

Article

Applied Big Data Analysis to Build Customer Product Recommendation Model

Rong-Ho. Lin ¹, Wei-Wei Chuang ^{1,*}, Chun-Ling Chuang ² and Wan-Sin Chang ¹

¹ Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei 106, Taiwan; rhlin@ntut.edu.tw (R.-H.L.); lemontree970@gmail.com (W.-S.C.)

² Department of Information Management, Kainan University, Taoyuan 338, Taiwan; clchuang@mail.knu.edu.tw

* Correspondence: chuanglucia@gmail.com; Tel.: +886-2-2771-2171 (ext. 3414)

Abstract: With the development of the Internet environment, the trend of the retail industry in the future. It cannot be separated from the community, data and experience. Consumers' lifestyles and purchasing behaviors are constantly changing and retailers must adopt policies to understand consumers. This research analyzes supermarkets most commonly touched by consumers in daily life. In order to find hidden information behind customer transaction data, it helps supermarkets to learn about the habits of customers to help them Formulate marketing strategies and improve the profitability of supermarkets and maintain long-term relationships with customers. Thus, the RFM model is used to convert customer transaction data into R, F, and M values and then clustering using the Ward's method to combine with K-means, fuzzy C-means, and self-organizing maps. Using discriminant analysis find out the grouping method with the highest accuracy rate to calculate the customer lifetime value score. In terms of product recommendation, customers can be recommended to buy products in the top five categories or to use rules found in association rule to make recommendations. In terms of customers, we maintain long-term relationships with customers by recommending other related products, products for bundling sale, giving gifts or discount coupons, and regularly organizing promotional activities.



Citation: Lin, R.-H.; Chuang, W.-W.; Chuang, C.-L.; Chang, W.-S. Applied Big Data Analysis to Build Customer Product Recommendation Model. *Sustainability* **2021**, *13*, 4985. <https://doi.org/10.3390/su13094985>

Received: 22 March 2021
Accepted: 24 April 2021
Published: 29 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: supermarket; RFM model; ward's method; K-means; fuzzy C-means; self-organizing maps; discriminant analysis; customer lifetime value score; association rules

1. Introduction

Scoring customers based on their total spend in the current year has formed the basis for businesses to select their target customers for the following year. This method primarily evaluates customers based on their total spend on purchases. Businesses often believe that 20% of customers contribute 80% of total revenues and the relationship between resource input and customer contribution is not equivalent. Italian economist Pareto [1] stated in 1897 that a nonequivalent relationship exists between causes and results as well as between efforts and effects.

With the recent growth of the network environment and advent of information explosion, people have become inextricably linked to big data. The Global Powers of Retailing 2017 released by Deloitte & Touche indicates that future retailing trends can be encapsulated into three major paradigms: social networks, data, and experiences. Social Networks: With the advent of the social networking age, the starting point for consumer management should be social networks. Chatbots are expected to become highly popular in e-commerce. Consequently, big data will become the key decision-making point for retailers. Cyber Physical Systems (CPS) will become mainstream; brick-and-mortar stores face major changes; and data play an integral role in making the narrative facts specific, clear, and more convincing. Emotion-oriented consumer experiences, interactive experiential consumption, and games are expected to win over the hearts and minds of consumers. The new retail

era has led to constantly changing consumer lifestyles and purchasing behavior. Retailers must adopt corresponding policies to “attract customers” before consumption, “retain customers” during consumption, and “provide after-sales service” after consumption. Consumers’ shopping dynamics must be ascertained to grasp the proclivities of the customer base, thus illustrating the importance of big data. This is because knowing how to interpret data can facilitate better understanding of the customer base.

Consumers are most exposed to retailing in their daily lives. Retailing can be divided into convenience stores, supermarkets, mass merchandisers, and department stores in terms of sales methods, commodity pricing, commodity types, and modes of operation. This research explores and analyzes supermarkets that sell “mid-price commodities” and “target at residential communities and community households”, which are closer to the lives of the general public.

Table 1 shows that the annual growth rate is positive, indicating that the annual growth of the turnover of supermarkets has been positive, with the turnover increasing year to year. The annual growth rate refers to the difference compared with those in the previous year.

Table 1. Supermarket turnover and annual growth rate ¹.

Year	Turnover (Billions)	Annual Growth Rate (%)
2008	121.3	9.33
2009	126.8	4.55
2010	133.6	5.32
2011	143.4	7.35
2012	151.9	5.92
2013	158.7	4.52
2014	167.2	5.34
2016	180.4	7.89
2017	197.3	9.35
2018	209.6	6.24

¹ Data Source: Department of Statistics, Ministry of Economic Affairs of Taiwan (2018). Organized by author.

The first step for a quality company to develop over the long term is digitization, while maintaining good customer relations is the primary focus. Safeway, the third largest supermarket chain in the United Kingdom, records the types of goods purchased by customers using store credit with every purchase. Sales can be increased by identifying the correlations between different product purchases from customers’ consumption data and placing two related products in close proximity. This research holds that the hidden information behind customer transaction data is highly important to understand customers’ buying habits and increase sales volume by placing two highly correlated products together.

1.1. Statements of the Problem

The prevalence of the Internet makes product costs visible to the world. In an era of meagre profit, business management is in an increasingly competitive environment. After a period of operation, a consistent decision is a must in business management. It is of great importance to successful businesses about selecting which customers as primary marketing targets, how to verify the marketing resource is distributed to the selected customers equally, that is how to distribute the marketing resources to various types of customers (proportions and methods are not necessarily the same every year and would have to be modified based on the social boom and the political-economic atmosphere of the year), and how to screen customers and foster them to enhance the corporates future profitability.

Many businesses have been using the customer spending score (CSS) to evaluate their customers. CSS is defined as the anticipated revenue from customers’ consumption, with which potential target customers for the next year are evaluated and differentiated based on one or two deciles ranging from the top to the bottom. Since CSS mainly relies on customers’ consumption to differentiate different customers, without taking into considerations other

influential factors of customer service. Therefore, many businesses hope for a better alternative to replace the oversimplified CSS evaluation. It is the same with the retail business. When customers shop in a store, they are encouraged to apply for membership with all kinds of rewards, such as a one-hundred-dollar coupon for future consumption with the completion of membership application. For this, the researchers of the current study have designed a membership application form to collect customers' data, used as a basic marketing differentiation. Customers are divided into three tiers (i.e., A, B, and C) according to their annual expenditure, and are further provided with tier-specific marketing activities based on their preferences. However, problems come with the fact that some loyal frequent buyers in the stores with little annual expenditure might be ignored.

The Pareto principle (also known as the 80-20 rule) is widely quoted in many businesses to account for the unbalanced relation between resource investment and customer contribution. Discovered by Italian economist Vilfredo Pareto in 1897, the 80-20 rule suggests that there exists an unbalanced relation between causes and effects, or efforts and gains. Typical examples would be "20% of the gains come from 80% of the efforts, and 80% of the effects are made by 20% of the causes". In other words, 80% of a person's efforts only result in 20% of the gains. It is a rule of thumb in business that 80% of a company's profits come from 20% of its valuable clients.

The 80-20 rule has been proved worldwide. For example, ARC Management Consulting Ltd. (London, UK) analyzed the customer market with Activity Based Costing (ABC), finding that 80% of the overall interests come from top 20% of the 400 companies (2001). Ref. [1] further extended the rule into the 80/20/30 rule, in which the number "30" represents that most of the profit is wasted supporting the 30% of the customers who are not generating any revenue for the business. Therefore, the top 20% should get significant focus and receive most of the resources as they are bringing in the bulk of your revenue (80%). Knowing the 80% that make efforts out of proportion to gains enables the businesses to either suit methods to specific situations or simply delete them, which in turn reduces cost. This implies that giving up on the customers who do not make much contribution enhances a company's profits.

1.2. Scope and Review

Whether it is the 80/20 rule or the 80/20/30 rule, both rules imply the centralization and key management of resources. In other words, 80% of a company's resources should be distributed to the most valuable 20% customers, and the rest 20% to less valuable customers. With this strategy, customers who make much contribution to the profits would become loyal customers because of special preference or delicate services. On the other hand, customers who do not make much contribution to the profits would increase consumption and become those who actually make much contribution in the hope of receiving advanced preferences or services. They might also gradually move to other competitors because of not being able to receive advanced preference or service in the original company. This kind of transfer is beneficial to a business for that it not only lowers marketing expenditure to those less-valuable customers but, also, increases competing companies' cost. To fulfill this kind of customer differentiation and management, a successful database analysis is a must. First, administrators should develop a business culture with continuous creativity and R&D and support the required resources into the company-customer interactive network. Second, customers should be differentiated into different tiers with the use of database analysis. The last part lies in the most important assets of long-term business operation; that is, to keep high customer satisfaction and to maintain customer lifetime value. Specialized membership databases and multi-functional marketing activities that knitting consumption and after-sales services together would boost lasting and stable profits. For example, Capital One used database and analytical models to locate consumers who spend, borrow, and return money. This type of consumers receives more marketing activities and resources. On the contrary, for other consumers, their loan interests were raised, and they were forced to move to other credit card companies. This strategy made

Capital One a professional credit card company with the highest growth rate. Another example is an index research of customer lifetime value (CLV), customers' profitability, and distribution of marketing resources conducted by International Business Machines Corporation (IBM) in 2008. The study tested one simple management principle: Could low-valuable clients create high-value profits with enhanced accuracy of interacting with clients? The aforementioned examples have been proven and have helped businesses save resources.

1.3. Objective

In light of the previous statements, pinpointing those key customers who make contribution to a company's profits is vital to business management. It helps businesses target on valuable customers and establishes long-term positive relationships with those customers. In addition, it enables businesses to effectively distribute limited marketing resources.

1.4. Organization of the Paper

In such a competitive management environment, important issues in business management have been centering on how to examine the cost, enhance profits, and identify customer values to boost profits, effectively distribute business resources, and improve performances. Sleeping customers were deleted to ease the heavy load of data analysis. Then the integration of RFM model and customer lifetime value model allowed for a thorough examination of customer lifetime value. In addition, the use of clustering technique to differentiate markets aimed to pinpoint high-value customers. Given the effect of customer lifetime value on customers' potential value, the current study proposed the use of membership databases of retailers to predict customer lifetime value, in the hope of receiving more accurate customer value.

1.5. Differentiating Active and Sleeping Customers

It is very difficult to analyze membership databases of retailers, partly because the large membership as well as its massive basic information. Membership databases include the consumption track of every member. In addition to the card number, relevant information includes purchased item numbers, unit prices, discounts of each item, and overall discounts. Of course, many members do not purchase products after applying for the membership. They may also change their minds and purchase in competitors' stores. Therefore, the researcher in the current study also wanted to understand the analysis of targeting on active customers and of deleting sleeping customers to save resources. Due to the lack of a differentiation standard, the sleeping customers in the current study referred to those who did not purchase for six consecutive months. The deletion of them in the data analysis also excluded marketing activities designed for them. However, this method has been questioned since the identification method that lacked theoretical supports might have excluded many valuable customers.

1.6. Predict Customer Value with Enhanced Precision

In the analyses and investigations of customer purchasing behavior, the most important part is the customer-based analysis, where a customer's purchasing history is used to predict his/her potential purchasing behavior in the future. Then the predicted customer lifetime value is employed to differentiate customers in the database, thus facilitating the choice of target customers [2,3]. More precise and real customer value means more accurate customer potential value, with which businesses could accurately identify customers with higher potentials and effectively develop relevant marketing strategies.

1.7. Research Objective

This study takes supermarket members as the research object to uncover hidden information behind customer transaction data and help supermarkets understand customers' habits, develop marketing strategies, increase profits, and maintain long-term customer

relationships. In this study, we first employed the recency, frequency, and monetary (RFM) model to encode customer data and converted customer transaction data into R, F, and M values. Next, we used Ward's method along with K-means, fuzzy C-means (FCM), and other clustering methods to cluster existing customers. Following this, we calculated the customer lifetime value score (CLVS) of each group to obtain the customer lifetime value for each customer. Finally, we employed association rules to determine the correlations between customers buying one product and those buying another. According to each process step, this study is expected to achieve the following results:

1. Converting customer transaction data into a calculable value for measurement and comparison;
2. Grouping customers with similar RFMs to help supermarkets understand customer habits and develop marketing strategies for each customer group;
3. Calculating the customer lifetime value of each customer group to arrive at the value brought by each group of customers to the supermarket;
4. Determining the correlation between the products purchased by customers to juxtapose the products and thereby increasing the profitability for supermarkets.

2. Materials and Methods

2.1. Material

2.1.1. Supermarkets

Supermarkets are large retail stores, usually chain stores, selling not only daily necessities but also many different types of food, including fresh food, snacks, cooked food, and so on. They are based on self-service, have low gross margins and selling prices, and seek to sell more at lower margins.

American scholar [4] pointed out in his book *The Super Market* that, "a supermarket is a highly departmentalized retail establishment dealing in foods and other merchandise, either wholly owned or concession operated, with adequate parking space, and doing a minimum of \$250,000 annually".

Supermarkets are defined as markets that retail daily household necessities and food products with fresh and prepared food, including those that operate in the form of employees' consumer cooperatives or consortia and any business that engages in vending or catering in the premises of supermarkets [5].

The Ministry of Economic Affairs of the Republic of China [6] defines supermarkets as follows: "Those engaged in the provision of sectoral retail of daily household necessities and food products with mainly fresh and prepared food, including small-scale catering services provided on site".

Ref. [7] defines a supermarket in the following manner: "Supermarkets are departmentalized stores that sell daily foods for people's livelihood. They sell a variety of fresh foods (meat, fish, poultry, fruits, vegetables, and processed products) and daily necessities, most of which are sorted, neatly packaged, and labeled with weight, price, and date of manufacture. They are retail stores with self-service purchases, unified checkout, cash registers, and refrigeration equipment".

2.1.2. RFM Model

The three indicators in the RFM model, i.e., R, F, and M stands for recency, frequency, and monetary, respectively. Recency (R) refers to the time of a customer's last purchase and can be regarded as the customer's degree of activity toward the firm. Frequency (F) refers to how often a product is purchased during a period of time and can be regarded as the customer's loyalty to the firm. Monetary (M) refers to the amount of goods purchased by customers over a time span and can be regarded as the extent of the customer's contribution to the firm and customer value. Ref. [8] pointed out that RFM is a widely used analysis technique, and the RFM model can be used to run simple and quick customer analysis. With the three measurement indicators in the RFM model, customer consumption behavior can

be outlined simply and clearly to provide firms consumer behavior analysis and develop customer relationship management strategies [9].

Recency

Recency (R) means how long it has been since the last time a customer visited a store for the purpose of consumption. When the time of a customer's last visit to a store is closer to the present, the recency value will be smaller; conversely, when a customer's last visit to a store is farther from the present, the recency value will be larger. Assuming that a customer's recency value is small, the firm will be able to discern that there is a greater chance of a repeat visit; in other words, the customer's activity level is high.

Frequency

Frequency (F) means the number of times a customer purchases a product within a certain time span. When the customer purchases a greater number of times, the frequency value will be higher; conversely, when the customer purchases a fewer number of times, the frequency value will be lower. Given a high frequency of purchase, firms can discern that the customer's repeat purchase rate is high, i.e., the customer loyalty is high.

Monetary

Monetary (M) means the amount of money a customer spends on a product in a certain time span. When a customer spends more money on a product, the monetary value will be higher; conversely, when a customer spends less money on a product, the monetary value will be lower. Given a high monetary amount, firms can discern that the customer value is high, i.e., the customer contribution is high.

RFM Score

The RFM scores are divided into two parts: Hughes' RFM score and Stone's RFM score.

- Hughes' RFM Indicator

Ref. [10] sorted the recency, frequency, and monetary values for all customers and divided them into five parts: 0–20%, 21–40%, 41–60%, 61–80%, and 81–100%. Customers with percentages between 0–20% were assigned a score of "5", customers with percentages between 21–40% were assigned a score of "4", customers with percentages between 41–60% were assigned a score of "3", customers with percentages between 61–80% were assigned a score of "2", and customers with percentages between 81–100% were assigned a score of "1". The Hughes RFM code is shown in Table 2.

Table 2. Hughes' RFM ¹.

Percentage	Score
0–20%	5
21–40%	4
41–60%	3
61–80%	2
81–100%	1

¹ Data Source: Hughes (1994). Organized by author.

- Stone's RFM Indicator

Ref. [11] divided the recency values into five parts. When the customer's last visit to the store is within three months of the current time, the R value is given a "24" point ranking. When the time between the customer's last shopping and the current time is three to six months, the R value is given a "12" point ranking. When the last time the customer consumed in the store is six to nine months from the current time, the R value is given a "6" point ranking. When the time between the last time the customer visited the store and the current time is 9–12 months, the R value is given a "3" point ranking; if the time between

the last time a customer visits the store and the current time exceeds 12 months, the R value is given a “0” point ranking. The number of purchases (F) code is the number of times a customer purchases a product in a period of time multiplied by “4” points. When the number of purchases is 3, the F value will be $3 \times 4 = 12$ and scored as a “12”. The monetary (M) code is the amount of product purchased by the customer over a period of time multiplied by 1/10; however, the highest M value is limited to 9 points to avoid the occurrence of excessive monetary amount and insufficient number of purchases. When the monetary amount is NT\$100, the M value will be $100 \times 1/10 = 10$. However, due to M value’s highest score being 9 points, M value code is scored “9”. The Stone RFM codes are shown in Table 3.

Table 3. Stone RFM Codes ¹.

