

Cell2Cell Churn Management Assignment

This case requires you to develop a model for predicting customer churn at “Cell2Cell,” a fictitious wireless telecom company, and use insights from the model to develop a proactive incentive plan (a proactive churn management marketing action or offer) for enticing would-be churners to remain with Cell2Cell.

The data for the case are a scaled down version of the full database generously donated by an anonymous wireless telephone company. There are still 71,047 customers in the database, and over 65 potential predictors. Logistic regression is the best choice to develop your predictive model.

This case requires both statistical analysis and creativity/judgment. I recommend you not spend too much time fine-tuning your predictive model; make sure you spend enough time interpreting and using the results to come up with your plan.

Questions for the write-up and for class discussion

Your task is to execute the 3-stage process for **proactive churn management** as stated by Charles R. Morris on page 7 of the case. In your write-up, please include all the steps specified in “How to proceed” and answer the following questions:

1. Briefly describe your predictive churn model. How did you select variables to be included in the model?
2. Demonstrate the predictive performance of the model. Is the performance adequate? (Hint: Page 22 in the UCL Class 13 2019 lecture note. Please check the most recent one on Moodle.)
3. What are the key factors that predict customer churn? Do these factors make sense? What is their relative importance?
4. What offers (incentives) should be made to which customers to encourage them to remain with Cell2Cell? Please specify at least 3 incentive or offer plans in detail.
5. For a least 2 plans, simulate the effect of your offer/incentive/action on the predicted probability of churn. Also show that your offer/incentive/action is profitable. How much can you spend on your offer/incentive/action?
6. How much reliable are your offer/incentive/action plans?

The Data

- The data are called “cell2cell_UCL.dta” and can be found on the class website.
- The data are documented in the “Cell2Cell Data Documentation UCL.xls” spreadsheet, which is on Moodle.
- The data consist of 71,047 customers divided into training/validation and representative sub-samples contained in the data file called “cell2cell_UCL.dta”. The training/validation sample contains 40,000 customers, 20,000 of whom churn. There is a 60/40 split in the training/validation sample. The training sample has 24,000 customers and the validation sample has 16,000 customers. The representative sample contains 31,047 customers, 609 of whom churn. The 50% churn rate in the training and validation sample is not realistic, but the “over sampling” of churners makes it easier to identify the profile of churners as distinct from the profile of non-

churners. The representative sample contains 1.96% churners, which is the current monthly churn rate at Cell2Cell. (On page 7 of the case, it says "calibration database consisting of 40,000 customers and validation database consisting of 20,000 customers". **But please ignore this and follow the sample classification in this assignment guideline.**)

How to proceed:

Run a Logistic Regression Predictive Model

1. Read in the "cell2cell_UCL.dta" file.
2. Run a logistic regression model with "churn" as your dependent variable. All the variables from "revenue" to "retcall" can be used as independent variables. You can select all of them by typing the name of the first variable you want to include, then a dash ("-") and then the last variable you want to include in the logistic regression.
3. Make sure you only run the regression on the training sample (by using "if training==1" in the regression command).
4. Think about how to put weights into your logistic regression (see Class 13 Retaining Customers lecture note).
5. Check overfitting.
6. Use the model to predict attrition probabilities.

Calculate the lift

You should use the ***lift for the representative sample*** to determine if your model is adequate (which you should take to be a 70% higher than average probability of churning, i.e. a lift of 1.7 or 170%; on page 7 of the case it says 75% (or a lift of 1.75, please ignore this). If you can't get a lift close 1.7 or 170%, pick the decile with the highest lift to which to apply your churn management suggestions.

You can use the "lift" command in Stata to calculate the lift. (See the "User Menu" / "Marketing Menu" / "Model Performance." If you cannot find the command, please update your menu.) Please make use of the option in the command that allows you to calculate lift **only for the representative sample**. You might be confused about which options you should choose for the lift command in the Stata menu because the options say about "training" or "validation", not "representative". Because we need to get lifts for "representative sample", choose "representative" variable for "Distinguish between training and validation sample" option and choose "training sample only" for "Graph lift for" option. "training sample only" option means that you want to get lifts for the representative sample because you specify "representative" variable in the option as a "Dummy variable describing training (1) and validation (0) samples:" and training (1) indicates the condition that representative = 1.

Notes :

- The graph created by Stata is the “cumulative lift.” You want to know the “lift,” not the cumulative lift. This is displayed in the table displayed by the lift command in the results window but not graphed.
1. Please report the *lift for the representative sample* for all deciles and choose the first decile with the highest lift to apply your churn management suggestions. You are required to suggest multiple marketing programs. Therefore, regarding your marketing offers/incentives/actions that can be applied to all customers, you can mention that and decide to target all customers, not just for the first decile, for those marketing programs.

Determine and Rank the Economic Importance of the Predictor Variables

1. Determine the economic importance of each variable for predicting attrition. This calculation is best done in Excel. You can find the standard deviation of each variable in the spreadsheet “Cell2Cell Data Documentation UCL.xls”.¹

Create an Incentive Plan and Develop a Way to Estimate the Impact of Program

1. Looking at the importance-ranked list of predictor variables, decide which of them suggest a retention marketing action by Cell2Cell.
2. (Insights to develop incentives) For each “actionable” predictor variable, specify what retention marketing action you suggest, i.e. what type of incentive you plan to give to consumers to encourage them to remain with Cell2Cell (e.g. new phone, rebate of \$X, new plan, etc.).
3. (Incentive/retention Plan) Please specify at least 3 incentive or retention plans in detail. Specify the exact rules for making incentive offers to customers. For at least 2 incentive plan, you should simulate the effect of a churn driver on the probability of churn. If you can simulate the effect of a churn driver using your logistic model, you can calculate a projected churn rate.
4. Determine whether your incentive/marketing/offer is profitable based on LTV calculations.

Hint: In determining how much you can spend on incentives notice (1) that revenue data are in the dataset, and (2) the marginal cost of providing cell phone service is 0.

5. Do you think that your conclusion based on simulation and LTV calculations is reliable? If you think that it is reliable, explain why. If you do not think that it is reliable, explain why it is not reliable first and discuss what you can do further to make a final decision on whether to execute each of your churn management plans.

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

¹ You can also calculate the standard deviation for each significant variable in Stata and then cut and paste these standard deviations into Excel. You can use the following command in Stata to get the standard deviation for all variables:

```
tabstat revenue-retcall, statistics(sd) columns(stat)
```