Using RFM model to help mobile phone dealer to find their target customers in China

Junyang Wang *

School of Liberal Art, University of Minnesota, Twin City, United State

*Corresponding author: Wang9546@umn.edu

Abstract. Target customer positioning is the first step for enterprises to develop market strategy. After the completion of user positioning, through the analysis of specific groups, we can accurately know the consumption habits and thinking process of users, which will bring great help to the precision marketing of businesses and reduce operating costs. This paper focuses on using RFM model to help Chinese mobile phone dealer to find their target customers and create a portrait of the target consumers. This paper analyses 263352 different customers with 10 different brands and discusses the valuable cell phone consumers. The result shows Guangdong, Shanghai and Beijing have the highest sales, volume, and number of customers. The top five mobile phone sales are Samsung, Apple, Xiaomi, Huawei and OPPO, among which Samsung accounts for more than half of the share, and Apple accounts for about a quarter. The analysis found that the purchase of Samsung and Apple phones by all age groups and by men and women was relatively average. It was also found that Guangdong, Shanghai, and Beijing accounted for more than 50% of the sales. The percentage important value customers accounts for about 50% of the important customers, which is higher than the average percentage of other e-business products. Despite the sluggish sales figures, mobile e-commerce is still an industry with great potential.

Keywords: RFM model; Target customer; Portrait of the consumer; Customer segmentation; Pareto principle.

1. Introduction

Online purchases in the UK were predicted to have totaled 50 billion in 2011, a more than 5000 percent increase from the year 2000, according to the Interactive Media in Retail Group (IMRG). This extraordinary rise in online sales suggests that customers, purchasing and using habits have fundamentally changed [1]. With the development of technology and Internet, More and more people choose to use e-commerce websites for shopping, and mobile phone as one of the most important electrical devices for people will also be an important part of online shopping orders. One of the key success factors in the mobile phone market is a quick and detailed understanding of consumer needs. In the face of fierce market competition, mobile phone manufacturers must understand the characteristics of consumers and analyses their purchasing behavior. So, it is valuable to find the features of the target customers and create a portrait of them for the online mobile phone dealers.

RFM is the common model for sealers to find their potential and target customers. By delivering accurate and timely data to companies operating in the retail industry, Anitha and Patil employed this technique and the K-Means algorithm to discover potential clients [1]. Real-time trade and retail data sets are examined in order to execute and put scientific approaches to use. The dataset,s values and parameters offer a systematic understanding of consumer purchasing trends and behavior across geographies over the course of a particular commercial transaction. His study not only increases sales and profitability but also gives me the idea to construct a customer portrait and offers clever insights into predicting client behavior.

Pavel, from University of Economics, used RFM model to make customer segmentation purposes and to hit customers with high probability of repurchase. He proposed an innovative way to model RFM not only with purchase data, but with website visits and social network interactions as factors of valuable customer activity to predict user behavior. His way to integrate data from individual customer interactions on Facebook and Twitter into the RFM analysis can help to understand how to use RFM model to predict customers behavior.

A loyalty program is described by Leenheer&Bijmolt as "an integrated system of marketing actions, which aim to increase member customers, loyalty." [3]. To benefit from the reward program, a consumer must sign up and show his loyalty card at every opportunity to make a purchase [2]. Advanced data mining tools have made it simpler and more practical to segment clients and create more effective loyalty programs in recent years. Customers, buyback behavior and frequency are factors in customer loyalty, according to authors like Kandampully&Suhartando.

2. Methodology

2.1 Data source

The data used in this paper is from model whale [4] and China Bureau of Statistics. In the data file, 263352 different customers purchase information in 2020 who have brought mobile phone or electrical devices from Ali buy in China can be obtained. The population information from different provinces and their average income in 2020 are also included

2.2 Data clean

This study used python IDLE to process the data. The data has already arranged by user ID, order ID with time, order information and customer information. The first thing is to clean the data, which means to outliers and inaccurate values are removed to generate the initial dataset [5]. The data entry initially consists of 12 components, such as user ID, product ID, quantity purchased, and consumption date. Then, this study use “date” function creates new columns, including--date, month, hour, day of week, local and sex in order to compare the different customers feature and summarize the law of consumer purchase (see Table 1). After some new columns are added, the missing data and duplicate data are checked by reviewing the data output, and find the columns with “null”, and duplicating information.

Table 1. Part of new columns of date information

User_id	age	sex	local	date	month	hour	weekday
1515915625441993984	24	F	Hainan	0424	4	11	5
1515915625441993984	24	F	Hainan	0424	4	11	5
1515915625447879424	38	F	Beijing	0424	4	14	5
1515915625447879424	38	F	Beijing	0424	4	14	5
1515915625443148032	32	F	Guangdong	0424	4	19	5

From the table 1, it is obvious to find that there are some duplicate values, but from another perspective, these duplicate values are orders of multiple quantities placed in the same order, so the duplicate values are not deleted, and then a column of purchase quantity and total price is added for the statistic.

