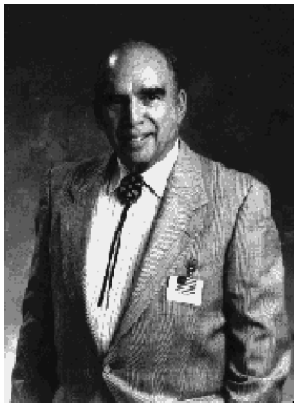
Ed Burnett

The ABC's of Modeling for "Business-to-Business"

During the course of a career that spans over 40 years in the direct mail marketing business, Ed Burnett has helped mailers select over 10 billion names for direct mail and telemarketing campaigns. He is President of Ed Burnett Consultants, Inc., one of the five major direct marketing firms that makes up THE DATABASE AMERICA COMPANIES. His firm specializes in direct mail consultation, list compilation and list brokerage. He is widely recognized as the pioneer of many of the list marketing concepts and techniques utilized today throughout the industry. These concepts and techniques are freely shared in the numerous articles he contributes to all of the major trade publications, and are gathered together in the book he recently authored for Prentice Hall, *The Complete Direct Mail List Handbook: Everything You Need to Know About Lists and How to Use Them for Greater Profit*. He is also the author of *Database Marketing -- The New Profit Frontier*.



Ed Burnett

ABSTRACT

Modeling of a Business-to-Business Customer file proceeds on two levels. This Ed Burnett article covers the steps necessary for the production of a penetration analysis. Wondering what kind of differences it pay to analyze? Here is your answer, given in numerous examples with detailed explanations. Below, you will read how penetration is figured and analyzed in various SIC circumstances, and how to make use of levels of data. And review what advantages can acrue to a large Business-to-Business model.

Introduction

Modeling of a Business-to-Business Customer file proceeds on two levels. This article covers the steps necessary for the production of a penetration analysis. Here are the answers to your questions about penetration, given in numerous examples with detailed explanations. Below, you will read how penetration is figured and analyzed in various SIC circumstances, and how to make use of levels of data.

Modeling of a Business-to-Business Customer file proceeds on two levels - the first is required for all businesses - and thus serves as a platform for the second. While any business list can profit from the complete two-pronged approach, it is clear that some mailers will obtain enough intelligence about their customers from step one, that they can safely stop at that point. They will have, in effect, such a wealth of data concerning "where they have been" and "where they next should go" that immediately moving to a "scoring" model may well be an uneconomic use of prospecting dollars.

This treatise covers Step I in detail, and then outlines what now can be done and why, if Step II is embraced.

Step I calls for the production of a penetration analysis (a share of market).

But before even one penetration is run, certain basic decisions need to be made:

A. The penetration is based on establishments, not executives or contact personnel. This means the customer file must be merge-purged of all duplicates. (A file with multiple executives at some or all establishments will skew or sully the validity of the penetrations.)

B. The level of analytic detail required must be determined. Each major variable needs its own penetration analysis (otherwise the file is handled as a single set of data, and any and all differences will be smoothed out and "homogenized" in a single set of reports.)

What kind of differences might it well pay to analyze?

1. Many companies incorporate in their customer file, several kinds of business. (A classic example is a firm with two basic lines - a Senior Product line selling for thousands of dollars, and a Jr. Product selling for under $100. The markets, obviously will be quite different for each product line.)

2. There may be a need to distinguish between inquiries and buyers, or between previous buyers and recent buyers.

3. Some businesses incur conditional sales - in which some cancel, while others carry on to term. The operative has a "need to know" how these differ.

4. The size of each of the differing cells of the customer file needs to be taken into account, for a cell with only a few thousand establishments may not provide the same wealth of detail as a cell consisting of many thousands of establishments. (For comparison, however, each individual file or cell is run exactly as every other cell or file.)

Because so little has been done to date to model Business-to-Business customer files and because the terms utilized tend to "sound" alike, it seems best to provide definitions of the major terms which need to be employed.

These major terms are:

A. Distribution
B. Business Universe
C. Penetration
D. Prospect Universes (at various levels of Penetration)

A. Distribution is another word for "counts" - counts primarily of demographics (of the customer's file or files):

By Classification (2-4-8-digit SIC) plus Major SIC Clusters

By Employee size (from 1 to 4 employees, to over 1000 employees.)

