DEVELOPING MODELS TO PREDICT GIVING BEHAVIOR OF NONPROFIT DONORS

A Research Paper submitted to the Department of Engineering Systems and Environment
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Engineering Systems and Environment

Josh Eiland
Clare Hammonds
Sofia Ponos

By

Shawn Weigand

April 12, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR
William T. Scherer, Department of Engineering Systems and Environment

Developing Models to Predict Giving Behavior of Nonprofit Donors

Josh Eiland, Clare M. Hammonds, Sofia M. Ponos, Shawn M. Weigand, William T. Scherer
Department of Engineering Systems and the Environment, University of Virginia, Charlottesville, United States
jhe5ah, cmh3ed, smp7we, smw5gn, wts@virginia.edu

***Abstract* - Organizations in the nonprofit space are increasingly using data mining techniques to gain insights into their donors' behaviors and motivations. Data mining can be costly but can also be valuable in retaining and obtaining donors. Throughout the course of this project, we have prioritized two objectives. One is to increase the ratio of funds raised to dollars spent on fundraising from current donors, making these efforts more profitable. The other is to determine how to most effectively solicit new donors. To accomplish these goals, we have used statistical modeling and data analysis to gain insights and create recommendations related to donor optimization and acquisition. To learn about the current donors, it is important to identify which unique traits make donors more likely to donate and whether those traits are related to an individual's demographic information or giving history. Our team is classifying donors into "states" of giving based upon different metrics, including how recently, how much, how often, and for how long they have donated. We are using various data models to create actionable recommendations on how to tailor fundraising appeals specifically to different donors, which will increase the Inn's overall donations and their return on fundraising investment. We are also mapping the transitions between these giving states so that donors dropping from higher states can be re-engaged, while donors with a high chance of moving into a more profitable state can be flagged and targeted. We will present these results in a dashboard that the Inn can use moving forward to better solicit each donor and maintain a steady fundraising revenue stream.**

***Index Terms* - nonprofit analytics, data-driven nonprofits, data mining, donor mining, predictive modeling, Markov, RFM analysis**

I. Introduction

As the nonprofit space grows more competitive, with over 1.6 million registered 501(c)3 organizations in the United States and a 4.5% increase in this number from 2006 to 2016, each organization must find ways to set itself apart, retain its current donors, and attract new donors [1]. Organizations in the nonprofit space have started to recognize and take advantage of the benefits of data-driven decision making in recent years, with data mining providing insights that would otherwise be unnoticed [2, 3]. Extensive analysis can be difficult to undertake, as many nonprofits do not possess the resources themselves to perform high-level data-driven work, and outsourcing can be expensive. As data mining is relatively new to this sector, some nonprofits do not even understand the potential benefits of this type of work [4]. Many nonprofit organizations worry that in incorporating more technology and data into their operating structures, they will lose alignment with their organization's mission and become out of touch. However, while advanced models can be unfamiliar and may seem like a "black box," if used correctly, data-driven methods can help a nonprofit realize more of their goals in a shorter period [5]. The nonprofits that both have the resources and the desire to incorporate this work into their operating models achieve a competitive advantage and can more easily identify individuals with a high propensity to give to their organization.

Over the course of this past year, the team worked with The Children's Inn (The Inn) at the National Institutes of Health (NIH) to provide insights on their current donors, as well as recommendations for attracting new donors with substantial giving propensity. The Inn is a residential "Place Like Home" for families with children participating in clinical research studies at the NIH. The Inn provides families with room and board, meals, and activities free of charge. One of the Inn's key goals is to expand, diversify, and retain its base of supporters to ensure that they can continue to provide important services to the youth patients of the NIH and their families. Although the Inn already performed and outsourced some data-driven work, the team hoped to bring a fresh analytical perspective and help the Inn achieve their goals with new methods and techniques that the organization had not previously leveraged.

A. Prior Work

Data mining became relevant in the late 1980's and by the 1990's was being used as a subprocess of Knowledge Discovery in Databases (KDD). KDD is the use of data analysis to create new information, which is different from parsing through a database to find an existing answer; new insights must be created that could not be originally seen in the data. For example, finding a minimum or maximum value would not be considered KDD. However, finding out from a statistical test that donors are significantly more likely to give in December than other months would be considered KDD [6].

