Erasmus University Rotterdam

Tatev Karen Aslanyan

Customer Segmentation: RFM Model

We use the RFM model as a benchmark segmentation method for clustering customers into different clusters. There are several reasons for the popularity of RFM model as an approach for ranking, comparing and classifying customers. Firstly, this method is straightforward and contains only three variables making it easy to understand. Secondly, the transaction data used in the method is usually stored in a database system which makes it easy to access and extract (Lumsden, 2008). Finally, RFM is very valuable in predicting response and can boost a company's profits in the short term (Khajvand, 2011). However, this model also has some disadvantages. Wei, (2010) argue that the simplicity of RFM model has been overvalued as the model can only use a limited number of variables, while other characteristics can influence a customer's response and purchase. Moreover, it ignores the analysis of new firms setting up in a short period but also customers that only purchase once and placed small orders. Finally, as most customers do not buy recently and regularly and spend little on their purchase, the model may also underestimate this large proportion of customers Miglautsch, (2002). Using the transactions data of the customers as an input, RFM model determines and gives as an output R-score, F-score, and M-score based on those variables described in Section 3.2. The score variables are determined based on the Pareto Principle (80-20 rule) which implies that 80% of the company's revenue come from 20% of customers, we divide each variable into 5 groups using the 20% threshold.

Recency	**R score**	**Frequency**	**F score**	**Monetary**	**M score**
quantile-1 (20%)	5	quantile-1 (20%)	5	quantile-1 (20%)	5
quantile-2 (20%)	4	quantile-2 (20%)	4	quantile-2 (20%)	4
quantile-3 (20%)	3	quantile-3 (20%)	3	quantile-3 (20%)	3
quantile-4 (20%)	2	quantile-4 (20%)	2	quantile-4 (20%)	2
quantile-5 (20%)	1	quantile-5 (20%)	1	quantile-5 (20%)	1

Table 1: RFM scores quantiles

Table 1 presents this transformation procedure from Recency, Frequency and Monetary numerical variables to categorical variables of scores. The first quantile of variable Recency includes 20% of most recent customers (R-score 5) while quantile-5 includes 20% of least recent customers who have made their last purchase in the early stages of the two years (R-score 1). Quantile-1 of Frequency includes the 20% of most frequent buyers (F-score 5) and quantile-5 represents the 20% of least frequent customers (F-score 1). Finally, the customers who are in quantile-1 of Monetary variable are the 20% highest spenders (M-score 5), and the last quantile of Monetary variable represent the least 20% spenders (M-score 1). The higher the R, F, and M scores are, the greater the potential value a customer can possess.

RFM-based segmentation

A simple and widely-known approach for customer segmentation is combining R, F, and M scores in different ways to obtain some meaningful segments. Table 2 presents one possible setting defining the segments. The segment "Best customers" is the customer segment with the highest scores for R, F and M. These customers are the most frequent and recent buyers but also spend the most money compared with the rest of the customers. The "Loyal customer" segment contains customers with high spending and low recency but slightly less than "Best customers". However, the frequency of purchases of these type of customers highly depends upon the promotion period. "Potential loyalist" is the group of customers having high recency (R-score 3-5) along with spending a fair amount of money on purchases (M-score 1-3). The frequency of this category's customers is expected to be average (F-score 1-3), so they have the potential to become "Loyal customer". "New customers" are those customers involved in most recent purchases (R-score 4-5) but the frequency and spendings are on the low level (F- and M-scores 1-2). The "Need attention" segment refers to customers having above average recency along with frequency and spendings on purchases being above average as well (R-, F- and M-scores 2-3). This segment of customers needs to carefully examined as they are the ones who are most likely to churn and go from Better to Good if they are not provided proper attention. The "At risk" segment contains customers who have ordered frequently and spent a good amount of money on purchases, but that happened a long time ago. This segment might represent the customers who are not any more happy with the present deals and discounts offered by the company and their loyalty towards the brand is fading. The "Lost customers" is the one with mediocre recency, frequency and monetary amount (R-, F- and M-scores 1-2).

	Segment	**Description**	**R**	**F**	**M**
1	Best Customers	Have ordered recently,buy often and spent the most	4-5	4-5	4-5
2	Loyal Customers	Spent lot of money,are responsive to promotions	3-5	3-5	3-4
3	Potential Loyalist	Recent customers that order more than once and spent large amount of money	3-5	1-3	1-3
4	New Customers	Bought more recently but not often	4-5	1-1	1-1
5	Promising	Recently ordered but didn't spent too much money	3-4	1-2	1-2
6	Need Attention	Above average recency,frequency,monetary value	2-3	2-3	2-3
7	About to sleep	Below average recency,frequency,monetary value	2-3	1-2	1-2
8	At Risk	Spent lot of money and ordered often but long time ago	1-2	2-5	2-5
9	Can't be lost	Have spent lot of money and often but very long time ago	1-2	4-5	4-5
10	Lost Customers	With lowest frequency,monetary and recency scores	1-1	1-1	1-1
11	Others	Remaining customers that don't fall under any of the above defined categories			

Table 2: Segment definitions

This segmentation method, which is usually used as a benchmark method, has two significant disadvantages. Firstly, RFM-based segmentation is based on R, F, and M scores by applying the Pareto assumption(20% rule) but one could also use 10% or 30% decision rules for defining the scores, and there is no universal decision rule for this. All these rules lead to different results, and one is not able to state which one is better. Secondly, this clustering approach assigns equal weights to Recency, Frequency and Monetary value as an indicator of importance or allows to assign random weights which again leads to inaccurate results. Secondly, RFM-based segmentation assumes that all three model variables Recency, Frequency, and Monetary have equal weights. According to Hughes, (2000), this should also be the case, and all three variables should have the same weight. However, there are many ways of assigning different weights to R, F and M value variables. Miglautsch, (2000) argues that the weights need to reflect the order of importance in the model such that Recency has weight 3, Frequency has 2, and Monetary has 1. Khajvand, (2011) applied the Analytic Hierarchy Process method based on experts' views to get a relative weight for each variable. So, very regularly users assign different weight to these variables based on their perception or just by their assumption without providing the reason behind their choice of weights which leads to inaccurate estimation and segmentation results.