Rental Customer Segmentation Based on Length, Recency, Frequency, Average-Monetary and Satisfaction Value Model and Cluster Analysis

Fu Tao[1,*] Xindi Wang[1]

[1]School of Economics and Management, Beijing Jiaotong University, Beijing, China
*Corresponding author. Email: tf751267941@163.com

ABSTRACT

As the rental market continues to grow, it is increasingly important to subdivide rental customers and formulate targeted marketing management strategies. Aiming at the characteristics of large number of renting customers, long period, and large potential value, a length, recency, frequency, average-monetary and satisfaction value (LRFAS) model of customer segmentation based on recency, frequency and monetary value improvement is proposed. On this basis, the K-means clustering algorithm is used to subdivide customers, and the entropy weight method is used to determine the weight of each indicator, and the obtained indicator weight is used to calculate the value of the customer. An example of an intermediary company was used to verify the feasibility and effectiveness of the improved model in the field of housing leasing. The results show that the improved length, recency, frequency, average-monetary and satisfaction value model can more effectively and accurately segment rental customers, and at the same time formulate corresponding marketing strategies for different types of customer needs, helping intermediary companies to gain great core competitiveness in the market.

Keywords: renting customers, customer segmentation, LRFAS model, K-means clustering

I. INTRODUCTION

With the increase in home ownership costs and government policy support, the rental industry has entered a new era of rapid development. Data show that the rental population in China's service rental market has maintained steady growth since 2017, with 210 million in 2018, 220 million in 2019, 220 million in 2020, and 240 million in 2022. As urbanization continues to advance, the continued increase in the size of the floating population will provide the basis for the rental population. It is estimated that by 2025, the scale of China's rental market will grow to 2.9 trillion, and by 2030 the domestic rental population will reach 270 million, and the market size will exceed 4.6 trillion [1]. With the increasingly fierce competition in the rental market and the increasing number of rental users, the rental market field needs to segment customers like traditional marketing to meet the growing personalized needs of customers, and through differentiated services, targeting different User needs formulate corresponding marketing strategies, form long-term buying behavior, improve customer loyalty, and enable leasing companies to gain core competitiveness in the fierce market competition. Chen Shuqian et al.[2] applied the traditional RFM model to the logistics field, and used the k-means algorithm to subdivide logistics customers

and calculate customer value. Bao Zhiqiang et al.[3] applied the RFA model improved based on the RFM model to the Baidu takeaway field to subdivide customers; Ren Chunhua et al.[4] proposed the LRFM model of automobile loyalty customers and adopted a layer of K nearest neighbor The density peak algorithm finds the initial clustering center to subdivide auto loyalty customers. Xu Xiangbin et al.[5] introduced the total profit attribute on the basis of the traditional RFM model, and proposed that the RFP model be applied to the e-commerce field to subdivide it.

This paper proposes an improved RFM model based on the combination of customer consumption behavior and customer satisfaction with the leasing company, aiming at the characteristics of large number of rental customers, long cycle and large potential value. At the same time, the entropy weight method is used to determine the variables in the model Weight, and on this basis, K-means clustering algorithm is used to segment customers, calculate and determine the customer's personal value, and formulate corresponding marketing strategies for different customer needs.

II. CUSTOMER SEGMENTATION MODEL

A. Traditional RFM model

As a means of customer segmentation, the RFM model was first used in the field of direct sales, with the purpose of increasing the number of customer purchases. With the continuous improvement and innovation of the RFM model, the current RFM model has become an important tool and means for measuring customer value and the ability of customers to create profits for enterprises. The RFM model measures the value of a customer from the perspective of a customer's consumption behavior through three indicators: a customer's recent purchase behavior (Recency), the overall frequency of purchase (Frequency), and the purchase amount (Monetary). Where R (Recency) represents the time since the customer's last purchase from the statistical deadline. If the value of R is smaller, the customer has consumed the company's products or services in a short time, which means that the customer has kept Customer's concern, customers are more likely to buy in the short term. F (Frequency) represents the frequency with which customers purchase enterprise products or services during the statistical period. Generally speaking, the greater the F value of a customer, the higher the frequency with which customers purchase enterprise products or services, and the higher the customer loyalty. The greater the value. M (Monetary) represents the total amount of customer consumption in the statistical period. Generally speaking, the larger the M value of the customer, the higher the customer loyalty, and the higher the profit the customer brings to the enterprise. Through these three indicators, enterprises comprehensively evaluate customer value, and identify customers with short consumption cycles, frequent consumption, and high consumption amounts. Furthermore, for different types of customers, according to their circumstances and different needs to develop corresponding marketing strategies to help companies get the highest return with the lowest investment.