In 2005, a study was conducted on a nonprofit organization comparing the Recency, Frequency, Monetary (RFM), Chi Square Automatic Interaction Detector (CHAID) and logistic regression methods for increasing the response rate of donors when predicting the most profitable donors from a list. The study was conducted at different "depths," using the top 20%, 30%, 40%, and 50% of donors from a given file. The dataset included 99,200 members. The data was randomly split into a test group of 49,600 people and a holdout sample of 49,600. The overall response rate to the solicitation was 27.4%. The response rate for the test group was 27.3% and the response rate for the hold out sample was 27.6%. The methods were evaluated in two ways. The first was comparing the increase in response rate for a particular depth of the file in the test group with the increase in response rate for the same file depth in the hold out sample. The second was comparing the increase in response rate for a particular depth of the file across the three segmentation methods. The study results suggested that the difference in proportions between test and hold out samples for the three methods were very small and none of them were statistically significant. This suggests that for this study with a relatively large response rate, all three methods provide an accurate prediction of the response rate when the results of a test mailing are applied to the full file [7].

A survey of nonprofit organizations in Canada and Australia was conducted in 2017 to better understand the key tools they used in their knowledge management strategies. Over 95% of nonprofit organizations used physical documents; the two next most popular tools were public websites (such as Charity Navigator and other evaluation websites) and commercial productivity software (such as Microsoft Excel). According to the survey, commercial cloud computing services were not as popular among nonprofits because of their cost. There was a high level of use of low-cost or no-cost cloud computing services such as Google Docs among all nonprofits that were surveyed. Organizations that the study defined as "very small" had relatively lower use of tools for knowledge management activities across almost all sectors (religious, environmental, etc.). One of the key conclusions drawn from this study was that within the nonprofit sector there is an emerging focus on cloud computing solutions for knowledge management. Much of the investment and research in data mining to date has been by for-profit institutions. Although nonprofits have begun increasing their efforts in this space, they still lag the depth of achieved by their commercial peers [8].

B. *Challenges*

One of the main challenges the team faced throughout the course of this project was the attainment of all relevant data. Different data sources were used throughout the year, coming from both inside The Inn and external sources. These sources also had to stay as up to date as possible, especially given the COVID-19 pandemic, which has shifted giving patterns among Inn donors dramatically, influencing the insights and recommendations provided by the team. This resulted in the team acquiring new information constantly, with models, figures, and other results having to be updated accordingly.

Additionally, because separate teams oversee The Inn's direct mail, online, and events-based fundraising, it was difficult to analyze revenues and expenses across various channels to create an omnichannel optimization plan. Therefore, the group focused on direct mail campaigns, for which the most historical and consistent data were available. There was some work done on data from events, primarily to see which types of donors were likely to give high amounts at events rather than through other methods, but event data were not incorporated into most of the models built.

C. *Insights*

Historically, donor mining work has been heavily based upon identifying different donor groups, classified by a variety of variables. In our work, we focused on using recency, frequency, and monetary (RFM) metrics to categorize these donors. We also classified donors using demographic and historical giving data, which were narrowed down after models suggested which factors were most important in determining propensity to give to The Inn. What prior work often lacks, however, is showing how a donor can move between different categories. The team worked to model these transitions and discover how likely an individual is to move between certain states. This enables us to predict behavior and better account for both donors who may be about to fall off The Inn's radar and those who are poised to become large monetary donors.

II. Methods

A. *Data Collection and Infrastructure*

The data obtained directly from The Inn came mostly from Blackbaud Raiser's Edge, a cloud-based database for nonprofits. The team never interacted with the software itself, but instead received data pulls from Raiser's Edge with requested information. This included a list of over 6,000 historical donors of The Inn, dating back to The Inn's founding in 1990, as well as a list of all donations made to The Inn since then. This list of transactions included both responses to appeals sent by The Inn, as well as donations received at specific events such as golf tournaments, galas, and auctions. The two lists were merged using donor ID numbers, a unique identifier for every individual who has ever donated to The Inn.

For each donation within the database, the amount of money involved in the transaction, the date of the transaction and the name of the appeal that prompted the donation are all included, as are an ID for the gift, the payment method, and the type of gift. For each donor, key information provided includes the donor's full name, birth date, marital status and spouse ID, whether or not they are deceased, ethnicity and gender, phone number and email, and last known address. A more in-depth view of donors' demographic information was then compiled using external data sources.

