Haas School of Business
UC Berkeley

The BookBinders Book Club: Basic Customer Analysis

DATABASE MARKETING

Direct Marketing (DM) is among the fastest growing channels of marketing. Reports estimate that more than half of the adult population, or 52.5%, ordered merchandise by phone or mail in 1994[1]. Sales volume estimates for 1991/92 range from $200 billion in mail order sales to $275 billion in telephone sales.[2] Advertising expenditure on direct mail alone have risen by nearly 150% over a decade from $11.79 billion in 1983 (15.6% of the total advertising cost) to $27.43 billion in 1993 (19.9% of the total).[3] Expectations are that sales through DM will grow at a rate of 10% a year, versus only 6% for all other types of retail sales.[4]

Within DM, Database Marketing (DBM) is assuming an increasing role, because it allows the firm to reach its target markets more effectively, to personalize and customize marketing communications at the segment or even the individual level, to build/maintain a relationship with each customer, and to easily test and measure the results of the marketing efforts.[5] In DBM, one uses the power of data and information technology in order to target audiences for promotions of goods and services from among a list of potential customers in a database, often the company's "housefile" (or houselist), to cut down expenses and increase profits. This trend was made available by recent advances of computers and information technology which allows even the smaller marketing outfits to gather, compile and process a wealth of information about their customers, from which they can learn about their customers' preferences and wants and fulfill their individual needs. The results are more focused promotions, lower costs and higher

[1] DMA's "News & Information", Spring 1995.
[2] DMA's "Grassroots Advocacy Guide", Feb. 1995
[3] DMA's "News & Information", Spring 1994.
[4] Kolter, Philip (1991), Marketing Management, 7th edition, Prentice Hall: Englewood Cliffs, New Jersey, p. 622.
[5] Kotler, Philip (1991), Marketing Management, 7th Edition, Prentice Hall: Englewood Cliffs, New Jersey, pp. 623-626.

This case was originally prepared by Nissan Levin and Jacob Zahavi, Faculty of Management, Tel Aviv University and subsequently modified by Professor Charlotte Mason and Melissa Martin of the University of North Carolina. The case was further modified by Florian Zettelmeyer for use in his course and for use with Stata instead of SPSS. This case was developed to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. Names and data may have been disguised to assure confidentiality.

responses, which translate, in the bottom line, into more effective use of the promotional budgets and higher profits.
DBM encompasses two main aspects:

- Database management
- Marketing management

The database management aspect is concerned with building and processing of a marketing database to include all the relevant information about the firm's customers, from which one can learn about the customer's purchasing habits and patterns and then offer appropriate products and services. This information may include purchasing history, demographic characteristics, lifestyle attributes, financial and credit data, psychographic data, survey data, and other related information. Often, the data comes from several sources and in different formats, and the role of database management is to put the data together in a coherent and unified format, to facilitate storing, manipulating, processing and retrieving the data. Several Database Management System (DBMS) packages are currently available in the market for database management, almost all of them have a PC version, to name a few: ACCESS, PARADOX, ORACLE, DB2, FOXPRO, MAGIC, and others. Most DBMS use relational database architecture to represent and link data elements in the database. In a relational architecture, raw data is stored in Tables, with each table usually corresponding to specific type of information (e.g., customers' name and address table, purchase history table, etc.), which are linked together by means of one or more KEY fields (e.g., the customer's Account Number).

Marketing management is concerned with analyzing the data in the database and interpreting the results for decision-making. Often, the marketing analysis phase consists of two major steps:

- Response modeling to assess customers' reactions to product/services offered to them
- "Scoring" and assigning a likelihood of purchase for each customer in the database

These scores then drive the decision making process, which may include:

- Targeting audiences for promotion
- Prediction
- Offer and price optimization
- Order quantity
- Optimal number of contacts/mailings/emails and others

The analysis stage generally uses a statistical package, such as STATA, SAS, SPSS, GAUSS, and others.

THE BOOK INDUSTRY

About 50,000 new titles, including new editions, are published in the US each year, giving rise to a $20 billion industry (in 1994). This industry is segmented as follows (in terms of percentage of sales):[6]

[6] 1994 US Industrial Outlook, Department of Commerce, Washington, DC.

27% textbooks
21% trade books sold in bookstores
21% technical, scientific and professional books
10% book clubs and other mail-order books
8% mass-market paperbound books
13% all other books

Book retailing in the 1970's was characterized by the growth of US chain bookstore operations in concert with the development of shopping malls. Traffic in bookstores in the 1980's was enhanced by the spread of discounting. In the 1990's, the superstore concept of book retailing was responsible for the double-digit growth of the book industry. Generally situated near large shopping centers, superstores maintain large inventories of anywhere from 30,000 to 80,000 titles, and employ well-informed sales personnel. Superstores are putting intense competitive pressure on book clubs and mail-order firms as well as other retail outlets. In response to these pressures, book clubs started looking at alternative business models that are more responsive to their customers' preferences.

