Organic Recommendation System for MATSUSEI
Team: CLWC

Hsuan-Yu Chang[1], Wan-Yi Chou[2], Ching-Yu Li[3], Chih-Hui Wang[4], and Prof.Chiu-Hsing Weng Supervisor[5]

[1]Department of Statistics, National Chengchi University
[2]Department of Management Science, National Chiao Tung University
[3]Master Program in Statistics, National Taiwan University
[4]Department of Statistics, University of California, Berkeley
[5]Department of Statistics, National Chengchi University

May 27, 2018

Abstract

In this report, we discuss the business model and background for MATSUSEI, concluding that MATSUSEI should promote the organic products strongly to fit the market situation. First, with the 2-year sale data through 2009 and 2010, we make the market segmentation. Second, considering the quantity of each categories purchasing by every customers, we cluster customers by non-negative matrix factorization to grab the potentially most valuable customers of MATSUSEI, we select the number of groups by the criteria that could distinct the target customer well. After clustering the customers, we utilize sequential pattern mining to find out the purchasing pattern in the data. We could implement strategies with the purchasing pattern, such as limited time coupon, to stimulate potential purchasing desire. Finally, in order to capture the customers heterogeneity, we use the recommendation system technique to model the quantity a user might purchase a product. Our workflow can be applied to different target products in a specific period to not only increase sales but also improve corporate image.

Keywords: NMF, LIBMF, SPM, RFM, MATSUSEI, recommendation system, organic products

1 Introduction

The report organizes in following ways: We first give an overview and discuss the business model of MATSUSEI by SWOT analysis. We describe how we cleaned the data, explored the pattern in the data, and implemented RFM model in Section 2. For Section 3, we present how we performed the Recommendation System, including nonnegative matrix factorization and sequential pattern mining. Also, we designed the strategy for MATSUSEI based on the Recommendation System. Lastly, we make a conclusion in Section 4.

1.1 Business Baackground

Jhong Guo Cing Nian Store (中國青年商店), which was the former name of Sung Ching Commercial Co., Ltd.(松青商業公司) had appeared before other supermarkets started their business in Taiwan. In 1986, Wei Chuan Foods Corporation(味全食品公司) cooperated with FRESSAY (a Japanese company), introducing their experience and management to open the first "MATSUSEI"[1].

In 2009, although 76 branches of MATSUSEI started to turn around their business[2], however, from 2010 to 2015, MATSUSEI had gotten loss yearly, and the number of their branches had decreased. In 2010, MATSUSEI had lost 265 million, and in 2015, they also had lost 193 million. The total of loss in five years was 1 billion. What's worse, people had refused to buy their products, since the food safety scandal of Ting Hsin Group(頂新集團) in 2013. Finally, MATSUSEI could not afford all of these, merged by Pxmart Company (全聯) in 2016[3]. In this report, our primary aim is to use two-year sales data, 2009 and 2010, and help MATSUSEI make a better business strategy. If we can make use of all two-year data we obtain, and try to help MATSUSEI make some changes in the beginning of 2011, can everything be different for MATSUSEI now in 2011, 2012, or even 2016?

1.2 Business Model: SWOT Analysis

In order to set the goals and strategies for MATSUSEI, we utilize SWOT analysis to realize what resources MATSUSEI could allocate as well as do market research to help them understand whole market situation and development. The Figure 1 shows the overview of MATSUSEI's SWOT analysis in the end of 2010.

- Strength

In this part, we lay stress on what MATSUSEI owned, or its strength. To put it differently, if we can take advantage of MATSUSEI's strength, we can find out the key to success! The following are four vital strengths we investigate for MATSUSEI:

1. High Quality and Variety

MATSUSEI provided with health ingredients, information transparency, many kinds of fresh ingredients and daily necessities to their customers. Firstly, for special local ingredients, MATSUSEI cooperated with Hualien(花蓮) farmers, providing with non-toxic fruits and vegetables to make customers buy safe food. Furthermore, MATSUSEI also collaborated with farmers from Tainan(台南), Pingtung(屏東), and so on. to serve with special local ingredients as well as their detailed information about places and farmers of the production[1][4]. Secondly, for daily rigorous "green shield" test(綠盾檢測), MATSUSEI tests the quality of all kinds of fruits and vegetables before putting them on the shelves for purchasing[5]. For various imported products, MATSUSEI sells popular and unique Japanese food, rice, organic seasoning. Additionally, we can also see Spanish olive oil, Australian premium beef, and so on[6].

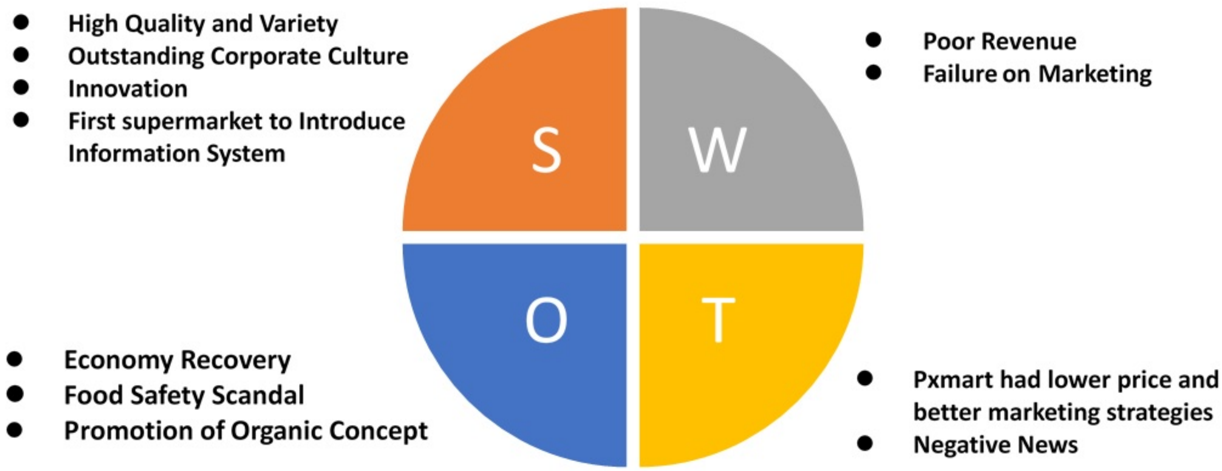

Figure 1: SWOT Analysis

In sum, customers has a wonderful shopping environment to buy high quality of commodities. Besides, in 2010, MATSUSEI obtained first place in the services survey of magazine, “Global Views Monthly”(遠見雜誌), about fresh food quality[1].

2. Outstanding Corporate Culture

Corporate spirits for MATSUSEI are:「快樂工作」、「團體合作」、「品質第一」、「創新突破」、「客戶滿意」. Besides, MATSUSEI uses the following words to represent their company culture:「今天我們要以信心與熱忱，提供顧客最高的滿足，對商店與商品賦予最深的愛心，莫忘奉獻的精神，為達成自己的希望而努力。」Therefore, when employees serve customers, they will keep this in mind and provide the best service[7].

3. Innovation

In 2010, MATSUSEI opened its first “Delicate Life Style” supermarket in Da-an District(大安區) in Taipei, trying to offer fresh and cooked food such as daily-made sushi, which made not only small families, but also single families enjoyed the services. It is MATSUSEI that always puts their customers in their shoes before their customers need something.

Furthermore, in 2010, MATSUSEI cooperated with Yahoo to start its “E-MATSUSEI Online Supermarket”, so customers could order and receive fresh food in one day. From this, we can know that MATSUSEI was willing to do something different and innovative, making their customers enjoy more convenient shopping experience[8]. However, Pxmart doesn’t have their online supermarket service until now in 2016[9].

4. First Supermarket to Introduce Information System

In 1998, MATSUSEI was the first supermarket in Taiwan that introduced information system SAP ERP, so it collected a lot of historical data about purchasing, operations, financial, and so on, which could combine with their Business Intelligence System to dig more useful information, helping MATSUSEI make better strategies[1].

- Weakness

In this part, we discuss what MATSUSEI lacked, which could give us some thoughts to improve its weakness. We find that there were two significant weaknesses for MATSUSEI:

1. Poor Revenue

In 2009, MATSUSEI's revenue was only 5.1 billion, however, the revenue of MATSUSEI's biggest competitor, Pxmart, achieved 50 billion. What's worse, in 2010, MATSUSEI had a loss of 265 million[3][10].

2. Failure on Marketing

From 2008, to enhance the corporate image of MATSUSEI, which had same pronunciation of Japanese could make customer think of high quality of Japanese products, MATSUSEI drafted marketing budget about thirty million. Furthermore, MATSUSEI launched some commercials to market their organic food and special local food, however, not many people were impressed with their commercials or what they emphasized.

- Opportunity

We find that there were some opportunities for MATSUSEI in the end of 2010, so MATSUSEI could adjust their strategies in the beginning of 2011. Therefore, we decide to push MATSUSEI to focus on "Organic Market". Moreover, we discuss details for MATSUSEI's opportunities from several aspects.

1. Economy Recovery

Firstly, for economy, after Financial Crisis of 2008, in 2009, average gross profit margin for retailing market was low to -4.45% [11]. Besides, economic growth rate was low to -1.57 %, however, in 2010, economic growth rate increased significantly to 10.63%. Moreover, in 2008, per capita National Income was NT485,347; however, in 2009, per capita National Income decreased slightly to NT476,000. In 2010, we are pleased to see that per capita National Income increased NT526,963. According to the information above, we can roughly regard the trend of 2011 as an optimistic growth[12]. Therefore, if MATSUSEI tried to promote and sell more high-priced commodities, we think people would not reject to buy.

2. Food Safety Scandal

Secondly, for food safety scandal, we can see that many kinds of food in Taiwan were not safe at all. For instance, in 2009, dead sick pork made of meatballs, stores sold rice for livestock to people, Dioxin(戴奧辛)was found in duck meat, and so on. Moreover, in 2010, black tea with Carcinogen(致癌物) was sold in famous beverage stores, Trans Fat's (反式脂肪) label was unclear, and so on[13].

