

Investigation on the Development Status and Consumption Demand of China's Bottled Water Market Based on the Consumption Market of Youth Groups in Anhui Province

Xiaoyan Feng¹, Zejiong Zhou^{2,*}

¹School of Finance, Anhui University of Finance and Economics, Bengbu, 233030, China

²School of Economics, Anhui University of Finance and Economics, Bengbu, 233030, China

* Corresponding author: Zejiong Zhou (Email: aczzj123456@163.com)

Abstract: With the development of economy and society and the improvement of people's living standards, people have higher and higher requirements for drinking water quality. Therefore, the bottled water market has great development space in the future, but it is also accompanied by many prominent problems such as many manufacturers and fierce competition. In order to understand the young people's cognition and purchase preference of bottled water in Anhui Province, the research group used online questionnaires, expert interviews, field research and other methods to explore the development status, consumption needs and main problems of the bottled water industry. Firstly, we conduct descriptive statistical analysis on the survey data to understand the distribution of samples, and use contingency analysis and decision tree analysis to analyze the purchase preference of youth bottled water in Anhui Province. Secondly, from the four aspects of price, advertising, publicity and sales, AHP is used to explore the influencing factors of young people's purchase of bottled water. Thirdly, the structural equation model is used to explore the transmission mechanism of each influencing factor of bottled water purchase. Finally, from the dimensions of purchase habits, purchase scenes and purchase needs, this paper uses cluster analysis to describe the purchase behavior of five types of youth in Anhui Province.

Keywords: Youth bottled water market, Neural network, Decision tree, Structural equation model.

1. Introduction

In recent years, the scale of China's bottled water market has been expanding. In 2019, the scale of China's bottled water market was close to 200 billion yuan, and the per capita bottled water consumption was 62 liters, lower than that of the United States and the United Kingdom, basically the same as that of Japan. However, this does not mean that the growth space of China's per capita consumption of bottled water has peaked. First of all, in terms of the overall per capita consumption of soft drinks, China's per capita consumption of 62 liters in 2019 is far lower than that of other developed countries, which has considerable room for improvement. The difference between China and Japan is mainly reflected in the category structure. The per capita volume of carbonated drinks, fruit juice, instant coffee and instant tea in Japan is much higher than that in China [1]. Secondly, from the historical growth rate of per capita consumption of various categories, it can be found that since 2013, the driving force of China's per capita soft drink consumption mainly comes from bottled water. With the strengthening of health awareness and exercise habits, the demand for bottled water or sugar free drinks is expected to grow faster than sugary soft drinks. Therefore, from the per capita perspective, China's bottled water consumption is expected to exceed Japan's per capita level in the short term [2].

Youth groups are defined as people aged 18 to 35. As an important part of bottled water consumers, young people's consumption preference and consumption ability are of great significance to the development strategy of bottled water in the future. In the context of income increase and consumption upgrading, youth are the main force of consumption in the

bottled water market at present and in the future. "Healthy drinking water + convenience" will be the core driving force for the sustainable growth of the bottled water market in the future. The rise of high-speed rail and self driving travel is a new growth point of the bottled water market [3].

Nongfu Mountain Spring, Master Kang, Baisui mountain, Wahaha and Bingle account for 80% of China's bottled water market. According to the prospectus and channel feedback of Nongfu Mountain Spring, in 2019, the CR5 of bottled water reached 56.2%, of which Nongfu Mountain Spring, the leading enterprise, accounted for 20.9%, and the second in the industry was China Resources Yibao, with a market share of 13%. According to the historical data of Euromonitor, the share increase of Nongfu Mountain Spring and China Resources Yibao mainly comes from the share replacement of Master Kang and Wahaha [4]. In terms of overall product structure, China's bottled water is still dominated by medium and low-end products, accounting for about 70%. With the evolution of consumption upgrading, the high-end trend of bottled water industry is obvious, which is mainly reflected in the increase in the proportion of high-end bottled water and the growth rate is significantly higher than that of low-end bottled water.

