Vegetable Merchandise Analysis from a Quarterly Sales Data Perspective

Yuxin Sang [1, *], Peiyan Li [2], Ziliang Yin [1]

[1] College of Software, Henan Normal University, Xinxiang, China, 453007

[2] College of Life science, Xinyang Normal University, Xinyang, China, 464000

* Corresponding Author Email: 18439007839@163.com

Abstract. In the field of fresh vegetable retailing, the category diversity and time-sensitive characteristics of commodities pose unique challenges to inventory management and pricing decisions. Comprehensively analyzing the distribution pattern of sales volume of different categories and individual items of vegetable commodities and their interrelationships is of great significance to retailers' scientific replenishment and pricing decisions. By categorizing vegetable sales every quarter, a comprehensive analysis is carried out on a category and individual item basis. Visualization was used to show the distribution pattern of sales volume by category and individual item. Subsequently, Kendall correlation analysis was utilized to explore the relationship between the sales volume of different vegetable categories. In addition, the interrelationships between the sales volumes of different vegetable individual items were investigated in conjunction with waterfall diagrams. Through in-depth analysis of sales data, understanding the sales relationship between different vegetable categories can help retailers develop a more refined pricing strategy, make pricing adjustments based on the relative sales of goods, and improve sales revenue and profit levels. Such scientific inventory management and pricing strategies not only help meet customer demand and enhance customer experience, but also strengthen customer loyalty and promote long-term stable business development.

Keywords: Kendall correlation analysis, normality tests, visualizing sales changes, big data.

1. Introduction

In the current retail market, fresh vegetables occupy an important position with their richness in category and timeliness. As consumers pay more attention to healthy diets, the sales of fresh vegetables have become the focus of retailers' superstores [1]. However, due to the perishability and seasonal fluctuations of vegetable commodities, it has become a challenge to effectively manage inventory and pricing to maximize sales revenue and reduce wastage [2]. This study aims to provide a basis for retailers' scientific replenishment and pricing decisions by deeply analyzing the sales volume distribution patterns of different categories and individual items of vegetable commodities and their interrelationships [3].

Kendall correlation analysis was mainly used in this study. This method was introduced to explore the correlation between the sales of different categories of vegetables, which was proposed by Maurice Kendall, a British statistician, and its analytical coefficients are mainly used to measure the linear correlation between random variables, which is widely used in the fields of sales forecasting, medical diagnostic analysis, as well as stock market analysis and forecasting, etc [4] [5] [6]. In addition, since it can deal with non-linear relationships, and categorical data, as well as considering multiple factors and quantifying correlations at the same time, which in turn can help decision makers to adjust their strategies in real-time, Kendall correlation analysis is therefore feasible to study the sales distribution pattern of vegetables [7].

To study the distribution law and the interrelationship between the sales volume of different categories and different single products of vegetable commodities, we first use Kendall correlation analysis to explore the interrelationship between the sales volume of different categories of vegetables based on the data characteristics of the sales volume of each category of vegetables divided by quarter, and then calculate the sales volume of vegetable commodities based on the division of the category

and the single product as a unit, and draw line graphs and waterfall charts for visual analysis to arrive at the distribution law of the sales volume of different categories and different single products [8][9].

2. The Sales distribution patterns by vegetable category

2.1. Distribution of sales of six types of vegetables over the quarter

Data for this article was obtained from http://www.mcm.edu.cn/,with the dataset of sales volume of various vegetables from the third quarter of 2020 to the first quarter of 2023 provided by the official website, we can get the distribution of sales volume of six categories of vegetables, namely aquatic roots and tubers, flowers and leaves, cauliflower, eggplants, peppers, and edible mushrooms, as shown in Figure. 1. Where 2020-Q3 represents the third quarter of 2020, and the rest by analogy.

