Using RFM Model to Construct Customer Value by Making Segment in Different Service Industries

Hui-Hsin Huang

Department of Advertising &Public Relations,
Fu Jen Catholic University,
No.510, Zhongzheng Rd., Xinzhuang Dist., New Taipei City 242062, Taiwan (R.O.C.)
E-mail : hoyasophia2020@gmail.com

Abstract: This paper uses RFM model to make customers segment by their purchase behavior. Different from previous researches, the author compares various service industries consumptions with the same customer database. It can indicate the dynamic pattern of RFM analysis and portray the diversity of customer value toward a specific customer in all his or her transaction in different situations of service industries. The characteristics of each industry are demonstrated through RFM analysis with customers' demographic data. Finally, it is shown the application for RFM model which is used in different industries in the conclusion.

Keywords: RFM; service industry; customers segment; customer value;

1 Introduction

To distinguish high contribution customer in customer segment and obtain customer value are both important when discussing the topic of customer relationship management (CRM). Since RFM framework was introduced by Alden's catalog company to decide which customer should receive a catalog (Roel, 1988). RFM models are widely used to conduct segmentation and estimate customer value.

R is the length of time since the latest purchase (or transaction), F shows the number of purchases within a time period, and M indicates the total amount of all transactions during a time period. Customers that simultaneously have high R, F, and M are considered as high-value ones (Hu et al. 2013). Thus, RFM models are considered as customer behavior dimensions, a large number of previous studies have used it to taken into account as a widely-used technique for analyzing customer value (Heldt et al., 2021) such as measuring customer lifetime value(CLV) or using it as a data mining tool to predict customer future purchase behavior.

Heldt et al.(2021) integrate the product and customer marketing perspectives propose a new approach to predict customer value based on an RFM per product model (RFM/P). To combine RFM index and product category, Heldt et al.(2021) provide a more complete overview of the future cash flow of a firm. In their model, there is no need to choose between the product and customer perspectives. The customer values are first estimated for each product (or product category) and then aggregated to obtain the overall customer value.

Anitha and Patil(2019) apply business intelligence in identifying potential customers by providing relevant and timely data to business entities in the Retail Industry. The data furnished is based on systematic study and scientific applications in analyzing sales history and purchasing behavior of the consumers. The curated and organized data as an outcome of this scientific study not only enhances business sales and profit, but also equips with intelligent insights in predicting consumer purchasing behavior and related patterns. In order to execute and apply the scientific approach using K-Means algorithm, the real time transactional and retail dataset are analyzed. Spread over a specific duration of business transactions, the dataset values and parameters provide an organized understanding of the customer buying patterns and behavior across various regions. They deploy dataset segmentation principles using K-Means Algorithm based on the RFM model. A variety of dataset clusters are validated based on the calculation of Silhouette Coefficient. The results of their research obtained with regard to sales transactions are compared with various parameters.

Rahimac et al.(2021) focuses on the extraction of the repurchase behavior of a customer through RFM features and statistical analysis of the purchase data. They use RFM features bases on repurchase behavior with relation customer to stock on the Point-of-Sale (POS) data .Because POS data is handily available to retail store owners and its data source is readily available at a business. They use customized multi-layer perceptron (MLP), decision tree classification (DTC), and support vector machine (SVM) as supervised learning methods to implement the proposed scheme on a publicly available retail data set to show the validity of the scheme. Their proposed scheme in this research identifies behavioral patterns that compose enough data to classify customers individually. Thus, the behavior pattern of their research in the form of the RFM features in combinations with statistical analysis is innovative and of minimum hardware or data collection cost. The results of their research show a high customer classification rate of more than 97% for the different numbers of the customers and eight transactions are sufficient to classify a customer with high

Abbasimehr and Shabani(2021) leverage time series clustering for detecting similar customer groups and forecasting future behavior of each group. Moreover, time series clustering facilitates the management of a large population of customers. Their proposed RFM methodology contains a time series forecasting component. And via modeling customer behaviors using the time series concept, they capture customers' dynamic behavioral patterns through integrating the concept of time series clustering and forecasting into customer behavior analysis. The representation of customer behavior as a time series additionally allows forecasting of future behavior. As well, the proposed concept is superior to previous approaches for customer behavior analysis.