The team acquired Consumer Insights data from marketing experts Jerry and Jamie Montgomery of 5W Strategists, who were able to find matches for around 60% of The Inn's living past donors based upon donor names and addresses. This appended data included economic factors such as household income, credit score, and home value, as well as demographic factors such as age, ethnicity, marital status, and profession. For some donors, data was available pertaining to their number of recent charitable donations and likelihood of being a charitable donor.

The group obtained donor acquisition lists from MINDset direct, a marketing consulting firm that works with The Inn to solicit current donors and appeal to potential new donors. These lists are the combination of names and addresses acquired from other organizations across the nonprofit industry, with a focus on those in similar sectors, filtered by MINDset to select the names they believe will be most profitable. The team acquired these lists for fiscal years 2020 and 2021 and obtained summary data such as response rate and average gift for acquisition appeals as far back as fiscal year 2016. MINDset also provided the team with appeal lists used to mail current donors of The Inn, which were used to identify who was sent each appeal and, when combined with transaction data, who responded.

The team's final data source was a recent survey by the Inn of their top donors and volunteers, in which 1000 email surveys and 1000 mailed surveys were distributed. This survey was focused on estate giving, both specifically to The Inn and in general, to gauge interest in future planned gifts. This survey, which generated 142 responses, was run through Stelter, a company that helps nonprofits with planned giving marketing.

B. *Modeling*

One of the main goals for this project was to create a profile for each donor by combining our various data sources in order to better predict each individual's probability and expected amount of giving. These profiles could then be used to better solicit each donor and could also be used to predict which prospective donors would be the most profitable to The Inn.

The first step in creating these profiles was compiling all of the information available for each individual. This included their demographics, which appeals they had been sent and had responded to, past donations, and whether or not they had left an estate gift to The Inn. With this information assembled, there was a clear image of the donor's lifetime giving history.

Using compiled demographic and historical donation data for individuals, a Classification and Regression Tree (CART) model was used to identify the donor characteristics most predictive of Inn donors. This model was also compared to one created based solely on FY20 acquisition lists to evaluate whether the random population same used in the first model is representative of individuals who were specifically targeted by the Inn. A third model was developed to differentiate between low-value and high-value Inn donors, as defined by total monetary donations to The Inn.

Another model was based on the survey data and was designed to help determine whether a current donor would be likely to leave an estate gift. The data from the survey was combined with the aforementioned donor profiles so that the most useful data points could be extracted to use in the CART model.

Along with the CART models, the team employed another one of the most common modeling techniques used within the nonprofit sector: RFM analysis. RFM stands for Recency, Frequency, and Monetary. These three attributes describe how recently a donor has donated, how frequently a donor donates, and how large of a donation a donor makes on average. For analysis of The Inn's donors, each of the RFM components were scored within a range of 1-3, with 1 being the least optimal and 3 being the most optimal. Table I describes what each of the values represents in correspondence to Recency, Frequency, and Monetary. When all three values are appended together, the result represents a state that a given donor is classified into during a given year. These scores can then be used to segment donors into mutually exclusive, collectively exhaustive states of giving. The team also experimented with additional metrics, including consistency, a measure of how often donors gave in consecutive years, and longevity, a measure of the range between the donor's first and last gift. Although these metrics were significant in predictive tests, they were less significant and not entirely independent from the traditional RFM components and other predictive factors (such as donor age), and were therefore not incorporated within models.

TABLE I. RFM STATE DEFINITIONS

Level	1	2	3
Recency (R)	No gifts in past 2 years	Gave at least once in the past 2 years	Gave at least once in the year
Frequency (F)	Did not give that year	Gave once that year	Gave more than once that year
Monetary (M)	Gave a total <$250 that year	Gave a total of $250-$1000 that year	Gave a total of $1000+ that year