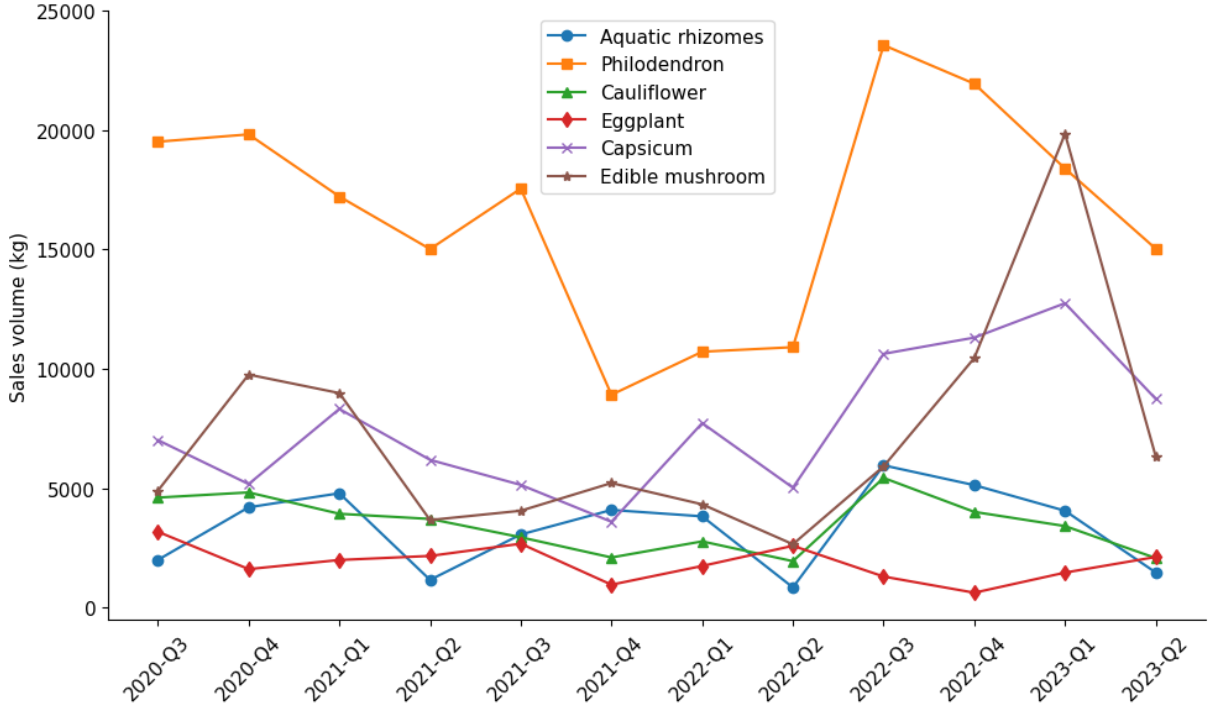

Figure 1. Distribution of sales of six types of vegetables.

The distribution pattern of sales of the six categories of vegetables in the above chart shows that the overall sales of aquatic roots and tubers, cauliflowers, and eggplants are low and fluctuate very little with the seasons. Leafy flowers, chili peppers, and edible fungi sales were generally medium to high but with seasonal fluctuations.

2.2. Selection of data for normality test and Kendall correlation analysis

Only from the distribution of sales of these six types of vegetables get the relationship between the sales of various types of vegetables too one-sided, want to further explore the relationship between the sales of these six types of vegetables, first we need to explore the sales of various types of vegetables in each quarter from these 12 quarters whether there is some kind of pattern. Therefore, we conduct a normality test on the sales of these six types of vegetables through QQ plots [10] [11]. Some of the test results are shown in Figure 2:

