**Predictive Analytics Project Data**

A leading US nonprofit organization has agreed to release detailed transaction and contact history of the donors it acquired in 2000. The transaction and contact history extend through November 30, 2006.

The dataset consists of two files.

**Appeals.dta**: The organization contacts donors by sending appeals to solicit donations (by mail, phone, email etc.). Each month an appeal is sent to a group of donors. There can be multiple appeals sent out in some months – each appeal being part of a different fund-raising campaign. All appeals that are part of the same campaign in a given month have the same appeal id.

The variables in the dataset are:

Donor\_id: unique donor id.

Appeal\_id: unique appeal id.

Appeal\_date: date that appeal was sent.

Appeal\_cost: cost of that appeal.

**Donations.dta**: This dataset provides all donations made by each donor. The variables in the dataset are.

Donor\_id: unique donor id.

Gift\_date: date of donation.

Gift\_amount: donation amount.

Appeal\_id: appeal that donor responded to.

Zipcode: zip code of donor.

First\_gift\_date: The date of the first donation by this donor

Note: The first donation may not be captured in the donations database for all donors.

**Project Objective**

The organization would like to understand what factors influence and can best predict donation behavior (propensity to donate and amount donated) in the period 2004 to 2006. Specifically, they want to investigate three sets of factors – zip code level demographics (from the 2000 Census), appeals sent to donors and past donation behavior (i.e., donation behavior in the period 2000-2004).

**Project Questions**

Demographical factors analysis

* Do (zipcode level) demographical factors correlate with the past donation behavior?
* How well do they perform in predicting future behavior?
* How well is customer lifetime value (total donations over the duration of the dataset) correlated with demographics?
* Will you recommend certain demographics / geographical areas for acquiring new donors? Why?

Appeals analysis

* Do appeals have an impact on donation propensity of donors?
* Does this vary by cost of the appeal? Timing of appeal?
* The organization would like to optimize when and whom to send appeals to. What are some limitations of using the current data to optimize the targeting of appeals?
* Describe a field study that can overcome these limitations and can be used to optimize the targeting of appeals.

Behavioral factors analysis

* Is past donation amount of a donor a good predictor of the future donation amount by that donor?
* Two commonly used behavioral factors to predict future behavior in direct marketing settings are *recency* and *frequency*.
  + To measure recency and frequency for each donor, you need to aggregate the data on a monthly / quarterly / annual basis (for all donors). Select an appropriate time duration for conducting the analysis and explain why you chose that particular duration.
  + The frequency of past donations for a donor is the number of past periods in which that donor has donated (for example in 10 of the 16 quarters from 2000 to 2004).
    - You can count multiple donations by a donor within a given period as one donation. Discuss / explore other ways in which you can treat such observations.
    - Donors differ in when they made their first donation in 2000. Should you / How will you handle this when calculating frequency?
  + The recency for a particular donor is the most recent period in which the donor made a donation.
* What patterns do you observe in the data between recency and frequency of past donations and the number of future donations? How will you summarize this (descriptive analysis)?
* Discuss the possible reasons for the patterns that you observe above.
* Do recency and frequency have a significant impact on future donations behavior?