3. Promotion of Organic Concept

In 2009/1/31, new organic produce law started to be implemented, which had the regulation of "compulsory labeling", to prevent consumers buying fake organic produce. Although the central ideas of organic foods are to foster cycling of resources and promote ecological balance and conserve biodiversity, consumers could have incentives to buy more organic produce[14][15].

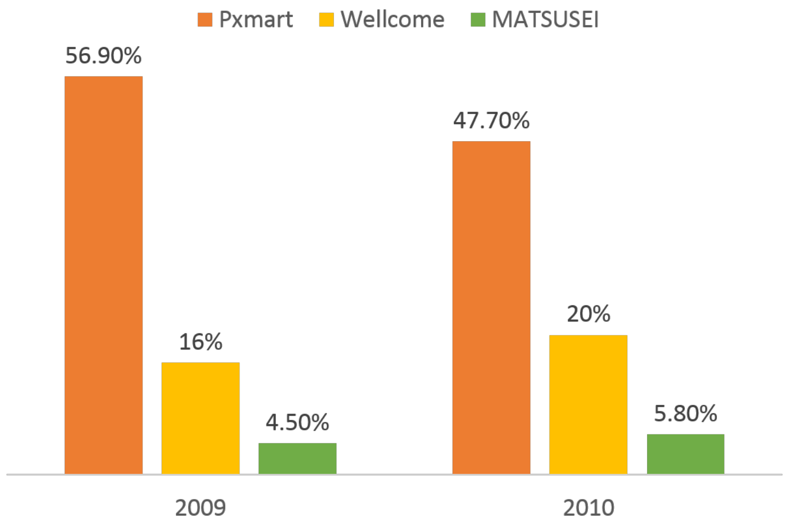

Figure 2: Which supermarket do customers visit frequently in 2009 and 2010? (Multiple Choice)

Furthermore, in 2009, organic-food lovers from all walks of life set November 11th to be "Organic Agriculture Day"(全國有機農業日)[16]. Besides, the growth rate of households who planted organic produce increased year by year. We can see that, from 2008 to 2010, growth rate are 0.04%, 0.31%, and 0.39%, which represents "Organic Market" was promising so more and more people started to invest in organic agriculture[17].

- Threat

In this part, we probe into external threat, to see who would be MATSUSEI's biggest competitor, or which events would hurt MATSUSEI. Here, we found there were two significant things would be threats for MATSUSEI.

1. Pxmart had lower price and better marketing strategies

According to the survey using E-ICP (Eastern Integrated Consumer Profile, 東方消費者行銷資料庫), in 2009 and 2010, people visited Pxmart, the biggest competitor of MATSUSEI, more frequently. Furthermore, this survey showed that there were 1,479 samples in 2010, and 1374 samples in 2009, containing 12 brands. However, the Figure 2 only shows top three brands of Taiwan supermarkets, Pxmart, Wellcome(頂好) and MATSUSEI[18].

We think there are some reasons why people visited Pxmart more frequently. We use two viewpoints to discuss. On one hand, for Pxmart, in 2006, to establish lower price impression, Pxmart marketing Mr. Pxmart to be their brand ambassador. Afterwards, almost everyone in Taiwan knew Mr. Pxmart, which meant that Pxmart made a considerable nice strategy. In 2006, Pxmart started to sell fresh food owing to the fact that they knew they should not only focus on dried food market. Therefore, they hired experts from Japan. Moreover, in 2009, they even found 30-year expert in fresh food area to be their consultant. Most importantly, in 2010, they started to sell organic products which were inexpensive than any supermarket[18]. On the other hand, for MATSUSEI, its concept was to sell special local ingredients, therefore, with low production and high quality, which led to high cost. Besides, general manager said that MATSUSEI wouldn't utilize price-oriented strategy which was similar to Pxmart. The reason was that MATSUSEI would like to promise their customers with higher quality instead of low price[19].

2.Negative News

In 2009, after new organic produce law started to be implemented, the organic market was flooded with about 90% organic food without qualified organic labels, including some famous brand, for example, MATSUSEI, Wellcome, and so on. However, we think the reason was that MATSUSEI had no time to adjust the labels of their organic products immediately. For more details, original organic produce law mentioned that products labeled organic must be verified, however, it didn't mentioned that produce not proven to be organic cannot be put organic on the packaging[14]. Therefore, we think that, for MATSUSEI, this event wasn't serious. In other words, if they could adjust organic labels to fit the law, customers would always believe in the quality of MATSUSEI.

- Summary

In conclusion, after making SWOT analysis to think about what we can do for MATSUSEI in the beginning of 2011, we decide to utilize a new opportunity in "High-priced Organic Market" for MATSUSEI as well as combine its strength in "Quality". Actually, according to the survey, among top three supermarkets, Pxmart, Wellcome and MATSUSEI, MATSUSEI had the highest quality of organic food as well as more "true food" instead of "artificial food" in their supermarkets (true food ratio for Pxmart, Wellcome and MATSUSEI are 4%, 4%, and 16%). Which means MATSUSEI beat others among supermarkets[19]. Moreover, we can find that there are too much "artificial food" in the organic store(有機專賣店) instead of "true food", which means organic store provided with products without nutrient, as well as candies or cookies with high calories and sugar. Even if these things were organic, however, they're all bad for our health. Compared with organic store, we can see that MATSUSEI also had an upper hand on quality[19].

In other words, for our business goal set for MATSUSEI is help MATSUSEI focus on "Organic Market", which includes finding who will buy or will be possible to buy organic products, and realizing what customers are eager to buy.

Therefore, if our customers purchase more organic products in MATSUSEI, the brand image would also be raised up because of Self-Perception Theory. In brief, Bem(1972[20])said that people tend to infer attitudes that are consistent with their past behavior. Moreover, when a customer thinks of organic products as well as he(she) used to purchase in MATSUSEI, he(she) will leave a good image on MATSUSEI, according to Balance Theory(Fritz Heider , 1946[21]).

Moreover, we utilize STP Theory to depict who are target audience of MATSUSEI and decide its position(Figure 3). By doing this, we can also help them make profit on these high-priced products, as well as establishing strong brand image by word of mouth by doing better customer relationship management and building customer loyalty.

2 Data Preprocessing and Description

Our datasets contain only six shops among about 100 supermarket of MATSUSEI. We assume our datasets are sampled from the database randomly so that they could use our sample to estimate and infer our population. We have four datasets. First, for customer data, we have 4,377 records and each contains the basic information about the customer including birthday, gender and job. Second, we have 105,170 transaction records. It includes the store number, member ID, quantity and total price. Third, we have the products names and categories in category dataset. Finally, there is a dataset for products names and their IDs.

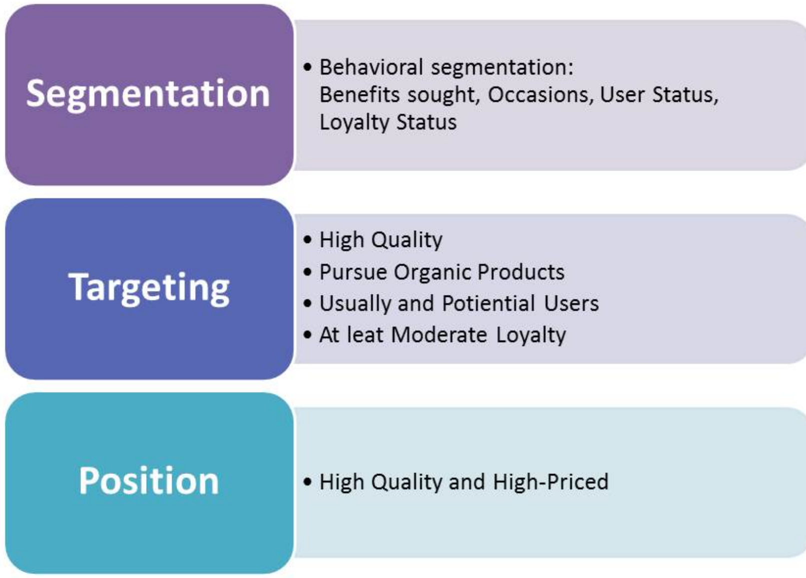

Figure 3: STP

2.1 Data Cleaning

First, the original transaction dataset has 105,170 records from 2009/1/1 to 2010/12/31, containing messy information. For example, some customers came to the store to recycle empty wine bottle to gain 2 dollars reward. Also, there are several wrong records such as negative amounts which were because products were accidentally added transactions and were balanced later. Before any further analysis, we should clean the data. For the wine bottle, we added back the money to total purchase dollars and then deleted the records. For wrong product records, we sum up the quantity of the same product in each receipt to solve this problem. Then, we removed all negative quantities records. In other words, in each receipt ID, there would be an unique product ID, whose purchase quantities are the sum of quantities for both positive and negative quantities. Furthermore, there are some products are sold by weights, meaning that the same product would have a different price because of different weights. To deal with this problem, we used the average price to replace the original price. Finally, the transaction dataset after preprocessing has 100,972 records.

Moreover, for product dataset, which contains the products' Chinese names, products coding and their categories of 11,171 products, 124 products are simultaneously corresponding to two categories. We deleted the inappropriate one. Besides, there are also some products which had never been bought during this two years, so we removed those records. Then, 11,018 products remained. For category dataset, 431 categories, some products are organic but their categories were not assign to organic categories. There are 11 organic categories. We derived and combined suitable coding and Chinese names for these organic in our dataset. We also delete those categories has never been purchased and left 409 categories. Afterward, we combined these two datasets into the transaction dataset by product ID and category ID.

