

John R. Miglautsch, PDM

There's Such a Thing as Too Much Data

John Miglautsch is founder of Miglautsch Marketing, Inc. which provides database marketing consultation and modeling. He received his MA and BA from the University of Wisconsin-Madison and PDM from the University of Missouri. He is the founder and Executive Director of the International Society for Strategic Marketing with members in 70 countries. John is on the Board of Directors of the Wisconsin Direct Marketing Club. John speaks at Marketing conferences all over the world. He was a major contributor to the 8th Edition of *Successful Direct Marketing Methods*.



ABSTRACT

ISSM's Executive Director shares his thoughts on The wide gulf between the techniques advocated by leading database marketing experts and the simple mailing methods of even the largest direct marketing practitioners. With cheap disk drives, it is possible to build a massive pile of data which may not dramatically improve your marketing success.

Whatever statistics and reporting, you must still have the ability to spot anomalies. This may mean fancy query tools, but for most of us, it means programs which allow us to select customers and look at their related data. Summaries will always cover up the facts. Make sure there is a way to look under the covers. If you must build a marketing database, focus on the crucial questions. Make a list of things you want to answer . . . make it very specific.

When I began my career in direct marketing consulting in 1981, I would show clients how to modify their computer systems so they could gather marketing information - keycodes, average order, etc. In those days, we made decisions by the seat of our pants because there was precious little hard data to go on. In the mid-80's things changed dramatically; clients started asking for help in understanding all the data they had accumulated. We typically couldn't get at the real data, so we asked for totals from the data-processing department (that's what it was called before IS and MIS).

Of course, each report came back looking a bit different than we expected. We always went through several versions before they understood what they wanted. At the time I thought it was because there was a language problem between marketing and MIS. But then I noticed that even when we got what we wanted, additional information opened new vistas and generated more questions. The new questions suggested new report requests and the cycle began again. One day I realized that whether we got what we asked for or whether we didn't, we still seemed to "need" more information.

The problem is not really in the data, if getting more data only expands the scope of the problem. When the time comes to make a decision, we're a bit more timid about facing the future. Decisions, after all, have consequences. We tell ourselves that if we knew with certainty that one additional number, we'd be comfortable making the decision. "But we deceive ourselves and the truth is not in us."

The fact is that mountains of data can immobilize corporations. Think about the U.S. government if you want a good example of endless data, (and you can be a judge as to the effectiveness of the government's decision-making abilities). Much has been made of the emerging Quality movement in America. "You will never control what you do not measure" leads us to believe that total quality management means measuring everything. Yet companies that win the Malcolm Baldrige Awards typically struggle for years afterward. A colleague who does ISO consulting says that they become so fixated on measurement, they often lose sight of their core business.

If you think, marketing databases are simply about data, you'll make terrible mistakes when you build one. One firm has about 14 million customers in its mainframe, and about 300 million transactions. Another has 27 million customers and a billion transactions. Both have said they needed real-time access to all that information. Another actually put 11 million customers in a special data warehouse, but it is now so cumbersome that they can't make sense of it and they are looking for consulting help. All these and many other companies are filling disk drives without thinking about what exactly they want to accomplish with the data.

This means that data quality and flexibility are far more important than the sheer volumes of data. For the moment, the 14 million name firm is working with several smaller databases and finding that, with only a few exceptions, using less than 100,000 names works very well for analysis.

Marketing is perhaps justifiably suspicious about settling for less than everything they think they need. For years, their only access to data was through the veil of paper reports. When they finally got crude query tools, they found out that much of their detailed data was fundamentally flawed. Because no one had been using or looking at the details, no one realized the problems that were being created. If not seeing all the data leads to corruption, then it "follows" that we must always look at all the data for every analysis.

Unfortunately, this misses the point of analysis: that patterns and trends are easier to spot in a limited and controlled environment. You build a model plane to put it in a wind tunnel, you can use a smaller tunnel, and it is much easier to hang the small plane. The point of models is to simplify complicated things so decisions can be made.

Historical data will never exactly match what you're trying to do in the future. Economic, competitive and a multitude of other variables will have to be changed. On the whole, the concept still works, but the tighter you try to match past and future, the harder the process becomes, and it's very easy to come to the wrong conclusions. A blood center spent thousands on a penetration analysis to find out that the vast majority of its donors came from very few lifestyle segments. They failed to realize that they had already targeted all their promotions to those types of people in selected neighborhoods.

Let's say you want to sell jewelry to your consumer customers. You start by looking at who has bought cosmetics in the past. You use modeling tools to see how these people are different from those who didn't buy. But this only works if you mailed broadly in the past. If you used this trick before and only mailed to people you thought would be likely prospects (let's say you mailed only to women), then you must be careful about what your current model will tell you (past buyers won't, for example, include men because they weren't selected for previous offers). Though your database consultant might not admit it, segmentation analysis works best with sloppy clients, who have been mailing to everyone in the past.

