Kellogg School of Management
Northwestern University

S-Mobile:
Predicting Customer Churn

Shu Ying Seng was a member of the first graduating class of the new Master in Analytics program of the National University of Singapore (NUS). In contrast to many of her classmates who had little prior work experience, Shu Ying had worked for the last 7 years for S-Mobile, a leading cellphone carrier in Singapore. Because she had loved working there and S-Mobile had helped her pay for her degree, she had returned to the company half a year ago.

Shu Ying's last job at S-Mobile before she went to NUS was to manage the retention desk at the company's call center. Her team's task was to persuade consumers who called to leave S-Mobile to stay with the carrier. While Shu Ying's team could prove high "save" rates, she had always felt uneasy about this form of "reactive" churn management --- she felt that it trained consumers to threaten to leave in order to get discounts.

One of the key moments during her master's program was when Shu Ying discovered that data analytics provided a compelling alternative to reactive churn management: Instead of waiting until a customer tried to leave the company, the company could be proactive and predict for each customer how much at risk they were of churning---before they ever called and threatened to leave. This seemed like a much better approach because it allowed a company to act before customers were dissatisfied enough to want to leave, and retention offers had a much better chance of delighting customers. After all, how good could a retention offer look if it was given in response to a customer's threat to quit?

Given Shu Ying's experience in customer relations and her new data analytic skills, she now managed an analytics team in charge of improving customer churn. She realized that this was a big task. In the future she might have to change what data needed to be collected, how customers were treated, what plans were available, etc.

Professor Florian Zettelmeyer prepared this case to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. The data used in the case was either publicly available or was simulated. Certain events and names are fictionalized for pedagogical purposes.

The first step, however, was to take the data that existed and to see whether it could be used to identify whether some customers were more likely to churn than others. And if so, what marketing actions and offers could be used to reduce churn.

Shu Ying asked her team to pull data on a random sample of customers in order to build a predictive churn model. The dependent or target variable was whether a customer had churned in the last 60 days. The predictor variables or features described customer characteristics and behaviors over the 4 months preceeding the last 60 days. The idea was to see whether these data could be used to predict whether a customer was likely to churn (see the data descriptions at the end of this document).

The data consisted two datasets (both contained in smobile_churn.RData):

1. smobile: A sample of 10000 customers with a training variable to split the sample into a training sample with 7000 observations and a test sample with 3000 observations.
2. rollout: A roll-out dataset with 8012 obs.

The model was going to be used to “score,” i.e. generate churn predictions for 8012 customers for whom Shu Ying had been authorized to test the proactive churn management program.

The task

As Shu Ying briefed her team she laid out what they would have to accomplish:

1. Develop a model to predict customer churn using the smobile dataset
2. Use the model to understand the main drivers of churn
3. Use insights to develop actions/offers/incentives
4. Estimate the impact of these actions/offers/incentives on the probability of churn in the rollout dataset
5. Decide which actions/offers/incentives to target to which customers in the rollout dataset
6. Evaluate the economics

Assignment guidelines

1. Develop a model to predict customer churn
   - Feel free to use any technique you like to predict churn
   - Make sure that one of your models is a logistic regression so that you end up with a set of coefficients/odds ratios for the interpretation required in step 2
   - You can use a different model for prediction and for interpretation
   - Build the model using training and test and explain your model choice or use caret with cross-validation

   Hint: Don’t forget to exclude the “training,” and “customer” variables as predictors in your model.

2. Use the model to understand the main drivers of churn
   - Report on the key factors that predict customer churn and their relative importance

3. Use insights to develop actions/offers/incentives
   - Actions do not have to be tied to individual variables.
   - Instead, consider variable groupings that make sense, e.g. “Equipment characteristic,” “Customer usage,” etc.

4. Estimate the impact of these actions/offers/incentives on the probability of churn
   - Either
     (i) simulate the effect of a churn driver (see the hint at the end of this section)
     or
     (ii) test the action/incentive/offer in the field
   - Perform any simulation on the rollout dataset.
   - Since you cannot execute a test, describe how you would set up a test and then assume a projected churn that results from the test.

5. Decide which actions/offers/incentives to target to which customers
   - For each action/offer/incentive specify the exact criteria to select customers.

6. Evaluate the economics
   - For each action/offer/incentive make the case for profitability.

   Hint: If you want to simulate the effect of an action, do this:
   - Train your model
   - Create a prediction for the dataset for which for which you want to simulate variable changes (here the rollout dataset).
   - Create a copy of the rollout.
   - In this dataset change any variable values you want with mutate()
   - For example, you want to know what giving people new phones does to attrition? Change eqpdays to a low number!
   - Predict using this modified dataset.
   - Compare the attrition for the groups of customers you are interested in between the original prediction and the new prediction.

Deliverables

1. Please upload an R Notebook (in pdf) describing the above steps.

2. E-mail me a Powerpoint or pdf file containing, for your <u>top</u> action/offer/incentive
   - A description of the action/offer/incentive
   - An estimate of the impact of the action/offer/incentive on the probability of churn
   - The exact criteria to select customers for the action/offer/incentive
   - The economic justification for the action/offer/incentive

   Please send me the slide latest 2 hours before class and please name the file with the first names of your team, so I can later find it on my e-mail and hard drive!

If you have any questions, please don’t hesitate to ask!

Data Description

Variable	Description
customer	customer id
Churn	churn between 31-90 days after observation date
revenue	mean monthly revenue
mou	mean monthly minutes of use
overage	mean overage minutes of use
roam	mean number of roaming calls
changem	% change in minutes of use
changer	% change in revenues
dropvce	mean number of dropped voice calls
blckvce	mean number of blocked voice calls
unansvce	mean number of unanswered voice calls
custcare	mean number of customer care calls
threeway	mean number of three-way calls
months	months in service
uniqsubs	number of unique subscribers in household
phones	# handsets issued
eqpdays	number of days of the current equipment
age	age of first household member (0 means missing)
agemiss	age data is missing - always include together with age variable.
children	presence of children in household
creditaa	high credit rating (aa)
refurb	handset is refurbished
occprof	occupation - professional
occcler	occupation - clerical
occcrft	occupation - crafts
occstud	occupation - student
occhmkr	occupation - homemaker
occret	occupation - retired
occself	occupation - self-employed
travel	has used phone outside of Singapore
retcalls	number of calls previously made to retention team
refer	number of referrals made by subscriber
incmiss	income data is missing - always include together with income variable
income	income (0 means missing)
mcycle	owns a motorcycle
training	Variable to create training/test samples