John R. Miglautsch, PDM

Thoughts on RFM Scoring

John Miglautsch received his Master of Arts from the University of Wisconsin-Madison in 1985, and his Bachelor of Arts with Distinction from UW-Madison in 1976. He is a Senior Fellow of The North American Society for Strategic Marketing and the founder and Executive Director of the International Society for Strategic Marketing. John is the former Secretary of the Board of Directors of the Wisconsin Direct Marketing Club. John speaks at Marketing conferences all over the world and is a frequent contributor to *DMNews*.

ABSTRACT

ISSM's Executive Director shares his thoughts on Recency, Frequency and Monetary (RFM) scoring and other marketing techniques, e.g. weighting, life-to-date and quintiles.

Thoughts on RFM Scoring

1.   RFM Basics

Direct marketing is fundamentally the scientific control of customer acquisition and contact. The recurring question is whether Customer A merits an additional contact based on his past purchase behavior. This question applies equally to direct mail, catalog, phone, field or Internet contact[1]. The process of making this decision is customer segmentation. Not all customers have purchased identical amounts.   Some have ordered more often, some have ordered more recently. Consequently, not all customers should be contacted with the same effort and expense. The cornerstone of direct marketing segmentation is RFM (Recency, Frequency and Monetary values).

Since direct marketing segmentation is a science, it is important to quantify customer behavior so that we can test the short and long term effect of our segmentation formulae. The purpose of RFM is to provide a simple framework for quantifying that customer behavior. Once customers are assigned RFM behavior scores, they can be grouped into segments and their subsequent profitability analyzed. This profitability analysis then forms the basis for future customer contact frequency decisions.

2.   RFM Scoring

The purpose of RFM scoring is to project future behavior (driving better segmentation decisions). In order to allow projection, it is important to translate the customer behavior into numbers which can be used through time.

Too often, direct marketers will use static customer selections. When initially building their segmentation system, they may consider their best customers to be those who have purchased more than, say, $100. If the mailer is relatively new, this definition will degrade rapidly. The initial selection of >$100 may have encompassed 20% of the customer file. After a year or two, it may easily identify the top 30%-40%.   If we evaluate the profitability of our first contact based on the $100 cutoff, we will see very positive results. Given the successful segmentation experiment, the direct marketer will again use the $100 cutoff. After substantial time elapses, that same $100 will yield poorer results. It is not that the best customers have significantly changed, it is simply that over time, more customers will have repeat purchases and achieve the $100 threshold.

---

[1]The cost per contact varies substantially between field sales and email (Internet) contact. In the later case, customer behavior is not so much a justification for additional contact expense as much as an indication of the depth of customer relationship. The greater the purchases, the less likely the customer will be annoyed, and the greater the risk that he is.

Two common scoring methods are used to avoid the bracket creep problem.

        a.   Customer Quintiles

The most common scoring method is to sort customers in descending order (best to worst). Customers are then broken into five equal groups or quintiles. The best receive a score of 5, the worst a 1 (see Figure 1). For Recency, customers are sorted by days since last purchase, the lower the number of days, the better the score. For Frequency, customers are sorted by number of purchases, the higher the number of purchases, the better the score. And for Monetary, customers are sorted by the amount spent.[2] T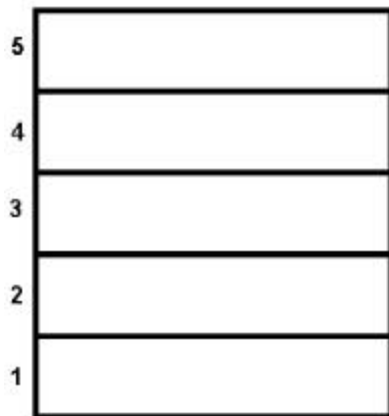he higher the amount, the higher the score. Each time customers are scored, a new relative segmentation scheme is created. This has the advantage of quantifying customer behavior which can be projected into the future. The relatively best customers would always fall into the 5,5,5 category. It is necessary to identify where the cutoff points fall, since they automatically change with each customer scoring.

