BIAM510 Final Project

Applications of Business Analytics II

BIAM510 (session)

April 22, 2021

Christine Baxter

Selected Lab for This Project (Week 4, 5, 6, or 7) : Week 5

Analytic Method Used in the Selected Lab: Survival Analysis

1. **Management Section - COs A and B**

**Application of this analytic method to Customer Life Cycle stages:**

1. Prospect
   1. Who are prospects?  
       We have three types of customer prospects. First are individuals in larger market areas that fall into the lower rate plan with a preference for dealer contact. The second type consists of individuals within the Smallville market area that prefer the dealer and store channels as they seem to have a higher percentage of active customers. The third type of customer prospect includes existing customers and utilizing up-selling, cross-selling, and usage stimulation to help ensure they come back for more.
   2. When is a customer acquired?  
       A customer will be considered acquired once their first order/purchase has been completed which will be labeled with ‘start\_date’. This includes receiving the credit card approval or money in-hand for the sale.
   3. What is the role of data mining?  
       Survival analysis is an excellent data mining technique that will guide our future marketing efforts by focusing on customer value, risk, and longevity. For example, customer lifetime averages can be established by rate plan, market segments/channels and cancel types giving us data to determine if one channel retains customers that lasts two times longer than another channel. Or, for another example, survival analysis will help us determine if our company performs better in one market over another market. Both, along with many more examples, will provide our company the ability to focus on areas that produce the best results and, ultimately, increase our customer retention which has a direct impact on sales/successfulness of the company.
2. New Customer

Listed in the following figure (Berry, 2010) is the high-level process for activating a customer. The ‘new customer’ begins with a sale coming in from various channels/markets and with the first payment being successfully completed. A prospective customer shows interest in acquiring a subscription when they provide billing/shipping address as well as payment information. This information may be obtained via phone call, web form, or mail-in response card. The next stage in the process includes creating account combined with a preliminary verification on the provided address/payment information. The subscription is provided via mail or electronic avenue. Confirmation of order and payment are received by the customer; the start\_date field is filled with the first order purchase date.

Chart, diagram

Description automatically generated

1. Established Customer

A customer is deemed established when they currently hold one or more monthly subscriptions, have been with our company for more than 6 months, and were active on date the data retrieval was initiated and completed. Data generated by these customers reflects actual behavior and are the best source for our data mining, answering questions like what our customers’ potential value is, what is the likelihood they will end relationship with our company or stop paying their bills. Using logistic regression and decision trees enables us to model involuntary churn and survival analysis helps us answer the ‘How long will it be until half the customers have been lost to involuntary churn?” question (Berry, 2010).

1. Former Customer

Within our company, a former customer is defined as someone who has voluntarily canceled their monthly subscription(s) and paid last bill or someone who was involuntarily canceled due to non-payment of subscription fee(s). For those customers who leave voluntarily, it is important that we track their reasons for leaving. With this collected information, data mining is used to help determine why our customers left, focusing on customer service complaints and behavioral data. Lastly, win back strategies including ‘Save Teams’ have been created and/or built to help valuable customers who have left return to our company by using incentives and pricing promotions.

1. **Methods Section - COs E, F, G, and H**

**Explanation of Analytic Method Used**

1. How this method works

Survival analysis in its simplest form is a model/analytic method for time until a certain event, also referred to as time-to-event analysis. Stated another way, it is the duration in terms of how long a person stays in a certain state. Within this analytic method, groups can be compared when time is an important factor and survival time can be measured in a variety of time measurements (e.g., years, months, days). This type of analysis is valuable for understanding customers and, by using past customer behavior, can assist with determining how long customers will remain customers as well as provide insights into the marketing process. When moving to the more complicated and detailed aspects, survival analysis is a set of tests, graphs, and models encompassing censoring (customers who survived to the data pull, no stop/cancel date is known), hazards (probability a customer may stop at a particular tenure), and survival curves (probability of customer surviving to any tenure).

