

User Behavior Path Analysis Based on Sales Data

Wangdong Jiang, Dongling Zhang*, Yapeng Peng, Guang Sun, Ying Cao and
Jing Li

Hunan University of Finance and Economics, Changsha, 410205, China

*Corresponding Author: Dongling Zhang. Email: 1571393048@qq.com

Received: 10 February 2020; Accepted: 10 February 2020

Abstract: With the rapid development of science and technology and the increasing popularity of the Internet, the number of network users is gradually expanding, and the behavior of network users is becoming more and more complex. Users' actual demand for resources on the network application platform is closely related to their historical behavior records. Therefore, it is very important to analyze the user behavior path conversion rate. Therefore, this paper analyses and studies user behavior path based on sales data. Through analyzing the user quality of the website as well as the user's repurchase rate, repurchase rate and retention rate in the website, we can get some user habits and use the data to guide the website optimization.

Keywords: User Behavior Path Analysis; visualization; conversion rate

1 Introduction

User behavior path analysis is an analysis method that monitors the user flow and statistics the depth of product use. It mainly analyzes the circulation rules and characteristics of each module in the App or website according to the click behavior log of each user in the App or website, and mines the user's access or click mode, so as to realize some specific business purposes. User access path analysis is a very important part of website analysis [1,2]. By analyzing user access path, we can help specific visitors to improve the efficiency of completing the visit tasks in different stages on the premise of achieving the business goals of the website [3,4].

The primary purpose of user access path analysis is twofold. The first purpose is to fulfill the visitor's task on the premise of achieving the business goals of the website. Generally speaking, a website has only one business goal, while visitors may have multiple tasks during their visit [5,6]. For example, a website's business goal is to make money from users downloading its documents, so helping users find the documents they need faster and more accurately is the purpose of user access path analysis [7,8]. The other main purpose of user access path analysis is to optimize the business objectives of the site, so as to improve the efficiency of users in completing access tasks [9,10]. This goal builds on the previous one.

This paper mainly studies the user access path when accessing the site, by analyzing the behavior way of user access to web users of consumer behavior, get some habit of users, with the data of the web site have some guidance effect optimization, make the operations department in marketing is more targeted, and to help users to improve the efficiency of access, the purpose of improving the user's experience, thereby saving costs, improve efficiency.

2 Research Status at Home and Abroad

With the rapid development of science and technology and the increasing popularity of the Internet, the number of network users is gradually expanding, and the behavior of network users is becoming more and more complex. A large number of studies have shown that users' actual demand for resources on the network



application platform is closely related to their historical behavior records [11,12]. As an abstract concept, user behavior refers to the behavior law of users when using application service business. User behavior analysis refers to the use of a variety of scientific knowledge to study and analyze the characteristics of user behavior in the network platform, and to dig out the behavior characteristics and behavior rules implied by users in the network application [13,14]. The aim is to combine this law with the network service strategy to provide the basis for further optimizing the network resources and providing high-quality service strategy [15]. At present, the user behavior studied mostly refers to the network user behavior of a single website platform, as well as the analysis of the website target user behavior.

In recent years, more and more scholars at home and abroad have studied Web user behavior mining, and the analysis of user behavior is mainly based on data mining technology [16]. As early as the early 1990s, foreign countries began to conduct research on the information analysis of user behavior under the Internet environment, and the earliest research results were Chennells et al.'s [17,18] research on the British academic network users. Relatively speaking, China began to pay attention to the behavior of Internet users and began to study the relatively late. At present, the methods of user behavior analysis and research mainly include statistics and mining based on server logs, statistical mining based on traffic usage and statistical mining based on users' web browsing paths. Common data mining analysis methods include statistical analysis, clustering, classification analysis, association rules or frequent set mining. In terms of the user behavior analysis method, Chen et al. [19] realized the personalized recommendation function based on the content collaborative filtering and augmented matrix by using the traditional classifier and heuristic scoring mechanism based on the user behavior log. Wei et al. [20] proposed a collaborative filtering algorithm based on joint clustering smoothing to solve the problem of sparsity of user behavior data, which improved the prediction accuracy to some extent. Based on mobile communication user behavior, Li et al. [21] proposed a multidimensional analysis method combining network data and market development and operation data, and verified that the method achieved good results in mobile network.

To sum up, the user behavior of each application domain is unique, and the analysis method is different. For example, due to the large number of users, extensive behaviors and diversified emotional factors in social networks, the analysis method combining multiple algorithms is generally adopted. Some professional scientific sharing platforms, such as geosciences data sharing platform, have their own professional user behavior, which is generally single and simple to predict, and their user behavior analysis model methods are targeted. However, at present, many researches in China focus on user characteristics analysis of social networks, and lack of analysis on user behavior path conversion rate. Therefore, it is necessary to analyze the user behavior path conversion rate.

