CA - S5: RFM

Josep Curto, IE Business School

June 06, 2018

Abstract

This technical note introduces what is RFM and its benefits and limitations.

Contents

The problem

- **Problem**: we don't understand our customer. We don't know if a specific customer is relevant for our organization.
- **Goals**:
  - We want to know how much value a customer is generating for the organization.
  - We want to know how the generated value is going to evolve.
  - We want to segment customers based on value.
- **Why?** To perform specific actions for different customer groups

But attention, our organization is **starting** the customer analytics journey and we can not propose really sophisticated techniques. What we can do?

How we solve the problem

- If our company has contractual customer and data about every transaction, we can propose RFM
- If not, then we need to use statistic to predict the customer's behaviour.

What is RFM

RFM Analysis is a customer segmentation method based on:

- Recency
- Frequency
- Monetary

It is easy to implement and understand. It is based on pareto principle: **20% of customers generate 80% of revenue.**

We want to spend our efforts on the best customers.

The origins of RFM can be found in the brick-and-mortar retail industry, where customers couldn't be identified with certainty. It is based on assumption that a valuable customer has a combo of recent, frequent, and expensive purchases.

Main Concepts in RFM

The main concepts are:

- **Recency**: the number of days that have passed since the customer last purchased - How recently did the customer purchase?
- **Frequency**: number of purchases in a specific period (for example, last 12 months) - How often do they purchase?
- **Monetary**: value of orders from a given customer in the specific period - How much do they spend?

Let's try to understand these concepts in detail:

- Recency implies that a client who has recently interacted with your business, all else equal, is more valuable than one you hasn't.
- Frequency assumes that a repeat buyer is inherently more valuable than a new buyer.
- Monetary = the assumption that valuable customers will continue to be valuable.

How to measure the value

- [**BU**] Discuss whether RFM fits as a metric in our business
- [**DP**] Identification and understanding of sources and metadata
- [**DP**] Extract, clean and load data
- [**M**] Create R,F,M variables per customer
- [**M**] Assign RFM percentile (use from 5 -offline- up to 10 -online- percentile segments) (and split values into bins)
- [**E**] Analyze results
- [**D**] Present and explain the results

Different ways of binning

When you divide your R, F or M datasets into chunks, that's also called 'binning', and each bin is a ranked category. There are two different methods of binning that you may come across: independent and nested binning.

- **Independent binning**: Simple ranks are assigned to recency, frequency, and monetary values. The three ranks are assigned independently. The interpretation of each of the three RFM components is therefore unambiguous; a frequency score of 5 for one customer means the same as a frequency score of 5 for another customer, regardless of their recency scores. For smaller samples, this has the disadvantage of resulting in a less even distribution of combined RFM scores.
- **Nested binning**: In nested binning, a simple rank is assigned to recency values. Within each recency rank, customers are then assigned a frequency rank, and within each frequency rank, customer are assigned a monetary rank. This tends to provide a more even distribution of combined RFM scores, but it has the disadvantage of making frequency and monetary rank scores more difficult to interpret. For example, a frequency rank of 5 for a customer with a recency rank of 5 may not mean the same thing as a frequency rank of 5 for a customer with a recency rank of 4, since the frequency rank is dependent on the recency rank.

Number of Bins

The number of categories (bins) to use for each component to create RFM scores. The total number of possible combined RFM scores is the product of the three values. For example, 5 recency bins, 4 frequency bins, and 3 monetary bins would create a total of 60 possible combined RFM scores, ranging from 111 to 543.

- The default is 5 for each component, which will create 125 possible combined RFM scores, ranging from 111 to 555.
- The maximum number of bins allowed for each score component is nine.

