Vegetable sales data mining: association rule analysis

Jiahe Liu*

International business college, Henan University, Kaifeng, China, 475001

*Corresponding author: ljh12345@henu.edu.cn

Abstract. Supermarkets sell a large number of vegetable categories, and there is a promotion or inhibition relationship between the sales volume of each product. In order to explore the law of sales volume between different products, this paper first preprocesses the data, counts the total sales volume of single products and the total sales volume of each category, and draws a bar chart to reflect the distribution law of the sales volume of each single product through visual analysis, for the correlation between different single products and categories, the correlation between categories is examined first, and the correlation between categories is pre-adopted respectively by using the Pearson and Spearman correlation analysis, and the correlation between the categories and sales volume is examined by drawing a Q-Q chart, and it is found that the distribution law does not satisfy the normal distribution, and thus the Spearman correlation analysis is adopted. The Q-Q diagram was drawn for testing, and it was found that the distribution pattern did not satisfy the normal distribution, thus Spearman correlation analysis was used to analyse the correlation, and it was found that the correlation between the categories showed a strong correlation. For different single product, we carry out Kendall consistency test for all single products, and the test results show that there is a good consistency between the single products. In order to make the analysis more comprehensive, this paper concludes with some suggestions for data collection.

Keywords: Spearman's Correlation Coefficient, Kendall's Consistency Test, Vegetable Marketing.

1. Introduction

Due to the super sales of many types of vegetables, origin is not the same, and the general freshness of vegetable commodities are relatively short, the quality will be with the increase in sales time and deterioration, the general pickers in the morning from 3:00 to 4:00 am to carry out the purchase of vegetables trading, so some of the vegetables such as the day did not sell, then the next day can not be sold for shipping losses and poor quality of the goods are usually sold at a discount. This paper gives the historical sales and demand of each commodity for correlation analysis. This paper analyses the correlation between different kinds of vegetable products or different single products according to the commodity information of 6 vegetable categories distributed by a supermarket; at the same time, it analyses the distribution pattern and correlation of the sales volume of each vegetable category and single product according to the sales flow details of the supermarket from July 1, 2020 to June 30, 2023[1].

2. Modelling and solving

In order to study the distribution pattern between different individual items and different categories, we firstly counted the total sales volume of different individual items from July 2020 to June 2023, and we normalised the total sales volume in the data pre-processing stage, then in order to further reflect the distribution pattern, we plotted the bar chart between the sales volume of each individual item and the name of the individual item through the visual analysis, as shown in Figure 1.

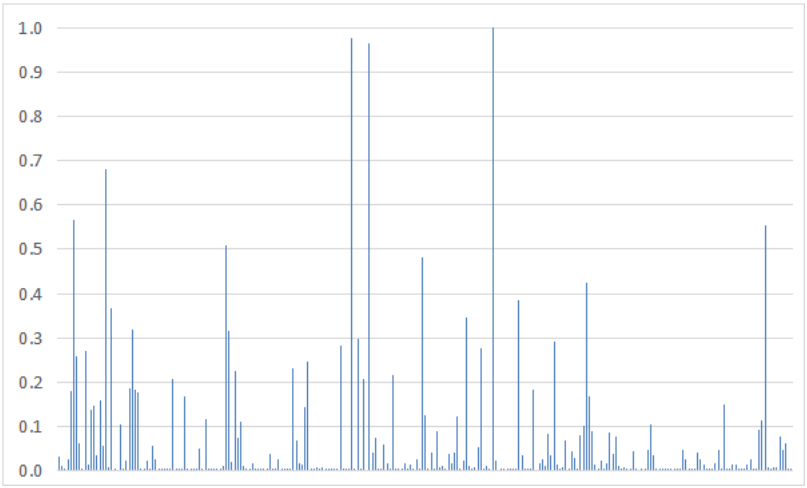

Figure 1. Histogram of sales and individual items

For the distribution law between different categories we in order to consider its trend with the seasons, the sales date is filtered, screened out the data of a whole year from July 2020 to June 2021, respectively, statistics of spring, summer, autumn and winter four seasons of the sales volume of each category, draw a line graph for visual analysis. As Figure 2 and Table 1.

