# National Charity donors Machine Learning project

Background:

The project requirement is to deliver actionable insights to our national charity client organisation.

The mail campaigns soliciting the donors for donations are necessary to highlight the work of the charity and to attract the donations for the to function. The mail campaigns are a significant cost for our client. The client wants to reduce the cost of the mail solicitation campaigns and improve the outcomes of the mail campaigns.

Our objective is to identify the characteristics of the most ‘profitable’ donors for our client, so that the next and future mail campaigns are focussed on the ‘profitable donors’. Focussed mail campaigns should result in reduced costs. Identifying the characteristics of the high value donors is a pre-requisite for increased amounts of donations being received by our client.

Data:

The available donor data identifies some general characteristics of the individual donors. The personal name, address details are not available in the data, possibly withheld in order to comply with the data protection law.

The available data has information which could be used for classification purposes i.e.

* Demographic data: Indicators denoting whether the donor is Urban, City, Suburban, Town, Rural resident
* Proprietary data: Indicators denoting whether the donor is a Home owner or not. For donors who are home owners, data relating to the value of their home is available
* Gift Numeric data: Last Gift Amount, average gift amount, life time amount, minimum/ maximum, average gift amounts, and the number of gifts data is available. Recent Gifts data is also available
* Income Numeric data: Per capita income, median household income data is available
* Income groups: Income group, socio-economic indicators are available

The most appropriate target variable appears to be FREQUENCY\_STATUS\_97NK. This variable the frequency of the donor as at June 1997. There appears to be a significant correlation between this variable and the RECENT\_CARD\_RESPONSE\_COUNT assuming that this variable indicates the positive response of the donor.

Business Analysis:

The initial business objective is to determine whether to contact the prospect or not. The secondary objective is to identify high value donors for increased donation amounts.

The first phase of the problem is a classification problem. This is because we have to determine whether or not the prospects should be contacted as the costs, time and effort are significant.

The second phase of the problem is a regression problem. This is not within the scope of this assignment as documented in the requirements. After determining which prospects are to be contacted, regression models would determine which prospects are likely to be most profitable donors. The last gift amount, lifetime gift amounts, and the lifetime gift count are good indicators of the profitable donors.

Both the first and the second phase are Supervised forms of machine learning because we already have labelled data.

Metrics/KPI:

The metrics used to evaluate the model is the Accuracy score of the model.

Donors data insights:

1. Most donors respond to the mail promotion campaigns. According to the data below, the count of the mail promotions sent out is approximately 15,000 promotions and just over 10,000 donors responded in the last 12 months (Card\_Prom\_12, Number\_Prom\_12 . In effect two thirds of the donor base responds to the mail promotions. This is to be expected.

There is direct correlation between Life time gift count of the donors responding and the count of the donors responding to the mail campaigns (Card\_Prom\_12, Lifetime\_Gift\_Count ).

1. The distribution of the Median Household income and the Per Capita Income is similar.
2. The pattern of the distribution of the Life time Gift Amount and the Recent Average Gift Amount is too similar. The data appears to be very limited and the relationship between these two variables as the data appears to be the same for the two variables.
3. The distribution of the Donor Age suggests that donors are aged between 35 – 80 form the core of the income for the national charity organisation. 75% of the donors are in this age band.
4. 

  