By Product Group

By Geographical Area

B. Business Universe - There are in the United States almost 10,000,000 unduplicated non-household establishments (by address) as opposed to some 100,000,000 individual households by address.)

The Database America all Business file includes 9,410,000 businesses, institutions and offices of professionals coded by 2-4 and 8-digit SIC and Number of Employees. It includes every listing of every establishment in America with a business telephone, and in addition to encompassing every line in every column of almost 5000 classified (yellow pages) telephone directories, includes over 100 additional business sources, for size data, for such fields as Banking, Savings & Loans, Lodgings, Governments, Hospitals and Rest Homes and every Educational Institution in America. This is one of just two all-business files available which breaks down SIC data by 8-digits. (The U.S. Dept. of Commerce provides the basic two and four digits - but that system leaves over half of American business operations uncoded in a "catch-all" titled "Not Elsewhere Classified.")

C. Penetration - This involves matching distribution counts by establishments for the customer file - by SIC - and by employee strength - against the Business Universe to provide percentages (by each SIC and each employee strength.) This match against the Business Universe provides - in effect - a "share of market" and since over 8,000 8-digit SIC's are available and 9 employee size groupings, data in this analysis is thus available for up to 72,000 individual "nodes."

This detailed set of percentages (penetrations) is then reported:

1. By 13 major business groups of two-digit SIC.
2. By individual 2 digit SIC.
3. By 4-digit SIC - within each 2-digit SIC.
4. By 8-digit SIC - within each 4-digit SIC.
5. By 9 employee strength ranges.

It should be noted that penetration, overall, is provided for each individual SIC classification and each is then provided in two distinctly different sequences:

1. In straight Ascending Zip Code Sequence (from 01 to 99).
2. In descending array by penetration (the range here is almost always astronomical from a high of 25, 50, 100% down to a low of 1/20 or 1/10 of 1%.)

D.Prospect Universes

Since penetration percentages are provided for each 8-digit SIC (which then embraces every 4-digit SIC and every 2-digit SIC) the data is easily at hand to accumulate all prospects (on the Business Universe which match the Customer File) at selected penetration percentages.

Typically for a file which may concentrate into a small number of 4 digit SIC's, but show some sales (ergo some penetration) in hundreds of 4-digit SIC's, the prospect universes are run at:

1 to 2%
2 to 3%
3 to 4%
4 to 5%
Over 5%

It is then feasible to combine data for the chosen penetration ranges to provide a Prospect Projection total at 1%+, or 2%+, or 3%+ (or for any overall penetration figure desired.)

It is of some interest to note that a penetration analysis done this way will include a sizable number of individual penetration runs:

3 for SIC (2- 4- and 8-digit SIC)

2 for Sequence (SIC and Penetration)

3 or 4 or 5 for Penetration Ranges - at various percentages plus Penetration against the Business Universe by Employee Ranges

A written report of findings makes clear:

1. Concentration by SIC
2. Spread by SIC
3. Prospect Universes at Various Penetrations
4. Analysis of Employee Strength Distribution
5. Highlighting of selected important 2- to 4-digit SIC's.

This report, if the customer file shows above average penetration of the computer markets in America (Manufacturing, Wholesaling, Retailing, Services, Education, Repairs) can include a comparison (by 2-digit only) of a file of some 4,000,000 mail-sold customers' records of PC computers for which close to 3,000,000 have SIC, and almost 2 million have number of employees.

Let us stop at this point and review some findings which prove the great value of a penetration analysis:

1. A major manufacturer had been renting for several years every manufacturing record - some 600,000 - without learning that while Manufacturing overall rated very high, at least one of the nineteen 2-digit SIC's was a total waste and this one 2-digit SIC represented over 100,000 establishments, over 1/6 of all manufacturing establishments in America. (And over 1/6th of his promotional dollars!)

2. A business mailer who had never rented wholesalers for his prospecting found that one-quarter of all his inquires, which converted into customer status, were wholesalers.

3. A major company in electrical products found that Electric Utilities were the ideal target for a rush promotion.

4. A mailer of Communication fear found "spikes" of great penetration in segments of the computer market it just had never considered as good potential prospects.

5. A company which had incorporated sales and transactions into its construct of 6 different levels of purchase, found the major difference - insofar as SIC and Number of Employees was concerned - was in Penetration. This was a bit of an eye opener to a fair-sized Marketing and Research Department.

