Teaching the Use of Database Research: A Customer Relationship Management/Direct Marketing Case Example

Joseph L. Orsini, College of Business Administration
California State University, Sacramento
Sacramento, CA 95819-6088; [tel] 916-278-6992

ABSTRACT

Continuous growth in computer storage capacity has led to the development of methods which utilize these large databases in marketing efforts. The purpose of this paper is to provide a case illustrating this process for use in marketing and marketing research classes. It integrates data mining and business analysis to provide a realistic example of how database research may be used in marketing.

BACKGROUND

The last two decades have seen an explosion in the capacity of business organizations to retrieve and store operational data. When combined with customer data, these massive databases constitute a valuable source of information for marketing purposes (Hughes 1996). Businesses are increasingly using this information in their marketing efforts, particularly in promotion (Benson, et. al, 2000).

In the late 1990's, editions of marketing research and business research textbooks have typically begun to contain discussions of databases and Data Mining (DM). For example, Churchill (2001) vs. Churchill (1996) and Cooper and Schindler (1998) vs. Cooper and Schindler (1995); both have DM discussions in their new editions, but not in the previous ones. However, an application of the integration of marketing research and business analysis has been lacking. The purpose of this paper is to provide an example of that process.

RELEVANT CONCEPTS

Customer Relationship Management

Customer Relationship Management (CRM) is "a process that manages the interactions between a company and its customers " (Schulz 2000; Berson, et. al, 2000). It typically uses Data Mining methods to find "patterns that are good predictors of purchasing behaviors", then marketers "feed the results into campaign management software that ... manages the campaign directed at the defined market segments" (Berson, et. al, 2000, p. 10). A typical example is finding which prospects in a database are most likely to respond to a particular offer, then conveying that offer to them (ibid, p. 11). The case illustrated in this paper is an example of an application of this type.

Direct Marketing

Overlapping the CRM concept is that of Direct Marketing (DMk), a concept which has been in existence long enough to have its own professional organization: the Direct Marketing Association. This organization has defined DMk as:

> "Direct Marketing is an interactive system of marketing which uses one or more advertising media to effect a measurable response and/or transaction at any location." (McDonald 1998)

A somewhat narrower concept than CRM, it is focused more specifically on selling rather than addressing the overall relationships between customer and seller. However, the CRM tasks of determining which customers are most profitable, and which promotions are most effective (Berson, et. al, 2000) fits equally well in DMk.

Databases

Hair, et al. (2000) note that the second half of the 20th Century has seen two shifts in the "fundamental character of data analysis" (p. 671). The first was the mid-1970s, when mainframe and personal computers allowed multivariable methods to gain widespread acceptance and use.

The second shift occurred in the 1990s, with two related developments: the emergence of large-scale databases ("data warehouses"), consisting of hundreds of millions of lines of data; and the development of new methods to analyze this "information avalanche". Compared to survey-derived data, the preparation of data warehouses

for analysis is no simple task, and may be the most time consuming part of any DM analysis (Groth 1998). For classroom pedagogical purposes we may use "cleaned" sets of data; i.e., data that is ready for analysis.

Data Mining

With a new and evolving discipline, definitions are not firmly established, and Data Mining is very much in accordance with this model. For this discussion, DM will be defined as: "the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules" (Berry and Linoff 2000, p. 12). The following is a brief listing of analytical tools that may be applied after a suitable database has been prepared (Orsini 2000).

Standard Statistics: a sample of a database obtained and put into a spreadsheet; all the "traditional" statistical techniques are available for use (Hair, et al., 1998).

Queries: investigative questions relating to specific items in the database; may be in the form of Structured Query Language (SQL) for use in On Line Analytical Processing (OLAP) methods (Berry and Linoff 2000).

Data Visualization: displaying data for visual review have been found to be increasingly useful, particularly for assisting managerial decisions (Hair, et al., 1998). Visualization methods range from simple graphs to quite complex multidimensional structures, which are combinations of art and mathematics.

Data Mining Tools: Data Mining Tools include the following: Market Basket Analysis, Memory Based Reasoning, Genetic Algorithms, Cluster Detection, Link Analysis, Decision Trees, and Neural Networks (Berry and Linoff 1997).

This case discussed in this paper uses the standard statistics aspect of DM. It is applied to a sample of transaction and demographic information originally contained in a very large database.

THE CASE EXERCISE

Background

The case used in the class exercise, and its accompanying data, is taken from Lilien and Rangaswamy (1998). It concerns the Bookbinders Book Club (BBBC), a catalog retailer of books that has made a trial mailing of a brochure about one of their books. The brochure is sent to a sample of 20,000 of its file of 500,000 customers; see Appendix. Product, promotion and overhead costs of the pilot mailing are included in the case background. This information allows two things:

- The development of a predictive model regarding book purchase, based on 1 demographic and 9 purchase history variables
- Computation of the profitability for different promotional approaches (model-based targeting vs. mass mailing, in this case).

