Data Science and Engineering (DSE) Record, Volume 3, issue 1.

Using data visualization for supermarket retail analysis

Shuming Wang [1] and Phisanu Chiawkhun [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

wang_shuming@cmu.ac.th

Abstract. The key words of this study are data visualization, RFM model and customer relationship management. Taking Wal-Mart as a research case, analyze the number of goods and stores in Wal-Mart supermarkets in four years (2011-2014) through data and visualize the results. Using visualize tool to analyze sales and commodity and present the results in a chart. It is particularly important to retain customers and improve customer retention rates in an environment of increasingly fierce competition among supermarket retailers. First, the supermarket retail data is visualized to make the data set more intuitive and easy to understand. Then the supermarket can choose the RFM (Regency, Frequency, Monetary) analysis method to distinguish customer value, so as to provide marketing services for different customers. The application of RFM first conducts basic evaluation through data visualization methods, and has a general grasp of retail data, then conducts RFM modeling to obtain the customer's RFM score, finally differentiates customer value.

Keywords: customer segmentation, retail enterprises, RFM model.

1 Introduction

The key words of this study are data visualization, RFM model and customer relationship management. Taking Wal-Mart as a research case, it is particularly important to retain customers and improve customer retention rates in an environment of increasingly fierce competition among supermarket retailers. First, the supermarket retail data is visualized to make the data set more intuitive and easy to understand. Then the supermarket can choose the RFM (Regency, Frequency, Monetary) analysis method to distinguish customer value, so as to provide marketing services for different customers. The application of RFM first conducts basic evaluation through data visualization methods, and has a general grasp of retail data, and then conducts RFM modeling to obtain the customer's RFM score, and then differentiates customer value, so that supermarkets can design different products with different RFM scores. customer's marketing strategy. If supermarkets want to better maintain customers, they must first identify the value of different customers, and then carry out targeted marketing and services, so as to increase the rate of return and even improve customer loyalty.

For general retention customers, general retention customers and general development customers that create low value for supermarkets, their activity is low and may be lost, so marketing costs should be reduced, and they can be reawakened by giving "electronic coupons" and email reminders to stimulate the next consumption. For customers of general value, you can set up a package purchase, how much you will get free when you spend, and apply for stored-value cards or joint bundled sales to increase the unit price of customers. It is important to retain customers and develop customers. You can exchange gifts at the store, participate in free activities, Increase the frequency of visits to the store by means of member activity days. It is important to keep customers, remind them whether they need to repurchase by time period, or conduct a return visit to the product appropriately to remind customers of the existence of the online store. Important value customers with high three indicators are very important resources for supermarkets. Supermarkets must maintain good interaction with these customers, regularly provide reminder services and design some promotional activities for them, so as to increase their loyalty.

2 Literature Review

2.1 Changes in retail statistics in the era of big data

Big data refers to a large amount of diversified data that is difficult to handle with traditional database management tools. At present, it is generally believed that big data has three characteristics (3V): the first volume is that the data capacity is very large, and the second Velocity is that the data growth rate is very fast, and the processing speed is very fast.

The third Variety is that data types are becoming more and more diversified [1]. With the advancement of data collection and storage technology, the retail industry has gradually formed big data in the retail industry. Through mining and analysis of these data, it can bring huge commercial value and service innovation to retail companies, such as better understanding and insight into consumers, so as to achieve precision marketing, or change the supply chain model to achieve refined management of goods Wait. But before that, it is necessary to understand which aspects of big data competition have brought changes to the retail industry in order to make good use of retail big data in a targeted manner.

In terms of data collection in the retail industry. The data collection work carried out by the retail industry mainly focuses on a series of data such as the purchase behavior of its customers, the sales of goods, and the financial status of the enterprise. The enterprise uses these data to make analysis and decision-making on daily production and operation. In the traditional statistical method, the original data collection behavior of enterprises is often mainly based on random sampling of customer groups to conduct questionnaire surveys, collect various sales reports and financial statements, and then process, analyze and apply these original data. Under this data collection method, the data collected by the retail industry is mostly structured data, and its application has great limitations. Now and in the future, the information generated by various daily behaviors is in a state of explosion. The efficiency of traditional data collection methods is getting lower and lower, and it can't keep up with the rhythm of information generation. It can no longer meet the needs of enterprises to innovate in production and operation in the new era. Therefore, retail enterprises urgently need to undergo major changes in data collection methods. It now appears that the direction of this change will be carried out simultaneously in online and offline data collection, and the structure of the collected data will shift from structured data to semi-structured or even unstructured data. Specifically, we can collect a series of online behavior data of customer groups on the Internet and store them for sorting and analysis. At the same time, due to the development of the Internet of Things and cloud computing technology, we can also collect consumers' physical information through electronic devices. A series of offline behavior data of the store [2]. In terms of online data collection, we will mainly collect information left by customers on retail websites, including a series of click behaviors stored in browser cookies and website transaction behaviors directly stored in the backend of the system.