Table 1 Vegetable Sales Chart by Season

	Foliage	Edible Mushrooms	Cauliflower	Aquatic Roots	Chilli	Eggplant
Spring	38130	36614	3655	33586	22767	22983
Summer	40374	38066	4495	32033	17934	18082
Autumn	38237	36558	4924	27721	15782	15958
Winter	147131	140371	3987	118946	73559	74295

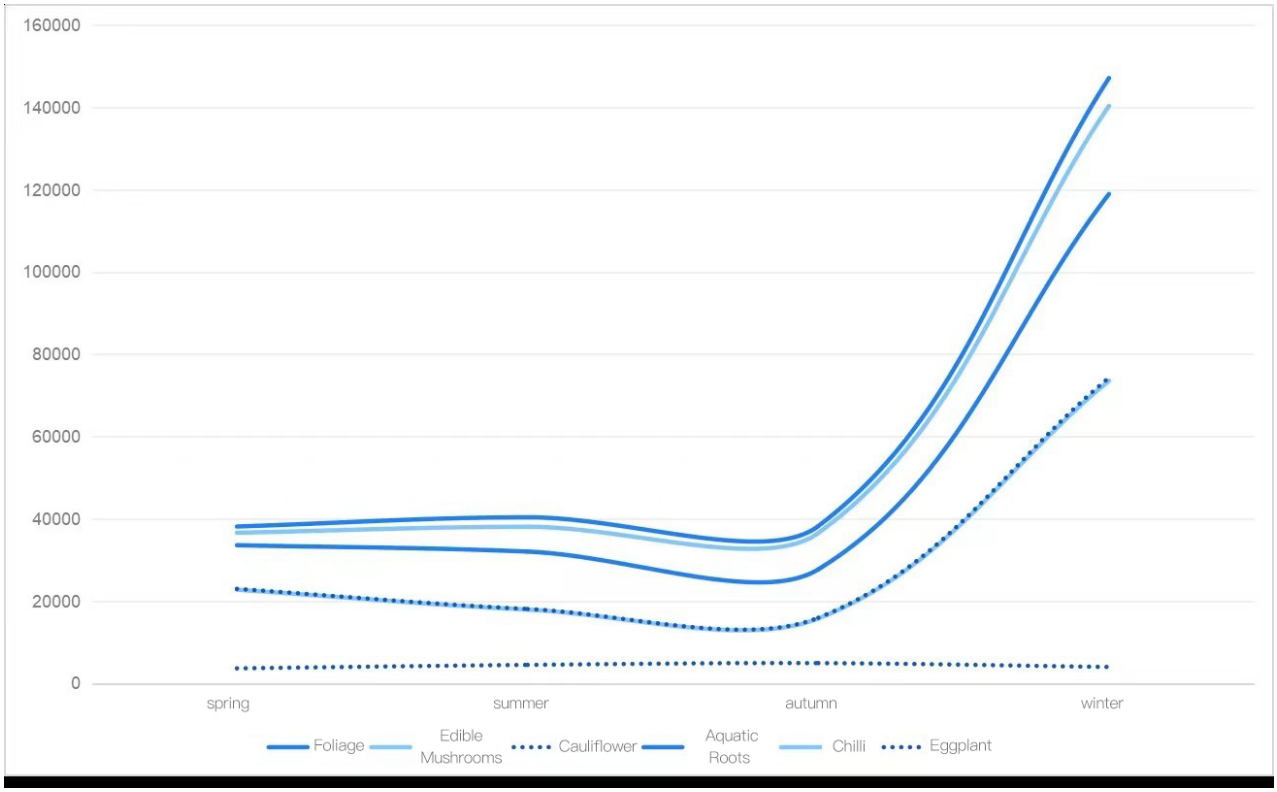

Figure 2. Vegetable sales volume line graph for all seasons

By observing the charts, it can be seen that these six vegetable category sales have consistent sales volume changes with the seasons, which suggests that the distribution of vegetable category sales may be seasonally related. There is a significant increase in the sales volume of vegetables in each category from autumn to winter, which is not in line with the daily perception of life[2]. We try to explain this anomaly:

The period from July 2020 to June 2021 is the period of the epidemic, and given the uncertainty associated with the epidemic, some consumers tend to stockpile food in case of emergencies. Vegetables can be refrigerated and frozen to extend their shelf life, and thus may trigger panic buying behaviour among consumers.

