

## Case Questions for Cell2Cell Part 1

### *Case Background*

Telcos are in the middle of a dramatic change in their average revenues per user; increased competition along with similarity of offerings and diversity of cheap phones have created a lot of customer churn in their user base. Cell2Cell is a telecommunication firm that seeks your help to manage its churn. Your task is in Part 1 of this exercise is to (a) develop a statistical model for predicting customer churn, (b) use the model to identify the most important drivers of churn, and in Part 2 of this exercise you will use these new insights to recommend a customer churn management program to the CBM (Customer Base Management) Group.

Currently, Cell2Cell uses mass marketing techniques for building brand awareness and acquire customers. While the explosive growth of cell subscribers helped this strategy, as the growth slowed, the net additions minus exiting customers has fallen into the red. To counter this, the CBM group is trying to increase revenue per user, up-sell new products and decrease churn.

When churn was first measured, the rate of customers leaving the company was close to 5% per month, refreshing about half the subscriber base every year. In response, Cell2Cell implemented a reactive retention program when customers called with the intention to leave. While this was partially successful, a truly data-driven churn management strategy that is proactive is still missing.

The client recommends a three-stage process for implementing the proactive churn management strategy. First is to develop an accurate predictive churn model. Second is to identify the most important factors that drive subscribers churning. Third, use the insights from the second stage to identify offers that should be targeted to customers with a high (at least 75% higher than average) risk of churning. Offers might include cash incentives, price discounts, or any other enticement that could be expected to work best. There is no need to target the same incentive to each of the high-risk customers, although the offer should be financially sound.

To assist your prototype analysis, your contact Sarah has assembled a “calibration” database consisting of 40,000 customers and a “validation” database consisting of 31,047 customers. Each database contained (1) a “churn” variable signifying whether the customer had left the company two months after observation, and (2) a set of 75 potential predictor variables that could be used in a predictive churn model. Following usual model development procedures, the model would be estimated on the calibration data and tested on the validation data. At the time, Cell2Cell’s churn rate was about 2% per month. However, Sarah created the calibration database so that it contained roughly 50% churners. This was to make it easier for whatever data mining tool she used to identify the factors influencing customer churn. The validation data contained 2% churners. The data are available in one data file with 71,047 rows that combines the calibration and validation customers.

### *Instructions*

This exercise should be completed in your assigned group (or if you are unable to work with your team you may complete it individually).

The questions for this exercise are based upon the assigned reading for this class: "Cell2Cell: The Churn Game". Please complete this reading before answering these questions. In this exercise you want to design a model that will predict which customers are most likely to churn (e.g., not renew their mobile phone contracts) and discuss the implications of this model. The data is provided in the file “cell2cell\_data.csv” and an R script to perform a suggested analysis called

“cell2cell\_Part1.R” has been provided on the course website. Additionally, definitions for the variables is provided in “cell2cell\_doc.csv” and “cell2cell\_doc.xlsx”.

Please provide a clear, concise, and well organized essay that addresses at least the following questions. You are free to address other issues in the case as well. The intent of the assignment is to have you think critically about the business problem faced in the case and how it can be solved through using a predictive model. Analyze the quantitative material in the case to support your answers. Spend most of your time in defining and defending your recommendation for what should be done.

Good answers may require assumptions of facts that may not be presented in the case. You are welcome to make these assumptions, but please state these assumptions and briefly justify why that are reasonable. Also, you may use whatever resources you can locate to provide further information about this industry or the web in general. Please reference your sources.

Your response must be typed, double spaced, with one-inch margins, and a 10 to 12 point font size, and must be a PDF file. This writeup must not exceed 3 pages in length. You may attach exhibits, tables, and/or graphs to support your arguments. These supporting materials must be referenced in the text and do not count toward the 3 page limit.

### ***Required Questions for Part 1***

Analyze customer churn using Classification & Regression Tree (or Decision Tree) and a Logistic Regression using the information provided to you. To help you complete this analysis briefly answer these questions:

1. Purpose – what is the marketing purpose of your task? How do you think the company could use your results to target customers who are likely to churn? Before you being the analysis name three relationships that you expect to see between churn and the predictive variables. (Hint: focus on the direction of their influence, do you think a high/low value of this variable will result in more or less churn?)
2. Estimate at decision tree that predicts CHURN. Prepare a graphic, visualization or table that summarizes the relationships that you have found.
3. Estimate a logistic regression that predicts CHURN. Prepare a table that summarizes the relationships that you have found. At a minimum include the following columns in your table: Variable, Parameter Estimate, Importance, and Meaning. The meaning should provide a description of the effect in plain English (e.g., if the parameter for “Eqpdays” is “0.0010527”, you could make a statement like “For every extra month (30 days) that a customers odds of churning versus not churning goes up by 3%.”).
4. Which model do you find is best? Be sure to justify your assessment based upon the performance of each of these models in the test sample using a confusion matrix and/or lift in the top decile.