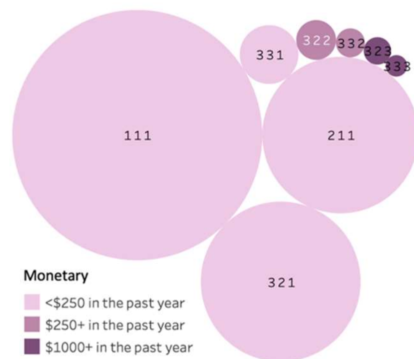

Fig. 1. Number of constituents in RFM states

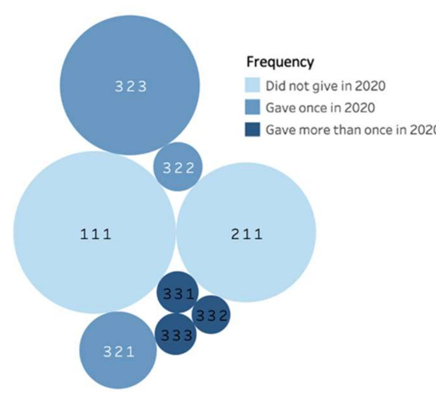

Fig. 2. Total donations in each RFM state

Fig. 1 and Fig. 2 outline some important features of individual RFM groups. In Fig. 1, the size of the bubbles corresponds to the number of donors in that RFM category, and the color of the bubble corresponds to the group's monetary component. From this figure, it is clear that most of the Inn's donors fall into the lowest RFM groups, as all four of the largest bubbles are in the lowest monetary group. In Fig. 2, the size of the bubbles corresponds to the sum of total donations that donors in that RFM category have contributed, while the color corresponds to group's frequency value. The figure illustrates that because so many donors are in the low-level RFM groups, those groups are responsible for large percentages of the total donations to the Inn. In addition, recently active donors who donate once a year in the highest monetary group (RFM group 323) also contribute a large portion of donations. However, those in RFM group 333 are a very small portion of donations, because of the low number of donors in those states. Insights generated from this initial look at RFM states informed later analysis and understanding of results.

The team's ultimate goal was to be able to map donors' journeys through different giving states and identify what causes donors to transition into more or less profitable states. For example, perhaps a common theme between donors who transition into higher states of Frequency is that they have a decrease in the number of dependents in their house. This can be explained by their children graduating college, leaving the donors with more disposable income. The Inn can then better solicit donations by approaching this specific donor group in a more personalized way. The different RFM traits that donors possess can also separate them into different "user groups". Some examples would be "those who give annually through a mutual fund" or "college students who donate online".

III. Data Analysis and Results

A. How Geography Affects Giving

Families who stay at The Inn come from all over the country. However, The Inn's main donor base is in the D.C., Maryland, Virginia area (DMV). The Inn provided the team with the home ZIP codes of families who have previously stayed at the Inn. Due to the Health Insurance Portability and Accountability Act (HIPAA) it was not possible to link Inn families to donors through names and addresses. However, the team compared Inn family ZIP codes with donor ZIP codes over time to observe trends. The results suggest that new donors often start giving after families from nearby ZIP codes stay at The Inn. Over time, the ZIP codes with Inn families create "hubs" that generally increase in donor concentration over time. An example of this is shown in Fig. 3, where blue dots represent individual donors to The Inn and red dots represent current and past families staying at The Inn. The size of red dots represents the number of Inn families from that ZIP code. The figure shows the growth in Inn family and donor prevalence from 2016 to 2019. In places where red Inn family dots appear, blue donor dots tend to cluster and increase in number.

