Why Knowing About Our Customers Helps Our Business:  
A Practical Application of Business Intelligence Software  
John McIntyre, SAS Institute Inc.  
Marc Halsted, Ph.D., NationsBank

ABSTRACT

The proliferation of affordable and powerful hardware in modern enterprises, the increase in the number of knowledge workers capable of making valuable use of these tools, and the realization that critical examination of data about the enterprise can really have bottom line impact, have combined to drive modern businesses to examine their decision support infrastructure. Of these business drivers, the most compelling is the connection between business intelligence analysis, and measurable impacts on the activity of an enterprise. This paper will step through a case-study which examines how the use of business intelligence tools from SAS Institute helps identify new opportunities for the enterprise.

BUSINESS INTELLIGENCE, DATABASE MARKETING AND BANKING

Advances in technology move fastest when they show bottom line impact to organizations. Historically, some industries have been quicker to identify the cost effectiveness of emerging technology than others, reaping rich rewards for their shareholders. The insurance industry is an example of an industry which has made great use of computerization of routine tasks, like claim processing, to expand its reach and scope. The reduction of cost for claims processing, and the faster response to customers became the competitive edge for insurance companies. Many industries followed suit, seeking ways to automate the transactions of their work. Over the last 20 years, the primary focus of computing has been to deliver facilities to enhance the efficiency of transactional and operational systems, offering less labor intensive, more accurate, and thus less costly ways of doing business.

Increasingly, however, organizations are facing pressures to increase their competitiveness which cannot be met by further streamlining of operational processes. Where IT investments had yielded consistent increase in competitiveness, now, a search for new ways to serve the enterprise is underway. How to continue to be relevant, when the central area of contribution, savings from increased efficiency of operational systems, is no longer a sufficient answer?

The answer has come from the intersection of IT technology expertise and an increasingly customer centric philosophy in Marketing Departments. The evolving of a customer relationship based model for marketing, one which focuses on the profile of customer interactions with the enterprise, has created a demand for new ways to understand the needs of their customers. With that information often locked in legacy operational systems, the challenge is to organize systems to allow access to these data. Not for mere reporting, but for the sophisticated analysis required to gain new insights into customer behavior.

The economics of this are compelling. Enterprises are responding to studies which have shown that it costs 4 to 10 times as much to acquire a profitable customer, as it does to retain an existing one. This means enterprises are increasingly interested in retaining customers. To do this, they must know much more about what customers do, like, don’t like and what they look like.

In a recent two year long study, the banking industry learned some interesting facts about it’s customers. Some of these new facts were counterintuitive. For example, not only were banks losing money on more than half of their customers, but better educated customers were not the most profitable ones². Banks were further alerted to the fact that only about 10 to 20 percent of a bank’s customers accounted for nearly 50 percent of the costly transactions requiring teller assistance. Not surprisingly, this has prompted some banks to look at improving their “tele-banking” services, to encourage customers to do their transactions in the less costly “teller-less” environment.

The point of this section was to indicate that, business intelligence as a class of computing, and database marketing as a specific instance of it, is showing impact on the bottom lines of modern enterprises. Successfully implemented, it requires a joint effort between Marketing Analysts, capable of “mining” their corporate data warehouses, and IT professionals, capable of bringing the technology and data at their disposal to bear on this opportunity. The following case study will help illuminate some of the critical capabilities necessary to make this happen for a modern bank, and how SAS Software can be a key contributor.
DATABASE MARKETING AT NATIONS BANK: BACKGROUND/HISTORY

Approximately two years ago, NationsBank set up a Database Marketing group under the leadership of Timothy Wishon. This group managed the analytical aspects of direct mail campaigns, while also developing measures to do such things as gauge the value of each customer to the bank. By the summer of 1995, it was clear that the day-to-day tactical analytical work that went along with direct mail campaigns was distracting the group from its mandate to also carry out strategic analyses. Accordingly, last summer the Database Marketing group was split into two halves: a Direct Marketing group focusing exclusively on direct marketing campaigns, and an Analytical Applications group (headed by Timothy Wishon) that would conduct analytical work with more medium to long term, bank-wide relevance.