2 The Model

It is denoted each R, F, M scale is in the utilityfunction $u\colon \mathrm{K} \rightarrow \mathbb{R}$ ranks in the set $\mathrm{K} = \{1, \ldots, n\}$. And the utility function $u(k)$ for each k is $u(k=1)=1$, $u(k=2)=2 \ldots u(k=n)=n$. We denote that $u(1) > u(2) \ldots > u(\mathrm{n}-1) > u(\mathrm{n})$.

For RFM modeling, we should divide customers into different segments by different threshold levels. When dividing customers into n segments in which S is the numbers of segments and is indicated $S=n$, there are n-1 threshold levels.

In the first step of R scaling, x is the value of raw data by transaction recency in different industries. The threshold level of R is considered as T_i, $i=1 \ldots n-1$ and $T_1 < T_2 < \ldots < T_{n-1}$. Then the scales of R value is

$$\mathrm{R(x)} = \begin{cases} u(1), x \leq T_1 \\ u(2), x > T_1 \\ \vdots \\ u(n), x > T_{n-1} \end{cases}$$

In the second step of F scaling, y is the value of raw data by transaction frquency in different industries. The threshold level of F is denoted as Q_i, $i=1 \ldots n-1$, and $Q_1 < Q_2 < \ldots < Q_{n-1}$.Then scales of F value is

$$\mathrm{F(y)} = \begin{cases} u(1), y > Q_{n-1} \\ u(2), y \leq Q_{n-1} \\ \vdots \\ u(\mathrm{n}), y \leq Q_1 \end{cases}.$$

In the third step of M scaling, z is the value of raw data by transaction monetary in different industries. The threshold level of M is considered as W_i, $i=1 \ldots n-1$, and $W_1 < W_2 < \ldots < W_{n-1}$. Then scales of M value is

$$\mathrm{M(z)} = \begin{cases} u(1), z > W_{n-1} \\ u(2), z \leq W_{n-1} \\ \vdots \\ u(\mathrm{n}), z \leq W_1 \end{cases}$$

3 The research method

3.1 The Empirical Database

The empirical data from the database are 5236 sample size. There are 2122(40.5%) female and 3114(59.5%) males. For the education level, there are 969(18.5%) bachelor's degree, 247 graduated degree(4.7), 111(2.1%) Ph.D. degree, 603(11.5%) college degree, 1406(26.9%) high school degree and 1900(36%)lower than high school degree. For the marriage, there are 4776 (91.2%) sample size have married.

For the purpose of this paper is to use RFM framework to portray the purchase behavior in different service industries. The transaction data in this data base includes the last purchase date, the last consumption amount, the highest amount of monetary spending, the average amount of monetary spending, and the frequency during the transactions. These transaction data includes customer making consumptions in Chain Catering, Transporter and fast fashion service industries. In the customer database, we consider the numbers of segments are S=2 and k=1, 2.

3.2 Chain Catering industries

For the chain catering industry, it is denoted that $\mathrm{R^{cc}}$, $\mathrm{F^{cc}}$ and $\mathrm{M^{cc}}$ as the three indexs of R, F and M. Then, to divide 5236 sample size into two groups(it is consider S=2) by R which is defined as the latest consumption days to this chain catering restaurant. According to the segmentation rule, the number of groups in each segment cannot be too small. And the number of people in each group cannot be too various.

Thus, it is used "Median" as the threshold level of each $\mathrm{R^{cc}}$, $\mathrm{F^{cc}}$ and $\mathrm{M^{cc}}$.