The theme of our market research is to understand the market development prospect of bottled water in the youth group. Through consulting a large number of literature, on the basis of preliminary understanding and combined with the actual situation, we clarify the following five purposes: (1) summarize the consumption characteristics and demand levels of the youth of the target group of this research. The group conducted in-depth research on young people everywhere by using various investigation methods such as

field visits, issuing QR codes, filling in online questionnaires and interviews, and using analysis methods such as logistic regression, decision tree and case analysis, so as to more objectively and truly reflect young people's attitudes, opinions and consumption needs of bottled water products. In order to better understand the market development prospect of bottled water in youth groups. (2) To explore the influencing factors and demand levels of young people's purchase of bottled water products. This group carries out multi-dimensional information analysis on the young subjects. After summarizing the consumption characteristics and needs, through literature research and actual situation, this group tries to analyze the consumption psychology of the target subjects one by one, so as to master the consumption motivation of the target consumers. Due to the influence of various internal and external factors, consumers' purchase behavior of a brand often shows the characteristics of variability. Only through in-depth investigation and understanding of the purchase behavior and purchase habits of this special consumer group can we make good use of the situation to maintain the loyalty of this consumer group to the products of the enterprise. (3) Reveal the existing shortcomings and avoid detours. Based on the integrated research on the influencing factors and demand levels of young people's purchase of bottled water products, based on the problem orientation, this group will analyze the possible deficiencies in the process of entering the market in the past, suit the remedy to the case, provide improvement suggestions, and strive to solve the possible problems in the process of seizing the market and promote the smooth progress of market occupation [5]. (4) Cultivate brand loyalty of target consumers. Our target users are young people, from 18-year-old college students to 35-year-old young people. Our purpose is to cultivate a batch of long-term customers with high brand loyalty. If you own a brand, you will own the market, the present and the future, and you can obtain consumer loyalty. This is the rules of the game in the brand era. Therefore, for developing the youth market and allowing young consumers to buy their own products, it requires early market investment to allow young consumers to fully contact the products and cultivate their brand taste, so as to form their consumption preference.

The main significance of this survey is shown in the following four aspects: (1) listen to the voice of target customers and understand the demand and target market environment. By conducting a questionnaire survey on young people, the group can directly understand their consumption habits, consumption needs and target market environment, which is conducive to exploring young people's preference and acceptance of advertising, packaging and taste, so as to open the market. At the same time, promote youth to understand their consumption habits and demands more directly and effectively, so as to better safeguard their own interests. Further, the group will also summarize the opinions and suggestions, so that we can hear the true voice of target consumers and promote the benign interaction between enterprises and the masses. (2) It is conducive to catering to the consumption habits of young people and deeply developing the target market. From a long-term perspective, the youth bottled water market has great profit space and good development prospects. With the gradual establishment of youth's world outlook and outlook on life and the improvement of their income level, their purchasing power level will continue to improve. If enterprises can start to pay

attention to the potential target customer group when they are young, and make them grow into long-term customers and cultivate their brand loyalty through various effective measures, these seemingly insignificant consumer groups will bring long-term and huge economic benefits to enterprises in the future. (3) Summarize the advantages and disadvantages of the past, and provide reference and guidance for the next development. The main purpose of market research is to collect and analyze data to help enterprises make better decisions, so as to win valuable time and reduce mistakes in decision-making. Through data collection and sorting, we can more intuitively and specifically summarize the marketing advantages and disadvantages and unnoticed loopholes of enterprises in young consumer groups in the past, so as to provide reference and guidance for the layout and implementation of the next strategic plan. (4) Combine theory with practice to promote experience accumulation and cognitive deepening. Members of the group use their professional knowledge, give full play to their practical ability and enthusiasm for unity and cooperation, carefully collect data, continuously pay attention to the target market environment and dynamics, use the network to launch questionnaires and conduct research and analysis. This helps to improve our attention and understanding of enterprise policies and youth bottled water consumption; It helps to realize the transformation from theoretical knowledge to practical application, further consolidate professional knowledge, promote their own growth, and contribute to the realization of the enterprise strategy occupying the youth consumption market.

2. Sample of Bottled Water Market Survey

2.1. Investigation Content and Scheme Planning

2.1.1. Investigation Purpose

With the continuous enhancement of China's comprehensive national strength and the gradual improvement of people's living standards, people pay more and more attention to health. The water source known as the "source of life" is increasingly favored by consumers. In the process of pursuing a healthier life, more and more consumers realize the importance of water to life and life. Moreover, with the consumption upgrading brought by the improvement of consumers' per capita income and the impact of continuous water pollution accidents, the sales scale of bottled water in China has increased year by year since 2013. Innovation leads the future, and youth are the main new force of water consumption in the future. Through the research on the interests, values, media exposure, dietary characteristics, drinking water behavior and obstacles of youth groups, this paper gives suggestions on the communication selling points of the product creativity and media publicity channels.