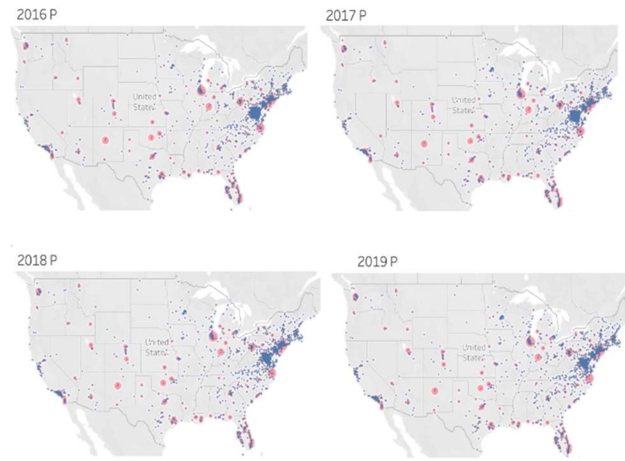

Fig. 3. Geographic patient and donor data over time

B. Giving Trends Analysis

By reviewing historical donations, the team discovered that many of The Inn's most valuable donors started by giving lower amounts or stopped giving for periods of several years before later giving large amounts. Fig. 4 shows an example of one such donor, who is one of The Inn's top 50 donors in terms of total dollars donated. Each red dot indicates a donation made to The Inn. The graph shows that although giving started in 1993 at a low monetary amount, there was a ten-year pause in giving, before donations recommenced with higher donation amounts. This indicates that appealing to lapsed donors and fostering strong donor-nonprofit relationships with individuals in all RFM states is important.

Fig. 4. Example giving journey of a top donor

C. CART Modeling

CART models were developed to identify characteristics most typical of Inn donors, high-value donors, and donors leaving estate gifts to The Inn. In Fig. 5, a tree diagram is depicted that was used to predict Inn donors from a sample population. This figure shows that individuals with a household net worth greater than $125,000 and with either a home value greater than $343,093 or at least two known recent charity recipients are most likely to be donors to the Inn. In addition to using these models to gain key insights about The Inn's existing donors, they can be used on random samples of individuals to predict giving behaviors. Fig. 6 shows the receiver operating characteristic (ROC) curve for this model, which indicates how well the model performs on a training data set. With a fairly low false positive rate, a high true positive rate, and an area under the curve close to 1.0, the model is a strong predictor of donor likelihood and can be used on other data sets confidently. Fig. 7 depicts the relative importance of the top 27 variables in predicting whether an individual from a random sample is an Inn donor. These attributes measure both individual and household-level measures as well as some median measures for the household's immediately surrounding region. The most important predictors pertain primarily to wealth, as measured by household net worth, home value, credit score, and income. Similar models can be applied to acquired donor lists to determine which individuals should be targeted in appeals as likely donors to the Inn, making new donor acquisition more cost effective for The Inn.

