Customer Analytics

—Group Research Project—

1. Background Information

Bank X has previously sold products to its customers via rather broad sales campaigns where customers were cold-called by call-center personnel. In retrospective the bank found that those campaigns rarely paid-off because the costs for such untargeted campaigns turned out to be immense given the small number of customers for which the campaign triggered a purchase.

The bank therefore decided to adopt a more targeted approach in its future campaigns. You are one of the bank's marketing analysts and you are asked for your help in that matter.

In particular, the bank plans to run a campaign to sell term deposits to its existing customers. A term deposit is a fixed-term investment which involves a money deposit at a financial institution. When buying a term deposit the money is usually locked in for the period for which the deposit is advertised. Funds can only be withdrawn after the term ends. The advantage of term deposits over other investment products is that the return on investment is very safe and the interest rate offered is usually more attractive than the one offered by alternative products. A term deposit contract is worth $180 for the bank.

The bank is reconsidering the call-centre approach through which customers will be contacted in the next campaign. Amongst other options, the bank is considering to out-source the call-centre activity and is now negotiating the maximum costs per customer contact. The call-center agency wants to have $7 per customer contact (successful or not).

Fully convinced by your skills as a marketing analyst, the bank's management wants you to develop a model that will be used to identify prospects in future campaigns. Management is also keen to learn about the campaign implications of the $7 dollar costs per customer contact that the call-center agency asks for.

2. Data, Sample and Variables

Since the bank has run previous campaigns you are provided a rich dataset that contains information about a random set of existing bank clients, details about the contacts made in the last campaign and other previous campaign attributes. The dataset is called "bank campaign". The dataset contains the following variables in particular.

2.1. Bank client information

age: (numeric)

job: type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')

marital: marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)

education: (categorical: primary, secondary, tertiary and unknown)

default: has credit in default? (categorical: 'no','yes','unknown')

housing: has housing loan? (categorical: 'no','yes','unknown')

loan: has personal loan? (categorical: 'no','yes','unknown')

balance: Balance of the individual.

2.2. Details about the last contact made in the most current campaign

contact: contact communication type (categorical: 'cellular','telephone')

month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')

duration: last contact duration, in seconds (numeric). Important note: Obviously the duration of the last call is not known before the call is performed. Consider very well, therefore, if this is a variable that is useful to identify the customers that shall be contacted in a marketing campaign and to fine-tune the campaign to them.

2.3. Campaign attributes

campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted)

previous: number of contacts performed before this campaign and for this client (numeric)

poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

2.4. Output variable

Respons - has the client subscribed a term deposit? (binary: 'yes','no')

3. Your Task

Your research goal is to develop a model that can be used to target prospects in future campaigns. Based on your model and the current data set you also want to determine what percentage of your customers could be contacted based on the price that the call-center agency asks for per customer contact to avoid a loss from the campaign.

Since your work shall lay the foundation for future work on customer prospect identification by yourself and by your marketing analytics colleagues at the bank, you are asked to provide excellent documentation of your work in the form of a research report.

Please specify and estimate a logistic regression analysis that serves to fulfils the research goal. Produce a report that contains the substance for other marketing analysts to critically evaluate your research approach but is still appealing to a managerial audience.

It is recommended to check the suggested structure for the report on Moodle; you are free, however, to choose whatever structure you think works best.

The report should not be longer than 10 pages including all graphs and tables but excluding references (font size 12, double-spaced). Please use the form provided on Moodle as a cover page for your report. Submit a digital copy via Moodle.

For the deadlines please see the updated timeline on Moodle.