The new Analytical Applications group conducts customer segmentation, predictive modeling, and statistical analysis working within the Corporate Marketing Division of NationsBank. We perform some quick ad hoc analyses for other parts of the Marketing Division, but spend the majority of our time on a few large scale analytical projects.

One of the activities of our group has been the development and continual refinement of a dollar-based customer value measure. With this measure we are able to determine which customer households are most valuable to the bank and which are losing the bank money. By tracking this measure we can determine how a customer's value to us changes over time, as well as assess whether different bank campaigns are actually enhancing customer value.

We are also developing a customer segmentation scheme that will allow the bank to better service distinct populations within our large customer base. Obvious uses of the segmentation scheme include: 1) being able to better tailor messages and products to different groups of customers, 2) helping the bank to design and provide the optimal mix of access channels (e.g., tellers, ATMs, phone banking, PC banking) to different segments of customers, and 3) identifying portions of the customer base that are most at risk of attriting while gaining insights into what may be causing them to attrite.

A third activity involves developing models to predict price elasticity and market demand for a wide array of bank products. With these models we can predict what will happen to loan or deposit volumes if NationsBank or a competitor raises or lower rates a specific amount. These models will help NationsBank remain a market leader in an increasingly competitive financial services environment.

Our computing environment consists of mainframe computers (OS/MVS) running SAS 6.07 and fast Pentium PCs running SAS 6.10.

As a fairly new and still small group, we currently rely on outside vendors to perform some of our more time- and resource-intensive analyses and model development. Accordingly, we end up spending a good deal of our time pulling together disparate data sources from both inside and outside the bank, then cleaning, transforming, aggregating, and massaging the data into a form that can be used easily by our vendors. To date, it has been in the area of data manipulation that SAS has proved most beneficial to us.

CURRENT SOLUTIONS AND BENEFITS

From day one, the original Database Marketing group (and now the Analytical Applications group) has relied on SAS to manipulate and analyze financial data. As such, in almost all cases it is difficult to speak of past (non-SAS) solutions or non-solutions that SAS has enabled us to improve upon. There have been instances where a more sophisticated application or feature of SAS has made it possible for us to improve upon an earlier SAS solution, and these instances are noted below. Otherwise, what follows in this section is a description of the rather ordinary ways we are currently using SAS to get our work done.

The tasks to be accomplished in our environment, fall into four categories: the acquisition of and cleaning of data, statistical profiling within the customer population, applying scores and categorizations to customer records, and management reporting of results. At present, SAS is employed in the data management, categorization and management reporting tasks. Profiling is currently being done as an outsourced activity, though the use of SAS in this area is discussed later in this paper.

As much of our use of SAS is in the initial data manipulation stages of projects, it is difficult to come up with definitive business benefits that have resulted directly from our use of SAS. Our mandate is typically something like "build this segmentation scheme" or "conduct this statistical analysis." Our bosses do not really care what tools we use to get the job done. But as I hope will become clear in this section, on a large number of fronts it would be
difficult (if not impossible) for us to perform our assigned duties if we did not have a powerful tool like SAS at our disposal.

DATA MANAGEMENT

Working with multiple data sets

Perhaps the feature of SAS we most depend on is the ability to work with multiple data sets at the same time. It is quite common for us to take a variety of flat files, along with some of our often used SAS data sets and extract key variables across the various data sets to produce a new SAS data set or flat file to use for an actual analysis. Other packages (e.g., SPSS) don’t seem to offer this kind of I/O flexibility.

Given the way data exist at NationsBank (numerous sources and formats, residing on a variety of platforms and media), the ability to work easily with multiple data sets at the same time has been critical to the success of our analytical efforts. If we had to use other software or programming languages to accomplish the same results, a lot of our work would either not get done or else take months longer to complete (we have training and experience as researchers and statisticians not hard core non-SAS computer programmers).

FIRST.OBS and LAST.OBS processing

Frequently, the data sets we use must be transformed to a different level of analysis. For example, we may have a file of account-level information but need to roll up all the accounts of a single household into one record. With FIRST and LAST OBS processing such rolling up becomes almost effortless. As so much of our analytical work is done at the household-level, it would be a big hindrance to us to have to use some other programming approach to roll up all the account-level raw input data files we use.

