Teradata 2019 Data Challenge

Team Number: 6

Team Members:

Hamsa Rajasekhara

Jacqueline Venigalla

Lakshmi Lavanya Nakka

Pragna Yelamanchili

Tejaswini Naredla

Course: DSBA/ MBAD 6211- Advanced Business Analytics

1 Executive Summary:

1.1 Business Problem Statement

Hire Heroes USA, (HHUSA), strives to help place veterans and their families in civilian sector jobs. HHUSA is funded through grants and donations from an individual, corporation, and government sources. While grants and some donations are often received via solicitations, (such as fundraisers or events), or other procurement processes, many donations come from individual or corporate contributors who are not solicited. Many of these donors are one -time donors that, perhaps, chose to give to HHUSA only after being prompted. For example, job seekers often donate to HHUSA but may fail to become repeat donors once they no longer need their services. HHUSA should find a way to retain these donors so that they become repeat donors year after year.

1.2 Business Goal

The goal of our project is to determine the likelihood that a donor will become inactive, as a function of time, (going forward known as "donor survival rate"). The term "inactive" refers to the status of a donor that no longer donates to HHUSA. Our Team intends to reveal whether the survival rate of HHUSA donors decrease at a decreasing rate after some time, or instead increase or remain flat. The former could indicate that once a donor becomes a repeat donor, it is more likely that they will stay a repeat donor. This would also imply that HHUSA should focus on retaining their one-time donors. However, it is possible that the latter is true. Whatever the result, knowing this information would allow HHUSA to revise their retention strategy with the goal of increasing the survival time of each donor, and therefore donor lifetime value.

1.3 Data Profile

Our Team obtained the dataset from the Teradata competition, (Data Challenge), which provided 13 different datasets and one data dictionary to its competition participants. For our project, we used 3 of those datasets including Accounts, Opportunity, and Contacts datasets. The Accounts dataset included 16,859 records and 168 columns; the Contacts dataset included 132,446 records and 391 columns, and the Opportunity dataset included 10850 and 130 columns.

In our Survival Analysis, we used a recoded column that shows the total number of years a donor has donated, and an existing column that shows the status of the donor as either Active or Inactive. The target attribute, Active or Inactive, was missing for most of our records and had to be imputed based on logical criteria our Team set forward. Active donors are considered to be those who were still donating at the end of the date of our analysis, and Inactive donors are those who had stopped donating by the end of our analysis. Further details regarding these criteria are discussed in the Methodology section of this paper.

1.4 Results

From the survival analysis, our interpretation of the results was that approximately half of the donors become inactive after donating for only one year. Of the remaining donors, (classified as “active”), those whose donation frequency was “monthly” had the highest probability of having the greatest survival time.

2 Project Report

2.1 Introduction

This semester, our Team worked on the Teradata Challenge which involved analyzing numerous data sets from Hire Heroes USA, (HHUSA). HHUSA is a non-profit organization whose mission is to assist current U.S. military members, veterans, and their spouses with job placement by providing counseling, mentoring, access to employer networks, coaching, and other services.

Teradata Questions and their Relevance to HHUSA’s Organizational Goals

In this project, our focus was on analyzing donor engagement, as Hire Heroes is fully funded by donations from individual and corporate sponsors. Thus, these donations are needed to maintain the organization where, without them, HHUSA would cease to exist.

Our team was given a set of questions that related to HHUSA’s donors and their donation behavior. These questions did not explicitly ask us to create a model to predict or describe behavior, but instead, prompted us to do more exploratory data analysis within a particular domain – monetary donations. In short, our Team was asked to perform analytic reporting on the data. As there were many questions, we decided to tackle three questions in particular:

Question 1: What is the average donor lifespan?

Question 2: Where are individual donors located?

Question 3: Which state has the maximum number of donors?

These are specific questions raised by Teradata. In regards to question 1, our Team believes this question is an anchor to a larger question involving donor lifetime value. We imagine that one of HHUSA’s goal is to increase donors and donations. We believe that HHUSA would like to retain or increase donations from current donors, as well as attract and develop a relationship with new or potential donors.

In regards to question 2, our Team believes this question is an anchor for a larger conversation involving the geographical trends related to needs for HHUSA services. This could ultimately lead to a shift in where HHUSA focuses their time and attention, demographically. For example, if most donations come from one area, further analyses (outside the scope of this project), could be performed to understand if this is a function of the area's average income, an average number of veterans, or average population. By determining the answers to the questions above, we hope to find the effects of donors to Hire Heroes overall mission.

For our project, we will use the average donor lifespan to determine the time it takes for a donor to become inactive. We will provide the definition of active and inactive donors, and the algorithm we will use to perform our analysis later in this paper.

A Bit About Hire Heroes USA and Donors

Hire Heroes is headquartered in Georgia and offers vets and their families help in finding full-time or part-time work, internships, or training opportunities throughout several states in the U.S. The majority of Hire Heroes' clients register for services online. During the registration process, demographic and geographic information is collected via the MyTrak portal, in addition to detailed information regarding the client's preferences. Hire Heroes relies exclusively on public and private donations to support their initiatives. These donations are made online, via mail, or from a retail sponsor willing to donor a portion of their proceeds to HHUSA.

Donor analytics and website analytics can be fused together to perform a more comprehensive analysis. An analysis can be performed to expose behaviors of visitors that ultimately decide to be donors to an organization. Based on these insights, informed business decisions can be made, and strategies can be created or revised.

2.2 Background

There have been numerous analyses of donor and donations, or other philanthropic behaviors, performed in academic and industrial environments in the past. The vast majority of analyses have been on the likelihood of giving, which in turn has made logistic regression the go-to method for many of these cases. In addition to logistic regression, our Team was able to find evidence of several other types of techniques including Customer Lifetime Value and Clustering. A brief summary of how these 3 methods have been applied to previous donation research is provided in the following paragraphs, in addition to a brief discussion of donor analytics.

Donor Analytics in General

All organizations should get to know their supporters, and prospective supporters, better. Without out a clear understanding of this, organizations are unlikely to see a high return on their fundraising efforts. Over many years, organizations have sourced donation data from various channels, surveys, and notes collected by the staff. As a result, they have access to vast amounts of information that could be used to gain meaningful insight as it relates to the donor/donation domain. Briefly stated, four factors that are often key when analyzing donor behavior include their engagement history, giving habits, prospective giving capacity/propensity, and communication preferences.

