DB Week 9

You are a senior marketing analyst in a large bank. The company has products found in any typical large bank (checking and savings accounts, loan products, credit cards, etc.). The vice president of the credit card division is concerned with the high rate of attrition of credit card customers. He has asked for your advice to minimize the attrition rate. You have decided to build a predictive model to predict who will attrite. Discuss your decision and modeling process on the following topics:

* What type of model are you building? The type of model that I would be building is a duration model, and this is based on the subject of study which is the attrition rate, or time until failure.
* What is your dependent variable? The dependent variable in this situation is the credit card account associated with customers.
* What data will you pull from the database to build the model? The data I would pull from the database to build the model would include: customer id, age, income, loan information, credit score, length of time with bank, debt to income ratio, credit limit, revolving debt, number of credit cards with other banks.
* How big of a sample will you pull for the modeling? I would prefer the sample size to be statistically valid without including too much data.
* How will you validate your model? The model would be validated using basic goodness-of-fit tests as well as analysis of the usual plots.

Discuss how you would try to convince your manager that duration analysis models are more powerful and more appropriate than logistic regression models.

At the end of the day, management is going to make decisions and integrate data models that impact the bottom line and make relative sense. My strategy would be to start by outlining logistic regression (LR), and what it accomplishes:

* LR is great for showing the relationship between non-linear variables.
* The outcome variable is binary.
* In LR, it is hard to ascertain how the coefficients are related to changes in the dependent variable.

Duration analysis on the other hand can handle predictor variables that are time related, categorical, and continuous. This contributes to the bottom line by having models that are more precise and versatile to the real world situation ie, this model is less wrong than logistic regression. In addition to the predictor variables, survival analysis allows for censoring. In a nutshell, censoring allows one to see a specific event through different time intervals, thus it gives more relevant information.

When analyzing data, the goal is to see the full picture with the best camera possible. While logistic regression gives a snapshot of the data, it is not as vibrant and colorful as duration analysis. Time is a factor with the vast majority of data, and duration analysis has this additional dimension covered where logistic regression does not.

Depending on the data, either tool could be used, but presenting the fullest picture will ultimately provide the most relevant information to affect the bottom line.

Sources:

Allsion – Survival Analysis Using SAS

Ajmani – Applied Econometrics

University of Delaware