2.2 Customer value segmentation method

From the perspective of customer relationship management, there are three main methods for customer value segmentation: ABC classification, CLV analysis, and RFM analysis.

(1) ABC classification

The ABC classification method is based on the twenty-eight rule [3]. The ABC analysis rule is an upgraded application of the twenty-eight rule. The common point of

the two is to distinguish the primary and secondary analysis objects. The difference is that the two categories are different. The analysis objects of the twenty-eight rule are two types. ABC analysis can be divided into three categories. According to the value created by customers for the enterprise, customers are divided into different categories such as high-end customers, large customers, medium customers, and small customers. High-end customers and major customers are classified as category A, their number is 20% of the total number of customers, and their sales accounted for 70% to 80% of the total sales; category B customers contribute 10% to 20% of the sales, and the rest are the total number of customers of C, B and C customers is about 80%. After the customers are classified by the ABC classification method, the management and maintenance of the customers can be carried out in a targeted manner. The focus of resources and time will be tilted to the A customers and the maintenance will be carried out, while the secondary maintenance will be carried out for the B customers and the C customers. Only simple maintenance is required.

(2) CLV analysis method

CLV refers to the Customer Lifetime Value (Customer Lifetime Value) [4], which refers to the value created by the customer for the company during the entire life cycle of the company. In a broad sense, CLV refers to the present value of all profits obtained from a customer during the entire process of maintaining a trading relationship with a customer. CLV includes current value and future value. What companies really care about is the future profit of customers. Therefore, CLV in a narrow sense only refers to the future value of customers. The CLV analysis method uses the current value and future value of customers as the horizontal and vertical coordinates, and can divide customers into: VIP customers, improved customers, maintenance customers and abandoned customers. The divided VIP customers will be regarded as the most valuable customers and become the core of the business; the improved customers are called the most growing customers and can focus on training; the maintenance customers are ordinary customers, and general maintenance is enough; The type of customer is considered to be a negative customer who can't bring enough to balance the related service costs for the company, and should be abandoned.

(3) RFM analysis method

RFM is a customer segmentation method based on online store customer consumption behavior data proposed by Arthur Hughes of the American Database Marketing Research Institute, which takes the most recent consumption (Regency), consumption frequency (Frequency) and consumption amount (Monetary) as important Indicators to analyze and segment customers. The most recent consumption (Regency) refers to the time from the last consumption among the customers' multiple consumptions. In theory, the closer the customer is to the last consumption, the better, and marketers can measure consumer loyalty from this. Frequency refers to the number of purchases made by consumers in a certain period of time. According to this indicator, customers who buy most often have the highest loyalty. Monetary refers to the total amount of products purchased by customers in a certain period of time. The higher the

sum of the customer's purchase amount over a period of time, the greater the value that the customer creates for the company [5].

3 Data and Methodology

3.1 Data

The data source is the kaggle platform, which is a four-year retail data set of global large supermarkets with detailed data. The data set is "superstore_dataset2011-2015.csv", with a total of 51,290 records and a total of 24 features. 24 features are row-id, order-id, order-date, ship-date, ship-mode, customer-id, customer-name, segment, city, state, country, postal-code, market, region, product-id, category, sub-category, product-name, sales, quantity, discount, profit, shipping-cost, order-priority.

Out[2]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	...	Product ID	Category	Sub-Category	Product Name	Sale
0	42433	AG-2011-2040	1/1/2011	6/1/2011	Standard Class	TB-11280	Toby Braunhardt	Consumer	Constantine	Constantine	...	OFF-TEN-10000025	Office Supplies	Storage	Tenex Lockers, Blue	408.30
1	22253	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	...	OFF-SU-10000618	Office Supplies	Supplies	Acme Trimmer, High Speed	120.36
2	48883	HU-2011-1220	1/1/2011	5/1/2011	Second Class	AT-735	Annie Thurman	Consumer	Budapest	Budapest	...	OFF-TEN-10001585	Office Supplies	Storage	Tenex Box, Single Width	66.12
3	11731	IT-2011-3647632	1/1/2011	5/1/2011	Second Class	EM-14140	Eugene Moren	Home Office	Stockholm	Stockholm	...	OFF-PA-10001492	Office Supplies	Paper	Enermax Note Cards, Premium	44.86
4	22255	IN-2011-47883	1/1/2011	8/1/2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	...	FUR-FU-10003447	Furniture	Furnishings	Eldon Light Bulb, Duo Pack	113.67

5 rows × 24 columns

Fig. 1. Data Display

Data cleaning refers to the process of discovering and correcting identifiable errors in data files, including checking data consistency and dealing with invalid and missing values. The consistency check is to check whether the data meets the requirements based on the reasonable value range and mutual relationship of each variable. This article first checked the data consistency and found that the data this time did not exceed the normal range, logically unreasonable or contradictory data. Afterwards, the data was checked for missing values, and data with missing values was deleted.

