Agricultural & Forestry Economics and Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/agrfem.2024.070210
ISSN 2616-2202 Vol. 7 Num. 2

Research on Seasonal Changes and Interrelationships of Vegetable Sales Volume Based on Data Mining

Han Lin*

College of Mathematics and Computer Science, Shantou University, Shantou, 515063, China
**Corresponding author: 21hlin3@stu.edu.cn*

Keywords: Sales volume, Seasonal variations, Correlation analysis, Association rule analysis

Abstract: This study analyzes the sales volume and pricing of various vegetable products. Data visualization and analysis were performed on the sales volumes of different categories and individual products within the vegetable category. The analysis reveals that Wuhu green pepper ranks highest in single product sales, followed by broccoli, net lotus root, Chinese cabbage, and Yunnan lettuce. Leafy vegetables, chili peppers, and edible fungi constitute 77% of the total vegetable sales, with eggplant vegetables having the lowest sales volume. Seasonal variations in sales volumes were observed, with flower and leafy vegetables peaking in February 2023 and reaching their lowest in April 2022. The sales volumes of eggplant, cauliflower, and aquatic root vegetables remain relatively stable throughout the year. Correlation analysis using the Pearson coefficient and Apriori algorithm identified relationships between different vegetable products. The heat map of correlation coefficients shows complementary and substitute relationships among top-selling vegetables. Association rule analysis suggests strategic placement of certain vegetables to enhance sales. This comprehensive analysis provides valuable insights into vegetable sales trends and relationships, offering practical recommendations for improving sales strategies.

1. Introduction

The vegetable sales industry is a critical component of the agricultural sector, directly impacting both the economy and the dietary habits of the population[1,2]. With the rapid development of information technology and data analysis methods, the integration of data-driven decision-making has become increasingly important in optimizing supply chains and enhancing market efficiency[3,4]. This study focuses on the analysis of vegetable sales data, which is essential for understanding market trends, consumer preferences, and the overall performance of different vegetable categories and products.

The primary objective of this research is to visualize and analyze the sales volume of various vegetable categories and individual products. This includes identifying the top-selling products, understanding the seasonal variations in sales, and examining the correlation between different vegetable products. By extracting sales records for products labeled as "returned" or "discounted," the study also highlights the dynamics of lower-quality product sales.

Furthermore, the research investigates the interrelationships between sales volumes of different vegetable categories and individual items using correlation analysis and association rule mining techniques[5]. The results provide insights into complementary and substitute relationships among products, which can help retailers optimize product placement and marketing strategies to boost sales.

2. Visualization Analysis of Vegetable Sales Data

The data in this paper originates from http://www.mcm.edu.cn. Based on the related vegetable data, this study merges the vegetable product names from different datasets with the sales transaction data and matches the product codes to obtain daily sales volumes and wholesale prices (in yuan per kilogram). This approach clarifies the data on product loss rates, making it more straightforward and reducing the complexity of information retrieval. Additionally, sales records for products labeled as "returned" or "discounted" are extracted as a separate category of lower-quality products. Based on the data that has been processed, the data visualization and analysis of the sales volume of different categories of goods and different individual products in the vegetable category is carried out.