The data included in the original case consists of these 11 variables, but for 3900 observations (development and test datasets), a considerably larger dataset than the more typical survey data used in research classes.

There are a number of variations possible in utilizing this case, two of which are described below. The fundamental questions are:

- Based on the trial sample, will it be profitable to do a mass mailing of 50,000 of these brochures?
- How much more profitable is it to mail the special brochure based on customers' scores on a model, rather than doing the mass mailing?

Case Computations

The following steps assume that students are able to take the data and perform the analyses. This assumption may be relaxed for research classes that do not include hands-on data analysis, or marketing classes that do not engage in any analysis; see below. This process is:

- Create a predictive model using the development dataset
- Compute the "Purchase Probability Index" (PPI) - scores of the dependent variable (purchased or not purchased) - using the test dataset
- Categorize the test dataset by PPI into a limited number of categories

- For each PPI category, compute a table of the % of customers in the data, and the % of customers who purchased
- Compute the "Breakeven %" of purchases for a mass mailing
- Compute the profit in each category for the two approaches to promotion (mass mailing and the model)

Two approaches were used in this example: a "best customer" model (RFM) widely used in industry, and a Binary Logistic Regression model (BLR). Results of the RFM model are shown in the case, the results of the BLR model may be obtained from the author.

For classes that do not include statistical analysis, the first four steps may be provided by the instructor and included in the case information; an example is included in the case in Appendix I. It is useful to at least demonstrate this process of steps in class, as it beneficial to show students that the computations are not beyond their potential capabilities.

The analysis of the case in the Appendix finds that any RFM score category with less then 6.37% of the persons in that category purchasing the book would result in a loss, rather than a profit, in mailing brochures to that category. That is, the cost of brochures would be greater than the profits derived from the book sold.

The 7 categories were then combined into two groups: those buying less than the breakeven %, and those buying more. Profitability is computed for two alternatives: the mass mailing, and the RFM model. While the mass mailing approach does yield a profit (about $0.27 per person mailed to, in this case), the generic "good customer" RFM

model indicates a 6% profit improvement over mass mailing, while the more specific BLR model produces a 95% profit improvement.

Benefits to Students

There are several benefits of the exercise to student learning. Most obvious is seeing a clear connection between marketing research and the impact on organization profits. In the author's experience, most textbook research examples use survey research, which lacks a direct connection to profits, or only through a complex set of connections. In the BBBC case, reducing promotion costs through improvement of predicting likely buyers has a direct and immediate impact on profits, typical of Direct Marketing applications. The profit computation should also be illustrated in the classroom, as discussed below.

From the methodology perspective, not only are bivariate and multivariate methods able to be demonstrated, but the practical value of predictive models, even when they are not strongly predictive, is also demonstrated. The Coefficient of Determination of the RFM model, for example, is a mere 6%, while the BLR model is only about 23%.

The use of spreadsheets for business analysis is also demonstrated, as these computations are easily accomplished in this way, but laborious when done for each category by hand calculator. Further, it is the author's experience that many marketing students, both undergraduate and graduate, do not clearly understand how to conduct even this simple business analysis, and that without several "helpful hints", many students could not correctly calculate the profits or the breakeven sales level for this exercise.

Extension to Additional Research Methods

Several other aspects of marketing coursework may be demonstrated in a classroom context through the use of a database research case. Most of the examples below apply to research classes.

Information Value: As indicated above, the models developed were understandably not strongly predictive, given the nature of the independent variables included. Would improving the predictive ability of the model be worth the cost; e.g. "How much would we be willing to pay to add credit card information to the database?" Further, supposing there was the ability to perfectly predict who would buy the book? The case provides the potential for a clear application of the value of perfect information.

Sensitivity Analysis: Varying several factors involved in the determination of profits not only helps the student to better understand the concepts, but also develop an appreciation for how quality differences may be transformed into "bottom line" profits. For example, what is the tradeoff between brochure sales effectiveness and the cost of brochure production?

Data Transformation: Unlike surveys, which typically have precoded data and are designed for future statistical analysis, transaction-derived data may need work prior to being able to be analyzed; see the Groth (1998) quote above. Alphabetic data (e.g. "m" and "f" for gender); transaction dates instead of elapsed time; and dollar signs in spending variables, are examples of data needing worked on prior to any analysis.

Sampling Methods: Random sampling tends to be a difficult concept for many to grasp. This case may be used to illustrate various sampling concepts. For example, simple, systematic and stratified sampling may be discussed in context of the question "How would you select the sample of the 50,000 who would receive the mass mailing of the brochure?"