Figure 1 Equal numbers of customers in each group make analysis simpler and easier to understand.

The customer quintile method has the advantage of yielding equal numbers of customers in each segment. There are five equal groups for RFM, generating 125 equal size segments overall. Initial analysis would be to contact all customers, look at the performance of each individual cell (cells would have definitions like: 4,3,5 or 2,3,3) and understand how different segments of the customers perform.

With 600,000 customers there would be 4,800 in each cell. A response rate of 2% would yield 96 orders giving you an acceptable sample for analysis[3]. With less than 600,000 customers, it would be highly questionable to evaluate each cell independently. Instead, the RFM would be evaluated by looking at the relative performance between the R scores, the F scores and the M scores. This may not be as satisfying, but it would provide statistically significant results. Thus, a 100,000 customer mailing would have 20,000 in each grouping (looking only at one dimension at a time). This method extends the usefulness of RFM down to the neighborhood of 10,000-25,000 customers[4].

---

[2] Substantial variation can be made in Monetary scoring. If there is wide variation in cost of goods percent, it is better to accumulate gross margin rather than gross revenue. If there is a high percentage of returned merchandise, it would be wise to subtract returns from customer totals. In the second case, customer service costs are also often subtracted from the total.

[3] Though each cell would have the same quantity, response rate would vary. It is beyond the scope of this paper to explore confidence interval calculations at the cell and the mailing level.

[4] 2,000 customers in each cell at an average 2% response would generate 40 responses. Confidence level would be ±10%. Most mailers would consider this minimally projectable. However, it is important to note that the 2%
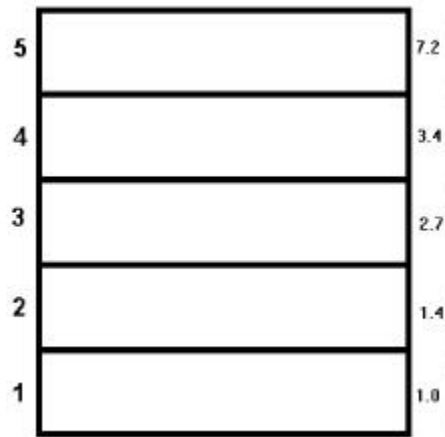
Figure 2 In this example, about 28% of the customers have only ordered once. This means over 1/3rd of the customers in score 2 have the identical behavior as score 1.

The customer quintile method does encounter some scoring challenges in the area of Frequency. In most direct marketing customer files, a high percentage of the customers have only ordered once. This percentage is often as high as 30%-60%. If more than 20% of the customers have only one purchase, then the lowest Frequency group will have a purchase amount of 1(see Figure 2). Since that group cannot hold all the customers with only one purchase, some of them will be sorted into the 2 score group. Note, their behavior is identical to those in the 1 score, they simply spilled over. If 40% of the customers had only one purchase, then both 1 and 2 score groups would have identical behaviors. If the percentage ran as high as 60% (which is not that unusual) then three of the five quintiles would have the identical behavior. Remembering the purpose of RFM (which is to quantify behavior), this would be a less than satisfying result.

A second concern with the quintile method is its relative sensitivity. At the high end of our Frequency model (see Figure 2) customers average 7.4 purchases. That is considerably more than the 1.0 purchases at the bottom and almost twice as great as the 3.4 purchases in the 4 score group. However, the Paretto Principle (commonly called the 80/20 rule) still applies inside the 5 score group. This means that there are a small number of very large customers and a larger number of relatively smaller customers who make up that 7.4 average.

As long as our segmentation scheme is primarily built for mailing purposes, this distinction is moot. Undoubtedly the 5 and 4 groups would be mailed. However, if our RFM model is being used to facilitate telemarketing or field sales contact, additional sub-segments would be crucial to identify the super customers.

