

The Hype and the Reality of Database Marketing Software

*By Boris Gendelev
Principal, Wheaton Group*

Original version of an article appeared in the September 12, 1994 issue of "DM News"

[Note: Despite dramatic increases in raw computing power and a proliferation of end-user software tools since the publication of this article, virtually all of the content remains highly relevant.]

You've heard the pitch: for you, the seeker of new database marketing heights, there is a software package that will take you there. With a click of a mouse, you will count and profile your customers, select your names and, when you are done mailing, analyze your results. Wow!

But after the initial thrill is over, you might be disappointed for one or more reasons:

- Data manipulation capability is not powerful enough.
- Data-driven modeling is supported only superficially.
- No training is provided for navigating in the dangerous waters of data analysis.
- Nobody addressed the issue of data integrity.
- Simple queries may be simple, but complicated ones are nearly impossible.
- Even if you manage to express your complicated question in the language of the package, to get an answer in a reasonable amount of time takes expensive and exotic hardware.

Of course, nothing can be perfect. But to minimize disappointment, ask the fundamental questions.

Is Important Functionality Missing?

To a direct marketer, automated counts and selects are very important. For a database marketer, that alone would not do.

Database marketing is first and foremost the process of using customer history to predict, under different scenarios, future productivity. So defined, the practice of database marketing centers on understanding the relationship between what was known about a customer at one point of time and subsequent results.

Because database marketing is about data-driven models of customer behavior, counting and profiling only qualifies as the first stage of the modeling process – getting familiar with a business and its data.

Marketers new to database marketing are likely to stop at counts and profiles. Does this mean they do not use models of customer behavior? Of course they do – their models are judgmental, not data-driven.

Judgmental models reflect one's intuitions, the subconscious sum of one's experiences. There is nothing wrong with that. Yet, when there is hard data to give or deny credence to a hunch, it makes business sense to use it. Software that doesn't go much beyond counts and profiles doesn't unlock the full potential of database marketing.

Can a Software Package Really Build Models?

Many packages claim to perform modeling – feed it your mailing results and after cranking away with logistic regression, neural network, fractals or some other statistical wizardry out pops the result.

But all they are offering you is model calibration, the calculation of parameters. Who decides what variables to toss into the magician's hat? You do! But how? By intuition alone? Is that data-driven? No! Is that true databased marketing? No!

The process of modeling is foremost the process of deciding, by business and data analysis, what data to use and how to transform it to illuminate patterns:

- Modeling subset – What time frames and business segments are relevant? Last fall? This spring? All spring seasons? Last five years? General media? Specialty media?
- Dependent variables – What should you try to predict and how it should be measured? Response Rate? Average Order? Demand per media? Per marketing dollar? Return Rate? Net sales? Long term value?
- Independent variables – What variables should be tried as predictor variables? Here, the list of possibilities is endless, considering combinations of variables (differences, sums, ratios, percentages). The creating and testing of predictors is where interesting data analysis takes place.

Once past the first round of these questions, you will be ready to calibrate a model. Then you have to validating its performance. The results might suggest fine-tuning and send you back to data analysis in search of new ideas.

Model calibration does not develop new concepts. You have to. Software that restricts you to variables you were foresighted enough to record during your mailing select is a package that hinders the practice of database marketing.

Will a “101” Standard Report Be Enough?

It might well be that the bulk of marketers' daily needs can be satisfied with a stack of standard reports and "fill-in the blanks" queries. Yet, just as surely, the remainder – reports produced ad hoc in search of understanding of changes from business as usual – are what provide your company with a competitive edge.

The standard reports help you monitor the business through the lenses of your existing models (intuitive and statistical) as well as monitor the robustness of the models. They serve to trigger new questions. The ad hoc, never-anticipated queries help answer the questions and move both the models and your business forward.

Therefore, while prepackaged report templates are often useful, good database marketing software should excel in ad hoc reporting of any depth and complexity. If the answer can be found in the data, the tool should give you power to formulate the question.

How Do You Avoid Discovering Useless Things?