3 User Behavior Path Funnel Model Overview

Funnel model can disassemble and quantify each link in the process, help us analyze and monitor the key link in product operation, find the weak link, optimize through user guidance or product iteration, and improve the transformation effect. There are three types of funnel model.

3.1 AIDMA Model

AIDMA, one of the mature theoretical models in the field of consumer behavior, was proposed by the American advertising scientist E. S. Lewis in 1898. According to this theory, consumers will go through the following five stages:

A: Attention—fancy business CARDS, embroidered advertising slogans on handbags, etc.

I: Interest—the general method used is to cut and paste refined color catalogues and news bulletins about products.

D: Desire—the person who sells tea must prepare a tea set at any time and brew the customer a cup of strong tea with strong aroma. Sell the house, to show the customer the house. The entrance of the

restaurant should display the refined samples with full color, fragrance and fragrance, so as to make the customer feel the charm of the product and arouse his desire to buy.

A successful salesman says, “every time I promote my company’s products, I bring along catalogues from other companies and compare them in detail. Because if you keep saying how good your product is, your customers won’t believe you. Instead, they want to learn more about other companies’ products, and if you come up with other companies’ products first, customers will recognize your own.”

A: Action—the salesman must be confident all the way through the sales process from drawing attention to making A purchase. Overconfidence can also cause resentment among customers who think you're bluffing. So I do not trust your word.

The theory says that consumers’ purchase behavior is modeled, which is helpful for advertisers to conduct more effective product publicity after studying consumers systematically. However, the theory is not specific to different categories of goods. In fact, the theory is more suitable for goods with high involvement (high price, need to make decisions carefully), while for goods with low involvement, the decision-making process of consumers is often less complicated.

The model is shown in the following figure (Fig. 1):

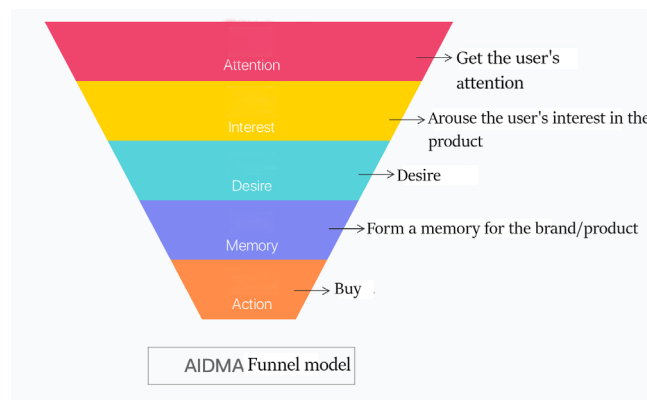


Figure 1: AIDMA model

Fig. 1: From top to bottom. Attract users’ attention. Draw the user’s attention to the product. Desire. Form a memory of the brand/product. Buy.

3.2 AISAS Model

AISAS model is a new consumer behavior analysis model proposed by dentsu for the change of consumer lifestyle in the era of Internet and wireless application. Emphasize the entry of each link, close to the user experience. In the brand new marketing law, the emergence of two “s” with network characteristics—search and share points out the importance of search and share in the Internet era, instead of blindly inculcating the one-way concept to users, which fully reflects the influence and change of the Internet on people's lifestyle and consumption behavior.

AISAS model is a new consumer behavior analysis model proposed by dentsu for the change of consumer lifestyle in the era of Internet and wireless application. Emphasize the entry of each link, close to the user experience. In the brand new marketing law, the emergence of two “s” with network characteristics—search and share points out the importance of search and share in the Internet era, instead of blindly inculcating the one-way concept to users, which fully reflects the influence and change of the Internet on people's lifestyle and consumption behavior.

In the traditional AIDMA model, consumers pay Attention to products, generate Interest, Desire to buy, leave Memory and make purchase actions. The whole process can be controlled by traditional marketing methods.

Based on the reconstruction of the network age characteristics of market AISAS (Attention note Interest Interest Search Search Action Action Share Share) mode, will consumers in Attention and Interest of information gathering (Search), and purchase information sharing (Share), after considerations as two important link, the two links are inseparable from the consumers in the Internet, including wireless Internet applications.

The new consumer behavior model (AISAS) determines the new consumer Contact Point. Management on the basis of dentsu's Contact Point Management (Contact), the media will no longer be limited to a fixed form, no longer fragmented, different media types for the media form, delivery time, delivery method, first of all, from the consumer the feasible Point of Contact with the product or brand recognition, in all of the Contact Point and consumers to communicate information. At the same time, in the center of the information communication circle, the consumer website that explains product features in detail becomes the deep end of information communication with consumers at each contact point. Consumer websites not only provide detailed information, so that consumers understand the product more deeply and influence their purchase decisions; it also facilitates interpersonal communication among consumers. At the same time, marketers can develop more effective marketing plans by analyzing visitor data.