In order to achieve the purpose of the survey, this survey needs to collect the following information: (1) collect the basic information of bottled water purchase of young people in Anhui Province, including gender, age and monthly disposable expenses, and analyze the characteristics of young people's purchase behavior. (2) Collect the bottled water purchase preferences of young people in Anhui Province, including the types of bottled water they like, whether they will repeatedly buy the same brand of mineral water, the preferred packaging form and style, what aspects of bottled

water will affect the purchase decision and the reasons for purchasing bottled water, analyze the interests and eating characteristics of young people in Anhui Province, and give breakthrough and innovative product ideas. (3) Collect the influencing factors of young people's purchase of bottled water in Anhui Province, including acceptable bottled water price, whether they will pay attention to the production place of bottled water, whether they will buy it because they have seen the advertisement of bottled water, whether they like the production process and technology of the product reflected in the advertisement, whether they will be externally affected by the environment and service attitude, and the most attractive publicity method. How to choose bottled water brands, favorite promotion methods, channels to obtain bottled water information, choose where to buy bottled water and frequently buy brands, study their contact with the media, and give suggestions on communication selling points of bottled water creativity and media publicity channels. (4) Collect the basic information about the purchase behavior of young people in Anhui Province, including the number of times they buy bottled water every month, the bottled water expenses they are willing to pay every month, whether the purchase behavior of bottled water will be affected by others, whether they will often buy Promotional bottled water, and what scenarios they will buy bottled water in, and analyze the values of young people in Anhui Province. (5) Collect the hot issues affecting the bottled water drinking behavior of young people in Anhui Province, including the most worried problems when buying bottled water, the aspects that need to be improved in the development of bottled water market, whether they are satisfied with the existing bottled water and the dissatisfied with the current situation of bottled water, and analyze the bottled water drinking behavior and obstacles of young people in Anhui Province.

2.1.2. Investigation Method

(1) Literature research method. Literature research method mainly refers to the method of collecting, identifying and sorting literature, and forming a scientific understanding of facts through the study of literature. Literature method is an ancient and vigorous scientific research method, which transcends the limitations of time and space, and has the remarkable advantages of saving time, money and high efficiency. Collect relevant policies, statistical data, academic papers and research reports of the current bottled water market, and summarize the contents and data.

(2) Intercept access method. Intercepting access refers to intercepting access objects in fixed places and conducting face-to-face access to qualified persons. According to different interception locations, it can be divided into street interception and fixed-point visit in the central block. Street interception, also known as random access, is to select the appropriate place in the block, generally commercial street, entertainment place, living community and other places, select the interview object according to certain procedures and requirements (such as intercepting one every few minutes or intercepting one every few pedestrians), and conduct a brief survey on the spot according to the questionnaire after obtaining the consent of the other party. The fixed-point visit in the central block is to select a relatively fixed place in the commercial block. Generally, the place with enough seats, good environment and can make the respondents feel safe (temporarily rented by the survey company). The interviewer intercepts the qualified respondents near the selected point and guides the respondents to this fixed point for visit. The

survey adopted the interception interview method in the preliminary pre survey stage. 50 young people were interviewed before RT Mart in Bengbu City to preliminarily understand the purchase status of young people for bottled water, so as to make preliminary preparations for the subsequent modification of the questionnaire and research.

(3) Network survey method. Network survey method, also known as online survey method, generally refers to the survey method of publishing survey information on the network and collecting, recording, sorting, analyzing and publishing netizens' feedback information on the Internet. The large-scale development of online survey originated in the 1990s. With the further development of China's Internet industry, online survey will be more widely used. Network survey method is a new survey method. Its emergence is a supplement to the traditional survey method and the application and development of the traditional survey method on the network. Compared with traditional survey methods, network survey method has obvious advantages in organization and implementation, information collection and survey effect. At the same time, network survey also has the characteristics of voluntariness, orientation, timeliness, interaction, economy and anonymity. We choose to publish the questionnaire directly on the network for investigation, because young people are active in various social media platforms. The way to obtain information through mainstream media such as QQ and Wechat has the advantages of simple organization, low cost, good objectivity, no space-time and regional restrictions, fast speed and so on. In February 2021, a small-scale pre survey was conducted on 50 young people in the form of online questionnaire. Compared with other public opinion poll interview methods, online questionnaire has the advantages of short time-consuming, low cost and high interview efficiency, which effectively improves the interview efficiency.

