**Understanding Donation Patterns in UT Alumni**

**EXECUTIVE SUMMARY**

Goals of the Analysis

* Identifying the characteristics of a “typical” alumni donor who has donated between $1-$10,000 since graduation
* Developing a model that accurately predicts the lifetime donation amount
* Quantifying the predictive utility of 3rd party information

Alumni Target Recommendation

We recommend targeting alumni who have donated at least once in the last 10 years, who got their first degree at UT before 2010, who first came to UT for a Master’s degree, and whose likelihood score of making a bequest, or legacy, donation is over 713 out of 1000. Some ways to target them might be personalized emails or merch giveaways in return of a donation.

Major Findings

* Alumni who have donated at least once in the last 10 years have an average lifetime donation of $1,264.24, while alumni who haven’t donated in the last 10 years have an average lifetime donation of $518.57
* Alumni who graduated in the 1960s have the highest average lifetime donation at $1,693.50, while alumni who graduated in the 2010s have the lowest average lifetime donation at $211.25
* PhD graduates have the highest average lifetime donation at $1,124.69, Bachelor’s graduates have an average lifetime donation at $975.48, and Master’s graduates have the lowest average lifetime donation at $858.65
* When predicting the overall average of lifetime donations, we would be off by the actual amount by factor of about 6.83, while our best model would only be off by a factor of about 4.33
* Using the 3rd party information when predicting lifetime donation amounts improves estimates from a factor of 4.47 to a factor of 4.33 in the best model
  + The 3rd party information variables should be reevaluated, and only the most important variables should be kept since a slight improvement is found in the model’s predictive power

Data

Analysis was done on a dataset of 7000 alumni with data collected from 2012-2020 and the total cash donation as of 2021. Variables include info on the alumni’s degree, social media, club activity, demographic info like gender and marital status, alumni-related activities, email interactions, sporting attendance, contact info like phone and email, 3rd party info like net worth, assets, and likelihood calculations, and past donation history.

Limitations

Data on sports attendance outside of football and basketball, data on other school activities like major-specific student organizations or plays, and college-specific data would be beneficial in more accurately determining “typical” alumni and predicting lifetime donation amounts.

Our model results would be fairly accurate in predicting 2022 lifetime amounts, but updated data will be needed to maintain accuracy, especially for the most impactful drivers.

**DISCUSSION OF FINDINGS**

Relevant Audience

Findings of this analysis are more relevant to the UT Development office in better understanding common characteristics of typical alumni donors. This information is also beneficial for UT Athletics and our 3rd party information providers, since typical alumni donors have characteristics that are relevant to their work.

Individual Driver Analysis

After fitting several models, we analyzed the most influential drivers towards higher cash donations of the best model.

Chart, box and whisker chart

Description automatically generated

Figure 1 shows the difference in average cash donation between alumni who have donated at least once in the last 10 years versus alumni who haven’t. 50% of donations in the “No” category were between $25-$350, while 50% of donations in the “Yes” category were between $75-$1,455. This highlights the importance of recency in donations, and alumni who have donated more recently are expected to have higher lifetime donation amounts.

Chart, box and whisker chart

Description automatically generated

Figure 2 shows average lifetime donation amounts by the decade the alumni graduated from UT.

Alumni graduating from 1950-1989 had higher variation in their lifetime donation amounts, with the 1960s being the most varied.

Alumni graduating in the 2010s have the lowest lifetime donation amounts in their box, likely because they’ve been an alumnus for less time. Alumni graduating in the 1940s and 1950s have little variation and low lifetime donation amounts, probably due to the economic ramifications of the Great Depression and World War II, or simply from a lack of alumni that fall in these categories in the data.

Chart, scatter chart

Description automatically generated

Figure 3 shows the linear relationship between bequest likelihood and total lifetime cash donations of alumni. There is a statistically significant relationship, where about 14% of the variation in lifetime donation amounts is explained by the bequest likelihood value.

Chart, box and whisker chart

Description automatically generated

Figure 4 shows the individual cash values by the type of degree the alumni obtained. If the alumni has more than one degree from UT, then the data accounts for the first degree they received. Although the middle 50% range is quite similar for each degree type, the interquartile range shows more variation at higher values.

Decision Tree

Figure 5 shows a decision tree that models combinations of alumni characteristics that generate an average lifetime donation amount. The highest average lifetime donation amount of $1,445.44 is expected to be from alumni who have a bequest likelihood number greater than or equal to 713 and who have donated at least once in the last 10 years.

**Figure 5**

Diagram

Description automatically generated

Important Predictor

Figure 6 shows how increasing values of the bequest likelihood variable affect how total lifetime donation is explained by the year the alumni graduated with their first degree at UT. The trend for each decade shows that as the value of bequest likelihood increases, higher total cash donations are predicted in the model for each decade of graduation years.

The most notable jump is in the 2010s decade, where predicted lifetime donations jump from about 1.8 to about 2.3 on the logged scale, which is from about $63 to about $200 in the actual cash amount when bequest likelihood is over about 700.