Due to the irreplaceable information integration and interpersonal communication functions of the Internet, all information will be aggregated on the Internet to produce multiple communication effects, and a cross-media full communication system with the network as the aggregation center will be born.

The model is shown in the following figure (Fig. 2):

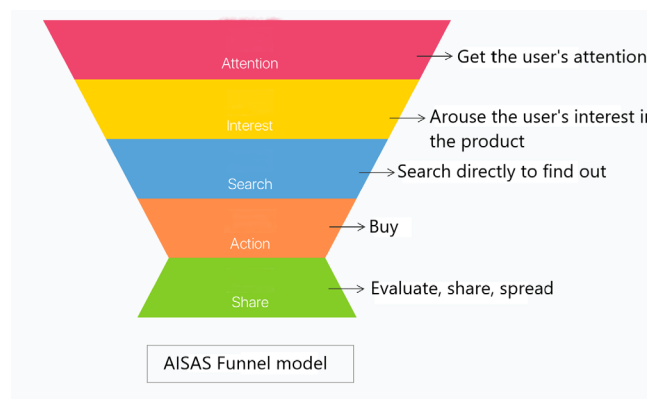


Figure 2: AISAS model

Fig. 2: From top to bottom: Draw the user's attention. Arouse users' interest in the product. Search directly and understand. Buy. Evaluate, share, spread.

3.3 AARRR Model

AARRR is an acronym for Acquisition, Activation, Retention, Revenue, and self-propagation, which correspond to the five key segments of a mobile application's life cycle.

Acquisition: the first step in running a mobile app is, of course, Acquisition, or promotion. If there are no users, there is no operation.

Activation: many users may have entered the application through different channels such as terminal presets, advertising, etc. These users entered the application passively. How to turn them into active users is the first problem operator face.

Retention: Some apps have solved the liveliness problem and found another one: "users come and go quickly." Sometimes we say the app is not as sticky.

Revenue: Revenue acquisition is actually the core of application operation. Very few people build an app out of pure interest, and most developers are most concerned with revenue. Even free apps should have a profit model.

There are many sources of revenue, and there are three main types: paid apps, in-app payments, and advertising. Paid apps are poorly received in China, including Google Play Store, which only offers free apps in China. In China, advertising is the source of income for most developers, and in-app payment is currently more widely used in the game industry.

In either case, the revenue comes directly or indirectly from users. Therefore, the aforementioned increase in activity and retention is necessary to generate revenue. The user base is big, the revenue just is possible on the quantity.

Refer: the previous operational model ended at the fourth level, but the rise of social networks has added another aspect to the operation, namely the viral spread of social networks, which has become a new way to obtain users. The cost is low, and the results can be very good; the only prerequisite is that the product itself is good enough to have a good reputation.

From self-propagation to acquiring new users again, the application operation forms a spiral. And the best apps take advantage of that trajectory and expand their user base.

The model is shown in the following figure (Fig. 3):

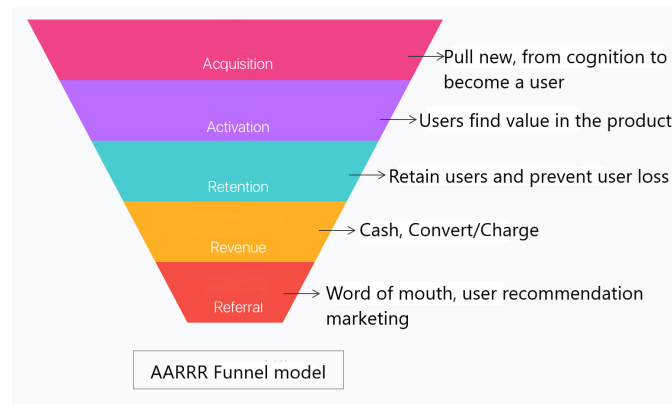


Figure 3: AARRR model

Fig. 3: From top to bottom: Pull new, from cognition to become a user. Users discover product value, Retain users, prevent user loss. Realization, conversion/charge. Word of mouth, user recommendation marketing.

Funnel model is actually an overview of user path, which can describe various processes, such as marketing purchase process, customer acquisition growth process, invite share process, add purchase conversion, operation bit conversion, repurchase and so on. Funnel model is widely used in data analysis, which can help us understand the running status of products in the current period, track the path of user behavior, realize the refined operation of products, and evaluate the results of each event.

4 Experiment and Analysis

This experiment analyzes the consumer behavior of CDNow website by analyzing the purchase details of users, including the overall consumption trend of users, individual consumption data, consumption cycle, user stratification and user quality to analyze the characteristics of consumer behavior.

