The Power of Detailed Data

How detailed data can drive more powerful targeting models

Dr Barry Leventhal, BarryAnalytics Ltd

How often do you hear the question, “What’s more important for a good analytical model – the technique or the data?” And the usual answer, “Having strong predictive data, of course!” Yet it seems surprising that when many marketing analysts build models, they start from standard data extracts containing aggregated variables rather than exploiting the detailed records that fill up considerable space in their data warehouses.

The reasons for relying standard extracts are clear and understandable. Looking at retail banking as an example, a typical data extract in this sector will contain around 500 to 1000 variables aggregated to customer (or account) level. This data set will cover topics such as customer attributes, product holdings, volume and values of transactions by product category, channel, type and time period, and so on. Such an extract will have taken a substantial amount of IT time to develop and so, once working, tends to be produced every month with hardly any modifications.

This extract provides a good general summary of customer behaviour, but may not contain the most predictive variables when the analyst is trying to model a particular activity, such as which customers are most likely to close their mortgages? In this situation, some specific mortgage-related factors may be more valuable, such as mortgage size, monthly repayment, or interest rate vs. competitors.

In the case of mortgage attrition, the need for detailed inputs may seem obvious, and bordering on 'teaching Grandma to suck eggs', however the same holds true for any situation where predictive variables are required.

So how could detailed data help marketing analysts to build more powerful targeting models, and how much difference could this make?

Detailed data comes in different flavours, depending on which industry you work in – some examples are given in the panel below.

	Some Examples of Detailed Data
Retail	POS transactions Market baskets
Finance	Debit and credit transactions
Telco	Call Detail Records (CDR's)

However the common theme is that detailed data provides individual instances of customer behaviour or interactions. Just as person or household classifications are more powerful than area classifications, detailed data is more predictive than summary data. This predictive power can be harnessed in several ways, for example:

- Codes may be built to indicate usage preferences; for example a bank may wish to hold each customer's preferred channel for making transactions, in order to help target communications and plan resources.

- Significant behaviours or 'events' may be identified, which can be used to trigger an immediate follow-up activity by the organization. For example, a customer pays an abnormally large amount into their bank account, which could usefully be put on deposit for them.

- Detailed data can improve the performance of analytical models, resulting in more accurate customer segmentation and targeting.

All of these applications will involve some ‘up front’ analytics to understand the detailed data, design and test different ways to transform it, and create analytic datasets for model building.

The key benefit of using detailed data for analytics is its flexibility – starting at the finest level of granularity, any number of behaviour patterns can be built that may help to predict the target outcome. The decision as to which measures will be useful requires an understanding of customer behaviour in that industry, combined with analytical expertise.

In the mobile phone industry, for example, churn analysis may benefit from knowledge about each customer’s ‘calling circle’ – such as how many different numbers are regularly called, what proportion of call time these account for, and whether these numbers are on the same network or with a different operator. These measures can be obtained from analysis of call detail records (CDR’s) but could never be derived from aggregated data extracts.

However, after initial exploration has identified the key ‘calling circle’ metrics, those variables can be calculated for a sample of churners and non-churners and used as part of a churn model.

The use of detailed data in this way should not alter the actual model-building process – which will still include stages for developing the model on a representative sample of churners and non-churners, and validating on a hold-out sample, prior to scoring the population of eligible customers.

