Research on APP Intelligent Promotion Decision Aiding System Based on Python Data Analysis and AARRR Model

Juan Liao1*, Yunfei Ruan2

1 School of Big Data Engineering, Kaili University, Kaili, Gui Zhou, 556000, China
2 School of Greater Health, Kaili University, Kaili, Gui Zhou, 556000, China

*Corresponding author’s e-mail: liaojuan@kluniv.edu.cn

Abstract. With the development of AI technology and the current situation of the epidemic, online shopping is increasingly popular all over the world. The mobile phone terminals also occupy a larger share, and the use of various APPs has also emerged rapidly in line with the trend of the times. How big data technology is used to analyze the effective data generated by APPs? The paper conducts research from the following three aspects: (1) Finding out customer’s behavior patterns from the analyzed e-commerce data, and analyzing the reasons for customer churn rate and providing effective marketing strategies for merchants; (2) according to the time dimension, from different time dimension of customer behavior rules, find out customer purchase rules for merchants; (3) from the point of view of commodities, finding out customers' preferences for different commodities, and give different marketing schemes for commodities with different preferences. Through the three aspects of the study, we use Python for data analysis, combined with AARRR model and traditional expert system, the system finally sets up an intelligent decision aid system recommended by APP, the system provides different marketing schemes for merchants. The experiment shows that through data analysis and simulation of the use of an APP, in the experimental environment, the APP can better serve customers under the same conditions, and the transaction rate has increased by 7.3%.

1. Introduction
With the rapid development of artificial intelligence, especially since the global epidemic in 2019, people's shopping habits have already changed from traditional shopping mode to online shopping, and shopping on mobile phones occupies a larger proportion. Entrepreneurs use this psychology of people to quickly flood the market with various types of apps. However, with the increase in APPs, how to apply APPs to the market and have a competitive advantage is an urgent problem for such entrepreneurs. CHEN Qianling uses the relationship between customers and managers to establish a data framework. The data mainly comes from China Telecom. It uses data mining to model the AARRR model, analyzes how to improve the conversion rate, and provides some guidance for managers. However, this article There is no analysis of other influencing factors, mainly the marketing strategy given from a single conversion rate[1]. Literature 2 uses the AARRR model to analyze the current popular online teaching, mainly analyzes the learning form of the MOOC website, uses the aarr model and E-learning, mainly analyzes the model, and does not analyze the data[2]. References 3 to 7 all study the application of AARRR model in different fields from the five elements of the model.
itself, and they all aim at different industries without fundamentally analyzing the data to analyze the operation situation of each industry.

The development of e-commerce in recent years has also been changing with each passing day, especially for mobile clients, the more mature development such as Taobao, JD, Vipshop, and Pinduoduo, the share of which is obviously showing an irreversible trend. Data is still difficult to adapt to the development of the market. And competitive environment in the field of App, will eventually lead to App products business to develop in the direction of more vertical and segmentation, at the same time, it also means the development of the global App will make the App constantly seek function, experience and mode of innovation, and vertical user requirements will be digging deeper and deeper, will also be further refined in the vertical direction.

This paper mainly analyzes the use of shopping APP and the problems encountered in the use of APP and the suggestions for improvement through the analysis of consumers' behavior of APP. The research is mainly carried out from the following aspects: (1) first, we analyze several technical indicators commonly used in shopping APP e-commerce, and perform linear analysis on each indicator to find out the customer churn rate in APP use; (2) second, This article divides the data into different time periods, and uses clustering algorithms to find out the rules for different periods of data, and gives different marketing strategies through the rules; (3) for different kinds of products, different suggestions are given based on the amount of customers' attention to them, or the value of customers' preference; (4) lastly, through the establishment of the AARRR model, combined with the traditional expert database system, this article establishes an intelligent auxiliary marketing system.

2. AARRR model and Python Data Analysis

AARRR model was first proposed by Shawn in the book Growth Hacker. This model can not only be applied to the major Internet to analyze the usage of the Internet, but also plays a significant role in the analysis of APP with the continuous development of mobile phone terminals in recent years.