4.1 Data Processing

Import the data and view the basic information of the data (Fig. 4 and Fig. 5):

	User ID	Purchase date	orders	order amount
0	1	19970101	1	11.77
1	2	19970112	1	12.00
2	2	19970112	5	77.00
3	3	19970102	2	20.76
4	3	19970330	2	20.76
5	3	19970402	2	19.54
6	3	19971115	5	57.45
7	3	19971125	4	20.96
8	3	19980528	1	16.99
9	4	19970101	2	29.33

Figure 4: Data basic information

	User ID	Purchase date	orders	order amount
count	69659.000000	6.965900e+04	69659.000000	69659.000000
mean	11470.854592	1.997228e+07	2.410040	35.893648
std	6819.904848	3.837735e+03	2.333924	36.281942
min	1.000000	1.997010e+07	1.000000	0.000000
25%	5506.000000	1.997022e+07	1.000000	14.490000
50%	11410.000000	1.997042e+07	2.000000	25.980000
75%	17273.000000	1.997111e+07	3.000000	43.700000
max	23570.000000	1.998063e+07	99.000000	1286.010000

Figure 5: Data statistical description information

Fig. 4 and Fig. 5: The number of columns from left to right is Id, Purchase date, orders and Order Amount. It can be seen from the statistical description information of the data that the user purchased 2.41 commodities on average per order and spent 35.89 yuan on average per order. The standard deviation of the quantity of goods purchased is 2.33, indicating that the data has certain volatility. The median is 2 items and the 75th quantile is 3 items, indicating that most orders are purchased in small quantities. The maximum is 99, which is pretty high. The amount of purchase is similar, with most orders concentrated in small amounts. In general, the distribution of consumer data is a long tail. Most users are small, while a small number of users contribute the majority of the revenue, commonly known as the 28.

The monthly total sales, times of consumption, sales volume and number of consumers are shown in the figure below (Fig. 6).

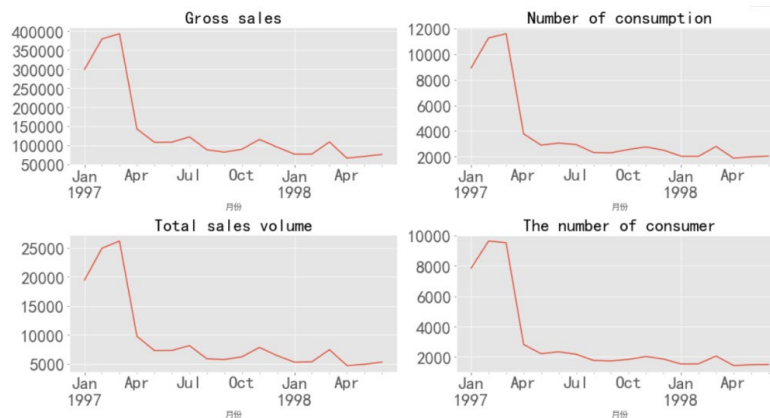


Figure 6: Consumption chart

Fig. 6: It can be seen that the sales volume in the first three months of 1997 was particularly high, which dropped suddenly after March. During the period from February to March, the number of consumers declined slightly, but the total sales volume and the total sales volume still rose. Users in March may have high-value customers that we need to focus on developing.

Draw the user scatter diagram as shown below (Fig. 7):

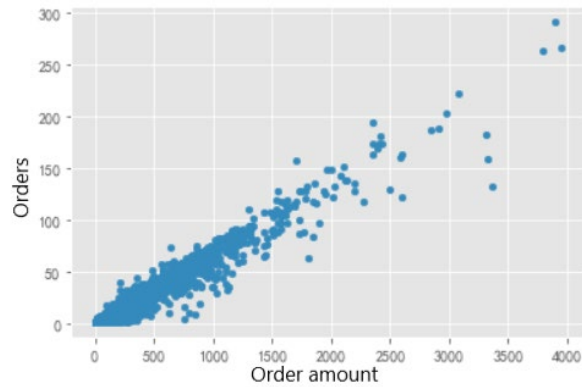


Figure 7: User scatter diagram

Fig. 7: As can be seen from the figure, users are relatively healthy and regular. Since this is the sales data of CD website, the commodities are relatively single, and the relationship between amount and quantity of commodities is linear, with few outliers.

According to the user’s consumption amount, the distribution diagram is shown below (Fig. 8):

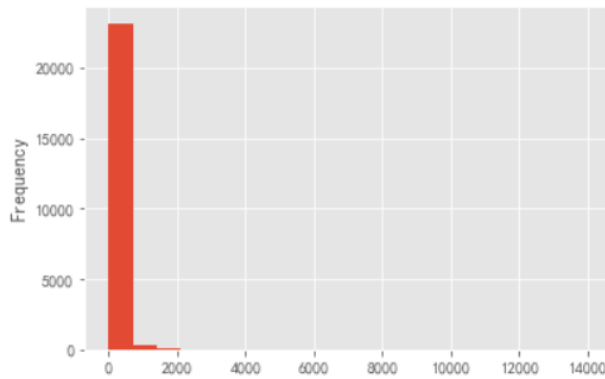


Figure 8: Distribution of consumption amount

Fig. 8: As can be seen from the figure, the consumption of users shows a central trend, which may be caused by individual maximum interference.