The first hierarchy is "R", which is "the duration of latest consumption days" less or more than 8 days. It can be demonstrated as

$$\mathrm{R}^{cc}(\mathrm{x}) = \begin{cases} 1, x \leq 8 \\ 2, x > 9 \end{cases}.$$

The second hierarchy is F, which is the frequency that the customer makes consumptions in this chain catering restaurant. In this part, the standard in both first and second level is less or more than 18. It can be demonstrated as

$$\mathrm{F}^{cc}(\mathrm{y}) = \begin{cases} 1, y > 19 \\ 2, y \le 18 \end{cases}.$$

The third hierarchy is "M", which is "the average of monetary spending amount in this chain catering restaurant. The threshold levels of eight groups are less or more than 37359, 25159, 16479, 10780, respectively. It can be demonstrated as

$$\mathrm{M}^{cc}(\mathrm{z}|R^{CC} = 1, F^{CC} = 1) = \begin{cases} 1, z > 37360 \\ 2, z \le 37359 \end{cases}$$

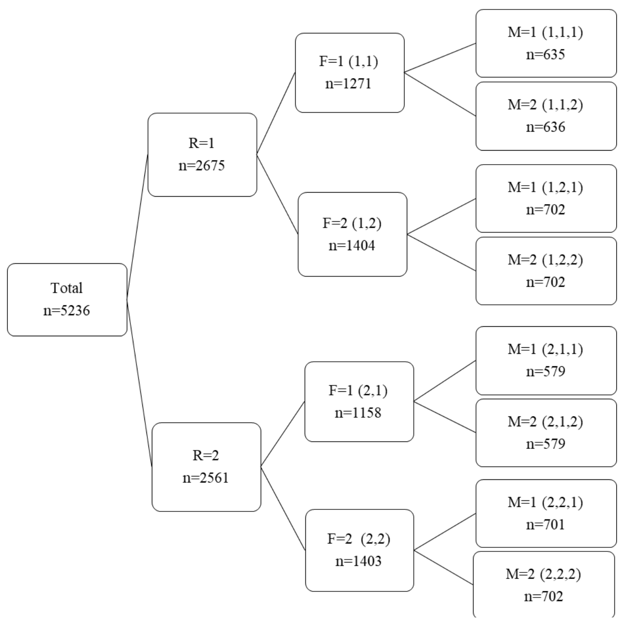
$$\mathrm{M}^{cc}(\mathrm{z}|R^{CC} = 1, F^{CC} = 2) = \begin{cases} 1, z > 25158 \\ 2, z \le 25159 \end{cases}$$

$$\mathrm{M}^{cc}(\mathrm{z}|R^{CC} = 2, F^{CC} = 1) = \begin{cases} 1, z > 16480 \\ 2, z \le 16479 \end{cases}$$

$$\mathrm{M}^{cc}(\mathrm{z}|R^{CC} = 2, F^{CC} = 2) = \begin{cases} 1, z > 10779 \\ 2, z \le 10780 \end{cases}.$$

The tree diagram is shown as figure1.

Figure 1 The tree diagram of RFM in chain catering industry



We use statistical analysis to make comparison between (1,1,1) and (2,2,2) groups to find the various of demographic variables of these low value and high value customers. For owning master's degree or above, it shows significant difference for female that there are 1.4% and 4.0% of women in group (1,1,1) and (2,2,2) respectively.

According to the data, the average frequency of men is about 14.21, and women are about 11.6. The consumption frequency of men is far from women. It means men patronize this chain catering restaurant more frequently than female. Based on the concept of consumption power, men have higher monetary spending, it means that men would spend more money on high class restaurant.