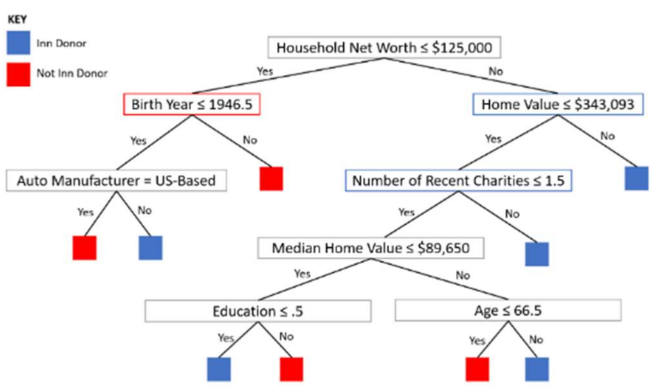

Fig. 5. Simplified tree diagram for predicting Inn donors

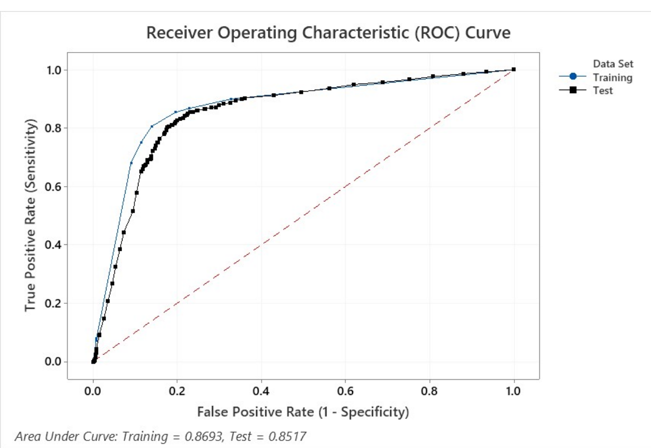

Fig. 6. ROC curve for tree diagram (Fig. 3) predicting Inn donors

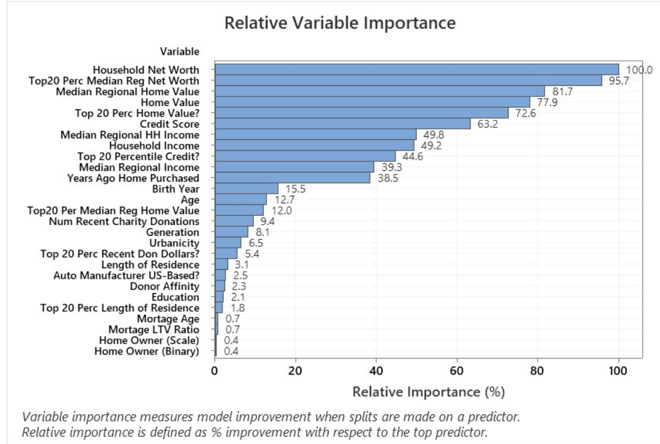

Fig. 7. Relative Variable Importance in Predicting Inn

D. RFM State Transitions

A transition matrix can be created to map the flow of donor RFM states from year to year. This matrix consists of the probability that a donor in one state at the beginning of a given year ends up in another state at the end of that year. RFM states for each donor in each year were calculated using the Inn's transaction data from 2015 to 2020. Eight possible states existed among donors: 111, 211, 321, 322, 323, 331, 332, and 333. The transition matrix in Fig. 8 displays both the number of donors who moved from one state to another and the probability that they moved from one state to another.

from(↓) to (→)	111	211	321	322	323	331	332	333
111	392611	0	18099	1050	680	1866	454	268
	94.60%	0.00%	4.36%	0.25%	0.16%	0.45%	0.11%	0.06%
211	15888	0	2184	241	144	547	128	65
	82.76%	0.00%	11.38%	1.26%	0.75%	2.85%	0.67%	0.34%
321	0	16181	5535	136	21	1361	258	47
	0.00%	68.74%	23.51%	0.58%	0.09%	5.78%	1.10%	0.20%
322	0	1018	93	440	39	23	95	68
	0.00%	57.32%	5.24%	24.77%	2.20%	1.30%	5.35%	3.83%
323	0	665	30	31	596	4	11	155
	0.00%	44.57%	2.01%	2.08%	39.95%	0.27%	0.74%	10.39%
331	0	1614	1132	21	3	1282	191	20
	0.00%	37.86%	26.55%	0.49%	0.07%	30.07%	4.48%	0.47%
332	0	478	181	85	13	111	297	75
	0.00%	38.55%	14.60%	6.85%	1.05%	8.95%	23.95%	6.05%
333	0	315	23	41	129	18	47	397
	0.00%	32.47%	2.37%	4.23%	13.30%	1.86%	4.85%	40.93%
	Number of transitions							
	Percent of "from" state that goes to "to" state							

Fig. 8. Transition matrix of RFM states

TABLE II. MEDIAN GIVING OF EACH RFM STATE

State	111	211	321	322	323	331	332	333
Median Giving	$0	$0	$35.00	$356.51	$2,200.00	$90.00	$500.00	$1,892.12

The transition matrix is useful to determine the value added or lost from each transition. Table II shows the median giving of donors in each state. Based on this median gift, Fig. 9 shows the amount of dollars that are lost or gained from each donor when transitioning between states. This type of matrix enables identification of state transitions that account for large gains and losses in donations. In this scenario, a donor in state 323 has a roughly 45% probability of dropping to state 211 in the next year, effectively surrendering $2,200 of donations for each of the occurrences. By combining these state transition rates and values with demographic data, we can statistically determine which attributes are significant in indicating whether a donor will remain in a given state or move out of it.

from(↓) to (→)	111	211	321	322	323	331	332	333
111	$0	$0	$35	$357	$2,200	$90	$500	$1,892
211	$0	$0	$35	$357	$2,200	$90	$500	$1,892
321	($35)	($35)	$0	$322	$2,165	$55	$465	$1,857
322	($357)	($357)	($322)	$0	$1,843	($267)	$143	$1,536
323	($2,200)	($2,200)	($2,165)	($1,843)	$0	($2,110)	($1,700)	($308)
331	($90)	($90)	($55)	$267	$2,110	$0	$410	$1,802
332	($500)	($500)	($465)	($143)	$1,700	($410)	$0	$1,392
333	($1,892)	($1,892)	($1,857)	($1,536)	$308	($1,802)	($1,392)	$0

Fig. 9. Monetary transition matrix of RFM states

To ensure RFM model accuracy, the next steps will be to examine individual transactions that contribute to the state changes in question in order to validate calculations and adjust for outliers. In this example, that would entail filtering the donors to only those who have transitioned from 323 to 211 in at least one year between 2015 and 2020. Donors with extremely large gifts can be removed from the dataset in order to account for outliers in calculating expected transition values.

