RFM Analysis in R

YI FAN

80/20 Rule

80% of your sales come from 20% of your customers

1. How to segment the customers in the database to find out the right customers who are more likely to response to mails or buy products?
2. For which type of customers, sales people should send the mails to so that the business can reach breakeven and make profit?

RFM Analysis roadmap

- What is RFM Analysis?

- How does RFM works?
  ◦ Example: CDNOW_master.txt

- The result of RFM analysis

- Issues

- Summary

What is RFM?

- R, F, and M stand for
  - ➤ Recency – How recently did the customer purchase?
  - ➤ Frequency – How often do they purchase?
  - ➤ Monetary Value – How much do they spend (each time on average)?

- Recency is the most important factor for higher response rate

- For each RFM score, there are three digital numbers. For example "532"

- Sometimes users want to determine the breaks for each dimension by themselves according to their own business requirement.
  - ➤ For example, 0-30 days, 31-90 days, 91-180 days, 181 – 360 days and more than 360 days as the 5 breaks for Recency

CDNOW_master.txt

- This dataset file contains the purchase history of 23,570 customers who made their first-ever purchase at CDNOW from first quarter of 1997 to the end of June 1998.

- There are 69,659 transaction records in this dataset, and each record comprises four field: the customer's ID, the date of the transaction, the number of CDs purchased, and the dollar value of the transaction.

```
00001 19970101   1    11.77
00002 19970112   1    12.00
00002 19970112   5    77.00
00003 19970102   2    20.76
00003 19970330   2    20.76
00003 19970402   2    19.54
00003 19971115   5    57.45
00003 19971125   4    20.96
00003 19980528   1    16.99
00004 19970101   2    29.33
00004 19970118   2    29.73
00004 19970802   1    14.96
00004 19971212   2    26.48
```

Predict & identify valuable customer → Higher response rate → Calculate breakeven to select customers

RFM Analysis Plan

1. Explore Data - construct a data frame with the necessary columns of customer ID, transaction data, and money amount paid by a customer per transaction.

2. Segment the Customers into RFM cells

3. Calculate the Recency, Frequency, and Monetary

      - Recency data – set up start date and end date, then calculate how many dates from purchase date to end date.

      - Frequency data – calculate the quantity of translations of a customer

      - Monetary data – sum the amount of money a customer spend and divide it by Frequency, to get the amount per transaction on average

4. Independent RFM Scoring – "getIndependentScore" function to implement the independent scoring. And returns a new data frame with four new columns of "R_Score", "F_Score", "M_Score", and "Total_Score"

RFM Analysis Plan (con't)

5. RFM Scoring with Breaks
  - 0-30 days, 31-90 days, 91-180 days, 181-360 days and more than 360 days as the 5 breaks for Recency.
  - 0-2 times, 3-5 times, 6-8 times, 9-11 times, and more than 11 times as 5 breaks for Frequency
  - 0-10 dollars, 11-20 dollars, 21-30 dollars, 31-100 dollars, more than 100 dollars as 5 breaks for Monetary value

6. Estimate response rate for each RFM cell – the response rate is based on historical responding data

7. Calculate the breakeven point to select valuable customers

$$P - C/R = 0$$

8. Select the Target Customers

Assumption

Response Rate Table of CDNOW

RFM cell (Total Score)	...	Response Rate
555		8.5%
...		...
441		5.6%
435		4.8%
...	...	...

Price or revenue per deal or per response is 100 dollars
The cost per mail sent out is 5 dollars

Result

```
> head(df)
   ID        Date Amount Recency Frequency Monetary
1   1  1997-01-01  11.77     546         1 11.77000
2   2  1997-01-12  12.00     535         2 44.50000
9   3  1998-05-28  16.99      34         6 26.07667
13  4  1997-12-12  26.48     201         4 25.12500
24  5  1998-01-03  37.47     179        11 35.05545
25  6  1997-01-01  20.99     546         1 20.99000
```

```
> head(df1[-(2:3)])
          ID Recency Frequency Monetary R_Score F_Score M_Score Total_Score
65954  22279     108        13 345.4338       5       5       5         555
4427    1412      68         5 323.1440       5       5       5         555
14026   4410      54         4 265.6050       5       5       5         555
15402   4881     122         5 263.9660       5       5       5         555
22526   7246      14         6 261.7783       5       5       5         555
6190    1903     124         4 223.6300       5       5       5         555
```

Result (con't)

```
> r <-c(120,240,450,500)
> f <-c(2,5,8,10)
> m <-c(10,20,30,100)
> df2<-getScoreWithBreaks(df,r,f,m)

> S500<-df2[df2$Total_Score>500,]
> dim(S500)
[1] 4250    10
> drawHistograms(df2)
> S400<-df2[df2$Total_Score>400,]
> dim(S400)
[1] 6575    10
```

P-C/R = 0

P = 100 dollars

C = 5 dollars

=> R = C/P = 5/100 = 5%

We should chose the customers in the RFM cells that have a response rate equal to or more than 5 %, to send out the direct marketing mails, to make money.

```
> target <- df2[df2$Total_Score>=441,]
> dim(target)
[1] 4427    10
```

RFM Analysis Issues

- It provides little meaningful scoring on R, F, and M when most customers do not buy often, spent little and have not purchased lately.

- RFM method also ignores the analysis on new firms setting up in a short period and customers that only purchase once and placed small orders

- Using limited number of selection variables

Summary

- What is RFM analysis?

- How does RFM analysis works?

- CDNOW_master.txt

- The result of RFM analysis for CDNOW

- Issues

References

Jo-Ting Wei, S.-Y. L.-H. (2010). A review of the application of RFM model. African Journal of Business Management, 4203-4204.

Mutyala, S. (2011, 01). Using RFM to Identify Your Best Customers. Retrieved from EightLeaves: http://www.eightleaves.com/2011/01/using-rfm-to-identify-your-best-customers

Peck, J. (2013, March 20). SPSS: Pros and Cons of Nested and Idependent Bining for RFM Analysis. Retrieved from Quora: https://www.quora.com/SPSS/SPSS-Pros-and-Cons-of-Nested-and-Independent-Binning-for-RFM-Analysis

RFM (Customer Value). (2016, November 8). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/RFM_(customer_value)

RFM Customer Analyisis with R Language. (2013, July 7). Retrieved from Data Apple: http://www.dataapple.net/?p=84

Thank You!