Historically, book clubs offered their readers continuity and negative option programs that were based on an extended contractual relationship between the club and its clients. Under a continuity program, a reader signs up for an offer of several books for a few dollars each (plus shipping and handling), and an agreement to receive a shipment of one or two books each month thereafter. This is a "low maintenance" arrangement since a single contract guarantees a sequence of sales. It is most common for children's books, where parents are willing to delegate the rights to make a selection to the book club, and in fact much of the club's prestige depends on the quality of its selection. In a "negative option" program, readers get to choose which and how many additional books they would receive, but the default option is that the club's selection will be delivered to them each month. The club informs them of the monthly selection and they are specifically required to mark "no" by a deadline date on their order form if they do not want to receive it. Negative option programs sometimes result in customer dissatisfaction and always give rise to significant mailing and processing costs.

In an attempt to reverse these trends and combat the success of superstores, some firms are beginning to offer books on a positive option basis, but only to selected segments of their customer lists that are deemed receptive to specific offers. Thus, book clubs are beginning to use database marketing techniques to work smarter rather than expand the coverage of their mailings. They target individual customers based on data in their databases to select only customers who are likely to be interested in their offers, and they differentiate their offers across their customer population.

One example of this database marketing trend is Doubleday Book & Music Clubs Inc. that recently started using the results of predictive modeling to target its mailings. The company uses their database to identify good customers early in their membership while cutting costs attributed to poor members. According to Doubleday president, Marcus Willhelm, "the database is the key to what we are doing ... We have to understand what our customers want and be more flexible. I doubt book clubs can survive if they offer the same 16 cycles, the same fulfillment, to everybody."[7]

Doubleday's predictive modeling looked at more than 80 variables during tests, including geography and what types of books customers purchased. Three to five variables were

[7] DM News, May 23, 1994. "Doubleday Rollout Eyes Top Members", Beth Negra P. 2

eventually chosen as the model's basis. "A whole battery of tests were run," Willhelm said. "The whole idea is to target subsets in the membership files" of about three million names. "If a customer only buys two or three books a year, do they need 16 catalogs? Maybe they need to be sent only eight. We look at profitability, not sales. " With the use of database marketing techniques, Doubleday is planning to switch to positive option plans that perform well in its market tests.

THE "BOOKBINDERS BOOK CLUB" - BBB

The BBB Club was established at the end of 1986, for the purpose of selling specialty books through direct marketing involving a variety of channels, including media advertising (TV, magazines, newspapers) and mailing. BBB is strictly a distributor and does not publish any of the books it sells. In anticipation of using database marketing, BBB made a strategic decision right from the start to build and maintain a detailed database about its club members containing all the relevant information about their customers. Readers fill out an insert that is returned to BBB which then enters the customer into the database. By now the company has built a database of 500,000 readers through advertising in specialty magazines.

Initially, and as long as the customers' database was still relatively small, BBB contacted all its club members with each appeal to purchase books. Every other month, the company sent out solo mailings for its latest offering (that is, each offer is for one book). BBB's sales have grown steadily, but profits began falling when the database got larger and when the company diversified its book selection and increased the number of offers sent to customers. The falling profits have led BBB to switch to database marketing in order improve its mailing yields and stay profitable. BBB's management has decided to base the database marketing operation on the following principles:

- New members would be acquired by advertising in specialty magazines, newspapers and TV.
- Existing club members would be contacted by direct mail and telemarketing.
- The mailing / telemarketing list would be chosen using database marketing technology.
- Customer response, whether purchase or no purchase, will be recorded and maintained in the database, to be used in future targeting of audiences for promotion (in fact, the management considers this database as the main asset of the company).
- Every new book would be offered to the club members first - prior to advertising it in the media.
- The price in media advertising would always be higher than the one in the direct mailing/telemarketing offer.
- Live market tests, involving a random sample of customers from the database, would be conducted for new book editions in order to analyze customers' response and calibrate a response model for the current book offering. The response model's results will then be used to "score" customers in the database and select customers for the mailing campaign.

The idea is to run simultaneous targeted campaigns where each target audience will receive appropriate solo mailings.

Parallel to selling books, BBB management has also been taking advantage of its database to offer its members non-book products. Based on the success that the company had in the past with the non-book operation, it plans to continue this operation, even expand it, in the future.