RFM	Score	Weight
Recency	Most recent 3 months = 24 points	Medium
	Most recent 3–6 months = 12 points	
	Most recent 6–9 months = 6 points	
	Most recent 9–12 months = 3 points	
Frequency	Most recent 12 months or more = 0 points Number of purchases \times 4 point	High
Monetary	Monetary amount \times 10 percent (The maximum point limit is 9)	Low

¹ Data Source: Stone (1995). Organized by author.

2.1.3. Clustering Method Application

This section reviews and discusses the methods used in this research: Ward’s method, K-means, Fuzzy C-means, FCM, self-organizing maps (SOM), and discriminant analysis (DA).

Ward’s Method

Ref. [12] proposed his method, also known as minimum variance method, which initially treats each individual as a group and then merges the clusters. The cluster with the smallest variance within the group is merged first. The earlier the cluster is merged, the higher is the similarity.

Ref. [13] applied Ward’s Method and K-Means to cluster non-regular employees. Results show that they are divided into four groups: “Work is Work”, “Work is Money”, “Newborn Tigers”, and “Youth is Capital”. This classification is based on their characteristics and respective group segments as well as describing them for management reference and suggestions.

Ref. [14] targeted the surrounding large-scale hypermarkets in Tainan and Kaohsiung as the main subjects of a survey. The research data were clustered through Ward’s Method to ascertain the competition among hypermarkets, thereby arriving at the following strategies: (1) business strategies wherein stores in the same business district tend to emphasize “future-oriented attributes”, and the operating performance deriving from aggressive strategies is superior to risk aversion and defensive strategies and (2) differing regions with significant differences in the time of establishment and degree of importance attached to business strategy attributes by position within the hierarchy.

To ascertain consumer behavior in wealth management between industries, Ref. [15] provided his successful strategic plan, utilizing Ward’s Method, K-Means, and variance analysis to explore consumer behavior in wealth management. The results of the study found that cross-sectoral wealth management consumers prefer to invest in stocks, funds, insurance plans, and bonds. While items such as currency exchange, trusts, futures, structured notes, and derivatives are less popular with consumers, it can be used as a frame of reference for financial professionals to promote cross-sectoral wealth management products.

The subject of a study by [16] was primarily the consumption behavior of tourists utilizing hotel services. The foundation of the study was tourist motivation treated as a

demographic, which was then divided into three consumer groups using Ward's Method and K-Means. Said groups were labeled "emotional exchange group", the "experience leisure group", and the "spiritual precipitation group". The tourist motivation and service quality of tourist hotels could then be analyzed through factor analysis, thereby yielding five factors that were analyzed through variance analysis to identify tourism motivation clusters. For the differences in the five tourism agencies, we use Scheffe's test to discover the clusters that had significant differences. This ultimately provided suggestions for positioning, marketing, and pricing strategies for pertinent consumer groups, which can serve as a reference for future business directions and strategies in the industry.

K-Means

K-means is one of the simplest unsupervised learning algorithms to solve common clustering issues. The process first specifies K clusters following a simple and streamlined method to group the given data points [17].

Ref. [18] proposed a facial recognition system based on ant colony optimization algorithms and K-means theory. Through the use of ant colony optimization to improve the flaw of K-means getting bogged down in local minimums to improve K-means clustering and correct facial recognition. The experimental results show that this improved the accuracy of facial recognition. There are three research contributions: (1) versatility using the Adaboost system to detect faces in complex environments, (2) reformability over the shortcomings of the traditional K-means algorithm getting bogged down in local minimums and improve the accuracy of facial recognition; and (3) scalability that has real-time facial recognition applications such as in anti-theft systems.

Ref. [19] proposed a mathematical algorithm based on K-means clustering algorithm and Bayesian belief networks that collects and extracts the characteristics of the packet information sent by the user during gaming and constructs a game add-on identification mechanism. It employs K-means clustering to reduce feature data dimensionality and assign attributes, establishing a Bayesian network architecture through those attributes to calculate the conditional probability to determine whether the user is an add-on. It was implemented in the game Ragnarok Online, and the experimental results show that the false positive and false negative rates can be lower than 5% at the same time, effectively resolving the issues caused by the add-on.

Ref. [20] applied K-means clustering analysis to the grouping of 98 sixth-grade students in a Changhua county elementary school. The students were divided into two groups: active learner and poor. Students' ability to concentrate and attitude toward learning mathematics are important indicators of grouping. Furthermore, intelligence and attitude toward learning mathematics can effectively predict and explain student learning achievement in decimal division. This is anticipated to provide teachers and researchers reference for teaching and grouping.

Ref. [21] applied RFM value analysis and K-Means grouping to segment customers into VIP and non-VIP groups and Formulated different marketing plans for different groups such that companies could concentrate resources on serving important customers and reduce the important customer attrition rate. In addition, it is imperative to understand customer needs and behavior to analyze customer buying habits and discern the items usually purchased by VIP customers. If a customer purchases a key item sequence, there is a high chance of the customer being a VIP customer.

Ref. [22] used the customer transaction records provided by a food processing plant to perform RFM and associative analysis of customers' patterns combined with individual RFM scores and associative patterns and divided customers into groups through K-Means, thus interpreting the purchase behavior characteristics for each cluster and examining the RFM weightings using decision tree analysis. The research results show that customers' buying behavior and monetary amount (M) are more important variables from which companies can Formulate marketing strategies.

Ref. [23] examined the customers of multi-level marketing companies using the RFM analysis method as an input value and K-Means to segment customers, proposing customer relationship management strategy recommendations based on the results of these groupings, finally ascertaining the characteristics of the customer groups through C5.0 decision trees and providing references for subsequent customer value prediction, thereby increasing revenue.

Ward's Method Plus K-Means

Ref. [24] pointed out that the number of clusters searched by the hierarchical Ward's Method should be used as the initial value, followed by searches performed by the non-hierarchical K-means, such that a satisfactory number of clusters can be found.

For more accurate clustering results, hierarchical and non-hierarchical clustering methods can be used. In the first stage of hierarchical clustering, the number of clusters can be determined in advance, and in the second stage, the non-hierarchical grouping method is used for clustering [25].

Ref. [26] believes that Ward's method plus K-means corrects the previous hierarchical clustering method, once clustering analysis is produced, it cannot be incorporated into inappropriate cluster observations for regrouping. At the same time, it overcomes the shortcomings in nonhierarchical clustering. The number of clusters and the center point need to be determined ahead of time.

Fuzzy C-Means

Ref. [27] proposed the fuzzy C-means (FCM) approach, while [28] extended it. FCM works primarily on the concept of fuzzy logic. Data points will not belong to any absolute cluster, but an integer between 0 and 1 is used to indicate the degree of its affiliation with a certain cluster, further enhancing the effectiveness of the clustering.

Ref. [29] pointed out that fuzzy C-means is a highly efficient tool in clustering analysis, with applications in medical diagnoses, pattern recognition, image processing, and geological model analysis.

Ref. [30] pointed out that FCM has wide-ranging applications in agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis, and target discrimination.

Ref. [31] utilized Fuzzy C-Means to execute clustering and calculated the central value vector of each cluster, using two cases, "identifying arrhythmias in heart patients" and "identifying quality of motor skill classifications", to verify the FCM algorithm's efficiency. The example results show that the average correct recognition rate for identifying arrhythmias of heart patients is 92.63%, and the average correct recognition rate for identifying motor skill categories is 96.0%, indicating that Fuzzy C-Means is a highly efficient, speedy, and practical clustering method.

Ref. [32] employed the Fuzzy C-Means (FCM) to process the data fed back by probe vehicles and recommended corresponding clustering and calculation methods under different road conditions. The results indicate that, when the average travel speed on a certain section of road is greater than 30 km/h (or the service level is at level B or higher), a clustering method that involves "dividing the high-speed group into two clusters" or "dividing a high-speed group into three clusters" can be utilized. Furthermore, when the average travel speed of a road section is less than 30 km/h (the road grade is below C level), it is recommended to "divide into three clusters to arrive at a medium speed group" or to take the "straight average".

Ref. [33] obtained the survey data of the various evaluation factors of a guesthouse after being a consumer at a bed and breakfast. Through Fuzzy C-Means, the weighting concept was added to develop a set of "guesthouse evaluations" and a "ranking model" to provide future reference for domestic guesthouses and reduce the gap between tourist need recognition and consumer expectations to increase willingness to stay and willingness to recommend.

Ref. [34] prepared a case study on auto parts services provided by automobile general agencies in Taiwan and introduced the RFM variable into Fuzzy C-Means, clustering the data into three groups referencing the basic statistical variables of the three clusters to establish customer patterns, serving as a reference basis for companies to quickly classify new customers, and provide management advice based on the transaction characteristics of each cluster.

Self-Organizing Maps

Self-organizing maps (SOM) were proposed by [35] and are commonly known as Kohonen network. The self-organizing map network is a high-quality data exploration tool that can map high-dimensional input data images to low-dimensional output topologies, providing users a cluster of visual or query data. In the updating process, researchers can update the input and output link weights through the steps of the learning process in response to future data updates [36].

Ref. [37] employed the clustering characteristics of the self-organizing map to explore the relationship between rainfall and river flow in the Choshui River Basin and groundwater changes in Choshui River fanhead, mountain area, and river mouth area, discussing SOM correlations between temporal-spatial factors in the model to groundwater changes.

Ref. [38] used RFM to analyze the consumer behavior in hotel duty-free shops and employed SOM for clustering, ultimately conducting decision tree and association analyses for accurate marketing.

Ref. [39] utilized a self-organizing map to pinpoint the sample types of customer groups, analyzed the clustering status to find out the types of customers, and finally developed marketing methods based on customer types.

Ref. [40] uses the basic information and historical transaction records of members in the database of outdoor activity supplies stores to conduct RFM customer value analysis, utilizing the Two-Step, K-Means, and SOM methods to maximize customers, ensure appropriate clustering, identify important and potential customers, and understand their purchasing behavior characteristics for outdoor products.

Discriminant Analysis

Discriminant analysis (DA) is mainly used to distinguish differences between groups in data. The earliest concept was put forth by eminent British statistician Ronald [41]. The discriminative analysis method involves creating two or more groups based on the characteristic data of the predictive variables. The goal is to find a linear combination of predictive variables and establish a set of discriminant models such that this linear discriminant analysis model can most effectively discriminate the groups [42].

Ref. [43] employed discriminative analysis as the primary tool to determine attributes combined with the newly designed Kano questionnaire and the Delphi method as an auxiliary tool when selecting elements and establishing discriminative functionality. The empirical results show that the 28 store quality factors include 4 glamour qualities, 5 one-dimensional qualities, 7 essential qualities, and 12 undifferentiated qualities. Finally, the indicators of increased probability of satisfaction and decreased probability of dissatisfaction were redefined to provide suggestions for improved order to the convenience store chains.

Ref. [44] used ECMWF Interim Reanalysis data and SSM/I microwave satellite inverted thermal synthesis to analyze statistical methods by differential analysis of hurricanes with secondary eyewalls (18) and strong hurricanes without secondary eyewalls (18) in Pacific Northwest from 2000 to 2011 and verify the degree of contribution of various environmental parameters for the hurricanes from 2012 to 2015.

Ref. [45] used the financial data of 27 public companies in Taiwan's plastic industry from 2004 to 2009 and TCRI credit score reports combined with the decision tree CART algorithm and step-by-step differentiation to conduct corporate credit score research, examine its correct prediction rate, and construct predictive models. The results of the

study found that the correct prediction rate using stepwise discriminative analysis was 72%, and the important variables were “operating profit ratio”, “accounts receivable days”, and “total assets”, indicating that they play an important role in predicting the company’s credit score.

Ref. [7] uses a quadratic discriminant function to classify consumers in Hsinchu City into “SOGO Department Stores”, “Sunrise Department Stores”, “Shin Kong Mitsukoshi Malls”, and “Far East Department Stores” groups, based on eight distinguishing variables. According to their respective customer loyalty groups, research results show that the probability of misclassification of male data is 10.29% and that of female data is 5%.

Ref. [46] used a case bank as a case study to explore the characteristics of how credit risk affects individual consumer credit loan application. Employing logistic regression, differential analysis, and linear programming models, research results show that the factors affecting the credit risk of personal consumer credit loan applications include education, position, average annual income, other bank loans, recent three months credit inquiries, and number of houses.

2.1.4. Customer Lifetime Value Score

Ref. [47] pointed out that the RFM method evaluates customer lifetime value (CLV). Ref. [48] pointed out that RFM information can be extracted from the general customer transaction data of firms. Therefore, the RFM model can be considered one of the most commonly used customer value analysis methods in the corporate world. Ref. [49] defines customer lifetime value as the total income that each customer brings to the firm. Ref. [50] pointed out that customer lifetime value has broad applications in performance measurements [51], target demographics [52], market allocation of resources [53], product offerings [54,55], price [56], and customer segmentation [57,58].

Ref. [59] added the card-carrying benefits and the two indicators of the average customer order based on the member information of a hypermarket case, calculated the customer’s value score, and then performed clustering, association analysis, and decision tree analysis to make predictions using a high value customer attrition model pyramid.

Ref. [60] considered the factors of individual and qualitative and group information simultaneously and calculated customer lifetime value on this basis. This can identify the most valuable customers and potential customers to Formulate corporate strategies. In addition, the understanding of customer consumption patterns helps the issuing bank build on the basis of customers when designing promotional activities.

Ref. [61] introduced customer lifetime value in the art industry, compared the gap between the current theory and art sales practices, calculated the initial retention probability of each customer by the drivers, and imported the Markov chain for the probability of future customer retention introduced in the modified customer lifetime value model to derive the customer equity model. Finally, the company’s marketing investment return rate was calculated as 16%.

Ref. [62] used the RFM-based PLANET Framework Model (PFM), case-based reasoning, Pearson correlation analysis, and association rules to establish a customer lifetime value management system to assist companies in developing good customer relation management strategies to maintain and improve customer value and improve corporate profitability.

Ref. [63] analyzed the case data of a domestic financial holding bank employing the RFM model to calculate customer lifetime value. The two-stage method was used to group customers to understand their characteristics and ultimately analyze the association rules to explore the differences between customers’ portfolios under the varying clusters, which are used as a basic reference for cross-selling of banking financial products.

2.1.5. Association Rules

Ref. [64] proposed the concept of association rules, which is a technique used to discover the relationships between large numbers of variables within a data set. It is

widely used in business decision-making processes. A classic example is the market basket analysis [65].

Ref. [66] captured head rotation angle and facial feature points from 6785 selfies to design 45 proportional features and employed the Apriori algorithm to find high-frequency item sets and their association rules. He also proposed five recommended strategies to make horizontal and vertical directional adjustments and applied the Kappa value to evaluate the performance of the recommendations.

Ref. [67] used association rules to analyze school entrance scores and grade point averages over three school years. The results showed that the English entrance scores were good, and the GPAs for all subjects were good. Past entrance score records were good, while GPAs for all subjects were weak. Past entrance scores records were weak, while GPAs for all subjects were good. Therefore, the schools can utilize this method to find the admission standards of each school department to achieve its characteristics and goals.

Ref. [68] mainly explored the online shopping behavior of customers, used association rule analysis technology to ascertain customers' online shopping characteristics, and employed clustering technology to group customers with the same characteristics for verification. Research results show that the largest online shopping demographic was 21–30-year-old students. Customers who often go online and are willing to recommend shopping platforms and products repurchase products more frequently; thus, this demographic can be categorized as loyal customers.

Ref. [69] used customer deposits of a private bank in Taiwan as data. Employing association rules to estimate customer deposit behavior, discovering a correlation between customer attributes and product attributes, showing that customer deposit patterns can assist banks in introducing related product mixes and act as a reference for marketing strategy.

Ref. [70] utilized association rules to explore the patterns of students' product purchases with the factors of "term" and "temperature". Results of the study show that they buy milk in the morning and tea at night, and when the temperature is low, they buy hot Ovaltine drinks. Through analysis, decision makers can understand the correlation between different products and match promotions to increase store profitability.

Ref. [71] treated the e-commerce industry, which places great importance on customer lists as an example case. By segmenting customers with RFM, customers with similar consumption patterns were discovered through association rules, after which different products were recommended for different customer groups.

2.1.6. Summary

According to the review of the supermarket industry, the RFM model, application of the clustering method, customer lifetime value, and association rules, literature on the application of clustering methods indicates that the application of clustering methods in other industries is far greater than in the retailing industry. This study applied the customer groupings of supermarket retailers. In terms of clustering methods in literature, it was found that Ward's Method plus K-means is more common. Fuzzy C-Means is an efficient, high-speed, and practical grouping method, and self-organizing maps is a kind of neural network-like method; therefore, this study determines which of the three clustering methods is the most effective. The basic concepts of monetary value calculations present in literature on customer lifetime value were applied to score customer lifetime value.

2.2. Research Method

This study constructs a customer lifetime value model to measure customer value. Armed with clustering models, this study helps companies find their meaningful target demographics by discovering the correlations between customer groups' purchase of various products. This section is divided into six parts:

1. research framework
2. data description
3. RFM model

4. grouping
5. customer lifetime value score
6. association rules.

2.2.1. Research Framework

This research first utilized the RFM model to convert customer transaction data into R-values, F-values, and M-values, which were divided into two groups, namely, customer RFM and product RFM. Next, Ward's Method plus K-means, Fuzzy C-Means, and self-organizing maps were utilized. These clustering methods are used for grouping, and the method with the highest classification accuracy rate of the three methods was ascertained via differential analysis. The clustering method continued to be used over the next steps, wherein the customer lifetime value score of each group was calculated and the clusters were ranked from high to low. Finally, association rules were utilized to ascertain the correlations between customers who buy various products to group the highly correlated products and increase supermarket profits. The research framework is shown in Figure 1.