3.2 Methodology

As an important resource of an enterprise, customers also have a life cycle. Customer life cycle value refers to the value created by customers for the company during the entire process of contact with the company. The typical customer life cycle value includes the investigation period, the formation period, the stability period and the recession period. At present, one of the more common and recognized models used to evaluate customer value is the RFM model. The RFM model was first proposed by Hushes, and it includes R (Recency), F (Frequency), the three variables of M (Monetary) are all consumption data from customers. Proximity (R) represents the time between the customer's last transaction and the current time. The smaller the value, the greater the possibility that the customer and the company will have another transaction; the frequency (F) represents the customer's interaction with the company within the analyzed time period. The number of transactions, the larger the value, it indicates the customer's loyalty to the company.[6]

(1) Customer value model based on product category

Based on the above analysis, this paper proposes a multi-level customer value segmentation model based on the RFM model (see Figure 2).

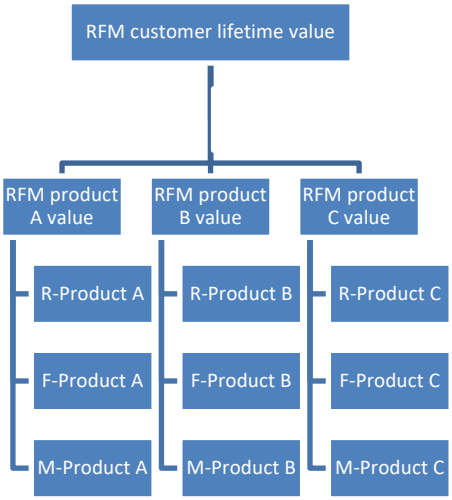

Fig. 2. RFM ustomer lifetime value

According to the nature (price, life cycle or use) of each product of the retail enterprise, the products are classified, assuming that they are divided into three categories: A, B, and C, and then the RFM model based on the three categories of A, B, and C products are constructed respectively, as shown in the figure As shown in 1, the lifetime value of an individual customer is the sum of the customer value of the three types of products A, B, and C. The customer value of various products is to consider only one

category of products, and then use the traditional RFM model to calculate the customer value of the corresponding product. The calculation formula for the lifetime value of an individual customer is as follows:

$$Vold=WRR+WFF+WMM$$
$$Vnew=\alpha AVA+\beta BVB+\gamma CVC=\alpha ARFMA+\beta BRFMB+\gamma CRFMC \quad (2)$$

Among them: Vold represents the lifetime value of individual customers in the traditional RFM model, R, F, and M represent the standard value of the proximity, frequency and amount of individual customers who regard all the products of the company as a whole, and WR, WF and WM represent three The relative weight of the variable, the sum is 1;

Vnew represents the lifetime value of individual customers of an enterprise with a multi-level model. VA, VB, and VC represent the value of product A, product B, and product C of individual customers. RFMA, RFMB, and RFMC represent product A, product B, and product C of individual customers, respectively The RFM value, the previous coefficient represents the relative weight of various products.

3.3 Data analytic process

(1) Sales situation

a. Sales

At this part, first, build a sub-data set of overall sales. Group and sum the sub-data set, according to year and month. And then, split the data into one table per year. Build sales table, rename the row and column names and show as a picture. From this picture, we can know: sales in the second half year are higher than the first half. And with the increase in years, the sales have also increased. And then, calculate the growth rate and total annual sales:

Calculate the growth rate：

```
rise_12=sales_sum[1]/sales_sum[0]-1

rise_13=sales_sum[2]/sales_sum[1]-1

rise_14=sales_sum[3]/sales_sum[2]-1

rise_rate=[0,rise_12,rise_13,rise_14]
```

	Sales-2011	Sales-2012	Sales-2013	Sales-2014
Jau	138241	162801	206459	268266
Feb	134970	152661	191063	244159
Mar	171456	201609	230548	347721
Apr	128833	187470	233181	302133
May	148147	218960	304510	304799
Jun	189338	249290	341162	372577
Jul	162035	174394	223643	278672
Aug	219223	271670	323877	432731
Sep	255238	256568	326897	405437
Oct	204675	239321	270122	406659
Nov	214934	270723	383039	508955
Dec	292360	291972	371245	427757

Fig. 2.sales situation

And then, Check the Area stacked chart.

