

## **RFM Cells: The “Kudzu” of Segmentation**

*By Jim Wheaton  
Principal, Wheaton Group*

*Original Version of an article that appeared in the July 15, 1996 issue of “DM News”*

“Kudzu” is a four-letter word in the Southeastern U.S. A plant native to Japan, it grows like crazy and – the point of this analogy – is difficult to eradicate.

The same is true of Recency-Frequency-Monetary (“RFM”) Cells, which have thrived for years despite the existence of more sophisticated statistics-based predictive models. Statisticians argue about which modeling technique is superior – regression, neural networks, genetic algorithms, and the like. But they generally agree that RFM should be relegated to history’s dustbin, to paraphrase a famous nineteenth century analyst.

I took my own shot at RFM in a two-part *DM News* article (December 11, 1995 and January 15, 1996), a financial simulation that illustrated how switching from RFM to a properly conceived and executed predictive model often generates a positive ROI on the first mailing – even for moderate-sized database marketers.

RFM remains in the news – *DM News*, to be exact. A recent issue advertised software that helps create and implement an RFM-based segmentation strategy, and ran yet another “how to” article.

Here, I’ll focus again on RFM, but from another perspective. I’ll show why there exist only two possible end-results of RFM:

1. A stable and easy-to-implement, but crude, segmentation strategy.
2. A strategy that is complicated but not particularly sophisticated, and that is both unstable and a nightmare to implement.

Let’s assume that a retailer has four years worth of point-of-sale transactions, consolidated in a database of one million customers. Also, our retailer has decided on RFM Cells to determine who should be mailed the monthly sale flyer. The following is the process required to define these Cells:

1. Five by-month Recency ranges are selected: 0-6, 7-12, 13-24, 25-36, and 37<sup>+</sup>.
2. Four Frequency categories are settled on: 1, 2, 3-4, and 5<sup>+</sup>.
3. Five Average Order Size Monetary groups are established: \$0-\$25, \$25.01-\$50, \$50.01-\$100, \$100.01-\$200, and \$200.01<sup>+</sup>.

(Note: The reader can substitute his or her industry of choice – financial services, telecommunications, catalog, fundraising, and the like. The concepts to be discussed are the same.)

Our retailer has defined 100 RFM Cells (5 X 4 X 5), a manageable quantity. With an average Cell size of ten thousand (1 million / 100), sufficient sample size is available to analyze past promotions and construct a selection hierarchy. Also, 100 Cells will be relatively easy to implement.

But wait a minute! Our retailer sells several thousand SKU's, with price points ranging from \$0.99 to \$2,995. Clearly, not all merchandise is created equal, and a given customer's past purchase patterns will be a strong predictor of future behavior. Although segmentation by SKU isn't practical, our retailer decides to create six Merchandise Categories. With this, we're up to 600 Cells (5 X 4 X 5 X 6).

However, we're not done yet! One of the best determinants of retail loyalty is Distance from the store. A customer who lives two miles away, for example, generally will spend more than one who's thirty miles away. After some thought, our retailer comes up with five distance categories: 0-2 miles, 2.01-5, 5.01-10, 10.01-20, and 20<sup>+</sup>. With this addition, we now have 3,000 Cells (5 X 4 X 5 X 6 X 5).

And finally, our retailer realizes that customer life cycle is critical to predicting purchase behavior. It's impossible to include all of the demographic overlay variables that are likely predictors of behavior – Age, Income, Marital Status, and Presence of Children, to name a few. Therefore, our retailer settles on just Age, divided into five groups: 18-25, 26-40, 41-50, 51-65, and 65<sup>+</sup>. This results in a final Cell count of 15,000 (5 X 4 X 5 X 6 X 5 X 5).

Consider the ramifications of a 15,000 Cell segmentation strategy:

1. It will be difficult to determine the correct selection hierarchy, because the average Cell size is a mere 67 (1 million / 15,000). This, of course, is far from being statistically significant. In order to attain workable quantities, our retailer will have to undertake the tedious task of manually combining many of the Cells.
2. The implementation will be daunting, because each Cell that's selected for a given promotion will require a separate line of programming code.

With all of this instability and complexity, our Cell strategy isn't particularly sophisticated. Our retailer had to compromise by collapsing several thousand merchandise SKU's into a mere six Merchandise Categories, and by not including several relevant demographic variables. Nor was it possible to consider many other likely predictors of future purchase behavior, such as merchandise returns, method of payment, and out-of-stocks.

None of these difficulties would exist with a statistics-based predictive model. All potentially predictive customer characteristics could be input to the modeling process. There would be none of the sample-size issues that are inherent in RFM Cells. And the result of the model – a rank-ordering of customers by their predicted future purchase volume – would result in a straightforward implementation: every customer above a certain point score would be promoted, and the others would not.

In summary, predictive models are more stable than RFM Cells. They're easier to implement. And, they're substantially more powerful.

As a final note, even database marketers who prefer a Cell-driven segmentation strategy that can be created on-site should dump RFM. This is because there exist sophisticated, statistics-based tree analysis

tools, such as CHAID or CART that are far superior. But that's another topic, and the subject of next month's article.

*Jim Wheaton is a Principal at Wheaton Group, and can be reached at 919-969-8859 or [jim.wheaton@wheatongroup.com](mailto: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](http://www.datauniversity.org).*