The customer quintile scoring method generates some unsatisfying results at both the top and bottom of the scale. It tends to group together customers who have vastly different buying behavior (at the top) and arbitrarily break apart customers who have identical behavior (at the bottom).

 b. Behavior Quintile Scoring
An alternative scoring method has been developed by John Wirth[5], PhD. It also sorts customers by behavior but, instead of creating arbitrary cutoffs at a certain percent of the customers, it generates cutoffs on percentage of behavior. This method seems to overcome the sensitivity

---

would not be consistent across cells.

[5] John did his course work in aeronautical engineering. He is the founder of Woodworker's Supply of New Mexico, a leading hand tool catalog and retailer.

problems mentioned above. Five groups are still created, but Monetary score would generate equal amounts of sales in each quintile.

Behavior scoring has the advantage of grouping customers by similar behavior. Since segmentation decisions are based on past customer behavior, this allows better segmentation.
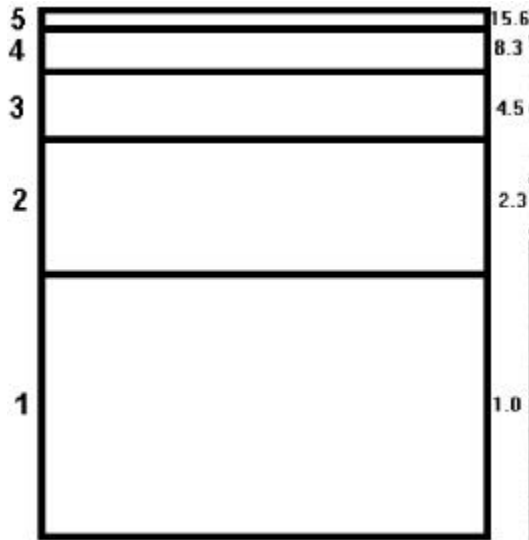

Figure 3 The Mean method yields sensitivity at both top and bottom but also isolates single purchasers.

i.   Frequency

The Behavior method does suffer from similar problems when tackling Frequency score. If we start at the top of the Frequency sort and subtract each customer's frequency from total Frequency, the customers who have purchased only once may not equal 20% of total Frequency. In that case, some of the customers who have purchased twice will be included in the 1 score group with this method.

It is also troublesome to sort customers from top to bottom in a computer generated scoring system. A special sort file must be created and each scoring process must be accomplished uniquely. The Mean scoring method, a further enhancement of the John Wirth method has been developed by Ted Miglautsch, V.P. Development, Miglautsch Marketing, Inc. When scoring Frequency, the single purchasers are given a score of 1. The system then averages the remaining Frequencies to determine the mean. If a customer total falls below the mean, he receives a score of 2. This process is repeated two more times giving us quintiles of behavior which approximate the John Wirth method, have sensitivity on both ends of the scale and allow scoring of many variables at the same time.
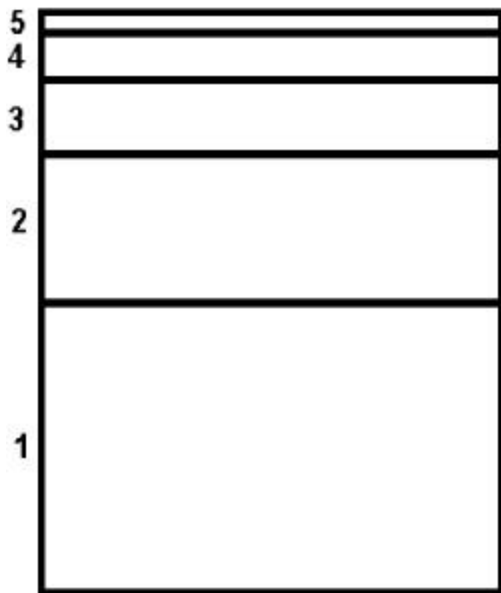

Figure 4 Though 20% of sales is in each score, the 5 score group may represent only 4% of the customers and the 1 score group may represent more than 40%.

ii.   Recency

Since past behavior is the best predictor of future behavior, Recency is typically considered the most powerful of the three variables. Many direct marketers make contact decisions based solely on Recency.