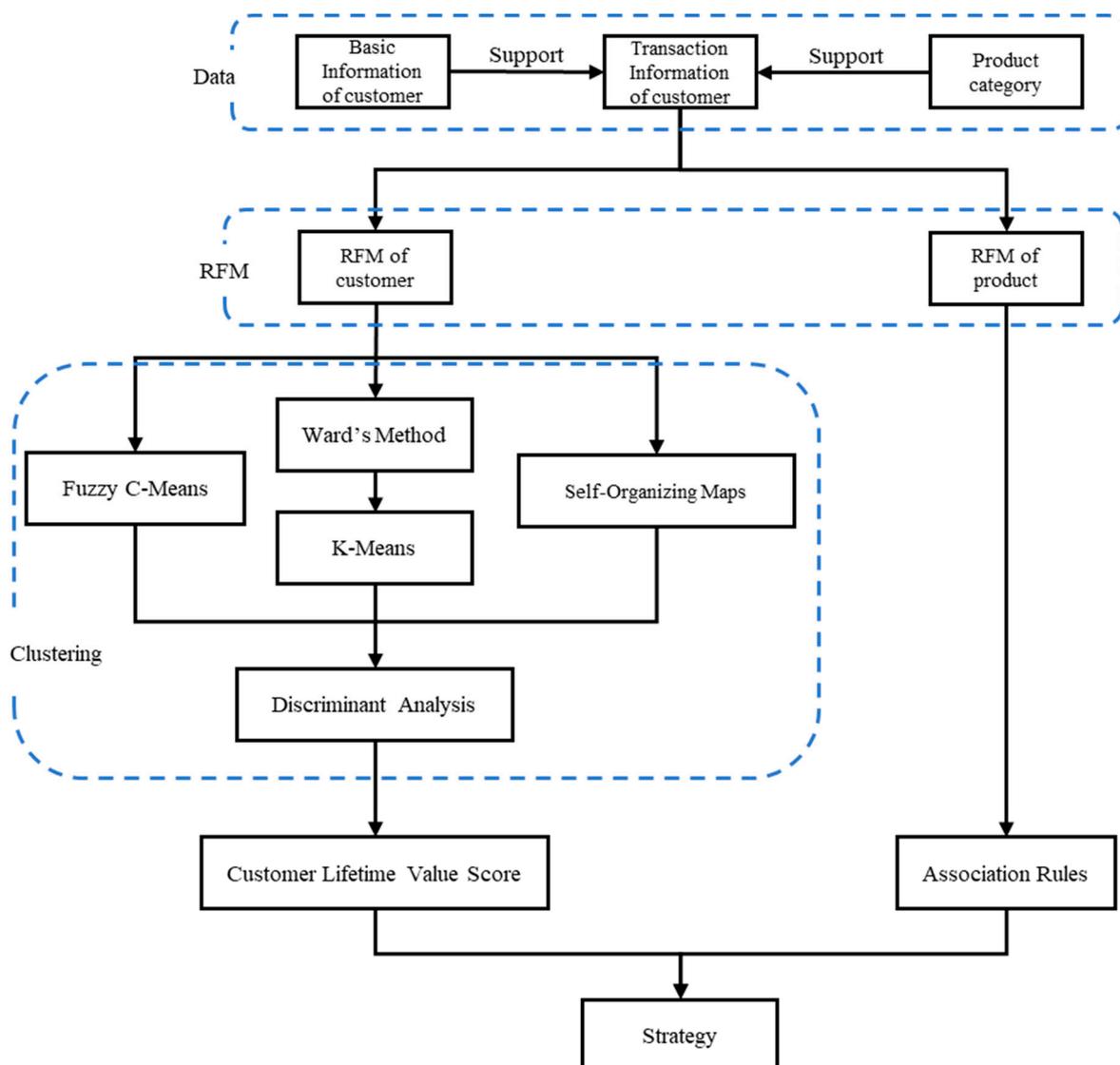


Figure 1. Research framework.

2.2.2. Data Description

A retailing mart database of over one million members was analyzed in the present study. The primary data were remarkably large, which included not only the member's card number but also the purchase history of each member on each item, such as the quantity of items purchased each time, unit price, item discount, and discount information after subtotaling. The present study focused on identifying regular patterns from these data that exceeded 1.5 PB (1 PB = 1024 TB). In this study, when customers made purchases, various reward schemes were used to attract them to sign up as members. Customers received an NT\$100 discount coupon for future redemption after filling in their member demographic data. Member information was tabulated, in which every purchase for which their membership card numbers were entered, and their purchase details were recorded accordingly. Information thus collected served as the primary data for marketing differentiation. The customers were then classified according to their annual purchase amount. This facilitated targeting of marketing activities at various levels based on customers' preferences.

2.2.3. RFM Model

This study has two parts. The part I is the customer's RFM score, and part II is the product's RFM score. According to [10] quintile method, it is found that the data tends to be unequal. Since the box plot is not affected by outliers, box plot can describe the discrete distribution of data in a relatively stable way. Therefore, this study uses the box plot method to separate the data into five parts.

2.2.4. Box Plot

In descriptive statistics, the box plot is a method for graphically depicting groups of numerical data through their quartiles. The box plot is a standardized way of displaying the dataset based on a five-number summary, the minimum, the maximum, the sample median, and the first and third quartiles (Figure 2).

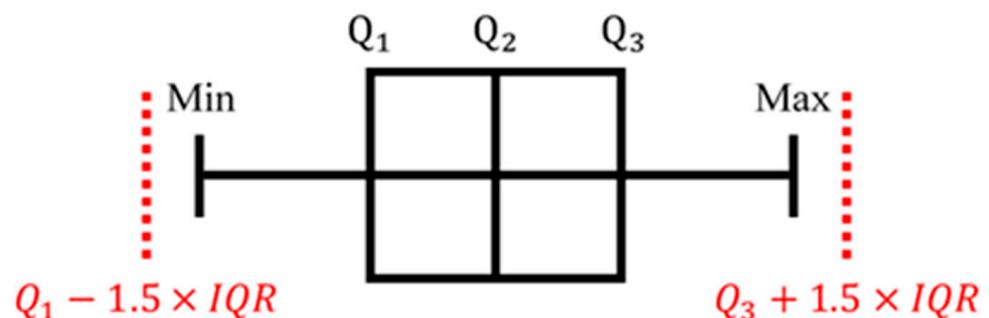


Figure 2. Box plot.

The steps are as follow:

Step 1, Quartile

In statistics, a quartile is a type of the quantile which divides the number of data points into four parts, or quarters, of more-or-less equal size. The data must be ordered from smallest to largest to compute quartiles; as such, quartiles are a form of order statistic. The three main quartiles are as follows:

1. The first quartile (Q_1) is defined as the middle number between the smallest number (minimum) and the median of the dataset.
2. The second quartile (Q_2) is the median of a dataset.
3. The third quartile (Q_3) is the middle value between the median and the highest value (maximum) of the dataset.

4. Calculate the quartile as Formula (1). If the quartile is an integer, calculate with Formula (2).

$$Q_k = n \times \frac{k}{4}, \quad k = 1, 2, 3 \quad (1)$$

$$x = \frac{Q_k_data + (Q_k + 1)_data}{2} \quad (2)$$

Step 2, Interquartile Range, IQR

The interquartile range is Formula (3). The third quartile is Q_3 , and the first Quartile is Q_1

$$IQR = Q_3 - Q_1 \quad (3)$$

Step 3, Inner Fence

We find the inner fences. We start with the IQR and multiply this number by 1.5. We then subtract this number from the first quartile. We also add this number to the third quartile. These two numbers form our inner fence. Formulas (4) and (5), Q_3 is the third quartile, Q_1 is the first Quartile, IQR is the Interquartile Range and remove the Outlier.

$$U = Q_3 + 1.5 \times IQR \quad (4)$$

$$L = Q_1 - 1.5 \times IQR \quad (5)$$

Step 4, The maximum and minimum

When we remove the Outlier. We can find the maximum and minimum.

2.2.5. Customer's RFM Score

According to Box Plot and weight score of [11],

Step 1, Check the data of purchase (e.g., purchase date, average point per transaction, and average cost) according to the box plot displaying the dataset based on a five-number summary of R, F, and M.

Step 2, Combine the customer's RFM score is 5-5-5, 5-5-4, 5-5-3, . . . , 1-1-1.

Step 3, Follow the weight score of the stone.

Step 4, Setup the weight scores: F = 5, R = 3, and M = 1

Step 5, Multiply the customer's RFM value by its corresponding weight value, Formulas (6)–(8):

$$CR^* = R \times 3 \quad (6)$$

$$CF^* = F \times 5 \quad (7)$$

$$CM^* = M \times 1 \quad (8)$$

CR^* , CF^* , CM^* are the weighted value of customers' R, F, and M.

2.2.6. Product's RFM Score

Many studies only explore the customer's last purchase date, purchase frequency, and purchase amount. However, this study believes that product's RFM scores are also very important. Therefore, the research add the score of the products' RFM.

Step 1, Divide into all products five categories: "row food", "instant food", "dry good", "people's livelihood food", and "others".

Step 2, Divide into all products five categories by the customer.

Step 3, Check the transaction data of customers. Sort the product's last purchase date, purchase time and purchase amount.

Step 4, Permutation and combination of product values. 5-5-5, 5-5-4, 5-5-3, . . . , 1-1-1.

Step 5, Based on the weight value of [11], the weight value.

Step 6, Setup the value of the weight, where F is 5, R is 3, and M is 1.

Step 7, Multiple the customer's RFM score by its corresponding weight value in Formulas (9)–(11):

$$PR^* = R \times 3 \quad (9)$$

$$PF^* = F \times 5 \quad (10)$$

$$PM^* = M \times 1 \quad (11)$$

2.2.7. Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning (Figure 3).

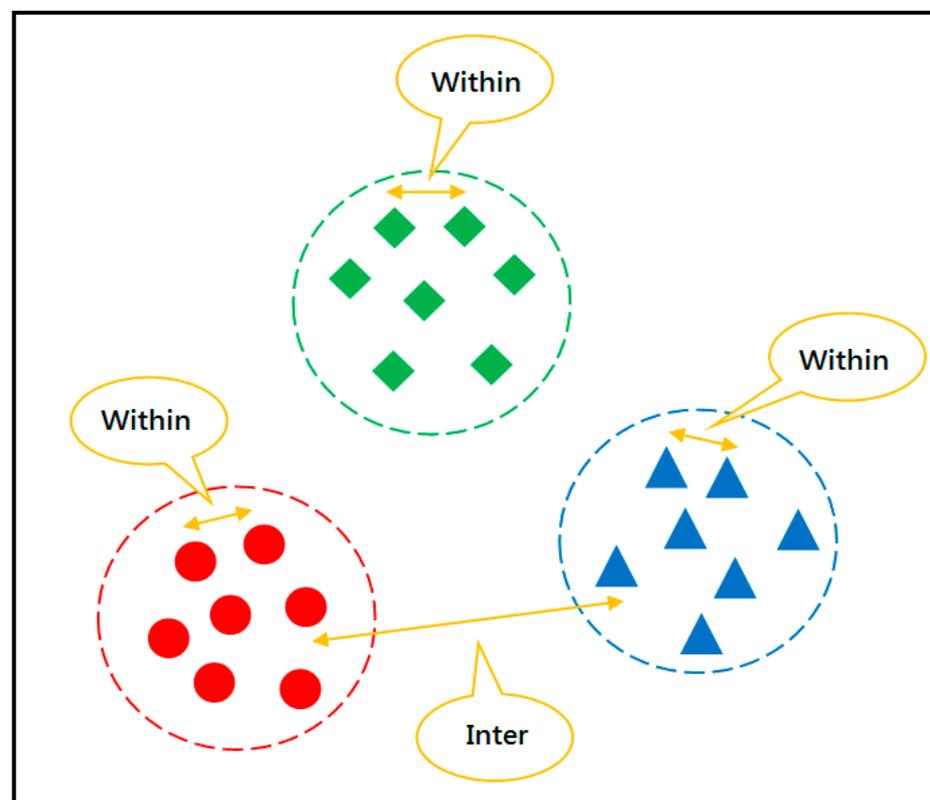


Figure 3. Clustering.

2.2.8. Euclidean Distance

The Euclidean distance between two points in either the plane or 3-dimensional space measures the length of a segment connecting the two points. This is shown in Figure 4. It is the most obvious way of representing distance between two points (Formula (12)).

$$D_{AB} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (12)$$

Example: $A(0,0)$, $B(3,3)$, Euclidean Distance: $D_{AB} = \sqrt{(0 - 3)^2 + (0 - 3)^2} = \sqrt{18}$

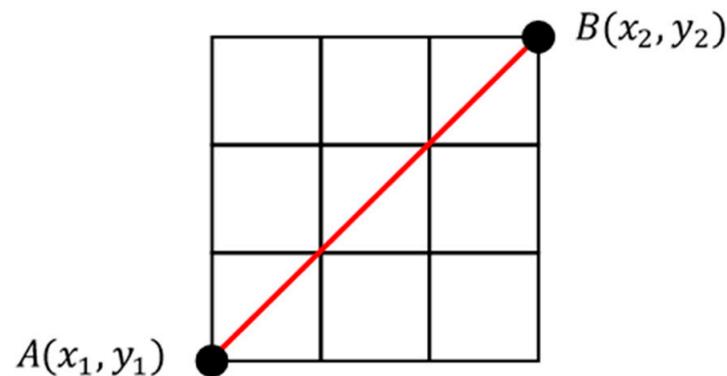


Figure 4. Euclidean distance.

2.2.9. Ward's Method

In statistics, Ward's method is a criterion applied in hierarchical cluster analysis. Ward's minimum variance method is a special case of the objective function approach originally presented by Joe H. Ward. Ward suggested a general agglomerative hierarchical clustering procedure, where the criterion for choosing the pair of clusters to merge at each step is based on the optimal value of an objective function. Many of the standard clustering procedures are contained in this very general class. To illustrate the procedure, Ward used the example where the objective function is the error sum of squares (Sum of Squares Error, SSE), and this example is known as Ward's method or more precisely Ward's minimum variance method (Formula (13)).

$$SSE = \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^k (n_i - 1) S_i^2 \quad (13)$$

i shows the group size, j shows the sample size, x_{ij} is the value of group i , \bar{x}_i is the average of i the group. n_i is the number of samples, and S_i^2 is the variation of the group.

Step 1, Treat each customer as a separate group, the error sum of squares is 0.

Step 2, Calculate the individual error measurements (Formula (12)).

Step 3, Combine the two groups with the least increase in the sum of squared errors.

Step 4, Repeat Step 3. Combine the group until all are merged into a group.

Step 5, Find the smallest sum of squared errors and merge them together.

This is best group.

2.2.10. Customer Segmentation Using K-Means Clustering

K-means clustering reduces the data by categorizing or grouping similar data items together. Such grouping is common to how humans process information. Furthermore, clustering algorithms are used to provide automated tools to help in constructing categories or taxonomies. These methods may also be used to reduce the effects of human factors in the process.

K-means clustering is a common method for partitioning factors, as it is closely related to the SOM algorithm. In K-means clustering, the criterion function is the average squared distance of the data items X_k from their nearest cluster centroids.

Step 1, Customer Segmentation to k groups and random designation the center of K groups (Figure 5).

Step 2, Calculate every customers' Euclidean Distance (Figure 6).

Step 3, Initial clustering (Figure 7). Find the new clustering center of customers (Figure 8). Reallocate customers to the nearest cluster (Figure 9).

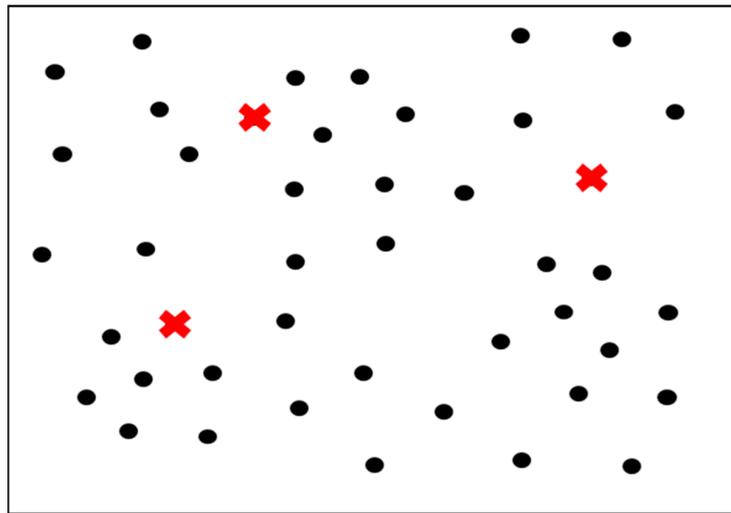


Figure 5. Random designation of the center of the K groups.

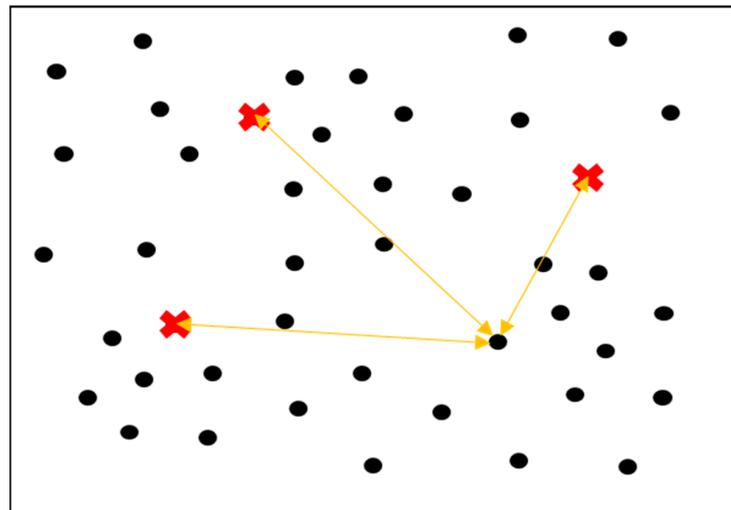


Figure 6. Calculate every customers' Euclidean distance.