6. One small company found that 34% of its customers were in establishments with over 500 employees. They planned large increases in mailings - but were shown that the total establishments in America with over 500 employees is just 18,000.

7. One major mailer mailing only into firms with 10 or less employees was selecting the SIC's primarily on the basis of the number of entities in given 4- and 8-digit SIC's. A brief review of penetration (conversion of prospects mailed to customer status) clearly pointed out the pitfalls in this practice.

8. One company which knew "Engineering Establishments" were prime prospects, was surprised to find that only 11 of 33 8-digit Engineering SIC's (by discipline) were really attractive markets for their product line.

9. A major producer of products for the educational market found that the key to successful prospecting was pupil enrollment. Below a given size, no school, or even a school district could be mailed at a profit.

Cost -

1. If only one file is analyzed.

and

2. If the customer is already mailing business

records in quantity.

then

3. Cost of the Penetration Analysis (all parts as detailed.) is zero if the customers commit to rent at least 100,000 SIC records taken from the DBA Business Universe File.

4. For penetrations of additional cells or selections, please inquire.

5. If the mailer cannot commit to mailing 100,000 or more DBA compiled names at this time, the Penetration Analysis on a single file will run $3,000 to $5,000 - depending mainly on the size of the file.

II.The scoring or weighted model starts with production of the Penetration Analysis.

It then makes use of 4 levels of data (all taken exclusively from the computerized files of the customer.)

These four levels will be described briefly after first reviewing just what advantages can accrue to a large business-to-business mailer who proceeds to create a Business-to-Business model - rather than stopping at the Penetration Analysis level:

A. One company selected its 1992 rentals on the basis of penetration - mailing all records - 531,000 - with a penetration of 2%. This was highly successful.

In 1993, the company added to each customer by 8-digit SIC two important pieces of data for the year 1992 -
a. Total sales dollars
b. Total number of transactions

It then selected all records from the 531,000 mailed in 1992, which matched the following parameters - for the year 1992:

1. Total sales for a given 8-digit SIC of $70,000.
2. Total transactions for a given 8-digit SIC of 350 or more.
3. Penetration for the 8-digit SIC of 2.5 or more.

This reduced the 531,000 prospect records to 345,000. This then became the distillation from 9,400,000 records of the 345,000 best prospects. And this company mailed each of these names 6 times, profitably, in 1993.

B. A $100,000,000 publishing entity on the West Coast, with 275 territorial sales representatives, was suffering from a turnover of over 50% per year - a dead loss of over $3,000,000. When the sales potentials of these fixed territories were scored, and arrayed from high to low, the range was from a score of 535,000 to a low of 34,000. It did not take a rocket scientist to recognize that the territories at the lower end made losers of their reps no matter how good they were, or how hard they tried. The potential was not there.

As a result, the territories were reduced to 175 - each with a sales potential guaranteed to provide a good living to a reasonably energetic sales representative. The turnover of representatives the following year was reduced to 10%.

C. A major seminar provider, nominally selecting sites for multiple seminars over a given period has done quite well using demographic data as to density of population, and of households by 3-digit Zip Code.

By adding a score for every establishment in a given market area (by 8-digit SIC, and no. of employees) the actual potential of each site can now be based on business potential, not simply number of firms. (The correlation is reasonably close - but there is far more confidence with the addition of scored business potential to the mix.)

(In addition, some markets eschewed in the past are now being added, basically on the basis of business potential.)

D. Those firms which have utilized Penetration Analysis for improved prospecting are now

inspecting the concept of scored potential to:

1) Compare one territory with another.
2) Compare one representative with another.
3) Provide a realistic quota system for new hires.
4) Compare for the first time sales vs. potential for Divisions. Users of this system are aware that they can provide four meaningful indexes based on:

1. The actual score per fixed territory (be it Zip, County, SCF, Metro Area, State, Section of the country.)

2. That score divided by sales to define specifically sales against potential.

3. That score divided by total customers to define specifically customer production over a period of time against potential.

4. That score divided by New Customers in last given period to define specifically new customers production against potential.

E. One variant of this is now being utilized by a manufacturer (who sells exclusively to retailers) to measure the efforts of a group of manufacturing agents. (The data is proving instrumental in substitution of company salesman provided exclusive territories with rational quotas and access to all lines. (The outside agents rarely present data, or obtain sales on more than 10% of the product line.)