The barrier to building robust actionable customer behavior models cannot be overcome by software alone. Data analysis expertise is equally essential.

The data, not the software, interacting with an analyst's logical faculties and imagination, drive the course of analysis. The decisions based on the analysis, once set in motion, may have profound, irreversible and long lasting impact on your business. Training and experience in data analysis and interpretation is what stands between you and disaster.

The modeling process is full of potential pitfalls. Two examples are:

Example #1: In the exploratory phase of modeling, there is a danger of selecting predictor variables that are contaminated by the dependent variable.

Suppose you have a hunch that customers with children are your better buyers. You decide to add a question to your order entry script and a new field – "presence of children" – to your database. The field is initialized to "no." After a while, you start analyzing if those who answered "yes" bought more frequently. And sure enough, "yes" customers are more frequent buyers than "no" customers.

Did you just find an important key to your business? Before you start paying for demographic overlays and over-circulating households with children, consider this: those who buy more frequently were more likely to be asked and give an answer. Therefore, being a frequent buyer makes a "yes" more likely, and not necessarily the other way around.

The specific lesson: a single code should not mean two different things. In this example, "unknown" should be a separate code. Moreover, a new segmentation variable must be evaluated while holding

constant other variables already known to be good segmenters; for example, RFM or your current scoring model.

A more general lesson: without being keenly aware of how your business is reflected in the imperfect mirror of your data, and how to evaluate an incremental value of a new idea, it is easy to "discover" useless things.

Example #2: In the calibration phase of modeling, exposing the statistical procedure to the "best" cross section of data is complicated.

Techniques that produce black boxes are particularly troublesome. A black box might fit the data used to build it, but as your business evolves and produces previously rare combinations of variable values the black box may start to spew out nonsense.

Suppose your new software presides over a newly built database. While your analyst knows to be careful due to having access to only eighteen months of data, did anyone tell the software? How do you tell the software that some customers are three, four, or ten years old and their history is partial? Will there be problems a year from now when the maximum recency is thirty months?

Neural net techniques are particularly vulnerable, because in addition to producing inscrutable black boxes, they demand a trade-off between computing power and sample sizes. Without specialized add-ons, several thousand cases may be the sample limit. For a good size business, this might be only several percent of the buyer file. Samples are likely not to contain the full range of variable values and their combinations.

True, better packages use sophisticated mathematics in order to avoid overfitting. For example, they could insure that higher frequency of buying consumables always leads to a higher score. A neural network can easily produce a result that contradicts this common sense reality and risks losing you sales.

A capable package without expert data analysts to train you and to help you use it is at best an incomplete, and at worst a dangerous, solution.

Who is Verifying What and How?

As previously discussed, it is easy to be misled by poorly conceived and carelessly executed data analysis. But even full awareness of the business, data and analytical issues does not protect you from simple mistakes. Because you will have to translate your thoughts into organized instructions a computer can execute, there is plenty of room for error.

The art of result verification is a specialized branch of data analysis. Multi-million dollar mistakes await those who believe that computers print nothing but gospel. Certain software features may prove helpful to minimize the occurrence of error, but again, training is the real answer.

Who Minds Data Integrity?

A database marketing package without data is an empty shell. With data that is inaccurate, incomplete or inconsistent, it is a ticking bomb.

In checking data integrity, there is no avoiding a human expert. But beware! Often, people who can make sense of system and file structure do not have a clue how to analyze and interpret data. Assuring data integrity is a data analysis task not a programming job.

Should you believe MIS when they say the data is clean and all you need to do is load the files? If they are not familiar with the tools and techniques needed to answer your business questions, how were they able to query the integrity of the data? How much has been learned from an audit of the data? If many basic facts about the business await the running of the "standard" reports, then on what basis did MIS reach its clean bill of health.

When it comes to construction of a database with accurate, complete and consistent data and, just as importantly, development of a process that maintains that integrity, it is hard to imagine a cookie cutter solution. Nor is it advisable to entrust anybody other than business-aware, computer-literate, seasoned data analysts with the task.

How does the Interface Deal with Query Complexity?

