

---

# Making a database pay off using recency, frequency and monetary analysis

Arthur M. Hughes

*Received (in revised form): 9th June, 1995*

---



**Arthur M. Hughes** is Executive Vice President of ACS, Inc., a database marketing company in Reston, Virginia. Arthur has been designing, building and maintaining customer marketing databases since 1985 for Fortune 500 companies including banks, insurance companies, automobile companies, oil companies, pharmaceutical companies, telephone companies, etc. His book 'The Complete Database Marketer — tapping your customer base to maximize sales and increase profits' has sold more than 5,000 and his second book, 'Strategic Database Marketing', was published in April 1994.

Arthur lectures and gives seminars on database marketing in the United States, Canada and throughout the world, including Malaysia, Portugal, Japan and Australia. His two-day marketing seminars for senior marketing executives in Canada are given four times a year at York University. In the US he gives two-day seminars four times a year for the Database Marketing Institute in Arlington, Virginia. He has been teaching Economics at the University of Maryland for the past 30 years.

---

## ABSTRACT

*This paper describes the application of RFM (recency, frequency, monetary) analysis to customer databases in order to use such databases most profitably. It explains how to code a database for RFM, the theory underlying it and some practical examples of how to make it actually pay off.*

## INTRODUCTION

The principal obstacle to effective database marketing is the development of profitable strategies for use of the database. It is relatively easy to construct a workable

marketing database. Many service bureaux are experienced at this work and can do a very satisfactory job. What the service bureau normally cannot help with, however, is working out how to make the database pay off.

One of the oldest, and still one of the best, techniques is Recency, Frequency, Monetary (RFM) analysis. Using this method, any marketer with a large customer database can almost guarantee profitable promotions to the customer base time after time.

There are alternatives to RFM analysis such as modelling based on demographics,

cluster coding, market research, focus groups or customer surveys. RFM analysis is preferred because all the components of RFM are behavioural, ie when the person purchased last, how often he or she purchased and how much money has been spent. These are the factors — which are readily available — which are used to predict customer behaviour. Predicting behaviour based on recent behaviour is a much more accurate and powerful method than predicting behaviour based on any other known factor. Analysis in this paper will demonstrate why this is so.

## **CODING A DATABASE**

To carry out RFM analysis, all customer records must contain certain purchase history data and be properly coded. This assumes, of course, that the customer database can keep track of customer response. This is true of telephone companies, oil companies, retail stores which use a proprietary card, catalogue mailers, insurance companies, travel and leisure, banks and many others: probably more than half of the database marketing situations.

In each customer record three pieces of information must be maintained:

- the most recent date that the customer has requested a change in service, purchased a discretionary item, etc.
- a counter for the frequency — the

number of times the customer has made a purchase, or continued service with you. For a telephone company, for example, it might be the number of months of continuous service; for a retail store, it would be the total number of store visits. This counter is incremented by one every time a purchase is made

- a counter for the monetary amount — the total amount of money the customer has spent with the company.

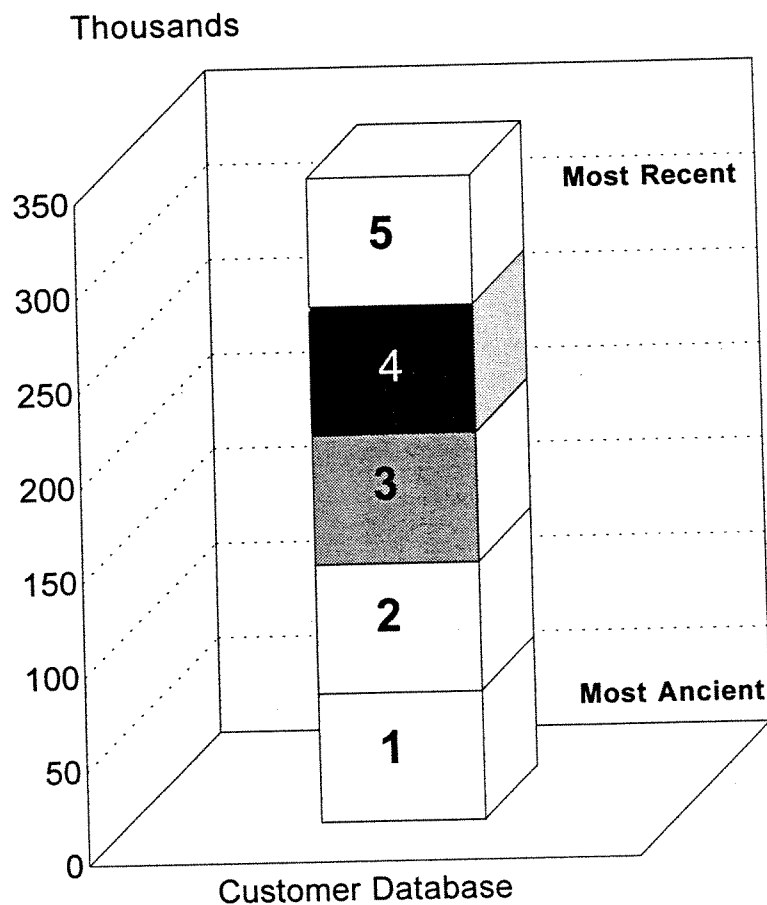
Using these three pieces of data, the RFM codes will be constructed.