Likelihood Models

Researchers at the University of Groningen's Marketing Department conducted research on over 300,000 donation observations to find out the drivers of charitable giving decisions. Over a period of 10 months, 20,000 individuals were given a choice on what to do with the money they'd received for completing a survey. Ultimately, these participants were partitioned into 3 groups, including Keepers, Donators, and Switchers. Univariate tests were used to investigate the differences between these three donation types. Further, a multinomial logit model was applied to simultaneously assess the relationships between the donation type and multiple independent variables. In essence, researches used logistic regression to develop a model for the 3 different donation classes. The binary outcome variable was "donated/did not donate", *(NIH.gov).*

In a different example, Blackbaud, a company which specializes in targeted analytics for their customers, creates three types of likelihood models including *Generic, Predictive,* and *Custom*, (part of their proprietary product offering). The *Generic* model is used to build a profile of the typical donor for that client. In this model, prospective donors are given a score within a range that indicates whether they are good or bad prospects. The *Prescriptive* model, Blackbaud's second type of model, is used to quantify a company's most likely prospects based on industry data. In addition to variables used on the *Generic* model, the *Prescriptive* model incorporates indicators such as affinity. This is a traditional likelihood model which predicts the likelihood of a donor giving to an organization.

Blackbaud's *Custom* model is similar to their *Prescriptive* model, but it constructs a profile of giving behavior unique to a particular organization. The *Custom* model is basically more comprehensive, as it takes a look at all solicited donors (donors and non-donors), along with certain behaviors that may indicate the strength of an individual's relationship with the organization. Examples include whether or not they are a volunteer, and how often they log into the organization's account, *(Blackbuad.com)*.

Clustering, CLV, and Time Series:

A Quick Word About Donor Lifetime Value

One of our Teradata business tasks is to find the donor lifespan. Our team believes this metric will be used to find the Donor Lifetime Value (DLV), an analytics methodology that uses contributions and other variables to predict or project the value of a donor/donee relationship, (*fundraisingreportcard.com).* Donor lifetime value is really an adaptation of the even more widely used customer lifetime value (CLV) algorithm. CLV's application to predictive analytics will be discussed in the following paragraphs.

Through our research, we found that predicting the future amounts, (not only cost or price but also spending or donation amounts), is often comprised of 3 particular analytics methods that work together. It should first be stated that the overall goal of predicting future amounts is to understand how to target certain segments to increase their CLV. Thus, Segmentation is often the starting point, where we want to learn how to target *segments* and not individual customers. The RFM (recency, frequency, monetary) model a common segmentation methodology, but other sophisticated clustering techniques such as K-Means Clustering are also popular. Once segments are defined and clustered, their lifetime value is determined, *(Optimove.com)*-. In the context of donations, these segments can be defined as high, low, medium, etc. giving groups.

Customer Lifetime Value is a valuation technique used to quantify customer segments. CLV has its own complexities when considering the time value of money. Thus, the CLV model also incorporates the future value of money. Finance and accounting algorithms often find the future value based on single sum or annuity amounts, but sophisticated data science methods incorporate analyzing trends of customer value in different seasons with time series analysis. This allows analysts to not only predict the CLV, (or DLV) of segments but also to predict their value at different times of the year. The graph below is an example of how, in a study conducted at the Toosi University of Technology in Iran, the aforementioned techniques are combined to produce an analysis.

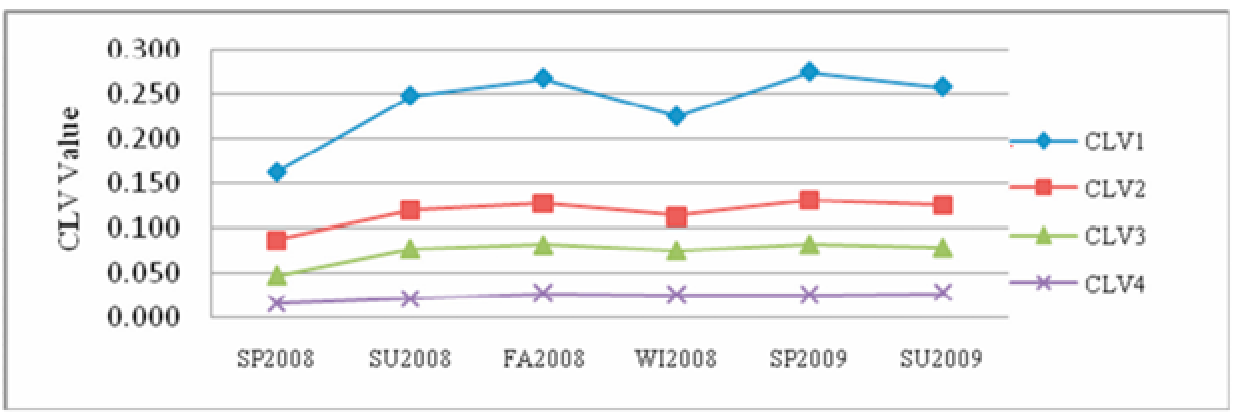


Fig. 3. Trend of CLV value in different segment during six seasons

The graph above shows the estimated customer future value of each segment, (CLV1 – 4) to reflect the seasonal behavior of each segment in a nonstationary status. The multiplicative seasonal ARIMA time series prediction method is used, *(Semanticscholar.org)*. As mentioned before, HHUSA asks us to find the donor lifetime value. Our Team believes Toosi University's analysis could have been an excellent guide for our analysis.

It should be said that, despite our extensive findings on how donations have been analyzed in the past, our Team decided to go in a different direction, and perform a Survival Analysis on our data.

The reasons and details of this technique will be discussed in the proceeding sections of our paper.

2.3 Data

Source of Data: Several structured datasets were provided by Hire Heroes USA. Some were not relevant, while several others required extensive cleaning and preparation.

Below is a brief description of each dataset we used, as it relates to donations:

Opportunities Dataset: Includes all sources of monetary contributions, including grants, donations, prospects, etc. Data is at the transaction level.

Contacts Dataset: Includes all employers who work with HHUSA to place job seekers.

Accounts Dataset: Primarily consists of information collected from job seekers. Some job seekers also chose to make donations at some point, which is recorded in this dataset. We used this dataset to obtain "state" which was often unavailable in the Opportunities dataset.

Below is an additional short snapshot of our initial data:

Dataset	**Number of Columns**	**Number of Records**
Accounts	16,859	391
Opportunities	10850	168
Contacts	132,446	391

2.4 Data Pre-Processing and Preparation

There were many small steps involved in the data preparation process. To complete our data preparation, we used a combination of leveraging Tableau functions, performing manipulations in Excel, coding in Python, and using features in SAS Enterprise Miner. We were not required to use imputed values to handle missing numerical values, as much of our data was available, only spread across spreadsheets, and across columns within those spreadsheets. Most of the cleaning, therefore, involved understanding which columns could be consolidated, which columns could be ignored, and which columns required recoding or aggregation. We did, however, have many missing values for our State variable.