For the correlation between different categories, this paper wants to use Spearman and Pearson correlation coefficients to analyse the correlation, and we first use Q-Q plots to test for normal distribution as shown in Figure 3.

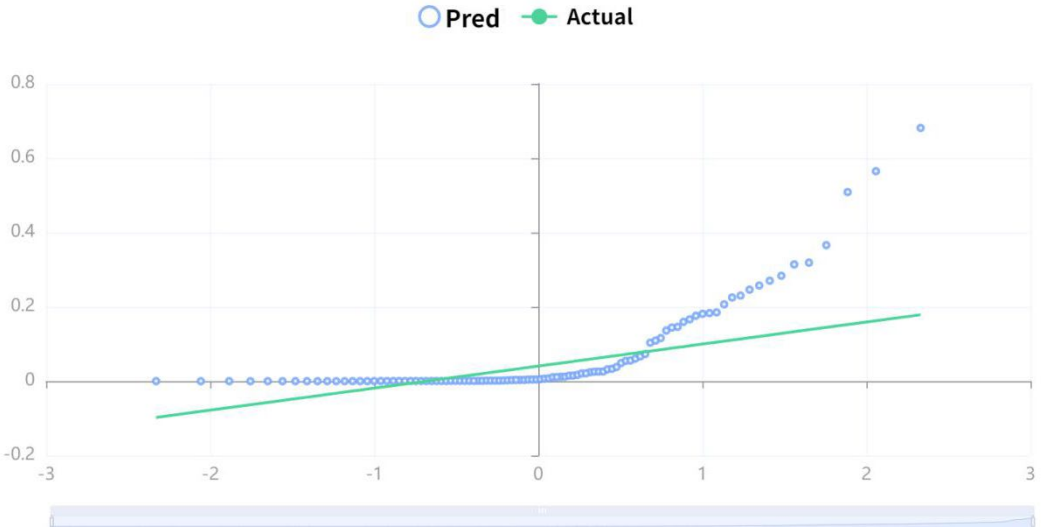

Figure 3.Q-Q diagram

It is observed that the individual scatter points have a large deviation from the straight line, so it is initially judged that it does not conform to the normal distribution. Therefore, the method of Pearson correlation analysis was excluded and the sales volume of each category was analysed using the Spearman correlation coefficient[3].

The results of the analysis are shown in Table 2 (significance p-value taken as 0.05)

Table 2 Results of Spearmans correlation coefficient analysis

	Foliage	Edible Mushrooms	Cauliflower	Aquatic Roots	Chilli	Eggplant
Foliage	1(0.000***)	0.943(0.005***)	0.829(0.042**)	0.771(0.072*)	0.829(0.042**)	0.829(0.042**)
Edible Mushrooms	0.943(0.005***)	1(0.000***)	0.943(0.005***)	0.6(0.208)	0.943(0.005***)	0.943(0.005**)
Aquatic Roots	0.829(0.042**)	0.943(0.005***)	1(0.000***)	0.486(0.329)	1(0.000***)	1(0.000***)
Cauliflower	0.771(0.072*)	0.6(0.208)	0.486(0.329)	1(0.000***)	0.486(0.329)	0.486(0.329)
Chilli	0.829(0.042**)	0.943(0.005***)	1(0.000***)	0.486(0.329)	1(0.000***)	1(0.000***)
Eggplant	0.829(0.042**)	0.943(0.005***)	1(0.000***)	0.486(0.329)	1(0.000***)	1(0.000***)

Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.

