

Predictive Modeling @ Vancity

October 2016

John Lo gazed thoughtfully at his computer monitor but was not thinking about the email messages that had appeared while he was in his last meeting. He had just received the green light from his supervisor to begin developing a propensity model to support the promotion campaign for the approaching 2016-2017 RRSP campaign. John had only joined VanCity earlier that month, and this project presented a great opportunity for John to demonstrate his abilities and his value to the department. At the same time, he was still adjusting to his new surroundings and was aware that this project would pose several challenges for him.

John's key concern was the time frame for the development and deployment of the project. With a deadline for delivering the mailing list within the next four weeks, John knew that he would be hard pressed to properly validate any model he would build. Fortunately, he would not be working alone. An outside consulting company had been retained to build a propensity model and John would be supervising their work while developing his own model.

Credit Unions

Credit unions developed as the result of cooperative activities in 19th century Europe. The main premise in organizing the first credit unions was that people should be able to pool their money together and make loans to each other. Even today most credit unions are governed by three main guidelines:

- Only members can borrow from the credit union
- Loans can only be made for 'prudent and productive' purposes
- A person's desire to repay, or his/her character, is considered more important than the ability to repay.

The first Canadian credit union was organized by Alphonse Desjardins in Levis, Quebec in 1900. It was a small credit union with the first savings deposit of only 10 cents and the total collection from all members of only \$26. Desjardins persevered and continued to organize other credit unions in North America. Today, total credit union membership in Canada is over 5.8 million. These credit unions are subject to provincial government regulations and as such cannot extend their operations beyond their own provincial borders.

Vancity

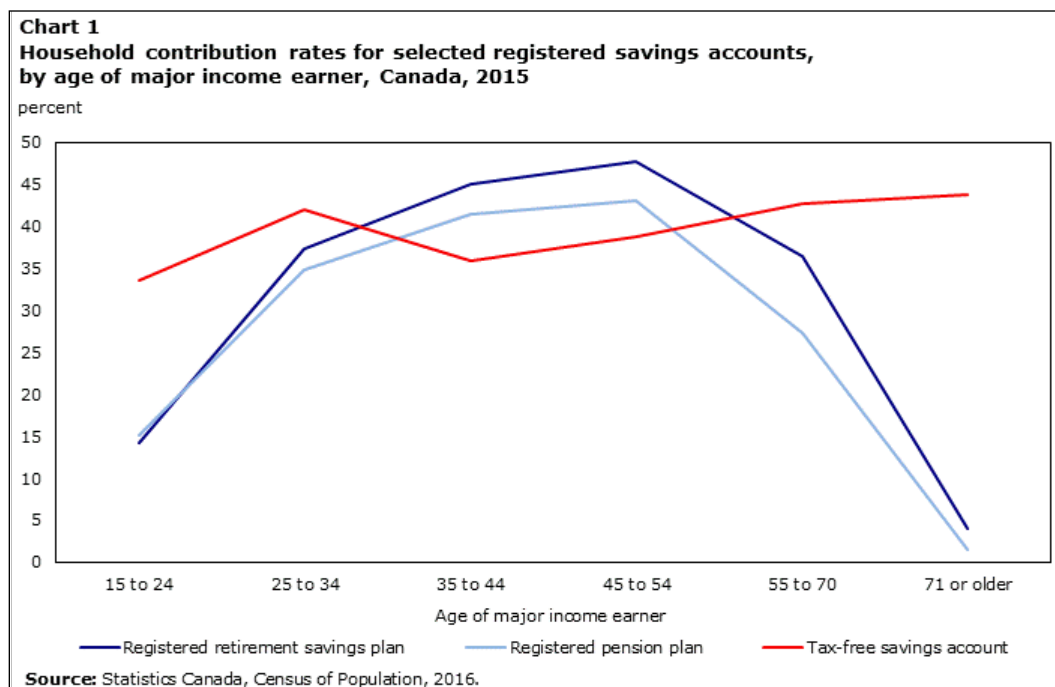
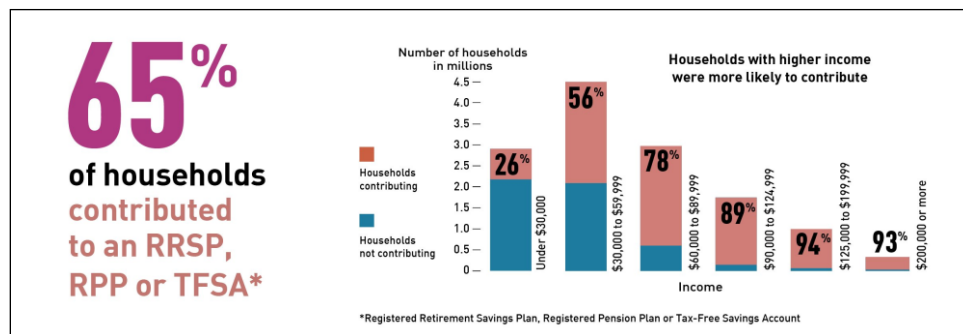
The Vancouver City Savings Credit Union, or Vancity, is Canada's largest credit union and the third largest in North America. Its assets in 2016 were \$25.6 billion, with membership base of 523,000 and 59 branches in British Columbia.