B. LRFAS model

The traditional RFM model has the following problems in customer segmentation:

- F and M have the defect of multiple linear relationships;

- The RFM model is a segmentation index determined from the perspective of customer consumption behavior. Limitations of satisfaction.

Therefore, in view of the characteristics of the large number of customers in the rental market, long cycle, and large potential value, an LRFAS model for customer segmentation based on the improved RFM model for the housing rental market is proposed. Among them, the meaning of each indicator in the LRFAS model is shown in "Table I". In the LRFAS model, the L index is added to reflect the loyalty of renters to the enterprise, which solves the problem of long consumption cycles for renters. Replacing M index with A index can eliminate the collinearity problem between frequency F and value M in the RFM model. Increasing the S index will solve the limitation of the traditional RFM model that lacks user satisfaction evaluation of the enterprise.

TABLE I.   LRFAS MODEL EACH INDEX MEANING TABLE

Index	Abbreviation	Unit	Meaning
Length	L	month	The time interval from the first rent to the most recent rent of a customer to reflect the customer's loyalty to the company
Recency	R	month	The time interval from the customer's most recent rent to the current time
Frequency	F	Times	Represents the number of rentals within a certain time range
Average Monetary	A	yuan	Represents the average rent amount within a certain period of time
Satisfaction	S	/	Renters' satisfaction with corporate leasing services

C. Weight analysis of LRFAS model

Regarding the weight of each indicator in the RFM model, Hughes et al. [6] believed that when using the RFM model, each indicator should be given the same weight. Miglautsch [7] believes that the indicators in the RFM model should have different weights for different problems. Entropy weight method is a method widely used in objective weighting method. The weight of each indicator depends on the entropy value of the indicator, and the size of the entropy value depends on the stability of each indicator. Entropy can measure the stability of an indicator. If the uncertainty of an indicator is greater, the entropy is greater. Conversely, the smaller the uncertainty of an indicator, the smaller its entropy.The weight is determined according to the amount of information provided by the entropy value, in order to transfer sufficient and effective information,

to avoid the subjectivity of traditional weighting methods, such as analytic hierarchy process and expert survey method, and to better find the information characteristics of the evaluation indicator data Enhance the credibility of weights[8]. Therefore, this paper uses the entropy weight method to analyze the weight of the LRFAS model. The steps are as follows:

1) Data standardization: Because the measurement units of each variable index are different, it will affect the evaluation results, so it is necessary to standardize before using them to calculate the comprehensive score, that is, convert the absolute value of the variable index into a relative value to achieve homogeneity of heterogeneous indicators Dimensionless processing. However, each variable index value represents a different meaning. Some index values are as large as possible, such as F, A, and S, which we call positive indicators. Some index values are as small as possible, such as L and R, which we call negative indicators. Therefore, different algorithms are used to standardize the positive and negative indicators.

Positive indicator:

$$x'_{ij} = \frac{x_{ij} - \min\{x_{1j},...,x_{nj}\}}{\max\{x_{1j},...,x_{nj}\} - \min\{x_{1j},...,x_{nj}\}}$$

Negative indicator:

$$x'_{ij} = \frac{\max\{x_{ij},...,x_{nj}\} - x_{ij}}{\max\{x_{1j},...,x_{nj}\} - \min\{x_{1j},...,x_{nj}\}}$$

Where is the standardized value of the j-th indicator of the i-th customer (i = 1,2..., n; j = 1,2,... m)

2) Calculating the specific gravity matrix: Calculate the proportion of the i-th customer under the j-th indicator to the indicator based on the standardized results:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \quad i=1,..,n, j=1,...,m$$