### **Constructing a recency code**

To create a recency code, all the records in the database are sorted by most recent date, with the most recent at the top and the most ancient at the bottom. Once this has been done the database is divided into five exactly equal parts (quintiles), as shown in Figure 1.

To the top group (the most recent) the number 5 is assigned and inserted in each customer record. To the next group, a 4, etc. Everyone in the database then has a recency code of 5, 4, 3, 2, or 1. It is important not to fall into the trap of determining the most recent group to be 0–3 months, the next 4–6 months, etc. This will cause problems later on. If a promotion is offered to the customer base a breakdown of responses by recency code will probably look something like that shown in Figure 2.

**FIGURE 1: ASSIGNING REGENCY QUINTILE NUMBERS DATABASE SORTED BY MOST RECENT PURCHASE**



Why is this so? There are very few absolutes in marketing, but one of them is that the people most likely to respond to a new offer are those people who have made a purchase from the company most recently. A customer is more likely to act on a promotion if they have recently had a satisfactory transaction with the company. This is true of retail stores, software houses, automobile companies, insurance firms. It is a universal phenomenon. It can

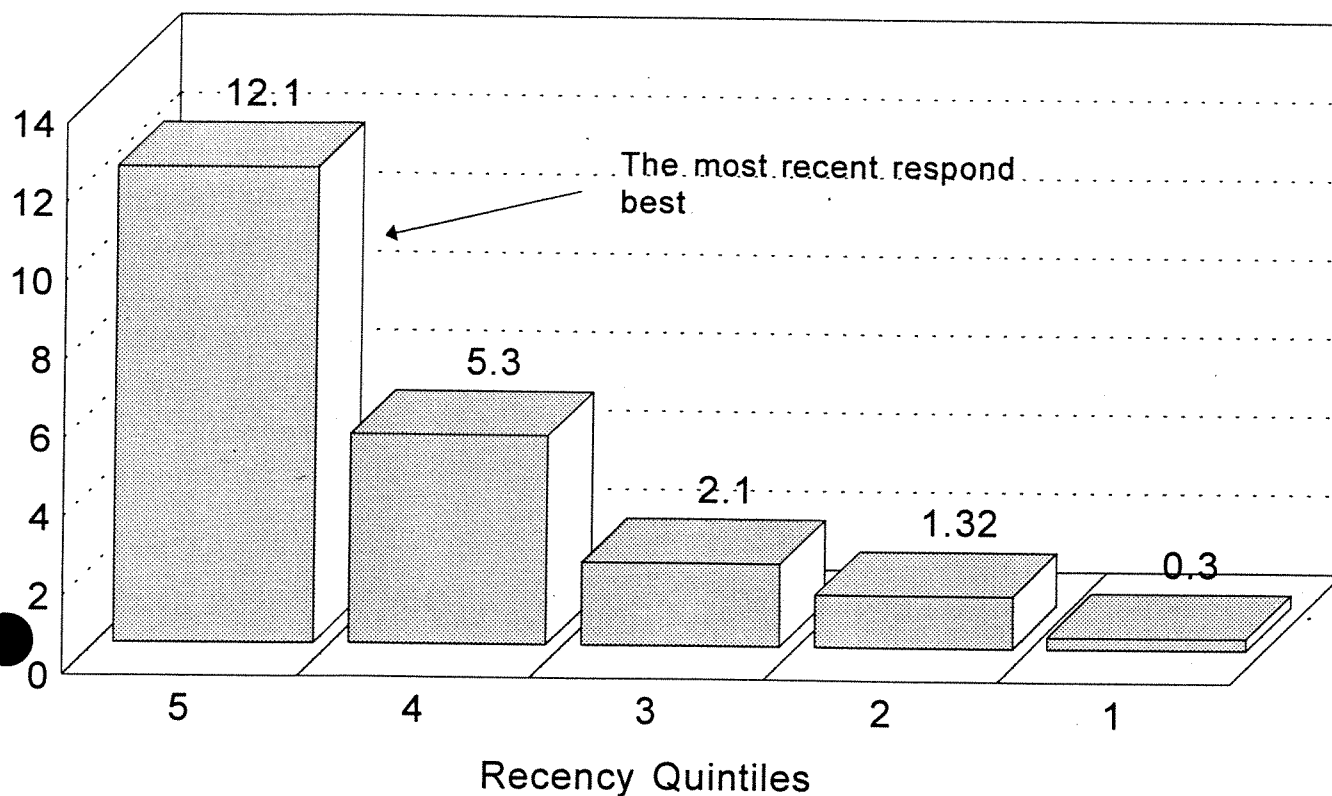
be tested at very cost little by running a program to insert regency codes into the database and, the next time a routine promotion is run, the response by regency code can be checked.