Recent customers are considered viable for a certain length of time. They are often mailed heavily

in the first 12 months and increasingly less often until, say, 36-48 months. After that they are considered dead.

Unlike Frequency and Monetary, customers reset themselves. A three-year-old reordering customer who has purchased an average amount only once moves up in an orderly manner from a 1 in F&M to a 2 in F&M. But in Recency, he jumps from a 1 to a 5! Customers who order often may never have anything other than a 5 score. At the core of Recency is the fact that most of the customers fall into two groups: hot and dead.

Though Recency can be scored by sorting customers by days since last purchase, industry list convention suggest a more calendar based method. "Hotline names" typically mean purchasers within three months or 90 days. Often, marketers work very hard to make sure the most recent names are pulled out of the order processing system, sometimes within hours of the cutoff deadline. Borrowing from the Wirth method of segmenting the hottest names more finely, the most common R score breaks Recency: 0-3 months, 4-6 months, 7-12 months, 13-24 months and 25+ months. Business-to-business direct marketers often extend these time frames since their customers can remain viable even though individuals change.

       c.   Weighting

With relational, database-driven marketing databases becoming more common, most marketers can select RF&M scores independently. However, others are not as fortunate and require a single field to do the work of all three variables. The advantage of a single variable is that customers can easily be segmented by a single query on one field.

Donald R. Libey, in his book Libey on RFM, suggests that Monetary, Frequency and Recency values can be added together[6]. Scoring is not explicitly discussed but he does offer a formula for creating a single RFM value. His method includes adding average order and Frequency per year. An example is a customer who bought 60 times, with an average order of 300, total purchases of $18,000 and a Recency of 1 (year). Adding them all together, you get a score 18,373. This method is actually a form of weighting. Monetary value, because it is so large in comparison with Recency, Frequency and Frequency per year overpowers the weighting.

An alternative would be simply to add together the RFM scores discussed above. The best customers would have a composite score of 15 (5+5+5) and the worst customers would have a minimum score of 3 (1+1+1). Many of the customers would have a score of 7 or 8 and it would be difficult to sort them. Further, the experience of decades of direct mail marketing suggests that the most recent customers are of greater value than those who have ignored more than a few repeated mailings. To enhance this composite formula, many mailers multiply Rx3, Fx2 and Mx1. This would give the best customers a composite score of 30 (5x3)+(5x2) +(5x1). This not only gives more power to the most Recent names, it also gives a bit of a boost to Frequency.

---

[6] Libey, Donald R., Libey on RFM, e-RFM.com

The logic behind weighting Frequency is that if two customers have equal Recency, spent the same amount but one ordered several times and the other only once, the more frequent buyer is much more likely to respond. If the choice were which one to mail your last catalogue to, the choice would be the more frequent buyer.

One additional enhancement is often employed in creating a composite score. Instead of multiplying by 3,2&1, substitute 9.9, 6.6 and 3.3. This yields a range of composite scores between 99 and 19.8. It preserves the approximately 3x weighting of R, it also creates more of a 100 point scale.

       d.   Life-to-Date

In general, RFM scoring is based on life to date totals. It is often asked whether it would improve RFM scoring to shorten up the time frame. The concept is that if Recency is so powerful, perhaps we should consider only the recent behavior of the past few years. An excellent suggestion but fraught with danger. The basic concept again is quantifying behavior for the purpose of customer segmentation. It is obvious that high RFM customers are easily identified. The true challenge is to identify viable customers beyond the 12 month window. Should any of them be mailed? Certainly some should - those who have spent a considerable sum and those who have ordered more than once (or perhaps twice). To gain this wider perspective, it requires that all available customer history be analyzed.

3.   Conclusion

RFM continues to provide the foundation for customer behavior quantification. As noted, there is considerable variation in scoring methods; each has its own strengths and weaknesses. It is hoped that this discussion assists marketers in forging a solid base for measuring, understanding and executing customer segmentation.