For R index, if the duration between last purchase day and the observed deadline is smaller, then it means the customers are still active and patronize this restaurant. The mean of group (1,1,1) is 7.14 days and mean of group (2,2,2) is 65.27 days. For F index which refers to the consumption frequency of this restaurant. The mean of group (1,1,1) is 24.33 and which of group (2,2,2) is roughly 4.33 times. The consumption frequency of(1,1,1) is far from (2,2,2). It means patronization of (1,1,1) is more frequently.For M index which refers to the consumption amount in the past one years, the larger amount means higher consumption power. There is a great different amount between (1,1,1) and (2,2,2), the average consumption amount of (1,1,1) is about 291475.48 N.T. dollars and (2,2,2) is about 4731.91 N.T. dollars.For the customer monetary spending amount in recent, the customers of (1,1,1) spends 28842.54 N.T. dollars and (2,2,2) spends 1302.06 N.T. dollars. And from the following comparison, we can find out that (1,1,1) not only consume in higher amount on average in the recent past, but also spending high amount in the last(recent) time. For the average number of days that customer first patronize this restaurant, (1,1,1) is about 1043.61 days, while for (2,2,2) is 2245.14 days. It means the customers that have longer relationships with the restaurant do not contribute high customer value such as high monetary spending and more frequently patronization.

For group (1,1,1), the average amount of each transaction is 17392.9128 N.T. dollars. As for group (2,2,2), the average amount of each transaction is 1343.6333 N.T. dollars. The average amount of each transaction for group (1,1,1) is much higher than group (2,2,2), so (1,1,1) has amazing purchasing power, they are high-spending group in patronization.The maximum spending amount in past180 days for group (1,1,1) is 61894.23 N.T. dollars on average, group (2,2,2) is 25108.66 N.T. dollars on average, so the maximum spending amount in past 180 days for two groups has huge difference. In other word, the limit of consumption for group (1,1,1) is much higher than group (2,2,2).

Combining the information above, the average consumption power of (1,1,1) is greater than that of the (2,2,2). For the education level of female, group (1,1,1) has higher percentage of master's degree or above, the higher education female may have better consumption power .

3.3 Transporter industries

For the transportation industry, this database includes the behavior of a certain brand of taxis. It is denoted that R^T, F^T and M^T as the three indexs of R, F and M. Then, it divides sample size into two groups(it is consider *S*=2) by R which is defined as the days since last boarding. According to this industry,the first hierarchy "R" is 14 days. It can be demonstrated as

$$\mathrm{R}^{T}(\mathrm{x}) = \begin{cases} 1, x \le 14 \\ 2, x > 15 \end{cases}.$$

Frequency is number of rides in the past month. It is used 8 times as standard. It can be demonstrated as

$$\mathrm{F}^{T}(\mathrm{y}) = \begin{cases} 1, y > 9 \\ 2, y \le 8 \end{cases}.$$

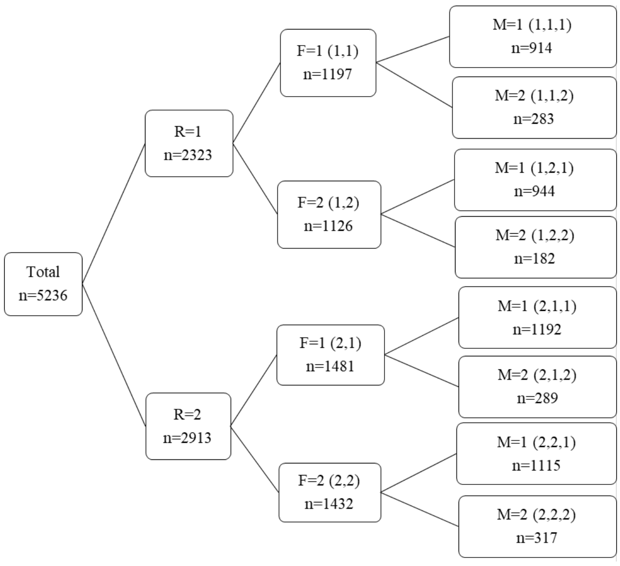
Monetary is the total amount of consumption in the past month. It is used 2700 N.T.dollars as standard. It can be demonstrated as

$$\mathrm{M}^{T}(\mathrm{z}) = \begin{cases} 1, z > 2701 \\ 2, z \leq 2700 \end{cases}.$$

In this industry, we don't use different threshold levels of each hierarchy R, F and M. We use the same threshold levels in each hierarchy.

The tree diagram is shown as figure2.