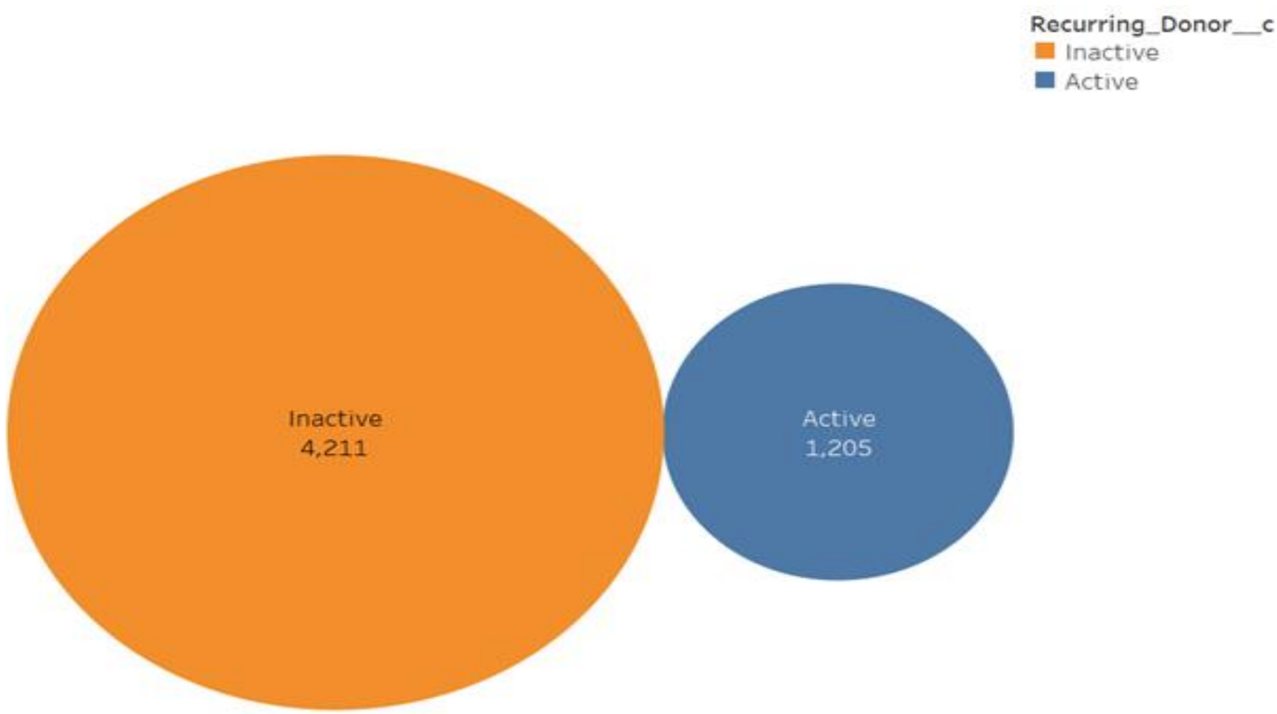
We have summarized the steps and rationale below:

Original Dataset	Field	How it was used	Included in Final Dataset?
Opportunities Contact Account	AccountID Id Id	unique account ID for each donor; renamed Id same as Opportunities AccountID. Unique account ID for each donor same as Opportunities AccountID. Unique account ID for each donor	yes, joined yes, joined yes, joined
Opportunities	COUNT_of_Id *(Recoded Column)*	counts the number of transactions for that unique ID, grouped by ID; this is the total number of times the donor donated.	Yes
Opportunities	Amount_Join *(Recoded Column)*	Joined "Amount" and "npe01__Payments_Made__c" columns for records where Amount = 0. Still at the transaction level.	*no
Opportunities	SUM_of_Amount_join *(Recoded Column)*	sum of transactions (Amount_join), grouped by ID	Yes
Opportunities	MAX_of_FiscalYear *(Recoded Column)*	maximum year ID appears in dataset	Yes
Opportunities	MIN_of_FiscalYear *(Recoded Column)*	minimum year ID appears in dataset	Yes
Opportunities	Ttl_Years_of_Donations *(Recoded Column)*	max - min + 1 (eg. 1 year would be 2017 - 2017 +1)	Yes
Opportunities	Avg_Amount_Per_Year *(Recoded Column)*	sum of transaction amounts divided by total years of donations	Yes
All datasets	State_joined *(Recoded Column)*	Joined on donor level ID's for across all 3 datasets as state information was available sporadically	Yes
Contact	Donor__c		Yes
Contact	Donor_Type__c	Donors classified as Corporate, Individual, or Event*	Yes

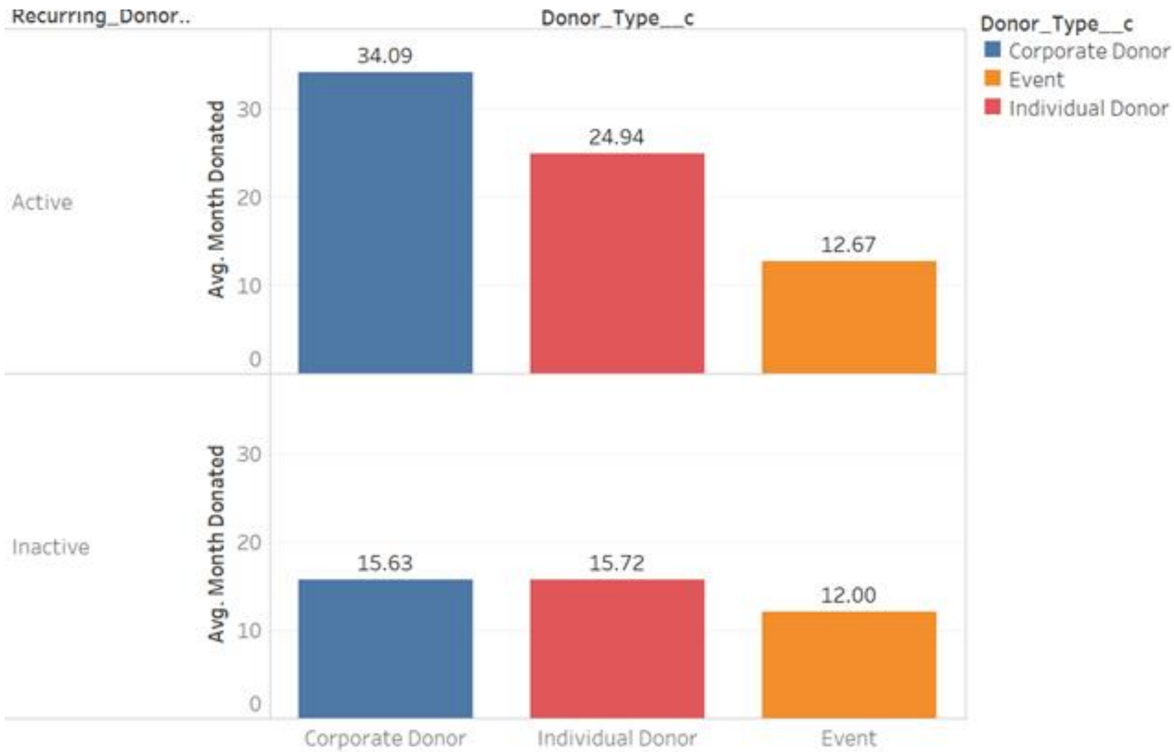
Contact	Recurring_Donor __c	Donors classified as Active or Inactive. **Inactive Donors were determined to be those whose last donation was before 2017, (which means that haven't made a donation since 2017 or earlier and are therefore inactive).* **Active donors are those whose last donation was after 2018, but this was not their first donation.* **We removed donors whose first and last donations were both between 2018 and 2019. Based on the fact that data was pulled around 2019, we cannot assume that they would not have donated again.*	Yes
Contact	Recurring_Donor_ Frequency__c	Donors classified as Monthly, Quarterly, Annual, or One-time	Yes
Opportunities	Type	Removed records classified as "Grant" or grant related.	No
Opportunities	StageName	Retained only records classified as "ClosedWon". * other stages included: closed lost, pledged, refunded, prospecting, search in progress, etc.	No
Opportunities	Amount; npe01_Payments_ Made_c	"Amount" and "npe01__Payments_Made__c" are two ways donation amounts can be represented in the dataset. Removed records where Amount = "0" and "npe01__Payments_Made__c" = 0. (114 records).	No
Contact	State	Joined to final dataset	No
Account	stayclass_cc_ state_c	Joined to final dataset	No
Opportunity	Billing State	Joined to final dataset	No
Opportunities	ID	unique transaction ID. Not used, but needed for clarification	No