### **Constructing a frequency code**

A frequency code is constructed in exactly the same way. The entire file should be sorted from most frequent down to least frequent and frequency codes applied to

**FIGURE 2: RESPONSE BY RECENCY**

Per cent of those mailed who responded



each quintile, so every record now has a frequency code of 5, 4, 3, 2, or 1. The frequency code in the customer records should be put right next to the recency code. In effect, a two-digit code has been created in every customer record, which varies from 55 (most recent and most frequent) down to 11 (most ancient and least frequent). Each group will have exactly the same number of records.

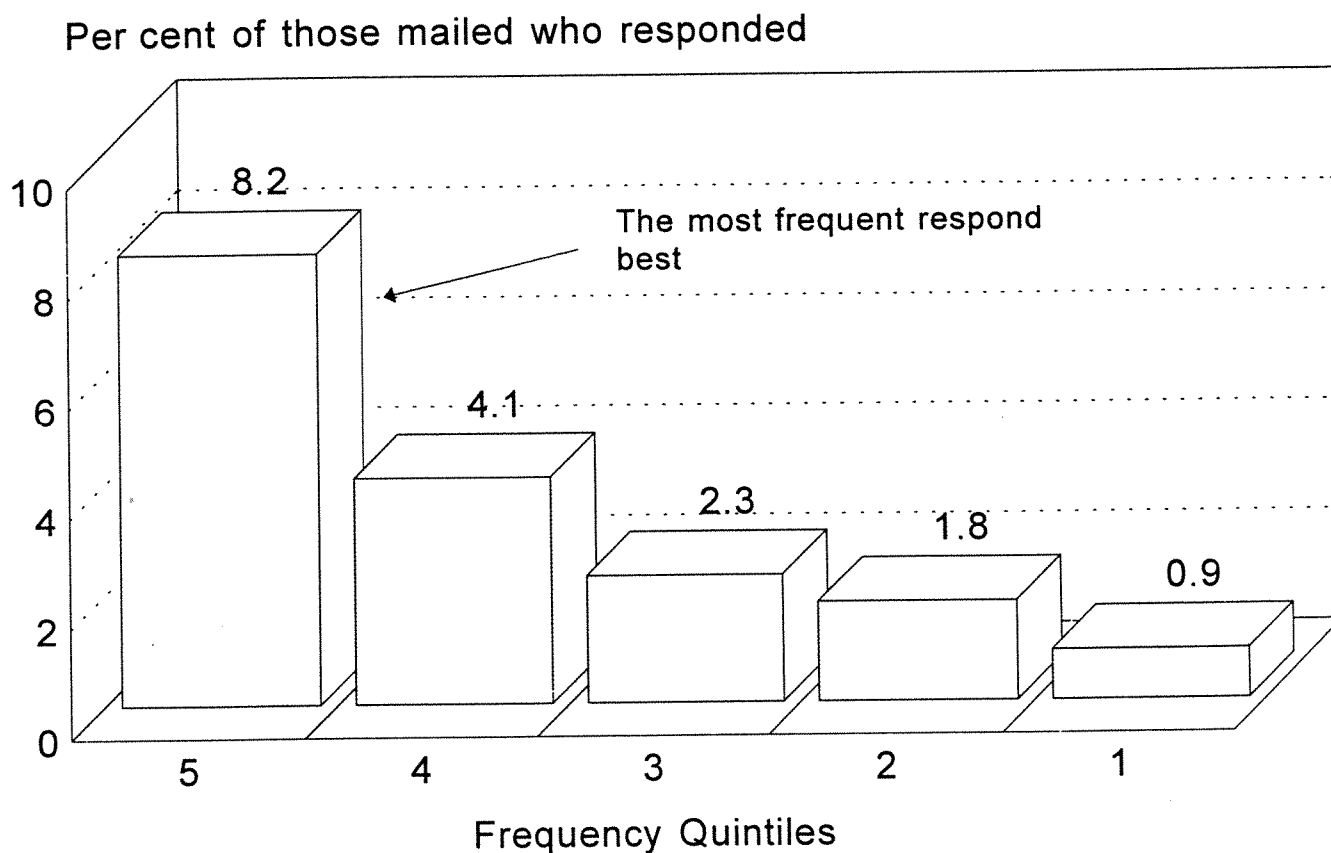
On the next promotion a graph of responses by frequency can be drawn. It