Second, for customer dataset, three customers bought something and returned on the same day and they did not come after purchasing, so we eliminated these three customers data. Also, the authenticity of demography variables in this dataset are also doubtful because there are inconsistency among their birthday, education, and Marital status. For example, there is a

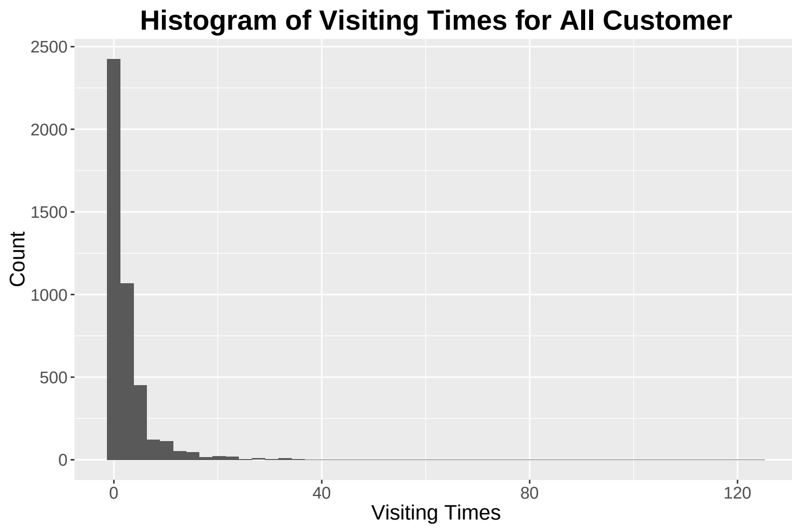

Figure 4: Distribution of Visiting Times for Customers

person who is seven years old but married or fifteen years old with a college degree. Moreover, there are some uninterpretable variable names. This is understandable that a person would not take the applied form seriously because the privacy concern or there is no incentive for them to fill the real data. Therefore, we had concerns about whether we should use these demography variables. To sum up, we combined four datasets into two to do the following analysis.

2.2 Data Exploration

First of all, we examined what the distribution of visiting times is. It is essential to the following analysis because more frequently customers come to the store, the more confidence we have when implementing the analysis and strategy. However, from Figure 4, we found that the distribution is highly right-skewed. There are 2,426 customers only visiting the market once. The most frequent visiting customer is member 3860 who went to the market 125 times.

In Figure 5, we summarized the purchasing records of all the organic products by categories. We can find that most of the customers came to MATSUSEI to buy organic vegetable and rice. However, for other organic products, they seemed to be not so popular. For Figure 6, we show the detail record of customer visiting time. The x-axis is the time and y-axis is the member ID sorted by visiting times(the lower the bigger). If a member went to the store at that day, the corresponding coordinate will have a point. We can find out that there were some new memberships which started from August 2010. Originally, we thought that it should follow a certain pattern such as more people on weekend. We did not see a regular pattern besides the store number 4001 and 4002, which are new-opening store. Contrary to our concept that these customers of these two stores are more concentrated than other because the customers may feel new and fresh to visit MATSUSEI, most of these customer still come MATSUSEI once or twice only.

In table 7, we find that most of the sales come from rice, and there are no seasonal pattern or explainable reason for the pike because those are caused by some outliers. Table 1 shows top 5 purchasing dollars customers. The average purchasing dollar is NT2264.45, however, customer 2464 purchased NT180,229. To explore more specific about if there is time series trend. We delete 5 outlier in table 1 and category rice. The result is in figure 8. We could see that there is neither significant patterns among the category. We think the strange trend are

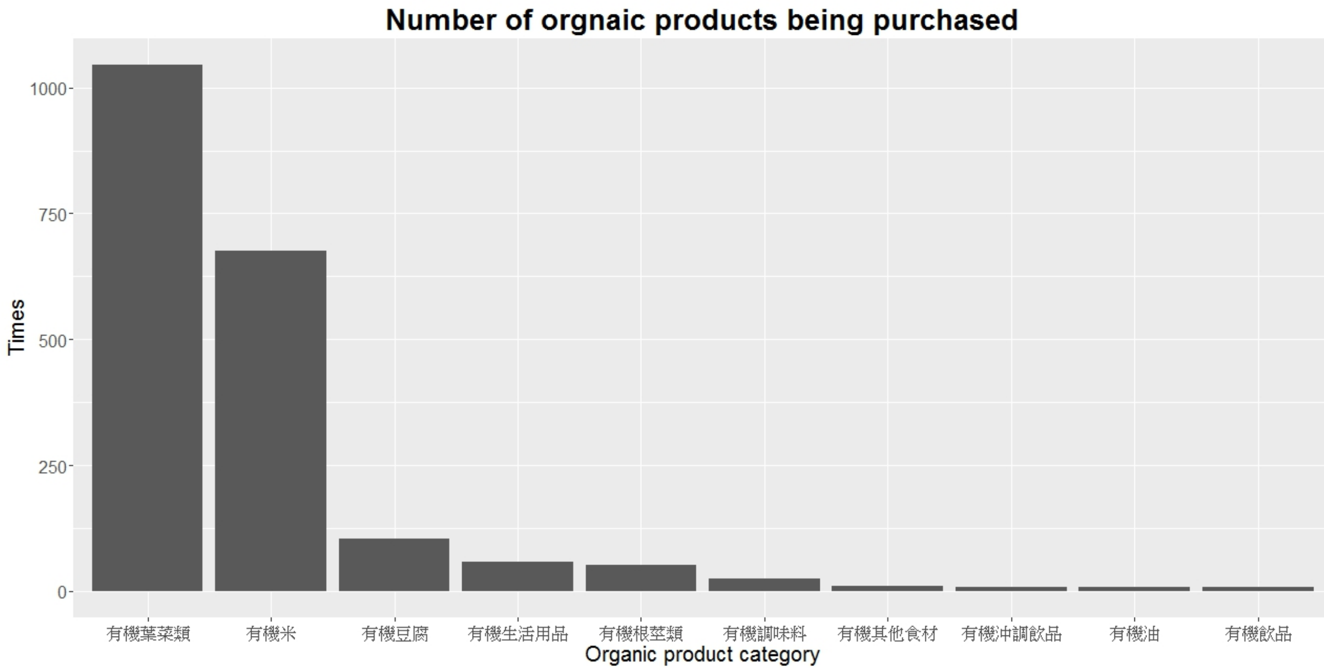

Figure 5: Total Number of Organic Product Sold Out During Two Years Group by Categories

Figure 6: Plot of Members with Transaction Dates

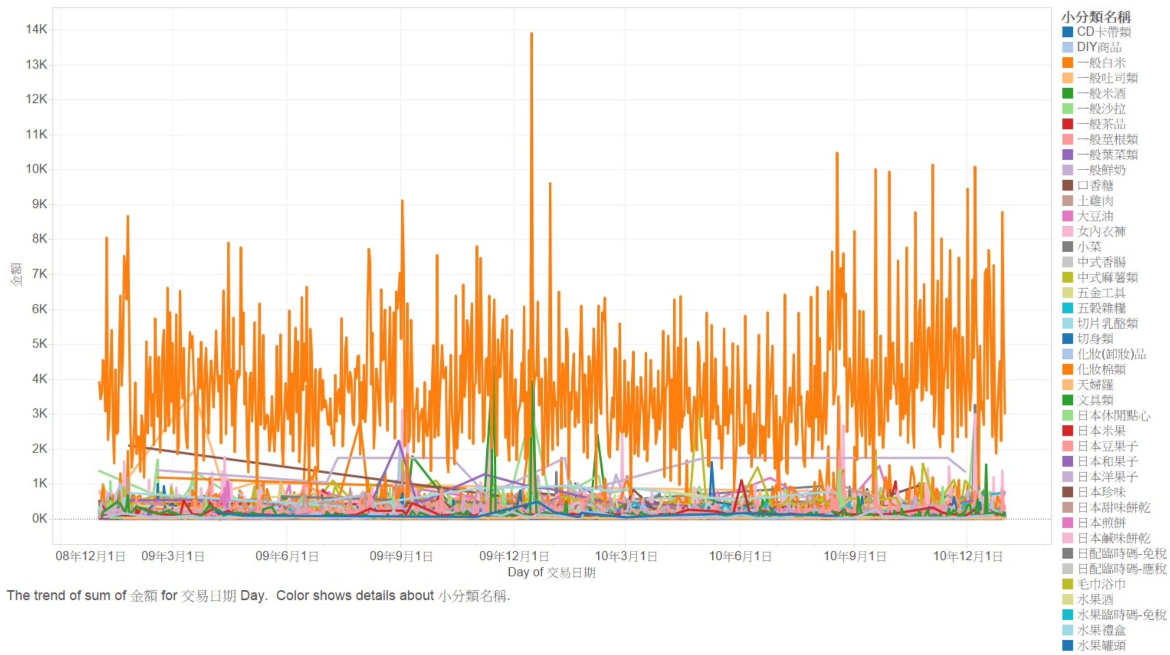

Figure 7: Line Chart of Transaction Dates with Monetary Amounts and Categories

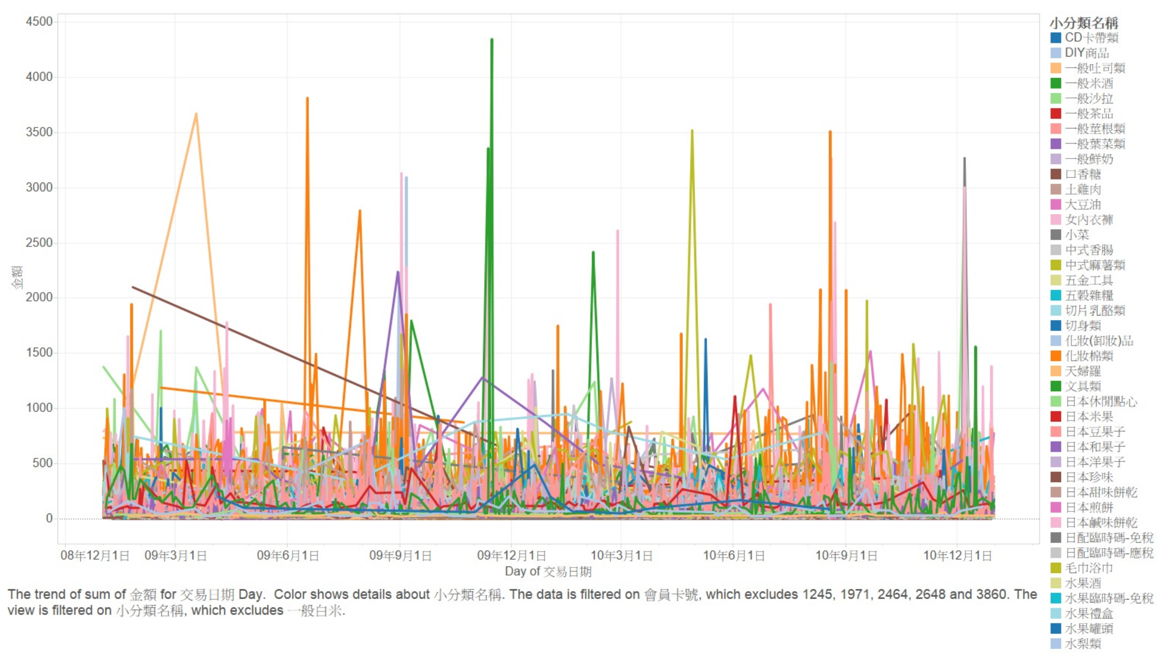

Figure 8: Line Chart of Transaction Dates with Monetary Amounts and Products without 2464, 3860, 1971, 2648, 1245 and rice