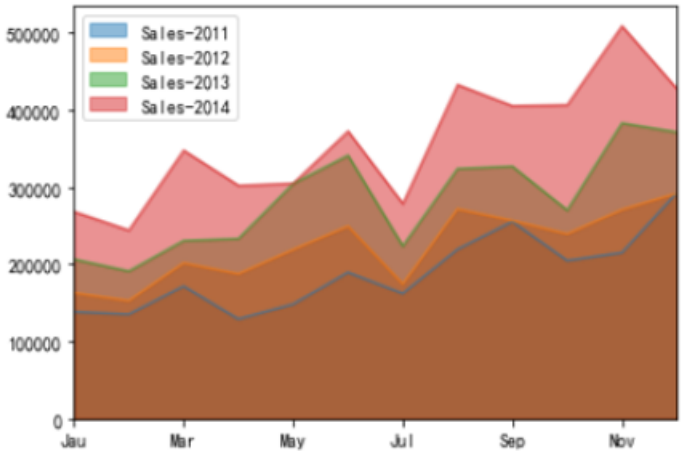

Fig. 3. Area stacked

From this picture as we can know: The supermarket's sales are seasonal. The first half year is off-season, the second half year is busy-season. In the first half year, sales in June were high, in the second half year, sales in July were low.

b. Sales quantity

Build sales volume table.

	Quantity-2011	Quantity-2012	Quantity-2013	Quantity-2014
Jau	2178	2150	2766	3656
Feb	1794	2156	2535	3450
Mar	2183	2880	3130	4820
Apr	2181	2506	3161	4397
May	2057	3239	4115	4415
Jun	2715	3508	4948	5528
Jul	2266	2551	3594	3867
Aug	2909	3451	4564	5506
Sep	3357	3752	5067	5741
Oct	2615	3399	3977	5915
Nov	3165	4497	5274	7020
Dec	4023	4022	5005	6307

Fig. 4 sales volume

And then, calculate sales quantity growth rate and the total annual sales quantity.

Calculate sales quantity growth rate：

rise_12=quantity_sum[1]/quantity_sum[0]-1

rise_13=quantity_sum[2]/quantity_sum[1]-1

rise_14=quantity_sum[3]/quantity_sum[2]-1

rise_rate=[0,rise_12,rise_13,rise_14]

c. Profit

Build profit table.

	Profit-2011	Profit-2012	Profit-2013	Profit-2014
Jau	13457.2	19627.4	26052.4	31447.7
Feb	17588.8	17828.2	31554	29454
Mar	16169.4	22752.9	34873.7	50097
Apr	13405.5	20804	26008.1	35709.5
May	14777.5	22865.4	39053.3	34335.6
Jun	25932.9	34358	43298.8	40869
Jul	10631.8	21725.2	27019.9	26450.7
Aug	19650.7	36511.4	32977.8	46443.5
Sep	32313.3	25039.9	18850.1	52533.8
Oct	30745.5	27773.9	27872	52342.5
Nov	21261.4	26160.6	51720.9	55561.7
Dec	33006.9	31968.3	47654.3	48920.8

Fig. 5 profit table

And then, calculate annual total profit and profit rate.

profit_rate= profit_sum/ profit_sum

	profit_sum	year	sales_sum	profit_rate
0	248940.81154	2011	2.259451e+06	0.110178
1	307415.27910	2012	2.677439e+06	0.114817
2	406935.23018	2013	3.405746e+06	0.119485
3	504165.97046	2014	4.299866e+06	0.117252

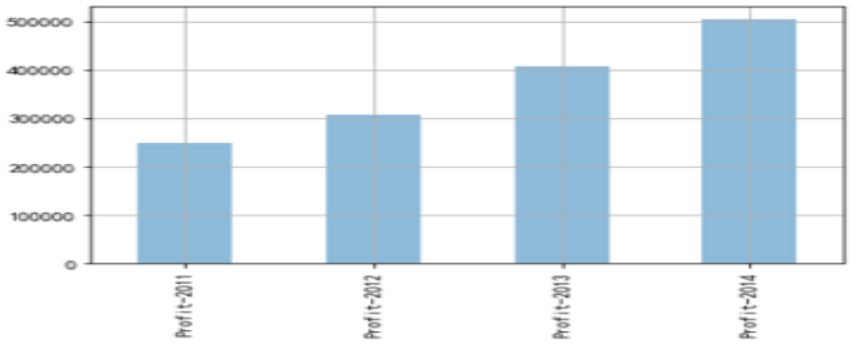

Fig. 6 annual total profit and profit rate

The annual profit, like the sales, is increasing year by year. It shows that the business situation is reasonable. Profit margins are generally stable.