will certainly look something like the graph shown in Figure 3. It will be noted that the difference in response between quintile 5 on frequency is not as great as the difference in the first and second recency quintiles. This is because recency is, generally, a more powerful predictor of customer response than frequency.

#### **Constructing a monetary code**

Construction of a monetary code is done in exactly the same way as for the

**FIGURE 3: RESPONSE BY FREQUENCY**



previous two codes. The entire file should be sorted by money spent with the greatest amounts at the top. A 5 is assigned to the top quintile, 4 to the next, etc. If the promotion response is measured by monetary amount, the response pattern is likely to resemble that shown in Figure 4.

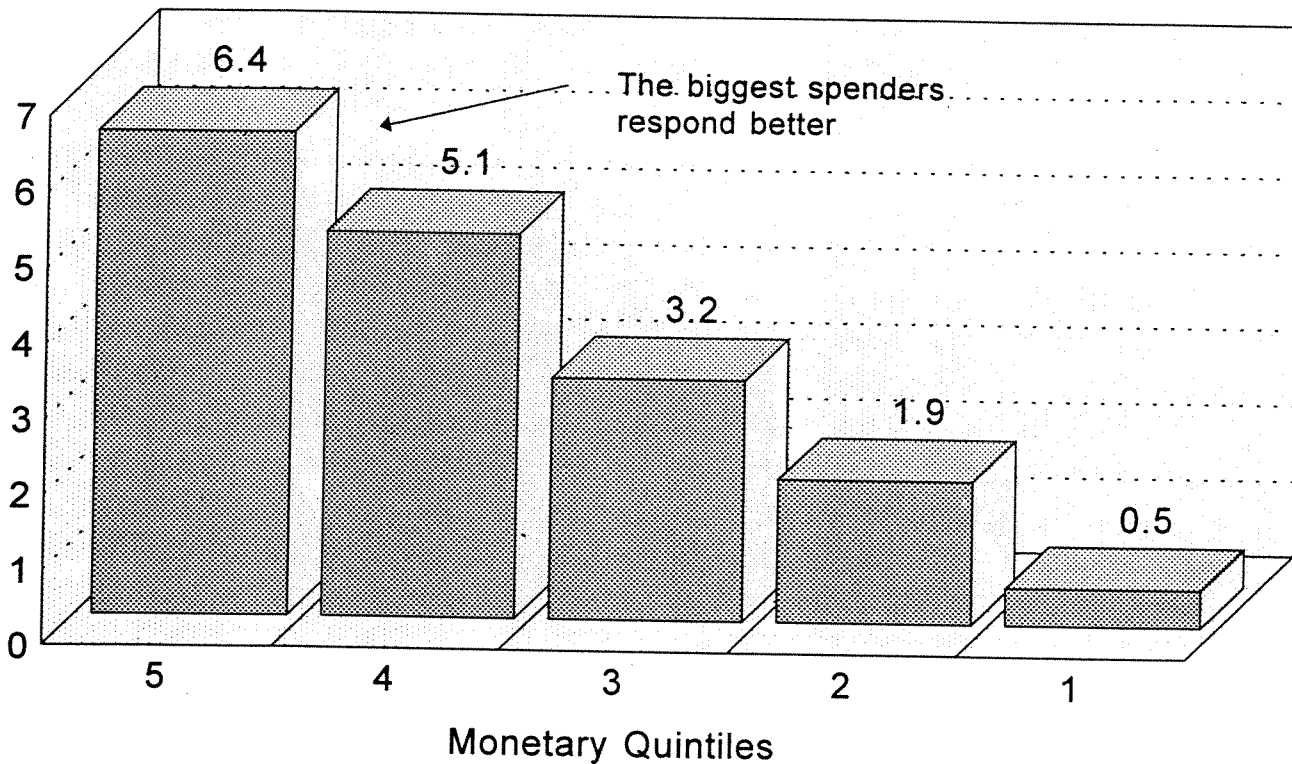
The monetary codes should be inserted next to the frequency codes. Everyone in the database will now have a three-digit code in their customer record, from 555

