

Chapter 12

RFM Analysis

Abstract Recency (R), Frequency (F), and Monetary Value (M) are the most popular database marketing metrics used to quantify customer transaction history. Recency is how recently the customer has purchases; frequency is how often the customer purchases, and monetary value is the dollar value of the purchases. RFM analysis classifies customers into groups according to their RFM measures, and relates these classifications to behaviors such as the likelihood of responding to a catalog or other offer. RFM analysis was probably the first “predictive model” used in database marketing. This chapter discusses the RFM framework, how it can be used and various extensions.

12.1 Introduction

How do you select customers for target mailing? Or whom should you send your catalogs or direct mail offers to? The need to mail smarter is always among the top concerns of direct marketers. The direct mail promotion that results in sales to 2% of the mailed universe is considered a success. Identifying and targeting the customers who are most likely to respond are therefore of prime concern.

Because of the nature of their businesses, direct marketers including catalogers have been collecting customer data, analyzing them, and developing models for several decades to improve their business performance. One popular approach used to improve mailing efficiency is the RFM – Recency, Frequency, Monetary amount – model. The primitive form of the RFM model was used about 50 years ago by catalogers of general merchandise. For example, as early as 1961, George Cullinan promoted the use and understanding of RFM customer data analysis. Recognizing his contribution in advancing the direct marketing industry, the DMA (Direct Marketing Association) inducted him into the DMA Hall of Fame in 1989.

The core concept of the RFM model is based on the empirical evidence. Direct marketers have found that the response to a mailing offer is heterogeneous across customers. And they also found that customers who have

responded well in the past are likely to respond in the future. More specifically, direct marketers have found that customers' purchase response can be predicted using their previous purchase histories. The three most important variables to summarize customers' purchase histories are recency (R), frequency (F), and monetary amount (M). That is, using RFM measures for each customer, one can predict his or her propensity to respond. Once identifying who is going to respond, the direct marketer sends catalogs to customers with high propensity.

This chapter first discusses the fundamental concepts of the RFM model. Second, using the notion of a breakeven point, we study how the model can be used to determine the number of customers to mail to, in order to maximize profits. In Sect. 12.2, highlighting the relationship of RFM model with other statistical tools, we criticize its current status and investigate the possibility of extending its potential. We conclude that while simple RFM analysis may provide a good starting point, statistical model-building using the raw customer data is the better option.

12.2 The Basics of the RFM Model

Suppose that a direct mail company has its house list with a million customers. Each season it decides to mail catalogs to a subset of customers from the house list. Sending catalogs to all customers will maximize its revenue or sales. But it may lose profits if the average response rate is too low. There are some customers who will not purchase the product whatever the company does. The company would like to select good customers who are likely to respond to the catalogs. The goal of an RFM analysis is to predict the response (or purchase) probability of each customer. Sending catalogs to only perhaps 20% of its customers based on these predictions (the good customers), the company can now make a profit.

12.2.1 Definition of Recency, Frequency, and Monetary Value

Based on their experience, direct marketers have found three important purchase-related variables that will influence the future purchase possibility of their customers. The first variable, recency (R), represents the last time the customer purchased from the company. It stands for the elapsed time (measured in days, weeks, months or years) since the last purchase. For example, suppose you randomly select 10,000 customers from a cataloger's house list. Your objective is to find the relationship between the recency and the response probability. You first choose a particular catalog mailing, say the June