Where n is the total number of customers and m is the number of variables

3) Calculating the information entropy of each index: According to the result p_{ij} obtained by the proportion matrix, the information entropy value of the j-th index can be calculated:

$$E_j = -k\sum_{i=1}^{n} p_{ij} \ln(p_{ij})$$

Where $k = 1/\ln(n)$, $E_j \geq 0$

4) Calculating information entropy redundancy: The information entropy redundancy is also the information utility value, which is negatively correlated with the information entropy. The greater the information entropy value, the smaller the information entropy redundancy, the lower the disorder of information, the greater the weight of the index [9]. The calculation formula is:

$$\delta_j = 1 - E_j$$

5) Calculating the index weight: The index weight is calculated according to the information entropy redundancy, and its calculation formula is:

$$w_j = \frac{\delta_j}{\sum_{j=1}^{m} \delta_j}$$

6) Calculating the value score of renting users:

$$s_j = \sum_{j=1}^{m} w_j * p_{ij}$$

III. EXAMPLE ANALYSIS

A. Statistics

By studying the R value of rented customers and whether customers will trade again in the future, it is found that when the R value of the customer is greater than 6, the possibility of using the rental service in the future period is very small, this part of the customer can be called Lost customers. Even if the enterprise spends a lot of costs on such users, the benefits are very small, so the enterprise does not have to pay too much attention to such users. Therefore, this article mainly selects customers who have had a transaction with the company within a half year (January 1, 2019 to July 1, 2019) of a housing agency company in a city as the research object.

- L: Months from the first rent to the most recent rent

- R: The number of months of the last transaction time of the rented customer from the statistical deadline

- F: The total number of transactions of F rental customers in half a year

- A: The average price of the order of the tenant's transaction within six months

- S: Asked the customer to fill out the "Customer's Satisfaction Survey Form for Intermediary Company L's Rental Behavior", and counted the average rating of rental customers who had transactions within half a year on the rental

service, with a single rating range of 1-5 points 280 copies, effective recovery of 226 copies)

According to the statistical calibers of the above indicators, data statistics are carried out. The statistics of some customers are shown in "Table II" below:

TABLE II. SOME CUSTOMER STATISTICS

Customer ID	L	R	F	A	S
10225364	5	3	1	6600	4
10234859	6	5	2	12000	3.5
...	...	...	...	...	...
20423456	3	1	2	14400	4.5

B. Data pre-processing

Before using the data for empirical research, the data must be sorted out to obtain clean data, which is conducive to our empirical research. For outlier data with incomplete customer information, failure to obtain customer satisfaction scores for orders, and outlier data with a customer's average transaction amount being negative, the customer's transaction frequency being too high within six months, and the customer's transaction date being clearly wrong Use the method of eliminating data. At the same time, the min-max normalization [10] method is used to eliminate the influence caused by the dimension and the range of values between the indicators. By mapping the data between [0,1] to ensure that the data is in the same dimension or the same data level for comparison, the conversion formula is as follows:

$$x^* = \frac{x - \min}{\max - \min}$$

Where max and min are the maximum and minimum values of the sample data. Part of the data after the min-max normalization is shown in "Table III" below:

TABLE III. THE NORMALIZED PART OF THE DATA

Customer ID	L	R	F	A	S
10225364	0.364	0.4	0	0.208	0.5
10234859	0.455	0.8	0.5	0.583	0.25

C. Result analysis and marketing strategy

Using the entropy weight method for quantitative analysis, the weights of the five LRFAS indicators are WL = 0.124, WR = 0.028, WF = 0.112, WA = 0.426, WS = 0.256 and use the formula to obtain the value score of each type of customer. The K-Means clustering algorithm is used to perform cluster analysis on the four indicators L, R, F, and A of the renting customer. By clustering the L, R, F, and A indicators multiple times to specify the number of clusters, Significance test Sig value is less than 0.01 and F value test is used. Finally, the number of clusters that are most suitable for L, R, F, and A index clustering are all 4, which is K = 4. Through the LRFAS model clustering, renters are classified into four types: high-value customers, developmental customers, medium-value customers, and low-value customers.

After consulting the opinions of senior practitioners in the field of house leasing, this article believes that when the average rating of a renting customer's transaction order is not less than 4 points, the customer is considered satisfied with the rental service provided by the enterprise. The LRFAS evaluation model can be obtained based on the classification of the LRFA model and the customer's satisfaction with the enterprise's rental service. The customer segmentation results based on the LRFAS model are shown in "Table IV" below:

TABLE IV. CUSTOMER SEGMENTATION RESULTS BASED ON LRFAS MODEL

LRFA layering	Satisfaction stratification	Number of people	Customer classification	value of customer
High-value customers Developmental customers	Satisfaction	48	VIP customers	0.47963
High-value customers Developmental customers	Dissatisfied	62	Diamond customers	0.32192
Mid-value customers Low-value customers	Satisfaction	72	Enriching customers	0.17893
Mid-value customers Low-value customers	Dissatisfied	38	Retaining customers	0.09254

The It can be seen from the above table that there are 48 VIP customers, accounting for 21.8% of all respondents. This type of customer has the most transactions in a certain period, the shortest transaction

interval, the largest average transaction amount, and the highest customer value. And the average rating of such users on transaction orders is 4 points or more, and they are very satisfied with the company's leasing services. Therefore, they believe that such customers are loyal customers. You should take the initiative to keep in touch with such customers in a timely manner, give certain preferential activities appropriately, and adopt marketing strategies that recommend high-end housing rental and long-term rental to them.