THE DECISION PROCESS

Two core decisions are usually supported by DBM: targeting and prediction. Targeting is concerned with selecting the audience that is most receptive to the current offering. Prediction is concerned with forecasting the number of orders generated. The decision process in DBM usually consists of several steps:

- Testing
- Response modeling
- Scoring
- Decision making

The process starts with a TEST mailing, where a sample of customers are selected randomly from the database and mailed the new offer.

The responses to the offer are then analyzed along with customers' past purchase history, demographics and other relevant data to determine how the response varies as a function of the customers' attributes and history. Purchase history includes variables such as: how recently they have purchased ("Recency"), how often they have purchased ("Frequency"), and how much money they have spent on buying the company's products in the past ("Monetary") (the so-called RFM variables). Other pieces of purchase history are the number of books bought by various categories, and whether or not a customer has bought specific related products in the past. The resulting model is referred to as a Response model.

The response model calibrated on the basis of the TEST results is next used to assign a "score" for each customer in the balance of the list (i.e., for customers who were not part of the TEST mailing, which constitute the majority of the list), reflecting the customers' "likelihood" of purchasing the current product offering.

Finally, customers are selected for the promotion based on their expected likelihood of purchase, the revenue generated by a sale, and the cost of mailing the offer. The final mailing is referred to as the ROLL mailing.

YOUR TASK

To evaluate the suitability and performance of the various response models currently used in the DBM industry, BBB plans to use the results of its latest offer to do some 'what-if' analysis. This book, "The Art History of Florence", was offered to a sample of 50,000 customers, chosen randomly from the list, and generated an impressive response of 4522 orders (a response rate of 9.044%).

Your first step is to 'get to know' the data by conducting some statistical analysis using BBB's customer database. Summary information about the BookBinders Book Club's customers' purchasing history and demographics is in the Python dataset called *BBB.csv*.
Exhibit 1 contains a listing of the variable names and descriptions of the data types that you should use to answer the following questions:

1. What percent of BookBinders customers are female?

2. Which three states account for the largest percentage of BookBinders's customers?

3. What is the average Total $ spent, the average Total # of book purchases, and the average number of months since last purchase?

4. Calculate the correlation between customers' total spending on books and their total spending on non-book products. Is the correlation statistically significant?

5. Use a regression as another way to determine whether customers' total spending on books and their total spending on non-book products is correlated. Is the coefficient significant?

6. Which book categories have sold the most books? Which have sold the least?

7. Create a bar chart showing the average total spending on books for males and females.

8. Is the average total spending on books statistically different for males and females? Answer this question using a t-test.

9. Answer question 8 using a dummy variable regression. Hint: First transform the gender variable into a 0/1 dummy variable: `generate female=(gender=="F")`

10. For both males and females, find the total number and also the percent who bought "The Art History of Florence."

11. Are women more likely to buy the "The Art History of Florence" than men? Use a statistical test to answer the question.

12. For both males and females, determine the total number of purchases and the average number of purchases by males vs. females.

13. Determine the minimum, the maximum, and the average number of months between customers' first purchase and their most recent purchase.

14. What percent of repeat customers (those with two or more total purchases) bought "The Art History of Florence?"

Exhibit 1

The BookBinders Book Club
Python Dataset

Summary information about the BookBinders Book Club's customers' purchasing history and demographics is in the Python dataset called *bbb.csv.* The dataset is available on Haas Jupyter Hub.

Below is a listing of the variable names and descriptions of the data types:

Contents of bbb.csv – contains records for 50,000 customers			
Variable name	Type	Size	Description
acctnum	Numeric	5	Customer account number
gender	String	1	Customer gender – M=male, F=female
state	String	2	State where customer lives (2-character abbreviation)
zip	String	5	ZIP code (5-digit)
zip3	String	3	First 3 digits of ZIP code
first	Numeric	3	Number of months since first purchase
last	Numeric	3	Number of months since most recent purchase
book_	Numeric	8	Total dollars spent on books
nonbook_	Numeric	8	Total dollars spent on non-book products
total_	Numeric	8	Total dollars spent
purch	Numeric	5	Total number of books purchased
child	Numeric	5	Total number of children's books purchased
youth	Numeric	5	Total number of youth books purchased
cook	Numeric	5	Total number of cook books purchased
do_it	Numeric	5	Total number of do-it-yourself books purchased
refernce	Numeric	5	Total number of reference books purchased
art	Numeric	5	Total number of art books purchased
geog	Numeric	5	Total number of geography books purchased
buyer	Numeric	1	Did the customer buy "The Art History of Florence?" (1=yes, 0=no)