down to 111. There are 125 RFM cells in all. These should be recalculated and revised every time that the database is updated — typically once a month. A space should be provided in the customer record for previous RFM cells so how people have moved during the past month can be measured — it is possible to keep track of RFM cells even earlier, six months for example.

The power of the codes can now be illustrated by an example involving a test

FIGURE 4: RESPONSE BY MONETARY AMOUNT

Per cent of those mailed who responded



promotion to be sent to 40,000 customers.

The first step is to select 40,000 customers from the database using an Nth. This is a computer program that automatically selects every Nth record. To determine an Nth, the number of records in the test group is divided into the number of records in the database universe. If there are 800,000 records in the customer database, dividing by 40,000 is 20. That means that every 20th record is selected so the 1st, the 21st, the 41st, etc., will be picked. The 40,000 test will then

be an exact statistical replica of the main database — no matter in which order the main database had been sorted (alphabetical, by zip code, customer number, etc.). There will be an exact statistical sample of the RFM codes as well. There will be 125 different RFM cells, and each one will have exactly the same number of customers in it.

In this example the offer made to these 40,000 customers is for a product that costs about \$120 and the net variable profit from a successful sale is \$35. The

---

**TABLE 1: RESPONSE TO 40,000 TEST MAILING**

<i>Cell position</i>	<i>RFM cell</i>	<i>Number mailed</i>	<i>Number responses</i>	<i>Response rate</i>
A	B	C	D	E
1	555	320	31	9.69%
2	554	320	30	9.38%
3	553	320	28	8.75%
4	552	320	20	6.25%
5	551	320	19	5.94%
6	545	320	26	8.13%
7	544	320	20	6.25%
8	543	320	18	5.63%
9	542	320	16	5.00%
10	541	320	12	3.75%
11	535	320	14	4.38%
12	534	320	10	3.13%
13	533	320	10	3.13%
14	532	320	9	2.81%
15	531	320	7	2.19%
16	525	320	13	4.06%
17	524	320	10	3.13%
18	523	320	8	2.50%
19	522	320	8	2.50%
20	521	320	7	2.19%

---

cost of the mailing (including creative, printing, personalisation and postage) is \$0.62 per piece. The response to the offer might look like that shown in Table 1.

These are only the first 20 cells. There are 125 in all. Lower cells usually have lower response rates. Those towards the bottom may have no response at all. (All

cells are of exactly equal size because of the method by which they are created. Each quintile (R, F, and M) is exactly 20 per cent of the entire file. It follows that each of the 125 cells will be exactly  $\frac{1}{125}$  of the entire file. If the coding is done correctly, there will be no larger or smaller cells.)