There are 62 diamond customers, accounting for 28.2% of all respondents. This type of customer has a large number of transactions in a certain period, a short transaction interval, and a large average transaction amount. This type of customer also has great customer value, but the score of the transaction order is below 4 points. The rental service is not very satisfactory. For such customers, they should actively communicate with them, understand their dissatisfaction in the order transaction process, and resolve and correct them in a timely manner.

There are a total of 72 gaining customers, accounting for 32.7% of all respondents. Such customers have fewer transactions within a certain period, shorter transaction intervals, and fewer average transactions. Although the value of such customers is relatively small, they are satisfied with the company's leasing service and can still be regarded as the The company's relatively loyal customers. For such customers, on the basis of the original service quality, more promotional activities should be added, and a marketing strategy that turns long-term rental into short-term rental can be adopted.

There are 92 retention customers, accounting for 17.3% of all respondents. This type of customer has the least number of transactions within a certain period, the shortest transaction interval, the smallest average transaction amount, and the lowest customer value. At the same time, he is not very satisfied with the company's rental service, and there is a high probability of such users losing. For such users, they can adopt the marketing strategy of providing low-end housing rental or daily rental housing to them. At the same time, they should improve the quality of service during the rental process, and strive to keep this part of customers to avoid loss.

IV. CONCLUSION

This paper combines the characteristics of long transaction cycle, large number of users and large potential value in the housing rental market, and proposes a LRFAS customer segmentation model that is more suitable for the field of housing rental. This model not only can more accurately subdivide rental customers, but also overcome the limitation of the traditional RFM model that lacks customer feedback on

enterprise satisfaction. Taking A house intermediary company as an example, its customers are subdivided into 4 categories, and the weight of each indicator is determined using the entropy weight method to calculate the value of each category of customers. For different types of customers, formulate corresponding marketing strategies according to their situations and needs, and carry out precise marketing for each type of customer to help the housing intermediary company improve its core competitiveness in the fiercely competitive housing rental market.

References

[1] General Office of the State Council. Several opinions of the General Office of the State Council on accelerating the cultivation and development of the housing rental market: Guobanfa [2016] No. 39 [A / OL].

[2] Chen Qianshu, Fang Xiaoping. Research on logistics customer value based on RFM model [J]. Logistics Technology, 2019,42 (07): 19-22.

[3] Bao Zhiqiang, Zhao Yuanyuan, Zhao Yan, Hu Xiaotian, Huang Qiongdan. Baidu takeaway customer value analysis based on improved RFM model [J]. Journal of Xi'an University of Posts and Telecommunications, 2019,24 (01): 105-110.

[4] Ren Chunhua, Sun Linfu, Wu Qishi.Automotive loyal customer segmentation method based on LRFAT model and improved K-means [J]. Computer Integrated Manufacturing System, 2019,25 (12): 3267-3278.

[5] Xu Xiangbin, Wang Jiaqiang, Tu Huan, Mu Ming. E-commerce customer segmentation based on improved RFM model [J]. Computer Applications, 2012,32 (05): 1439-1442.

[6] WEI J T, LIN S Y,WU H H. A review of the application of RFM model[J]. African Journal of Business Management,2010,4(19):4199-4206.

[7] MIGLAUTSCH J R. Thoughts on RFM scoring[J]. Journal of Database Marketing,2000,8:67-72

[8] Li Jingkang, Wang Xindi, Guo Xiaochun.Prediction of personnel loss risk of training institutions based on matter-element extension method [J]. Hebei Industrial Science and Technology, 2020,37 (01): 5-10.

[9] Zhao Hanxiao. Research on customer segmentation based on Wi-Fi detection of passenger flow data [D]. University of International Business and Economics, 2019.

[10] Zhang Liangjun, Yun Weibiao, Wang Lu, etc. R language data analysis and mining combat [M]. Beijing: Mechanical Industry Press, 2015: 46-47.