(2) Commodity situation

First, check the top 10 products of sales volume and products of sales. It can be seen that most of the highest sales volume are office supplies, and most of the highest sales

are electronic products and furniture with higher unit prices. And then, check the top 10 products of profit, the result is half of them are electronic products in top 10 products of profit, and we can focus on increasing the sales of these products so that can increase overall profits.

Next step is analyze the sales of different types products. According to the product type and sub-category, combine into a new category. And then, group the new category, calculate the sales and profit. And sort. Cumulative proportion of sales of each category of goods.

	Category_Sub_Category	Profit	Sales	cum_percent
16	Technology_Phones	216717.00580	1.706824e+06	0.135007
14	Technology_Copiers	258567.54818	1.509436e+06	0.254401
1	Furniture_Chairs	140396.26750	1.501682e+06	0.373181
0	Furniture_Bookcases	161924.41950	1.466572e+06	0.489184
11	Office Supplies_Storage	108461.48980	1.127086e+06	0.578335
4	Office Supplies_Appliances	141680.58940	1.011064e+06	0.658308
15	Technology_Machines	58867.87300	7.790601e+05	0.719931
3	Furniture_Tables	-64083.38870	7.570419e+05	0.779811
13	Technology_Accessories	129626.30620	7.492370e+05	0.839075
6	Office Supplies_Binders	72449.84600	4.619115e+05	0.875611
2	Furniture_Furnishings	46967.42550	3.855783e+05	0.906110
5	Office Supplies_Art	57953.91090	3.720920e+05	0.935542
10	Office Supplies_Paper	59207.68270	2.442917e+05	0.954865
12	Office Supplies_Supplies	22583.26310	2.430742e+05	0.974091
7	Office Supplies_Envelopes	29601.11630	1.709043e+05	0.987610
8	Office Supplies_Fasteners	11525.42410	8.324232e+04	0.994194
9	Office Supplies_Labels	15010.51200	7.340403e+04	1.000000

Fig. 7 the top 10 products of sales volume and products of sales

As we can see, different types of goods have different sales. Nearly half of the total sales of goods accounted for 84%.The profit of Tables is negative, that means this product is belonging to loss status. By checking the original data, l found that most of Tabels are discounting. I think that can explain why the Tables profit are negative.

（3）User situation

Find out the proportion of different types customers. As we can see, in the past four years, ordinary consumers accounted for the largest proportion of customers, reaching 51.7%. And then check the number of different types of customers every year.

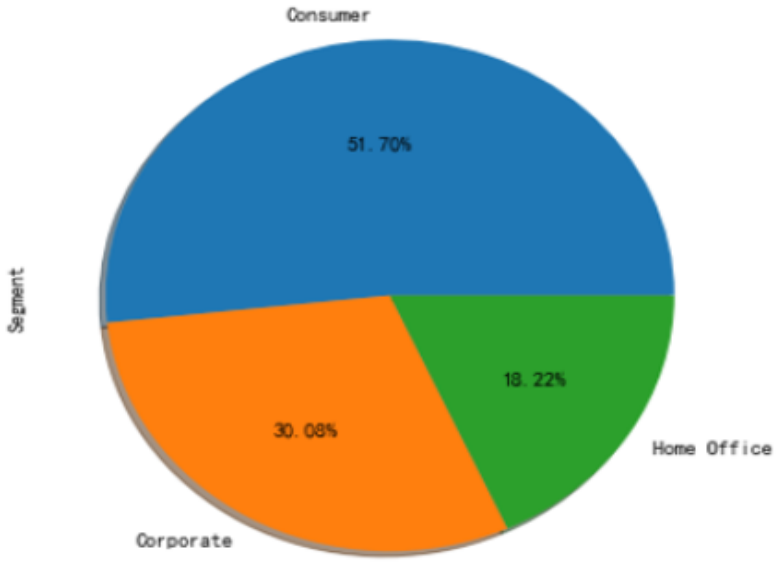

Fig. 8 proportion of different types customers

Fig. 9 the number of different types of customers every year

As we can see, each type of customer maintains a growth trend every year, and the customer structure is still very good. And then, check the annual sales contributed by different types of customers.