**TABLE 2: TEST PROMOTION USING A RETAIL CUSTOMER FILE**

	<i>Number</i>	<i>Cost/each</i>	<i>Cost/revenue</i>
Number mailed	40,000	\$0.62	\$24,800
Number responded	711	\$35.00	\$24,885
Response rate	1.78%		
Net profit from promotion			\$85

The results for this test promotion could look like those shown in Table 2 which were actually taken from a retail customer file where the promoted item was an article of men's and women's apparel which sold for about \$125.

Most people would consider this promotion a failure. Almost \$25,000 has been spent on a mailing which has netted a profit of only \$85 after all expenses have been paid. That certainly does not cover time spent planning the promotion, even if it does cover the mailing costs. However, something very valuable has been learned from this unsuccessful promotion — the response rate by each RFM cell. This is worth a great deal.

The basic theory behind RFM analysis is that the best predictor of future behaviour is past and present behaviour. It is certainly better than demographics, psychographics or geographics. Since the RFM cells selected for the test are identical to the RFM cells in the entire database, it

can be assumed that these unmailed customers will respond in the same way if given the same offer. Since it is known whether a given cell will respond or not, the promotion to the unprofitable cells can be suppressed and the profitable cells concentrated on.

### **DETERMINING PROFITABLE RFM CELLS**

The profitable cells are those whose response is greater than breakeven. There is a simple formula to calculate breakeven:

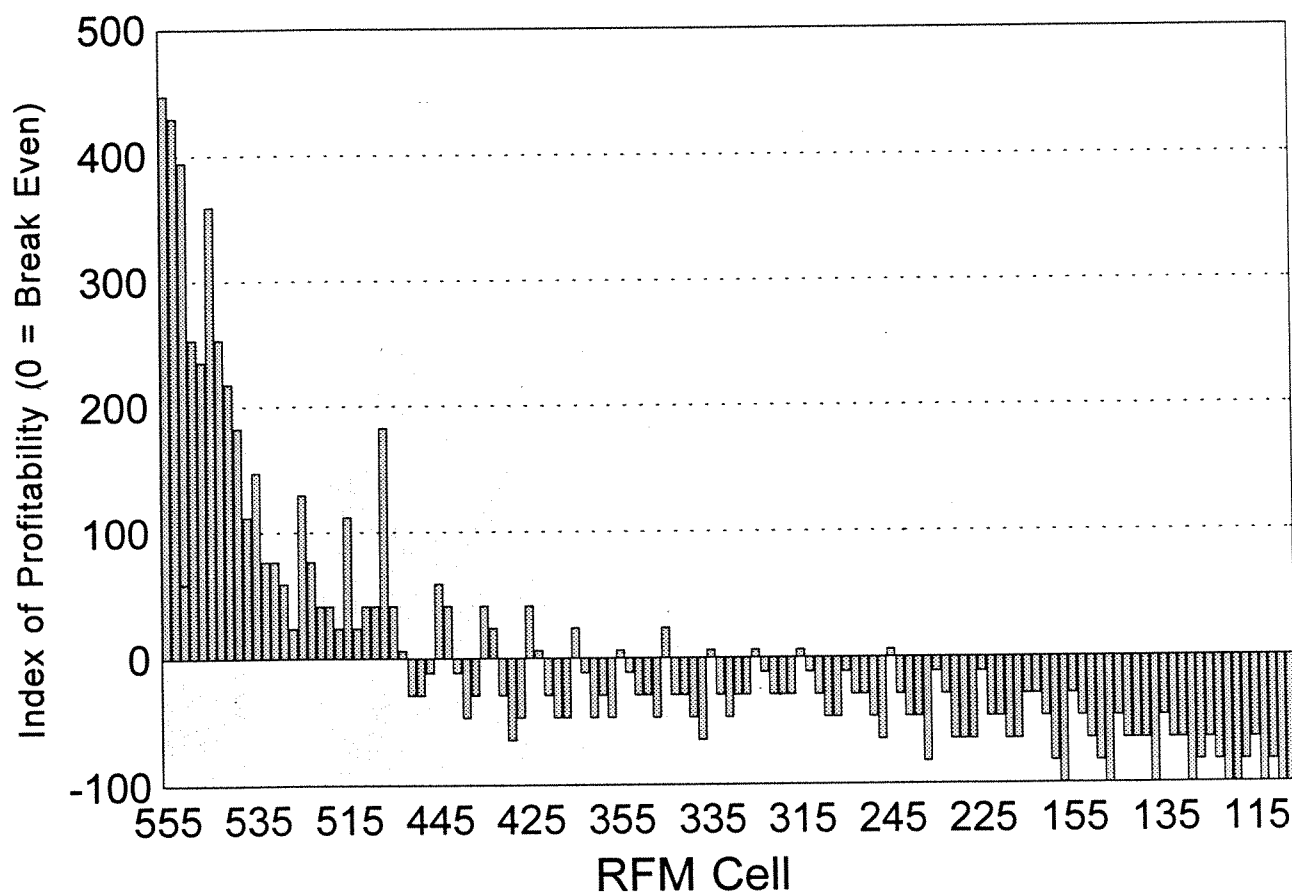
$$\text{Breakeven} = \text{mail cost per piece} / \text{Net revenue per sale}$$