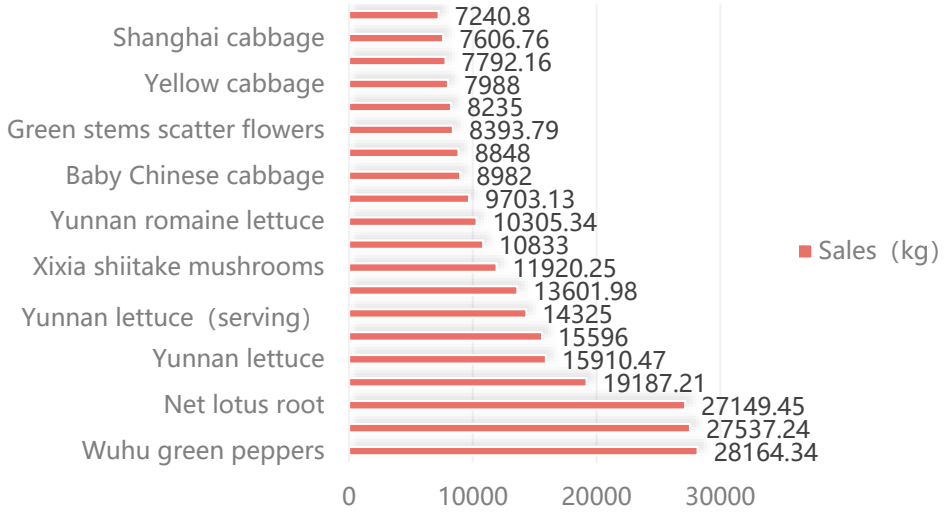

Figure 1: Top 20 List Product Sales Chart

Figure 1 is the top 20 single product sales chart. In the same cycle, Wuhu green pepper ranked first among all single products with total sales of 28164.34kg, which shows that Wuhu green pepper is well loved in the market. The top five vegetable items in total sales are: Wuhu green pepper > broccoli > net lotus root > Chinese cabbage > Yunnan lettuce.

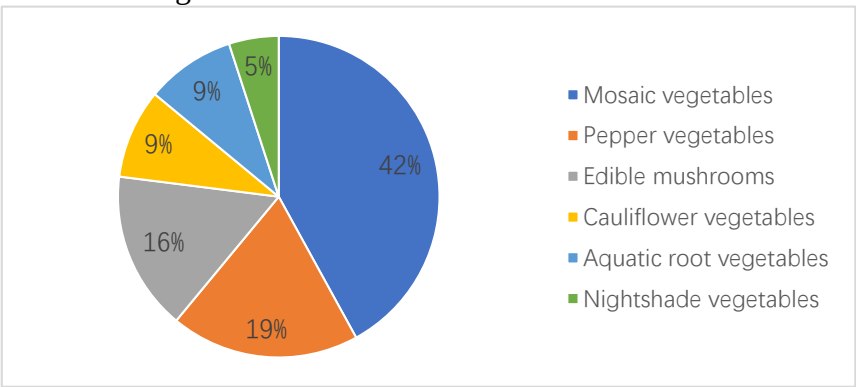

Figure 2: Sales distribution by category

As shown in the figure 2, in the vegetable category, the sales of leafy vegetables are the most, accounting for 42% of the overall vegetable category, chili peppers account for 19% of the overall vegetable category, edible fungi account for 16% of the overall vegetable category, leafy vegetables, chili peppers and edible fungi account for 77% of the sales of all the categories, which shows that the mainstream of the people's daily consumption of vegetable products is: leafy vegetables, chili peppers and edible fungi. The sales volume of eggplant vegetables is the lowest among the six categories, indicating that eggplant vegetables have a smaller audience.