Therefore, the distribution of the figure below is greatly worth excluding (Fig. 9).

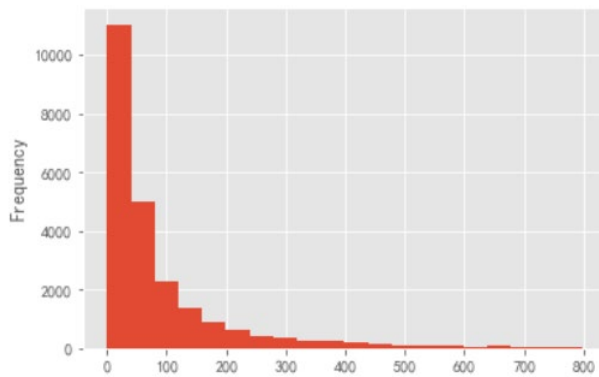


Figure 9: Distribution of consumption amount

Fig. 9: After selecting the users whose consumption amount is less than 800, it can be seen that the consumption capacity of most users is not high, nearly half of the users' consumption amount is less than 40 yuan, and the number of high-consumption users (>200 yuan) is less than 2,000.

It can be seen from the histogram of the figure above that most users' consumption ability is not high, and most of them focus on a very low consumption level. High consumption users can hardly be seen on the graph, which is also in line with the industry rules of consumption behavior. Although there is extreme data interference, most users still focus on the lower consumption level.

According to the user consumption times, the following distribution diagram is obtained (Fig. 10):

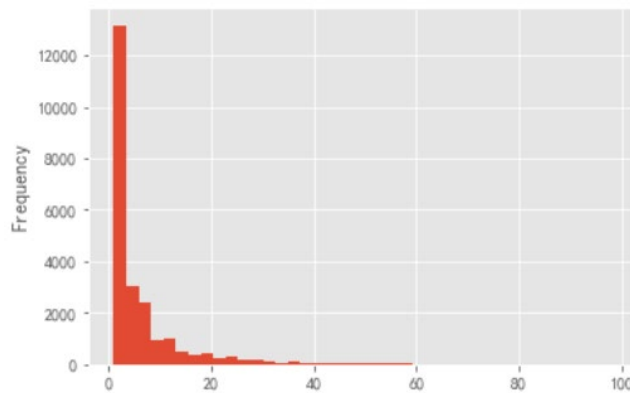


Figure 10: Distribution of consumption frequency

Fig. 10: It can be seen from the figure that most users buy CDS within 3 pieces, and the number of users who buy a large number of CDS is not large.

4.2 User Quality Analysis

How many users only consume once? (Multiple consumption within a day is recorded as one).

Draw the pie chart below (Fig. 11):

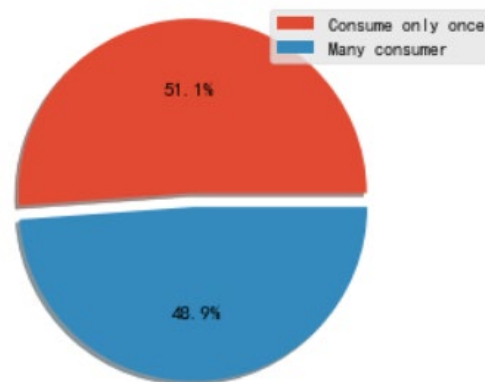


Figure 11: Consumption times pie chart

Fig. 11: More than half of the users only consume once, which also shows that the operation is not good and the retention effect is not good.

Repurchase rate, repurchase rate and retention rate.

Calculate the repurchase rate and draw the change in the repurchase rate as shown below (Fig. 12).