In the above example, the mail cost was \$0.62 and the net revenue per sale was \$35.00. Therefore, the breakeven response rate is:

$$\text{Breakeven} = \$0.62 / \$35 = 1.77\%$$



FIGURE 5: RESPONSE BY RFM CELL TO 40,000 TEST MAILING



Any RFM cell that has a predicted response rate of 1.77 per cent or better should be promoted. The balance should be dropped. A graph of the profitable and unprofitable cells is shown in Figure 5.

Data for this chart were prepared by dividing the response rate by the breakeven rate, multiplying by 100 and subtracting 100. Breakeven rates, therefore, are shown as zero. This is an

index of profitability. Any cell with a score above 0 will be profitable, all those below will be unprofitable.

Knowing this predicted profitability, it is a simple matter to make a profit from a roll-out promotion: mail only the profitable cells. There is one preliminary step, however, that will ensure success — the test results must be discounted.

It is well known in direct marketing that

**TABLE 3**

	<i>Test</i>	<i>Roll-out</i>	<i>Selected roll-out</i>
Mailed	40,000	800,000	211,200
Responded	711	12,104	7,293
Response rate	1.78%	1.51%	3.45%
Mail cost	\$24,800	\$496,000	\$130,944
Net sales revenue	\$24,885	\$423,640	\$255,255
Net profit	\$85	(\$72,360)	\$124,311

the test always does better than the roll-out. What this means is that if there is a 2 per cent response rate to a test, there will be less than 2 per cent when the entire file is mailed. The author's explanation for this is that marketers cannot stand to have an unsuccessful test. They will *always* take steps to make sure that the test has a high response rate. They assume that if their test does not work out, their reputation will suffer, so they pick the best lists, the best time of year, the best offer — in many cases things that cannot be replicated in a roll-out promotion.

In this case, the list cannot be fudged because an Nth has been used. However, there are subtle things that will creep in to prejudice the test results. For this reason, it is recommended that the test results be discounted by 15 per cent in determining the RFM cell response in a roll-out. 25 per

cent or some other number can be used. This means that if there was a 2 per cent response rate in some RFM cell in the test, it should be assumed that there will be only a 1.7 per cent ( $2.0 \times 0.85$ ) response by that same cell in the larger mailing.

Using these discounted test response rates, only those cells which score higher than breakeven should be concentrated on. The results can be quite startling, as Table 3 shows. What Table 3 shows is that if the entire database had been promoted, the response rate would have been 1.15 per cent (that is 85 per cent of the test rate of 1.78 per cent); 12,104 items would have been sold at a net profit of \$35 per sale. This would bring in \$423,640. But the total promotion would have lost \$72,360, after subtracting the mailing cost of \$496,000.

Using the discounted test results, there are only 211,200 customers that fall in

profitable RFM cells — cells whose discounted response rate is 1.77 per cent or higher. The discounted response rate is determined by discounting the test results from each RFM cell by 15 per cent. Mailing only these 211,200, there are only 7,293 sales for a net revenue of \$255,255. Since the mailing costs are only \$130,944, the net profit on the entire project is \$124,311.