Table 1: Top 5 Monetary

ID	# Frequency	Recency	# Monetary
2464	38	2009/12/19	180,229
3860	122	2010/12/14	91,968
1971	33	2010/12/13	71,384
2648	20	2010/9/29	64,966
1245	22	2010/12/9	59,140

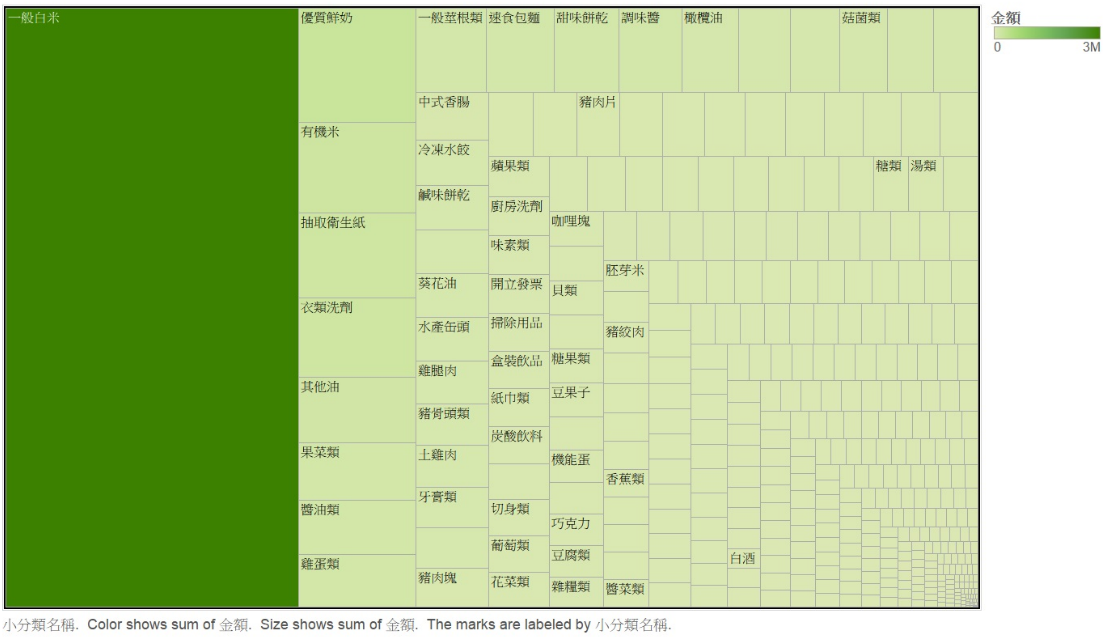

Figure 9: Block Diagram Of Product Selling Amounts

caused by either the specific promotion or customer idiosyncratic which we can not solve with this dataset so we ignore the time series in this research.

We show the overall amounts and revenues in Figure 9 and 10. You can see that most of the amounts and revenues are on two or three products and people. This may be a difficulty for us to implement the analysis.

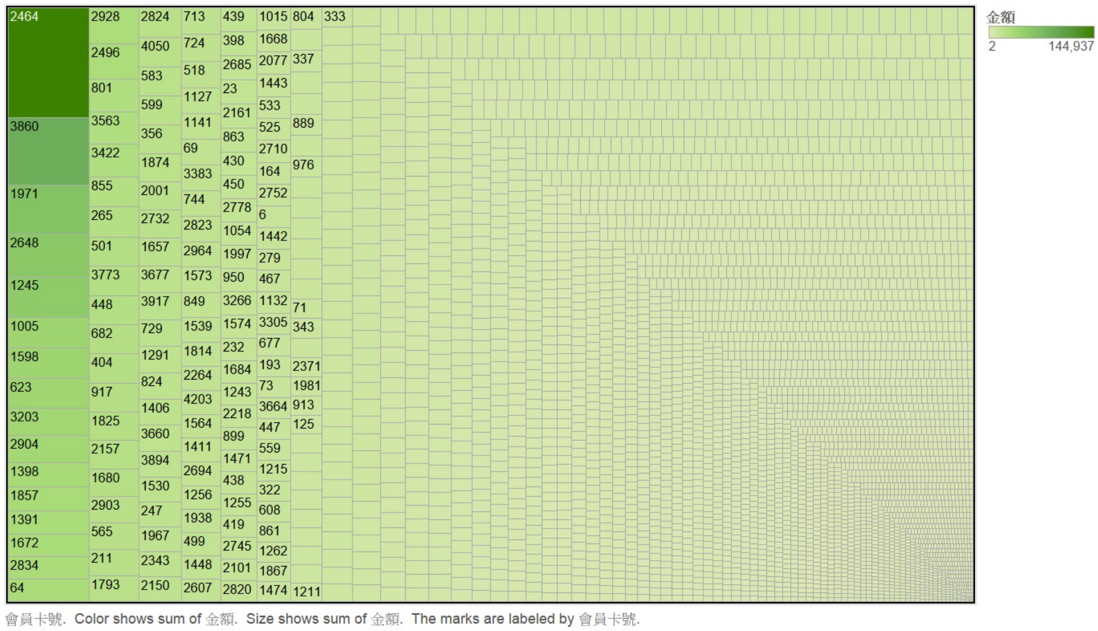

Figure 10: Block Diagram Of Members Buying Amounts

2.3 Evaluate Customer Value More than RFM

After visualizing customer profile with the previous plots, we found a few customers have a great impact on the sales of MATSUSEI. Therefore, we want to construct and index to evaluate MATSUSEI customer value to mine out who are the profile of MATSUSEI VIPs.

Arthur Hughes(1994)[22] defined an intuitive and popular model–RFM(recency, frequency, and monetary value), which is widely used in database marketing in many industries to quantify customer value and is is related to the 80/20 principle that 20% of customers bring in 80% of revenue.

Derya Birant has mentioned that integration of RFM analysis and data mining techniques provides useful information for current and new customers. For example, cluster analysis based on RFM attributes provides more behavioral knowledge of customer's actual marketing levels than other cluster analysis. Furthermore, RFM measures with association rule mining analyzes the relationships between product properties and customers' contributions to provide a better recommendation to satisfy customers' needs. Doubtlessly, RFM analysis can and has been applied to various industries. Aggelis and Christodoulakis have used it to score active e-banking users, Hsin Hsin Chang (2004) amended the elements in RFM and implemented in two stage data mining to a retailer, separated the customers into 16 types. We hope that it not only help us to distinguish customers but also to provide personalized recommendations for customers.

However, some people would discretize the continuous measurement and rank the result for the reason of approachable interpretation. Here, we use the original data not only to avoid information loss but also to combine another variable to create our own customer value[23]. For example, Bob Stone(1995)[24] placed highest weights on Frequency for the customer transaction data from a credit card company because the more frequent a customer consume the more credit card company would receive transaction fee. Catalog company knows that a customer's response probabilities to its three times per year catalog depend on both the customer's recency and frequency (Pfeifer and Carraway, 2000)[25]. In health examination industries, Monetary is the most important because customers would consume again in one to three years(趙景

明,2003)[26] . The influences of R and F are smaller.

We used the variables, the total purchase amount and the total visiting times. Moreover, we also used the total number of products purchased to complement total visiting times because there are more than three thousand customers visited less than three times in two years. We added this variable to distinguish someone might visit a few times but buy lots of products each time. Since the datasets we are analyzing is the sales data from the supermarket, Recency is hard to fit here because if a customer just consumed yesterday and his (her) consumption period is a week, the probability he (she) would come today is small. In other words, we do not want the small Recency. Instead, we hope the customer purchasing periods is short. Therefore, we used Consumer Activation Index (CAI). Moreover, our marketing target is to increase our sales on organic products which are symbols of high-priced and high-quality. We presume that those who buy expensive products are more likely to buy organic products. This gives us an idea to calculate that the proportion of every customer buy expensive products, compared to the average of its other products in the same category, among transaction dataset. Moreover, we add the quantity that customers to buy organic products to be another variable. The results are in table 2.

Table 2: Original Data of Variable to Construct Customer Value

ID	# purchases	amounts	# products	CAI	H/L price prop.	organic products
1	2	1687	31	-0.194	0.419	1
2	2	735	5	0.185	0.4	0
3	2	7912	5	0.123	0	0

Then, we want to construct our customer value index, but these variables cannot sum together directly because they may locate in a different dimension of the latent variable. As a result, we use factor analysis with varimax rotation to six variable after scale above to check whether these variable located in the same dimension. Because the price variation in the supermarket is not big, the number of products purchased is positively correlated with amounts (0.943). Moreover, number of purchases and quantity of organic product purchased are kind of similar. Sadly, the total purchase amount, the total number of purchases, the total number of products purchased, and the total number of organic products are are in the same factor with all the loadings (Table 3) are bigger than 0.65, and CAI and proportion of high-priced products are two unique factor(Table 4). Therefore, we sum up the value after scaling on the average of four purchase variable(called consume), CAI, and proportion of high-priced products to build our own customer value index called Green Value(GV) in Table 5.