Experimental Designs: Design of Experiments (DOE) is another research topic that experience has shown is difficult for many students to grasp, particularly the leap from the typical textbook discussion to an application. For example, the question may be asked "How would we test the effectiveness of two different brochures; how would we see if gender makes a difference in the response to those two brochures?"

DISCUSSION

As Orsini (2001) points out, teaching basic DM concepts in a marketing research classroom context is increasingly necessary. The purpose of this paper is to illustrate an example of the utilization of database information in marketing efforts. The integration data mining and business analysis in a case illustration of how database research works in a "real world" context has been found to provide a valuable learning tool for marketing students, both undergraduate and graduate.

The downside of this specific example is that it was "borrowed" from a textbook the author used in a graduate marketing class; clearly, permission to use this specific data would have to be obtained. The development of other similar and more widely accessible cases would be a very useful pedagogical contribution.

Appendix
The Bookbinders Book Club Case*

The Bookbinders Book Club [BBBC] was established in 1986 for the purpose of selling specialty books through direct marketing; they do not publish their own books. In anticipation of using database marketing, BBBC has established a database of 500,000 readers, and sends out informational mailings once a month. The data in its database consists of a demographic variable and transaction information, including the following for each customer: Gender, Recency [elapsed months since last purchase], Frequency [total number of books purchased], Money [total $ amount purchased], and several summaries of transaction-related variables, e.g. the number of art books the customer has purchased.

In a recent trial mailing, the company randomly selected 20,000 customers from their database and included, with their regular mailing, a brochure for the book *The Art History of Florence*. This resulted in a purchase response of 1806 orders for the book; the database for the 20,000 now contains a variable "Choice of *AHF*", coded 0 [did not purchase] and 1 [did purchase]. BBBC is now considering doing a similar mailing to another 50,000 of their customers. The marginal cost of the *AHF* book brochure mailing is $0.65/addressee [including brochure production, stuffing and postage]. Each book sold costs BBBC $15.00 to purchase and mail, with a selling price of $31.95; the company allocates organization overhead to each book sold at 45% of book cost.

A consultant for the BBBC has computed RFM scores for each of the 20,000 customers. The RFM score category (segment) of the RFM, the percent of customers in that RFM score category, and the percent of the customers who purchase in that category (Purchase Probability) is:

RFM Sc.	40	50	60	70	80	90	100
% Cust.	3.4	17.3	29.6	28.4	16.2	3.9	1.1
% PProb.	11.6	8.8	8.5	11.7	6.7	4.5	0

The student assignment is to compute the following:

- Profitability of a mass mailing
- Breakeven purchase % of mass mailing
- Profitability of the RFM model

*Based on Lilien and Rangaswamy 1998

REFERENCES

Berry, Michael J. A., and Gordon Linoff (1997), *Data Mining Techniques*, New York:

John Wiley & Sons, Inc.

Berry, Michael J. A., and Gordon S. Linoff (2000), *Mastering Data Mining*, New York: John Wiley & Sons, Inc.

Berson, Alex, Stephen Smith and Kurt Thearling (2000), *Building Data Mining Applications for CRM*, New York: McGraw-Hill.

Churchill, Gilbert A. Jr. (2001), *Basic Marketing Research*, Fort Worth: The Dryden Press.

Churchill, Gilbert A. Jr. (1996), *Basic Marketing Research*, Fort Worth: The Dryden Press.

Cooper, Donald R. and Pamela S. Schindler (1998), *Business Research Methods*, New York: Irwin/McGraw-Hill.

Cooper, Donald R. and Pamela S. Schindler (1995), *Business Research Methods*, New York: Irwin/McGraw-Hill.

Groth, Robert (1998), *Data Mining*, Upper Saddle River, N.J.: Prentice-Hall.

Hair, Joseph F. Jr., Ralph E. Anderson, Ronald L. Tatham, and William C. Black (1998), *Multivariate Data Analysis*, Upper Saddle River, N.J.: Prentice-Hall.

Hair, Joseph F. Jr., Robert P. Bush and David J. Ortinau (2000), *Marketing Research*, Boston: Irwin McGraw-Hill.

Hughes, Arthur M. (1996), *The Complete Database Marketer*, New York: McGraw-Hill.

Lilien, Gary L. and Arvind Rangaswamy (1998), *Tutorial To Accompany Marketing Engineering*, Reading, MA, Addison-Wesley.

Mcdonald, William J. (1998), *Direct Marketing: An Integrated Approach*, Boston: Irwin McGraw-Hill.

Orsini, Joseph L. (2001), "Teaching Data Mining and Other Database Research Methods in Marketing Research Classes: An Interim Assessment", *Proceedings*, Marketing Educators' Association, 121-125.

Schulz, Don E. (2000), "Learn to differentiate CRM's two faces", Marketing News, November 20, 11.