Table 2. Statistic values of the data

Variables	Count	Mean	Std	Min	1%	25%	50%	75%	Max
Price	536311	214.54	305.98	0	1.13	24.51	99.51	289.33	11574.05
Age	536311	33.18	10.12	16	16	24	33	42.00	50.00
Month	536311	7,72	2.56	1	1	6	8	10	11
Hour	536311	9.52	4.21	0	1	6	9	12	23
Weekday	536311	3.03	2.04	0	0	1	3	5	6
Buy_cnt	536311	1.00	0.04	1	1	1	1	1	4
Amount	536311	214.73	306.48	0	1.13	24.98	99.51	289.33	11574.05

The second step is to check whether the data is abnormal or redundant. The data is changed into a format that facilitates customer value analysis easier and more effective by removing unnecessary attributes[5]. The format of the data is changed in order to find if there are some problems with them. The id format is converted to “object” format and convert hour and weekday to “int” format. And then, the mean and stander deviation of them are calculated (see Table 2) in order to check if there are any outliers in prices and ages.

From Table 2, there are no outliers in the above 7 fields. The reason why price and amount have a minimum value of 0 is that these items should be the free items or gifts, so they are not outliers. It should be further analyzed whether the users who bought the goods with 0 value also bought other goods later.

After cleaning the data, this study can summary the customer information by Gross merchandises value(GMV), and unit price. The total GMV can be counted by summing each customer prices up and uses the total GMV over number of users, ID to get the unit price per person. The total GMV is about 0.115 billion RMB, and the unit price is 1240 RMB. In this way, this study summarizes the GMV in different month from January to November (see Figure 1). From the Figure, the GMV was basically on the rise until August, and the rise was even bigger in July and August, and after it peaked in August, it started to decline. This Figure also shows that most amount of Chinese people prefer to buy smart phone in August in 2020.

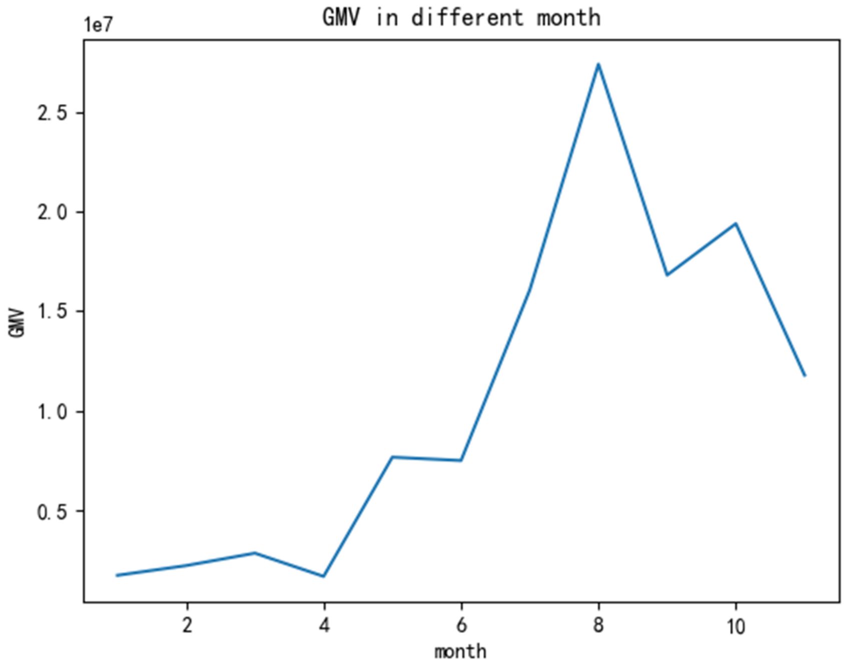

Fig. 1 GMV in different month in 2020

2.3 Customer Analysis

For customers, this study fucus on finding the online smart phone customers features including local, sex, ages, and times of purchase. One common fact is that not all collected features data useful or are genuine features. It is important to feature pruning, which aims to remove these incorrect features [6]. This study will compare the amount and quantity of sales under different independent variables of customer to draw a portrait of the mobile consumers in China and give some suggestion to the online mobile phone dealers.

2.3.1 Local Analysis

First step is to categorizes consumers by province (see Figure 2) and compared the number with the population of their province to find whether they have more potential customers. The province has the largest number of customers does not mean that the mobile phone is popular in that area. One more step is to count the percentage of users in the total population and the percentage of sales in that province in the total incomes to find the most cost-effective provinces and the most promising province which has the high percentage in customers but not high percentage in their incomes.

