

Enhance Lists with Overlay Data, Part 2

*By Jim Wheaton
Principal, Wheaton Group*

Original version of an article that appeared in the September 8, 2003 issue of "DM News"

Part 1 of this article (*DM News*, April 7, 2003) showed how demographic overlay data can improve your company's top and bottom lines. However, it is more difficult than one might think to interpret the profile reports that are generated by the overlay process. There are several analytical traps that frequently snare the untrained.

This month's article will focus on how to properly evaluate profile reports. It will also provide insight on how to determine the quality of the ones that have been supplied by your vendor. Unfortunately, many of the profile products on the market today have significant shortcomings.

Case Study

This article is based on a real analysis that was run for a specific automobile model. Because the analysis was done over a decade ago, we are able to share one of the profile tables (see Table 1). In this table, the buyers of Coupes (2-doors) were profiled on Age of Head of Household, and then compared with buyers of Sedans (4-doors) as well as a sample of the entire United States. (Shading has been added to the profile table for emphasis.)

The first three columns are straightforward. For each of fourteen Age ranges, the quantity of Coupe buyers is provided, followed by each range's percentage of the 62,492 Coupe buyers for which the Age variable was available. A critical component is the breakout of Coupe buyers for which Age was unavailable. This group numbers 50,230, or fully 44.6% of the 112,722 overall Coupe buyers. Unfortunately, a number of commercially available profile products do not include this breakout.

As was explained in the April 7 article, individuals who cannot be coded with a given data element are almost always demographically different from those who can. This is because representation on major overlay databases is skewed towards older individuals who reside in more geographically stable households. Non-coded individuals tend to be younger and more mobile. If this skew is not accounted for, then the resulting profiles will be misleading. The next two "Coupe Estimate" columns illustrate why this is true:

Generating Estimates for the Missing Component

In the first column, an approximation algorithm was used to allocate, across the fourteen Age ranges, the 44.6% of buyers for which no Age information exists. This is a sophisticated process whose

mechanics are beyond the scope of this article. The resulting allocations can be remarkably accurate, and are applicable to a number of overlay data elements. Unfortunately, many commercially available profile products do not include such estimates.

Table 1

	Coupe Actual		Coupe Estimate		vs. Sedan		
Age	Qty	Pct of Tot; Grand Tot	Qty	Pct of Total	Confi- Ratio	dence	Ratio vs. US
18-24	4,557	7.3%	27,755	24.6%	205	99%	204
25-29	8,126	13.0%	19,963	17.7%	132	99%	187
30-34	7,534	12.1%	10,656	9.5%	86	99%	93
35-39	5,888	9.4%	8,336	7.4%	77	99%	79
40-44	8,478	13.6%	10,261	9.1%	98	45%	106
45-49	10,227	16.4%	12,314	10.9%	123	99%	170
50-54	7,912	12.7%	9,554	8.5%	119	99%	159
55-59	3,886	6.2%	6,031	5.4%	93	73%	112
60-64	2,310	3.7%	3,444	3.1%	68	99%	63
65-69	1,412	2.3%	2,117	1.9%	57	99%	41
70-74	1,041	1.7%	1,165	1.0%	52	99%	27
75-79	511	0.8%	604	0.5%	50	95%	19
80-84	295	0.5%	262	0.2%	52	68%	13
85+	315	0.5%	260	0.2%	77	24%	16
Total Age	62,492	55.4%					
Missing	50,230	44.6%					
Overall	112,722						

Compare the estimated Percent of Total column with the “actuals” displayed in column three. For example, 18-24 year-olds account for an estimated 24.6% of the 112,722 overall Coupe buyers. This is dramatically higher than the “actual” 7.3% of the 62,492 Age-coded buyers. Clearly, focusing on the 7.3% would very misleading from a marketing perspective; and perhaps disastrously so! The “actuals” for the 18-24 and 25-29 Age ranges reflect significant under-representation. All other ranges reflect over-representation. For example, 45-49 year olds have an “actual” of 16.4% versus an estimate of 10.9%.

Coupe Versus Sedan Comparison

The next two columns compare Coupe versus Sedan buyers. The Ratio of 205, for example, means that the estimated 24.6% of Coupe buyers who are 18-24 is over twice that of Sedan buyers. Coupe buyers display higher concentrations among 18-29 and 45-54 year olds, as is apparent by the fact that the corresponding Ratios are over 100. Clearly, the Coupe appeals to two very different life-stages.

Interest in Coupes versus Sedans is highest at 18-24, with a Ratio of 205, and then consistently declines to a Ratio of 77 at 35-39. Then, there is increasing interest to 45-49, where a secondary “peak” occurs at 123. Next, there is a long decline to 75-79, where the primary “valley” occurs at 50. Finally, there appears to be a small up-tick beginning with 80-84 year-olds, although the corresponding Confidence statistics suggest that this is not clear-cut. (More on this later.)

From a life-style perspective, all of this makes good sense. The young, who often are single and have no children, find 2-door vehicles to be more appealing than 4-doors. So, too, do 45-54 year olds, who purchase the Coupes as secondary vehicles for themselves or their teenage children. There is lower interest in Coupes during the prime child-rearing years, when a premium is placed on 4-door practicality. Interest also declines during the senior years, because the 2-door design can be difficult to enter and exit.

Ratios are particularly important in the absence of an approximation algorithm to adjust for missing data and generate estimated Percent of Totals. This is because unadjusted “actuals,” although suspect from an absolute perspective, can be compared with each other to create valid relative metrics. For example, consider again that it is misleading to conclude that 7.3% of overall Coupe buyers are 18-24. However, even in the absence of the adjustment to 24.6%, we can be certain that the concentration of Coupe buyers in the 18-24 Age range is twice that of Sedan buyers.

The second column in this “comparison against Sedan” section is a Confidence statistic. This is based on something called a Z-Score, which is has been omitted from the table for the sake of simplicity. The Confidence statistic is frequently misunderstood. Consider, for example, the 99% Confidence that is associated with the 18-24 Age range. It does not translate to 99% Confidence that the Coupe-to-Sedan Ratio is exactly 205. Instead, it quantifies the likelihood that the Percent of Total for Coupe buyers is statistically different than for Sedan buyers.

Confidence is a function of the similarity between the Percent of Total Coupe and Percent of Total Sedan buyers, and of their corresponding sample sizes. By definition, a Ratio of 100 will have a 0% Confidence, regardless of the sample size. Notice that the 40-44 Age range has a Confidence of only 45%. The primary reason is that its Ratio of 98 is very close to 100. Conversely, the primary driver for the low Confidences associated with the 80-84 and 85+ ranges is small sample size. Therefore, the apparent up-tick in their Ratios might be nothing more than a statistical mirage.

Coupe Versus U.S. Comparison

The final column compares the Age distributions of Coupe buyers versus a sample of the U.S. population. Coupe buyers display higher Ratios among 18-29 and 40-59 year olds. The peaks and valleys for the U.S.-derived ratios are often more extreme than for the Sedan-derived. (The Confidence column has been deleted for simplicity. All Age ranges in this example displayed Confidences at or close to 99% because a very large sample of the U.S. population was employed.)

Conclusion

Demographic overlay data can improve your company's top and bottom lines. However, without robust profile reports, and their knowledgeable interpretation, it is easy to draw inaccurate marketing conclusions. Be particularly mindful of the effects of missing data, and the corresponding need to accurately adjust profile reports. (For additional reading on this topic, see "Individual/Household Demographics & Psychographics: Applications in Descriptive & Predictive Research," The Direct Marketing Association's *1997 Research Council Journal*, www.wheatongroup.com.)