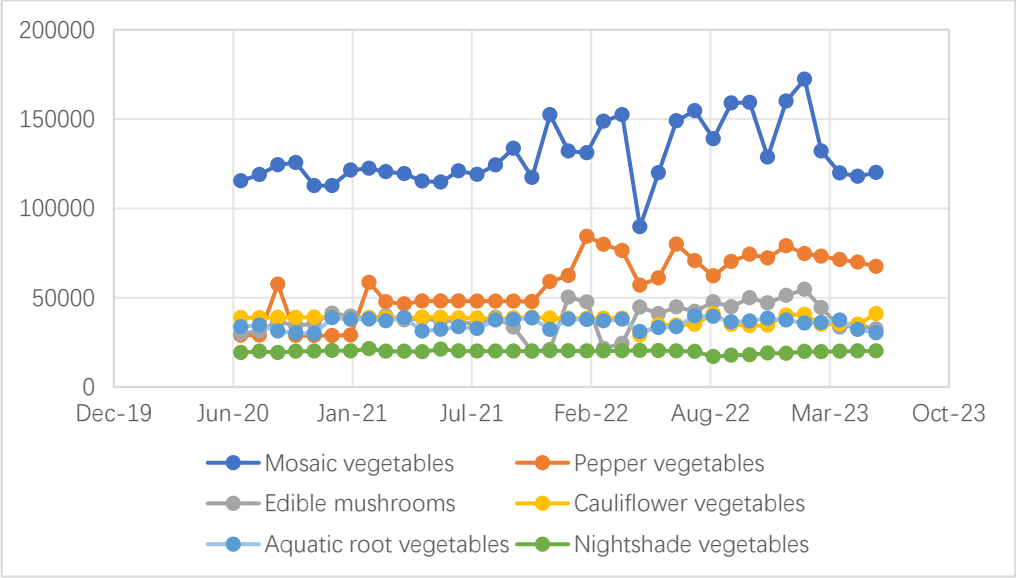

Figure 3: Line Chart of Category Sales Volume Changes

The figure 3 is a line graph of category sales change. Based on the data, it is observed that the sales volume of different vegetable categories show seasonal changes, for example, the sales volume of flower and leafy vegetables reaches the minimum in April 2022 and peaks in February 2023, which means that the demand for flower and leafy vegetables plummets in April 2022 and peaks in three years in February 2023, and the sales volume of eggplant, cauliflower, and aquatic root vegetables fluctuates more by season and shows seasonal changes. The demand reaches the peak in three years. The sales volume of leafy, chili and edible mushroom vegetables fluctuates a lot by the season, showing seasonal changes, and the sales volume of eggplant, cauliflower and aquatic root vegetables basically shows a stable trend throughout the year, indicating that eggplant, cauliflower and aquatic root vegetables are less affected by the season.

Table 1: Top 5 selling items (by month)

Months	Item Name	Sales volume (kg)
10	Lotus root(1)	4096.89
12	Brassica pekinensis	3711.21
1	Lotus root(1)	3484.30
12	Lotus root(1)	3256.35
3	Wuhu green pepper (1)	3206.50

As shown in the table 1, on the basis of the month as a unit, the top five sales of the single product in October of the net lotus root ranked first in sales, while in January and December of the net lotus root sales are in the top five, it can be said that the net lotus root in the aquatic root vegetables in the most popular, by the sales of Chinese cabbage can be learned that people in December of the demand for Chinese cabbage is higher in March of the demand for Wuhu peppers is at a high level.

Table 2: Top 5 selling items (by quarter)

Season (sports)	Item Name	Sales volume (kg)
4	Lotus root(1)	10354.45
1	Wuhu green pepper (1)	9026.79
3	broccoli	8217.89
1	Lotus root(1)	7973.45
4	Brassica pekinensis	7940.70

As shown in the table 2, on a quarterly time unit basis, the sales volume of net root under the fourth quarter was the largest among all the quarterly vegetable sales at 10,354.45 kg, which yielded the following order of popular vegetables in the sales volume calculated on an individual quarterly basis: net root > turnip peppers > broccoli > cabbage.

Table 3: Top 5 selling items (by year)

Vintages	Item Name	Sales volume (kg)
2021	Wuhu green pepper (1)	12280.69
2022	Wuhu green pepper (1)	11256.32
2022	broccoli	10294.53
2021	Lotus root(1)	9869.30
2022	Lotus root(1)	9513.22

As shown in the table 3, on the basis of year as a time unit, Wuhu green peppers had the largest sales in 2021-2022, which were 12,280.69 kilograms in 2021 and 11,256.32 kilograms in 2022, broccoli sales were ranked third in 2022, and net root sales were ranked fourth and fifth in 2021 and 2022, respectively. Analyzing the data on the basis of the whole year, Wuhu green pepper is the most popular vegetable of the year.