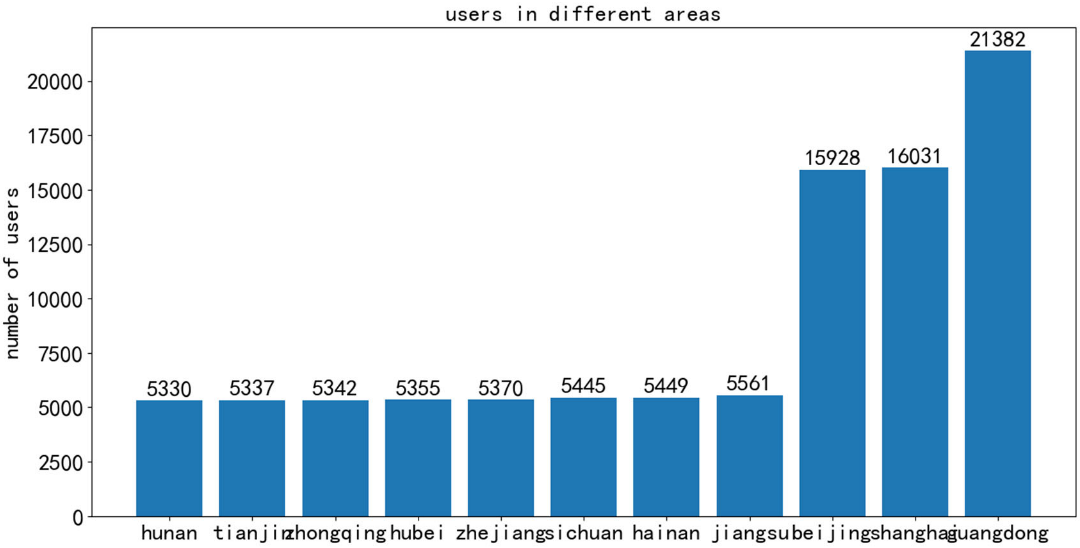

Fig.2 Number of customers in different areas

From the Figure 2, it is obvious that Guangdong has the largest number of users, followed by Beijing and Shanghai, and the other eight cities have an average number of users, all-around 5,400. The area has the large number of customers also has the large number of incomes. According to the national census in 2020, the population data of each province was obtained online, and the proportion of users in each province was analyzed to see which provinces could still attract new users. In terms of the number of users, the top three are Beijing, Shanghai and Hainan, the fourth is Tianjin, and Guangdong ranks the fifth, with only one fifth of the number of users in the first place, Beijing. In addition, Guangdong has the largest population, so the cost performance of new attraction activities in Guangdong is the highest.

2.3.2 Gender Analysis

Then, this study counts the gender information of consumers to confirm whether smartphone consumers have a gender preference (see Figure 3). From the Figure, the users are evenly split between male and female so that the gender will not be the effective feature for Chinese phone customers and the analysis in the next steps will not consider the effect of people in different genders.

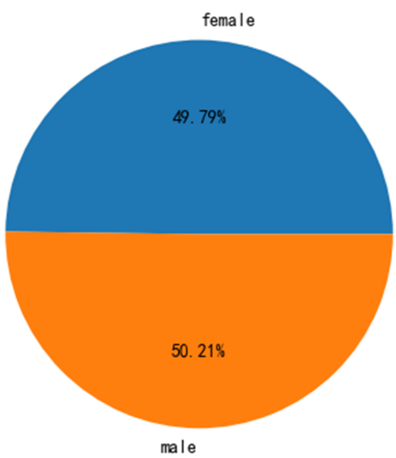

Fig.3 Genders of consumers

2.3.3 Age bins Analysis

The common view is that people in different ages have the different consumption preference. Andrew proofed that the closer in age the two cohorts are, the higher the correlation between the percentage changes in average consumptions is [7], which means people in the similar age group have the high probability of having the similar consumption habits. due to this theory, this study creates age bins in every 5 years, and the number of consumers in each age group was counted in Figure 4, in order to check whether the mobile phone has the ages preference. From the figure 4, the age distribution of users is relatively average, ranging from 16 to 50 years old. Then, this study counted the order quantity and total amount of each age group to confirm whether different age groups have different consumption expectations in purchasing mobile phones (see Figure 5).