Figure 2 The tree diagram of RFM in transportation industry



We also compare the highest and lowest customer values between (1,1,1) and (2,2,2) segment. To compare the total percentage of education level higher than bachelor's degree (including bachelor's degree) for each group, the total percentage of education level higher than bachelor's degree is 28.5% in (1,1,1) group. Group (2,2,2) have 20.8% total percentage of education level higher than bachelor's degree.

The difference of percentage might explain it is possible that education level could influence the result of customer base. Because the education level is related to the consumption ability.

3.4 The fashion industries

For the fashion industry, this database includes the purchase behavior of a certain brand of fast fashion chain store. It is denoted that R^{F}, F^{F} and M^{F} as the three indexs of R, F and M. It is also segment total data into two groups(it is consider S=2) by R which is defined as the days since last transaction. According to this industry,the first hierarchy "R" is 18 days. It can be demonstrated as

$$\mathrm{R}^{F}(\mathrm{x}) = \begin{cases} 1, x \leq 18 \\ 2, x > 19 \end{cases}.$$

Frequency is number of transactions in the past month. It is used 4 times as standard. It can be demonstrated as

$$\mathrm{F}^{F}(\mathrm{y}) = \begin{cases} 1, y > 5 \\ 2, y \leq 4 \end{cases}.$$

Monetary is the total amount of transactions in the past month. The threshold levels of eight groups are less or more than 3537, 2498, 1717, 1127 respectively. It can be demonstrated as

$$\mathrm{M}^{F}(\mathrm{z}|R^{F} = 1, F^{F} = 1) = \begin{cases} 1, z > 3538 \\ 2, z \leq 3537 \end{cases}$$

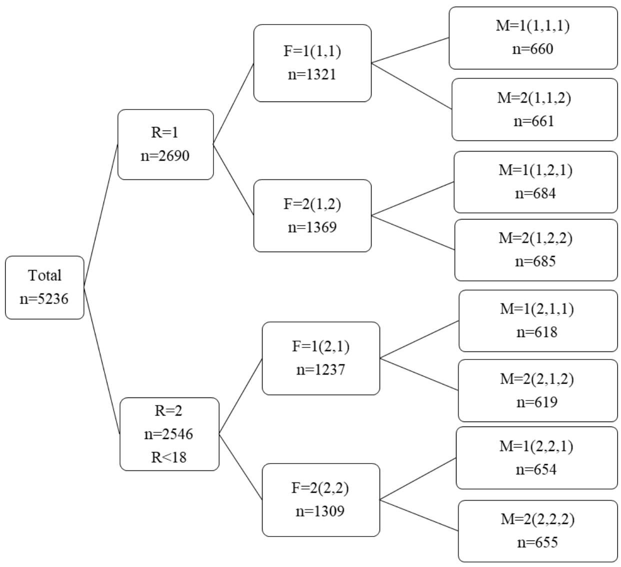
$$\mathrm{M}^{F}(\mathrm{z}|R^{F} = 1, F^{F} = 2) = \begin{cases} 1, z > 2499 \\ 2, z \leq 2498 \end{cases}$$

$$\mathrm{M}^{F}(\mathrm{z}|R^{F} = 2, F^{F} = 1) = \begin{cases} 1, z > 1128 \\ 2, z \leq 1127 \end{cases}$$

$$\mathrm{M}^{cc}(\mathrm{z}|R^{CC} = 2, F^{CC} = 2) = \begin{cases} 1, z > 1077 \\ 2, z \leq 1078 \end{cases}$$

The figure of RFM segment in fast fashion chain store is demonstrated as figure 3.

Figure 3 The tree diagram of RFM in transportation industry



In the fast fashion chain store data, we compare the group of (1,1,1) , (2,2,1) and (2,2,2) for the average consumption amount per transaction , the maximum consumption amount of transaction, the total consumption amount in the past month and the transaction frequency.