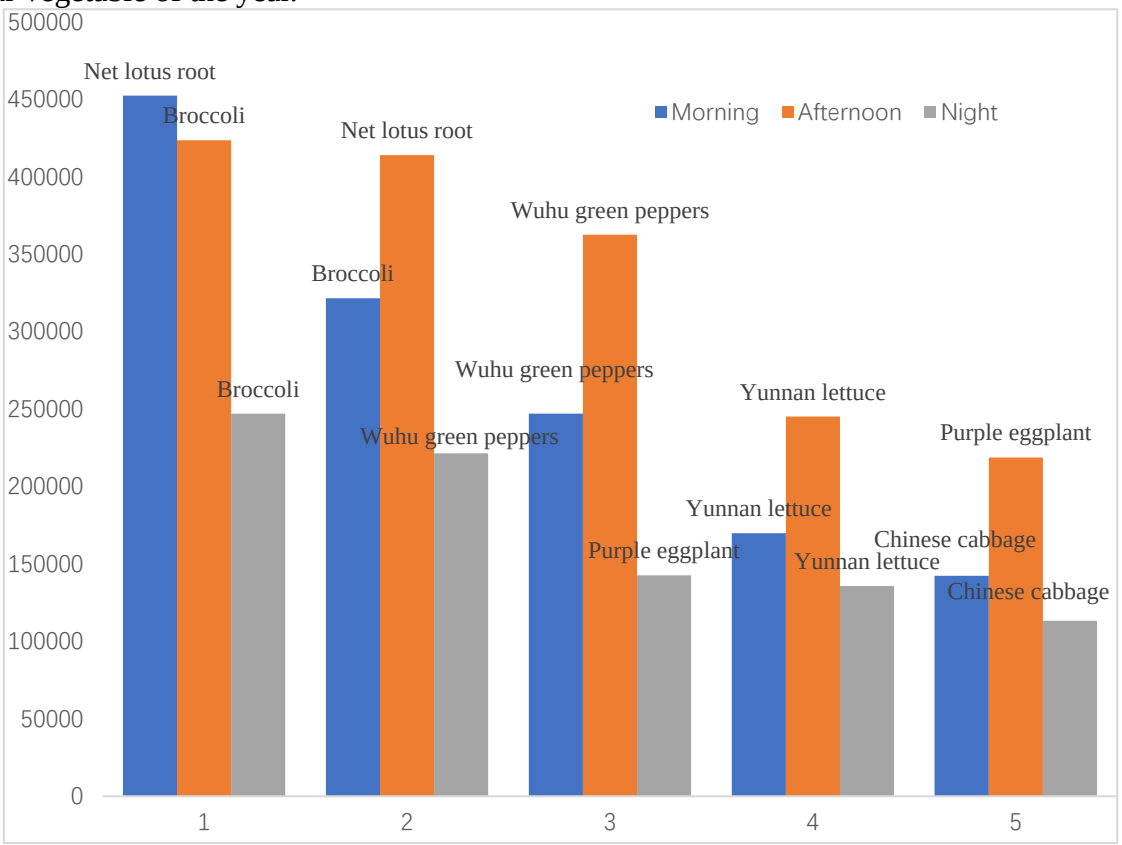

Figure 4: Top 5 selling items by time period

In order to explore the single product sales distribution pattern more carefully, this paper unifies the time data into morning, noon, and afternoon time slots.

As shown in Figure 4, in the morning session, the order of sales of single product is: net lotus root > broccoli > Wuhu green pepper > Yunnan lettuce > Chinese cabbage, in the afternoon session, the order of sales is: broccoli > net lotus root > Wuhu green pepper > Yunnan lettuce > purple eggplant, in the evening session, the order of sales is: broccoli > Wuhu green pepper > purple eggplant > Yunnan lettuce > Chinese cabbage. Combining the above graphs, the following conclusions can be drawn: net root sales are highest in the morning hours, indicating that net root is more popular with consumers in the morning, and broccoli is more popular with consumers in the afternoon and evening hours.

3. Analysis of Interrelationships in Vegetable Category and Product Sales

This part is to investigate the interrelationship between the sales volume of vegetable commodity categories and individual items, and will be solved using different models for vegetable commodity categories and individual items respectively.

For the analysis of vegetable commodity sales interrelationships, this paper uses the correlation analysis model based on the Apriori algorithm. Apriori algorithm is a classic data mining method for discovering frequent itemsets and association rules in transactional data[6]. Its original application is to mine associations between items in transactional data, most notably in the case of "diapers and beer". The Apriori algorithm scans the dataset to compute the support of the itemsets, and uses a support threshold to filter out frequent itemsets. It then generates candidate association rules and measures the strength of the rules by their confidence level.

Due to the large amount of data, the top 6 rules in terms of support ranking are displayed. The association rule is a relationship shaped like $A \rightarrow B$. A and B are called the antecedent and the consequent of the association rule, and the significance of its index is as follows:

1) Total support: the significance is the probability of A and B appearing at the same time. If the total support is greater, the relationship between A and B is considered greater.

2) Confidence level: the meaning is the probability of B appearing when A appears. If the confidence level is greater, the relationship between A and B is considered greater.

3) Improvement rate: the significance of {A->B} confidence / B support, the improvement rate reflects the use of rules relative to not using the rule, how much can be improved:

(1) If the enhancement rate is greater than 1, it means that the application of the association rule is positively influenced.

(2) If the lift rate is less than 1, the application of the association rule has a negative impact.