In our initial exploratory data analysis, we started with a cleaned dataset that included 13 columns and 5416 records, which we visualized to answer the three Teradata questions we discussed earlier. Graphs and summaries of this information are provided below:

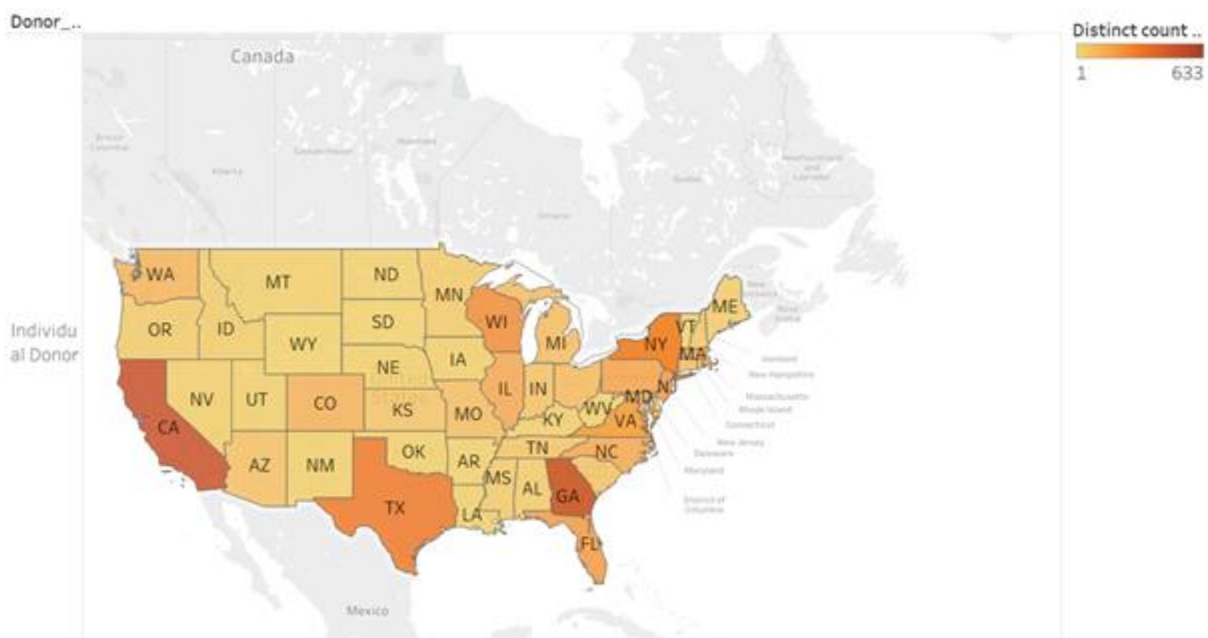
1. Distribution of Active Vs Inactive Donors: First we visualized the Distribution of Active and Inactive Donors based on Variable Recurring_Donor_c.



2. Average Donor Lifespan by Donor_type: Our first Teradata question focused on finding out the Average Donor Lifespan by Donor_Type_c for both Active and Inactive donors (Recurring_Donor_c).



3. Distribution of Individual Donors across various states: Then we focused to find out about the distribution of individual donors over the different States(geographical locations) in the USA. Below we have a list of states which had the highest counts of individual donors per state.



More Exploratory Data Analysis on our Final Cleaned Dataset

As we moved to our final dataset, it was reduced down to 12 columns and 4668 records. This is because we removed records of donors whose first donations were between 2018 and 2019.

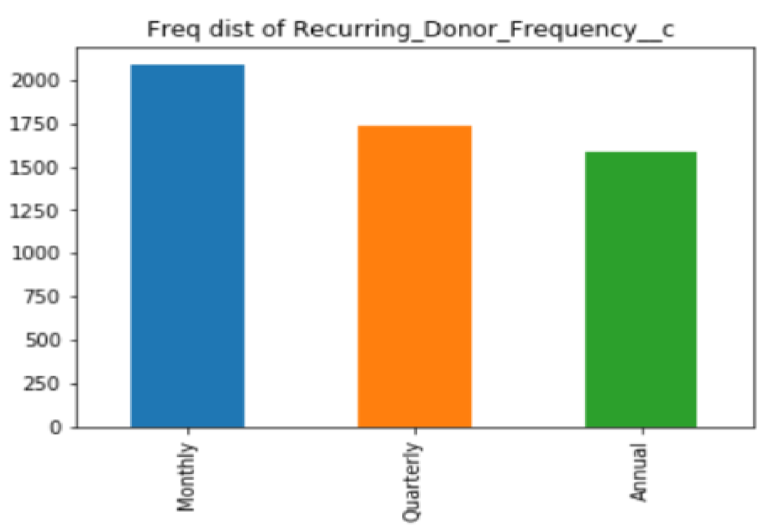
Below, we have provided some univariate analysis for variables within our final dataset.