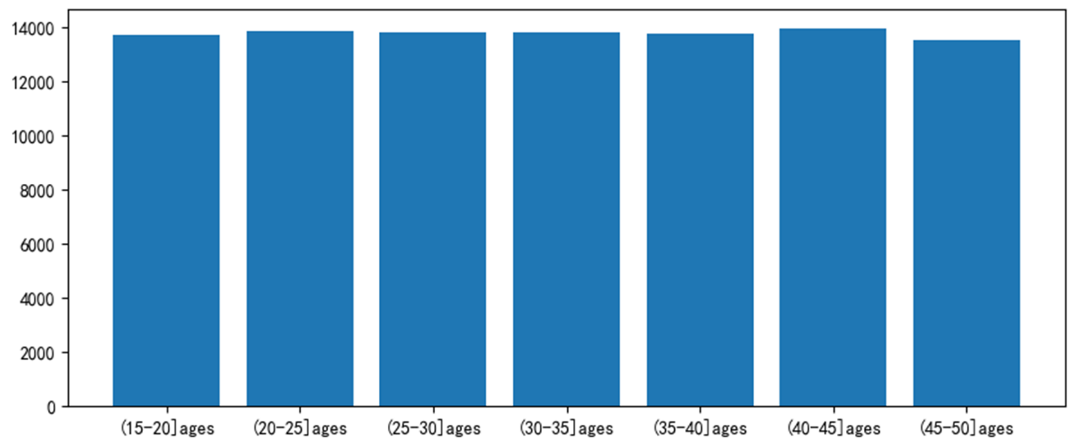

Fig. 4 Number of consumers in different age bins

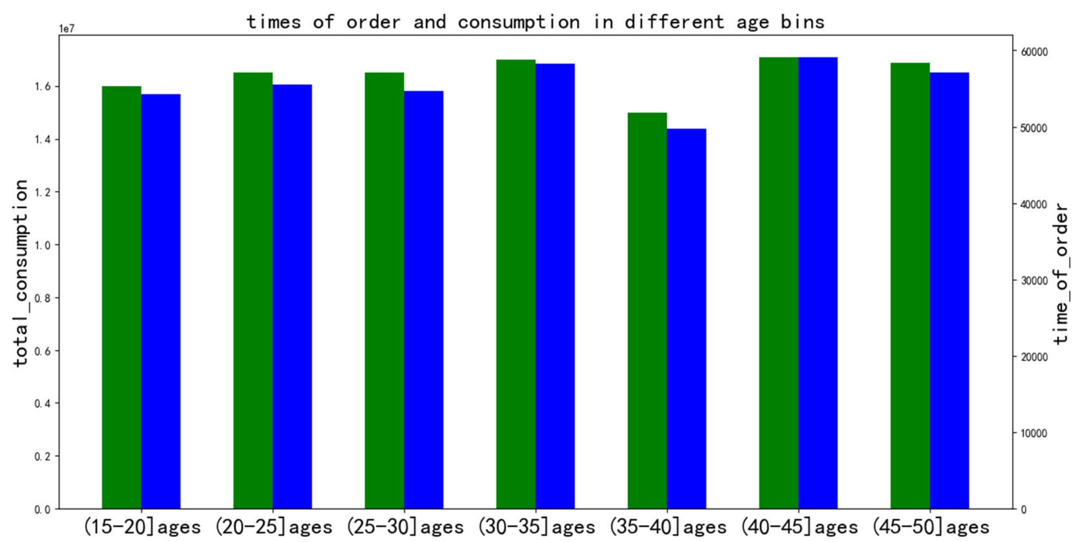

Fig. 5 Times of order and total consumption in different age bins

From the Figure, it is found that users aged 35-40 have the lowest amount of contribution and the number of orders, while other age groups are relatively average. Based on Figure 4 and Figure 5, for the e-commerce consumption of mobile phones, people have no age or income preference. In other words, the commodity attributes of mobile phones and people,s dependence on them make their consumers come from all ages.

2.3.4 Customer Dependence

According to the time when customers place orders, this study divide consumers into active and inactive consumers. Customers who place orders this month automatically become active users and become inactive users after not placing orders for 2 months. If an inactive user orders again, it will become a returned user. Finally count the number of different consumers (see Figure 6). As can be seen from the figure, the number of inactive users is accelerating with the improvement of platform users, while the number of returned users is always small, because the replacement cycle of most people,s mobile phones tends to be long-term. Therefore, attracting more new users is more important and urgent for the mobile platform than recovering old users.