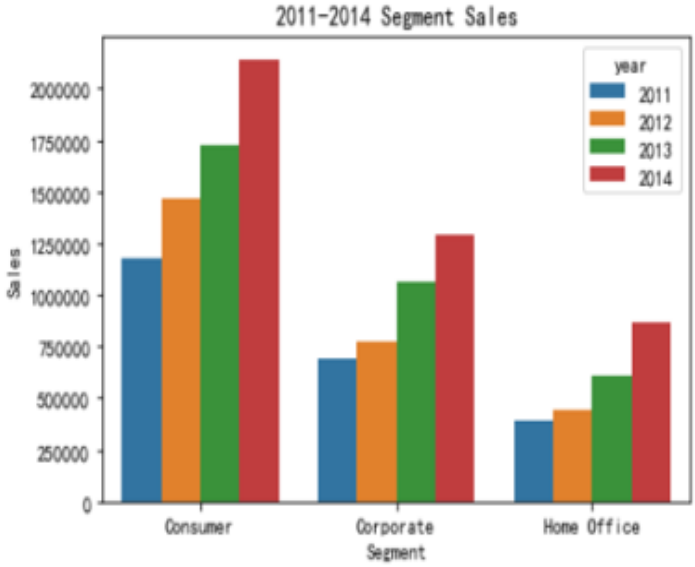

Fig. 10 annual sales contributed

Sales contributed by all types of customers are increasing every year. And ordinary consumers contribute is the most. Next step is customer order behavior analysis. Check the distribution of users' first and last purchase date

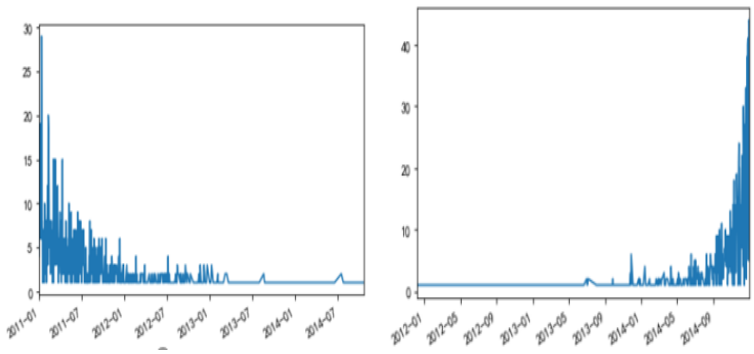

Fig. 11 distribution of users' first and last purchase dates

As we can see, after the beginning of 2013, the new users growth trend is slow.The number of new customers is decreasing year by year.

（4）RFM model

The RFM model is an important tool and method to measure customer value and customer profitability.

R (Recency): The interval of the customers's last transaction time. The larger the R value, the longer date of the customer transaction time.

F (Frequency): The number of customer transactions in the most recent period. The larger the F value, the more frequent customer transactions.

M (Monetary): The amount of customer transactions in the most recent period. The larger the M value, the higher the customer value, and vice versa.

First, build the RFM table.

Customer_ID	Order_Date	F	M	R
AA-10315	2014-12-23	42	13747.41300	8.0
AA-10375	2014-12-25	42	5884.19500	6.0
AA-10480	2014-09-05	38	17695.58978	117.0
AA-10645	2014-12-05	73	15343.89070	26.0
AA-315	2014-12-29	8	2243.25600	2.0

Fig. 12 RFM table

According to RFM Model define, the maximum transaction date of all users is the standard, and the time interval for each transaction is R. The number of orders for each customer is F, and the total sales is M. Then mark the customer value and divide the customer into 8 levels (2 x 2 x 2).[16]

Level 1: Important Value Customers

Level 2: Important Keep customers

Level 3: Important Retention Customer

Level 4: Important Development Customers

Level 5: General Value Customers

Level 6: General Keep Customers

Level 7: General Retention Customers

Level 8: General Development Customers

Level 1: Important Value Customers

Belonging to corporate loyal users, according to the life cycle, certain marketing strategies are required to extend the user loyalty time, such as VIP services, permanent discount strategies, etc., marketing strategies that give users a sense of honor;

Level 2: Important Keep customers

There has been a recent purchase behavior, and the consumption amount is higher than most users, but the purchase frequency is low. It is necessary to stimulate users to repurchase and increase their loyalty to the brand. You can regularly send information such as new products, popular items, discounts, etc. through text messages and emails to attract users. increase the frequency of purchases;

Level 3: Important Retention Customer

The most recent consumption time is long, but the consumption frequency and total amount are high, indicating that they are loyal customers who have not patronized for a period of time (will be lost soon), and we need to take the initiative to contact them to restore them. We can communicate by telephone to arouse users' awareness of the brand. Develop personalized marketing strategies based on user feedback, such as sending large coupons;

