeline. Where we make 2010 as Churn date (24 months), based on which we will analyze the survival of customer.

Ex: Two customer A & B start purchasing at the same date in 2008. Customer A's last transaction with company was in 2012 which is after our Churn date, so he is Censored and his Status marked as zero in our analysis.

Customer B's last transaction was at 2009, after that he did not deal with company within the mentioned time period till 2012. So we categorize him as 'Failed' and his Status marked as 1.

Duration is used to indicate the time that customer churn occurred, or for Censored observation the last time at which customer were observed any transaction with the company, both measured from the origin of time.

### **Procedure:**

LIFEREG procedure produces parametric regression models with censored survival data using maximum likelihood estimation.

PHREG procedure is a semi-parametric regression analysis using partial likelihood estimation.

**Modeling Process:** Explanatory data analysis(EDA) was conducted to prepare data for survival analysis.

A Univariate frequency analysis used to analyze value distribution, missing value and outliers.

Missing value for interval variables, replacement values were calculated based on the random percentiles of the variable's distribution i.e., value were assigned based on the probability distribution of the non-missing observations.

Missing value for class variables were replaced with the most frequent value (*count or mode*).

For variable reduction, multicollinearity checks are used. Other than that IV (*information value*) and WOE (*weight of evidence*) are used.

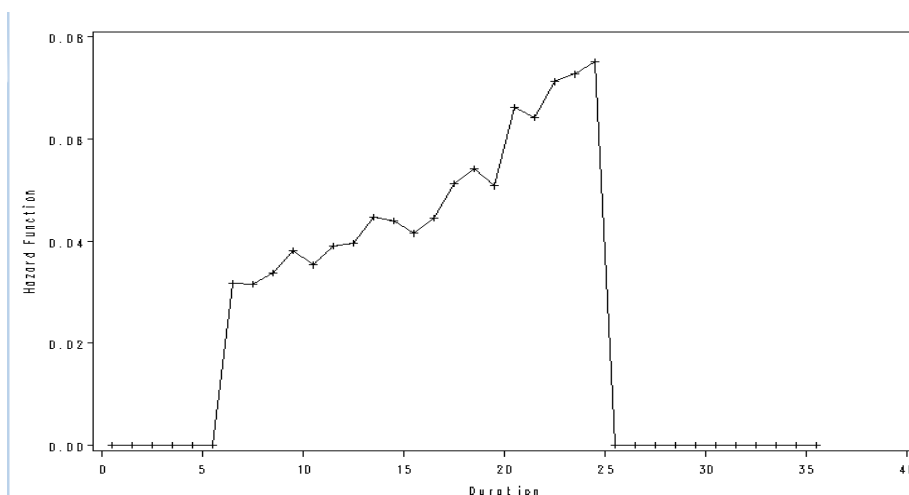
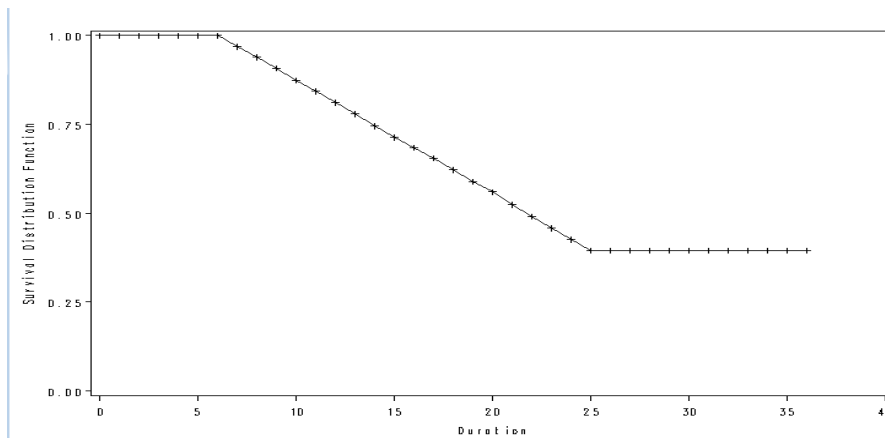
Phreg can be used to further reduce the number of variable based on the AIC criterion.

Based on the AIC values decide the final variables.

Check the model outcome with minimum AIC

After finalizing the variables spilt dataset into two parts one in training part and other is validation part.

Before applying Survival analysis procedures to the final dataset, customer survival functions and hazard functions should be estimated. These two functions give the knowledge of customer churn hazard characteristics.



From this analysis we get to know about the shape of the Hazard function, which can be considered as LogNormal in the above case

Using above information use LIFEREG (regression) to calculate customer survival procedure.

For using Lifereg procedure, we have to include a variable named "Use" to separate Training and Validation sample. Further to this, make status and duration for validation sample blank to generate score and validate within the process.

Now, append the above two datasets to form a single dataset with both training and validation sample.

Now on the final consolidated dataset we use LIFEREG procedure. This analysis gives us the probability and chi-square values for each variable and also convergence and likelihood value.

Next use the formula below for log normal to generate Scores / probability of survival within the given time period.

$$\text{PROB} = 1 - \text{PROBNORM}((\text{LOG}(T) - \text{LP}) / \text{\_SCALE\_});$$

Scale and LP comes as output of regression.

Scale comes along with the coefficients and LP comes as the structured part of the model.

We can get the customer probability of survival with every above mentioned time.

The above dataset gives us the probability to survive for each customer.

**Step 6:**

Now get these Survival numbers calculated in step5 and Profitability numbers calculated in step 4 together.

**Step 7:**

Calculate the customer life time value at each stage i.e. 6months after inception, 12months, 18months and 24 months.

This can be done multiplying probability of survival and total amount transacted at the end of each stage.

**Step 8:**

After deriving customer LTV at each stage, we can also calculate the Life time value as of now i.e. the current value of customer at this point in time.

For this to be calculated, we have to consider the discounting rate at which the value appreciates.