Vancity espouses a triple bottom line - social, environmental, and financial - and has a long history of community investing. Its vision ‘calls on us to measure success in terms of how we contribute to the well-being of our members and their communities—not just financial well-being, but social and environmental well-being as well’.

In 2016, Vancity was recognized as Canada’s top corporate citizen on the 15th-annual list of Canada’s Best 50 Corporate Citizens. Vancity’s member services centre also received a 2016 “Highest Customer Service - Banking Industry” award from Service Quality Measurement.

RRSP

The Federal Government introduced tax-sheltered Registered Retirement Savings Plans (RRSP’s) in 1957 as an incentive for Canadians to save for their retirement. Government statistics show that overall, higher income groups contribute more to both RRSP’s and TFSA’s, and contributions peak in middle age.



Vancity's RRSP Campaigns

Management set a goal of finding a more efficient way of cross-selling to its existing membership base. At the same time, increasing the return on RRSP promotions was not the only concern facing Vancity's managers. The new marketing campaign must consider Vancity's overall goals of social and environmental responsibility. Untargeted promotions reached many members who were not interested, and direct mail was especially nasty, consuming paper (and trees) and required recipients to recycle, or trash, the material.

Propensity Modeling

A propensity model is one kind of predictive model used in database marketing. One natural target behaviour is 'propensity to purchase' where purchasers and nonpurchasers are compared. 'Propensity' is used to distinguish such models from 'response' models derived from experiments, where there is an offer and the response (or not) is measured.

Vancity's RRSP Cross-Sell Propensity Model

Much of Vancity's RRSP marketing campaigns relied on mass media to get the message across to potential clients. These campaigns tended to attract new members but had little impact on existing members' RRSP purchases. At the end of September 2016, Vancity hired a consulting company to help them improve 'targeted marketing campaign response rates to cross-sell RRSP products to existing Vancity clients. At the same time, John joined VanCity and assumed a project management role primarily as translator, with responsibilities for supplying the consulting company with all necessary data and fielding questions that arose. John had graduated with a degree in business from Simon Fraser University, where he had concentrated on customer analytics, and he quickly suggested that there would be value in working in parallel with the consulting company. Management agreed, and approved John's suggestion to build a parallel in-house propensity model.

The decision to build two propensity models simultaneously served two purposes: 1) If the models agreed, it would instil confidence in their predictions; and 2) it helped the development and exploitation of the in-house capabilities of John and others, so that similar models could be built in-house in the future.

Model Objectives

The main goals of the cross-selling propensity models were to:

- Identify non-RRSP holding Vancity members who have the greatest potential for acquiring a new RRSP term.
- Develop profiles of the high-potential members that can support targeted message design.
- Increase RRSP volumes and new investment dollars, as well as product penetration and member entanglement.
- Improve targeting efficiencies and decrease campaign costs.

Since the plan was to apply the model during the following RRSP campaign period, from December 1st 2016 to March 1st 2017, to members who did not hold any RRSP with Vancity, John restricted the database for model development to those members who had not held an RRSP with Vancity *prior* to the 2015-2016 campaign. The observed target was the purchase of an RRSP *during* the 2015-2016 campaign. Model prediction would be the propensity of purchasing an RRSP for VanCity members who had not previously had one.

