

Merge/Purge Can Alter List Strategy

*By Jim Wheaton
Principal, Wheaton Group*

Original Version of an article that appeared in the September 3, 2001 issue of “DM News”

What could merge/purge possibly have to do with reading the results of test rental lists? The answer, quite surprisingly, is, “Plenty!”

Individuals who appear on multiple rental lists (“Multis”) generally have a significantly higher response rate than those who appear on only one (“Singles”). Therefore, the way in which Multis are allocated (“credited”) to the lists in which they appear will affect the corresponding reported response rates. This article will illustrate two methods of allocating Multis, and their effect on reported response rates.

Assume that 15,000 names are rented from each of five rental lists, and that every list-pair (for example, the intersection of Lists A and B) has 2,250 names in common. This overlap translates, as we will soon see, into 52,500 of the 75,000 total input names surviving the record matching step of the merge/purge. This is a “dupe rate” of 30%, which quite is typical for major merge/purges.

Exhibit 1 illustrates how this 2,250-name, list-pair overlap translates into 9,000 unallocated Multis, and 6,000 Singles, per list:

Exhibit 1: List Overlap

	Gross Names	List A	List B	List C	List D	List E	Unallo-cated Multis	Singles
List A	15,000	N/A	2,250	2,250	2,250	2,250	9,000	6,000
List B	15,000	2,250	N/A	2,250	2,250	2,250	9,000	6,000
List C	15,000	2,250	2,250	N/A	2,250	2,250	9,000	6,000
List D	15,000	2,250	2,250	2,250	N/A	2,250	9,000	6,000
List E	15,000	2,250	2,250	2,250	2,250	N/A	9,000	6,000
Total	75,000						45,000	30,000

Please refer to Exhibit 2 throughout the balance of this article, as we discuss the two methods of allocating Multis. We will assume that the Multi response rate is 75% higher than the Singles, or 1.40% versus 0.80%. (Typically for niche direct marketers, the response rate for Multis is between 50% to 100% higher than for Singles.)

**Exhibit 2:
Two Methods of Allocating Multis**

		----- Random -----			----- Hierarchical -----				
	Singles	Credited Multis	Total Net Names	Resp Rate	Credited Multis	Total Net Names	Resp Rate	Change Versus Random¹	Change Versus List A¹
List A	6,000	4,500	10,500	1.06	9,000	15,000	1.16	+ 9.7%	N/A
List B	6,000	4,500	10,500	1.06	6,750	12,750	1.12	+ 5.7%	- 3.7%
List C	6,000	4,500	10,500	1.06	4,500	10,500	1.06	0.0%	- 8.9%
List D	6,000	4,500	10,500	1.06	2,250	8,250	0.96	- 8.8%	- 16.9%
List E	6,000	4,500	10,500	1.06	0	6,000	0.80	- 24.3%	- 31.0%
Total	30,000	22,500	52,500	1.06	22,500	52,500	1.06		

¹ Apparent discrepancies are caused by rounding.

Random Allocation

Here, Multis are allocated on a random basis among their contributing lists. We saw in Exhibit 1 that each list-pair has 2,250 names in common. With random allocation, each list will be credited with 50% of the names per list-pair, or 1,125.

We see in Exhibit 2 that, with random allocation, each list is credited with 4,500 Multis; that is, 50% of 2,250 names across each of four lists. Along with the 6,000 Singles, each list corresponds to 10,500 total names. This results in a uniform reported response rate of 1.06%; that is, the weighted average of 0.80% for the 6,000 Singles and 1.40% for the 4,500 Multis. For illustrative purposes, this assumes that all lists are equally responsive.

Hierarchical Allocation

In this case, a list hierarchy determines the allocation of Multis. Whenever a name overlap appears between two lists, the higher-ranked list receives credit for 100% of the Multis.

In our Exhibit 2 example, the list hierarchy is A, B, C, D, E. As a result, List A is credited with 9,000 Multi's; that is, 2,250 from each of Lists B through E. Along with the 6,000 Singles, List A corresponds to 15,000 total names. Because of its large percentage of Multi's, List A's overall response rate is 1.16%; that is, the weighted average of 0.80% for the 6,000 Singles, and 1.4% for the 9,000 Multi's. This is 9.7% better than the overall response rate of 1.06% across all five of the lists.

Likewise, List B is credited with 2,250 Multis from each of Lists C through E, or 6,750 in total. List E, however, receives credit for no Multi's. Therefore, its response rate is 0.80%, which is a full 24.3% lower than the overall response rate, and 31.0% lower than List A.

The frequent motivation behind the hierarchical allocation of Multis is the minimization of list rental charges. If, for example, List A costs \$100 per thousand and List E \$110, then – all other things being equal – it is economical to credit List A for all of the Multis.

The downside of this approach is that it distorts reported response rates, and can result in unfortunate rollout decisions. Assume, for example, that List A has a rollout universe of 50,000 and List E a universe of 500,000. Assume also that List E's response rate is unacceptable on a hierarchical basis, but acceptable on a random basis. For the sake of \$10 per thousand, or one cent per piece, a major rollout universe will not be uncovered! This is, literally, "penny wise and pound foolish."

Final Thoughts

Random allocation among test rental lists generates response reports that accurately reflect the economic value of each list, and provides a basis for making rational rollout decisions. Although it does not minimize rental charges, random allocation is generally recommended because improved rollout decisions invariably outweigh the associated increase in rental charges.

Modern record matching algorithms offer a hybrid allocation option for Multis called "Random, Within Hierarchies." A future article will outline why it is typically the desired strategy for record matching jobs that include buyers, inquirers, established ("rollout") rental lists, and unproven test lists.

Jim Wheaton is a Principal at Wheaton Group, and can be reached at 919-969-8859 or jim.wheaton@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts. Jim is also a Co-Founder of Data University www.datauniversity.org.