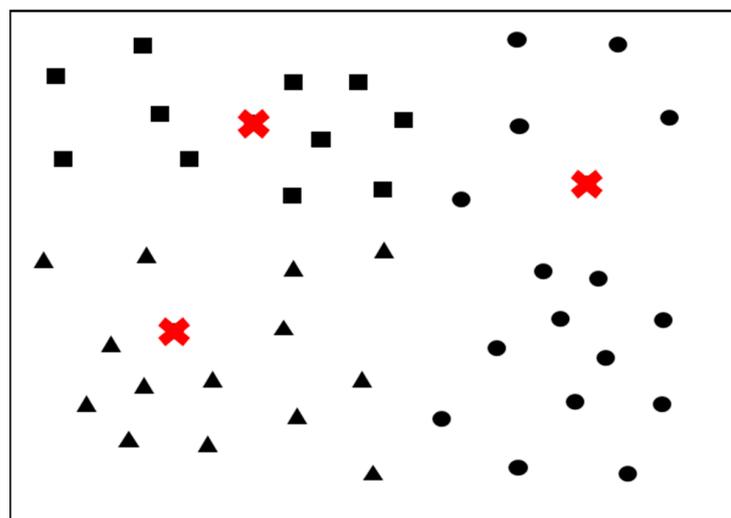


Figure 7. Initial cluster.

Step 4, Find the new clustering score of customers.

Step 5, Repeat Step 2 and Step 3, and reallocate customers to the nearest cluster.

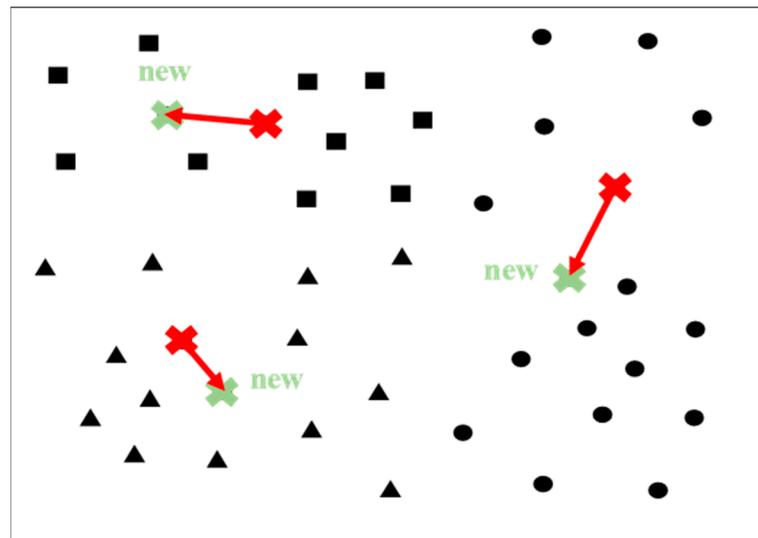


Figure 8. New clustering score of the customers.

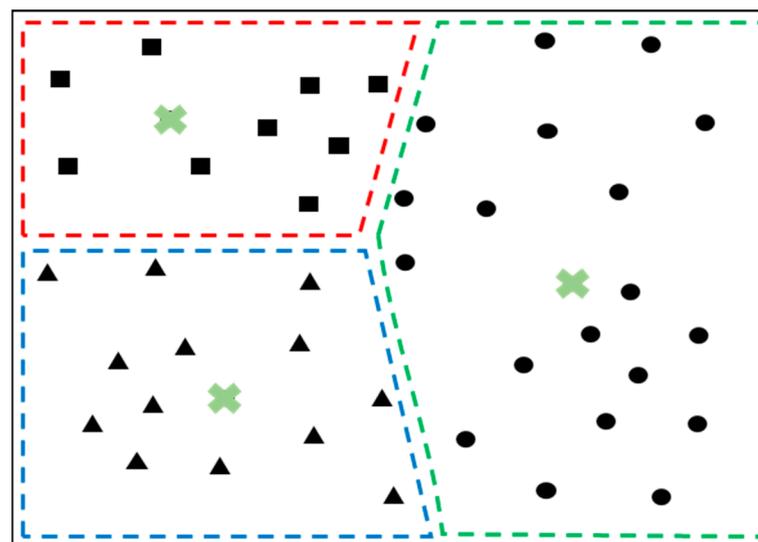


Figure 9. Reallocate customers to the nearest cluster.

2.2.11. Fuzzy C-Means

Fuzzy c-means clustering was developed by [27] and improved by [28]. Fuzzy clustering (also referred to as soft clustering or soft k-means) is a form of clustering in which each data point can belong to more than one cluster.

Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible. Clusters are identified via similarity measures. These similarity measures include distance, connectivity, and intensity. Different similarity measures may be chosen based on the data or the application. Through the concept of Fuzzy Logic. Fuzzy logic is designed to solve problems by considering all available information and making

the best possible decision given the input. So Fuzzy C-Means can further improve the effect of clustering (Formulas (14) and (15)).

$$U = \{u_1, u_2, \dots, u_c\} U_{c \times n} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ u_{c1} & \dots & \dots & u_{cn} \end{bmatrix}_{c \times n} \tag{14}$$

$$J_m(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2, \quad 1 \leq i \leq c, \quad 1 \leq j \leq n \tag{15}$$

Degree of members $m \in [1, \infty)$ $1 \leq m < \infty$, u_{ij} is the membership function of the observation i in the cluster j , v_i customer center point of the i group.

We hope to reduce the objective function by minimizing.

$$\min \left\{ J_m(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2 \right\}, \quad 1 \leq i \leq c, \quad 1 \leq j \leq n \tag{16}$$

Optimization of the problem by the Lagrange function (Formula (17)).

$$f = J_m(U, V; X) + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2 + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) \tag{17}$$

Get the best parameter v_i by partial differential.

$$v_i = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m} \tag{18}$$

Partitions a numeric dataset by using the Fuzzy C-Means (FCM) clustering algorithm [28]. The step as follows:

Step 1, Set the number of clusters to.

Step 2, Define the initial value, such as Formulas (19) and (20):

$$0 \leq u_{ij} \leq 1, \quad 1 \leq i \leq c, \quad 1 \leq j \leq n \tag{19}$$

$$\sum_{i=1}^c u_{ij} = 1, \quad 1 \leq j \leq n \tag{20}$$

Step 3, The objective function is zero

Step 4, Calculate the new v_i as Formula (18)

Step 5, Recalculate the value of u_{ij} (Formula (21)):

$$u_{ji} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{\frac{1}{m-1}}}, \quad 1 \leq i \leq c, \quad 1 \leq j \leq n \tag{21}$$

Step 6, Calculate the objective function as Formula (16)

Step 7, back to Step 4, while the value of the objective is less than the threshold value

2.2.12. Self-Organizing Maps

Self-organizing map (SOM) technique was developed by [35]. Thus, self-organizing maps are neural networks that employ unsupervised learning methods and mapping their weights to conform to the given input data with a goal of representing multidimensional data in an easier and understandable form for the human eye.

In the topology and related areas of mathematics, a neighborhood function is one of the basic concepts in a topological space. A neighborhood function can reduce the dimensionality of the dataset. Visualization can be used to present high-dimensional data structures with low-dimensional graphics. This is shown as Figure 10.

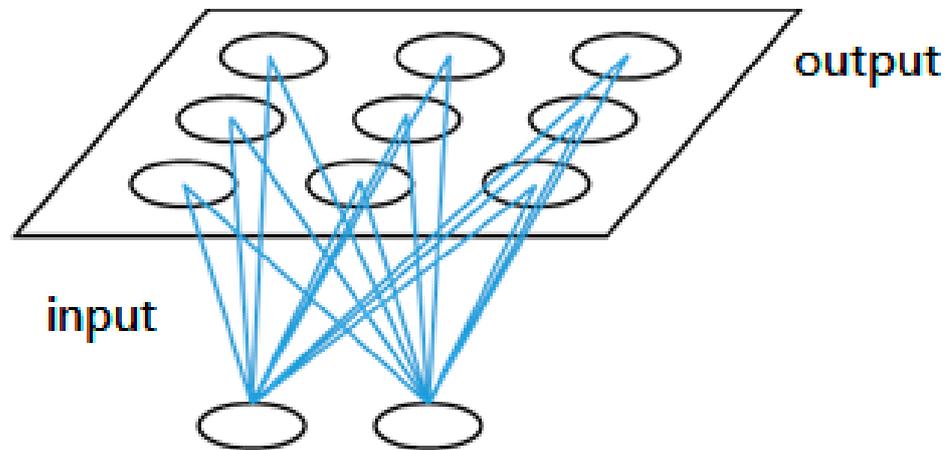


Figure 10. SOM schematic diagram. Data source: Cuellar & Neira (2013).

How to use the SOM to cluster the customers

Step 1, Input the dataset of customers. The datasets of customers are data vectors. V represents the input vector of n dimensions. A vector of n dimensions and input SOM system with Formula (22):

$$V = \{V_1, V_2, V_3, \dots, V_n\} \quad (22)$$

Step 2, Set up the learning parameters, initial weight W_j , Neighborhood Function K_{qj} , and learning speed η .

Step 3, Input Euclidean distance of data vector V and weight vector W_j to choose the winner. This is shortest distance j^* , Formula (23):

$$j^* = \min_{V_j} \{ \|V - W_j\| \} \quad (23)$$

Step 4, Use the Neighborhood Function to find customers who are close to the Winner, Formula (24):

$$K_{qj} = e^{-\left(\frac{\|r_j - r_q\|}{R}\right)^2} \quad (24)$$

K_{qj} is the value of Neighborhood from j th customer-to-customer q of the winner. r_j is the topology ordinate of j th, and r_q is the topology ordinate of the winner. R is a neighborhood radius, Figure 11.

- Neighborhood center: Focusing on winning customers, revise all customers in the neighboring area.
- Neighborhood radius: Take a larger radius value first, but as the number of learning time increases or time increases, the radius can be gradually reduced (Formula (25)):

$$R^{n+1} = \lambda \times R^n, \lambda < 1 \quad (25)$$

- λ is the factor of Neighborhood radius. n is learning times. Each time the network learns, the Neighborhood radius will decrease once (Formula (26)):

$$\frac{R^{n+1}}{R} = R^n \quad (26)$$

- Neighborhood area: Take Neighborhood center as the center point and the length of Neighborhood radius as the radius, and the area enclosed by it.
- Neighborhood distance: Distance between customer j and winning customer q (Formula (27)):

$$D_{qj} = \|r_j - r_q\| = \sqrt{(x_j - x_q)^2 + (y_j - y_q)^2} \quad (27)$$

$r_j(x_j, y_j)$ is topological coordinates of customer j , $r_q(x_q, y_q)$ is topological coordinates of winning customer.

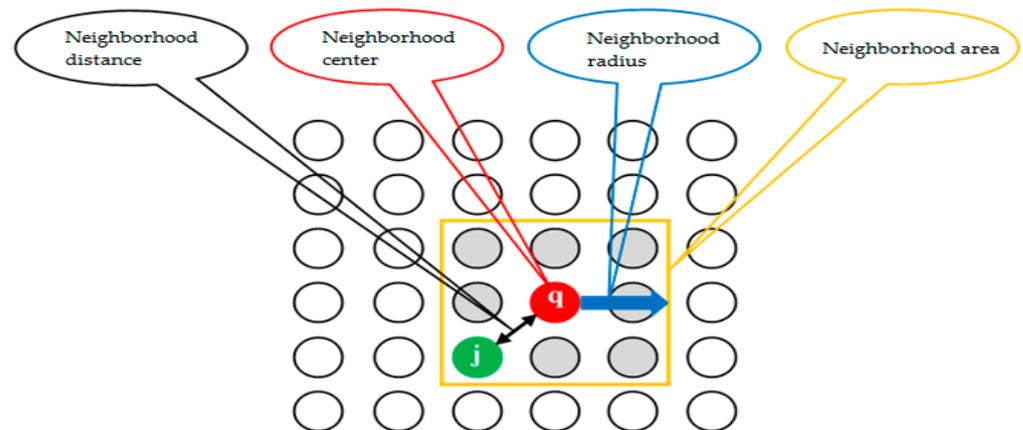


Figure 11. Parameter of neighborhood.

Step 5: Update the weight vector (Formula (28)):

$$W_j = W_j + \Delta W_j \quad (28)$$

ΔW_j is $\eta \times (V - W_j) \times K_{qj}$ is the weighted value correction matrix for the j th customer. η is learning speed. K_{qj} is the neighborhood function of customer q and customer j th. The greater the proximity distance between the customer and the winning customer q the smaller the proximity function and the smaller the weight correction.

Step 6: Repeat Step 3, Until the Self-Organizing Maps is formed.

2.2.13. Linear Discriminant Analysis

Linear discriminant analysis (LDA), normal discriminant analysis (NDA), or discriminant function analysis is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

Figures 12 and 13 are two groups. Figure 13 C_1 , C_2 be project at x_1 axis and x_2 axis. We found two types of data will overlap. Therefore, this is poor discrimination.

C_1 , C_2 project at FLD have best discrimination.

Linear Discriminant Analysis (Formula (29)):

$$y_i = b_0 + b_{i1}x_1 + b_{i2}x_2 + \dots + b_{ik}x_k \quad (29)$$

y_i is discriminant Function, b_0 is constant term, b_i Discriminant coefficient, k is quantity of independent variables x .

2.2.14. Customer Lifetime Value

The definition of customer value is different from the perspective of the enterprise and the customer. For the enterprise, it refers to how much the customer can contribute

to the enterprise. For customers, it refers to the products and services that companies can provide.

Customer lifetime value can also be defined as the monetary value of a customer relationship, based on the present value of the projected future cash flows from the customer relationship. This study wants to find the company's target customer.

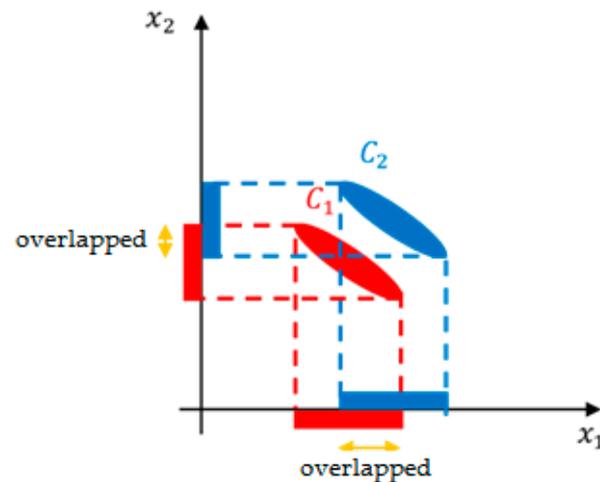


Figure 12. Project to the x_1 axis and x_2 axis. Data Source: PPV (2016).

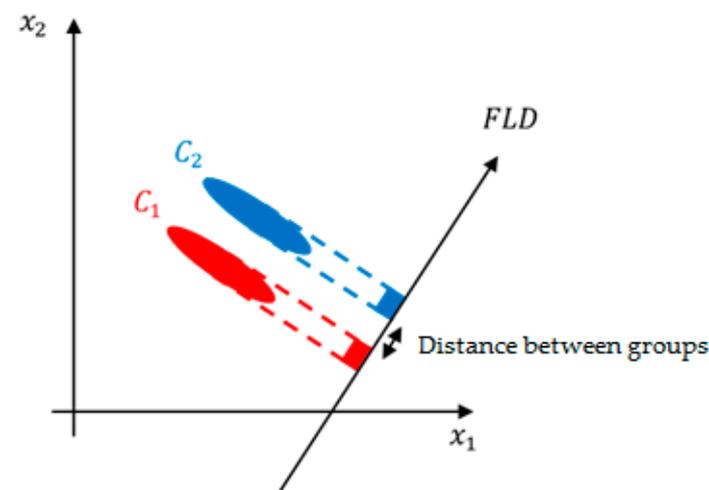


Figure 13. Project to the straight line of the FLD. Data Source: PPV (2016).

Calculate customer Lifetime Value Score after customers are grouped. The steps are as follows:

Step 1: Transform the purchase data to RFM value.

Step 2: Setup the value of weight is $(R, F, M) = (\text{Medium}, \text{High}, \text{Low})$.

Step 3: Substitute the value of RFM into the Formulas (6)–(8).

Step 4: Calculate the life value of customers, Formula (30):

$$CLVS = CR^* + CF^* + CM^* = R \times 3 + F \times 5 + M \times 1 \quad (30)$$

CR^* , CF^* , CM^* is weighted values of customers' R, F, M.

2.2.15. Apriori Algorithm

Based on the concept of strong rules, Rakesh Agrawal, Tomasz Imieliński and Arun Swami introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. Ref. [65]

Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, if you buy a pint of beer and do not buy a bar meal, you are more likely to buy crisps at the same time than somebody who did not buy beer.

2.2.16. Association Rules

Apriori Algorithm is an algorithm for frequent item set mining and association rule learning over relational databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis (Figure 14).

