Customer Segmentation Analysis

Before describing and comparing segmentation systems it might be interesting to review the history of the development of geodemographic segmentation analysis.

Begun in the early '70s in England, the first segmentation systems were keyed to data gathered at the neighborhood level using government-generated data. In England it was Postal data, but when these early systems were brought over the Atlantic the primary data source became the Decennial Census. Unfortunately, there were no mechanisms for producing lists for, or marketing to, the areas defined by the Census geographies of Block Groups and Census Tracts.

Initially the core methodology for clustering people into groups through these new systems was regression analysis. There were also several conditions existing that made these early systems better than anything that had been available before: Computing hardware and software, while incredibly expensive, was gaining more and more prominence in data manipulation; more data were being compiled than ever before; and the use of computers was gaining ground as a tool for corporate functions other than bookkeeping and accounting. The utility of these early systems for marketing applications was limited by the constraints of a product that required the user to be reasonably well-educated in its functionality; being able to apply it to marketable geography; access to data that enhanced the government-supplied data; as well as the time constraints to generate needed information at increasingly lower levels of geography.

A pioneer in the marketing of geodemographic segmentation analysis systems within the United States was CLARITAS and their (at the time) unique PRIZM product that clustered households into 48 segments and delivered the results based upon Zip Code geography. Their methodology drew from the Decennial Census data, they then tried to correlate data from Block Groups and Census Tracts to Zip Codes. It was not perfect but it was a step above anything else available at the time. In fact, their catch phrase, "*Birds of a Feather Flock Together*", has virtually become synonymous with describing these types of neighborhood clustering systems.

Even though the Personal Computer came into being in the early '80s their capacity and capabilities were, by today's standards, non-existent (I still have the Osborne I bought in 1982 that uses 5 ¼" diskettes, had no hard drive, a 6" screen, and 64K of RAM. The modem had an overwhelming speed of 128 Baud).

However, by the mid-'80s hardware and software had made dramatic progress and with that came the ability for data providers to massage larger and larger volumes of data at a fraction of the time, and cost, that it took only a few short years before. The net result was that data and geodemographic system suppliers were able to begin gathering and analyzing data at much smaller geographies.

By the early '90s there were approximately six companies competing for dollars being spent on segmentation (cluster) analysis. Today that number has dwindled to about half that number due to mergers and acquisitions. Too, many of the companies involved in providing these analytical tools were statisticians and programmers with few that really understood marketing, much less how to apply the data in meaningful ways.

The mid-'90s saw the introduction of Household-based systems. These systems were developed by list compilers that maintained databases with nearly every household in the country – by head of household name – lists that contained more than 100,000,000 records. To the records the compilers began appending data from a myriad of sources: public record information; auto registrations; magazine subscriptions; questionnaires and surveys; sweepstakes; directories; warranty cards; and a host of other resources. Yes, inferred data (i.e., from the Census) was a part of the data but not the focal point.

The key to this new approach was to create a "dynamic" database that changed as the information on individual records changed. The net result is a resource that has incredible recency and provides marketers with up-to-the-minute demographics and lifestyle information on the consuming public. By comparison, Decennial Census data – the key component in neighborhood-based systems – are questionable the further one gets away from the reporting of the Census information. So, systems relying upon Census data from the 2000 census are just now receiving the detailed data at the Block Group level (over two years from the taking of the Census) and these data are "snapshots" since the long form detailed surveys are sampled rather than required of all households reporting.

The level of confidence in the information is suspect for a number of reasons – not the least because it "infers" that all households within the geography share the averaged information. And, the further one gets away from when the data are collected and reported, the less reliable it becomes. By the midpoint of a decade the data are five years old, by the time new data are collected the base data for the neighborhood-based systems are 10 years old. True, the companies relying upon neighborhood systems attempt to "freshen" the data and that is done by using updates supplied by state data centers that "sample" areas of their respective states

for changes in the base Census Data. However, "sampling" does not accommodate specific shifts in actual household data.

One of the true values of a household-based system, other than the reasons given above, is the fact that, as household data changes, specific households may move from one cluster to another in the period of a year – rather than every 10 years with a geography-based system.

Another point of comparison, and reason for opting for one system over another, is the level of discrimination found between neighborhood- and household-based systems. Three of the more commonly used neighborhood-based systems have between 42 and 64 clusters. This suggests two things: they are using similar "clustering" methodology and their data sources are fairly uniform. On the other hand, the household-based system we employ has 112 clusters/segments that provides a much greater level of discrimination than the three referenced. For a marketer, the more you know about an existing or potential customer, the greater the effectiveness of your communications and marketing programs.

In the final analysis, which would you, as a marketer, prefer? A system that says that "You Are Where You Live"? Or, would you have more confidence in one that says "You Look Like Someone Who Shares Similar Demographics and Lifestyles – Regardless of Where You Live"?

Here's a simple exercise. According to neighborhood-based systems, when you walk out of your house in the morning and look around, you share the same demographics and lifestyle characteristics as all of your neighbors. Do you? Really?

I doubt it very seriously. In actuality, you share the characteristics of "*Home Value*" and, by inference, "*Household Income*" (because you need to have a certain income to afford the home you live in).

With a household-based system you are "matched" to other *Households* that have essentially the same characteristics across the board.

In head-to-head comparisons, the more accurate, timely, and functional household-based system provided much clearer information for developing creative concepts and copy, not to mention more targeted prospecting lists.

This system has been used with superior results for, among others: an automobile manufacturer, religious organization, non-profits, newspapers, real estate developers, political campaigns, retailers (grocery, clothing, furniture, office products), radio and tv stations across the country.

If you'd like to see how Mosbygrey llc can help **you improve your marketing and direct marketing efforts using a truly unique and ROI-generating approach, please give us a call at 804-467-4500 or you can drop an email to us at:**

info@mosbygrey.com