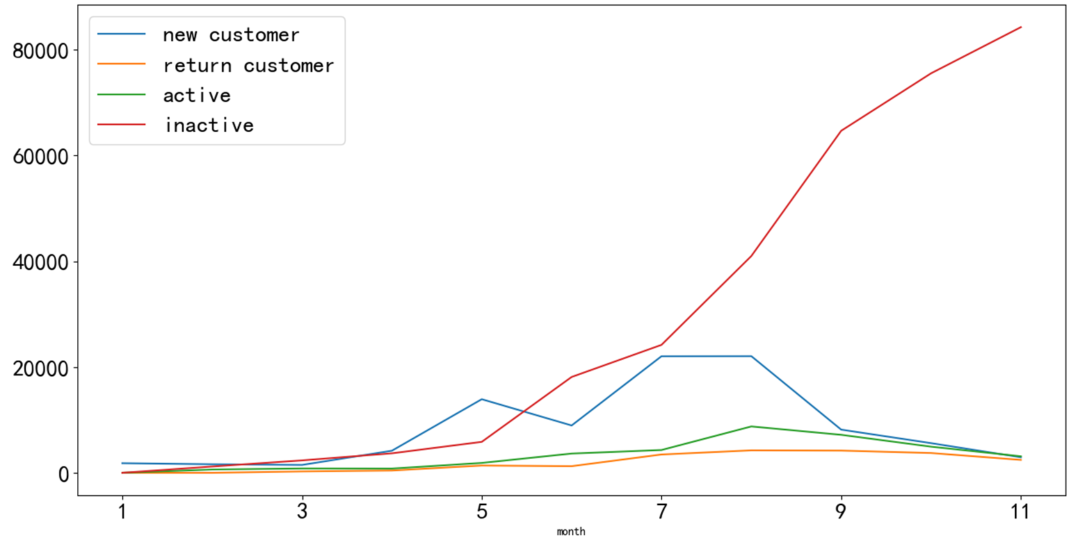

Fig. 6 Number of active and return customers

2.3.5 Gift and Pareto effect

Gift effect, which means a small concession in the interests of producers or dealers can lead to larger economic effects in the future. In the area of smart phone, the concession can be a free gift or service. From the data, it shows on the 0 value products counted before. As can be seen from the product category, the product worth 0 yuan should be the one who won the lottery. Among the 30 users who won the lottery, only 1 user did not buy other products, and the customer unit price of these 30 users was as high as 34,832 yuan (the total customer unit price was 1,240 yuan).

Pareto principle is also useful to find the most important customers. The 80%/20% rule states that for many outcomes, around 80% of consequences result from 20% of causes. In business, this means that 80% of sales result from 20% of customers. In the collected data of smart phone sales, the top 27% of users contribute 80% of the sales revenue. This group of users should follow up well and be sure to retain these big contributors.

2.4 Product Analysis

As the development of technology and manufacturing, smartphones of different brands are gradually converging in functions and product positioning. When "one face of a thousand machines" becomes the trend of The Times, the brand effect becomes the only way to occupy the minds of users, and the brand difference is the ultimate destination for the industry to mature. In recent years, various analyses have all pointed to the fact that young consumers have a more pressing desire to connect emotionally with brands and more emotional spending propositions, and often blessed by a small number of brands in the same category. McKinsey released a report shows that: nowadays, there are more and more young consumers begin to focus on only a few brands, unwilling to buy other brands besides their own attention, and even some young people willfully lock on to a certain brand.

2.4.1 Brands Analysis

For products, it is useful to analysis different brands, and their customers feature in different brands. this study considers the five most popular mobile phone brands for customers in China (see Figure 7), and compares the Apple and Samsung, which are the top 2 popular brands, customers information to find if they have the similar features such as local, age and sex information.

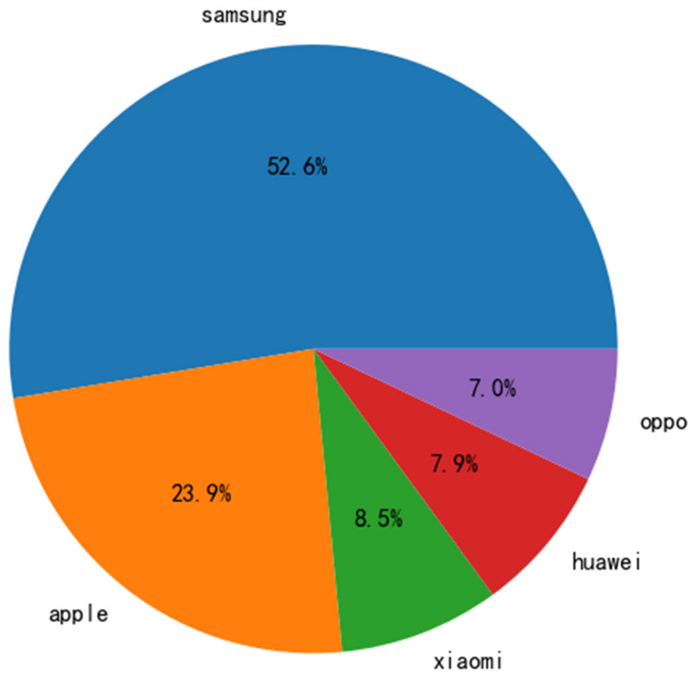

Fig. 7 Market share of top 5 mobile phone brand in China in 2020

In Figure 7, the top five sellers are Samsung, Apple, Xiaomi, Huawei and OPPO, with Samsung accounting for more than half of the market share and Apple accounting for a quarter. For hot selling products such as Apple and Samsung, we should always pay attention to their inventory to avoid out of stock situation. From all bands in data, there are 12069 products whose sales volume is less than 10. For this batch of products, we can consider promotion activities to clear the stock.