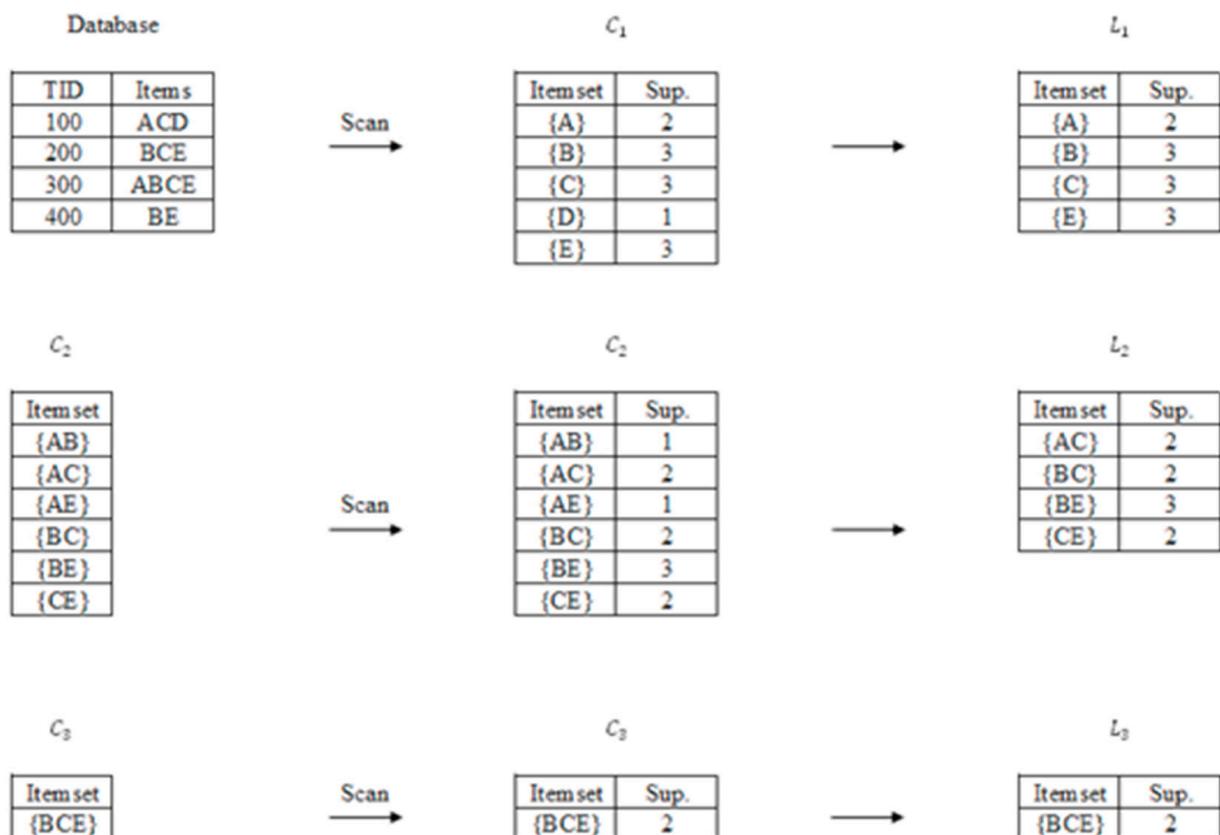


Figure 14. Apriori Algorithm.

Step 1: Scan the database and get the single dataset is C_1 . Extend one item at a time and then it prunes the candidates which have an infrequent sub pattern. The result is frequent dataset L_1 .

Step 2: Combine two dataset to candidate item sets C_2 contain two dataset.

Repeat the step 1 and step 2. The result is frequent item sets L_2 .

Step 3: Combine the dataset to candidate item sets C_3 .

Step 4: Get the frequent item sets of $L_1, L_2, L_3, \dots, L_k$.

3. Results

3.1. Data Description

Subsubsection

This study analyzed, eliminated, and compared the big data of member information retrieved from a retailing industry database. This study focused on the customer purchase behavior database of 300 member from 1 July 2013 to 31 December 2013. A total of 117,003

transaction records in six months. The primary database contains a large amount of information on the purchase history of each member, including member's card number, the quantity of items purchased each time, unit price, item discount, and discount information after subtotalling (Table 4). Product category information includes category number and category name. Customer transaction data contain a large amount of information on the purchase history of each member (Tables 5 and 6).

Table 4. Basic information of the customer.

Item	Remark	
Member's card number		
Birthday	YYMMDD	
Gender code	1: Male	2: Female
Gender		
Marriage code	1: Married	2: Unmarried
Marriage		
Number of children		
Occupation code	1: Business 2: government employees 3: Worker 4: Service industry	5: Housewife 6: Student 7: Other 8: Unemployed
Education code	1: PHD 2: Master 3: University	4: College 5: High school 6: Under
Education	Highest education	
Income class code	1: Under 40 thousand 2: 40~59 thousand 3: 60~89 thousand	4: Over 9 thousand 5: Other
Income		
Family population code	1: 1~2 2: 3~4 3: 5~6 or more	
Family population		
Address code	Postal 3 code	
Address	Counties/City	

Table 5. Product category.

Item	Remark
Category number	
Category name	

Table 6. Transaction information of the customers.

Item	Remark
Purchase date	YYYYMMDD
Member's card number	
Serial number	
Serial number of product	
Subcategory	Category number
Price	
Quantity	
Subtotalling	

3.1.1. RFM Model

According to [10], the quintile method, it is found that the data tends to be unequal. Since the box plot is not affected by outliers, box plot can describe the discrete distribution

of data in a relatively stable way. Therefore, this study uses the box plot method to separate the data into five parts.

3.1.1.1. Score of Customers' RFM

The box plot of last purchase date. Distance between the last purchase date to 31 December 2020 are 42 days, 34 days, 31 days, 29 days, 27 days, 20 days, 15 days, 13 days, 12 days, 10 days, 9 days and 8 days. These numbers are outliers. The maximum value is 7 days. The minimum value is 0 day. The first quartile is 0 day, and the third quartile is 3 days. The median value is 1 day.

This study divides the last purchase date into five parts and take its integer value.

"Minimum value 0 day ~ First Quartile is 0 day", "First Quartile is 0 day~The Median value is 1 day", "The Median value is 1 day ~ Third Quartile is 3 days", "Third Quartile is 3 days~The maximum value is 7 days" and "upper/down fence value is 7 days/0 day".

$R = 0$ day, $0 < R \leq 1$ day, $1 < R \leq 3$ days, $3 < R \leq 7$ days, $R > 7$ days, and $R < 0$ day.

$R = 0$ day, the score of customer last purchase date is "5" points. $0 < R \leq 1$ day, the score of customer last purchase date is "4" points. $1 < R \leq 3$ day, the score of customer last purchase date is "5" points. $3 < R \leq 7$ day, the score of customer last purchase date is "2" points. $R > 7$ day, the score of customer last purchase date is "1" point and delete $R < 0$ day (Figure 15).

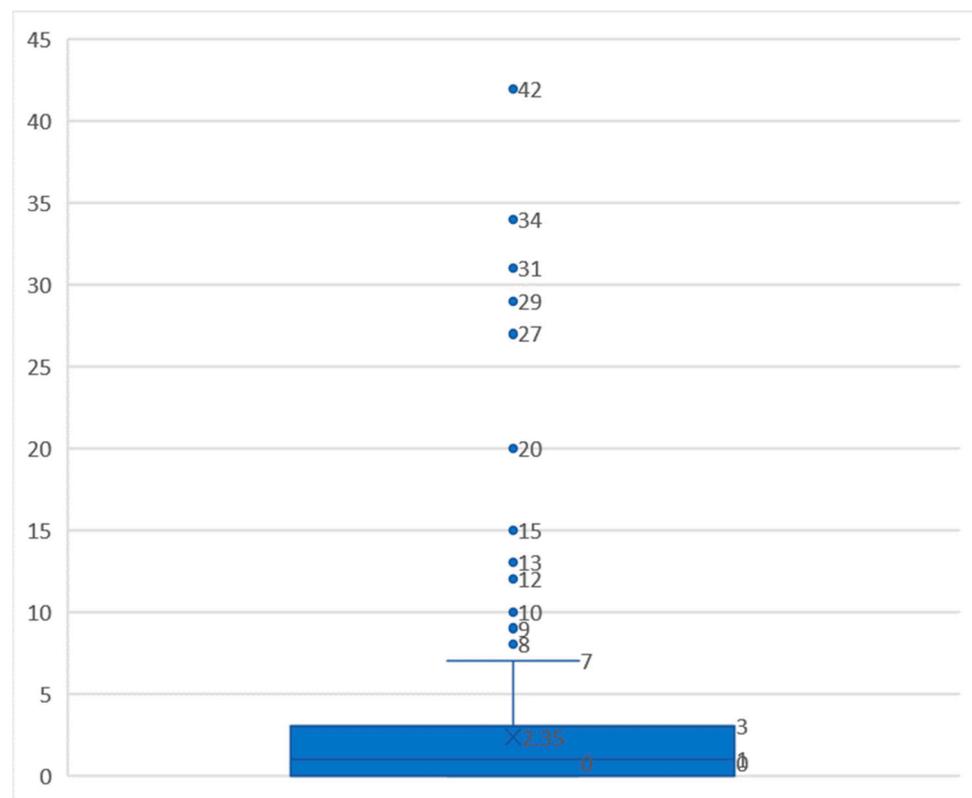


Figure 15. The box plot of last purchase date.

The box plot of purchase frequency, The number of purchase by customers are 168 times, 92 times, 88 times, 56 times, 54 times, 50 times, and 44 times within six months. These numbers are outliers. The maximum value is 87 times. The minimum value is 60 times. First quartile is 70 times, and third quartile is 77 times. The median value is 74 times.

This study divides the purchase frequency into five parts and take its integer value.

"Third Quartile is 77 times ~ The maximum value is 87 times", "The Median value is 74 times ~ Third Quartile is 77 times", "First Quartile is 70 times ~ The Median value is

74 times”, “The minimum value is 60 times ~ First Quartile is 70 times” and “upper/down fence value is 87 times/59 times”.

$R = 0$ day, $0 < R \leq 1$ day, $1 < R \leq 3$ days, $3 < R \leq 7$ days, $R > 7$ days, and $R < 0$ day.

$77 < F \leq 87$ times, $74 < F \leq 77$ times, $70 < F \leq 74$ times, $60 < F \leq 70$ times, $F > 87$ times, or $F < 59$ times.

When $77 < F \leq 87$ times, the score of customer purchase frequency is “5” points. $74 < F \leq 77$ times, the score of customer purchase frequency is “4” points. $70 < F \leq 74$ times, the customer purchase frequency is “3” points. $60 < F \leq 70$ times, the score of customer purchase frequency is “2” points. $F > 87$ times or $F < 59$ times, the score of customer purchase frequency is “1” points (Figure 16).

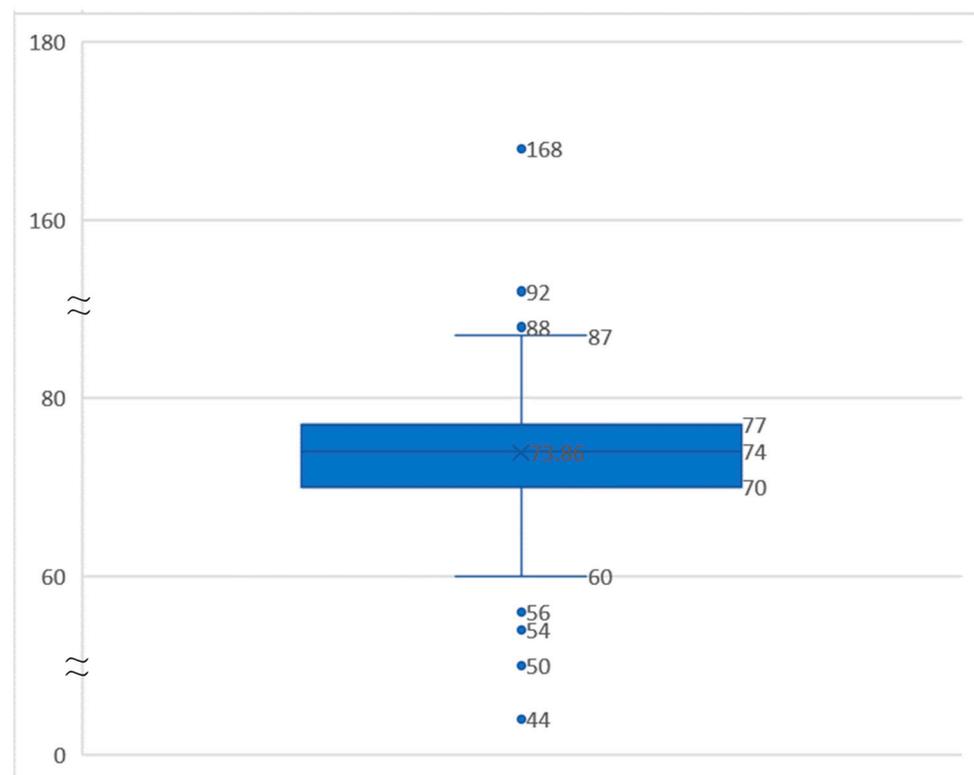


Figure 16. The box plot of the purchase frequency.

The box plot of purchase amount. The price of purchases by customers is \$18,666 dollars, 5457 dollars, 4746 dollars, 4278 dollars, 4085 dollars, 3956 dollars, 3880 dollars, 3780 dollars, 3527 dollars, 3261 dollars, and 3026 dollars. These numbers are outliers. The maximum value is 3022 dollars. The minimum value is 319 dollars. First quartile is 979 dollars and the third quartile is 1796 dollars. The median value was 1276 dollar.

This study divides the purchase amount into five parts and take its integer value.

“Third Quartile is 1796 dollar ~ The Maximum value is 3022 dollar”, “The Median value is 1276 dollar ~ Third Quartile is 1796 dollar”, “First Quartile is 979 dollar ~ The Median value is 1276 dollar”, “The minimum value is 319 dollar ~ First Quartile is 979 dollar” and “upper/down fence value is 3022 dollar/-246 dollar”.

$1796 < M \leq 3022$ dollar, $1276 < M \leq 1796$ dollar, $979 \leq M \leq 1276$ dollar, $319 \leq M \leq 979$ dollars, $M > 3022$ dollar or $M < -246$ dollars, and delete $M < -246$ dollars.

When $1796 < M \leq 3022$ dollars, the score of customer purchase amount is “5” points. $1276 < M \leq 1796$ dollars, the scores of the customer purchase amounts are “4” points. $979 \leq M \leq 1276$ dollar, the score of customer purchase amount is “3” points. $319 \leq M \leq 979$ dollars, the score of customer purchase amount is “2” points. $M > 3022$ dollars, the score of customer purchase amount is “1” points (Figure 17).

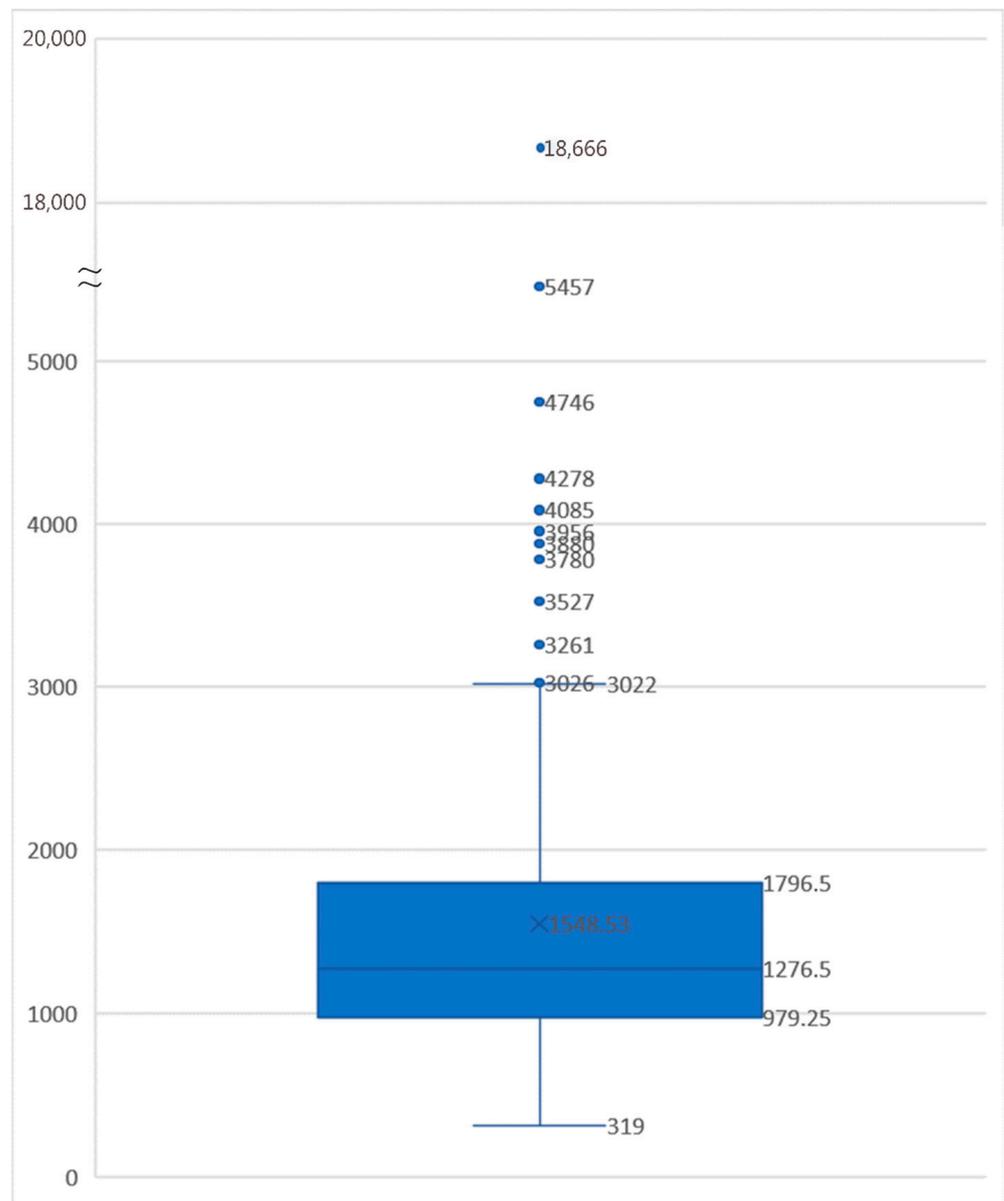


Figure 17. The box plot of purchase amount.