Data Preparation

The analysis data set consisted of:

- 2555 Vancity members who acquired a *first* RRSP during the 2015-2016 RRSP campaign period
- 2555 randomly sampled Vancity members that did not own an RRSP and did not purchase an RRSP term during the same campaign

This represented an oversampling of RRSP purchasers to an artificial 50% response rate. The actual RRSP purchase rate (for those not previously owning an RRSP) in 2015- 2016 was 2.2%. To develop the model John used predictor variables available through the core banking system. Variables from this system included:

- ***Member personal information*** such as age, tenure, branch, gender, value segment, behavior segment
- ***Member product portfolio*** that included products held by month and balances by month
- ***Member banking transactions*** by channel and by month

Finally, this data was combined with Tax Filer Data available from StatsCan that provided average data at the neighborhood level. When interpreting the estimated model, care must be taken to remember that these geodemographic variables are not measured at the individual member level.

In the end, John tested over 70 unique variables, half of which had 12 monthly permutations. In the process of modeling, hundreds of additional variables were created through transformations, such as logarithmic and binning.

ASSIGNMENT DIRECTIONS:

The data is provided in the csv file, **vcRRSP.csv**. Most of the original variables have been deleted from the data set, leaving 31 variables to work with (TABLE 1A below). This is based on real data and will require some thoughtful attention to missing values and other cleaning chores to get good results. Your objective is to develop a “good enough” model, and to recommend to VanCity the number of members who should be contacted to maximize profits, (contribution from RRSP sales minus contact costs). It is important to John to be able to interpret the model to check that it makes intuitive sense, and to be able to sell it to management.

ASSIGNMENT DELIVERABLES:

A word document:

1. Describe the key steps, following the “**rapid model development**” framework, that you went through in arriving at your final model, and **why** you took those steps.

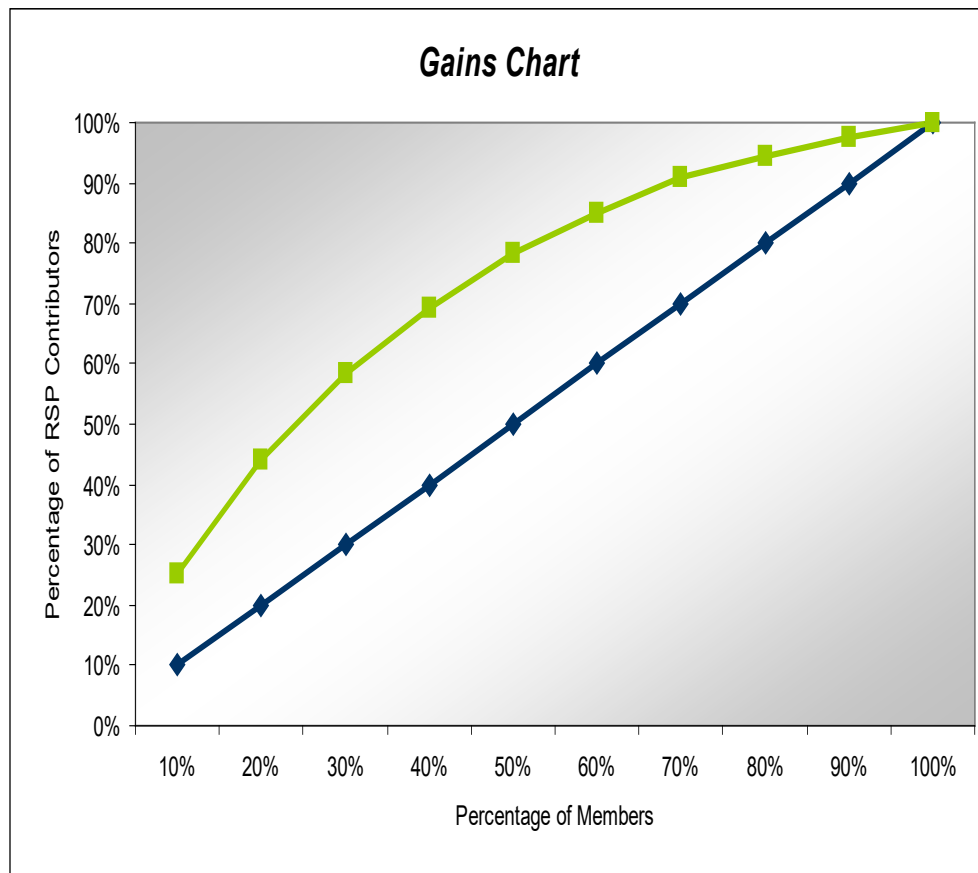
The key steps are those that advanced your data cleaning, model and assessment development. Clear communication is important. Your audience for this question understands technical jargon (like the case protagonist, John). For example, to explain how you make R recognize a blank space in a factor variable as NA, you refer to your use of `na.string = ""` in the command `read.csv("vcRRSP.csv", stringsAsFactors = TRUE, na.strings = "")`. Another example is a comment that the data set does not have a specific data issue discussed in Chapter 10 and you thus do not need to do anything about it. Such a comment can show to your technical audience the thoroughness of your work. As a rough guide, aim for 3 pages of 12-point text, plus as many pages as you need for figures, including the final model’s lift chart. Use point form where it seems appropriate (*60 points*).