Level 4: Important Development Customers

The historical consumption amount is high, but the purchase frequency is low. There has been no purchase behavior recently, and it has tended to be lost. The core demand is to stimulate users to repurchase and increase the impression of products and brands. Profits, such as the free trial, buy one get one free and other services

Level 5: General Value Customers

Recently purchased, the purchase frequency is high but the total consumption amount is low. The unit price of this part of users is low. You can try to advertise on the site and off the site, such as YouTube and other platforms, to increase brand exposure and increase the unit price of customers;

Level 6: General Keep Customers

There has been purchase behavior recently, and the purchase frequency and consumption amount are relatively low. From the perspective of the user life cycle, it is in the introduction period and growth period. It is necessary to take the initiative to contact the product by phone, care for the use of the product, provide good after-sales service, and increase the user's awareness of the brand. Reliability and amount increase the repurchase rate and consumption amount;

Level 7: General Retention Customers and Level 8: General Development Customers.

Users are already in the churn stage, they can either do not do marketing, or try to reach them at a lower cost.

The RFM analysis application is for customer grouping, that is, the three indicators are divided into two types: "high" and "low", those higher than the average value are "high", and those lower than the average value are "low". Therefore, there are three things to do: calculate the average of each indicator score; define each variable as "high" if the score is above the average and "low" if it is below the average; according to the three variables "high" and "low" to define customer types; such as “high”, “high” and “high” for high-value customers.

R	F	M	Level
High	High	High	Level 1: Important Value Customers
Low	High	High	Level 2: Important Keep customers
High	Low	High	Level 3: Important Retention Customer
Low	Low	High	Level 4: Important Development Customers
High	High	Low	Level 5: General Value Customers
Low	High	Low	Level 6: General Keep Customers
High	Low	Low	Level 7: General Retention Customers
Low	Low	Low	Level 8: General Development Customers

Fig. 13 8 levels of RFM model

Customer_ID	Order_Date	F	M	R	label
AA-10315	2014-12-23	42	13747.41300	8.0	Important keep customers
AA-10375	2014-12-25	42	5884.19500	6.0	General keep customers
AA-10480	2014-09-05	38	17695.58978	117.0	Important value customer
AA-10645	2014-12-05	73	15343.89070	26.0	Important keep customers
AA-315	2014-12-29	8	2243.25600	2.0	General development customers

Fig. 14 RFM model result

Compare based on the average value, if the value greater than average value, It is 1, otherwise it is 0. Then analysis of new users, active users, inactive users and regression users.

month	2011-01-01 00:00:00	2011-02-01 00:00:00	2011-03-01 00:00:00	2011-04-01 00:00:00	2011-05-01 00:00:00	2011-06-01 00:00:00	2011-07-01 00:00:00	2011-08-01 00:00:00	2011-09-01 00:00:00	2011-10-01 00:00:00	...	2014-03-01 00:00:00	2014-04-01 00:00:00	2014-05-01 00:00:00	2014-06-01 00:00:00	201 00:(
Customer_ID																
AA-10315	0	0	1	1	0	0	0	0	1	0	...	0	1	0	1	
AA-10375	0	0	0	1	0	0	1	0	0	1	...	0	0	0	0	
AA-10480	1	0	1	1	0	1	0	1	0	0	...	1	1	1	0	
AA-10645	1	0	0	1	0	1	1	0	0	0	...	1	1	0	1	
AA-315	0	0	0	0	0	0	0	1	0	0	...	0	0	1	0	

5 rows × 48 columns

Fig. 15 compare the average value result

And then define the state function and mark the state

	0	1	2	3	4	5	6	7	8	9	...
Customer_ID											
AA-10315	unreg	unreg	new	active	unactive	unactive	unactive	unactive	return	unactive	...
AA-10375	unreg	unreg	unreg	new	unactive	unactive	return	unactive	unactive	return	...
AA-10480	new	unactive	return	active	unactive	return	unactive	return	unactive	unactive	...
AA-10645	new	unactive	unactive	return	unactive	return	active	unactive	unactive	unactive	...
AA-315	unreg	unreg	unreg	unreg	unreg	unreg	unreg	new	unactive	unactive	...

5 rows × 48 columns

Fig. 16 state function and mark the state

（5）results

It can be found that the number of new customers is decreasing every year, that means the new customer obtainment rate is low. The results of user situation analysis:

1. Ordinary consumers account for the largest proportion of all customers, but the customer structure is still balanced, and the number of customers is increasing every year.