According to the score of Customers’ RFM compiled of this research. The result can be found. For example, the last purchase date of Customer number 20000154800 is 31 December 2013. The distance of is zero date. Therefore, the R score is “5”, purchase frequency is 69 times. F score is “2”, purchase amount is 4746 dollars, M score is “1” (Table 7).

Table 7. The scores of customers.

Score	Last Purchase Date R (day)	Purchase Frequency F (time)	Purchase Amount M (dollar)
5	R = 0	77 < F ≤ 87	1796 < M ≤ 3022
4	0 < R ≤ 1	74 < F ≤ 77	1276 < M ≤ 1796
3	1 < R ≤ 3	70 < F ≤ 74	979 < M ≤ 1276
2	3 < R ≤ 7	60 < F ≤ 70	319 < M ≤ 979
1	R > 7	F > 87 F < 59	M > 3022

The last purchase date of customer number 20000186600 is 31 December 2013. The distance of the zero date. R score is “5”, purchase frequency is 92 times. F score is “1”, purchase amount is 1094 dollars; M score is “3”. Multiply by weight value of F, R and M after the customer is coded. We can get the weighted R, F, and M values of customers (Table 8).

Table 8. The RFM scores of customers.

Member’s Card Number	Last Purchase Date	R	Purchase Frequency	F	Purchase Amount	M
20000154800	31 December 2013 (0)	5	69	2	4746	1
20000186600	31 December 2013 (0)	5	92	1	1094	3
20000344800	28 December 2013 (3)	3	75	4	2796	5
20000356000	28 December 2013 (3)	3	81	5	1181	3
20000361700	31 December 2013 (0)	5	76	4	3557	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21053653300	29 December 2013 (2)	3	76	4	2540	5
21053774000	31 December 2013 (0)	5	76	4	948	2
21053837700	29 December 2013 (2)	3	72	3	942	2
21055202600	21 December 2013 (10)	1	75	4	1103	3
21055261400	31 December 2013 (0)	5	72	3	1070	3

3.1.1.2. RFM Score of Product

First, divide the products of customers’ purchased into five categories. “Raw food”, “Instant Food”, “Dry Food”, “People’s Livelihood Products”, and “Others”. Then, find the last purchase date of products’ purchase frequency and purchase amount. The score of products is Table 9. For example: Last purchase date of five categories is also 31 December 2013. Therefore, the R score all is “5”, purchase frequency of five categories is 55,740, 16,775, 11,254, 20,120, and 13,114. Therefore, the score value is “5”, “3”, “1”, “4”, and “2”, average sales price of five categories is 69, 87, 80, 92, and 111. Therefore the score value is “1”, “3”, “2”, “4”, and “5”. RFM score of raw food is 5-5-1; RFM score of instant food is 5-3-3; RFM score of Dry Food is 5-1-2. RFM score of People’s Livelihood Products is 5-4-4. RFM score of others is 5-2-5. Multiply by weight the value of F, R, and M after the product is coded. We can get the weighted R, F, and M values of the product [11] (Table 9).

Table 9. RFM score of the product.

No.	Categories	The Last Purchase Date	R	Purchase Frequency	F	Purchase Amount	M Value
1	Raw food	31 December 2013 (0)	5	55,740	5	69	1
2	Instant Food	31 December 2013 (0)	5	16,775	3	87	3
3	Dry Food	31 December 2013 (0)	5	11,254	1	80	2
4	People’s Livelihood Products	31 December 2013 (0)	5	20,120	4	92	4
5	Other	31 December 2013 (0)	5	13,114	2	111	5

Further explore the characteristics of product sales, organize as in Table 10.

Observe the number and percentage of purchases by customers in each cluster based on product sales frequency. Purchase frequency in which customers buy raw food. Golden customer > Iron customer > Copper customer > Diamond customer > Silver customer. Purchase frequency in which customers buy instant Food: Iron customer > Golden customer > Silver customer > Iron customer > Diamond customer. Purchase frequency in which customers buy dry Food: Golden customer > Copper customer > Iron customer > Diamond customer > Silver customer. Purchase frequency in which customers buy People’s Livelihood Products: Golden customer > Copper customer > Iron customer > Diamond customer > Silver customer. Purchase frequency in which customers buy other

food: Golden customer > Iron customer > Copper customer > Iron customer > Diamond customer (Table 11).

Table 10. The characteristics of product sales.

No.	Categories	Description of RFM	Characteristic
1	Raw food	1. The last purchase date is 0 date. 2. Sales frequency is 55,740 times. First among the products 3. Average sales amount is 69 dollars. Ranked fifth among the products.	High activity, high loyalty, low contribution
2	Instant Food	1. The last purchase date is 0 date. 2. Sales frequency is 16,775 times. Ranked third among the products. 3. Average sales amount is 87 dollars. Ranked third among the products.	High activity, medium loyalty, medium contribution
3	Dry Food	1. The last purchase date is 0 date. 2. Sales frequency is 11,254 time. Ranked fifth among the products. 3. Average sales amount is 80 dollars. Ranked fourth among the products.	Low activity, low loyalty, low contribution
4	People's Livelihood Products	1. The last purchase date is 0 date. 2. Sales frequency is 20,120 time. Ranked second among the products. 3. Average sales amount is 92 dollars. Ranked second among the products.	High activity, high loyalty, high contribution
5	Other	1. The last purchase date is 0 date. 2. Sales frequency is 13,114 time. Ranked fourth among the products. 3. Average sales amount is 111 dollars. First among the products.	Low activity, low loyalty, high contribution

Table 11. Number and percentage of purchases made by customers in each cluster.

Categories	Diamond Customer	Golden Customer	Silver Customer	Copper Customer	Iron Customer	Total
Raw food	8298	16,011	7906	11,639	11,886	55,740
percentage	14.89%	28.72%	14.18%	20.88%	21.32%	
Instant Food	2624	4018	2799	4677	2657	16,775
percentage	15.64%	23.95%	16.69%	27.88%	15.84%	
Dry Food	1805	3204	1535	2635	2075	11,254
percentage	16.04%	28.47%	13.64%	23.41%	18.44%	
People's Livelihood Products	3290	5735	3030	4135	3930	20,120
percentage	16.35%	28.50%	15.06%	20.55%	19.53%	
Other	1830	3775	1913	2592	3004	13,114
percentage	13.95%	28.79%	14.59%	19.77%	22.91%	

3.1.2. The Result of Cluster

The study cluster the customers with Ward's Method, K-Means, Fuzzy C-Means, and Self-Organizing Maps. Finally, Discriminant analysis to check the result and verify which methods work best. Use the results to calculate customer lifetime value.

3.1.2.1. Ward's Method and K-Means

The study used SPSS 12.0 statistical software for the analysis. First, find the best quantity of cluster by Ward's Method. The coefficient increment to 101.277 from step 295 to step 296. Therefore, the best number of groups are five (Table 12).

Table 12. The agglomerative of Ward's Method.

Step	Groups		Coefficient	Increment	First Stage Cluster		Next Step
	Group 1	Group 2			Group 1	Group 2	
1	289	300	0.000		0	0	11
2	280	298	0.000	0	0	0	20
3	147	297	0.000	0	0	0	132
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
293	7	20	375.232	36.599	286	291	298
294	3	6	435.942	60.710	292	284	298
295	1	9	505.569	69.627	290	282	297
296	4	5	606.846	101.277	287	288	297
297	1	4	799.642	192.796	295	296	299
298	3	7	996.328	196.686	294	293	299
299	1	3	1282.703	286.375	297	298	0

This study found that the best number of groups are five based on the Ward method. The random designation of the center of K groups by K-Means. The center of initial clustering. Initial center of group 1 is 1-2-1. Initial center of group 2 is 5-1-2. The initial center of group 3 is 1-5-5. The initial center of group 4 is 5-5-4. The initial center of group 5 is 2-1-5 (Table 13).

Table 13. Initial center of groups by K-Means.

	Group 1	Group 2	Group 3	Group 4	Group 5
R	1	5	1	5	2
F	2	1	5	5	1
M	1	2	5	4	5

After iterations of five times can get the final grouping. Group 1 has 42 observations. Group 2 has 78 observations. Group 3 has 45 observations. Group 4 has 94 observations. Group 5 has 41 observations (Table 14).

Table 14. The number of observations by K-Means.

Group 1	42
Group 2	78
Group 3	45
Group 4	94
Group 5	41

Final center of group 1 3-3-2. Final center of group 2 is 5-2-3. Final center of group 3 is 2-5-3. Final center of group 4 is 5-4-4. Final center of group 5 is 2-3-5, and can name the group according to the central point nature of each group (Table 15).

Table 15. The final center of group by K-Means.

	Group 1	Group 2	Group 3	Group 4	Group 5
R	3	5	2	5	2
F	3	2	5	4	3
M	2	3	3	4	5

Group 300 customers into the group. The customer of No. 1 is group 2. The customer of No. 2 is group 2. The customer of No. 3 is group 5 and so on (Table 16).

Table 16. Each group member by K-Means.

No	Group	No	Group	No	Group	No	Group	No	Group
1	2	61	1	121	4	181	1	241	3
2	2	62	4	122	4	182	4	242	2
3	5	63	2	123	4	183	5	243	2
4	3	64	2	124	2	184	1	244	1
5	2	65	5	125	4	185	3	245	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	4	116	5	176	2	236	4	296	5
57	4	117	4	177	4	237	2	297	4
58	4	118	2	178	4	238	3	298	1
59	2	119	4	179	3	239	4	299	3
60	1	120	4	180	5	240	4	300	2

3.1.2.2. Fuzzy C-Means

This study used MATLAB R2018a (MathWorks, Natick, MA, USA) statistical software for analysis. First, setup the number of group and compare the effects of the three grouping methods. Experiment with five groups as the number of groups. Matrix U of $c \times n$. According to two conditions, these values are normalized between 0 and 1. For one point of n_j and j is equal 1~300. The total value is 1 (Table 17).

Table 17. Matrix U.

c	N					
	1	2	...	298	299	300
1	0.09178	0.07773	...	0.26066	0.40824	0.07227
2	0.08099	0.11212	...	0.17728	0.32350	0.06123
3	0.10016	0.18789	...	0.08854	0.08149	0.21501
4	0.56231	0.50994	...	0.31310	0.08974	0.40938
5	0.16474	0.11229	...	0.16040	0.09700	0.24208
Total	1	1	...	1	1	1

Customer center point of cluster 1 is (2.74587, 4.38981, and 2.99331). Customer center point of group 2 is (2.29605, 2.80328, and 3.96252). Customer center point of group 3 is (4.61544, 3.12470, and 4.48114). Customer center point of group 4 is (4.37029, 2.32204, and 2.38201). Customer center point of group 5 is (4.63284, 4.34444, and 2.61103) (Table 18).

Table 18. Customer center point of the group by Fuzzy C-Means.

	R	F	M
Group 1	2.74587	4.38981	2.99331
Group 2	2.29605	2.80328	3.96252
Group 3	4.61544	3.12470	4.48114
Group 4	4.37029	2.32204	2.38201
Group 5	4.63284	4.34444	2.61103

After iterations of 93 times. The objective function is reduced from 333.26600 to 224.47860 (Table 19).

Group 1 has 50 observations. Group 2 has 59 observations. Group 3 has 73 observations. Group 4 has 66 observations. Group 5 has 52 observations (Table 20).

Group 300 customers into a group. The customer of No. 1 is group 4. The customer of No. 2 is group 4. The customer of No. 3 is group 4 and so on (Table 21).

Table 19. After iterations of the objective function.

Iterations	Objective Function	Iterations	Objective Function	Iterations	Objective Function
1	333.26600	32	226.98985	63	224.62937
2	256.34525	33	226.96481	64	224.58932
3	256.18766	34	226.94051	65	224.55953
4	255.90851	35	226.91650	66	224.53764
5	255.42615	36	226.89243	67	224.52167
⋮	⋮	⋮	⋮	⋮	⋮
27	227.15133	58	225.06781	89	224.47866
28	227.11039	59	224.94304	90	224.47864
29	227.07531	60	224.83734	91	224.47862
30	227.04431	61	224.75089	92	224.47861
31	227.01612	62	224.68233	93	224.47860

Table 20. The number of observations by Fuzzy C-Means.

Group 1	50
Group 2	59
Group 3	73
Group 4	66
Group 5	52

Table 21. Each group member by Fuzzy C-Means.

No	Group	No	Group	No	Group	No	Group	No	Group
1	4	61	4	121	3	181	4	241	2
2	4	62	5	122	5	182	3	242	4
3	2	63	4	123	3	183	3	243	4
4	1	64	3	124	3	184	2	244	4
5	5	65	2	125	3	185	1	245	5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	5	116	2	176	4	236	3	296	2
57	3	117	3	177	3	237	3	297	5
58	3	118	3	178	3	238	1	298	4
59	4	119	3	179	1	239	5	299	1
60	2	120	3	180	2	240	5	300	4

3.1.2.3. Self-Organizing Maps

This study used MATLAB R2018a statistical software for analysis. First, setup the number of groups and compare the effects of the three grouping methods (Table 22).

Table 22. Setup of the parameter.

Parameter	Setup	Remark
Dimensions of map	[51]	Setup the map of weight [51] is 5 groups
Topology function	HEXTOP	Hexagon pattern (Preset)
Distance function	LINKDIST	Linkage function (Preset)
Ordering phase learning rate (OLR)	0.9	Preset
Ordering phase steps	1000	Preset
Tuning phase learning rate (TLR)	0.02	Preset
Neighborhood distance	1.0	Preset

Customer center point of cluster 1 is (4.6848, 3.4061, and 2.2788). Customer center point of group 2 is (4.6410, 3.4060, and 3.2820). Customer center point of group 3 is

(4.1958, 3.2208, and 4.2125). Customer center point of group 4 is 43.0633, 3.2616, and 3.8987. Customer center point of group 5 is (2.3795, 3.5436, and 3.0820) (Table 23).

Table 23. Customer center point of group by self-organizing maps.

	R	F	M
Group 1	4.6848	3.4061	2.2788
Group 2	4.6410	3.4060	3.2820
Group 3	4.1958	3.2208	4.2125
Group 4	3.0633	3.2616	3.8987
Group 5	2.3795	3.5436	3.0820

Group 1 has 55 observations. Group 2 has 55 observations. Group 3 has 69 observations. Group 4 has 47 observations. Group 5 has 74 observations (Table 24).

Table 24. The number of observations by self-organizing maps.

Group 1	55
Group 2	55
Group 3	69
Group 4	47
Group 5	74

Group 300 customers into a group. The customer of No. 1 is group 1. The customer of No. 2 is group 2. The customer of No. 3 is group 4 and so on (Table 25.)

Table 25. Each group member by self-organizing maps.

No	Group	No	Group	No	Group	No	Group	No	Group
1	1	61	1	121	3	181	5	241	5
2	2	62	2	122	2	182	3	242	1
3	4	63	2	123	3	183	3	243	1
4	5	64	3	124	3	184	5	244	5
5	1	65	4	125	2	185	5	245	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
56	1	116	4	176	1	236	3	296	4
57	3	117	3	177	3	237	3	297	1
58	3	118	3	178	2	238	5	298	5
59	1	119	3	179	5	239	1	299	5
60	4	120	3	180	5	240	2	300	2

3.1.2.4. Comparison of Accuracy by the Discriminant Analysis

The study used SPSS 12.0 statistical software (SPSS Inc., Chicago, IL, USA) for the analysis. According to the grouping results of the three methods by discriminant analysis. The correct rates are 96.7%, 98.3%, and 90.0%. Thus, the study chooses Fuzzy C-Means (Table 26).

Table 26. The three methods by discriminant analysis.

Method	Ward's Method and K-Means	Fuzzy C-Means	Self-Organizing Maps
Correct rate	96.7%	98.3%	90.0%
Prior choice	2	1	3

3.1.3. Customer Lifetime Value

According to the results of grouping by Fuzzy C-Means, calculate the customer lifetime value and sort from cluster 1 to cluster 5. Maximum is 45 points and minimum is nine

points. For example, the customer number is 20000356000 that belongs to group 1, and the RFM value is 3-5-3. Be multiplied by the weight value of F (high), R (medium) and M (low). The study set the weight value of F is 5, R is 3, and M is 1 [11]. The customer-weighted RFM value is 9-25-3. Finally, the weighted RFM values are summed up. The score of the customer lifetime value is 37 (Table 27).

Table 27. Customer lifetime value by Fuzzy C-Means.

No	Member's Card Number	Group	R	F	M	R*3	F*5	M*1	Customer Lifetime Value
4	20000356000	1	3	5	3	9	25	3	37
8	20000439600	1	2	5	2	6	25	2	33
9	20000440800	1	1	3	1	3	15	1	19
10	20000458700	1	3	5	4	9	25	4	38
14	20000879700	1	3	5	4	9	25	4	38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
261	21024048600	5	4	4	2	12	20	2	34
278	21040954000	5	5	5	2	15	25	2	42
287	21045079700	5	4	5	2	12	25	2	39
295	21053652800	5	5	5	4	15	25	4	44
297	21053774000	5	5	4	2	15	20	2	37

Sorting the score of customer life values. Name the groups in descending order as Diamond, Golden, Silver, Copper, and Iron (Table 28).

