



# A Simulation Game to Learn Predictive Analytics

## Fundraising Scenario

## Game's Scenario

You will be working on a fundraising campaign for a 12-year old, not-for-profit charitable organization (foundation) with a million members. The foundation has decided to add a direct contact campaign to its list of marketing activities. You will use predictive analytics to determine how many and which individuals to target in the campaign. The objective is to fundraise the most donation amount considering the costs of calling members (sum of predicted amount given minus costs).

You will be provided with the dataset of potential donors, and some pre-built templates (i.e., diagrams, pipelines, or codes depending on predictive modeling software or platform you are using).

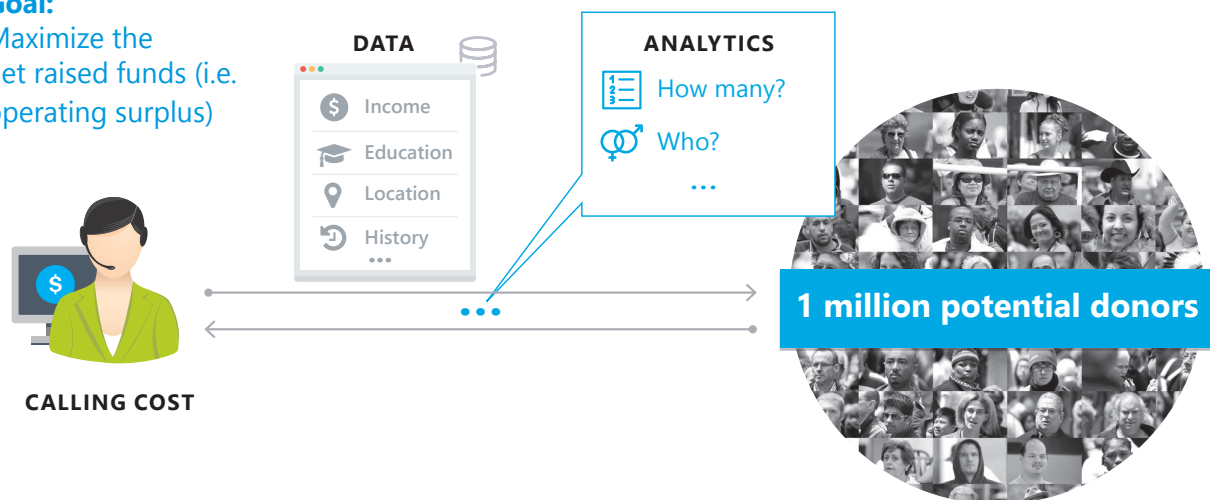
Please note that these are only templates to get you started. You will need to build new models, manipulate data, deal with missing values, and transform some variables to enhance the performance of your models and perform better than the baseline.

These pre-built templates will train and fit models based on previous behavior of donors (if they gave or not or how much they gave) and will also score donors to predict this year's donation. The list of scored donors will be exported to an output file/report (excel or csv).

Using this output, you will have to decide how many potential donors to target and will have to create a list of IDs of those potential donors. You will have to upload/submit the created list to the game's leaderboard which will rank the submissions based on **operating surplus – i.e., sum of donations minus the total cost of calling**.

### Goal:

Maximize the net raised funds (i.e. operating surplus)



## How to Play the Game

This game is designed to be played in 2 rounds. The cost schedule presented remain the same in both rounds. Please review each section below carefully for more details.

### *Round 1: Predict donation amount*

In Round 1, your mission is to predict the amount that each member would give next year to the foundation. You would then use this prediction to call the most valuable members (i.e., members who are predicted to give the most to the foundation).

### *Round 2: Perform a 2-stage modeling approach to calculate the 'uplift'*

In Round 2, the foundation still needs help to increase the net amount of the donations. To do so, the effectiveness of the modelling can be improved using a '2-stage modeling' approach to compute the 'uplift'.

We suggest the following steps to calculate the uplift:

1. Fit a model to determine the **Probability (P)** that an individual will give
2. Keep only the data of those who gave, fit a model for **Amount Given (M)**
3. Use both models to make predictions on the population
4. Compute **P\*M** to determine the 'Expected Donation' of each individual

Since the population is split into two groups - a group that was **contacted** and a group that was **not contacted** - we could calculate:

- 'Expected Donations **with a call**' = Probability of Giving\_Contact \* Predicted Amount\_Contact
- 'Expected Donations **without a call**' = Probability of Giving\_NoContact \* Predicted Amount\_NoContact

Then, the uplift could be calculated as the following:

- Uplift = Expected Donation with a call **minus** Expected Donation without a call

The idea is to contact people who yield higher uplift (value) when called.

## Available Information

In order to play the game and make decisions, you will have access to a dataset of 1 million potential donors as well as the costs associated to calling them:

**Table 1 Members' Data (please see table 3 for a full list of variables)**

Category of data	Example
<b>Membership activity (historical info)</b>	Minimum, maximum, & total donation, ...
<b>Demographic data</b>	Age, Gender, ...
<b>Socioeconomic status</b>	Salary, education...
<b>Previous behaviour (last year &amp; this year)</b>	If members gave and how much

**Table 2 Example of a Cost Schedule\***

Number of contacted members	Cost per person
0 - 60,000	5\$/person
> 60,000	25\$/person

*\* Please note that the cost schedule could vary from game to game. Make sure that you are using the correct cost schedule communicated to you by your teacher.*

**Table 3 List of Variables**

Variable Name	Description	
<b>ID</b>	Member number (unique ID)	ID data
<b>LastName</b>	Last Name	
<b>FirstName</b>	First Name	
<b>Woman</b>	Sex (1=woman, 0=man)	Socio-demographic
<b>Age</b>	Age (years)	
<b>Salary</b>	Annual salary in USD	
<b>Education</b>	Highest education level	
<b>City</b>	Type of neighborhood	History (please refer to the note below)*
<b>SeniorList</b>	Seniority for being on the VIP list	
<b>NbActivities</b>	Number of participations to annual meeting	
<b>Referrals</b>	Number of referrals	
<b>Recency</b>	Number of years since last donation	
<b>Frequency</b>	Number of donations	
<b>Seniority</b>	Number of years since first donation	
<b>TotalGift</b>	Total donation amount since a member	
<b>MinGift</b>	Minimum donation amount since a member	
<b>MaxGift</b>	Maximum donation amount since a member	
<b>Contact</b>	Direct solicitation this year (Only applicable to Round 2)	Target
<b>GaveLastYear</b>	Whether or not the individual gave last year	
<b>AmtLastYear</b>	Donation amount given last year	
<b>GaveThisYear</b>	Whether or not the individual gave this year	
<b>AmtThisYear</b>	Donation amount given this year	

**\* History (HIST) dataset gives the history of 10 years leading up to, but excluding, last year.**