Table1. Summary Statistics

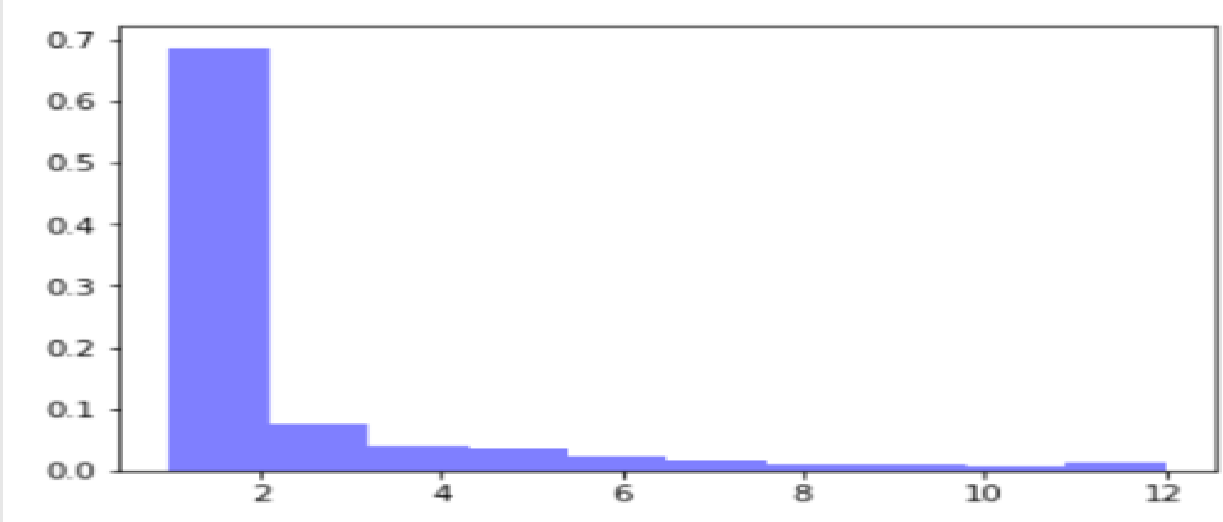
	Observations	Mean	Std. Dev.	Min	Max
Count_of_ID	4668.00	2.041131	5.7251	1.000	288.000
Donor_Type_c (0 is Corporate,1 is Event, otherwise 2 Individual)	4668.00	1.8777	0.4655	0.000	2.000
Recurring_Donor_Frequency (0 is Annual,1 is Monthly ,2 is One-time donation,3 is Quarterly)	4668.00	1.7170	0.46914	0.000	3.000
Ttl_Years_of_Donations	4668.00	1.5526	1.3874	1.000	12.000
Recurring_Donor_c (0 is Active, Otherwise Inactive)	4668.00	0.9020	0.2972	0.000	1.000

Here are a few distributions of some important variables in our Univariate analysis:

Frequency distribution of variable Recurring_Donor_frequency_c gives the rate at which most of our recurring donors made their contributions.



Frequency distribution of variable Ttl_Years_of_Donations gives the number of years a particular Recurring donor made his contribution.



Years

2.4 Methodology

After exploring various methods, we did not believe it was feasible to use regression or classification techniques to answer any of our questions. First, our data was heavily skewed, where most of our donors stopped donating after just one year. Second, a logistic regression analysis would have only given us the likelihood of donors being Active or Inactive based on several predictor variables. Instead, we believed that Teradata wanted to know just how long their donors would be their donors, not whether or not they were inactive based on certain criteria. Finally, clustering would not provide any insight to the questions being asked by Teradata in regards to the history of donator behavior. Throughout our project, our Team was fortunate enough to learn about other advanced algorithms, one of which was Survival Analysis. For our project, our Team believes that Survival Analysis can be used as a proper methodology for answering some of Teradata's business questions as it relates to donors.

To answer these questions, our team performed extensive data wrangling and preparation procedures, discussed earlier in this paper. After answering these questions, our team wanted to create a model that related directly to at least one of Teradata's questions. We decided to branch off of the question "What is the average donor lifespan?" to create a model that shows the percentage of donors that become inactive, as a function of time, year after year. To do this, we ran a Survival Analysis of our data.

In our final dataset, we classified donors as Active or Inactive. Our goal was to find out how long it takes a donor to become Inactive. In essence, we would like to determine the likelihood of that event – donors becoming inactive – happening in a specified amount of time.

To recap our explanation of Active vs. Inactive donors, an Inactive donor is one who has not donated since 2017. An Active donor is one that has donated as recently as 2018. In addition, our unit of time will be represented in years.

Survival Analysis: Definition

Survival analysis refers to the set of statistical analyses that are used to analyze the length of time until an event of interest occurs. The phrase survival time is used to refer to the type of variable of interest. It is often also referred to as failure time and waiting time. Such studies generally work with data leading up to an event of interest along with several other characteristics of individual data points that may be used to explain the survival times statistically.

The statistical problem (survival analysis) is to construct and estimate an appropriate model of the time of event occurrence. A survival model fulfills the following expectations:

- Yield predictions of a number of individuals who will fail (undergo the event of interest) at any length of time since the beginning of observation (or other decided point in time).
- Estimate the effect of observable individual characteristics on the survival time (to check the relevance of one variable holding constant all others).

It is often observed that the survival models, such as proportional hazard model, are capable of explaining the survival times in terms of observed characteristics which is better than straight-forward statistical inferences such as rates of event occurrence without considering characteristic features of data.

The survival data has to be right-censored which means that time at which the observation for the subject has to be censored and whether or not the subject's survival time has to be censored. A common feature of data on survival times is that they are censored or truncated. Censoring and truncation are statistical terms that refer to the inability to observe the variable of interest for the entire population. Ideally, a censored sample has more information than a truncated sample.

Basics of Survival Function: Algorithm

T — A random variable for survival time

t — A specific value of interest for T. e.g., ($T > t = x$) means whether the survival time is beyond x unit time. Time is measured in units of days, weeks, months or years.

d — Random variable denoting the occurrence of event or censoring.

($d = 1$ denotes the occurrence of the event, $d = 0$ denotes censoring)

Survival data are generally described and modeled in terms of two type of quantitative functions, namely Survival and Hazard.

Survival (or Survivor) function, S(t), is the probability that the event does not occur from the time of origin of study till the specific time t.

Following are the characteristics of a survivor function,

- S(t) decreases as t increases.
- S(t) = 1 at the start of the study. i.e at t = 0.
- S(∞) = 0. Sooner or later everyone will experience the event.

$$S(t) = P(T > t)$$

<u>Problem of Estimation</u>

As S(t) can be interpreted as a proportion, it can easily be estimated by the observed proportion of subjects surviving time point t:

$$S(t) = P(\text{Outcome} > t) \quad \longrightarrow \quad \hat{S}(t) = \frac{\#\text{ subjects surviving } t}{N}$$

<u>Kaplan-Meier Survival Curve</u>

The goal is to estimate a population survival curve from a sample. A Kaplan-Meier analysis allows estimation of survival over time, even when participants drop out or are studied for different lengths of time. For each interval, survival probability is calculated as the number of points surviving divided by the number of points at risk. Points who have died dropped out, or not reached the time yet are not counted as "at risk." Points who are lost are considered "censored" and are not counted in the denominator. Probability of surviving to any point is estimated from the cumulative probability of surviving each of the preceding time intervals (calculated as the product of preceding probabilities). Although the probability calculated at any given interval isn't very accurate because of the small number of events, the overall probability of surviving to each point is more accurate. The survival probability at any time is calculated by the formula given below:

$$S_t = \frac{\begin{matrix}\text{Number of subjects} \\ \text{living at the start}\end{matrix} - \begin{matrix}\text{Number of subjects} \\ \text{died}\end{matrix}}{\text{Number of subjects living at the start}}$$

The dependent variable in survival analysis is composed of two parts: one is the time to event and the other is the event status, which records if the event of interest occurred or not. To estimate the average lifespan of donors, we have dependent variables no of years of donations (time) and donor status (event). Kaplan-Meier Survival Curve uses survival function to estimate the survival probability.