For comparing (2,2,2) and (2,2,1) , the mean of average consumption amount per transaction in the past month of (2,2,2) is 1445.48 N.T. dollars which is 1.5 times higher than 99.46 N.T. dollars in group (2,2,1). For the mean of transaction frequency, group (2,2,2) is 4.07 times and group (2,2,1) is 21.12 times. It can be seen that the customers in (2,2,2) who have a high single consumption amount but low consumption frequency and the customers in (2,2,1) who have higher transaction frequency but low single consumption amount.

For comparing (1,1,1) and (2,2,1) ,the mean of maximum consumption amount of group (1,1,1) is 6098.3 N.T. dollars and which of (2,2,1) is 5605.2 N.T. dollars. There is little difference between these two groups. It is also little different for the mean of transaction frequency in group (1,1,1) and (2,2,1). Group (1,1,1) is

23.16 times and (2,2,1) is 21.12 times. But the total consumption amount and the mean of average consumption amount per transaction, (1,1,1) are both about twice as much as (2,2,1). The men of total consumption amount of (1,1,1) is 18035 N.T. dollars and which of (2,2,1) is 9384 N.T. dollars. The mean of average consumption amount per transaction of (1,1,1) is 5823 N.T. dollars and which of (2,2,1) is 944.61 N.T. dollars.

To sum up, it can be seen that the maximum spending power and consumption frequency of the two groups of customers (2,2,1) and (1,1,1) are very close, while the amount of each consumption of (1,1,1) is significantly higher.

4 Conclusion

This reteach uses the RFM framework to segment customers in different industries. It shows various appearance of purchase behavior in different industries among the same customer data. To focus on the higher contribution group which is defined as most active, most frequently consumption and spending highest monetary amount for RFM segment, female have higher education in chain catering restaurant and both male and female have higher percentage of education level higher than bachelor's degree in taxi such as uber riding. It also demonstrate higher consumption in fast fasion chain store no matter in average consumption amount per transaction , the maximum consumption amount of transaction, the total consumption amount in the past month or the transaction frequency. It is interesting that the highest contribution customers have shorter time relations with this chain catering restaurant than the lowest contribution group in chain catering industry. It can be indicated that the loyal customers do not necessarily have a high amount of contribution for the company.

In the future, it can use other RFM framework to make customer segmentation. For example, conduct stochastic model to consider Possion or exponential probability density to portray RFM indexes respectively. The research can also try to conduct other method such as investigating customer psychological variables and exploring the relationships of these variables with the evaluation outcome of customers such as satisfaction and loyalty to make customer segmentation.

Acknowledgements

The author would like to thank the National Science Council of Taiwan (MOST 105-2410-H-156-013) and Fu Jen University supporting this research.

References

[1] R. Roel, Direct Marketing's 50 Big Ideas, Direct Marketing, 50 (May), 45-52, 1988.

[2] Y. H. Hu, TC. K. Huang, Y. H. Kao, Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. Journal of System Software, 86, 779-788, 2013.

[3] R. Heldt, C. S. Schmitt, F. B. Luce, Predicting customer value per product: From RFM to RFM/P, Journal of Business Research, 127, 444-453, 2021.

[4]P. Anitha, M. M. Patil, RFM model for customer purchase behavior using K-Means algorithm, Journal of King Saud University - Computer and Information Sciences, 25, 1-8, 2019.

[5]M.A. Rahimac, M. Mushafiq, S. Tkhan, A. Z. Arain, RFM-based repurchase behavior for customer classification and segmentation. Journal of Retailing and Consumer Services, 61, 1-9, 2021.

[6]H. Abbasimehr, M. Shabani, A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques. Journal of Ambient Intelligence and Humanized Computing ,12, 515-531, 2021.

Biographical notes:

Dr. Hui-Hsin Huang received the Ph.D. in management sciences from Tamkang University, New Taipei City in 2008. Now, she is an associate professor with Department of Advertising &Public Relations, Fu Jen Catholic University, New Taipei City, Taiwan. She has published papers on International Journal of Information and Management Sciences (IJIMS), Information and Knowledge Management, and Advanced Materials Research. Her research interests include marketing model, consumer behavior and applied statistic.