Given some of the limitations of traditional statistics, users must be able to examine data that does not yet create a trend. With regression, we generally look for correlation (i.e., people who buy widgets are older than people who buy gizmos). My stats friends tell me they'll often remove odd data (like the person who ordered \$10,000 when your average order is \$100) in order to build more elegant models. However, some of the best marketing successes have come well before the trend became obvious (i.e., photo copiers, portable computers, Federal Express, Pine Sol and Arm & Hammer in the fridge). To understand the difference between irrelevant static in your data and interesting exceptions that reveal a new marketing breakthrough may take a human brain. There will always be room in database marketing for a good gut feel. A bad offer will seldom be saved by analytically finding enough "suckers," and a good offer will be broad enough to overcome a few bad list choices.

Often clients want the database to answer questions which do not even have an empirical basis. A major auto maker started a credit card which offered a rebate on a new car. After acquiring several million accounts, they asked if we could identify who would have bought their cars without the extra rebate. Now 'would of's' and 'could of's' are particularly challenging . . . a bit like asking if I had done better in life had I become a retailer rather than a marketer. That is to say, you can't split test your own life, you have to make decisions and choosing one thing eliminates all other possibilities. Your company may choose a direction and it may not be possible to test it (the offer goes to all customers and none can be held out as a control). This isn't to say we can't look back and make some educated guesses (but it is very hard to lose a bet on what might have been). It is never a level playing field; though we try to make every test statistically valid, without theoretical underpinnings, each test is only one event at one time.

Can Yesterday's customer data solve the issues? If these monumental problems are not at least considered as you think about your database design, you can be asking and answering the wrong

questions. Many marketing database experts give the impression that if there were just enough information, we could mail only to those who are likely to respond. An editorial in Telemarketing suggested that with enough information, we could afford target phone calls and be rid of "junk mail" forever. However, to be realistic, I don't even know exactly when I will buy particular items. I didn't buy my new van until that little red light started blinking on my low mileage special. Neither my mechanic nor I were able to predict the little light (turned out not to be serious but I bought a new car anyway). If I cannot predict my own behavior, then how are you going to fill up your database to predict it. The bottom line is that we will never achieve 100% response rates. We will always need to over promote (to communicate with people who don't yet know that they want our product) with great marketing offers.

The Great Middle

We have been talking about some of the areas which database marketing purports to address. Yet, at the same time, without some careful thought, the data technology can often cloud more basic issues. Perhaps this explains the wide gulf between what the priests of DM are preaching and what the parishioners are practicing. (See DM News, July 25, 1994 p. 2). Once we understand some of the limitations of DM, we should look at the fundamentals one last time. If the issue is how to promote most efficiently, we should step back and look at what we know without elaborate analysis.

I have clients who mailed successfully for thirty years without using key codes (and no DM), but they made good decisions. They had good products at fair prices. The fact is even without a marketing database, we know most of the answers already. We can easily divide our customers into three groups: the good, the bad and the middle. We know we should mail the hot line (most recent buyers) and we're quite sure that we don't have to mail customers who haven't bought for more than a few years.

The only real question is which of those in the middle are good enough to mail. Now most of the techniques whether sophisticated or seat-of-the-pants work on this great middle. If you have 10,000 customers, your middle is probably only about 4,000. This means that it is far cheaper to mail them all than to spend tens of thousands on sophisticated models. This illustration has used recency (the R in RFM). If you want to get really sophisticated, you can look at the number of times a customer orders and how much they have spent (that is frequency and monetary or F and M). But no matter how many additional variables and sophisticated techniques you add, you are still basically working with the great middle. If you have 20 million customers, this sophistication will be worth much, otherwise, take your time exploring new (and expensive) analytical silver bullets.

Conclusion

There is a wide gulf between the techniques advocated by leading database marketing experts and the simple mailing methods of even the largest direct marketing practitioners. With cheap disk drives, it is possible to build a massive pile of data which may not dramatically improve your marketing success.

If you must build a marketing database, focus on key data - what you really need. You don't need millions of customers to do statistically valid modeling. However, if you want to build models based on what people have actually purchased, you will want to include transactions in your system (so you can select

based on part numbers). You must have more than a quick count system to find the people who have bought X widget. It might be better to have all of the information on some of your customers than some of the information on all of your customers!

Whatever statistics and reporting, you must still have the ability to spot anomalies. This may mean fancy query tools, but for most of us, it means programs which allow us to select customers and look at their related data. Summaries will always cover up the facts. Make sure there is a way to look under the covers. If you must build a marketing database, focus on the crucial questions. Make a list of things you want to answer . . . make it very specific.

Finally, it is very difficult to anticipate exactly what should be included. Understand that whatever you build will only be a first step. It must change and develop. Look for tools which allow you to interact with your data, build your own reports, etc. Walk before you run and allow the system to do simple things well for a while, then phase in a second set of requirements. Attempt to build maximum system flexibility.