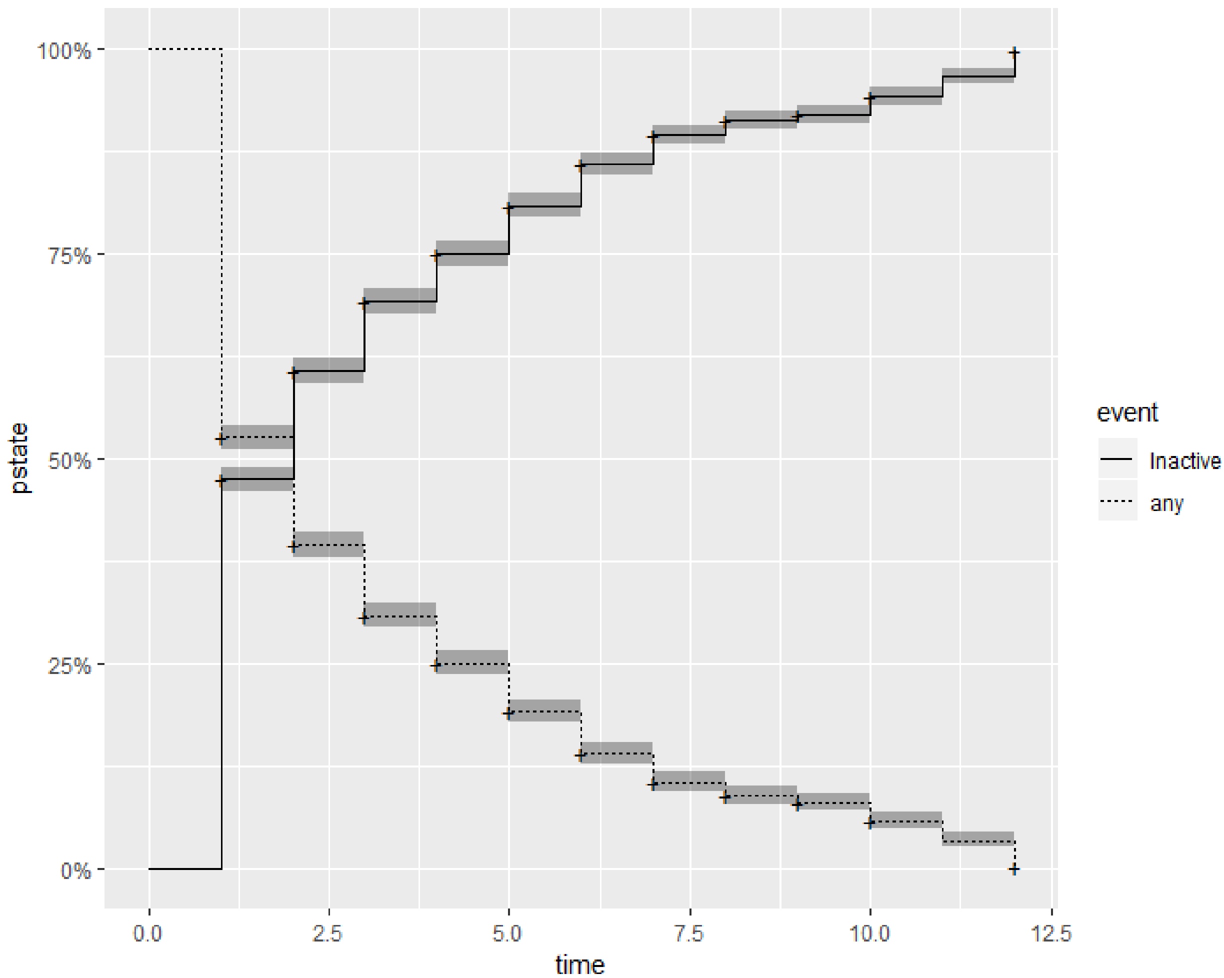
Hazard function, h(t), is the probability that an event occurring at a particular time t. To put it in another way it is the instantaneous rate of change of probability of an event at time t. Hazard function gives the risk of having an event at any instant, given the individual has survived till that instant.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t+\Delta t \mid T > t)}{\Delta t}$$

$P(t<T\leq t+\Delta t \mid T>t)$ is the probability that the event occurs somewhere between t and t+Δt. So as Δt approaches 0, we get the instantaneous risk of having an event. It is important to note here that Hazard function is not a probability as it is divided by time. Hence the hazard value ranges from 0 to ∞.