The essence of the Business-to-Business model includes four types of data in five separate and distinct layers taken from the computerized file of the list owner, namely:

A. Potential - as measured against the Business Universe - by SIC and No. of Employees.

B. Dollars of sales - by 8-digit SIC.

C. Transactions by 8-digit SIC.

D. Customers - by 8-digit SIC.

E. New Customers - by 8-digit SIC.

Potential has been rather thoroughly covered in the preceding pages. Dollars, transactions and customers by 8-digit SIC's require simple accumulation - over given periods of time. Ideally this should cover a period of 3 years, broken out into 6 half years, or better, 12 quarters.

A good structure as far as continuituy is concerned might look like this:

	Year Current				Year Prior				2nd Year Prior			
Score	8	9	10	12	3	4	5	6		1	3	2
Period	1	2	3	4	1	2	3	4	1	2	3	4
Score												
28	X	X	X		X	X			X			
20			X	X						X		X
31		X	X	X								
20		X		X								
9	X			X			X	X		X	X	
4				X				X			X	
11									X			
3					X			X				
-0-	X									X		
22	X			X				X			X	
10			X					X				X

By providing weights for each period, and cumulating them for each individual SIC record, a single weight can be ascribed to each record on the file. (This is a program which a computer can handle quite easily - if the data is consistent and dollars and dates are always in the same field in the record.)

For the above example, a weight of 12 is given for Year 1, Period 4 (the most current transaction), 10 for Period 3, 9 for Period 2, 8 for Period 1.

Then follows 6 for Year Prior Period 4, 5 for 3, 4 for 2, 3 for 1. 2nd Prior Year shows 2 for Period 4. Blanks, using a very simplistic algorithm are scored at -1 for any period between 2 or more periodic transactions. This methodology permits the establishment of a single "average" figure for each 8-digit SIC.

With one recent customer, the value of a new current customer was doubled - as the product mix had changed rather drastically over the last several years.

There is need, one finds, to include a weight for transactions, as well as for dollars. (one client, working quite ably on his own, decided average dollar per transaction would embrace transactions and dollars of sales. What this missed was the fact that for a surprising number of SIC's a small number of transactions provided abnormally high "average sales per SIC" which brought a handful of large prospect lists with poor potential (penetration) into play. On examination of some of the grevious results, dollars averaging was omitted.)

In general, weighting has favored Penetration by SIC (which has already been factored by employee size). A good rule of thumb is to weight Penetration for between 33% and 50% of the total. This leaves from 50 to 67% for dollars, transactions, total customers, and new customers. For low ticket sales, dollars can be weighted more heavily than transactions. In fact, in such averaging it is always best to "throw out" the extremes, such as any sales of individual transactions which represent an appreciable portion of annual volume.

A weighting for a company with a bell curve type of sales output by transactions will find 33% for

Penetration, 33% for Transactions, 20% for Dollars, and 14% for Customers will do well. Where sales, as in some companies provide a reverse sales distribution bell curve (high at both ends, lower in the middle), the subject company should be able to provide some insight, for the future to help guide the weighting process.

In all cases, the final weight is made universal, irrespective of area, or sales representation, or products sold, and is applied as a single score for each 8-digit SIC - by each of the 9 levels of employment. The score for any geographical area - the potential in other words for that area - is simply the accumulation of the scores for each SIC record therein. The range of 535 to 35 noted earlier is quite usual. The ranges for the 3 indexes which result from comparing the raw data to the scored potential have been:

For sales against potential - a Range of 20 to 1.

For customers against potential - a Range of 25 to 1.

For New Customers against potential - a Range of 10 to 1.

In the building of a computerized model for a given business, it is well, at the beginning, to discuss ways and means to improve the validity and usefulness of the model by overlaying the basic customer data, if possible, with outside factors which are known (or are expected) to influence predictability.

The specific information that is required for such outside improvement may be:

1. Ownership, or use of certain equipment - as for example, computers, trucks, conveyors, or numbers of phone lines.

2. Geographic conditions which may affect deliverability time, local stocking, overseas procurement.

3. Profitability pattern -More than one major company finds sporadic casual purchase at one end of the customer file, and high volume purchase with deep discounts at the high end, both result in very little profit. The profit comes from the steady but modest customers in the mid-range. And adding Gross Profit to transactions blazes a trail in this dimension.