Macros

In the days of the original Database Marketing group, a lot of the work involved performing similar analyses for different direct mail campaigns. We quickly came to appreciate the time saving flexibility of using SAS Macros to streamline and standardize the pre-mailing data preparation as well as the post-mailing response analysis.

In the new Analytical Applications group, although we no longer work in such a production-oriented environment, we still make good use of Macros when updating our monthly customer value calculation, and will inevitably continue to use Macros as we regularly re-score and update our new customer segmentation scheme.

RECORD CATEGORIZATION

Bucketing records

In our ongoing project to build market share/price elasticity models for different bank products, we regularly are confronted with tens of millions of records that needed to be aggregated into "buckets" on the basis of things like date of transaction, geographical location, and bank product. In this type of analysis the bucketed records become the basic unit of analysis, with variable values representing the averages or sum totals of all the records placed in each bucket.

In the early stages of the project only a couple hundred logical buckets existed and a previous programmer had produced the buckets by simply writing out a long DATA STEP, containing several hundred lines of intricate, nested IF statements. But even with extensive debugging, it was difficult to be 100% certain that his code was bucketing each record correctly.

As the project grew more involved, so did the number of logical buckets. Instead of a few hundred there were soon thousands of buckets that needed to be defined. Indeed, one of our most recent data sets contained nearly 50 million records that each needed to be placed into one of over 70,000 buckets.

Obviously with such a large number of buckets, it was no longer practical to even try to bucket the records using IF statements in a monstrously long DATA STEP. Given the size of the raw input file, it was also not practical to sort this enormous data file on all six or seven grouping variables and then use a first.obs/last.obs nested DO loop aggregation approach. Fortunately, after paging through SAS documentation, the Analytical Applications group devised a two-step solution centered around using PROC REPORT.

The enormous raw data file was run through PROC REPORT to produce a report containing 70,000+ rows, where each row corresponded to one of the desired buckets. Of course, the output from PROC REPORT could not be used directly as a final data file because redundant grouping variable values were not printed for every row of the report. However, by reading the PROC REPORT output into a subsequent
DATA STEP job as a flat file (after only moderate text editing to delete some header and page break lines), it was a simple matter to add the necessary grouping variable values to each row, and then PUT the modified rows out into a usable electronic flat file. This subsequent DATA STEP also allowed new variables to be created by manipulating the bucketed data variables PROC REPORT had produced in the earlier run. It also made it possible to precisely control the record layout of the resulting flat data file.

One of the advantages to this two-step approach is that there seems to be no logical limit to the number of buckets we can create. Furthermore, changes in the bucketing criteria can be easily made by altering just a few lines in the PROC REPORT job. In the old nested IF statement approach, a single change in the bucketing criteria would necessitate making numerous changes throughout the entire long block of code.

The two-step bucketing of records using PROC REPORT has made it possible to produce very fine-grained market share/price elasticity models containing many multi-level grouping variables. Without this solution it would have been necessary to build our models using far fewer variables, thus sacrificing a great deal of explanatory power. Even in the simplest (and not very useful) case, if we had used only time and location as predictive variables in our models, it would have taken nearly 1,000 lines of IF-THEN code to produce the necessary buckets. As the models have not yet been implemented, it is hard to put a precise estimate on the value they will have for NationsBank.

MANAGEMENT REPORTING

Graphs

One of the most important dictums in Statistical Analysis is "LOOK AT THE DATA!" Unfortunately, because we typically work with very large data sets in batch mode on the mainframe, it's difficult to get detailed, high resolution pictorial representations of the data. The character-based graphs we could easily derive tended to have such low resolution that they were more likely to blur our understanding of the data rather than clarify things. Most of the time we have had to content ourselves with studying non-pictorial printouts of frequencies, means, and other related descriptive statistics. Until recently, our only experience with graphs came from downloading SAS printout to our PCs and then using a spreadsheet package like Excel to produce detailed pictures of our data.

The benefits of using software like Excel or PowerPoint is that the resulting graphs are almost always presentation quality — even when the only person likely to see the graphs is the statistician or analyst. The big problem with producing graphs in this kind of environment is that a great deal of manipulation of the imported SAS printout file must usually take place before a decent graph of the data of interest can be produced. Furthermore, working with Excel in situations where a large number of very similar graphs need to be produced, each graph must either be created from scratch or else copied from an existing graph and then modified extensively. We often found ourselves thinking, "there has to be a better way."