2.4.2 Compare Apple and Samsung customer

Consider the customer features in the top 2 brands of the mobile phones, Apple and Samsung. This study research the statistical data of their customers local, sex and ages. From the figure, the age distribution of both two brands is fairly even and it is also evenly split between men and women Guangdong, Shanghai and Beijing accounted for more than 50% of the sales.

Similar consumer groups prove that the brand effect is another characteristic independent of consumer location and age. Like the sentences from Jun Lei, the CEO of xiaomi, if you want your product not to be seen as a tool for continuous use, and you want to further increase the psychological distance with your users, you have to make the brand itself provide additional value to users just like the product.

2.5 Sales Analysis

For sales, this study records the number of sales in different dates and compares sales and sales volume of new and old customers, in order to find the factors to influence the customers choice. From January to November, Sales rose from January to August, but began to decline after August. And the sales volume is generally in line with the trend of sales. The data show that during China,s shopping festivals (June 18 and November 11), the data fluctuates severely and does not increase significantly, indicating that mobile phones are still an important choice for people,s holiday gifts, or the promotion of mobile phones is not enough.

And this study counts the Monthly repurchase rate, which equal to Number of users who purchase more than 2 times in the month/total number of buyers (multiple purchases on the same day are regarded as one purchase), to define when and why people,s preference of repurchase a new smart phone. Then, this study considers the old customers who have purchased more than one times and compared the rate of purchase to the new customers (see figure 8). The fact is that basically every month, the contribution of new customers is greater than that of old customers, which means people buy new phones infrequently, and existing users are not platform dependent.

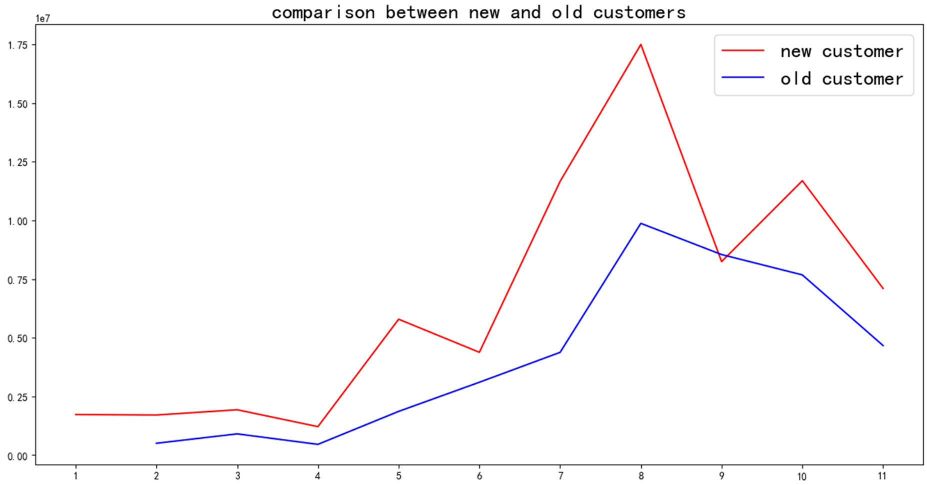

Fig. 8 Comparison between new and old customers

2.6 RFM Model

A fantastic way to discover groups of clients that should receive special care is through RFM segmentation.RFM analysis enables marketers to communicate with individual client clusters in a way that is much more in line with their behavior.Thus, this study results in significantly greater response rates, enhanced customer loyalty, and increased client lifetime value. Despite the fact that there are various methods for segmentation, RFM analysis is well-liked for three reasons:

• The RFM model makes use of objective, numerical scales to produce a clear and detailed high-level representation of customers.

• Since marketers can utilize it efficiently without the assistance of data scientists or complex tools, it is straightforward.

• This segmentation method produces results that are simple to comprehend and analyze.

The most common type of activity is a purchase, while other types are also utilized, such as the most recent website visit or mobile app usage. In most circumstances, a customer is more likely to respond to communications from a business the more recently they have interacted or made a purchase from it.

Customers who participate in activities frequently are undoubtedly more interested and devoted than those who do so infrequently. Customers who purchase something once are in a special category.