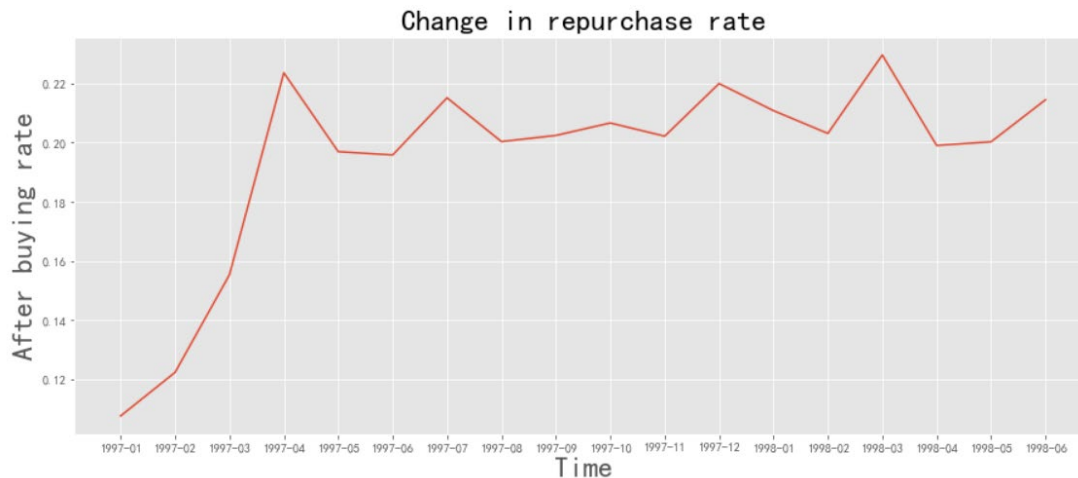


Figure 12: Change in repurchase rate

Fig. 12: It can be seen from the figure that the repurchase rate is in the early stage. Due to the large number of new users, the repurchase rate of new customers is not high. For example, the repurchase rate of new customers in January is only about 6%. And later on, at this time the user is the dafangtaosha remaining old customers, repurchase rate is relatively stable, about 20%. Look at new customer and old customer only, answer buy the difference of rate of about 3 times.

After calculation, the following figure shows the change in the buyback rate (Fig. 13):

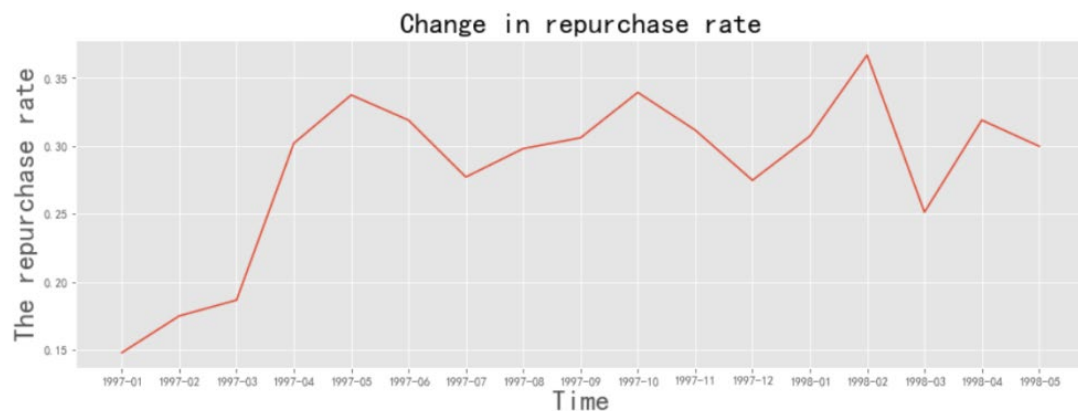


Figure 13: Repurchase rate variation chart

Fig. 13: As can be seen from the figure above, the repurchase rate of the initial users is not high. The repurchase rate in January was only about 15%, and since April the repurchase rate has been stable at about 30%.

As can be seen from the data of the number of users who have buyback consumption every month, the number of buyback users has a downward trend as a whole. The analysis of the buyback rate once again shows that for new users, three months after their first consumption is an important period, and marketing strategies are needed to actively guide their consumption again and continuously. In addition, for the continuous consumption of old customers, also should timely launch feedback of old customers preferential activities, in order to strengthen the loyalty of old customers.

After analysis, the following retention figure was obtained (Fig. 14):

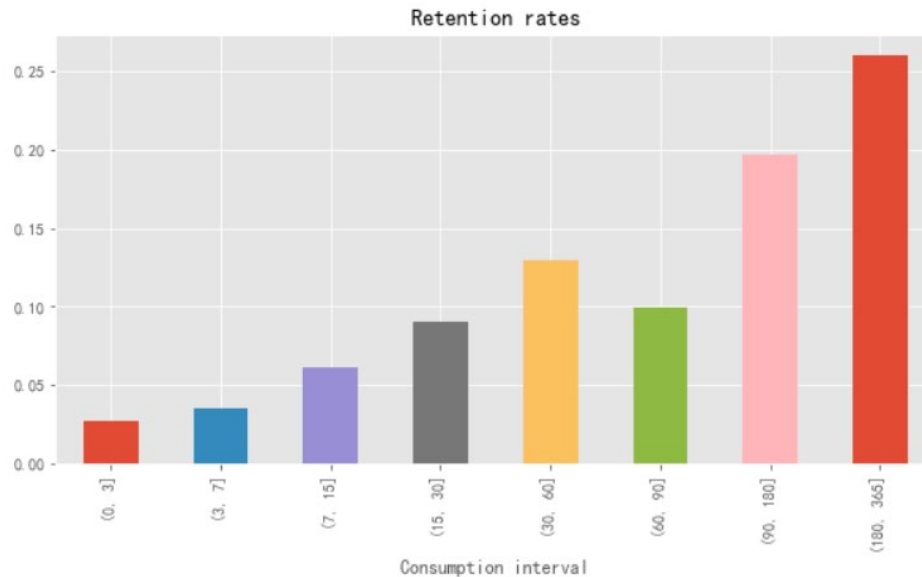


Figure 14: Retention rates figure

Fig. 14: Only 2.5% of users consumed within the next day to three days after their first purchase, and 3% consumed within three to seven days. The Numbers do not look good, and CD buying isn't really high-frequency consumer behavior. 20% of users made a purchase between three months and six months after the first purchase, and 27% of users made a purchase between six months and one year after the first purchase. From the perspective of operation, while serving new users, CD marketing should pay attention to the cultivation of user loyalty and recall users to purchase within a certain period of time.

