Characterizing New Prospects
Data Science Capstone Project Milestone Report
Missy Lee
September 9, 2016

Introduction to the project
After recently closing our second fundraising campaign, we need to identify new potential major gift donors ahead of the next campaign. Major gifts are sought, not only to provide funding for buildings, scholarships, professorships and other initiatives but also attract other gifts. A major gift donor is currently defined as one who has the capacity to make a $25,000 gift. How can what we know about people who gave before be used to identify those who may give in the future? What characteristics can be identified and used to find currently unknown prospects? Do they currently give in smaller amounts (less than $25,000)? Are they more likely to be alumni?

State funding has declined over the past decade at the same time as the need for new facilities and programs has increased. Private philanthropy is looked upon to help fill the funding gap. We have very loyal and generous alumni who made the last campaign a success by donating very large amounts. These people may not have the inclination or financial ability to make similar gifts in the next campaign. Results from this analytical effort will be used to identify people who may be likely major gift donors, but who are presently not identified as major gift prospects. These names can then be passed on to development officers for further assessment.

This study will have two parts. First, I will test variables in a training set of data using linear regression to create a model of known donors. This part of analysis will begin by validating the use of the following variables, or not validating their use, in identifying the set of major donors to create a list of hypothetical donor indicators. In the second part of the study, I will use the model on a second set of data that contains data from known donors and people without a giving history to see if giving can be predicted.

Variables that will be tested to develop the description of current donors
- Total giving
- Velocity score (velocity is calculated by taking the average of a constituents giving history from the most recent three fiscal years, divided by the average of a constituents giving history over the past five fiscal years)
- Recency score (when in the last five years was their most recent gift?)
- Number of years of giving
- Largest gift in past five fiscal years
- Confirmed assets
- Whether the donor is an alumnus, parent of a former student, or friend
- Number of relatives who have attended the university
- Participation in student activities
- Distance between the university and current address

Approach to this analysis
My approach to the analysis will be to use linear regression to examine the strength of the relationship between total giving (dependent variable) and the other variables listed above. Variables showing the best fit will be combined to create a predictive model. This model will be used on a test set of data, one that includes non-donors as well as donors, to predict who will donate.

What this analysis can and can't achieve
This model may not be able to predict a person's likelihood of becoming a donor with certainty, but the prediction it yields will be more likely than cold calling at random to result in identifying a donor or weeding out those unlikely to give.

This model cannot determine which variables are causing the giving and which are simply correlated with giving behavior. While it will be possible to see variables that major donors share, it will not be possible to identify all of the factors that influence non-donors to become major donors. This study will not identify the non-quantifiable human factor that makes someone decide to donate. That task will be left to the development officers who meet with potential donors.