In summary, survival analysis is about understanding customers, quantifying marketing efforts in terms of customer retention, providing a way of estimating how long it will be until something occurs with the understanding the event of interest will have usually occurred in some but not all, and estimating when customers will return after they have stopped.

1. Strengths

Survival analysis has many strengths including measuring customer retention (length of the customer relationship) and it is relatively easy to calculate the effects of different initial conditions such as number of times a customer has visited in the past. In addition, it provides a better understanding of customer half-life and average truncated tenure measurements as well as providing a better estimation of customer lifetime value by not just whether a customer will leave, but when they will leave.

Censoring is a unique feature within this method and, as noted above, enables customers who survived to the data pull or have no stop/cancel date known be dropped from the analysis or discarded from certain outcomes (Schober & Vetter, 2018). Censoring can be deemed as left censoring (customer tenure is less than certain duration) or right censoring (customer still active when data pull is initiated, and no stop/cancel date is known).

Overall, this type of analysis can provide a simplicity in analyzing survival data and ease of interpretation as well as provide a variety of tools for data that often cannot be analyzed any other way (Grace-Martin, 2018).

1. Weaknesses

It is possible that large outcome differences can be hidden within similar event rates and naïve analysis of untransformed times could produce invalid results. In addition, mistakes and distortions can frequently arise within the interpretation and display of survival plots (Bollschweiler, 2003). Model implementation within survival analysis is not as straightforward/clear and survival analysis assumes censoring is random with times varying across individuals. If the censored data is removed, the shape of curve will change, and biases will be created. Lastly, instances can arise where many variables cannot be correlated or monitored simultaneously (D41, 2020).

1. When to use  
   There are many uses for Survival Analysis, listed below are just a few:  
   ▪ Length of customer relationship  
   ▪ Measuring the probability of customers leaving  
   ▪ Fundamental component of customer value calculations  
   ▪ Forecasting customer levels  
   ▪ Predicting other types of events during the customer lifecycle  
   ▪ Understanding revenues  
   ▪ Determining customer’s average revenue that are active during the nth time unit  
   ▪ Forecasting the number of customers/stops on a given day into the future  
   All of the above uses have one thing in common…they all want to know how long it takes before a certain event happens. We discussed the uses, now here is a list of questions that can be answered with the survival data analytical method:  
   ▪ What is the probability that a participant survives 5 years?  
   ▪ Why did we lose customers?  
   ▪ When will a customer return?  
   ▪ Are there differences in survival between groups?  
   ▪ How long people stay employed?  
   ▪ How long before a lightbulb breaks?  
   ▪ When customers start on a minimum pricing plan, how long will it be before they upgrade to a premium plan?  
   ▪ When customers upgrade to a premium plan, how long will it be before they downgrade?  
   ▪ What is the expected length of time between purchases for customers, given past customer behavior and the fact that different customers have different purchase periods?  
   Using survival analysis, it is easy to calculate effects of different initial conditions (Berry, 2010)

**Alternative and Complementary Methods**

1. Alternative method

Thinking about an alternative method has proved challenging as the initial analysis was very thorough and encompassed several types of analytic methods. Survival/Retention Curves, Proportions, Distributions and Decision Tree analyses were implemented and discussed. However, to make a choice for an alternative approach, a chart could be constructed showing the original survival curve and the cumulative sums of the stops of different types. The overall survival curve could be used to determine the proportion of customers who leave at a given tenure (this is the hazard probability). These stops are then partitioned based on the hazards for each competing risk.

1. Complementary method

In regard to complementary methods, adding and/or showing hazard probabilities for three different groups of customers, one that lives/works in the Gotham market, one that lives/works in the Metropolis market and one that lives/works in the Smallville market then quantifying using one-year survival, median survival, or average truncated tenure. This would provide additional information that would prove useful in the overall analysis. Lastly, it would be great if the analyst could produce a chart or two covering what happens to subscription-based customers over time showing those customers who are still paying, who are missed opportunities, and those who are actual company loss.