Table 3: Loadings

	Factor1	Factor2
# purchases	0.763	0.176
amounts	0.863	0.500
# products	0.977	0.199
CAI		
H/L price prop.		0.138
organic products	0.857	

After calculating GV of every customer, we rank them by GV and extract the top 500 and countdown 1,000 customers to compare their demographics for discovering patterns. In

Table 4: Uniquenesses

# purchases	amounts	# products	CAI	H/L price prop.	organic product
0.388	0.005	0.005	0.999	0.981	0.263

Table 5: GV

ID	# consume	CAI	H/L price prop.	GV
1	0.114	-0.889	0.184	-0.591
2	-0.351	0.743	0.100	0.491
3	-0.348	0.476	-1.636	-1.508

both groups, we use Goodness of Fit test to find out whether the demographics in groups is distributed equally to all customers. If not, we state that the distribution of this attribute is different from the overall pattern. We test by six variable: Sex, the number of household members, occupation, education, marital status, and household income. So there are total twelve tests. Interestingly, in the top 500 group, the number of household members, occupation, marital status, and household income are significantly different. We summarize the levels whose proportion are larger than the proportion of all customers to profile the top 500 customers. Their number of household members more than 3 people, occupations are more homemaker and less in None/retired and students. They are married, and their household income is more than sixty thousand per month. On the other hand, in the countdown 1,000 group, occupation, marital status, and household income are significantly different. Their occupations are industry and students and less in homemaker. They are unmarried, and their household income is under forty thousand. Table 6 shows the summary of our customer profile.

Table 6: Customer Profile

	Top 500	Countdown 1,000
members	more than 3	
occupations	more housewife and less in None/retired and students	industry
martial status	married	unmarried
household income	more than sixty thousands	under forty thousands

3 Organic Recommendation System

We found out that our most valuable customer in table 6, however, we cannot examine the characteristic of every customer so far. As a result, for the next work, we focus on Micro-Marketing. We retrieve our target customer, then delve into their purchasing attribute. The method we use in this research has a restriction that the number of activation customers is not big enough, so we do not make the validation part. If we split the data into training and testing data by time, the dataset will be very sparse and hard to fit model. Therefore, we use all two years data and include all the customers even they just come 1 or 2 times. We hope that our model can exclude this customers. In future work, when we want to promote a special product in a small period, we could apply this work to the whole database of a retail company.

3.1 Non-negative Matrix Factorization

For clustering methods, many researchers have done the comparison between methods for clustering, Li and Ding empirically showed that Nonnegative Matrix Factorization is inherently related to kernel K-means clustering but it has clear clustering effects, and thus here we will use Nonnegative Matrix Factorization instead of K-means for clustering.

Here, we want to cluster customers by what they bought before. In other words, if two customers always buy the same thing with same quantity, we could assign these the two customers in the same cluster. Therefore, we transfer the data into customers row by category column with the amount that the customer buys the product the normalize by each category from 0 to 1. However, not every customer buy every product so this become a sparse matrix problem when we use k-means, hierarchical clustering, or SVD to cluster our customer. Jingu Kim and Haesun Park (2008) tested sparse NMF as a clustering method, and our experimental results with synthetic and text data show that sparse NMF does not just provide an alternative to k-means, but rather gives much better and consistent solutions to the clustering problem.[27]

Therefore, we use NMF to cluster our customer by what they bought before:

$$V \approx WH^T,$$

where $V \in R^{p\times n}$, which is a features by samples matrix. $W \in R^{p\times k}$, is the group center given the k group. $H \in R^{n\times k}$, sample n can be assign to group k by extracting the maximum value in a given k column. We then find the solutions of W and H are by solving the optimization problem:

$$\min_{W,H} f_k(W,H) \equiv \frac{1}{2}\left\|V - WH^T\right\|_F \quad s.t. \;\; W,H \geq 0$$

where $W,H \geq 0$ means that each element of W and H is non-negative, and the subscript k in denotes that the target reduced dimension is k. Often W is called a basis matrix, and H is called a coefficient matrix.

The cliche problem in clustering analysis is that how to chose k, number of cluster. The most common way is that researchers should have domain knowledge on the industry of this data. Here, since we are not quite familiar with customer profile of MATSUSEI, we use cross validation to solve this problem. However, contrary to most of criteria, helping us to find the most consistent and stable groups or minimize within sum of square and maximize betweeness sum of square, we want to find out the groups who bought most organic products. As a result, we construct our own criteria to fit the problem we encounter. We cluster with number of groups k from 3 to 15, each k we make 10 trials to determine which k is the most suitable for our data. Each trials of NMF cluster to k group, in one or two groups in the k groups, we want to maximize the quantity of organic product but minimize the people in a group. That is, we want to capture the patterns that a few of people buy most of the organic product, not just select specific customers but select group with similar purchasing habit. Therefore, if we compute the organic product quantity bought per person of k groups, the best result is that it is a bimodal distribution with some outliers in the right-hand side, meaning that some group contains our target customer. On the contrary, if the distribution is uniform, it's herculean for us to distinguish which group would be most valuable. We define:

$$P_{i,k,j} = \frac{po_{i,k,j}}{no_{i,k,j}}, \quad i = 1,2,...,10, \; k = 3,4,...,15, \; j = 1,...,k$$

Where $P_{i,k,j}$ is the proportion of organic product purchase per person of whom ever bought organic product, $po_{i,k,j}$ is the proportion of quantity of organic product purchased, and $no_{i,k,j}$ is the number of people ever purchased organic product. And i is the i trials, k is the number of group, j is the j^{th} group in a given i and k. For each k, we want that the distribution of $po_{i,k,j}$

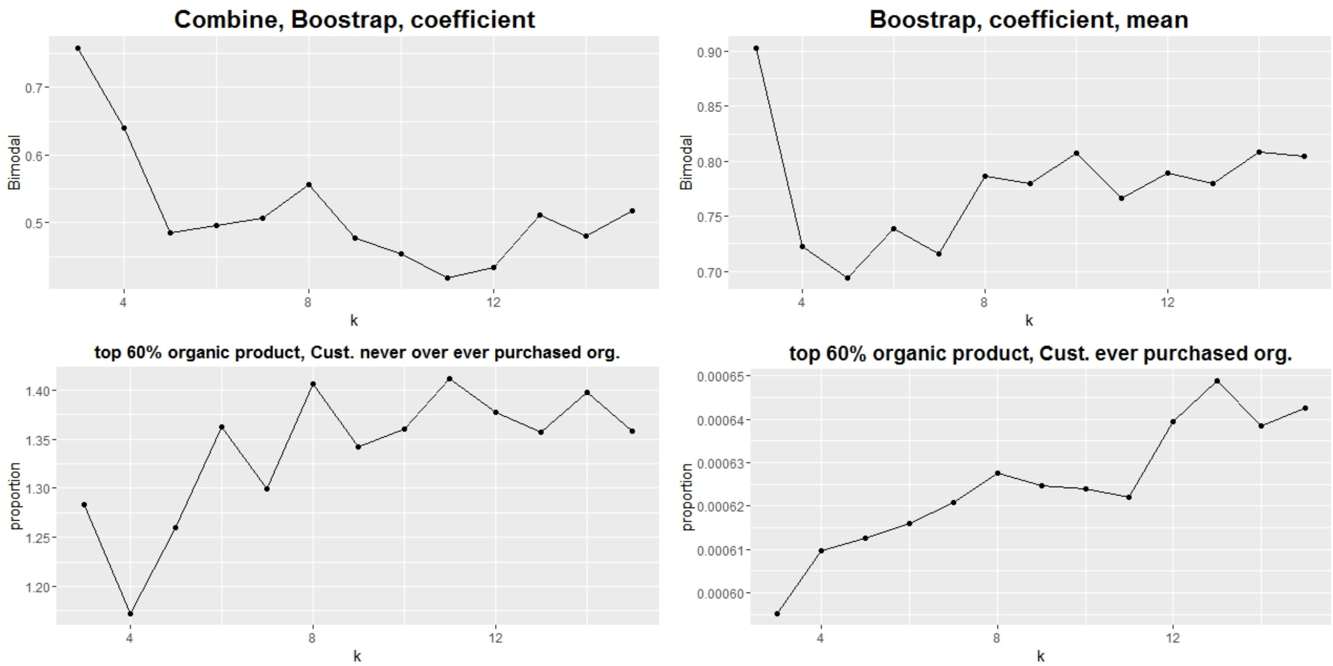

Figure 11: **Top-left:** Combine ten trials. Bootstrap 1000 samples in each k. Calculate the Sarle's Bimodality coefficient. We can see that k=8 is the first peak after fast drop from the beginning. **Top-right:** Bootstrap 1000 samples in each trials. Calculate the Sarle's Bimodality coefficient. Given k, take the average for ten trials. In the beginning, the coef. drop deeply but climb up after k=8, and become stable. **Bottom-left:** Total number of customers never purchased organic products in top n groups over total number of customers ever purchased organic products in top n groups. n is decide by the cumulative sum when bigger than 60%. This value is that how many customer would purchase with our information per customer. **Bottom-right:** Total quantity of organic product top n groups over total number of customers ever purchased organic products in top n groups. n is decide by the cumulative sum when bigger than 60%. The proportion will climb higher when k become bigger because the group will become more pure. That is, it is more possible to put every customer ever purchase organic product into same group.