This component, which is also known as "monetary value," represents how much a customer has spent with the brand during a specific time frame. Customers who spend a lot deserve different treatment than those who spend little. When segmenting clients, the average purchase value is an important secondary aspect to take into consideration by dividing the amount by the frequency of purchases.

2.6.1 Customer Segematation

According to the different value in R, F and M, customers can be divided into 4 different groups. Customers in the first group has high value in R which means they had spent more time since their last purchase. The last purchase is made earlier, the total purchase amount is smaller, the purchase frequency is lower, and the purchase data they leave on the platform is less easily noticed. Moreover, the number of such customers is relatively small. It is considered as a customer at risk of loss and requires further observation. Some resources need to be invested in the platform to further analyze and understand these customers [5].

The clients in Group 2 are the most significant ones because both their buy frequency and total quantity are higher than the average, and their most recent purchase was made only a short while ago. They represent a group of high-value clients and increase the platform,s cash flow and earnings. The platform should work more to keep in touch with users and develop their relationship with them. In group 3, purchases are made less frequently and overall, and the final sale closes sooner than is typical. This indicates that despite their recent platform purchases, they have not yet formed a spending habit or the capacity to significantly increase platform revenue.Of these two types of customers, the first has the strongest ability to create value for the enterprise, and the second has strong consumption potential although its temporary ability to create value is not as good as the first type of users. Therefore, these two types of users are the users whose needs need to be considered and met with the most emphasis.

Customers in Group 4 most recently made a transaction not too long ago. Their metrics for average purchase quantity and frequency are relatively close. These clients use the site frequently and have developed particular consumption patterns. They should be regarded as high potential clients because there is much space for development in the total quantity and frequency of their purchases. They should be promoted from group 4 to group 2, which makes more purchases and spends more money, as part of the platform,s marketing strategy[5].

2.6.2 Experimental Design

Table 3. The standard of RFM score

Score	R	F	M
5	(0.30)	5	Above 2000
4	[30,60)	4	(1000,2000]
3	[60,90)	3	(500,1000]
2	[90,120)	2	(200,5000]
1	Above 120	1	(0,200]

Table 4. Part of RFM result of each customer

Order	User_id	R	F	M
0	1515915625439951872	0	0	0
1	1515915625440038400	1	1	0
2	1515915625440051712	1	1	1
3	1515915625440099840	1	1	1
…	…	…	…	…
92753	1515915625514891008	1	0	1
92754	1515915625514891264	1	0	0

Customer R, F, and M values are determined through RFM analysis, and these metrics are used to define the company,s proposal cluster. (R) is the most recent capital letter, which refers to the time of last purchase. (F) represents the number of purchases, (M) represents the currency and represents the total cost to the customer [8].

Consider to the mean, medium, and standard deviation of each data, we give the score of each customer on R, F and M follow the standard in Table 3.

Then, this study focuses on counting the mean score of each customer in R, F and M. For customer has the score is greater than the mean value, it is recorded as 1 on that, else it will be recorded as 0. In the end, all customers that have 111 in all RFM model are the most valuable customers, and the customers who have 101 in RFM model is important development customer, and those who have 011 in RFM model are important retention customers (see Table 4).

3. Result

For customer perspective, at present, Guangdong has the largest number of customers (21,382), followed by Shanghai (16,031) and Beijing (15,928), while the other eight cities have the average number of customers (about 5,400). According to the 2020 census, Guangdong has the largest permanent resident population in China (120 million people), with customers accounting for only 0.017%. The proportion of the highest customers is Beijing (0.0728%), which is about 4 times that of Guangdong, the cost performance of increasing investment in new activities in Guangdong is the highest. The current clientele is evenly split between men and women and is generally evenly split between 16 and 50 years old. The analysis shows that the consumption number of men and women is basically the same as the number of orders. Users aged 35-40 have the lowest consumption amount and the number of orders, while other age groups are more average. The top 27% of users contribute 80% of the sales revenue, and the average consumption number of customers is greater than the 75% quantile (there are high consumption users). This group of users should be well followed up, and we must retain this group of customers with large contribution. Customers who consume twice or more have 50% consumption cycle within 7 days and 75% within 26 days, which is a short consumption cycle.

For product perspective, for the top 10 and top 10 products by sales, their inventory should avoid shortages. For products that sell less than 10 units, a promotion should be considered to clear the stock. The top five mobile phone sales are Samsung, Apple, Xiaomi, Huawei and OPPO, among which Samsung accounts for more than half of the share, and Apple accounts for about a quarter. Through the analysis, it can be found that the purchase of Samsung and Apple phones by all age groups and by men and women was relatively average. It was also found that Guangdong, Shanghai and Beijing accounted for more than 50% of the sales.