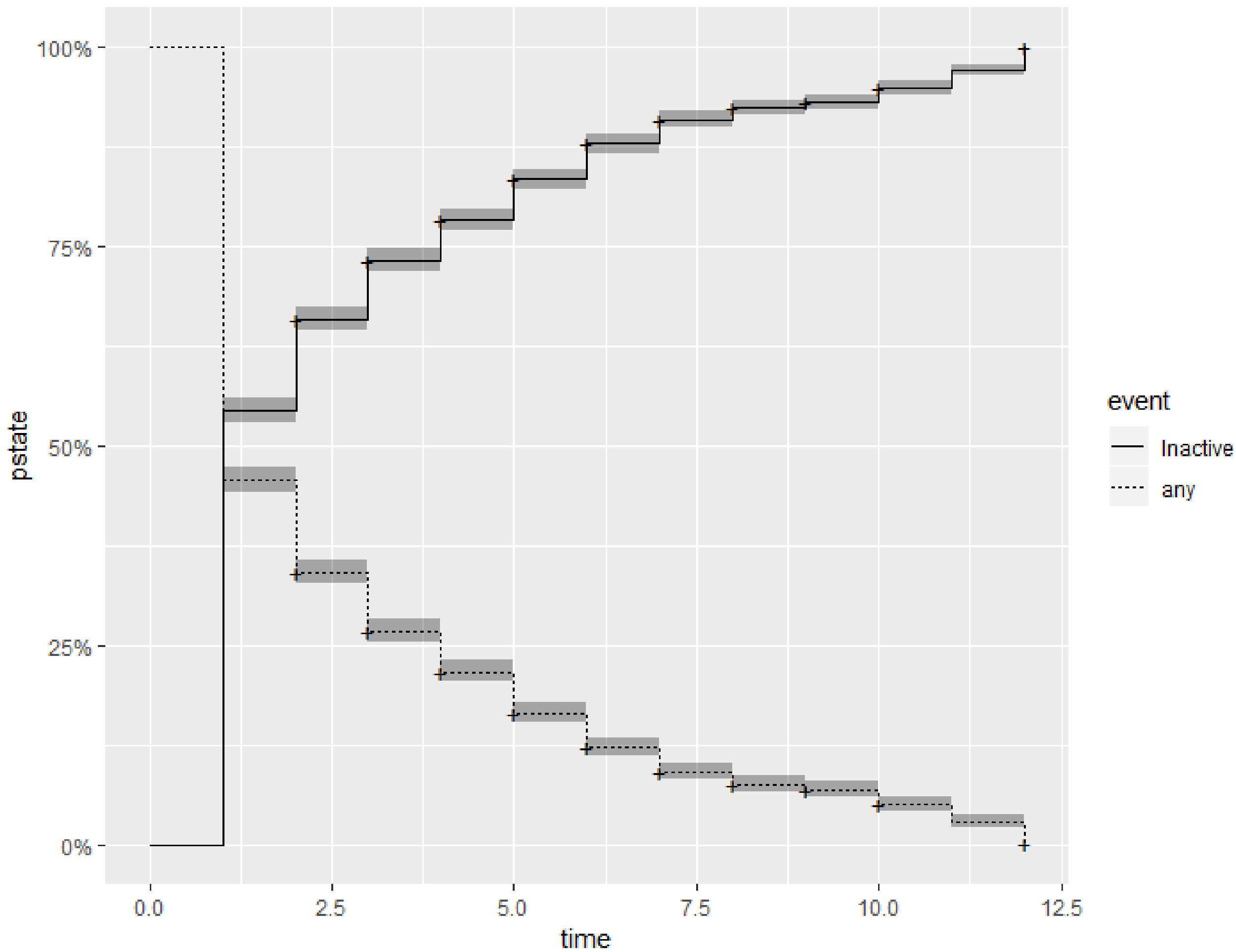
2.4 Results and Discussion

The visualization below is of a time series analysis that includes all of our observations, (including donors first donated in 2018 or later):



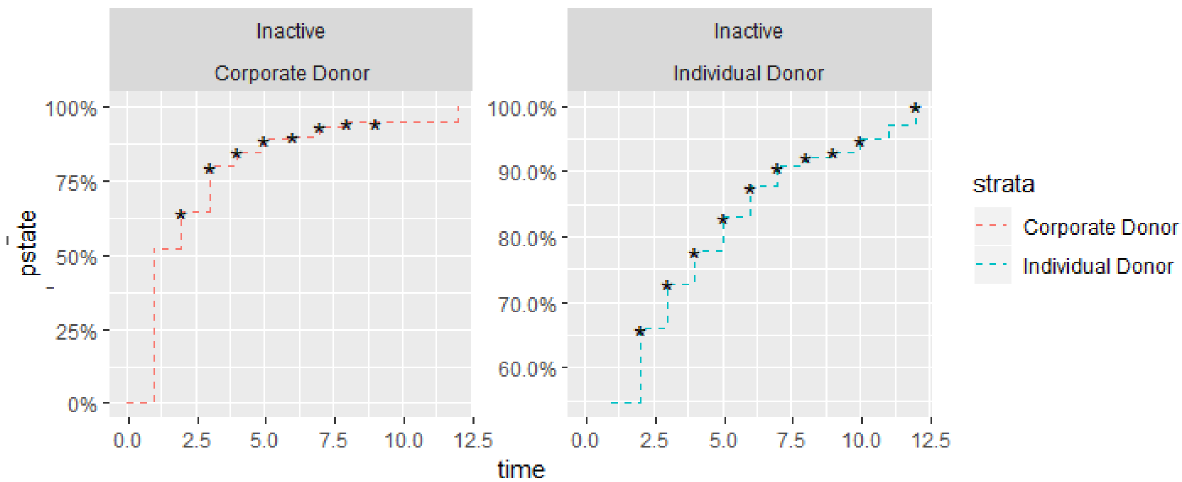
From the above plot, we can infer that around 47.4% of the donors were inactive after their first year of donation, (when we included the donors that started from 2018 and 2019).

As the per the business question, we thought considering the donors who have started from 2018 and 2019 would not give us effective results for the solution, as we are not sure if they would be donating a year later. Therefore, we also decided to run the model on the data excluding the donations that started between 2018 and 2019. The results are shown below:

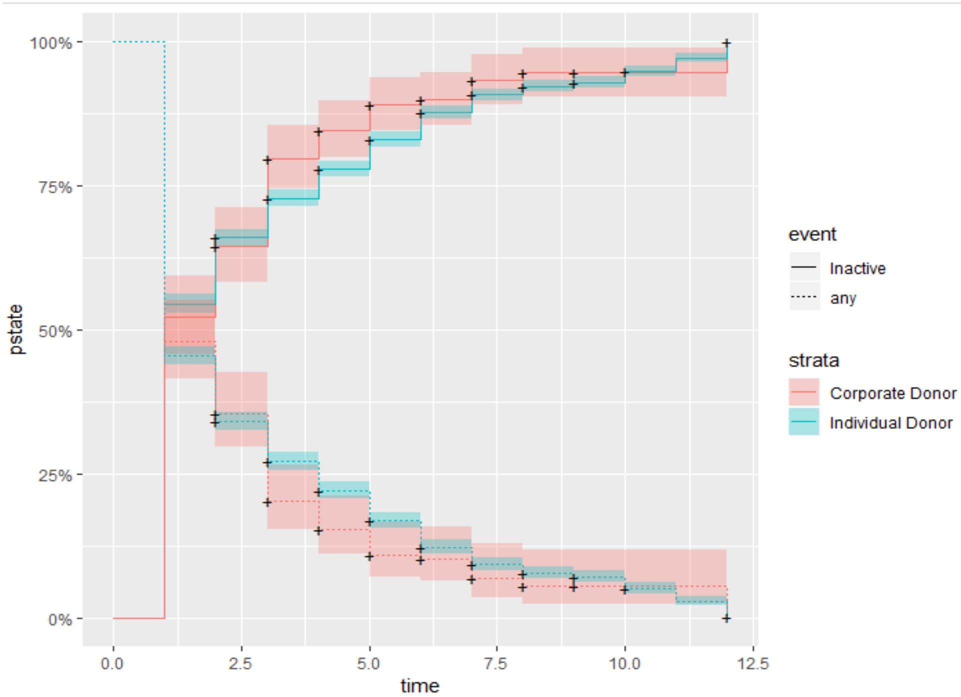


From the above graph, we can infer that around 54.4% of the donors become inactive after donating for a year.

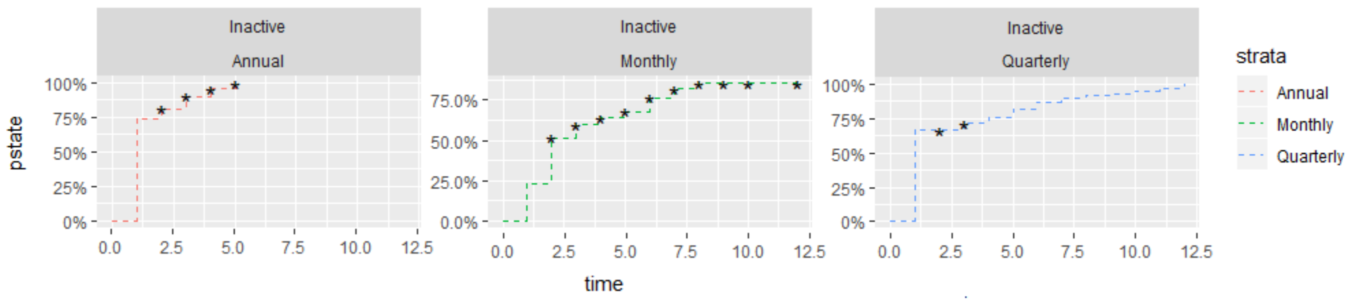
When compared to the other classes like Donor type, out of 211 Corporate donors, 110 of them have become inactive after one year, which accounts to be 52.1%. Looking into the Individual Donors, out of 3634 donors, 1981 have become inactive after a year totaling the percentage up to 54.5%.