(4) In depth interview. In depth interview method is also known as in-depth interview method. In depth interview is a kind of unstructured, direct and personal interview. In the process of interview, an investigator with advanced skills deeply interviewed a respondent to reveal the potential motivation, beliefs, attitudes and feelings about a problem. Deep interview is suitable for understanding complex and abstract problems. Such problems are often not clear in a few words. Only through free conversation and in-depth discussion of the topics of concern can we summarize the information we need to understand. Deep interview method has the advantages of more in-depth exploration of the respondents' inner thoughts and views, direct connection between responses and respondents, more free exchange of information and so on.

2.1.3. Sampling Design

(1) Survey object and scope. The research object of this survey is the youth of Anhui Province. Youth refers to people aged 15-35. The main reasons for choosing them as the research object are as follows: ① Young people have a strong sense of autonomy, have unique opinions and receive social information comprehensively; ② Young people are the main force of bottled water consumption; ③ In 2019, there were 17.9072 million young people in Anhui Province (data from the 2020 Anhui Statistical Yearbook). A sufficient number of young people provided more possibilities for the selection of survey samples, wider coverage and more representative survey results; ④ The members of the

municipal survey team are XXX undergraduates. The market survey for Anhui Province is more conducive to the development of the survey activities and the good results of the survey activities. Therefore, the youth are selected as the research object, and the overall goal is the youth of Anhui Province.

(2) Sampling method and specific implementation. The team used the method of random sampling to select the appropriate sample size from 17.9072 million young people in Anhui Province. The actual operation is divided into four steps. The first step is to use the method of repeated sampling in stratified sampling to calculate the questionnaire that should be issued by each of the 16 prefecture level cities in Anhui Province. In practice, there are a large number of young people, and we cannot obtain the personal information of each young person. Therefore, we obtain the number of young people in Anhui Province in 2020 and the distribution number of local cities through the official website of Anhui Provincial Bureau of statistics. Taking this as the sampling frame, we allocate samples according to the proportion of the number of young people in each prefecture level city in Anhui Province. The second step is to use the random number method in simple random sampling and Python code to select three districts and counties in each prefecture level city. In the third step, the number of young people in three districts and counties in each prefecture level city is taken as the sampling frame, and the stratified sampling method is used to calculate the number of questionnaires that should be distributed in these three districts and counties. The samples were distributed according to the proportion of the number of young people in the three districts and counties in the total number of young people in the three districts and counties. In the fourth step, the deep learning network recognition model is used to find the places with large traffic in the three districts and counties extracted from each prefecture level city. In the specific implementation process, the team went to places with large traffic and invited young people to fill in the online questionnaire.

(3) Sample size determination. In order to ensure the universality and representativeness of the respondents, the formal respondents are set as the youth bottled water consumption according to the relevant results of the pre survey. Therefore, the sample size determination process before the formal survey is to pay attention to the sample variance of the youth bottled water consumption survey. The formula for calculating the optimal sample size n_0 before correction is:

$$n_0 = \frac{\mu^2 PQ/d^2}{1 + \frac{1}{N} \left[\frac{\mu^2 PQ}{d^2} - 1 \right]} \quad (1)$$

N is the total quantity. When N is large, $n_0 \approx \frac{\mu^2 PQ}{d^2}$. Taking the value of μ when the confidence is 90%, it can be seen that $\mu=1.645$, μ is the bilateral of the standard normal distribution.

P is the sample proportion, d is the sampling error range, and set $d=0.02$. According to the results of 50 pre surveys based on youth groups in the province, $P=0.4$. In the actual statistical process, if P is near 0.5, the sample can be estimated according to the maximum of the overall variance when $P=0.5$. Therefore, $P=0.5$ is set. According to the statistical information released by Anhui Provincial Bureau of statistics, the total N of Anhui youth in 2019 is 17.9072 million. By substituting the data, the most appropriate approximate

sample size can be obtained according to the relevant formula, that is, based on 90% confidence and under the condition that the absolute sampling error does not exceed 2%, the most appropriate sample size for this formal survey is determined as 1691.

At the same time, considering that the survey method is questionnaire survey and has the characteristics of random sampling, according to the applied statistical theory, taking the overall information related to the purchase of bottled water by young people in Anhui Province as the content, the determination formula of sample size is as follows:

$$n_0 = \frac{N z_{\alpha/2}^2 \sigma^2}{(N-1) \Delta_x^2 + z_{\alpha/2}^2 \sigma^2} \quad (2)$$

In this formula: n_0 is the sample size, that is, the overall content of relevant information; N is the number of young people in Anhui Province; σ is the overall standard deviation; Δ_x is sampling error; $z_{\alpha/2}$ is the standard normal distribution table ($\alpha/2$) Corresponding value of. The most suitable approximate sample size is 1691 by substituting the above relevant data.