to be bimodal, with some values is extremely large because by this way, only a few groups can explain most of the pattern of organic product purchasing. We use Sarle's Bimodality coefficient:

$$\beta = \frac{\gamma^2 + 1}{\kappa + \frac{3(n-1)^2}{(n-2)(n-3)}}$$

Whereγ is sample skewness, κ is sample Kurtosis, most of the β lie between 0 and 1. We use this criteria to find out the maximum among each k, however, for the reason of compare different k $\frac{3(n-1)^2}{(n-2)(n-3)}$ will be amply a lot by when n is small. To solve this problem, we bootstrap $P_{i,k,j}$ to all equal size. In other words, in each group k, we amplify the size form k to 1000 in top-right of figure 11. Compared with the method above, We combine ten trials first and bootstrap 1000 samples and calculate Bimodality coefficient on each k in top-left of figure 11

To capture more information, we also use cumulative sum to extract top n cluster, order the group total quantity proportion descending, then find out which points that cumulative sum greater than 60%. Given k and i, then use the sum of quantity of organic product top n group divided by sum of the number of customer who ever bought organic products. In

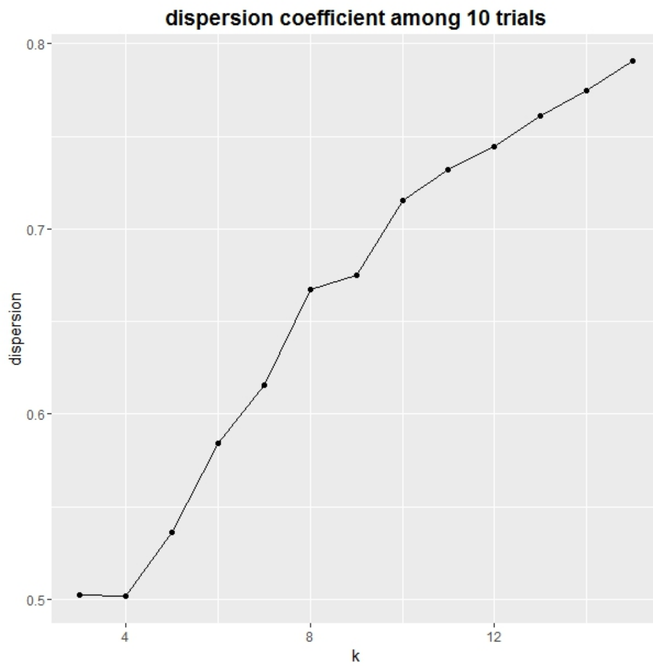

Figure 12: Dispersion Coefficient climb up slowly from 0.5 to 0.8, when k from 3 to 15. The dispersion coefficient when k=8 is 0.667

bottom-right of figure 11 We want to maximize:

$$P_{60\%,i,k} = \frac{\sum_{j\epsilon n} po_{i,k,j}}{\sum_{j\epsilon n} no_{i,k,j}}$$

Although k=13 id the peak, we want to find out minimal k to be explainable so we think k=8 can be a good choice. On the other hand, we use the total number of people do not purchased anytime divided by sum of the number of customer who ever bought organic products. In bottom-left of figure 11, we also maximize:

$$P_{60\%,i,k} = \frac{\sum_{j\epsilon n} nno_{i,k,j}}{\sum_{j\epsilon n} no_{i,k,j}}$$

Where $\sum_{j\epsilon n} nno_{i,k,j}$ is the summation of number of customer do not purchase any organic product before in a given trial i, group k, among top n groups.

On the other hand, for the cluster consistent, we use dispersion coefficient:

$$\rho_k = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} 4 \left(C_k \left(\hat{i,j}\right) - \frac{1}{2} \right)^2$$

where $0 \leq \rho_k \leq 1$, and $\rho_k = 1$ represents the perfectly consistent assignment(Jingu Kim and Haesun Park, 2008).[27] $C_k \left(\hat{i,j}\right)$ is the proportion that customer i and customer j assign into same group, if it is closed to 0 or 1, meaning that the cluster result is stable ρ_k and would be closed to 1. (Figure 12)

Based on the above criteria, we select 8 to be the number of clusters. Next, we are going to find out who is our target customer. We want to capture that how many organic products would be bought by the potential customers, who never bought organic products but shared the same purchasing habit with other who bought much organic products group. Optimize:

$$\max_j Pn_{i,8,j}(j) = \frac{po_{i,8,j}}{nno_{i,8,j}} \quad j = 1, 2, ..., 8$$

With bigger proportion of organic products $po_{i,8,j}$ are purchased by smaller per customer who never purchased organic products $nno_{i,8,j}$ could maximize $Pn_{i,8,j}$. In other words, a customer never purchase organic products but he(she) share purchasing habits with those bought a lot would be more possible to buy organic product. In each trials i, we find the target cluster and all the member in the group is our target customer, so there are ten times of selection for each customer. Table 7 shows an example.

Table 7: Ten Trials of Target Group

ID	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
1	0	0	0	0	1	1	1	1	0	1
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	1	1	0	1	1	1	0	1	0	1
6	1	0	0	1	0	1	1	0	0	0

$$\sum_{i=1}^{10} t_{i,n} \geq 5$$

Customer n will be assign to final target group if he(she) is assign to target cluster more than 5 times. There are 658 customers assigned into our final target group. Afterward, we compare the purchasing habit between target group and non target group. Then, we compare the mean of two group by t-test and extract the top ten t-statistics and countdown ten Z-statistics. 13 We know that our target customer prefer to buy more Vegetable and pork. Moreover, about customer profile of target customers, only number of household members and household incomes are significant different in chi-square test from non target customers. The number of household members of target customers are lower and household incomes are lower, too.

Furthermore, when a new customer join MATSUSEI and we got his(her) first purchase records, we can map the records into:

$$W^T V \approx Q^T,$$

Where $Q \in R^{1\times k}$ is the new customer Also, we can just assign this customer to highest weight in Q to be the group and repeated ten times of our model to see if he(she) is assign to the target group.

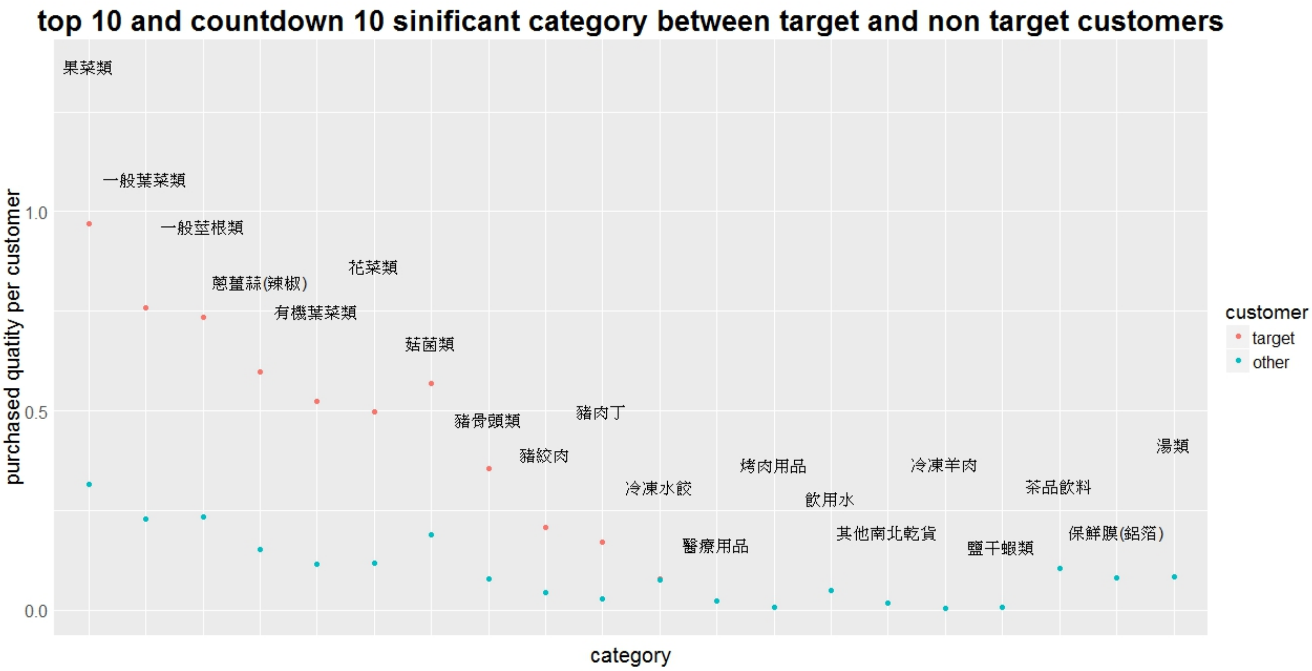

Figure 13: We use t-test to compare the mean of purchased quantity per customer in the group

3.2 Sequential Pattern Mining

Qiankun Zhao and Sourav S. Bhowmick (2003) have done a survey on sequential pattern mining, concluded that sequential pattern mining, which is widely used in analyzing of DNA sequence, is trying to find if there exist any specific order of the occurrences. We can find the sequential patterns of specific individual items, also we can find the sequential patterns cross different items. And here we are focusing on the potential customers to the organic products and the relationships between the item sets they have consumed. Therefore, We use the result of NMF to find the potential customers on organic products and try to find their consumption patterns.

Table 8 is part of our mining result with the support set at 0.0065 and the confidence is more than 0.5. Obviously, we found that there exists a basket of organic leaf vegetables that our potential organic customers would choose from. Thus, for the above items, We suggest that they could be bind together when selling. Also, we can provide the customers with time limited coupons for their next consumption to these goods with some related but relatively upset vegetable products in order to introduce their potential buyers. On the other hand, additional purchase discount (加價購) is another way to enchourage the efficient purchasing.

However, table 9 is another part of result as the mining we did for the previous table. Different from table 8, the consumers here are much preferred to a certain product of a certain brand that they will pay for the same product after trying twice. The result does make sense because for our potential organic customers, they paid more than normal customers for a better quality, and thus they are likely to be more picky on finding their loving products. Therefore, we can also figure out that, by the table, the supports for consumers to pay on the same product is negative related to the times they buy it. So what we focus here is to maintain the persistency, and turn it to a sustainable sales. Our suggestion here is to provide rewards cards on these goods, such as rice, to attract our customers to come again. Also, since these consumers are relatively loyal to certain products, it is considerable to bind the preferred products with other organic food on a time limited coupon to access these green consumptions to other options they might also interest in.