The promotion of shopping APPs continues to flood the market. However, no matter how various APPs develop, they cannot be separated from the operation of customers' browsing, adding to the shopping cart, collecting, buying and so on. The basic composition of the AARRR model is shown in Figure 1 below.

![AARRR model](image)

**Figure 1. AARRR model.**

User acquisition: For any APP promotion, the acquisition of users is the first and very important step, only with customers, any business can be done.

User activity: The use of an APP, only to reach a certain degree of activity, it is necessary to promote activity. The activity of users need to be counted every day, and the activity of users need to be counted every week and every month.

Retention rate: As the name implies, the retention rate is the customers that can be retained. The retention rate can be divided into customers that can be retained every day, those that can be retained for 5 days, and those that can be retained for 10 days.

Income capture: Merchants develop and promote APP, the ultimate goal is to have income, so income is also the key to the whole model analysis, and also plays a decisive role in the intelligent decision-making system.
Transformation diffusion: A good product requires the "word of mouth" of customers. Similarly, a useful APP also needs customer diffusion and customer communication.

3. Intelligent Expert Assistant Decision System

By calling the traditional expert assistance programs of e-commerce, including e-commerce services provided by traditional experts. By establishing the modification plan analyzed in the AARRR funnel model, and then calling the opinions of the expert library, the data between the two are similarly matched, and then the intelligent auxiliary decision-making system is obtained.

4. Simulation experiment

4.1. Data processing

The data studied in this article comes from a certain building material APP. The marketing model of this building material city is online and offline. The data obtained is the data from December 11, 2019 to January 11, 2020. Since the promotion of mobile apps in 2017, the company’s sales have been four times higher than offline transactions in 2015 and 2016, and it is also being promoted to many cities across the country. Then selecting the data for analysis has certain simulation significance.

The meanings of the data fields of each part are as follows: user ID number, product classification number, customer location information, customer behavior classification (browsing products, adding shopping carts, collecting products, purchasing products), record of purchase time, purchase amount, etc. We process the selected data as shown in the figure below: Due to the limited factors to be considered, we exclude the customer location information and do not consider it temporarily.

Table 1. Raw data.

| user_ID  | commodity_ID | Customer_location | Customer_behavior | Purchase_time | purchase_price |
|---------|--------------|-------------------|-------------------|--------------|---------------|
| 2288318 | 25020377     | bt                | add_ct            | 2018021441   | 604           |
| 2503348 | 2620771     | bt                | add_ct            | 2018021442   | 25.3          |
| 2576651 | 149192      | bt                | add_ct            | 2018021445   | 43            |
| 3039088 | 4813161     | bt                | add_ct            | 2018021446   | 80            |
| 4360666 | 25220777    | bt                | add_ct            | 2018021440   | 60.7          |
| 4609018 | 2785495     | bt                | add_ct            | 2018021441   | 187           |
| 2302806 | 411135      | bt                | add_ct            | 2018021431   | 597           |
| 3827999 | 2000476     | add_ct            | add_ct            | 2018021431   | 34.7          |
| 3741569 | 2891500     | add_ct            | add_ct            | 2018021440   | 79            |
| 1551078 | 2903457     | add_ct            | add_ct            | 2018021420   | 69            |
| 2298667 | 4198183     | add_ct            | add_ct            | 2018021421   | 69.9          |
| 2881066 | 1080780     | by_item           | by_item           | 2018021200   | 43            |
| 3189717 | 2350607     | by_item           | by_item           | 2018021203   | 55            |
| 3398256 | 149192      | by_item           | by_item           | 2018021206   | 1279          |
| 2296764 | 2465336     | by_item           | by_item           | 2018021460   | 456           |
| 5000108 | 2620377     | by_item           | by_item           | 2018021712   | 224           |
| 2739026 | 4198183     | by_item           | by_item           | 2018021731   | 1187          |
According to the data processed in the above figure, it can be seen that some of the data is the result of repeated processing. Of course, this part of the data also shows that there is a phenomenon of repeated purchases by some customers in a short period of time. In order to deal with it more convincingly, we temporarily ignore this part of the data. The processed data is shown in the figure below.