IV. Conclusion

Since a major goal of The Inn is diversifying its donor and volunteer base, the team's work modeling locations of donors provided useful insights. The fact that donors from new geographic areas begin giving after a family from that area stays at The Inn indicates opportunities to obtain even more donors from outside the Washington D.C., Maryland, and Virginia (DMV) area, as Inn families are rarely from the DMV.

In addition to geographic donor demographics, other information such as household net worth, education, and age were useful in determining how likely an individual is to donate to The Inn. These key indicators will be combined with RFM metrics and provided to The Inn to inform future decisions.

Being able to model Recency, Frequency, and Monetary values of donors allows for effective segmentation of The Inn's donor base. With clear states, it is easy to track the transitions of both individual donors and The Inn's donor base as a whole over a number of years. The team's work on understanding lifetime giving of donors, especially those that become high-value Inn donors in terms of total money given, illustrates that donors who have transitioned to lower states still have the potential to return to high-level states.

Identifying the value of each giving state and the probability of transitions into higher and lower states enables the targeting of donors who are likely to increase or decrease in value. After validating the accuracy of the state transitions without outliers, the result of the RFM analysis will combine the transition rates with demographic information to indicate who the Inn should target to encourage high-value transitions. This will allow The Inn to capture donors on the verge of entering higher-value states and to retain donors who are expected to lapse.

A. *Significance*

The aforementioned models and results will enable The Inn to take immediate action in prioritizing high-value donors and allocating their budget across various campaigns. In the future, The Inn will be able to continue using these frameworks to invest time and resources more strategically into their fundraising efforts to maximize total donations while balancing other goals such as diversifying their donor base and advancing their business model.

B. *Future Work*

Future work will focus upon finalizing and increasing the accuracy of models, which will then be provided to The Inn along with documentation enabling them to be reproduced or updated with future data sets. Additionally, the team will conduct statistical analyses on the most recently acquired data source: an overview from MINDset on all appeals sent in the past five years. We hope that the Inn will continue to rely upon these models to inform their decision making, especially in determining how to interact with various types of donors.

Acknowledgment

We would like to thank The Children's Inn for allowing us the opportunity to work with them throughout this past year. Special thanks to Liza Cole and Peter Spring for being our liaisons with The Inn. We would also like to thank Jerry and Jamie Montgomery of 5W Strategists for all of their guidance throughout this project, as well as the work that they have contributed to The Inn. Finally, we would like to thank Foxhaven Asset Management for their generous sponsorship of this project.

References

[1] NCCS Project Team. (2019, June). *The nonprofit sector in brief 2019*. Retrieved from National Center for Charitable Statistics website: https://nccs.urban.org/publication/nonprofit-sector-brief-2019

[2] MacLaughlin, S. (2016). *Data driven nonprofits*. Saltire Press.

[3] Bopp, C., Harmon, E., & Voida, A. (2017, May). Disempowered by data: Nonprofits, social enterprises, and the consequences of data-driven work. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3608-3619).

[4] Lenczner, M., & Phillips, S. (2012). From stories to evidence: how mining data can promote innovation in the nonprofit sector. Technology Innovation Management Review, 2(7),10-15. doi.org/10.22215/timreview/575

[5] Anheier, H. K., & Toepler, S. (Eds.). (2020). *The Routledge Companion to Nonprofit Management*. Routledge.

[6] Coenen, F. (2011). Data mining: Past, present and future. The Knowledge Engineering Review, 26(1), 25-29. doi:10.1017/s0269888910000378

[7] McCarty, John A. Manoj Hastak. "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression."Journal of Business Research, vol. 60, no. 6,16 Jan. 2007, pp. 656-662

[8] Rathi, Dinesh and Lisa M. Given. "Non-profit Organizations' Use of Tools and Technologies for Knowledge Management: a Comparative Study." Journal of Knowledge Management, vol. 21, no. 4, 10 Jul. 2017, pp. 718 - 740.