Considering that some young people may give up filling in the questionnaire during the process of filling in the questionnaire, or the questionnaire filling results cannot be uploaded timely and effectively due to unforeseen and irresistible factors, after obtaining the opinions of the instructor and careful discussion of the team, it is decided that assuming that the invalid proportion of this survey is 20%, the total number of samples that should be investigated should be 2114, This is the final questionnaire distribution.

$$n' = \frac{1691}{1-20\%} \approx 2114 \quad (3)$$

(4) Sample distribution. Because there are great differences in the number of young people among cities at all levels in Anhui Province, the team adopted the method of stratified sampling to obtain the sample number of cities at all levels in 16 prefecture level cities in Anhui Province. According to the sample size distributed by local cities, 2114 questionnaires were actually distributed and 1759 valid questionnaires were recovered, with an effective proportion of 83.21%.

2.2. Investigation Implementation

2.2.1. Investigation Organization

After being familiar with the background, purpose, research framework and significance of the investigation project, we understand that the quality of investigators is very important to the success of the investigation. Although before the formal market research practice competition, we all passed the examination of basic knowledge of market research and statistics, which means that we have certain market research ability. However, we still systematically learned the survey knowledge and interview skills that should be mastered in the survey process, so as to improve our own quality. In particular, those who act as investigators in in-depth interviews should strive to: avoid showing their superiority and superiority, and let the respondents relax; Detached and objective, but with grace and humanity; Ask questions in the form of providing information; Don't accept simple "yes" and "no" answers; Pry into the heart of the interviewee. In addition, each investigator is responsible for different parts of the investigation process according to

personal ability and comprehensive quality, strive to achieve orderly division of labor, develop strengths and avoid weaknesses, and ensure effective team communication throughout the whole investigation process. At the same time, in order to carry out the investigation activities efficiently, avoid the work delay caused by personal procrastination as far as possible, so that all relevant work cannot be carried out in time, and finally delay the investigation process of the whole team, the heads of each part shall supervise in turn to ensure that each part has a head and a supervisor, Ensure the smooth progress of the whole research process.

2.2.2. Quality Control

The operation technology and activities adopted to meet the quality requirements are called quality control. That is to say, the purpose of quality control is to eliminate the factors causing nonconformity or unsatisfactory effect in all stages of the quality ring by monitoring the quality formation process. Quality control refers to the activities of technical measures and management measures taken to make the products or services meet the quality requirements. The objectives of quality control include obtaining objective, accurate, true and effective data, reducing human errors, improving the scientificity of survey results, and avoiding the waste of people and property. Therefore, we have carried out strict quality control in the stages of questionnaire design, survey implementation and data sorting.

(1) Questionnaire design stage. Firstly, before the questionnaire design, each member has queried a large number of literature and materials, and referred to other questionnaire templates related to dairy products, so as to ensure the integrity and effectiveness of the questionnaire design content; Secondly, in the process of designing the questionnaire, we gave full play to the strength of all members. After the main person in charge of the questionnaire completed the questionnaire design, we handed it over to other members for review and supplement, further consulted the suggestions of other members, and conducted fierce discussion on controversial issues, so as to further optimize it. Then we sent the questionnaire to the teacher and repeated deletion and modification under the guidance of the teacher Weigh every word to achieve the rigorous and reliable design of the questionnaire; Finally, for some questions that may not be answered honestly, we set the same and opposite questions in different positions of the questionnaire to assist each other in verification, so as to improve the authenticity of the questionnaire design and ensure the data quality at all levels.

(2) Investigation implementation stage. We have carried out special and systematic learning before the survey, improved our own survey quality and ensured the acquisition of high-quality information, which is mainly reflected in the following aspects in the survey implementation stage: ① questionnaire survey stage. In the intercepting interview of pre survey, according to the characteristics of university town and intercepting interview, we chose to conduct the survey in supermarkets and shopping malls, which are crowded and conducive to rapid access. The expected sample size was achieved in a short time, which greatly shortened the time of pre survey and won more sufficient preparation time for follow-up investigation. ② Field research stage. During the field survey, we try to make the respondents feel relaxed to obtain more key information; Keep detached, objective, elegant and humane, which is conducive to the smooth development of the investigation; Ask questions in the way of providing information and strive for more information

sources; Do not accept simple "yes" and "no" answers, and strive to explore and discover the respondents' deep-seated values; Pry into the interviewee's heart and understand the most real views and opinions in his heart.