(3) When the enhancement rate is 1, A and B are considered independent of each other.

4) Leverage: the significance of {A->B} support - (A's support * B's support), used to measure the independence of A and B, that the leverage of 0 when A and B are independent, the greater the relationship between A and B the closer.

5) Belief rate: (meaning 1 - B's support) / (1 - {A -> B}'s confidence), the same for A and B's independence measure, the greater the relationship between A and B the closer.

As shown in the Table 4: the lift and leverage between broccoli and eggplant are both 1, so they are independent of each other and do not affect each other, and the total support between broccoli and net lotus root is 0.967 and the confidence level is 0.975, which are both close to 1. At the same time, the lift rate is also greater than 1, which can show that there is a strong positive relationship between broccoli and net lotus root, which suggests that, based on the other conditions certain, placing broccoli and net lotus root in the neighboring area can play a role in promoting sales. The total support between lotus root and purple eggplant is 0.927, the confidence level is 0.952 and the enhancement rate is more than 1, which indicates that there is a strong positive relationship between lotus root and purple

eggplant under certain other conditions, and it is recommended to place lotus root and purple eggplant in the adjacent area, in which the conviction rate of purple eggplant to lotus root is 1.648, which proves that purple eggplant and lotus root have a closer relationship.

Table 4: Association Rule Results Table

Antecedent (math.)	Eventualities	Degree of support for the preceding item	Degree of support for the latter	Total support	Confidence level (math.)	Promotion rate	Leverage	Conviction rate
Broccoli	Net lotus root	0.992	0.974	0.967	0.975	1.001	0.001	1.028
Net lotus root	Broccoli	0.974	0.992	0.967	0.992	1.001	0.001	1.096
Broccoli	Purple eggplant	0.992	0.942	0.935	0.942	1	0	1.008
Purple eggplant	Broccoli	0.942	0.992	0.935	0.992	1	0	1.06
Net lotus root	Purple eggplant	0.974	0.942	0.927	0.952	1.01	0.01	1.203
Purple eggplant	Net lotus root	0.942	0.974	0.927	0.984	1.01	0.01	1.648

4. Conclusion

This study provides valuable insights into sales volumes, distribution patterns, and the interrelationships among various vegetable categories and individual products through an in-depth analysis of vegetable sales data. The findings reveal that Wuhu green pepper is the top-selling single product, while leafy vegetables, chili peppers, and edible fungi dominate overall vegetable sales, accounting for 42%, 19%, and 16% respectively. Additionally, different vegetable categories show significant seasonal variations in sales volumes, with flower and leafy vegetables experiencing the lowest demand in April 2022 and peaking in February 2023.

Using Pearson correlation coefficients and the Apriori algorithm for association rule analysis, the study identified complementary and substitute relationships among certain products. For instance, red bell pepper and Xixia flower mushrooms are complementary, while Yunnan oilseed rape and Yunnan lettuce are substitutes. These results suggest that strategic product placement, such as positioning complementary products next to each other, can enhance sales. Overall, this study helps businesses understand consumer preferences, optimize inventory management, and improve sales strategies.

References

[1] BORSELLINO V, SCHIMMENTI E, EL BILALI H. Agri-food markets towards sustainable patterns[J]. Sustainability, 2020, 12(6): 2193.

[2] GU H, WANG C. Impacts of the COVID-19 pandemic on vegetable production and countermeasures from an agricultural insurance perspective[J]. Journal of Integrative Agriculture, 2020,19(12): 2866-2876.

[3] YU W, WONG C Y, CHAVEZ R, et al. Integrating big data analytics into supply chain finance: The roles of information processing and data-driven culture[J]. International journal of production economics, 2021,236: 108135.

[4] BECHTSIS D, TSOLAKIS N, IAKOVOU E, et al. Data-driven secure, resilient and sustainable supply chains: gaps, opportunities, and a new generalised data sharing and data monetisation framework[J]. International Journal of Production Research, 2022,60(14): 4397-4417.

[5] LIU S, LIU D, SRIVASTAVA G, et al. Overview and methods of correlation filter algorithms in object tracking[J]. Complex & Intelligent Systems, 2021,7: 1895-1917.

[6] XIE H. Research and case analysis of apriori algorithm based on mining frequent item-sets[J]. Open Journal of Social Sciences, 2021,9(04): 458.