Table 28. Name the groups in descending order.

Group	R	F	Average M	CLV	Name
5	4.7	4.6	2.6	39.7	Diamond
3	4.6	3.2	4.5	34.4	Golden
1	2.4	4.5	2.9	32.8	Silver
4	4.3	2.4	2.3	27.0	Copper
2	2.3	2.7	4.0	24.2	Iron

In order to further explore the characteristics of each cluster customer (Table 29).

Table 29. Characteristics of the each cluster customer.

No	Categories	Description of RFM	Characteristic
5	Diamond	1. The last purchase date ≤ 1 day 2. Purchase > 74 times and ≤ 77 times 3. Purchase amount ≥ 319 dollar and ≤ 979 dollar	High activity, high loyalty, low contribution
3	Golden	1. The last purchase date ≤ 1 day 2. Purchase > 70 times and ≤ 74 times 3. Purchase amount ≥ 1276 dollar and ≤ 1796 dollar	High activity, medium loyalty, High contribution
1	Silver	1. The last purchase date < 3 days and ≤ 7 days 2. Purchase > 74 times and ≤ 77 times 3. Purchase amount ≥ 319 dollar and ≤ 979 dollar	Low activity, High loyalty, Low contribution
4	Copper	1. The last purchase date ≤ 1 day 2. Purchase > 60 times and ≤ 70 times 3. Purchase amount ≥ 319 dollar and ≤ 979 dollar	High activity, Low loyalty, Low contribution
2	Iron	1. The last purchase date < 3 days and ≤ 7 days 2. Purchase > 60 times and ≤ 70 times 3. Purchase amount ≥ 1276 dollar and ≤ 1796 dollar	Low activity, Low loyalty, High contribution

4. Discussion

First, the study transforms customer's last purchase date, purchase sequence, and purchase amount to values of R, F, and M, by box plot. Second, find the best quantity of groups are five by Ward's Method. Third, cluster the customers by K-means, Fuzzy C-Means, and Self-Organizing Maps. Finally, validation results by discriminant analysis.

4.1. Research Results

Three research methodologies were applied in this study to classify customers, ultimately discriminant analysis was employed to see which clustering method was most effective, and to select results from which to further calculate customer lifetime value.

4.2. Ward's Method with K-Means

Three discriminant functions were extracted with this method, all of which are shown in Table 30. Since 58.1% of the information can be explained by the first discriminant function. 34.8% of the information can be explained by the second discriminant function, while 7.1% of the information can be explained by the third discriminant function. The greater the eigenvalue, the stronger the discriminative power. The eigenvalue of the first discriminant function is 3.544, whose discriminative power is the strongest of the three.

Table 30. Eigenvalues of discriminant functions of Ward's Method with K-Means.

Function	Eigenvalue	Explanatory Variance	Cumulative Explained Variance	Canonical Correlation Coefficient
1 (R)	3.544	58.1	58.1	0.883
2 (F)	2.123	34.8	92.9	0.824
3 (M)	0.430	7.1	100.0	0.548

Within the significance of discriminant function test as shown in Table 31, the Wilks' Lambda value is the ratio of the Within Groups Sum of Squares to the Total Sum of Squares (TSS), of which Within Groups Sum of Squares, also known as the residual Sum of Squares (RSS). The smaller the Wilks' Lambda value, the better its discrimination; the Wilks' Lambda value of the first discriminant function is 0.049, whose discriminant power is the greatest of the three. As shown in Table 31, the significance of the discriminant for all three was 0.000, indicating a relatively high significance in all of them.

Table 31. Significance testing on the discriminant function based on the Ward's method with K-Means.

Function	Wilks' Lambda Value	Chi-Square Value	Degrees of Freedom	Significance
1 (R)	0.049	887.934	12	0.000
2 (F)	0.224	441.393	6	0.000
3 (M)	0.699	105.478	2	0.000

Fisher's linear discriminant is shown in Table 32 and was separated into five clusters for this research, each of which corresponds to one discriminant function and are thus written as five separate discriminant functions, and they are, respectively, Formulas (31)–(35).

$$d_1 = -23.069 + 6.660 \times R + 6.551 \times F + 3.474 \times M \quad (31)$$

$$d_2 = -43.953 + 12.512 \times R + 5.641 \times F + 4.102 \times M \quad (32)$$

$$d_3 = -44.032 + 5.374 \times R + 11.683 \times F + 5.402 \times M \quad (33)$$

$$d_4 = -60.384 + 11.817 \times R + 10.168 \times F + 5.565 \times M \quad (34)$$

$$d_5 = -32.879 + 6.045 \times R + 7.422 \times F + 6.127 \times M \quad (35)$$

Table 32. Fisher's linear discriminant for the Ward's method with K-Means.

	Group				
	1	2	3	4	5
R	6.660	12.512	5.374	11.817	6.045
F	6.551	5.641	11.683	10.168	7.422
M	3.474	4.102	5.402	5.565	6.127
Constant	-23.069	-43.953	-44.032	-60.384	-32.879

It is apparent from the discriminant analysis as shown in Table 33 that there are 42 customers in Cluster 1, of which 41 were classified correctly, while 1 was misclassified for an accuracy of 97.6%. There are 78 customers in Cluster 2, all of which were classified correctly, for an accuracy of 100%. There are 45 customers in Cluster 3, all of which were classified correctly, for an accuracy of 100%. Ninety-four customers in Cluster 4, all of which were classified correctly, for an accuracy of 100%. Forty-one customers in Cluster 5, of which 32 were classified correctly and 9 were misclassified, for an accuracy of 78%. There is a total of 300 customers, of which 290 were classified correctly, for an accuracy of 96.7%.

Table 33. Discriminant analysis results based on the Ward's method with K-Means.

Cluster	Predicted Group Members					Total
	1	2	3	4	5	
Items	1	41	1	0	0	42
	2	0	78	0	0	78
	3	0	0	45	0	45
	4	0	0	0	94	94
	5	0	2	5	2	32
Percentage	1	97.6	2.4	0.0	0.0	100.0
	2	0.0	100.0	0.0	0.0	100.0
	3	0.0	0.0	100.0	0.0	100.0
	4	0.0	0.0	0.0	100.0	100.0
	5	0.0	4.9	12.2	4.9	78.0

4.3. Fuzzy C-Means

Three discriminant functions could be extracted with this method, the eigenvalues of the discriminant function are shown in Table 34. The first discriminant function can explain 49.6% of the information. The second discriminant function can explain 34.9% of the information. The third discriminant function can explain 15.6% of the information. The larger the eigenvalue, the stronger the discriminant power. The eigenvalue of the first discriminant function is 2.929, which has the strongest discriminant power among the three.

Table 34. Discriminant function eigenvalues of Fuzzy C-Means.

Function	Eigenvalue	Explained Variance	Cumulative Explained Variance	Canonical Correlation Coefficient
1 (R)	2.929	49.6	49.6	0.863
2 (F)	2.060	34.9	84.4	0.820
3 (M)	0.920	15.6	100.0	0.692

The significance test of discriminant function is shown in Table 35. The Wilks' Lambda value is the ratio of the Within-Groups Sum of Square to the Total Sum of Squares, in which the Within Groups Sum of Squares is also known as the Residual Sum of Squares (RSS). The smaller the Wilks' Lambda value, the greater the discriminant power of its discriminant function. The first discriminant function's Wilks' Lambda value is 0.043, which has the strongest discriminant power among the three. Table 35 shows that the significance of

the three discriminant functions are all 0.000, thus indicating that the three discriminant functions all have relatively high significance.

Table 35. Significance testing on the discriminant function for the Fuzzy C-Means method.

Function	Wilks' Lambda Value	Chi-Square Value	Degrees of Freedom	Significance
1 (R)	0.043	926.114	12	0.000
2 (F)	0.170	522.436	6	0.000
3 (M)	0.521	192.506	2	0.000

Fisher's linear discriminant, as shown in Table 36, was separated for this study separated into five clusters, each of which corresponds to a discriminant function and are thus written as five discriminant functions, each of which has been separated into Formulas (36)–(40).

$$d_1 = -31.259 + 5.891 \times R + 8.172 \times F + 2.850 \times M \quad (36)$$

$$d_2 = -23.738 + 5.309 \times R + 3.972 \times F + 5.407 \times M \quad (37)$$

$$d_3 = -48.175 + 10.949 \times R + 5.081 \times F + 5.803 \times M \quad (38)$$

$$d_4 = -31.546 + 10.304 \times R + 4.260 \times F + 2.428 \times M \quad (39)$$

$$d_5 = -51.111 + 11.554 \times R + 8.553 \times F + 2.017 \times M \quad (40)$$

Table 36. Fisher's linear discriminant functions based on the Fuzzy C-Means method.

	Cluster				
	1	2	3	4	5
R	5.891	5.309	10.949	10.304	11.554
F	8.172	3.972	5.081	4.260	8.553
M	2.850	5.407	5.803	2.428	2.017
Constant	-31.259	-23.738	-48.175	-31.546	-51.111

The discriminant analysis as shown in Table 37 shows that there are 50 customers in Cluster 1, all of which were classified correctly, for an accuracy of 100%. There are 59 customers in Cluster 2, of which 55 were classified correctly. 4 were misclassified, for an accuracy of 93.2%. There are 73 customers in Cluster 3, all of which were classified correctly, for an accuracy of 100%. There are 66 customers in Cluster 4, of which 65 are classified correctly, and 1 was misclassified, for an accuracy of 98.5%. There are 52 customers in Cluster 5, all of which were classified correctly, for an accuracy of 100%. There is a total of 300 customers, of which 295 were classified correctly, for an accuracy of 98.3%.

Table 37. Discriminant analysis based on the Fuzzy C-Means method.

Cluster	Predicted Group Members					Total
	1	2	3	4	5	
Item	1	50	0	0	0	50
	2	4	55	0	0	59
	3	0	0	73	0	73
	4	0	0	1	65	66
	5	0	0	0	0	52
Percentage	1	100.0	0.0	0.0	0.0	100.0
	2	6.8	93.2	0.0	0.0	100.0
	3	0.0	0.0	100.0	0.0	100.0
	4	0.0	0.0	1.5	98.5	100.0
	5	0.0	0.0	0.0	0.0	100.0

4.4. Self-Organizing Map

Three discriminant functions could be extracted with this method, and the discriminant function eigenvalues are shown in Table 38. The first discriminant function can explain 56% of the information. The second discriminant function can explain 43.5% of the information. The third discriminant function can explain 0.5% of the information. The greater the eigenvalue, the stronger the discriminant power. The eigenvalue of the first discriminant function is 3.929, which has the strongest discriminant power among the three.

Table 38. Discriminant function eigenvalues of the self-organizing map.

Function	Eigenvalue	Explained Variance	Cumulative Explained Variance	Canonical Correlation Coefficient
1 (R)	3.929	56.0	56.0	0.893
2 (F)	3.052	43.5	99.5	0.868
3 (M)	0.037	0.5	100.0	0.188

The significance testing of discriminant function is as shown in Table 39. The Wilks' Lambda value is the ratio of the Within-Groups Sum of Square to the Total Sum of Square, which is also known as the Residual Sum of Squares (RSS). The smaller the Wilks' Lambda value, the better the discriminant function. The first discriminant function's Wilks' Lambda value is 0.048, whose discriminant power is strongest among the three. Table 39 shows that the significance of two discriminant functions is 0.000, indicating that the significance of these two discriminant functions is higher than the third one.

Table 39. Significance testing of the discriminant function of the self-organizing map.

Function	Wilks' Lambda Value	Chi-Square Value	Degrees of Freedom	Significance
1 (R)	0.048	893.917	12	0.000
2 (F)	0.238	423.350	6	0.000
3 (M)	0.965	10.579	2	0.005

Fisher's linear discriminant, as shown in Table 40, was separated for this research into five clusters, with each cluster corresponding to one discriminant function, and are thus written as five discriminant functions; they are, respectively, Formulas (41)–(45).

$$d_1 = -42.484 + 13.999 \times R + 1.452 \times F + 6.138 \times M \quad (41)$$

$$d_2 = -53.750 + 14.291 \times R + 1.266 \times F + 10.327 \times M \quad (42)$$

$$d_3 = -68.786 + 14.742 \times R + 0.412 \times F + 14.402 \times M \quad (43)$$

$$d_4 = -42.560 + 8.613 \times R + 0.649 \times F + 13.244 \times M \quad (44)$$

$$d_5 = -23.854 + 7.332 \times R + 1.858 \times F + 7.680 \times M \quad (45)$$

Table 40. Fisher's linear discriminant function of the self-organizing map.

	Cluster				
	1	2	3	4	5
R	13.999	14.291	14.742	8.613	7.332
F	1.452	1.266	0.412	0.649	1.858
M	6.138	10.327	14.402	13.244	7.680
Constant	-42.484	-53.750	-68.786	-42.560	-23.854

The discriminant analysis in Table 41 shows that there are 55 customers in Cluster 1, of which 53 were classified correctly, 2 were misclassified, for an accuracy of 96.4%. There are 55 customers in Cluster 2, of which 49 were classified correctly, and 6 were misclassified, for an accuracy of 89.1%. There are 69 customers in Cluster 3, of which 61 were classified correctly and 8 were misclassified, for an accuracy of 88.4%. There were 47 customers in Cluster 4, of which 43 were classified correctly, and four were misclassified, for an accuracy of 91.5%. There are 74 customers in Cluster 5, of which 64 were classified correctly, and 10 were misclassified, for an accuracy of 86.5percent. There is a total of 300 customers, 270 of which were classified correctly, for an accuracy of 90%.

Table 41. Discriminant analysis of the self-organizing map.

Cluster	Predicted Group Members					Total	
	1	2	3	4	5		
Item	1	53	0	0	0	2	55
	2	0	49	6	0	0	55
	3	0	8	61	0	0	69
	4	0	0	0	43	4	47
	5	0	0	0	10	64	74
Percentage	1	96.4	0.0	0.0	0.0	3.6	100.0
	2	0.0	89.1	10.9	0.0	0.0	100.0
	3	0.0	11.6	88.4	0.0	0.0	100.0
	4	0.0	0.0	0.0	91.5	8.5	100.0
	5	0.0	0.0	0.0	13.5	86.5	100.0

According to the grouping results of the three methods by discriminant analysis, the correct rates are 96.7%, 98.3%, and 90.0%. Thus, the study chose Fuzzy C-Means to calculate the score of customer life value. This research divides the customers and name into five types: Diamond, Golden, Silver, Copper, and Iron. In terms of customers, we provide the following strategies in Table 42.

Table 42. Strategies of this study.

No	Categories	Description of RFM
5	Diamond	Increase the customer's purchase amount. 1. Place more preferential products behind the POS and facilitate customers to purchase. 2. Observe the products purchased by customers and recommend additional products to them. 3. Customers can get a discount if buy product of A and B at the same time.
3	Golden	Increase the customer's purchase sequence. 1. Bonus points are given directly. 2. Customers can get gifts or coupons when they spend a specific amount in a single time.
1	Silver	Increase the customer's last purchase date and purchase amount. 1. Regular and special day promotions. 2. Buy one get one promotion.
4	Copper	Increase the customer's purchase sequence and purchase amount. Regularly send product catalogs, coupons and event newsletter notification.
2	Iron	Increase the customer's last purchase date and purchase sequence. 1. Send the recommended top ten best-selling catalogs of product regular. 2. Send product's trial package. For example: Shampoo, facial mask, and so on

5. Conclusions

Ref. [72] Customer lifetime value (CLV) measurement is challenging, as it requires forecasting customers' future purchases [73]. The same treatment of all customers will cause customers who are not so valuable will be value destroyer rather than a value creator. Giving discounts and promos to all customers has not provided benefits to a company in retaining customers. Thus, Ref. [74] leading multinationals such as Alibaba, Lenovo, LG

Corporation, Mitsubishi, Nestlé, Tata Motors, and Tesco use customer value strategies to build rational and emotional bonds with their target markets.

The author of this paper works in the third largest supermarket in Taiwan and is responsible for marketing analysis work and in charge of a database with more than one million members. However, the author has always concerned about how to allocate supermarket marketing resources; therefore, the author enters the academic field to look for a solution. After substantive survey and literature review work, the author found that most of the academic community uses a single method to analyze large databases, such as the RFM method; however, in practice, this approach cannot fully identify the correct customers. Therefore, the author sorted out all the methods presented in studies one by one, conducted data analysis, and completed this paper, hoping to share with other scholars and inspire other research results.

This study uses Taiwan's third largest supermarket member database for analysis. Since it is easy to acquire data, there is no copyright issue, and the key is to share technology. In the future, if there are other databases of different nature to be combined in the research, then it could further prove the feasibility.

The recent food security scandals in Taiwan have caused substantial management difficulties for the retail industry. Subsequently, the optimal response to these challenges largely focused on the affluent top-level customers. Therefore, customer value analysis has become increasingly crucial. Valuable customer clusters were identified on the basis of the detailed value characteristics of all clusters. Comparative analysis and verification were conducted with regard to the life value of all customer clusters.