GUI's (graphical user interfaces) can be wonderful. They allow you to construct queries by pointing and clicking instead of typing. And for a two-finger typist, it is a relief. The problem is that most GUI's make it easier for you to construct 4GL (4th generation language) statements, but still require you to know the technical concepts of their system. They offer step automation by demanding less effort for mindless steps. You would rather the interface achieve a conceptual shift – elimination of whole groups of steps that, if it were not for the need to spoon feed the computer, would not even be a part of the way you think.

Electronic ignition instead of a crank saves muscle energy. However, you still have to start the engine; you cannot just jump in and step on the gas. Entering a query by dragging and dropping saves keystrokes, but does not eliminate the responsibility for learning the effect of what you are dragging and where you are dropping.

A conceptual shift is vastly more powerful. It is automation or even the complete hiding of a whole set of steps that together are conceptualized as a single task. It saves not only physical but also mental energy. Manual vs. automatic transmission is a good example. With an automatic transmission, you make choices from a limited, high-level set of options. You do not need to know that there is such a thing as transmission, let alone what gear is appropriate for what speed. A box that decides without your involvement of when, and how, to shift leaves more human brain neurons focused on the business of avoiding other automobiles and getting to your destination.

Marketing data analysis, arguably one of the more complex computer-assisted endeavors, is challenging enough in its own right. The more the software allows you to "show" or "lead" it to a desired final result, in a manner natural to you, the better. Approaches that have you tell the software how to apply its own (foreign to you) operators lengthen the learning curve, increase the likelihood for mistakes and rework, and reduce the time you have to apply your brain to more creative activities.

Are You Spending Too Much on Hardware?

The real question should be: "Is the software optimized for the task to get the most out of the hardware?" Only if the answer is "yes" can you make intelligent trade-offs in cost vs. performance.

Unless you like expensive big iron, steer clear of the "state-of-the-art" in database technology – relational database management systems (RDBMS). On many levels the RDBMS model is a poor fit for database marketing. Inherently, RDBMS's are extremely inefficient for large numbers of complex calculations, processing of many linkages between records (joins, in RDBMS terminology), and massive aggregations (projections).

True, indexing can speed up certain kinds of joins and projections – those that are done on indexed fields. But indexing all of the fields is contingent on having lots of disk space. More importantly, the fields you may want to use in a join or a projection are probably calculated on the fly. They represent new concepts you want to try, or expressions that evolved along with your business. How could you index the file on such variables? No one should expect you to!

Indexes are useful for quick access to predefined aspects of the database much like you would use a library index. But, if your goal is to describe how form and content affect a book's popularity, a library index of title, author and subject would not be helpful. You would need to start reading books and their reviews and summarize your findings in ways that might be useful for cajoling the patterns. As you generate new hypotheses, you go back and reread the whole pile.

If a software vendor touts indexing capabilities, remember that you are in the business of reading all the books and related materials. The books better be arranged in a way that makes it easy to pick up one as well as related material, study them, add its gist to all of your different summaries, and move on to the next. Be prepared to go through this process over and over again. The number of times you have to get up and search for something on the shelf (disk access, particularly random access) should be cut to the minimum and your work should all be in one quick access working area (RAM).

With properly optimized software, a Pentium PC can produce performance on par with speeds claimed by some software vendors to be "incredibly fast" on hardware that costs 20 times more. When possible, spend your money on data analysis and modeling, not on hardware.

Was the Testing Ground the Same as Your Battle Ground?

Look at the past and present clientele of the software vendor. Ask which businesses were used as test beds for prototyping and stress testing the software. If the industries or companies mentioned are not known for their database marketing expertise and use of accountable marketing media, what makes you believe the software is appropriate for you?

Are These Questions in Order of Priority?

The questions above are applicable to most potential users of database marketing software. But specific circumstance of your business will dictate just how important each question is and generate other specific questions. The important thing is not to be overtaken by hype.

Boris Gendelev is a Principal at Wheaton Group, and can be reached at 847-205-0916 or boris.gendelev@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts.