CUSTOMER SEGMENTATION AND ANALYTICS

ANALYTICS SPECIALIZATIONS & APPLICATIONS
2019/20

Shiqi BAI (20219140)
Bixsb3@nottingham.edu.cn

Table of Contents

Executive Summary

Our task is to segment the market based on the data set provided. It is beneficial to select the target market and formulate marketing strategies. Cluster analysis is very common in various industries, and market segmentation is its most common analysis requirement. First, we pre-process the data to ensure the validity of the data. Then through feature extraction of different aspects of the data and analysis of the overall data through these features. We got some characteristics about the overall data. Then, the commonly used K-means method is adopted for clustering and the number of clusters is selected by Silhouette value. The scientific nature of clustering is guaranteed. Finally, through a separate analysis and comparison of each cluster, we find out the characteristics of each cluster.

The result is that we divided customers into 5 categories according to the above steps: Normal Customers, Losing Customers, Loyal Customers, Big Spenders and Good Customers. Through analysis, it is concluded that Losing Customers and Big Spenders have greater development potential and profit space. Therefore, I suggest focusing marketing on the Losing Customers and Big Spenders crowd.

Feature Engineering

First, we considered the direct consumption features, including total spend, total quantity and category spend. The total spend includes min_spend, max_spend and mean_spend. These are the data that most directly characterize the consumer. Category spend is the customer's spending on each category. The second is the RFM features, namely Recency, Frequency and Monetary, which reflect the behaviour of the consumer. According to the RFM features, I divide the customers into 4 levels (1-4). Recency is the consumer's most recent spending time. I just fix the date to be one day after the last entry in the database. Recency can effectively respond to customer attrition or additions,1 on behalf of customers who have visited recently. Frequency is the number of purchases made by the customer within a limited period.1 represents the most frequent customer, i.e. the loyal customer. Monetary is also divided into 4 levels according to the amount of money spent, 1 being the largest amount of money spent. Finally, we considered the time features. Time features can reflect the cyclical habits of customers' consumption, such as a preference for what time of the week it is.

The consumption characteristics and RFM indicators between the clusters show obvious differences, and the time characteristics also show some different trends, which means that my feature selection can successfully reflect consumer behaviour and distinguish them.

Customer Base Analysis

Spend and quantity

The average spend of customers per visit is mainly concentrated at 5 to 15 dollars. The average quantity of customers per visit is mainly concentrated at 0 to 10.This means that customers tend to buy a smaller number of products of moderate value at a time.

Visits time

46.8% of customers visit 20-60 times.18.1% of customers visit more than 100 times. A large number of customers are visiting normally. But there are also a considerable number of frequent customers. These customers are potential loyal customers. At the same time, customers who have fewer visits may spend a lot of money each time. This requires specific cluster analysis later.

Category spends

The most spent type is TOBACCO and the least spent type is DISCOUNT_BAKERY. The most quantity type is DAIRY, and the least quantity type is PRACTICAL_ITEMS. It can be seen that although cigarettes are not purchased in large quantities, they cost the most. Consumption is mainly concentrated on food.

Time

Month

The data is between March and August. It can be seen from the total monthly sales that it is showing a downward trend.

Weekday

Friday's spending is the highest. The consumption on Sunday is the lowest. But Saturday's consumption is relatively high. This may be because there is more time during the weekend, so you can buy more on Saturdays, and therefore less on Sundays.

Day

Daily spending during the month showed a cyclical trend. This may be related to holidays and weekends. At the same time, spending on the 31st are significantly lower than other dates, because some months do not have the 31st.

Hour

The daily spending presents a tidal shape. With the rapid climb to the peak of 11 in the morning. Then it showed a slow downward trend until it closed at 22 o'clock.

RFM

49% of customers have just consumed within the last 2 days. But a considerable number of customers have visited within 2-10 days, these may be potential lost customers. Most customers have visited in the past two days, indicating that overall customer loyalty is high.

All in all, the company's market is mostly ordinary consumers. Consumption of tobacco and food commodities are more. The overall market sales showed a monthly downward trend. Weekly consumption is highest on Friday and Saturday. The highest sales at 11 o'clock every day. Most customers have high loyalty but average consumption frequency and ability.

Segmentation Methodology

My choice is the k-means clustering method. The k-mean clustering algorithm is an iterative clustering analysis algorithm, the step is to pre-divide the data into K groups, then randomly select K objects as the initial clustering centre, then calculate the distance between each object and each seed clustering centre, assign each object to the nearest clustering centre. Clustering centres and the objects assigned to them represent a clustering class. For each sample assigned, the clustering centre of the cluster is recalculated based on the existing objects in the cluster. This process will be repeated until some termination condition is met. Termination conditions can be no (or minimum number) of objects reassigned to different clusters, no (or minimum number) of clustering centres changed again, and the error squared and local minimum.

I will choose from the given 5-7 cluster centres. In clustering problems, Silhouette analysis is used to study the distance between clusters. The Silhouette value measures how closely points in the same class are compared to points in different classes. I chose to use the average Silhouette value to select the number of cluster centres. The average silhouette value of 5-cluster-center is 0.11, which is larger than other values. Therefore, the number of cluster centres is determined to be 5.

Results

According to the indicators of clustering results, we classify customers into 5 categories: Normal Customers, Losing Customers, Loyal Customers, Big Spenders and Good Customers.

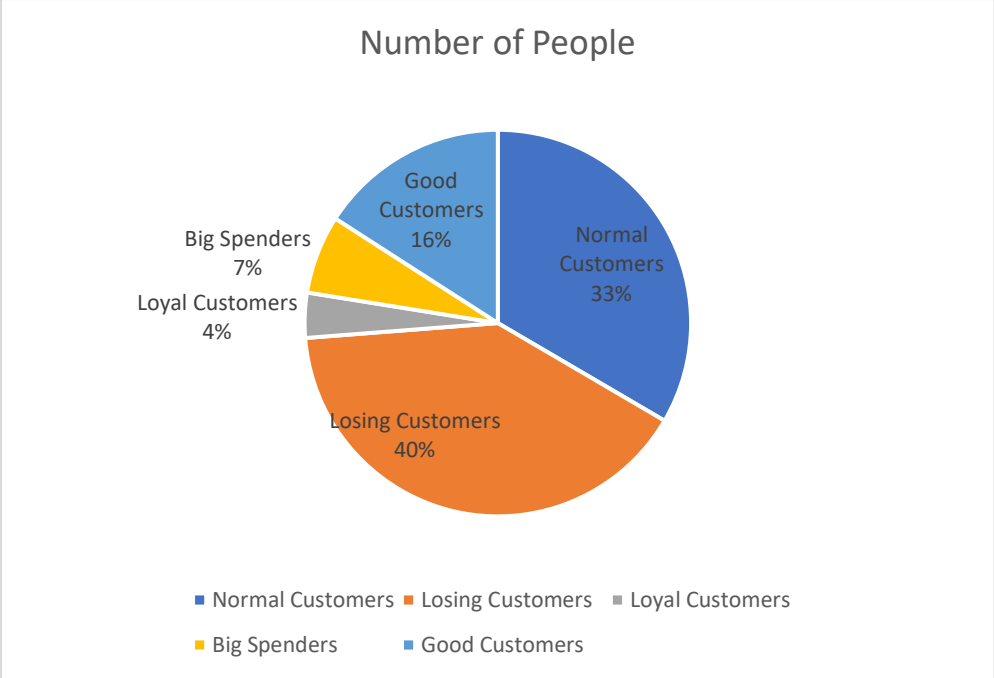

Figure 1: Number of people for each segment

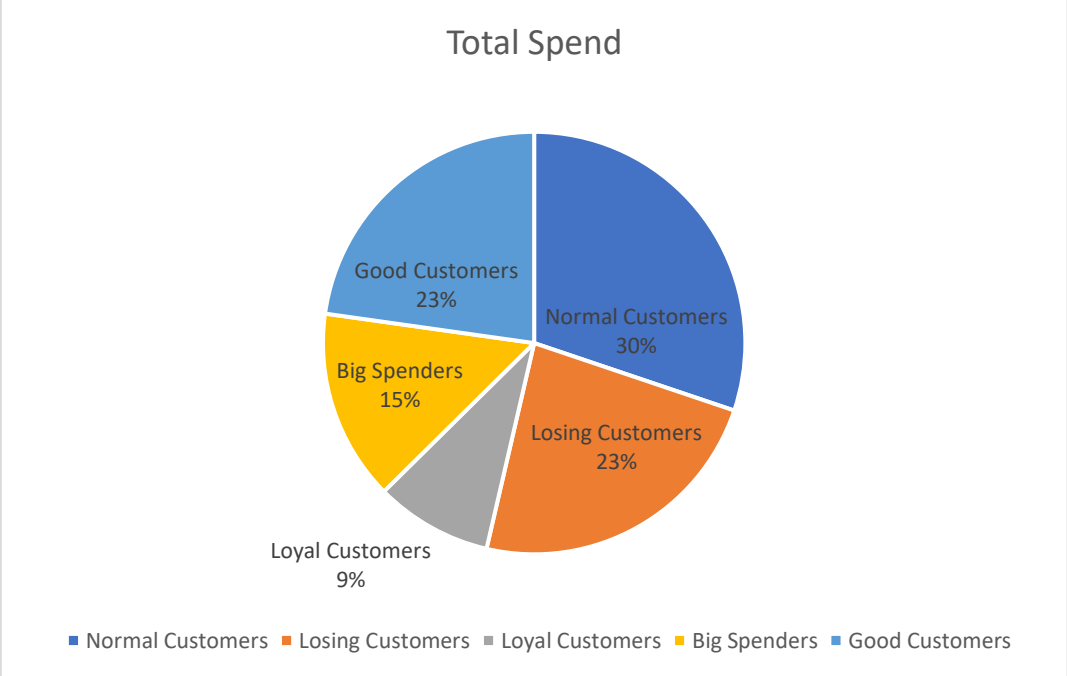

Figure 2: Total spends for each segment

The following is the Key Figures table of each cluster.

	Normal	Losing	Loyal	Big	Good
Total spend	698042.5	542389.5	207485.4	338521.7	526703.7
Min Basket Spend	1.49	3.82	0.87	4.34	1.10
Max Basket Spend	37.71	45.13	39.99	91.90	38.61

Mean Basket Spend	10.64	17.87	8.84	34.62	9.15
Average_spend	695	448	1836	1718	1104
Recency	3.85	11.09	0.58	3.38	1.70
Frequency	70.38	32.33	224.36	69.30	126.76
RFM Most	222	444	111	221	111

Table 1: Key figures

Normal Customers have the highest total spend, followed by Losing Customers, Good Customers, Big Spenders and Loyal Customers. Loyal Customers have the least consumption. Average_spend is the total spend per capita. Loyal Customers and Big Spenders have the highest average_spend. In terms of Recency and Frequency, Loyal Customers have the Best performance. Losing Customers` Frequency is very low, which means they have not patronized for a long time. The highest frequency RFM indicators for each cluster are: 222, 444, 111, 221, 111. Loyal Customers and Good Customers performed well, and Losing Customers performed poorly.

Normal customers

The number of normal customers' accounts for about one-third of the total. Normal Customers' RFM scores are mainly 2 or 3. The key figures are in the middle of the five clusters. This means that their spends, recent visit time and visit frequency are close to the average. Normal customers mainly consume tobacco, food and drinking products. Among them, tobacco consumes the most. In terms of time, the number of purchases per month and the overall spend amount are consistent with the overall trend, which also shows a significant downward trend. Friday and Saturday have more spends and quantities. Every day is also 11 o'clock as the peak period.

Losing customers

The number of Losing Customers accounts for about one-third of the total. Losing Customers' RFM scores are mainly 4 or 3.Although their consumption amount is not low, they have not patronized for a long time and the consumption frequency is very low. This means that we are very likely to be losing this part of the customer, that is, they are potential lost customers. Losing customers mainly consume food and drinking products. They consume the most fruits and vegetables while tobacco consumption is small, probably because although it is not attractive here, it is more convenient to consume daily necessities (maybe close).The time characteristics are not much different from normal users, but the monthly decline is even greater.

Loyal customers

The number of loyal customers is small. Loyal Customers' Frequency scores are almost 1. This means that they come often and are loyal customers. In the key figures, average spend is low, but Mean Basket Spend is high. This means that they buy less each time but buy more times. Recency and Frequency are in the top of the five clusters. Although their loyalty and spending power are high, due to the small number of people, the revenue generated is very low. Loyal customers mainly consume tobacco, food and drinking products. In addition, CASHPOINT is in the fourth position of consumption spends. This may be caused by a customer buying more of this item, and the overall consumption ability of this cluster is weak. Because of high loyalty, consumption has no obvious monthly downward trend.

Big spenders

The number of Big Spenders accounts is small. But Big Spenders' Monetary scores are almost 1. This means that their spending power is enormous. In the key figures, Recency and Frequency indicators are on average but the spends are high. Big Spenders spend most on GROCERY_HEALTH_PETS, not TOBACCO and foodstuffs. It may be because this part of consumers is a high-income group, and there is a difference in consumption habits from ordinary customers. Their monthly consumption shows a volatile trend, but not a clear downward trend. This is a good signal, which means that we can take action to explore their consumption potential.

Good customers

Good Customers` total spend and Monetary scores are high. Good Customers' spending power is close to Loyal Customers, but the loyalty (RF indicator) falls somewhere in between Loyal Customers and Normal Customers. Good Customers mainly consume tobacco, food and drinking products. There is also some non-food consumption. This is different from the consumption habits of Normal Customers. Good Customers` monthly consumption shows a declining downward trend. This is a bad signal, which means that we are losing them.

In short, Normal Customers are those with low spending power and average loyalty. Losing Customers are potential churn customers with higher spending power and lower loyalty. Loyal Customers are small part customers with high frequency and loyalty. Big Spenders are high-level customers with high spending power and average loyalty. Good Customers are mid-level customers with average spending power and high loyalty.

Recommendations

After feature extraction and clustering of customers, we divided customers into 5 clusters: Normal Customers, Losing Customers, Loyal Customers, Big Spenders and Good Customers. We find that Losing Customers and Big Spenders have great consumption power. For these two clusters, I have the following two suggestions. For Losing Customers, we can take targeted cash issuance activities to try to retain them. Because such customers' consumption power is not weak, but they are gradually flowing to competitors, so strong preferential measures are needed to attract these customers. For Big Spenders, such customers have strong spending power, and we can use high-quality services and membership policies to further tap their spending potential.