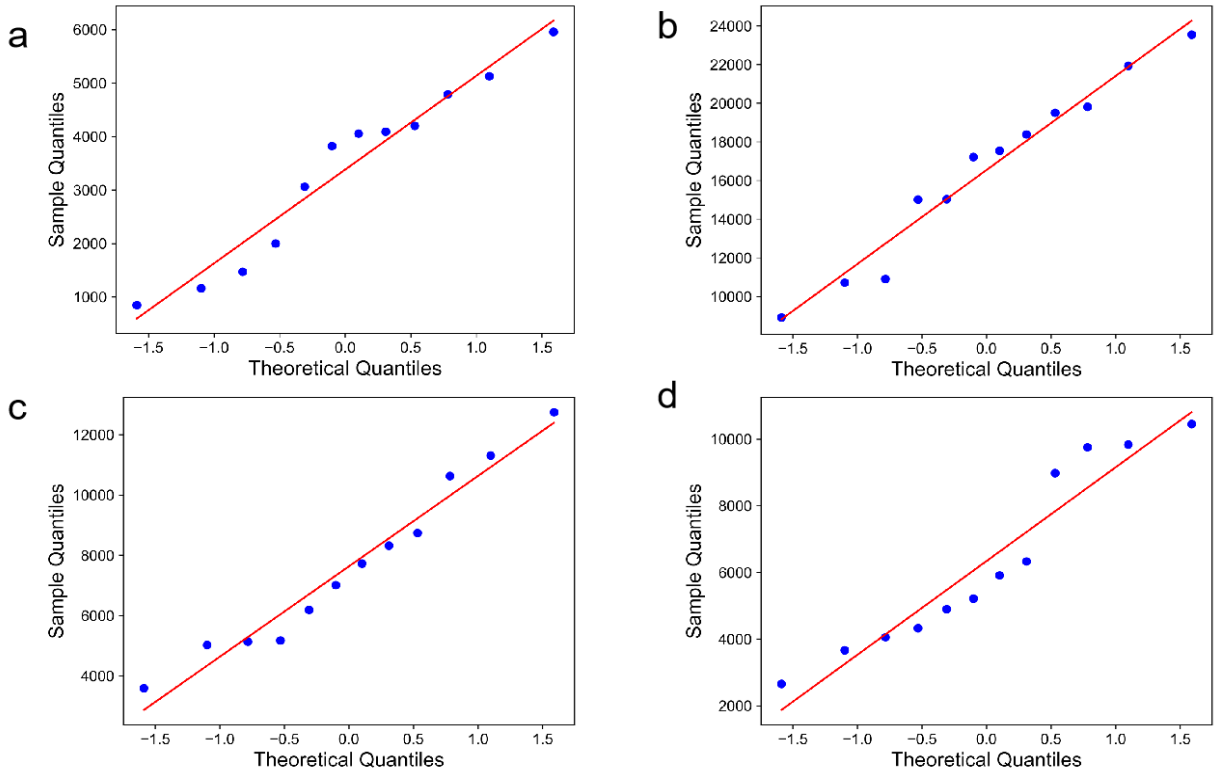

Figure 2. Plot of normality test for some dishes.

The resulting QQ plot shows that the sales of these six types of vegetables do not conform to a normal distribution between them, and the choice to use Kendall's Tau and Spearman correlation analysis rather than Pearson correlation coefficients is usually a more appropriate choice when the dataset does not conform to a normal distribution. This is because the Pearson correlation coefficient is more sensitive to skewed or non-normal distributions of the data, which may lead to inaccurate results [12][13].

Kendall's tau is a non-parametric statistic that does not depend on the distribution of the data and is therefore more robust and able to provide a reliable measure of correlation when the data does not conform to a normal distribution. At the same time, Kendall's tau may be preferred when dealing with small sample datasets, when higher computational efficiency is required, or when the consistency and inconsistency of data pairs are particularly important. In addition, Kendall's tau coefficient is slightly more resistant to outliers, and Spearman's correlation coefficient is slightly less resistant to outliers, so Kendall's tau may provide a more robust estimate of the correlation in some situations [14]. Next, we used Kendall's correlation analysis to explore the potential relationship that exists between the six categories of vegetables and reflected this relationship by plotting a heat map between the six categories, as shown in Figure 3.

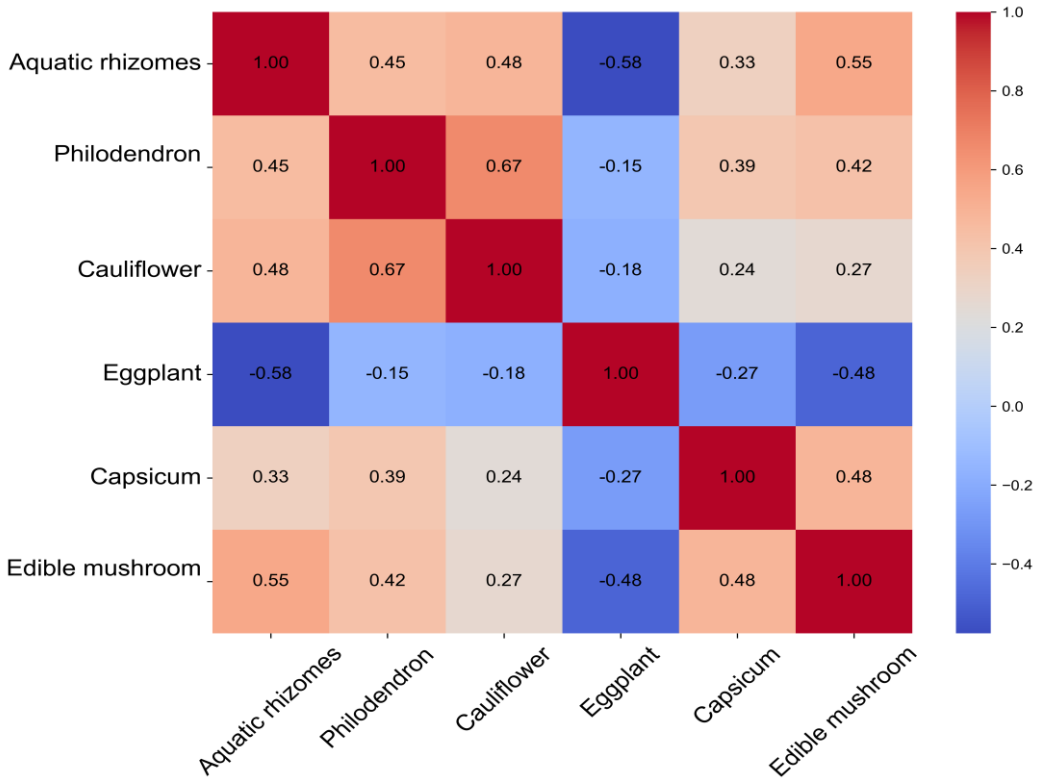

Figure 3. Heat distribution of Kendall correlation coefficients for six types of vegetables.