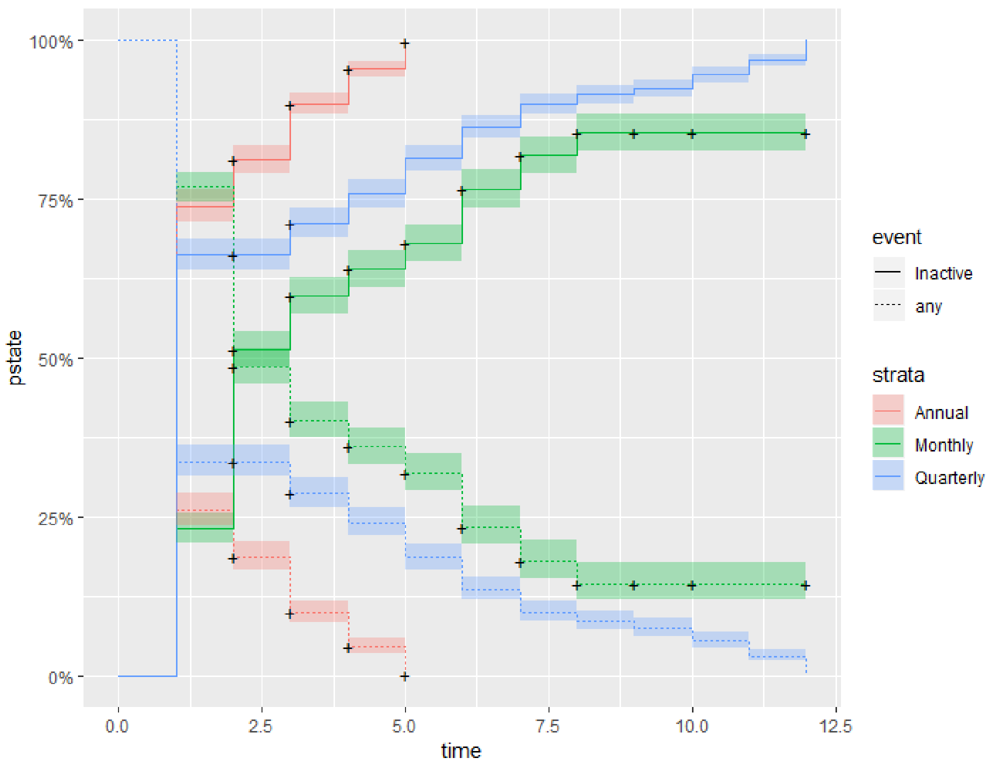


The visualization below shows the variation of donors across the years when we have considered the Donor type.



In a similar way when we considered the Recurring Donor Frequency class, we can clearly see from the graph below that 73.9% of the annual donors become inactive after one year. However, only 23.1% of Monthly donors and 66.2% of Quarterly donors become inactive after donating one year.





The graph above shows the variation of Recurring Donor Frequency class to determine if the Donor is active or inactive.

From the above analysis, we can conclude that almost more than 50% of the donors become inactive after their one year of donation. Of the remaining donors who are active, then a greater number of donations come from those that donate monthly.

Next Steps or Suggestions for HHUSA

HHUSA could possibly improve the survival time of their donors by improving the publicizing strategies to make them more visible. We can speculate that if a person or corporation has previously donated to HHUSA, they are likely to be receptive to the idea once again. However, a reminder or prompt may be necessary for them to do so. In addition, to generally increase their donation amounts, HHUSA could leverage the idea of having a fixed minimum donation amount, in particular for its recurring donors.

In addition, HHUSA should focus their attention on retaining new donors, as our analysis explains that most of our donors discontinue giving to HHUSA after just one year. HHUSA would need to develop a strategy specifically to target those donors. Some general strategies include continuing to deepen the relationship with donors after they have donated. This may be done by sending thank you cards, holiday cards, and/or providing quarterly newsletters, which not only keep donors informed about HHUSA but serve a reminder of HHUSA's cause.

References

1. Shefska, Zach. June 6, 2017 *Lifetime Value: What It Is & What It Isn't* https://fundraisingreportcard.com/donor-lifetime-value/

2. Khajvanda, M. and Tarokhb, M. 2010. *Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in Retail Banking Context*. https://www.semanticscholar.org/paper/Estimating-customer-future-value-of-different-based-Khajvand-Tarokh/84ed391cfaf1db37d2991976e27d89b5f598775b

3. Optimove Learning Center. N.d., *RFM Segmentation.* https://www.optimove.com/learning-center/rfm-segmentation

4. Target Analytics. N.d., *Using Statistical Modeling to Increase Donations.* Retrieved from https://www.blackbaud.com/files/resources/downloads/WhitePaper_TargetAnalytics_StatisticalModeling.pdf

5. Gauthier, J. February 27, 2017. *An Introduction to Predictive Customer Lifetime Value Modeling.* Retrieved from https://www.datascience.com/blog/intro-to-predictive-modeling-for-customer-lifetime-value

6. LeBlanc, L. and Rucks, C. September 27, 2009. *Data Mining of University Philanthropic Giving: Cluster-Discriminant Analysis and Pareto Effects*. Retrieved from https://link.springer.com/article/10.1057/ijea.2009.28

7. Leliyeld, M. and Risselada, H. September 20, 2017. *Dynamics in Charity Donation Decisions: Insights from a Large Longitudinal Data Set.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5606704/

8. Teradata. *2019 Data Challenge*. Retrieved from https://www.teradatauniversitynetwork.com/Community/Student-Competitions/2019/2019-Data-Challenge

9. Lityx. February 28, 2018. *Using Predictive Analytics to Maximize Fundraising Campaign ROI,* Retrieved from https://medium.com/nonprofit-analytics/using-predictive-analytics-to-maximize-fundraising-campaign-roi-de6e6192c45f

10. Diesing, Carl. April 17, 2018. *Donor Analytics Crash Course: 5 Nonprofit Data Essentials*. Retrieved from https://www.dnlomnimedia.com/blog/donor-analytics

11. Donges, N. February 22, 2018. *The Random Forest Algorithm.* Retrieved from https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

12.Shams S. July 23, No Year. *Introduction to Survival Analysis.* Retrieved from https://machinelearningmedium.com/2018/07/23/survival-analysis/

13. Vinodhkumar S. August 22, 2018. *Survival Analysis- An Introduction.* Retrieved from https://medium.com/zasti/survival-analysis-an-introduction-d6a6c2fdc54b