Ties

A "tie" is simply two or more equal recency, frequency, or monetary values. Ideally, you want to have approximately the same number of customers in each bin, but a large number of tied values can affect the bin distribution. There are two alternatives for handling ties:

- *Assign ties to the same bin.* This method always assigns tied values to the same bin, regardless of how this affects the bin distribution. This provides a consistent binning method: If two customers have the same recency value, then they will always be assigned the same recency score. In an extreme example, however, you might have 1,000 customers, with 500 of them making their most recent purchase on the same date. In a 5-bin ranking, 50% of the customers would therefore receive a recency score of 5, instead of the ideal value of 20%. Note that with the nested binning method "consistency" is somewhat more complicated for frequency and monetary scores, since frequency scores are assigned within recency score bins, and monetary scores are assigned within frequency score bins. So two customers with the same frequency value may not have the same frequency score if they don't also have the same recency score, regardless of how tied values are handled.
- *Randomly assign ties.* This ensures an even bin distribution by assigning a very small random variance factor to ties prior to ranking; so for the purpose of assigning values to the ranked bins, there are no tied values. This process has no effect on the original values. It is only used to disambiguate ties. While this produces an even bin distribution (approximately the same number of customers in each bin), it can result in completely different score results for customers who appear to have similar or identical recency, frequency, and/or monetary values – particularly if the total number of customers is relatively small and/or the number of ties is relatively high.

We can use different business/statistical rules rules to create the bins:

- Using **quantiles**: split *Recency*, *Frequency* and *Monetary* using 25th quantile, median and 75th quantile. We obtain 4 segments. We can use other quantiles splits to generate more segments. Traditionally, 5 or 10.

- Using **survival analysis**: split *Recency* at 25% and 50% of customer churn probability
- Identify **high-value customers** by splitting out the top 10% in *Frequency* and *Monetary*
- Separate **one-time buyers** from customers with **repeat purchase**

RFM limitations

- Easier to calculate (it only needs customer id, transaction amount and date), but less predictive power.
- Only for contractual settings
- Omit demographics, psychographics

Contractual vs. Non-contractual Settings

- In non-contractual settings, organizations can not know wheteher a customer is still active or not (alive). The status is a (latent) variable that needs to be predicted.
- In non-contractual settings, the time at which a customer becomes unactive is unobserved.
- How we differenciate customers in an ended relationship vs. a long hiatus between transactions.
- In reality we are dealing with two problems:
  1. The prediction of "Alive" or "Dead"
  2. The probability of purchase/s
- To understand these approaches is to imagine each customer as having two separate coins one that they flip to "decide" to answer 1. And the other for 2. (if applicable).
- For example, the basic Pareto/NBD model simulates two events. It uses a "coin" to determine whether a customer churns and then it uses "dice" to determine how many items a customer will order. The coin is modeled using a Pareto distribution and the dice are modeled using a negative binomial distribution. The more information you have on a customer, the better the models can fit them to a specific distribution and the more accurate the predictions end up being. The Pareto/NBD model uses order history as the primary input, and in particular takes into account the frequency and the recency of orders.
- The more information you have on a customer, the better the models can fit them to a specific distribution and the more accurate the predictions end up being.

Benefits

- Forecasting and Sales reporting based on RFM
- RFM profiling
- Churn Analysis
- Marketing actions: mailing, call center,...

Use Cases

- Create market strategies based on RFM. For example, targeted email lists or A/B testing within groups.
- Customer segmentation based on RFM
- Forecasting and customer evolution per segment
- Create different communication, services and loyalty programs based on RFM
- Awake "non-active" customers

Customer segmentation with RFM

You can create different types of customer segments with RFM analysis, but here are 11 segments we recommend. Think about what percentage of your existing customers would be in each of these segments. And evaluate how effective the recommended marketing action can be for your business.