2.2.3. Pre Questionnaire Test

(1) Pre survey reliability test. According to the previous literature collection and sorting, combined with in-depth interviews, this survey finally determines to explore the five dimensions of bottled water purchase ability, bottled water purchase preference, influencing factors of bottled water purchase, bottled water purchase behavior and hot issues of bottled water drinking behavior of young people in Anhui Province. Because the reliability can only be analyzed on the scale, Therefore, before the reliability analysis of this survey, the text options are assigned according to various papers and documents in the preliminary preparation. In order to ensure the validity of the questionnaire, it is necessary to test the reliability of each scale on the basis of the pre survey questionnaire, and modify the questionnaire accordingly to form a formal questionnaire. Reliability is reliability. Reliability test refers to the reliability of the questionnaire, that is, the reliability of the questionnaire. It refers to the consistency of the results obtained when the same method is used to repeatedly measure the same object, that is, the degree of reflecting the actual situation. Firstly, the reliability index mainly represents the consistency, consistency, reproducibility and stability of the test results. It is a good measurement tool. If the same thing is measured repeatedly, the result should always remain unchanged. Secondly, reliability indicators are mostly expressed by correlation coefficients, which are mainly divided into three categories: stability coefficient (consistency across time), equivalence coefficient (consistency across forms) and internal consistency coefficient (consistency across items). Finally, there are four methods of reliability analysis: test-retest reliability method, replica reliability method, half reliability method and α reliability coefficient method. According to the survey characteristics and specific needs, this pre survey takes Cronbach 'alpha as the basis to measure the effectiveness of the questionnaire. Cronbach 'alpha can deal with the calculation of the internal consistency coefficient of any test to test the consistency of the scores of various items in the scale. It is the most common reliability coefficient at present. It is more suitable for the reliability analysis of attitude, suggestion and other scales. Therefore, it is also suitable for our quantified questionnaires.

(2) Pre investigation validity test. Reliability and validity is the basis to measure the accuracy and stability of a questionnaire test result. Reliability and validity test are two tests, one is reliability test and the other is validity test. Reliability test is used to measure whether the sample answer results are reliable, that is, whether the sample actually answers scale items. The higher the reliability coefficient, the more consistent and stable the test results are; Validity test is used to measure whether the item design is reasonable, which is verified by factor analysis. The two test methods of reliability and validity analysis are independent of each other, but the test results complement each other. They need to be analyzed together to get the desired results. Reliability is the basis of validity, which must have reliability to be effective. If the reliability is low, the validity cannot be high. But the reliability and efficiency are not necessarily high. In addition, the research objects of the two are different. The research object of reliability is the respondent and the research object

of validity is the item. Among them, validity can be divided into content validity and construction validity. Construction validity is also known as structural validity. Generally speaking, exploratory factor analysis of content validity can be done with SPSS software. In the process of pre investigation validity test, there is no a priori exploration to divide the scale, so SPSS software is often used for exploratory factor analysis.

3. Analysis of Bottled Water Market Survey

3.1. Data Overview

In this data analysis, we collected 1759 valid data, involving 42 questions, including basic information, purchase behavior, purchase preference, influencing factors, bottled water hot issues and influencing paths.

3.2. Basic Information of Respondents

(1) Gender. Among the youth groups surveyed, 33.50% of the youth are boys and the remaining 66.50% are girls. There are more girls and fewer boys.

(2) Age. Among the youth groups surveyed, 34.40% were 18-24 years old, 33.76% were 30-35 years old and 31.84% were 25-29 years old.

(3) Monthly disposable expenses. Among the youth groups surveyed, nearly half (44.05%) of the youth have a monthly disposable amount of 3000-8000 yuan, nearly one fifth (20.72%) have a monthly disposable amount of less than 3000 yuan, accounting for 24.71% within 8000-15000 yuan, and

more than 15000 yuan (10.52%) are also distributed.

3.3. Purchase Preference

3.3.1. Descriptive Statistical Analysis

(1) Species preference. Among the youth surveyed, 1759 respondents chose refreshing taste, 789 respondents chose purified water and 1017 respondents chose mineral water. It can be inferred that young people prefer pure water or mineral water.