Table 8: Sequential Pattern: Organic Products I

LHS	RHS	Support	Confidence	Lift
{台灣台糖有機白菜}, {台灣有機小白菜}	{台灣台糖有機青江菜}	0.0079	0.6364	11.5584
{台灣有機青江菜}, {台灣有機小白菜}	{台灣台糖有機青江菜}	0.0067	0.6667	12.1088
{台灣有機小白菜}, {台灣有機青江菜}	{台灣有機小白菜}	0.0067	0.5455	4.9036
{台灣有機青江菜, {台灣有機小白菜}	{台灣有機小白菜}	0.0079	0.6364	5.7208
{台灣有機青江菜}, {台灣有機小白菜}	{台灣有機小白菜}	0.0067	0.6667	5.9933
{台灣有機青江菜, 台灣有機小白菜}, {台灣有機小白菜}	{台灣有機小白菜}	0.0067	0.8571	7.7056
{台灣有機小白菜}, {台灣有機青江菜, 台灣有機小白菜}	{台灣有機小白菜}	0.0067	0.8571	7.7056

Nevertheless, there are also some patterns that do not contain organic products. Since these are our target customers who are possible to purchase organic products, and we could provide incentives for them to purchase organic product in the same category when these patterns occur. For example, if a customer buy {台灣金絲菇(200g)} and {金墩米4.5kg} the next time, he(she) would be the most possible to purchase {中華超嫩豆腐} . Therefore, We can give he(she) a coupon or discount on {中華有機豆腐}, to attract he(she) to change to organic products buyer. However, for goods which can not be substituted by organic products, because there are no organics in the same category of the rule's RHS, we do not know which kinds of organic products to recommend, for instance, there is no organic leek in MATSUSEI to substitute {台灣青蔥}. To solve this problem, we introduce LIBMF in section 3.3 which can predict the quantity of a customer's consumption so that we could know which kinds of product he(she) prefers and introduce them the related organic products.

Table 9: Sequential Pattern: Organic Products II

LHS	RHS	Support	confidence	lift
{富麗有機白米2Kg}, {富麗有機白米2Kg}	{富麗有機白米2Kg}	0.0090	0.5333	5.4559
{富麗有機白米2Kg}, {富麗有機白米2Kg}, {富麗有機白米2Kg}	{富麗有機白米2Kg}	0.0079	0.8750	8.9511
{古早蛋(10粒)-L}, {古早蛋(10粒)-L}	{古早蛋(10粒)-L}	0.0079	0.5385	10.1964
{銀川有機長秈白米2kg}, {銀川有機長秈白米2kg}, {銀川有機長秈白米2kg}	{銀川有機長秈白米2kg}	0.0090	0.5000	4.1589
{銀川有機長秈白米2kg}, {銀川有機長秈白米2kg}, {銀川有機長秈白米2kg}, {銀川有機長秈白米2kg}	{銀川有機長秈白米2kg}	0.0067	0.7500	6.2383
{中華有機豆腐}, {中華有機豆腐}	{中華有機豆腐}	0.0101	0.7500	9.5357

To sum up, since our main goal is to attract buyers to green products, we will recommend them to related organic products with coupons, discounts and perhaps some slogans and advertisement that targeting on the benefits we can gain from organics. Better prices provide easier assessments for upset organic foods to customers, who can try new things with lower opportunity costs and find other suitable products with our store.

Table 10: Sequential Pattern: Nonorganic Products

LHS	RHS	Support	Confidence	Lift
{台灣金絲菇(200g)}, {金墩米4.5kg}	{中華超嫩豆腐}	0.0047	1.0000	25.7200
{台灣小白菜(裸)}, {台灣捕撈野生北海道秋刀魚}	{台灣青江菜}	0.0047	1.0000	16.0750
{中興金饌中興米-4kg, 台糖沙拉油大塑-2L, 春風抽取式衛生紙-110*6}	{台灣小白菜(裸)}	0.0047	0.7500	13.7786
{中興香米-3kg}, {中興香米-3kg}	{中興香米-3kg}	0.0093	0.7500	20.9674
{健康廚房橄欖油750ml}	{台南黑橋牌香腸(原味)-370g}	0.0047	0.7500	30.1406
{台灣青蔥, 台灣小黃瓜},	{台灣青蔥}	0.0093	0.5455	2.9473
{義進洗選蛋}	{台灣青蔥}	0.0093	0.5455	22.9494733
{義進洗選蛋}	{EGG冠軍蛋(L)}	0.0093	0.5455	6.8770

3.3 LIBMF

In this Section 3.3, we tried to make our recommendation be more perfect, therefore, we would like to predict our target audience's preference in a period, or 2009 to 2010. In other words, when we used sequential pattern mining for recommending in last Section 3.2, we can only see the pattern according to "time". Besides, if we extended our analysis to know customers' special preference in a fixed period, which wasn't related to "time" and would not change in this fixed time, we could improve our recommendation to work efficiently.

Firstly, we used the matrix factorization to explore the user purchases behavior and preference. The original application of it is to predict the movie rating, and the method is also known as collaborative filtering (Yehuda Koren, Robert Bell and Chris Volinsky, 2009). However, in 3.3, we tried to apply this method to predict how many quantities would MATSUSEI's target audience buy based on someone who had the similar purchasing behavior to them. However, we had an assumption here, when a customer buys a large amount of product, which means that he or she is favorite with it. In brief, with this assumption, we could predict the preference of our target audience.

Secondly, it's the same as Section 3.2, we only used the target customers found in Section 3.1 to fit the model. Besides, for analyzing conveniently, we transformed the data into a matrix form. The rows of the matrix represent customers, and the columns of the matrix represent products. Each cell in the matrix is the quantity of product j the customer i bought. In R, there is a package called recosystem [1] which enables us to implement the matrix factorization efficiently. It is an R interface to LIBMF [28].

Thirdly, let's see our model performance. there are several parameters such as the number of latent factors and the regularization cost for latent factors we can tune in the function. We summarized the range of tuning parameters in Table 11. We tried all the combination of the tuning parameters, and it turned out that the best combination is 30 for the number of latent factors, 0.5 for the regularization cost, and 0.01 for the learning rate. The RMSE of the model is 2.627364.

From the performance above, we think that, the potential drawback of our model was that

[1] Github: recosystem https://github.com/yixuan/recosystem

Table 11: Tuning Parameter

Number of Latent Factors	10	20	30
regularization cost	0.01	0.1	0.5
Learning rate	0.01	0.05	0.1

it applied to predict movie rating originally. The move rating was a measurement of liking of the movie. In our analysis, we used it to predict quantities. However, there were lots of factors which may influence how much a customer would buy. For example, if the product was the necessity, customers would tend to buy more on that product. This might be also a reason why our RMSE was high.

To sum up, even if the model performance above would not be excellent, MATSUSEI could still try to add this extension to their Organic Recommendation System to make recommendation work better. The reason was that if we can found more variables related to customers or got more historical data, the performance would be better. As a result, MATSUSEI could still apply with this idea. For instance, when MATSUSEI have no idea with target audience not having large patronage recently, they could assume customers' preference, or the quantities of their target audience, would not change in a fixed period, then make the personalized recommendation to them, to awake their purchasing desire.

3.4 Strategy

According to data analysis process above, we built an Organic Recommendation System for MATSUSEI, which was designed from our data analysis. At the same time, this matches MATSUSEI's corporate culture as well as their strengths in innovation[7]. To put it differently, through this recommendation system, MATSUSEI could recommend customers what they need before they think of that. It is MATSUSEI's spirit for its innovation that put its customers in its shoes before its customers need. We will illustrate our strategies in detail in the following paragraph.

Firstly, we used Non-negative Matrix Factorization to find out who must be MATSUSEI's target audience, or who will buy or be possible to buy organic products. Furthermore, from our SWOT analysis, we can identify customers who would consider high quality. Besides, we can also see who had been affected by the promotion of organic concept, or even who were afraid of food safety scandal. With these target audience, MATSUSEI did not have to worry about its biggest competitor, Pxmart with price-oriented strategy. In other words, MATSUSEI could take advantage of its differentiation between Pxmart, and tightened these high-valued customers to make them be MATSUSEI's loyal customers.

Secondly, using Sequential Pattern Mining to consider the "time" effect, or the purchasing order, we can forecast what the target audience would like to buy next time. Therefore, we can give them a personalized coupon or additional purchase discount for organic products based on different categories. For example, according to our Organic Recommendation System, we can make the targeting audience feel respected by recommend them what they need, which means MATSUSEI can maintain or improve its customer relationship management. At the same time, through this kind of Organic Recommendation System, we can sell the organic products by personalized precise marketing to make more profit on organic products. In other words, since our main goal is to attract buyers to buy green products, we will recommend them to buy related organic products with coupons, and for goods which cannot be replaced by organic products, we introduce LIBMF to predict the quantity of a customer's consumption so that we could know which kinds of product he(she) prefers and introduce them the related organic products.

Finally, for LIBMF, as we mention in last paragraph, we can use it to predict quantities of a product that customers may buy, or preference of target audience. Based on the prediction, we can also make a recommendation to customers. As we mentioned in Section 3.3, when MATSUSEI have no idea with target audience not having large patronage recently, MATSUSEI can arrange with their promotion period or anniversary or any other fixed period, to assume customers' preference will be consistent in this period, that is to say, customers' quantities of purchasing will be consistent in this period. Then, we can help MATSUSEI make the personalized recommendation to their target audience, and make direct marketing to them, by e-mail, message, and so on, to awake their purchasing desire. By doing this, these target audience will be reminded to visit MATSUSEI, and enjoy their personal selling service.