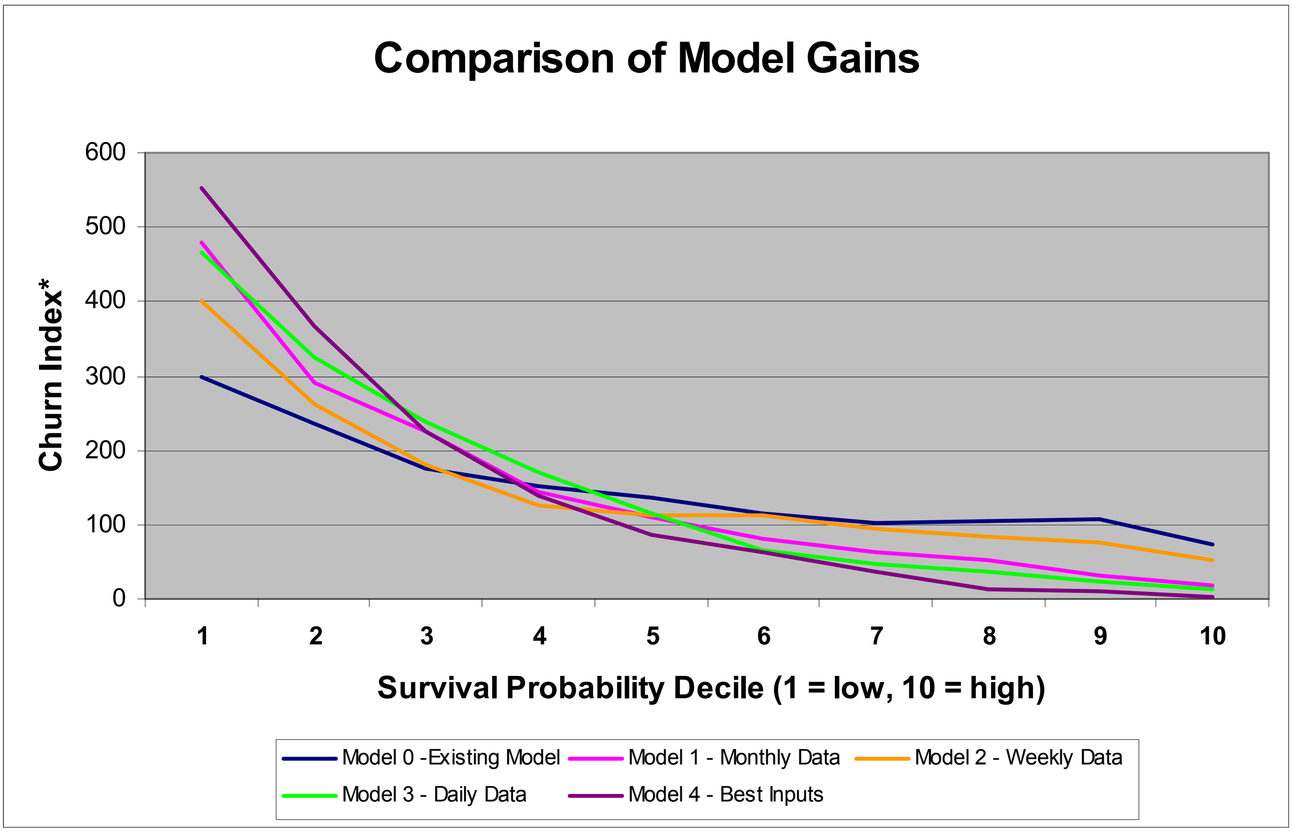
Example

A customer lifecycle model was built for a European mobile phone operator, using one segment of its customer base. The aim of this model was to predict survival probabilities over time for each individual, so that the predictions could be employed as part of the customer management process. The model was developed using a standard set of inputs, including monthly summaries of inbound and outbound calls, along with other variables including customer demographics, handset features and payment data.

Having built the initial model – which we'll name 'Model 0' –an extract of CDR data was employed to create additional sets of detailed call variables at monthly, weekly and daily levels. The variables were subjectively chosen by the analysts and were not optimized in any way. Using these variables in addition to the variables employed in Model 0, four further models were developed:

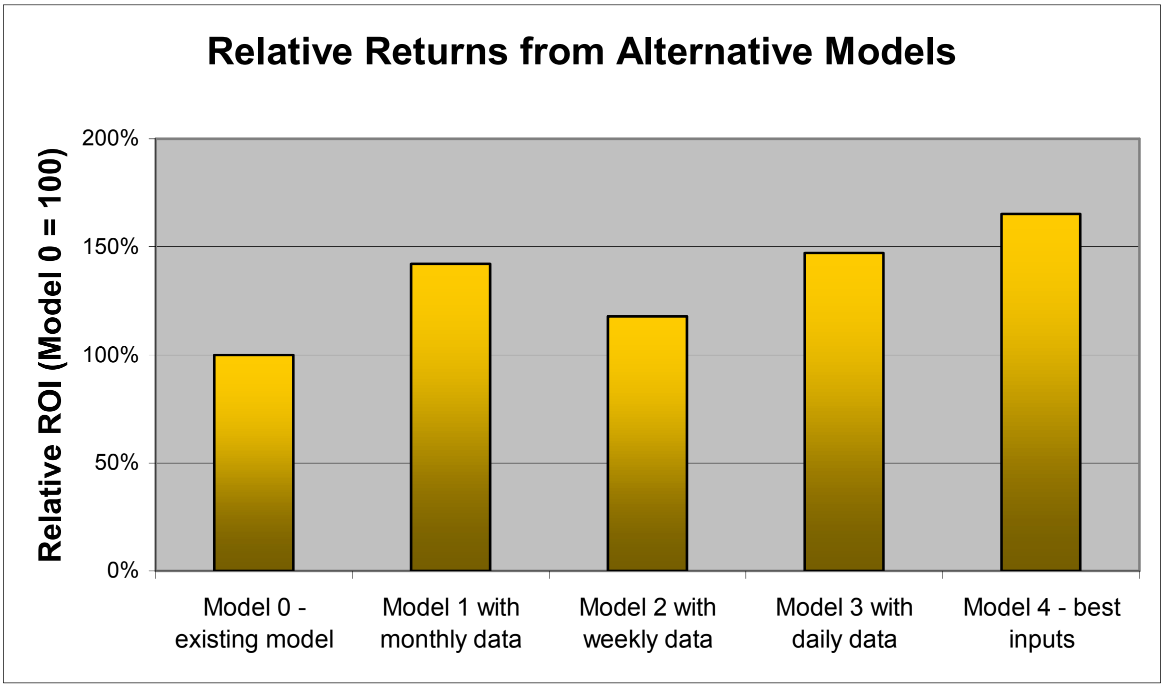
- Model 1 – including monthly variables built from detailed data
- Model 2 – including weekly variables built from detailed data
- Model 3 – including daily variables built from detailed data
- Model 4 – including the best inputs from all of the models

All of the models that used detailed data were dramatically more powerful than Model 0 – as can be seen from the comparison of lift curves below. Although the detailed variables did not replace all the Model 0 variables, the detailed data had massive importance in each model (with the exception of Model 2, where the weekly detailed variables were not quite as strong).



* The churn index scale has been disguised.

The benefits for churn management were assessed based upon assumptions made by the analysts. The resultant returns on investment are displayed below, relative to the existing model – showing that the models using detailed data all generate significant improvements in ROI.



The Way Forward

This article has discussed some of the benefits in using detailed data for building targeting models, and has demonstrated that the gains can be significant. As a result of recent developments in data mining software, detailed records can be investigated and converted into analytic data-sets more easily and effectively than ever before. This should come as good news for database marketing analysts!

Dr Barry Leventhal is director of independent London-based consultancy BarryAnalytics.
Tel: +44 (0)7803 231870
barry@barryanalytics.com