5.1. In Terms of Customers, Provide the Following Strategies

1. **Diamond-level customers:** Diamond-level customers account for 17.3% of the total members. The characteristics of this customer group are high degree of activeness, high customer loyalty, low customer contribution, and they are the customers with the highest lifetime value. In order to increase the purchase amount of this customer group, the supermarket can put discounted products at the back of the cashier, attracting customers to add to their purchase; observe products selected by customers and recommend additional products to them; hold some promotion events, such as mark-down for buying red-label and green-label products together.
2. **Gold-level customers:** Gold-level customers account for 24.3% of the total members. The characteristics of this customer group are high degree of activeness, moderate customer loyalty, high customer contribution, and they are the customers with the second highest lifetime value. In order to increase the purchase frequency of this customer group, the supermarket can combine amount spent by customers directly into the bonus points, which can be used for mark-down at the next visit; give gifts or coupons to a single purchase of a certain amount.
3. **Silver-level customers:** Silver-level customers account for 16.7% of the total members. The characteristics of this customer group are low degree of activeness, high customer loyalty, low customer contribution, and they are the customers with average lifetime value. In order to improve the last purchase date and purchase amount of this customer group, the supermarket can hold regular events such as special holiday promotions, buy one get one for free products, weekday benefits, etc.
4. **Copper-level customers:** Copper-level customers account for 22.0% of the total members. The characteristics of this customer group are high degree of activeness, low customer loyalty, low customer contribution, and they are the customers with below-average lifetime value. To increase the frequency and purchase amount of this customer group, the supermarket can send product catalogs, coupons and event newsletters on a regular basis to update customers about the latest of the supermarket.
5. **Iron-level customers:** Iron-level customers account for 19.7% of the total members. The characteristics of this customer group are low degree of activeness, low customer loyalty, high customer contribution, and they are the customers with the lowest

lifetime value. In order to improve the latest purchase date and frequency of this customer group, the supermarket can send catalogs featuring Top Ten Popular Products Recommended by Netizens, so as to attract customers to revisit; send product sample kits, such as: shampoo, mask, etc., in order to rekindle customers' interest in supermarket products.

5.2. In Terms of Products, Provide the Following Strategies

According to Association Rules. The shopping habit of the customer is Raw food and People's Livelihood Products, Raw food and Instant Food, Instant Food, and People's Livelihood Products. Therefore, we put Raw food and People's together, Raw food and Instant Food together, and Instant Food and People's Livelihood Products together.

The product can be tied for sale, gift or promotion. In term of products, product bundling can be carried out, or promotional gift. For example: If the customer bought the Raw food, People's Livelihood Products is free, and so on. This can increase supermarket profits.

5.3. Research Limitations

Due to the limitations of this study, information of certain aspects cannot be obtained, which are described as follows:

1. The database is based on a supermarket in Taiwan, so it is uncertain that whether other supermarkets in Taiwan also conform to the results of this study.
2. The database consists of members of the supermarket, so it cannot reveal the transaction status of nonmembers.
3. The database is not a questionnaire, so it cannot reveal customer satisfaction with products, services, environment, etc.
4. The amount information recorded in the database does not specify whether it is the promotional price, so it cannot tell if the amount is discounted.
5. Regarding research limitations, because many members use the same membership card for shopping, resulting in distortion in analysis, it is necessary to find solutions in the future.
6. This study only analyzes and discusses the data of a supermarket. In the future, it is recommended to combine customer transaction data of other supermarkets into a large database for study; it is also suggested to compare the results of foreign supermarkets, develop, and examine hypotheses, so as to conduct a variance analysis to discover if the effect is significant.
7. In terms of subject matter, scholars can study other retail entities, such as department stores, convenience stores, etc., or other industries, such as finance and insurance, leisure, and entertainment, etc.
8. In terms of customer segmentation, other clustering methods can be used, such as the Expectation Maximization method (EM), the Density-based Spatial Clustering of Applications with Noise algorithm (DBSCAN), the Mean Shift (MS), and so forth, and then analyze their differences to find the clustering method with the highest accuracy.

Author Contributions: Supervision, formal analysis: R.-H.L.; methodology, funding acquisition: W.-W.C.; conceptualization, project administration: C.-L.C.; writing—original draft, validation: W.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Matei, S.M.; Bruno, R.J. Pareto's 80/20 law and social differentiation: A social entropy perspective. *Public Relat. Rev.* **2015**, *178*–186. [[CrossRef](#)]
2. Fader, P.S.; Hardie, B.G.S.; Lee, K.L. Counting your customers the easy way: An alternative to the Pareto/NBD model. *J. Mark. Sci.* **2005**, *24*, 275–284. [[CrossRef](#)]

3. Reinartz, W.; Dellaert, B.; Krafft, M.; Kumar, V.; Varadarajan, R. Retailing innovations in a globalizing retail market environment. *J. Retail* **2011**, *87*, S53–S66. [CrossRef]
4. Zimmerman, M.M. *The Super Market*; Mass Distribution Publication Inc.: New York, NY, USA, 1955; p. 18.
5. Environmental Protection Administration, Executive Yuan. *Announcement of the Environmental Protection Agency of the Executive Yuan*; Environmental Protection Administration, Executive Yuan: Taipei, Taiwan, 2018. Available online: <https://oaout.epa.gov.tw/law/> (accessed on 10 March 2016).
6. Ministry of Economic Affairs, R.O.C. *Retrieval System for Company Bank Number and Limited Partner Business Item Code List*; Ministry of Economic Affairs: Taipei, Taiwan, 2018. Available online: <https://www.moea.gov.tw/Mns/populace/home/Home.aspx> (accessed on 10 March 2016).
7. Lin, Q. Customer Distinction Analysis of Department Stores-Taking Hsinchu City as an Example. Master's Thesis, Chung Hwa University, Hsinchu, Taiwan, 2002.
8. Kahan, R. Using Database Marketing Techniques to Enhance Your One-to-One Marketing Initiatives. *J. Consum. Mark.* **1998**, *15*, 491–493. [CrossRef]
9. Xu, Y.; Xu, W.; Li, S. Research and Application Practice of Using RFM Model to Analyze Customer Consumption Behavior and Contribution-Taking CNPC Membership Card Customers as an Example. *Pet. Q.* **2011**, *47*, 83–100.
10. Hughes, A.M. *Strategic Database Marketing*; Probus Publishing: Chicago, IL, USA, 1994; pp. 487–499.
11. Stone, B. *Successful Direct Marketing Methods*; NTC Business Books: Lincolnwood, IL, USA, 1995; pp. 37–59.
12. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]
13. Xiao, W. Research and Discussion on Grouping of Informal Employees. Master's Thesis, National Central University, Taoyuan, Taiwan, 2005.
14. Huang, T. Research on the Impact of the Establishment of Large-Scale Retail Stores on the Operation Strategies of Surrounding Businesses. Master's Thesis, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, 2013.
15. Li, M. Research on Consumer Behavior in Cross-Industry Financial Wealth Management. Master's Thesis, Taichung University of Science and Technology, Taichung, Taiwan, 2010.
16. Lin, G.; Tan, X.; Lin, M. Research on Tourism Consumption Behavior-Taking Hualien County's Five-star Tourist Hotel as an Example. *Bus. Mod. Chem.* **2012**, *6*, 211–242.
17. Rakhlin, A.; Caponnetto, A. *Stability of K-Means Clustering, Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2007; pp. 216–222.
18. Lin, Y. Face Recognition Based on Ant Optimization and K-means Method. Master's Thesis, Chung Yuan Christian University, Taoyuan, Taiwan, 2013.
19. Zhang, K. Using K-means and Bayesian Network to Detect Game Plug-Ins. Master's Thesis, National Taipei University, Taipei, Taiwan, 2015.
20. Shi, Y. Using K-Means Cluster Analysis to Explore the Influencing Factors of Mathematics Learning Achievement in Primary Schools. Master's Thesis, Chung Hwa University, Hsinchu, Taiwan, 2011.
21. Lin, Y. Research on Finding Key Item Purchase Sequences to Identify VIP Customers. Master's Thesis, National Central University, Taoyuan, Taiwan, 2016.
22. Huang, Y. Using Statistics and Data Exploration to Analyze Customer Purchase Behavior. Master's Thesis, Evergreen University, Tainan, Taiwan, 2013.
23. Huang, Q. Application of Data Exploration Technology in the Study of Customer Value and Consumer Behavior-Taking a Direct Selling Company as an Example. Master's Thesis, Chaoyang University of Science and Technology, Taichung, Taiwan, 2015.
24. Milligan, G.W.; Cooper, M.C. Methodology Review: Clustering Methods. *Appl. Psychol. Meas.* **1987**, *11*, 329–354. [CrossRef]
25. Sharma, S. *Applied Multivariate Techniques*; John Wiley & Sons: New York, NY, USA, 1996; pp. 232–233.
26. Yang, Z. A RFID-Based Positioning Mechanism. Master's Thesis, National Chiao Tung University, Hsinchu, Taiwan, 2006.
27. Dunn, J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *J. Cybern.* **1973**, *3*, 32–57. [CrossRef]
28. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Prentice Hall: New York, NY, USA, 1981; pp. 203–239.
29. Zhang, D. Application of Fuzzy Clustering Algorithm to the Determination of the Pollution Range of Kaohsiung Sea Area. Master's Thesis, National Sun Yat-sen University, Kaohsiung, Taiwan, 2002.
30. Yong, Y.; Chongxun, Z.; Pan, L. A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding. *Meas. Sci. Rev.* **2004**, *4*, 11–19.
31. Liu, Z. Effective fuzzy C-average Algorithm and Its Application in Cluster Analysis. Master's Thesis, Jianxing University of Science and Technology, Taoyuan, Taiwan, 2016.
32. Hong, C. Research on the Application of Fuzzy Grouping Method to Estimate Road Speed. Master's Thesis, National Chiao Tung University, Hsinchu, Taiwan, 2013.
33. Li, J. Research on Improved FCM Classification and Classification Optimization-Taking B&B Evaluation as an Example. Master's Thesis, National Dong Hwa University, Hualien, Taiwan, 2012.
34. Lin, Z. Establishing a Fuzzy Clustering Model Suitable for Customer Relationship Management-Taking Automobile Maintenance Service as an Example. Master's Thesis, National Taiwan University, Taipei, Taiwan, 2007.
35. Kohonen, T. The Self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. [CrossRef]

36. Lin, W. Research on the Performance of Self-Organizing Map Networks in Unsupervised Network Clusters. Master's Thesis, Shute University of Science and Technology, Kaohsiung, Taiwan, 2008.
37. Chen, Y. Research on the Interaction of Surface Water and Groundwater in Zhuoshuixi Basin Using Self-Organized Feature Mapping Network. Master's Thesis, Tamkang University, Taipei, Taiwan, 2013.
38. Sung, H.H.; Sang, C.P. Application of data mining tools to hotel data mart on the Internet for database marketing. *Expert Syst. Appl.* **1998**, *15*, 1–31.
39. Li, S. Research on Customer Cluster Analysis by SOM-Take the Customer Service of SMS Service as an Example. Master's Thesis, Chung Hwa University, Hsinchu, Taiwan, 2010.
40. Luo, Q. Application of Data Exploration to Explore Customer Loyalty and Value Analysis of Specialty Stores for Outdoor Activities. Master's Thesis, National Changhua Normal University, Changhua, Taiwan, 2007.
41. Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Hum. Genet.* **1936**, *7*, 179–188. [[CrossRef](#)]
42. Cooper, D.R.; Emory, C.W. *Business Research Methods*; Irwin: Chicago, IL, USA, 1995.
43. Zhu, W. Using Distinguishing Analysis to Determine the Attribute Classification of Two-Dimensional Quality Models-Taking Chain Convenience Stores as an Example. Master's Thesis, Chung Hwa University, Hsinchu, Taiwan, 2006.
44. Wang, L. Application of Discrimination Analysis Method to Discuss Generating Factors of Binocular Wall Typhoon in Northwest Pacific. Master's Thesis, National Central University, Taoyuan, Taiwan, 2017.
45. Lin, J. Credit Score Prediction-Taking Taiwan's Listed Plastic Companies as An Example. Master's Thesis, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, 2013.
46. Jian, Q. Analysis of Bank Personal Consumer Credit Loan Credit Risk Assessment Model and Lending Pricing Strategy-Taking a Domestic Bank as an Example. Master's Thesis, National Dong Hwa University, Hualien, Taiwan, 2006.
47. Shin, Y.Y.; Liu, C.Y. A Method for Customer Lifetime Value Ranking—Combining the Analytic Hierarchy Process and Clustering Analysis. *Database Mark. Cust. Strategy Manag.* **2003**, *11*, 159–172.
48. Ma, B.; Li, F.; Wang, G.; Li, C. Stochastic RFM Model and Its Application in Retail Customer Value Identification. *J. Manag. Eng.* **2011**, *25*, 102–108.
49. McCorkell, G. *Direct, Database Marketing*; Kogan Page: London, UK, 1997.
50. Ajvand, M.K.; Tarokh, M.J. Recommendation Rules for an Online Game Site Based on Customer Lifetime Value. In Proceedings of the 7th International Conference on Service Systems and Service Management, Tokyo, Japan, 28–30 June 2010; pp. 968–973.
51. Rust, R.T.; Lemon, K.N.; Zeithaml, V.A. Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *J. Mark.* **2004**, *68*, 109–127. [[CrossRef](#)]
52. Haenlein, M.; Kaplan, A.M.; Schoder, D. Valuing the Real Option of Abandoning Unprofitable Customers When Calculating Customer Lifetime Value. *J. Marketing* **2006**, *70*, 5–20. [[CrossRef](#)]
53. Reinartz, W.J.; Thomas, J.S.; Kumar, V. Balancing Acquisition and Retention Resources to Maximize Customer Profitability. *J. Mark.* **2005**, *69*, 63–79. [[CrossRef](#)]
54. Liu, D.R.; Shih, Y.Y. Integrating AHP and data mining for product recommendation based on customer lifetime value. *Inf. Manag.* **2005**, *42*, 387–400. [[CrossRef](#)]
55. Shih, Y.Y.; Liu, D.R. Product Recommendation Approaches: Collaborative Filtering Via Customer Lifetime Value and Customer Demands. *Expert Syst. Appl.* **2008**, *35*, 350–360. [[CrossRef](#)]
56. Hidalgo, P.; Manzur, E.; Olavarrieta, S.; Farias, P. Customer Retention and Price Matching: The Afps Case. *J. Bus. Res.* **2007**, *61*, 691–696. [[CrossRef](#)]
57. Rosset, S.; Neumann, E.; Eick, U.; Vatnik, N.; Idan, Y. Customer Lifetime Value Modeling and Its Use for Customer Retention Planning. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 332–340.
58. Haenlein, M.; Kaplan, A.M.; Beeser, A.J. A Model to Determine Customer Lifetime Value in a Retail Banking Context. *Eur. Manag. J.* **2007**, *25*, 221–234. [[CrossRef](#)]
59. Huang, B. Using the Technology of Data Mining to Explore the Churn Model of Pyramid Customers-Taking a Mass Merchandise Store as an Example. Master's Thesis, National Chung Cheng University, Chiayi, Taiwan, 2011.
60. Zhong, M. Research on Supermarkets in Villages and Towns-Taking Xiluo Town as a Case. Master's Thesis, National Taiwan Normal University, Taipei, Taiwan, 2002.
61. Xu, Y.; Chen, S. Constructing the Customer Lifetime Value and Customer Equity Model of the Art Industry and Calculating the Return on Marketing Investment. *Data Anal.* **2013**, *8*, 17–40.
62. Zhuo, Z. Application of Data Exploration Methods to Establish an Insurance Company's Customer Lifetime Value Management Model. Master's Thesis, National Taipei University of Technology, Taipei, Taiwan, 2009.
63. Chen, L. Research on Application of RFM, Cluster Analysis and Association Rules in Financial Commodity Portfolio Recommendation-Taking Bank A as an Example. Master's Thesis, Fu Jen Catholic University, Taipei, Taiwan, 2017.
64. Agrawal, R.; Imielinski, T.; Swami, A. Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 26–28 May 1993; pp. 207–216.
65. Goh, D.H.; Ang, R.P. An Introduction to Association Rule Mining: An Application in Counseling and Help-Seeking Behavior of Adolescents. *Behav. Res. Methods* **2007**, *39*, 259–266. [[CrossRef](#)] [[PubMed](#)]
66. Xie, Y. Using Association Rules to Adjust Selfie Angle. Master's Thesis, National Taiwan Normal University, Taipei, Taiwan, 2016.

67. Wen, Y. Applying the Law of Relevance to Data Exploration to Explore the Impact of University Entrance Scores on Academic Performance-Taking the Department of Asset Management as an Example. Master's Thesis, Nanhua University, Chiayi, Taiwan, 2005.
68. Zhao, T. Exploring the Association Rules of Customers' Online Shopping Behavior. Master's Thesis, National Yunlin University of Science and Technology, Yunlin, Taiwan, 2013.
69. Wang, Y. Research on the Use of Association Rules to Explore the Characteristics of Bank Customer Deposits. Master's Thesis, Universal University of Science and Technology, Taoyuan, Taiwan, 2015.
70. Li, C. Analysis of Association Rules of Convenience Store Transaction Data in School District with Semester Time and Temperature Factors. Master's Thesis, Nantai University of Science and Technology, Tainan, Taiwan, 2010.
71. Huang, J. Research on Application of RFM Grouping and Association Rules Technology in Product Recommendation-Take Telemarketing as an Example. Master's Thesis, Chinese Culture University, Taipei, Taiwan, 2009.
72. Jaime, R.; Ralf, V.D.L. A Partially Hidden Markov Model of Customer Dynamics for CLV Measurement. *J. Interact. Mark.* **2013**, *27*, 185–208.
73. Siti, M.; Putri, N.; Rice, N. Analysis for Customer Lifetime Value Categorization with RFM Model. *Procedia Comput. Sci.* **2019**, *161*, 834–840.
74. Art, W. Creating Superior Customer Value in the Now Economy. *J. Creat. Value* **2020**, *6*, 20–33.