This is not a fluke but a normal RFM analysis. It will always work, if there are the right data to construct valid RFM cells, and if the methods are followed rigorously. Almost any valid test can be turned into a profitable roll-out.

## TEST SIZE

One question that comes up is how large (or small) the test should be to be valid. A basic rule is that the expected responses should be 500 or greater. To know how large that is, it is necessary to have some idea as to the size of the expected response. If there is normally a 5 per cent response from customers, then, to produce 500 responses, only 10,000 customers have to be tested.

The expected responses should be 500 or more because there are 125 RFM cells; 500 responses mean that the average RFM cell will have four responders (500/125). When the average cell response is less than four, the law of chance becomes more important in the results than the law of

customer behaviour. Of course, the larger the test is, the more accurate will be the results. A test of 40,000 will be more accurate than a test of 30,000. It will also cost more. Will the extra cost of the 10,000 be worth it in terms of accuracy in roll-out selection? That decision will have to be made in the particular circumstances of each company and experiments done over time to learn the ideal test size for the customer database.

## SOME EXCEPTIONS

As can be seen from the analysis, RFM works only with customer databases — and only with customer databases that contain the necessary data. It is of no use with a prospect database. If the customer database has only part of the data needed (for example has recency and frequency, but lacks the monetary amounts) there can still be some benefit out of trying the technique.

Determining recency can be a problem. With an electric utility, telephone company, newspaper or insurance company, for example, customers often receive monthly statements and have to pay their bills every month. If last bill paid were used for recency, every current customer would be in the top quintile which would make it useless. In this situation, for most recent date, the last time the customer changed their service, or last time they called customer service or

---

some other similar date should be used. The idea is to get the last time that they thought about the company and did something. At such times, the company has been logged in to their conscious memory.

### **CELL PERSONALITY**

RFM has been illustrated by assuming that a direct promotion was to be run using either mail or telemarketing. There are many other uses. If, for example, it has been decided to create special customer segments which will get special treatment (gold-card holders, for example) RFM is an ideal way to find out who should go into the top category.

Each RFM cell has a personality of its own. All new customers, for example, enter as 511s. They are the most recent, but usually the least frequent and have the lowest dollar amount. After they enter, they can move either up or down, based on their subsequent behaviour. It is a good idea to track where people were last month. Goals can be set for each cell, to get new customers to advance in their second month from a 511 to a 512, for example, rather than moving down to a 411.

The lowest group, 111s, may not be customers at all. These could be archived, by taking them off the database, after trying a reactivation mailing on them. To keep them on the database year after year

can just waste company money and their time. The 155s should not be dropped without a struggle, however; special attention should be paid to them.

### **MAINTAINING CONTACT**

If RFM cells are used as a basis for all the company's customer contact, touch will be lost entirely with some customers — the ones below the breakeven point on the tests. This may be good, or bad. After all, the purpose of building a relationship with customers is to build profits. Eighty per cent of company revenue usually comes from the top 20 per cent of customers — those with a 5 in Monetary.

Discrimination is good for business — and will also be accepted, or even welcomed, by customers. Welcomed because some people do not want to be bothered by their suppliers all the time. A 111 may be trying to say 'Leave me alone, please'. Both parties may profit by a parting of the ways.

### **INFLUENCING BEHAVIOUR**

Does RFM analysis influence customer behaviour? Not really. RFM is used simply as a method of segmenting a customer file based on recorded behaviour from the past and using this segmentation method to predict the future. Customer behaviour will be influenced more by the communications that are generated in the

---

process of using RFM analysis. For example, a reactivation mailing may be directed at people whose recency quintile is a 1. A welcome letter may be sent to new customers, and they may be encouraged to refer new customers to the company. Customers in high RFM cells might be told that they have achieved 'gold' status which provides certain benefits to them, the benefits of 'platinum' status which requires only a little more effort on their part can be described. So, RFM does not influence behaviour by

itself, it just helps to target behavioural promotion efforts.

## CONCLUSION

There is really no limit to the interesting and profitable marketing experiments that can be performed with RFM analysis. As the author is continuing to do research on the subject, he would be delighted to hear from readers who have case histories or anecdotes which might be interesting for others.