Enterprise is faced with the problem of falling sales in the third quarter of 2020 wal-mart, in order to completely solve the problem, the company hired researchers, the survey found that every batch is due to the enterprise purchases unified category of goods, and some fundamental oversupply, lead to goods unsold, and fresh class will deteriorate, leading to economic benefits from the enterprise. To solve this problem, they took some statistics to figure out the categories of supermarket unmarketable products,

And the best selling category every day. To reduce or increase the better, in the fourth quarter of 2020 record revenues reached $152.1 billion, up $10.4 billion, up 7.3%, according to this method, we for the country's facing the same problem also took some small supermarket model method (mentioned in the article have) to be based on science and technology plan of data statistics and analysis, The conclusion is analyzed and reported to alleviate the problem faced by small and medium-sized supermarkets nearby.Preface: with the rapid development of retail industry, the competition of domestic retail industry is becoming increasingly fierce; At the same time, with the gradual relaxation of market access conditions for domestic retail industry, international retail giants have landed one after another. These super chain retail enterprises have created great competitive pressure on domestic peers with strong strength, advanced management concept and excellent service. In view of the problems affecting the sales volume of food commodities, we conducted practical investigation and research with supermarkets as the investigation location. We will design and summarize the models of these commodities, draw conclusions through the data, improve the adverse factors existing in the traditional business model, achieve reasonable and efficient operation, obtain the maximum sales volume and obtain the best interests, So as to put forward some suggestions that can improve the sales volume of food commodities, and provide a feasible basis for the sales means and methods of retail enterprises. Therefore, in this report, we study the top items purchased by members, numbers of orders received each day and other survey items to investigate the situation of supermarket consumer goods purchased by consumers and the business status of supermarkets, so as to find out what the commodities with the largest sales volume and the least sales volume are respectively. In this regard, it can provide businesses with some valuable information reference when purchasing goods, so as to attract customers and retain customers, so as to increase sales. Model analysis is also the top priority. Lu Hengtai believes that we must rely on the power of science and technology, skillfully use computers, carry out model analysis, make data statistics and draw charts, and cannot carry out analysis with the help of traditional theories, This consumes both material and human resources, and the results are not satisfactory (from using time series analysis); In addition, global economic integration has intensified people's requirements for data, and the quality of data is one of the important factors restricting this. The use of scientific statistical methods can improve the quality of statistical data to a certain extent. Statistical data quality is a systematic, comprehensive and multi-dimensional conceptual system (NIE Fuqiang's definition of international statistical data quality); Statisticians are also the bridge and link to transform statistical strategies into statistical data; Implementation is a necessary requirement for obtaining correct data. (Wang Hua, statistical quantity evaluation)

》）

From Chen Wenbing, Wal Mart Model vs Carrefour model It can be concluded that consumers have different choices when they go to each different supermarket. Convenience, cheapness, novelty and comfort are their reference standards. For example, customers choose to go to RT Mart because the things in it are good and cheap, provide high-quality services and amazing commodity discounts, so the analysis of consumer shopping trends is an essential link, and Consumer shopping tendency analysis is also an indispensable part of model analysis,

From - Chen Ling's consumption guide; by dividing these information monthly and annually, it can be seen that the enterprise's sales situation is good in a few months, while in the months with poor performance, measures are taken to avoid being robbed of customers by peers. From Chen Tao's proceedings of the Sixth Central China business and Economic Forum,

This study takes into account that the ordinary Apriori algorithm is too basic and has significant disadvantages: when the transaction database is large, the number of candidate frequent K itemsets is huge; when verifying the candidate frequent K itemsets, it needs to scan the whole database, which is very time-consuming.

Therefore, after discussion, we decided the following improvement scheme: using FP growth algorithm, we introduced a special data structure to temporarily store data to reduce the number of I / O; using MapReduce framework to map block the input original data set, the main process distributed these data blocks to each node in Hadoop cluster, and then each node processed these data blocks in parallel The FP Tree data structure and MapReduce program structure are referenced, and some of the tasks in each stage are shown in the code. The full name of FP Tree algorithm is frequentpattern tree algorithm, which is the frequent pattern tree algorithm. Like Apriori algorithm, it is also used to mine frequent itemsets. However, FP Tree Algorithm is the optimization processing of Apriori algorithm, which solves the Apriori algorithm There will be a large number of candidate sets in the process, and FP Tree algorithm finds frequent patterns without generating candidate sets. However, after frequent patterns are mined, the steps of generating association rules are the same as apriori. When customers shop, the following operations will be performed according to the map code:

Shopping basket T1: crackers, icecream, coke

Shopping basket T2: icecream, coke, crackers

According to the association rules, the following rules will be generated for T1

[(crackers,icecream),1]

[(crackers,coke),1]

[(icecream,coke),1]

For T2, the following rules will be generated

[(icecream,coke),1]

[(icecream,crackers),1]

[(coke,crackers),1]

We can see that there are six pairs of rules, but we find that (crackers, icecream) and (icecream, crackers) are the same. They will be divided into different rules here. Therefore, before generating rules, you need to sort the goods alphabetically to avoid this problem

Combinations in map is a simple tool. A given set can be generated by using combinations. Generatecombinations (S1, S2,..., Sn). Assuming that the shopping basket is {a, B, C}, and a rule with two commodities is generated, the classification results are as follows: [a, b], [a, C], [b, C], which further simplifies the statistical difficulty.

According to the above model, we can get the statistical results of top products purchased by members, products purchased by month and the number of orders received every day

We installed a control device for the supermarket business scanning system. As long as the customer buys and pays, the computer will classify the goods. From the model, it can be seen that this includes: milk, fruit, water, beer, egg, etc. In addition, in order to prevent omissions, we will hire computer personnel to check whether the system works normally, whether there are loopholes and vulnerabilities at the end of each day. You can get the daily purchase frequency of goods. Among them, January accounted for the highest proportion, reaching 8.6% of the pie chart, followed by August and may, 9.0% and 8.6%; (the data comes from items bought split by month) after one month, we can get the monthly report through statistics of the collected information. The monthly report can clearly see that the goods purchased each month are the most and the goods purchased are the least; And set the category of each commodity in the computer, so that the computer can automatically collect this information and obtain the annual report after receiving the prompt of successful payment. From the statistical report, we can see that in the daily purchase statistics, through the analysis of the numbers of orders received each day, it can be concluded that, in comparison, the purchase rate on Thursday is the highest.

From this statistical report, we can conclude that milk and soda are more popular drinks for customers; When investigating the items born by month, we can conclude that most of them were purchased in August. In the weekly report, we can conclude that the sales rate on Thursday is very high. Based on the above information, we can provide enterprises with useful business information on procurement and other related matters.