(2) Packaging preferences. Of the 1759 young people interviewed, 78.3% chose bottled water in plastic bottles. In addition to bag materials, 1084 people like simplicity and generosity in the choice of packaging style.

(3) Brand preference. More than half of the young people habitually choose the same brand of bottled water, while 24.2% said it doesn't matter.

(4) Time preference. Among the 1759 young people surveyed, more people would choose to buy bottled water for dinner, travel or business, while for most people, bottled water is more an occasional choice.

(5) Reason for purchase. From the choices of 1759 young people, we can see that the reasons why people choose to buy bottled water are different, but the convenience of buying bottled water when thirsty is the common reason why most people choose to buy bottled water.

3.3.2. Analysis of Youth's Preference for Bottled Water Based on Contingency Table Analysis

(1) Monthly disposable expenses for contingency analysis between packages.

Table 1. Two dimensional contingency table (1)

Monthly disposable expenses	Plastic bottle	Glass bottles	Other	Total
Below 3000 yuan	238	67	57	362
3000—8000 yuan	565	155	16	736
8000—15000 yuan	268	135	12	415
Over 15000 yuan	130	107	9	246
Total	1201	464	94	1759

The following is a two-dimensional contingency table of monthly disposable expenses and packaging:

Table 2. Two dimensional contingency table (2)

Revenue and packaging cross table				
Monthly disposable expenses	Plastic bottle	Glass bottles	Other	Total
Below 3000 yuan	238	67	57	362
Expected value	127.4372	26.2932	8.2696	
Difference	10.5628	-9.2932	-1.2696	
3000-8000 yuan	565	155	16	736
Expected value	342.9791	70.7644	22.2565	
Difference	22.0209	-15.7644	-6.2565	
8000-15000 yuan	268	135	12	415
Expected value	90.4647	18.6649	5.8704	
Difference	-22.4647	16.3351	6.1296	
Over 15000 yuan	130	107	9	246
Expected value	40.1991	8.2775	2.6034	
Difference	-10.1991	8.2775	1.3966	
Total	1201	464	94	1759

Calculate each expected value and difference, substitute it into the formula and perform chi square test:

Table 3. Chi square test

Chi square	Degree of freedom	P value
48.9288	6	<0.05

It can be seen from the table that the person chi square independence test corresponds to $P < 0.05$, that is to reject the original hypothesis, so there is a direct connection between the monthly living expenses and packaging, not independent. Therefore, there is a correlation between youth's preferred packaging and youth's monthly disposable expenses.

3.3.3. Study on Youth's Preference to Buy Bottled Water Based on Decision Tree Model

(1) Selection of calculation method. Decision tree algorithms based on information theory include ID3, CART and C4.5, of which C4.5 and CART are derived from ID3 algorithm. CART and C4.5 support the processing when the data feature is continuous distribution. It mainly uses binary segmentation to process continuous variables, that is, find a specific value - split value: if the feature value is greater than the split value, go to the left subtree or go to the right subtree. The selection principle of this split value is to reduce the "degree of chaos" in the divided subtree. Specifically, C4.5 and cart algorithm have different definitions.

The ID3 algorithm was invented by Ross Quinlan and is based on the "Occam razor": the smaller the decision tree, the better the larger the decision tree (be simple theory). ID3 algorithm evaluates and selects features according to the information gain of information theory, and selects the feature with the largest information gain each time as the judgment module. ID3 algorithm can be used to divide nominal data sets without pruning. In order to remove the problem of excessive data matching, adjacent leaf nodes that cannot produce a large amount of information gain can be cut and merged (for example, setting the information gain threshold). In fact, there is a disadvantage of using information gain, that is, it prefers attributes with a large number of values - that is, in the training set, the more different values taken by an attribute, the more likely it is to take it as a split attribute, which is sometimes meaningless. In addition, ID3 can not deal with continuously distributed data features, so there is C4.5 algorithm. CART algorithm also supports continuously distributed data features.

C4.5 is an improved algorithm of ID3, which inherits the advantages of ID3 algorithm. C4.5 algorithm uses information gain rate to select attributes, which overcomes the disadvantage of selecting attributes with more values when using information gain, and prunes in the process of tree construction, can complete the discretization of continuous attributes, be able to process incomplete data. The classification rules generated by C4.5 algorithm are easy to understand and have high accuracy. But the efficiency is low, because in the process of tree construction, the data set needs to be scanned and sorted several times. Also because multiple data set scans are required, C4.5 is only suitable for data sets that can reside in memory.