2. Discuss what the most important five predictors are, how each of them would affect the target, and whether it makes intuitive sense or not. Your audience for this question is smart managers with strong domain knowledge and limited technical knowledge (i.e. John’s bosses). (*20 points*)
3. Using the costs and contribution in Table 2A below, recommend the percentage of members (ie., the “best X%”) who should be contacted to maximize profit after contact costs, and report what this expected profit is. Included a table showing your calculations. (*20 points*)

Your R script and RData files:

Make sure that the script can be sourced in sequence from beginning to end.

The lift (called “Gains” at VanCity) chart from the final model John developed.



Can you do better? It is possible, but difficult!

Data dictionary and financial details below

TABLE 1A: Data Dictionary for *vcRRSP*

APURCH	=Y if RRSP purchase, N if not purchased
unique	Vancity member identification number
age	age
gendf	= 1 if female, 0 otherwise
gendm	= 1 if male, 0 otherwise
pcode	postal code
atmcrd	=1 if owns ATM card, 0 otherwise
paydep	=1 if member uses payroll deposit, 0 otherwise
BALCHQ	average monthly balance chequing account over previous 12 months
BALSAV	average monthly balance in savings account over previous 12 months
TOTDEP	average total monthly deposits over previous 12 months
BALLOAN	average monthly personal loan balance over previous 12 months
BALLOC	average monthly line of credit balance over previous 12 months
BALMRGG	average monthly mortgage balance over previous 12 months
NEWLOC	= 1 if acquired a new line of credit in previous 12 months, 0 otherwise
NEWMRGG	= 1 if acquired a new mortgage in previous 12 months, 0 otherwise
TXBRAN	number of in-branch transactions per month, previous 12 months
TXATM	number of ATM transactions per month, previous 12 months
TXPOS	number of Point of Sale transactions per month, previous 12 months
TXCHQ	number of cheque transactions per month, previous 12 months
TXWEB	number of online transactions per month, previous 12 months
TXTEL	number of telephone transactions per month, previous 12 months
TOTSERV	total number of distinct services (distinct product lines) held
CH_NM_SERV	12 month change in number of distinct services
CH_NM_PRD	12 month change in number of products (e.g. Loans, accounts, etc.)
valsegm	internal VanCity “value segment” code, using an ad hoc (i.e., human-judgement-based) method to rank all members; A = top 5%; B = top 6-10%; C = top 11-20%; D = top 21-70%; E = bottom 30%

GEODEMOGRAPHICS from tax file by Postal Code neighbourhood

N_IND_INC_1	number of individuals reporting income in the PC nbhd.
numrr_1	number of RRSP contributors
numcon_1	number of RRSP contributions (allows for “spousal” contributions)
avginc_1	average employment income
avginv_1	average investment income

TABLE 2A: Costs and Contribution (not the true proprietary figures!)

Contact cost (Mail and glossy brochure production cost) :	\$5.40
Number of potential contacts (members without an RRSP):	120,000
Estimated average contribution per RRSP purchase:	\$215.00