Table 2. Processed data.

| user_ID | commodity_ID | Customer_behavior(bt, add_ct, bm_item, by_item) | Purchase_time | purchase_price |
|---------|---------------|-----------------------------------------------|---------------|---------------|
| 2268318 | 2520377 | bt                                            | 2019121411    | 634            |
| 2333346 | 2520771 | bt                                            | 2019121413    | 25.3           |
| 3830808 | 4181361 | bt                                            | 2019121416    | 89             |
| 4366668 | 2520777 | bt                                            | 2019121620    | 66.7           |
| 4661008 | 2735466 | bt                                            | 2019121410    | 167            |
| 230880 | 411153  | bt                                            | 2019121511    | 567            |
| 3872999 | 902475  | add_ct                                        | 2019121920    | 79             |
| 3745169 | 2891590 | add_ct                                        | 2019121921    | 69             |
| 1531036 | 2902476 | add_ct                                        | 2019121211    | 589            |
| 2296567 | 4145813 | add_ct                                        | 2019121212    | 55             |
| 5108779 | 2355072 | bt                                            | 2019121823    | 55             |
| 1338625 | 419192  | bt                                            | 2019121920    | 1279           |
| 2296574 | 4635383 | bt                                            | 2019121618    | 465            |
| 5002015 | 2620377 | bt                                            | 2019121712    | 234            |
| 2774426 | 4145813 | bt                                            | 2019121711    | 1187           |
| 5102615 | 5002015 | bt                                            | 2019121212    | 368            |

Since each part of the data is directly imported from the database, it is difficult to understand the column names for processing with Python. We need to perform the following consistency processing on the data: (1) Processing of column names: the user ID Number, product classification number, customer behavior classification (browsing products, adding to shopping cart, collecting products, purchasing products), record of purchase time, purchase amount, etc. are recorded as use_id, class_id, use_shop, use_time, use_money in turn. (2) For the classification of customer behavior, the original data uses a notation method, which is not easy to identify. This article is convenient for clustering discussion during the research, and normalization is used to separate browsing products. (3) Time processing. The rule of time period requires separating the date and the number of hours. The data after consistent processing is shown in the following table.

Table 3. Date separated data is final data.

| user_ID | commodity_ID | Customer_behavior(bt, add_ct, bm_item, by_item) | Pur_date | purchase_price | pur_time |
|---------|---------------|-----------------------------------------------|----------|---------------|---------|
| 2268318 | 2520377 | bt                                            | 20191214 | 634            | 11      |
| 2333346 | 2520771 | bt                                            | 20191214 | 25.3           | 13      |
| 3830808 | 4181361 | bt                                            | 20191214 | 89             | 16      |
| 4366668 | 2520777 | bt                                            | 20191214 | 66.7           | 20      |
| 4661008 | 2735466 | bt                                            | 20191214 | 167            | 1       |
| 230880 | 411153  | bt                                            | 20191214 | 567            | 11      |
| 3872999 | 902475  | add_ct                                        | 20191217 | 34.7           | 19      |
| 3745169 | 2891590 | add_ct                                        | 20191217 | 79             | 20      |
| 1531036 | 2902476 | add_ct                                        | 20191229 | 69             | 21      |
| 2296567 | 4145813 | add_ct                                        | 20191221 | 58.9           | 21      |
| 5108779 | 2355072 | bt                                            | 20191218 | 55             | 23      |
| 1338625 | 419192  | bt                                            | 20191219 | 1279           | 9       |

4.2. Establishment of AARRR Hourglass Model

Analyze the four behaviors of users, such as browsing products, adding to shopping carts, collecting products, and purchasing products. From the processed data, using the data analysis in Python, we can get the share of each part of the four behaviors as follows Show.