The analysis of Figure n shows that the significance of each correlation coefficient is basically less than 0.05, indicating that there is a correlation between categories, and the correlation coefficient can reflect the size of the correlation between different categories, for example, for the foliage category the strongest correlation with it is edible mushrooms, aquatic roots and tubers, peppers and eggplants, and cauliflower category in the order of the strongest correlation. In order to represent the correlation more intuitively, the heat map of correlation coefficient is drawn in this paper, as shown in the figure 4.

Figure. 4. Graphical representation of the consistency test

For the correlation between different single product we use Kendall consistency test for correlation analysis of all single product, after the test found that the overall data of the significance of the P-value of 0.000***, the level of the level of significance, reject the original hypothesis, so the sales of each single product presents consistency, at the same time, the model of Kendall's coefficient of coordination W value is 0.777, so the degree of correlation of the sales volume of each individual product is high degree of consistency test part of the results are shown in Table 3.

Table 3 Consistency test results

Kendall's W Analysis Results					
Name	ordinal mean	median	Kendall's W factor	X^2	P
White Jade Mushroom(Bag)	37	396.5			
Spinach	44	599.068			
Vegetable Heart	46	698.001			
Cordyceps Flower	10	7.924			
Chinese cabbage	60.75	3711.221			
Dai Long Eggplant	31	244.962			
Lantern Pepper(1)	16	43.578			
East Gate Chinese Cabbage	30.25	144.179	0.777	192.818	0.000***
Tall Melon(1)	23.75	95.932			
Seafood Mushroom(1)	24.5	78.01			
Seafood Mushroom(Bag)(1)	32.5	237.5			
Red Bell Pepper(1)	20.25	64.517			
Red Bell Pepper	26	118.826			
Red Bell Pepper(1)	40	347.255			
Sweet Potato Tips	34	358.643			

Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.

In order to make our off-contact analyses more accurate and collect more data, we also need to consider the following aspects:

Consumption data on investment in preservation technologies

Investment in preservation technology has a significant impact on the quality aspect of storage, preventing the deterioration of vegetables and thus reducing the rate of loss, which can be more profitable[4].

The quality of vegetables, and thus reduce the rate of loss, you can get a greater profit. The investment in preservation technology requires statistics on its consumption costs.

The consumption costs of investments in preservation technologies need to be calculated in order to better select categories with lower preservation costs and higher profitability.

Customer impact

The size of customer traffic and the age of customers will also have an impact on product sales to a certain extent. For customer consumption data, statistics should be based on the age of customers, and the types of vegetables that customers of different ages like to buy and the number of purchases should be counted.

Sales data of other supermarkets

When selling goods, the pricing strategy of supermarkets also needs to take into account the influence of other supermarkets.

When the price of an item fluctuates within a certain range, people will consider the amount of effort and price spent and the freshness of the vegetables purchased.

As a result, it is necessary to collect data on food sales from other supermarkets to inform pricing in order to gain an advantage over competing merchants[5-6].

Data on seasonal changes in food sales

Vegetable sales are generally related to the time of the season, under the influence of the environment, the need for statistics in different seasons, time periods, holiday sales changes, for different times and climates will have an impact on the sales of vegetables, in different seasons and climates, the need to formulate different discount strategies to improve sales[7-8].

Vegetable storage intelligence

Supermarket warehouse purchases and outputs need to carry out detailed statistics, focusing on statistical replenishment of categories and sales, while observing their replenishment time, supermarkets also need to have an accurate understanding of the number of commodities on sale in order to timely replenishment and scheduling.

Customer evaluation and feedback

Customer evaluation and feedback is conducive to the superstore to further improve the various types of dishes need statistics.

Suggestions on the demand, as well as other super in the vegetable sales problems, and then improve or strengthen their own construction, in order to receive customer praise for the super, but also to increase the customer's goodwill for the super, to a certain extent, improve the customer's favour for the super, to improve the super vegetable sales.

Past replenishment of the wrong amount of products

In the past sales process, it is inevitable that there will be due to a variety of accidents appeared in the sales volume of the replenishment of the wrong, although a small part of the factors.

Although it is a very small part of the factors, but in order to ensure that the replenishment will not be under replenishment of the situation, it is necessary to count the category names of vegetables, so as to increase the sales of supermarket vegetables.