2. By observing the latest purchase date, it can be found that users have basically not lost and the rate of returning is high. This shows that the customer's stickiness is very high.

3. However, after 2013 year, the growth of new users has been slow, Managers can put advertising and other promotion strategies, to attract more new users.

4. Identify different customer groups by RFM, it can measure customer value and customer profitability, and that can make personalized communication and marketing

services, to provide strong support for more marketing decisions, and create greater benefits for the company.

4 Analysis conclusion and discussion

Through the above division and comparison of customers, we found that it is more reasonable to divide customers into four categories. We refer to the four types of customers as new users, active users, inactive users, and returning users.

Active customers obtain high value through the company's products and services. These customers also create high value for the company through high profit margins, strong customer loyalty and long-term retention. This kind of relationship is mutually beneficial, customers get higher value, and the company also gets customer loyalty and higher profits. Therefore, for such customers, companies should invest in marketing resources to the greatest extent, pay attention to their needs from the early stage of product development, and communicate effectively with them. In terms of marketing strategy, products or services should be provided according to their requirements, and personalized services can be provided in terms of transaction conditions, settlement methods, and delivery.

What the new customer represents is an unequal and unstable relationship.New customers can create higher value for the company. Such customers may include those newly acquired customer groups, their experience is unstable, they may be considering why you should choose your product in the first place; they may also be long-term stable customers, but they remain loyal to the company due to inertia. In short, these customers are very sensitive, if they do not adopt an active marketing strategy, they may turn to competitors. Companies can increase their value growth by paying more marketing efforts, providing better products and more service support. Companies should take the initiative to communicate with these customers and take corresponding special measures for their unmet needs. So that they will not develop into inactive customers.

With the gradual deepening of the "customer-centric" marketing concept in various industries, research on customer value and other related data mining-based customer relationship management will receive increasing attention. At present, major companies have fully realized the importance of customer value and satisfaction to the company, and are actively carrying out related projects. The customer segmentation in this research is one of the current research focuses. After having a certain understanding of customers, further research should be carried out on customers' consumption habits and spending power. It can also in-depth evaluation of customer loyalty and analysis of the reasons for the decline in customer loyalty, so that the company's customer relationship management level is more perfect. Therefore, the research of customer value based on data mining has broad research prospects. In order to improve the competitiveness of enterprises, enterprises need to analyze customer value, retain high loyal customers and tap potential value customer groups, formulate effective competition strategies, concentrate enterprise resources, and enhance their own advantages. The RFM method still has broad market application prospects. The recent purchase date,

purchase amount, and purchase frequency can not only be combined to identify customers, but they also have individual indicators. It has important meanings, for example, the customer obtains the customer's life cycle through the analysis of the most recent purchase date, etc.

The supermarket obtains consumer data through the back-end database and calculates the customer RFM level to distinguish the value of different types of customers, and formulates targeted marketing plans to promote more consumption by members with low consumption intentions and high consumption quotas, and promote members with low consumption quotas to increase Consumption quota, thereby creating greater value for the enterprise.

For general retention customers, general retention customers and general development customers who create low value for the supermarket, their activity is low and may be lost. Marketing costs should be reduced. They can be re-awakened through coupons or email reminders to stimulate them. Time consumption. For general-value customers, you can set up package purchases, how much you can get when you spend, store-value cards or joint bundling sales, etc. to increase the unit price of customers. Important customers and important development customers can be exchanged for gifts, participate in free activities, Increase the frequency of visits to the store by means such as member activity days. It is important to keep customers, remind them whether they need to repurchase according to the time period, or properly conduct product telephone return visits to remind customers of the existence of offline supermarkets. Important value customers with high three indicators are very important resources for supermarkets. Merchants must maintain good interaction with these customers, regularly provide reminders and design some promotional activities for them, so as to increase their loyalty.

References

1. Hughes AM. Strategic database marketing ［M］. Chicago，IL: Probus Publishing Company，1994 .
2. Zeng Hong, Feng Minxuan. Big data and statistical transformation [J]. China Statistics, 2013, (9): 49-5O.
3. Duan Shuyong. Research on recommendation methods based on customer value segmentation [D]. Jilin: Jilin University, 2011.
4. Wei Jianguo. Application of RFM Model in Commercial Bank Customer Behavior Analysis [J]. China Financial Computer, 2011 (6): 40-43.
5. An Xiangqian. Research on customer value identification in C2C environment based on RFM model [D]. Sichuan: Southwestern University of Finance and Economics, 2012.
6. Haughton D，Oulabi S. Direct marketing modeling with CART and CHAID ［J］. Journal of Direct Marketing，1993，7 (3)，16－26 .