Figure 3. Four parts of the data.
In order to give a more convincing model, when we deal with page views, we count the pages that are only viewed once, and compare them with the total number of pages visited. This is called the customer bounce rate, that is, the one-time visit. We ignore such customers. In our statistical data period, the total number of pages visited by the APP is 305013, the total number of customers who use the APP is 1,935, and the number of customers who only enter the APP is 13 people, that is, nothing is done when entering the mobile client, and the page is not To browse, I just clicked the APP to enter and then exited, without browsing any products. It can be concluded that the customer's skip rate: 13/1935=0.67%, which can also prove that the effectiveness of the APP in the three years of promotion is still very good, because the skip rate reflects the customer's app to a certain extent The recognition rate of 0.67% is almost negligible.

The hourglass model established by us in terms of browsing, collection and payment is as follows:

![Figure 4. Model established.](image)

It can be seen from the model that the functions of collection and addition are the same. There is no sequence and both of them have a certain conversion rate. They are just customers' personal hobbies. We separated the customers of each segment, as shown in the figure, and we can see the conversion rate of each segment to the next step.

Judging from the changes in the data analyzed by Python, the month selected here is the data from December to January. There are two more important subsystems in this time period, Double Twelve and New Year's Day. The number of purchasers is mainly concentrated on Double Twelve. On New Year’s Day, this day is also the two most active days for merchants. Based on the seven-day regular change curve of traditional experts, the number of people who join shopping carts and collectors is mainly concentrated before Double 12. Given the large number of shopping carts and collectors, the probability of conversion into purchases is also possible.

![Figure 5. Weekly change graph.](image)
5. Conclusion
Through the analysis of the model and the data, the following classification suggestions can be drawn for each link: (1) use the expert curry to reduce the price of goods, and the actual price is higher than the attractive price, combined with the optimization of the product recommendation strategy in the AARRR analysis, the two are combined with each other, using the linear relationship between the data, and actively recommending products according to customers’ favorites, and sorting the priority of the products; (2) advertising, advertising in the traditional expert database is mainly on holidays, combined with the AARRR model, we can put advertisements not only on holidays, but also product advertisements to increase the conversion rate of adding shopping carts to purchasing products; (3) expert database The discount promotions in China, using good data analysis models. Promotions can be through live broadcasts, videos and other forms.

In this paper, under the same amount of funding, the traditional experts only increased the benefit by 2.4%, but the combination of AARRR and the expert database increased by 7.3%. It shows that the use of big data analysis, combined with traditional expert opinions, and the establishment of an intelligent auxiliary decision-making system have a certain effect on the development of mobile phones.

Acknowledgments
This work was supported in part by the Guizhou Provincial Department of Science and Technology Joint Fund Project No. [2017]7163 and the Department of science project No. [2016]309.

References
[1] CHEN Q.L, Lan DU. Managing Mobile Market Users Based on the AARRR Model in the Age of Big Data. 2016, 10(1):58-66.
[2] Li Y.H. Research on Airbnb user growth strategy based on AARRR model [D]. Mentor: Ning Lianju. Beijing University of Posts and Telecommunications, 2018.
[3] Yi H.L, Li H, Xu L, Wang R. Optimized operation of IT technology WeChat public account based on AARRR model [J]. E-commerce,2020(03):73-75+84.
[4] Zong P. Analysis of Network Marketing of Catering Enterprises Based on AARRR Model [J].Journal of Sichuan University of Tourism,2019(02):40-43.
[5] Xu Q.R. Research on User Retention Strategy of Shipping APP Based on AARRR Model [J].E-commerce,2019(07):66-67.
[6] Chen G.H, Feng Xuecheng, Gan Sha, Li Liwei. Analysis of marketing strategy of Luckin Coffee based on AARRR model [J]. E-commerce,2020(04):61-62.
[7] Zhou J, Wei J. Research on User Growth Strategy Based on AARRR Model -- Taking Pinduoduo as an Example [J].Journal of Shanxi Economic Management Cadre Institute,2020,28(01):11-16.
[8] Wu Y.Y. Industrial and Commercial Bank of China E-life Platform Operation Improvement Scheme [D]. Dalian University of Technology,2019.
[9] Zhao, Z. Z., & Balague', C. (2015). Designing branded mobile apps: Fundamentals and recommendations. Business Horizons, 58, 305-315.