For sales perspective, the sales volume and sales volume showed an upward trend from January to August, reaching the peak in August, and increased significantly in May, July, and August. The number of new customers increased in these three months, and the range was also quite large, which proved that the new recruitment activity was effective. In September, sales and new customers began to decline, with new customer orders down by three-quarters, proving that the recruitment campaign in September was not efficiency.

Understanding different types of clients is critical for businesses to make decisions that will increase their profitability [9]. From RFM model, there are 44,129 important customers. Among them, there are 20,012 important value customers, accounting for about 50% of the important customers. For important value customers, key services are needed, VIP-style services should be given to them, and their purchase feedback should be paid attention to. There are 4415 important development customers. These customers need to be maintained. Their consumption frequency is not high enough. We should find ways to increase their consumption frequency, such as issuing full reduction coupons. There are 11,368 important customers, and the last purchase time of these customers is far away from now. We should send them information or contact them by phone, and we can also issue the full volume reduction to them, to wake them up. For 8334 important retention customers, we need to find a way to recover this batch of nearly lost customers and one of the most effective ways to analyze product categories and identify the best solutions is to use RFM models to artificial and real datasets for consumer segmentation analysis [10].

4. Conclusion

Customer segmentation based on lifetime value or value has taken the position of segmenting customers based on needs. The customer lifetime value is comprised of the estimation of the customer,s future worth as well as the computation of the customer,s past and present value. This study based on the data from Ali buy which is one of the largest e-commerce websites to help online mobile phone dealers to see the features of the value model rather than needs. Since smartphones have gone through the period of rapid growth, and the benefits of finding valuable long-term customers and increasing their dependence on the platform are greater than the benefits of reinventing new users. Customer retention has always been a major issue and a tough challenge for businesses. And apply RFM models to synthetic and real datasets to analyse customer segmentation is one of the most efficient way to analyse specific categories of products and find optimal solutions. This research verifies the feasibility of the RFM model in dividing the types of consumers in the smartphone market and generalizes characteristics such as region and age of customers based on customer value. The general characteristics of mobile phone consumers are summarized to provide reference and suggestions for Chinese mobile phone dealers.

Knowing customeris not an easy task, and marketers must study their target customers, desires, perceptions, preferences, and purchasing behaviours. Although the sales of mobile e-commerce have been sluggish in the past two years, the successful combination of smart phones, as a necessity of future life, and e-commerce platforms will be the trend in the future. Although it still has issues such as security and privacy, we can predict a huge potential market through the RFM model and consumer profiling.

References

[1] Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University - Computer and Information Sciences, 34(5), 1785–1792.

[2] Gómez, B. G., Arranz, A. G. &Cillán, J. G. (2006). The role of loyaltyprograms in behavioral and affective loyalty. Journal of ConsumerMarketing, 23(7), 387-396.More references.

[3] Leenheer, J. &Bijmolt, T. H. A. (2008). Which retailers adopt a loyaltyprogram? An empirical study. Journal of Retailing and ConsumerServices, 15(6), 429-442.

[4] Dame. Model Whale. (2021). Electronic product sales analysis. https://www.heywhale.com/mw/dataset/601e971ab233440015800bc7/file.

[5] Wu, J., Shi, L., Lin, W.-P., Tsai, S.-B., Li, Y., Yang, L., & Xu, G. (2020, November 19). An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm. Mathematical Problems in Engineering. Retrieved September 27, 2022, from https://www.hindawi.com/journals/mpe/2020/8884227/.

[6] Hu, M., & Liu, B. (2018, October 17). Mining opinion features in Customer Reviews. Microsoft Research. from https://www.microsoft.com/en-us/research/publication/mining-opinion-features-in-customer-reviews/.

[7] Abel, A. B., & Kotlikoff, L. (1988, January 1). Does the consumption of different age groups move together? A new nonparametric test of intergenerational altruism: Semantic scholar. undefined. Retrieved October 7, 2022, from https://www.semanticscholar.org/paper/Does-the-Consumption-of-Different-Age-Groups-Move-a-Abel-Kotlikoff/8acf867b359b5b73d889c49ee1a1134eb02f78e8.

[8] Doğan, O., Ayçin, E., & Bulut, Z. A. (1970, January 1). Customer segmentation by using RFM model and clustering methods: A case study in retail industry: Semantic scholar. undefined. from https://www.semanticscholar.org/paper/CUSTOMER-SEGMENTATION-BY-USING-RFM-MODEL-AND-A-CASE-Do%C4%9Fan-Ay%C3%A7in/11f8e998bdd759a8c576bad4e1cd66388a168920.

[9] Khajvand, M., & Tarokh, M. J. (2011, February 22). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. Procedia Computer Science. Retrieved October 8, 2022, from https://www.sciencedirect.com/science/article/pii/S1877050911000123.

[10] Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University - Computer and Information Sciences, 34(5), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011.