Thanks to Organic Recommendation System, MATSUSEI didn't have to be anxious about Pxmart competing with them. Additionally, they didn't have to draft a lot of budget on marketing to convey what they want to tell customers[4], because of word of mouth. In other words, when their target audience enjoyed their shopping experience, they would share with others naturally, which means their success on marketing. From above, we can see that if MATSUSEI could make some changes in the beginning of 2011, or utilizing our Organic Recommendation System, we think everything will be different. That was to say, MATSUSEI did not have to face with poor revenue problems from 2010 to 2015[3][10], and even didn't need to be merged by Pxmart now in 2016.

4 Conclusion

4.1 Further Application

On one hand, actually, we can also make recommendation to all customers and all products directly,or general recommendation. In other words, if MATSUSEI also wanted to seize non-targeting audience or also wanted to focus on non-organic product, they can still adjust our recommendation system to make other application.

On the other hand, for MATSUSEI, in the beginning of 2011, its opportunity was on "high-priced" and "quality", so we designed this kind of Organic Recommendation System for MATSUSEI. However, if future trend changed or MATSUSEI's SWOT analysis transformed, MATSUSEI could also adapt this Recommendation System to diverse situations. For example, when MATSUSEI found that they faced new threats, such as new competitors, or their original opportunities did not exist, they could also utilize this Recommendation System, and only have to adjust their target audience, and follow our whole analysis process to accomplish recommendation, that is to say, this kind of recommendation system will be an ongoing analysis. Furthermore, MATSUSEI could also make use of this application to different promotion activities. For instance, when MATSUSEI would love to promote the products of their brand in a period, they can build another recommendation, maybe called MAT-SU Recommendation System. We can give a precise suggestion about promotion in Promotion of 4P. With the concept of Integrated Marketing Communication (IMC), we can solve Sales promotion, Personal selling, and Direct marketing parts. Firstly, Sales promotion is the short-term incentive to encourage the purchase or sale of a product or service. For example, if we know a frequent pattern:

{台灣枋寮黑珍珠蓮霧,台灣A仔菜} $\rightarrow$ {銀川有機長秈白米2kg} with confidence equal to 1

During the wax apples season, we can combine {台灣枋寮黑珍珠蓮霧,台灣A仔菜} with a sales discount. Because customer who come to buy wax apples would buy 銀川有機長秈白米2kg definitely, we could use fresh wax apples to induce customer to consume. Or, during Hungry Ghost Festival(中元節), we can combine {台灣枋寮黑珍珠蓮霧, 銀川有機長秈白米2kg} to make customer purchased rice ahead of time. Secondly, Personal selling is the personal presentation by the firm's sales force for the purpose of making sales and building customer relationships. For example, we can print personalized limited time coupon or additional purchase discount after a customer's checkout. Besides, we can predict how many a customer would buy (Section 3.3)and if the customer prefer high price or low price by our Customer value in Section (Section 2.3). Finally, Direct marketing involves making direct connections with carefully targeted individual consumers to both obtain an immediate response and cultivate lasting customer relationships through the use of direct mail, telephone, direct-response television, e-mail, and the Internet to communicate directly with specific consumers. In section 2.3, we calculate the Green Value(GV), customer value on organic products, and we have the Customer Activation Index(CAI). Therefore, if our the most valuable customer is located at low CAI, we can send a special discount about a product predicted that he(she) would most likely to purchase(Section 3.2 and 3.3). Therefore, MATSUSEI could follow this kind of way to make profits on right customers to keep customers loyal to MATSUSEI.

4.2 Limitation

Limited to data, we cannot learn about the exact location of store as well as do overall data analysis for MATSUSEI without all data. However, if we can obtain more data or variables, such as location or brand of products(we tried to connect external data source, but we could not find many brand of products on the internet), we can dig more about data or through training and testing again and again, to make a greater recommendation system with exact prediction for MATSUSEI, just as we mentioned in the final part of Section3.3. Furthermore, we think that MATSUSEI had complete ERP system according to our previous SWOT analysis[1], not only with lots of historical data as well as well-developed. Accordingly, we think we can connect not only sales data, but financial or suppliers data as well. Furthermore, we cannot ignore the potential of internet market, according to Section1.2, we mentioned MATSUSEI's innovation, about which they had their "E-MATSUSEI Online Supermarket". Therefore, we can also make use of this strength to give online target audience the same Organic Recommendation System.

Last but not least, with all the basic of Organic Recommendation System, we can extend this application to any situation only have to adjust timely, then we can do deeper research for MATSUSEI to think about the best strategies to help them maintain their profit and corporate image. Not only for the beginning of 2011, but also for 2012, 2013, ..., and 2016.

References

[1] 味全60周年特刊,「訪松青商業（股）公司協理 王祿鏡提供買得到的精緻生活」, pp. 128–131, 2013, [Online].Available: http://www.weichuan.com.tw/monograph/mobile/index.html.

[2] 邱馨儀,「味全 浴火重生 兩岸爭霸」,經濟日報 , June 14, 2009.

[3] 楊雅民,「不堪虧損味全松青超市賣全聯」,自由時報, November 6, 2015, [Online].Available: http://news.ltn.com.tw/news/business/paper/930108.

[4] 陳建豪,「品質超值 兩年半弭平1億5000萬虧損 松青超市 多給20%的品質價值」,遠見雜誌,第266期, August, 2008, [Online].Available: http://store.gvm.com.tw/article_content_14367_2.html.

[5] 張寧馨,「超市個性化求生」,自由時報, July 8, 2010, [Online].Available: http://news.ltn.com.tw/news/consumer/paper/409578.

[6] 吳雨潔,「松青獨賣商品人氣推薦16款」,聯合報, February 28, 2009.

[7] 林珮萱,「松青超市用51字服務誓言灌頂員工」,遠見雜誌,第293期, November, 2010, [Online].Available: http://www.gvm.com.tw/Boardcontent_16939_1.html.

[8] 李宜蓁,「雜事外包網購用品免奔波」,親子天下雜誌,第11期, April 5, 2010, [Online].Available: http://new.cwk.com.tw/open_link.htm?type=journal&app=1&doi=K0000001-011-110-596.

[9] 楊美玲、曾桂香,「徐重仁翻轉全聯搶超市龍頭」,聯合晚報,第293期, November 6, 2015, [Online].Available: http://udn.com/news/story/7241/1297382-%E5%BE%90%E9%87%8D%E4%BB%81%E7%BF%BB%E8%BD%89%E5%85%A8%E8%81%AF-%E6%90%B6%E8%B6%85%E5%B8%82%E9%BE%8D%E9%A0%AD.

[10] 天下編輯部,「服務業行業別排名」,天下雜誌,第421期, May 6, 2009, [Online].Available: http://new.cwk.com.tw/open_link.htm?type=journal&app=1&doi=10152784-421-274-18066.

[11] 天下編輯部,「服務業營運狀況」,天下雜誌,第446期, May 5, 2010, [Online].Available:http://new.cwk.com.tw/open_link.htm?type=journal&app=1&doi=10152784-446-252-19433.

[12] 行政院主計處,「重要經濟指標」, 2010, [Online].Available: http://new.cwk.com.tw/open_link.htm?type=journal&app=1&doi=10152784-446-252-19433.

[13] 維基百科,「台灣食品安全事件列表」, 2009–2010, [Online].Available: https://zh.wikipedia.org/wiki/%E5%8F%B0%E7%81%A3%E9%A3%9F%E5%93%81%E5%AE%89%E5%85%A8%E4%BA%8B%E4%BB%B6%E5%88%97%E8%A1%A8.

[14] 有機農業全球資訊網,「有機農業法規及政策」, 2006–2012, [Online].Available: http://info.organic.org.tw/supergood/front/bin/ptlist.phtml?Category=105835.

[15] Wikipedia,“Organic food”, 2016, [Online].Available: https://en.wikipedia.org/wiki/Organic_food.

[16] 有機農業全球資訊網,「全國有機農業日系列活動快報」, 2009, [Online].Available: http://info.organic.org.tw/supergood/front/bin/ptdetail.phtml?Part=org20091014.

[17] 有機農業全球資訊網,「有機農業統計」, 2008–2010, [Online].Available: http://info.organic.org.tw/supergood/front/bin/ptlist.phtml?Category=105937.

[18] 盧昭燕,「超市一哥——全聯挖角日本大將進軍生鮮」, 天下雜誌, 第443期, March 24, 2010, [Online].Available: http://new.cwk.com.tw/open_link.htm?type=journal&app=1&doi=10152784-443-066-19231.

[19] 曾沛瑜,「賣有機，誰最貨真、價實、又在行？」, 康健雜誌, 第165期, August, 2012, [Online].Available: http://new.cwk.com.tw/open_link.htm?type=journal&app=1&doi=15603121-165-108-7589.

[20] D. J. Bem, "Self-Perception Theory ," 1972.

[21] F. Heider, "Attitudes and Cognitive Organization," pp. 107–112, 1946.

[22] A. M. Hughes, " Strategic Database Marketing," 1994.

[23] 葉進儀、吳泰熙、李凱平,「結合模糊理論及馬可夫鏈評估顧客價值」,資訊管理學報, 第十六卷, 第一期, pp. 109–134, 2009, [Online].Available: http://gebrc.nccu.edu.tw/jim/pdf/1601/JIM-1601-05-fullpaper.pdf.

[24] B. Stone, "Successful Direct Marketing Methods," 1995.

[25] R. L. Pfeifer, Phillip E; CarraWay, "Modeling customer relationships as Markov chains," in *Journal of Interactive Marketing*, 2000, pp. 43–55.

[26] 楊清潭,「應用資料探勘於顧客價值預測－以醫療器材業為例」,碩士論文, 2003, [Online].Available: http://handle.ncl.edu.tw/11296/ndltd/03342848562412362069.

[27] J. Kim and H. Park, "Sparse Nonnegative Matrix Factorization for Clustering," College of Computing Georgia Institute of Technology, 2005.

[28] M.-Y. Y. Y. Z. Y.-C. J. W.-S. Chin, B.-W. Yuan and C.-J. Lin., "Libmf: A library for parallel matrix factorization in shared-memory systems," 2015.