**Figure 6**

Chart, line chart

Description automatically generated

Breakdown of Individuals

Figures 7 and 8 provide individual breakdowns of two alumni: one who has a lower lifetime donation amount and one who has a higher amount. The individual with the lower donation amount in Figure 7 was negatively impacted in the model by almost all their characteristics. Most notably, their bequest likelihood number was below 713 at 219, and they graduated in the 2010s.

As seen in the decision tree in Figure 5, the combination of having a bequest likelihood number above 713 and having donated in the last 10 years increases the lifetime donation amount, which is also seen in the individual with a higher donation amount in Figure 8. Other characteristics that also increased the value were their favorite game, women’s basketball, and being an athletics donor.

**Figure 7**

Chart, waterfall chart

Description automatically generated

**Figure 8**

Chart, timeline

Description automatically generated

**TECHNICAL SUMMARY**

Data Cleansing Steps

The data provided was already divided into training and holdout samples, so they were combined for data cleansing. The following changes were made to the variables in the data:

* Variables kept as they were
  + GENDER, PHONE\_NUMBER, EMAIL\_ADDRESS, DEGREE\_1\_TYPE, LINKEDIN, TWITTER, FACEBOOK, NUM\_UT\_FAMILY\_RELATIONSHIPS, NUM\_STUDENT\_ACTIVITIES, FRATERNITY\_SORORITY, BAND, HONOR\_SOCIETY, STUDENT\_GOVT, ANNUITY\_LIKELIHOOD, BEQUEST\_LIKELIHOOD, and CRT\_LIKELIHOOD
  + For MARITAL\_STATUS, attempts were made to combine the NULL and Unknown categories, but they showed statistically significant differences in a connecting letter report, so they were left as is
* Variables with minimal changes to levels and cleaning
  + In ESTIMATED\_INCOME, the NULL and “Unable to rate” levels were combined, and dashes and dollar signs were removed to simplify the level names and reduce errors. The new level names are NULL, 1to50K, 50Kto100K, 100Kto250K, 250Kto500K, and 500K+.
  + In ESTIMATED\_TOTAL\_ASSETS, the NULL and “Unable to rate” levels were combined, and dashes and dollar signs were removed. Alumni with assets over $25 million were combined into one group. The new levels are NULL, <25K, 25Kto50K, 50Kto100K, 100Kto500K, 500Kto1MM, 1MMto5MM, 5MMto10MM, 10MMto25MM, and 25MM+.
  + In ESTIMATED\_GIFT\_CAPACITY, the NULL and “Unable to rate” levels were combined along with all redundant levels. All values over $1 million were combined as one level, and all dashes, spaces and dollar signs were removed. The new levels are UnableToRate, < 1K, 1Kto2K, 2Kto3K, 3Kto5K, 5Kto7.5K, 7.5Kto10K, 10Kto15K, 15Kto20K, 20Kto25K, 25Kto30K, 30Kto40K, 40Kto50K, 50Kto75K, 75Kto100K, 100Kto200K, 200Kto300K, 300Kto500K, 500Kto1MM, and Over1mil.
* Variables converted from quantitative to categorical
  + DEGREE\_1\_GRAD\_YEAR was converted to a categorical variable with each level corresponding to a decade from the 1940s to the 2010s.
  + NUM\_SPORTS was converted as a Yes/No categorical variable called SPORTS\_YES\_NO, where alumni with 0 values were given No and the rest were given yes. The same was done with the ALUMNI\_ACTIVITIES, which is now called ALUM\_ACTIVITIES\_YES\_NO.
  + NUM\_ALUMNI\_EVENTS was converted with the levels None, 1or2, and 3orMore.
* New variables created
  + Due to data error issues in the difference between DEGREE\_1\_GRAD\_YEAR and BIRTH\_YEAR, and new variable, AGE\_AT\_GRADUATION, was created with the levels Unknown, Before21, 21, 22, 23, 24, late20s, 30s, 40s, and 50sAndAbove. Rows with issues were placed in the Unknown level, and the BIRTH\_YEAR variable was deleted.
  + TOT\_MBB\_GAMES, TOT\_WBB\_GAMES, and TOT\_FB\_GAMES were combined to find each alumni’s favorite game of the three (FAV\_GAME). The original variables were then deleted.
  + Using TOT\_EMAIL\_CLICK, TOT\_EMAIL\_OPEN and TOT\_EMAIL\_RECEIVED, variables for ratio of clicks to open (CLICK\_OPEN\_RATIO), ratio of clicks to received (CLICK\_RECEIVE\_RATIO), and ratio of opened to received (OPEN\_RECEIVE\_RATIO). The original variables were then deleted.
    - These variables were left as numerical due to rare levels existing when trying to categorize them.

By the end of data processing and cleansing, 32 predictor variables were ready for predictive modeling

Model Building

6 models were run for this data analysis: regularized linear regression, K nearest neighbors, random forest, boosted tree, extreme gradient boosted tree, and neural net. The best model was a boosted tree model with a shrinkage of 0.009, and interaction depth of 15, 16 minimum nodes per tree, and 850 trees. Table 1 shows the generalization error of each of the 6 models run, with the boosted tree having the lowest generalization error at 0.636.

**Table 1**

Graphical user interface, text, application, email

Description automatically generated