Grouped by user id, the total amount spent by the user is summed up, and then compared to the total sales, so that the ratio abscissa is the user's id. Calculate the ratio diagram below (Fig. 15 and Fig. 16).

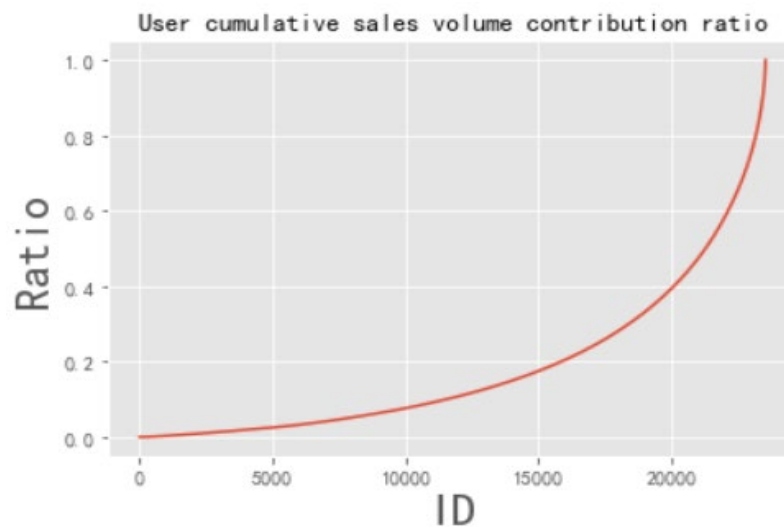


Figure 15: User cumulative sales volume contribution ratio

Fig. 15: The first 20,000 users, about 80 percent contributed 40 percent of sales, and 20 percent contributed 60 percent.

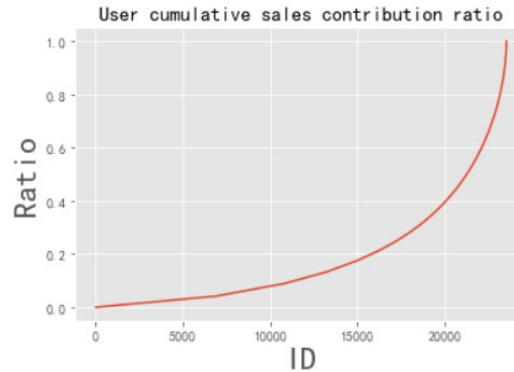


Figure 16: User cumulative sales contribution ratio

Fig. 16: Very close to sales. The first 20,000 users contributed 40% of the consumption, while the last 3,500 users contributed 60% of the consumption. In line with the trend of 28. In other words, as long as we maintain the 3,500 users, we can complete 60% of the performance KPI, and if we can better operate the 3,500 users, they can account for 70% to 80%.

5 Conclusion

1. Overall trend: the monthly trend sales volume and sales volume in January to march are relatively high, and then drop sharply, which may be related to the vigorous promotion during this period or the quarterly nature of the goods.

2. Individual characteristics of users: the amount and purchase amount of each order are concentrated at the low level of the range, and they are all purchased in small amounts and batches. This kind of transaction group can enrich the product line and increase promotional activities to improve the conversion rate and purchase rate.

3. The total consumption and purchase amount of most users are concentrated in the low segment and the long tail, which is related to user demand. It is possible to endow products with diversified cultural values, enhance their social value attributes and enhance users' value demands.

4. The repurchase rate of new customers is about 6%, and that of old customers is about 20%; The repurchase rate of new customers is about 15%, and that of old customers is about 30%. Marketing strategies are needed to actively guide their consumption again and continuously.

5. User quality: the individual consumption of users has a certain regularity. The consumption of most users is under 2000. So, pay close attention to high-quality user is eternal invariable truth, these high-quality customer are "member" type, need to optimize shopping experience specially for the member, such as special line answer, special discount and so on.

6. In terms of retention rate, half of the users will be lost, so attention should be paid to the cultivation of user loyalty, such as card check-in, point system, discount system for old users and membership upgrade system.