Although it is a small factor, it is necessary to count the category names of vegetables in order to ensure that there is no under replenishment when replenishing the products. It is also important to keep in mind that the product numbers in stock may be incorrect, which can lead to discrepancies when replenishing products.

Quantity and quality of the source of supply

Supermarkets often have multiple sources of supply at the same time and need to consider the quality of the vegetables from each source as well as the quality of the source of supply[9-10].

The quality of vegetables and supply sources of the road expenses, thus the purchase of goods to the location of the purchase of goods to screen the amount of goods, select the best sources of goods in order to enhance sales, while selecting their sources of goods to get the maximum price of the products, the profit.

3. Conclusions

This paper first preprocesses the data, draws a bar chart to reflect the distribution law of the sales volume of each single product through visual analysis, for the correlation between different single products and categories, first consider the correlation between categories, pre respectively using Pearson and Spearman correlation analysis, first draw Q-Q charts for the test of the categories and sales volume, and found that the distribution law does not satisfy the normal distribution, and thus using Spearman Spearman correlation analysis was used to analyse the correlation, and it was found that there was a strong correlation between the categories. For different individual products, we conduct Kendall consistency test for all individual products, and the test results show that there is good consistency between individual products. In order to collect relevant data to better analyse the relationship between sales data to promote product sales, first of all, you can make investment in preservation technology, to get the investment cost and profit changes, in order to seek higher profits; statistics on the age of customers to buy vegetables preferences, to facilitate the adjustment of the location of vegetable placement; collection of food sales data and pricing of other superstore to facilitate the adjustment of the price and the amount of stock; statistics on seasons, time periods and holidays Sales volume changes, in order to adjust the pricing and stocking strategy in response to the environmental impact; statistics on the purchase and output of the warehouse, the inventory, to facilitate the replenishment scheduling; statistics on the customer's evaluation of the super service and the amount of vegetables, vegetables, such as the evaluation and recommendations, to facilitate the adjustment of the improvement.

References

[1] BER Renliu, ZHAO Fei, HE Yinju et al. Pesticide residue detection of vegetables in Maonan Township of Pingtang Kapu to help rural revitalisation[J]. Yunnan Chemical Industry,2023,50(08):166-168+179.

[2] Li Xuehui. Shandong vegetable industry development countermeasures research[D]. Dalian Ocean University,2023.

[3] Zhang Ruqi. Analysis of Changchun residents' willingness to consume organic vegetables and its influencing factors[D]. Jilin Agricultural University,2022.

[4] WANG Shuren,CHEN Dan,HUANG Zhifeng et al. Cultivating Vegetable Leading Enterprises and Developing Order Farming--An Analysis of the Development Mode of Order Farming by Leading Vegetable Enterprises in Chongming[J]. Shanghai Rural Economy,2022(10):25-27.

[5] M. A A P,Shadia D,E. J H, et al. Effectiveness of a Multi-Strategy Behavioral Intervention to Increase Vegetable Sales in Primary School Canteens: A Randomised Controlled Trial[J]. Nutrients,2022,14(19).

[6] Chiranjit C,Kumar P G,Kumar R P, et al. Solving a Mathematical Model for Small Vegetable Sellers in India by a Stochastic Knapsack Problem: An Advanced Genetic Algorithm Based Approach[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,2022,30(05).

[7] Feng Guanghui, Zhang Haijuan, Yue Wei et al. Thoughts and Suggestions on Stable Production and Supply of "Vegetable Basket" Products in Zibo City[J]. China Agricultural Extension,2022,38(09):19-20.

[8] Zhang Jianwei. Science and technology to enhance the quality of brand-driven sales - Shouguang Hanxing Vegetable Co-operative Society to bring people out of the road to prosperity[J]. China Farmers' Cooperatives,2022(09):54-55.

[9] Wang Jing,Zhang Huihui. Affordable vegetables make the "food basket" full[N]. Qinghai Daily,2022-09-09(006).

[10] Zhu Baocai,Li Tiegang. Pinggang town to extend the service to vegetable development, production, sales of the whole process[N]. Tieling Daily,2007-07-17(006).