The full name of cart algorithm is classification and regression tree. Gini index (the feature s with the smallest Gini index) is used as the splitting standard. At the same time, it also includes post pruning operation. Although ID3 algorithm and C4.5 algorithm can mine as much information as possible in the learning of training sample set, their decision tree branches are large and scale is large. In order to simplify the scale of decision tree and improve the efficiency of generating decision tree, the decision tree algorithm cart, which selects test attributes according to Gini coefficient, appears. Based on the data characteristics, we use ID3 algorithm.

(2) Data analysis. ① According to the previous investigation and research, we selected four characteristics: Youth's monthly living expenses, whether to buy the same brand repeatedly, the number of purchases per week and packaging. ② Generate a decision tree according to the selected nodes, non leaf nodes, etc. The criteria for selecting which of the four characteristics of youth monthly disposable

expenses, whether to repeatedly purchase the same brand, the number of purchases per week and packaging is the root node is the maximum information gain. Information entropy: in probability theory, information entropy gives us a way to measure uncertainty. It is used to measure the uncertainty of random variables. Entropy is the expected value of information. If the things to be classified may be divided into n classes, namely X_1, X_2, \dots, X_n , and the probability of each is p_1, p_2, \dots, p_n , then the entropy of X is defined as:

$$H(X) = -\sum_{i=1}^n p_i \log p_i \quad (4)$$

Information gain indicates the degree to which the uncertainty of Y is reduced after knowing the information of feature X. Defined as:

$$g(D, A) = H(D) - H(D|A) \quad (5)$$

Calculate the information gain of each Eigenvalue according to the formula, as shown in the following table:

Table 4. Information gain table (1)

	Number of copies	Probability	Log value
Yes	1068	0.5563	-0.84615
No	691	0.4437	-1.17223

The information gain value under each feature is shown in the following table:

Table 5. Information gain table (2)

	Monthly living expenses	Packing	Number of purchases per week
Information gain	0.0091	0.0630	0.0074

Among the three features, the monthly cost of living information gain is 0.0091, the packaging information gain is 0.0630, the weekly purchase times information gain is 0.0074, and the repeat purchase information gain is 0.9908. According to the principle of maximum information gain, packaging is selected as the root node. That is, the primary feature of judging whether a young person purchases repeatedly is the cost of living of the young person.

Under the characteristic of youth monthly living cost, the next non leaf node is determined. Similarly, calculate the information gain under condition A (monthly living cost of youth). Assuming that there are random variables (X, Y), the joint probability distribution is:

$$P(X = x_i, Y = y_j) = p_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (6)$$

Conditional entropy ($H(Y|X)$) represents the uncertainty of random variable Y under the condition of known random variable X, which is defined as the mathematical expectation of the entropy of conditional probability distribution of Y on X under given conditions.

$$H(Y|X) = \sum_{i=1}^n p_i H(Y|X = x_i) \quad (7)$$

The calculation results are as follows:

Table 6. Information gain table (3)

	Packing	Number of purchases per week
Information gain	0.0212	0.0504

It can be seen from the above table that under condition a, the information gain of feature "purchase times per week" is 0.0504 and the information gain of "packaging" is 0.0212. Therefore, select "number of purchases per week" as the next non leaf node. That is, when we judge whether a young person will repeatedly buy bottled water of the same brand, we first consider the monthly living expenses, then consider the number of purchases per week, and finally consider the packaging.

(3) Build a decision tree. According to the decision tree, we can predict the youth's preference for daily bottle purchase. If the monthly disposable cost of youth a is less than 3000 yuan and buys bottled water 3-6 times a month, if he likes bottled water in plastic packaging or other packaging, he may not like to buy the same brand; If he likes glass bottles, he will

probably choose the same brand of bottled water.

References

- [1] He Xi. Why NongFu Spring messed up marketing this time [J]. China Food, 2021 (16): 68-69.
- [2] Cheng Xuefei. Good ecology, water conservation, good environment and ensuring water quality [N]. Chengde Daily, 2021-09-01(02).
- [3] Guo Wei. Exploration of brand design diagnosis in the new media era -- Taking Baishan brand as an example [J]. Popular Literature and Art, 2021 (08): 45-46.
- [4] Zhu Danpeng. The new generation of consumers is subverting the industrial development model -- ten industry development trends of China's food and beverage industry in 2021 [J]. China Food, 2021 (06): 40-49.
- [5] Qiao Mingli. Research on development strategy of SW seawater desalination technology company [D]. Shandong University, 2021.