Acknowledgment: This research work is implemented at the 2011 Collaborative Innovation Center for Development and Utilization of Finance and Economics Big Data Property, Universities of Hunan Province; Hunan Provincial Key Laboratory of Big Data Science and Technology, Finance and Economics; Key Laboratory of Information Technology and Security, Hunan Provincial Higher Education. This research is funded by the Open Foundation for the University Innovation Platform in the Hunan Province, grant number 18K103; Open project, Grant Number 20181901CRP03, 20181901CRP04, 20181901CRP05; Hunan Provincial Education Science 13th Five-Year Plan (Grant No. XJK016BXX001), Social Science Foundation of Hunan Province (Grant No. 17YBA049).

Funding Statement: This paper partly supported by the project 18K103.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] L. Wang, G. Laszewski. Von and A. Younge, "A cloud computing: a perspective study," *New Generation Computing*, vol. 28, no. 2, pp. 137–146, 2010.
- [2] G. Wang, X. Zhang and S. Tang, "Unsupervised clickstream clustering for user behavior analysis," *Chi Conference ACM*, vol. 35, no. 2, pp. 225–236, 2016.
- [3] M. Al-Qurishi, M. S. Hossain and M. Alrubaian, "Leveraging analysis of user behavior to identify malicious activities in large-scale social networks," *IEEE Transactions on Industrial Informatics*, vol. 27, no. 99, pp. 1, 2018.
- [4] M. Wang, J. Wang and H. E. Yuntao, "An approach for prediction of web user behavior and data recommendation for geoscience data sharing portals," *Journal of Geo-Information Science*, vol. 41, no. 9, pp. 524–531, 2017.
- [5] Y. C. Chen, C. C. Yang and Y. J. Liu, "User behavior analysis and commodity recommendation for point-earning apps," *Technologies and Applications of Artificial Intelligence*, pp. 170–177, 2017.
- [6] M. Li, J. Yin and J. Tan, "User behavior analysis and research based on big data in large-scale gathering scene," *International Symposium on Communications and Information Technologies*, pp. 360–366, 2016.
- [7] H. Yin and Z. Hu and X. Zhou, "Discovering interpretable geo-social communities for user behavior prediction," *IEEE, International Conference on Data Engineering*, pp. 942–953, 2016.
- [8] Y. Xiong, B. Wang and C. Chu, "Distributed Optimal Vehicle Grid Integration Strategy with User Behavior Prediction," *Intelligent Systems IEEE*, vol. 1, no. 2, pp. 1–5, 2017.
- [9] S. Chen, M. Ghorbani and Y. Wang, "Trace-based analysis and prediction of cloud computing user behavior using the fractal modeling technique," *IEEE International Congress on Big Data*, pp. 733–739, 2014.
- [10] I. Bordino, N. Kourtellis and N. Laptev, "Stock trade volume prediction with Yahoo Finance user browsing behavior," *International Conference on Data Engineering*, pp. 1168–1173, 2014.
- [11] P. Dai, S. S. Ho and F. Rudzicz, "Sequential behavior prediction based on hybrid similarity and cross-user activity transfer," *Knowledge-Based Systems*, vol. 77, no. 12, pp. 29–39, 2015.
- [12] M. Ghiassi, D. Lio and B. Moon, "Pre-production forecasting of movie revenues with a dynamic artificial neural network," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3176–3193, 2015.
- [13] W. Etaiwi, M. Biltawi and G. Naymat, "Evaluation of classification algorithms for banking customer's behavior under Apache Spark Data Processing System," *Procedia Computer Science*, vol. 113, no. 5, pp. 559–564, 2015.
- [14] G. Sun, X. P. Fan, S. Fu and Y. J. Song, "Software watermarking in the cloud: analysis and rigorous theoretic treatment," *Journal of Software Engineering*, vol. 9, no. 2, pp. 410–418, 2015.
- [15] M. D. Preda and M. Pasqua, "Software watermarking: a semantics-based approach," *Electronic Notes in Theoretical Computer Science*, vol. 331, pp. 71–85, 2017.
- [16] N. Zong and C. Jia, "Software watermarking using support vector machines," in *IEEE 39th Annual Computer Software & Applications Conference*, vol. 2, pp. 533–542, 2015.
- [17] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang *et al.*, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.
- [18] L. Zhou, "QoE-driven delay announcement for cloud mobile media," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 1, pp. 84–94, 2017.
- [19] S. Patel and T. Pattewar, "Software birthmark based theft detection of JavaScript programs using agglomerative clustering and improved frequent subgraph mining," in *International Conference on Advances in Electronics Computers and Communications, Bangalore, India*, pp. 1–6, 2014.
- [20] G. Hurel, R. Badonnel, A. Lahmadi and O. Festor, "Outsourcing mobile security in the cloud," in *Monitoring and Securing Virtualized Networks and Services*. Berlin, Germany, pp. 69–73, 2014.
- [21] K. Hashizume, D. G. Rosado, Eduardo Fernández-Medina and E. B. Fernandez, "An analysis of security issues for cloud computing," *Journal of Internet Services & Applications*, vol. 4, no. 10, pp. 109–114, 2013.