1. **Analysis Section - TCOs C and D**

**Review of your efforts**

1. Strengths of specific analysis performed in the selected lab

Overall, this analysis brought to light several areas that needed further attention and/or improvement which could ultimately increase our customer retention to lengthier timeframes. For example, the frequency/percentage of customers’ length of tenure before becoming inactive through involuntarily/voluntarily measures were shown in several of the generated graphs. In addition, several areas of concern/anomalies were brought to light. For example, in Month 1, Month 13, and Month 100, there are significant drops where we lose approximately 30,000 customers, 40,000 customers, and 75,000 customers, respectively.

Starting after 30 months, our customer’s tenure steadily and consistently decreased. As a result of these analyses, the sales/marketing department and business operation leaders have actionable data enabling them to investigate and determine next steps/action plan to address and resolve these concerning items.

1. Weaknesses of specific analysis performed in the selected lab

Overall, the initial analysis was solid as a variety of methods were used from histograms and mosaics to distributions and decision trees so not too many weaknesses to report. As mentioned above, there was significant customer loss in Month 1, Month 13, and Month 100; however, no further analysis was completed to help the business determine or provide insight into the cause of those large losses. Hazard probabilities were not included in the analysis to show quantification of one-year survival, median survival, or average truncate tenure. In addition, it would be helpful to know and/or understand what happens to subscription-based customers over time showing those customers who are still paying, who are missed opportunities, and those who are actual company loss.

1. Recommendations to improve the analysis performed (at least three)
   1. Calculate hazard probabilities by group. To get accurate hazards and survival curves, using only groups of customers defined on initial conditions.
   2. Show hazard probabilities for three different groups of customers, one that lives/works in the Gotham market, one that lives/works in the Metropolis market and one that lives/works in the Smallville market then quantify using one-year survival, median survival, or average truncated tenure.
   3. Understand what happens to subscription-based customers over time regarding who remains active (and still paying), those who left voluntarily (missed opportunity), and those who were forced to leave (actual loss). These curves are constructed in two parts. First, the overall survival curve is used to determine the proportion of customers who leave at a given tenure (this is the hazard probability). These stops are then partitioned based on the hazards for each competing risk. The chart shows the original survival curve and the cumulative sums of the stops of the different types.

**Demonstration and implementation of recommendations**

(Step-by-step demonstration of how recommendations were implemented through redoing parts of the lab or performing additional analysis, with screenshots)

Figure 1. Distribution of Cancel Type by Market

Chart, bar chart

Description automatically generated

Figure 2. Distribution of Cancel Type by Censor Status

Chart, bar chart

Description automatically generated

Figure 3. Distribution of Market Area by Censor Status

Chart, bar chart

Description automatically generated

References

Berry, G.S.L.M. J. (2010). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. [devry]. Retrieved from <https://devry.vitalsource.com/#/books/9781118275603/>

Schober, P., & Vetter, T. R. (2018). Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesthesia and analgesia*, *127*(3), 792–798. https://doi.org/10.1213/ANE.0000000000003653

Grace-Martin, K. (2018, -08-06T12:00:35+00:00). Six types of survival analysis and challenges in learning them. <https://www.theanalysisfactor.com/the-six-types-of-survival-analysis-and-challenges-in-learning-them/>

Bollschweiler E. Benefits and limitations of Kaplan-Meier calculations of survival chance in cancer surgery. Langenbecks Arch Surg. 2003 Sep;388(4):239-44. doi: 10.1007/s00423-003-0410-6. Epub 2003 Aug 14. PMID: 12920602.

Tolley, H. D., Barnes, J. M., & Freeman, M. D. (2016). Chapter 10 - survival analysis. In M. D. Freeman, & M. P. Zeegers (Eds.), *Forensic epidemiology* (pp. 261-284). Academic Press. <https://doi.org/10.1016/B978-0-12-404584-2.00010-0>

D41. (2020, -12-09T10:36:36+00:00). Introduction to survival analysis and kaplan meier estimator. <https://www.analyticsvidhya.com/blog/2020/12/a-brief-introduction-to-survival-analysis-and-kaplan-meier-estimator/>