Earlier this year, in conjunction with installing SAS 6.10 on our PCs, one of the members of our team took the time to learn how to use some of the functionality of SAS/GRAPH. We determined that when combined with a laser printer, SAS/GRAPH offered PC graphing capabilities that could help us do our jobs more efficiently. Our biggest roadblock to making regular use of SAS/GRAPH is that almost all of our data files reside on the mainframe. To use SAS/GRAPH on the PC we obviously have to transfer files between the two platforms.

For small data sets containing just a few variables we have found it often most practical to use a PUT statement to create a simple flat file, and then FTP the file down to a PC. For SAS data sets with many more variables, we use the PROC COPY utility (with XPORT engine) to create exportable SAS data sets. These XPORT data sets can then be FTPed directly down to a PC. Once on the PC and converted back to regular SAS data sets the files can be run through utilities like PROC GCHART, GPLOT or G3D.

Using the BY statement in a utility like PROC GPLOT, it is a trivial matter to produce a large number of graphs (differing only on something like date or geographical location code). A real plus with this approach is that the graphs can be piped directly to our laser printer, with no need to intervene (e.g., press RETURN) before each individual graph is printed.

This graphing functionality came in useful on a recent project where we needed to produce frequency distributions of regional mortgage interest rates for each month stretching back several years. With just a few lines of code we were able to create an entire series (100+) of crisp high resolution graphs in a matter of moments, rather than the days it would have taken if we had tried to do the same thing in a spreadsheet.
Being able to easily produce large numbers of high resolution graphs made it possible for us to detect subtle spikes and shifts in the distribution of mortgage interest rates from month to month; spikes that were not apparent when scanning through lists of frequency numbers. With this ability we were able to distinguish adjustable from fixed rated mortgages in unlabeled marketplace mortgage data. This knowledge has made it possible for us to produce separate market share-price elasticity models for the two types of mortgages, greatly enhancing the usefulness of the models to our mortgage division.

CONTINUING CHALLENGES & NEXT STEPS

As with all stories, this one is not finished. In the effort to improve our processes, we will be looking to SAS Institute to address a number of issues with their software, to help us more effectively do our job. Below, I have listed some areas where we will be looking for improvements or alternatives from SAS, as we move forward with our projects.

Sorting

In many of our projects before we can perform our analyses we first need to sort one or more data files on a variable of interest. This simple process however often turns into a nightmare as a simple PROC SORT job will bomb time and time again. Part of the problem is that PROC SORT jobs often default to a system sort routine, perhaps causing the job to not optimally use allocated SAS work space. Ultimately the problem seems to boil down to one of ignorance on the part of the users. It would be a great help if SAS could provide users with better documentation on setting up JCL to set aside the right kind of storage or work space so that sort jobs will not so frequently ABEND.

Alternatively, we will be looking into the use of PROC SQL. Because the ordering of data is so critical to our work, we will need a facility, whether it be PROC SORT, or something else. Use of PROC SQL should address two areas of concern, the multi-step process of sorting and merging, and the use of PROC REPORT to do our bucketing of records.

For sorting and merging, SQL should reduce the need to sort each set before merging, because the join activity in SQL does not require sequential matching of records. Where sorting the input data files ahead of time to do match-merging, the use of SQL will obviate that need.

In addition, we will be looking into the use of SQL to do our bucketing of records. The shift from using DATA STEP programming, to PROC REPORT was motivated by the need to reduce our maintenance of SAS code, as the number of "buckets" increased. By taking advantage of the summing available in PROC REPORT, we were able to avoid coding this ourselves, and only had to manage the post-processing of the text output file. PROC SQL has the automatic summing routines built in (i.e. ORDER BY and GROUP BY) which should duplicate the summing we were getting in PROC REPORT. In addition, we can include additional transformation logic in the SQL execution (should we need it) and have the result set saved as a SAS data table. This latter capability will allow us to skip over the post-processing of the text output file.