Customer Segment	Activity	Actionable Tip	RFM Code
Champions	Bought recently, buy often and spend the most!	Reward them. Can be early adopters for new products. Will promote your brand.	555
Loyal Customers	Spend good money with us often.	Responsive to promotions. Upsell higher value products. Ask for reviews. Engage them.	334, 335, 345, 355, 444, 445, 454, 455, 544, 545
Potential Loyalist	Recent customers, but spent a good amount and bought more than once.	Offer membership/loyalty program, recommend other products.	431, 432, 433, 434, 435
New Customers	Bought most recently, but not often.	Provide onboarding support, give them early success, start building relationship.	413, 414, 415, 421, 422, 423, 513, 514, 515
Promising	Recent shoppers, but haven't spent much.	Create brand awareness, offer free trials	411, 412, 511, 512
Customers Needing Attention	Above average recency, frequency and monetary values. May not have bought very recently though.	Make limited time offers, Recommend based on past purchases. Reactivate them.	223, 232, 233, 311, 321, 322, 323, 332, 333
About To Sleep	Below average recency, frequency and monetary values.	Will lose them if not reactivated. Share valuable resources, recommend popular products/renewals at discount, reconnect with them.	123, 133, 134, 211, 212, 221, 222
At Risk	Spent big money and purchased often. But long time ago.	Need to bring them back! Send personalized emails to reconnect, offer renewals, provide helpful resources.	144, 244
Can't Lose Them	Made biggest purchases, and often. But haven't returned for a long time.	Win them back via renewals or newer products, don't lose them to competition, talk to them.	145, 155, 245, 255
Hibernating	Last purchase was long back, low spenders and bought seldomly.	Offer other relevant products and special discounts. Recreate brand value.	112, 121, 122, 131, 132

Customer Segment	Activity	Actionable Tip	RFM Code
Lost	Lowest recency, frequency and monetary scores.	Revive interest with reach out campaign, ignore otherwise.	111

RFM Stages

We can consider 4 stages for RFM:

- **Stage 1 ("Triers")**: Customers are one-time buyers with positive expectancy.
  - Goal: transform them into multi-buyers (stage 2)
- **Stage 2 ("Buyers")**: Customers are multi-buyers with positive expectancy.
  - Goal: retain them (keep in stage 2)
- **Stage 3 ("High Value, but falling potential")**: Customers are multi-buyers with negative expectancy.
  - Goal: reactivate them (back to stage 2), fix what is wrong
- **Stage 4 ("Low Value, and falling potential")**: Customers are buyers with negative expectancy.
  - Goal: replace them.

RFM variations

- **RFD (Recency, Frequency, Duration)** is a modified version of RFM analysis that can be used to analyze consumer behavior of viewership/readership/surfing oriented business products.
- **RFE (Recency, Frequency, Engagement)** is a broader version of the RFD analysis, where Engagement can be defined to include visit duration, pages per visit or other such metrics. It can be used to analyze consumer behavior of viewership/readership/surfing oriented business products.
- **RFC (Recency, Frequency, Cost)** is a modified version of RFM analysis that can be used to analyze customers' cost. It is useful in the context of social services.
- **RFMTC (Recency, Frequency, Monetary, time since first purchase, churn probability)** is a modified version of RFM analysis that includes time since first purchase and churn probability. This model can estimate the probability that one customer will purchase at the next time and the expected value of the total number of times that the customer will purchase in the future.
- **RFMP (Recency, Frequency, Monetary, Product)** is a modified version of RFM analysis that includes one additional parameter, period of product activity, to classify customer product loyalty under B2B concept. The findings showed that the developed methodology for CRM produces better results than other commonly used models.
- **RFM-I (Recency, Frequency, Monetary Value – Interactions)** is a version of RFM framework modified to account for recency and frequency of marketing interactions with the client (e.g. to control for possible deterring effects of very frequent advertising engagements).
- **RFM-Apriori**: is a modified version of RFM analysis known as well as RFM sequential pattern. It combines RFM with Apriori Algorithm (association analysis) for generating all RFM sequential patterns from customers' purchasing data.

References

- Creating an RFM Summary Using Excel (Peter S. Fader, Bruce G. S. Hardie)
- Shweta Singh , Sumit Singh , (2016) "Accounting for risk in the traditional RFM approach", Management Research Review, Vol. 39 Iss: 2, pp.215 - 234

- Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-value Curves for Customer Base Analysis." Abstract PDF
- Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2004), " 'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model." Abstract PDF. Associated Excel spreadsheet and note
- Tkachenko, Yegor. Autonomous CRM Control via CLV Approximation with Deep Reinforcement Learning in Discrete and Continuous Action Space. (April 8, 2015). PDF
- Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan
- Probability Models for Customer-Base Analysis