Clustering

Currently our biggest activity is devising a segmentation scheme to categorize each of our eight million plus customers. Although SAS has a few algorithms that can be used for disjoint clustering, we are relying on a non-SAS approach to form our segments.

We did test SAS clustering methods but found them lacking. Initial experimentation with non-hierarchical procedures like PROC FASTCLUS produced solutions where one of the clusters might contain 70% or 80% of all the observations — but clearly a cluster containing five or six million NationsBank customers is not going to help us tailor products across segments very well. Furthermore, large segments were produced even when we followed the suggestions in SAS documentation to: 1) standardize the data or use Principal Components as the raw inputs, 2) use a multi-pass approach to identify optimum starting centroid seeds, and 3) eliminate outliers from the centroid estimation process. We did find that running FASTCLUS through dozens of iterations (and lots of CPU) brought the size of the largest cluster down to a little under 50% of all observations, but this was still unacceptable for our business purposes.

There are other clustering algorithms out there that can perform disjoint clustering on large samples with many variables and not end up with one big elephant cluster and a bunch of flea clusters (some do not even require that the desired number of clusters be specified a priori). When is SAS going to introduce a more efficient and effective disjoint clustering technique appropriate for very large samples?
Graphing Analytical Results

With the roll out of a new segmentation scheme, many parts of the bank are going to want to see how particular segments compare on a variety of characteristics and financial indices. To be effective this information is going to have to be presented in simple, clear, easy-to-read graphs, charts and tables. Producing these quickly and easily in SAS is going to be a challenge, even with the power of SAS/GRAF. Clearly, the one thing that SAS could do to help the Analytical Applications group at NationsBank perform our jobs more effectively would be to provide users with an integrated presentation graphics capability.

One strategy we will be looking into will be to have a catalog of these graphs actually built on our mainframe as a batch process, and have the graphics catalog transported to our PC's. Once on the PC, the Graphics Editor facility in SAS should allow us to manually edit any of the charts which need minor adjustments. We will continue to be able to send these finished graphs to our locally attached laser printers for production style hard copies. Additionally, we will be looking into the SAS interface to PC e-mail packages, as a vehicle for distributing these graphs, without having to make multiple hard copies.

In addition, we will be investigating the use of the Graph Object in SAS/AF, available with SAS/AF. This capability will allow us to present graphs along a BY dimension, using a scrollable action. This should allow viewing many graphs in a single "frame", without having to produce many single instances.

Data Visualization

Aside from producing presentation graphics suitable for non-technical audiences, we also have a strong need for sophisticated data visualization and statistical graphics tools. While we have made limited use of such utilities as PROC GPLOT and PROC G3D, it would be infinitely more helpful to have an interactive data visualization tool that would allow us to do such things as rotate data in three-dimensional space, zoom into specific regions of that space, and delete specific data points with a click of the mouse.

We understand that SAS/INSIGHT and SAS/SPECTRAVIEW may be able to fill some of our data visualization needs.

Conclusion

NationsBank has recognized for some time that the effective use of customer data will be a key to our competitive advantage in today's marketplace. Knowing about our customers is no longer an option. The more we can understand about who our customers are, the better we can profile this information in ways which can inform our product design and service delivery.

To accomplish this, we will need tools which can handle the data management tasks involved with gathering information from disparate sources. We will need state-of-the-art analytical capabilities to allow us to maximize our investigations. We will need efficient routines for re-scoring and categorizing our customer populations based on our analyses, and we will continue to need easy to use, effective tools for management reporting.

SAS Software has been an invaluable resource in this endeavor, and we expect to continue to see enhancements and improvements to the tools, to further assist us in moving our enterprise forward.


SAS, SAS/GRAPH, SAS/AF, SAS/SPECTRAVIEW and SAS/INSIGHT are registered trademarks of SAS Institute Inc., in the U.S. and other countries.

For Further Information

John McIntyre is a Program Manager for Business Solutions at SAS Institute, Inc. He can be reached at:

SAS Institute Inc.
SAS Campus Drive
Cary, North Carolina 27613
e-mail sasmi@unx.sas.com

Marc Halsted, PhD., is Vice President, Database Marketing at NationsBank. He can be reached at:

NationsBank
100 North Tryon Street
NC1-007-17-03
Charlotte, North Carolina 28225