From the correlation coefficients in the above graph, it can be seen that there is a significant positive correlation between sales of foliage and cauliflower that eggplant is negatively correlated with all the other five categories, and that this negative correlation is most significant with aquatic rhizomes. Meanwhile, excluding eggplant, aquatic rootstocks are positively correlated with the remaining four categories, and this positive correlation is most significant with edible mushrooms.

3. Results and discussion

3.1. Sales distribution trends for individual vegetables

The studied dataset is grouped according to "item name" and the sales volume of each group is summed up to find out the sales volume of each item in that quarter of the year, and a waterfall diagram is drawn to visualize the distribution pattern of the sales volume of each item of vegetables in each quarter from July 1, 2020, to June 30, 2023.

Looking at the 12 waterfall charts below shows that the data points are presented in a continuous ladder with key points and outliers highlighted, which leads to the top-selling vegetable items for each quarter:

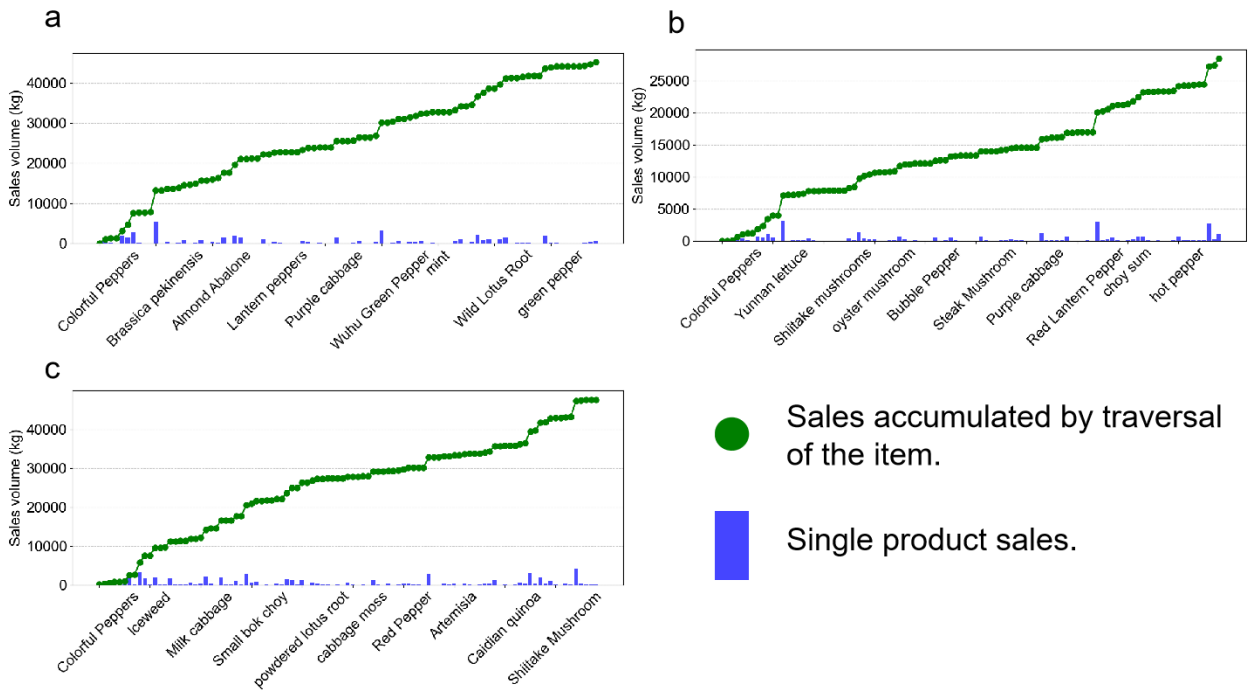

Figure 4. Q1 Vegetable Sales Volume Waterfall by Individual Product. a: First quarter of 2021; b: First quarter of 2022; c: First quarter of 2023.

As shown in Figure 4, net root and broccoli in the first quarter of 2021, enoki mushrooms, radish chilli and purple aubergine in the first quarter of 2022 and Yunnan lettuce and millet peppers in the first quarter of 2023 were the best sellers.

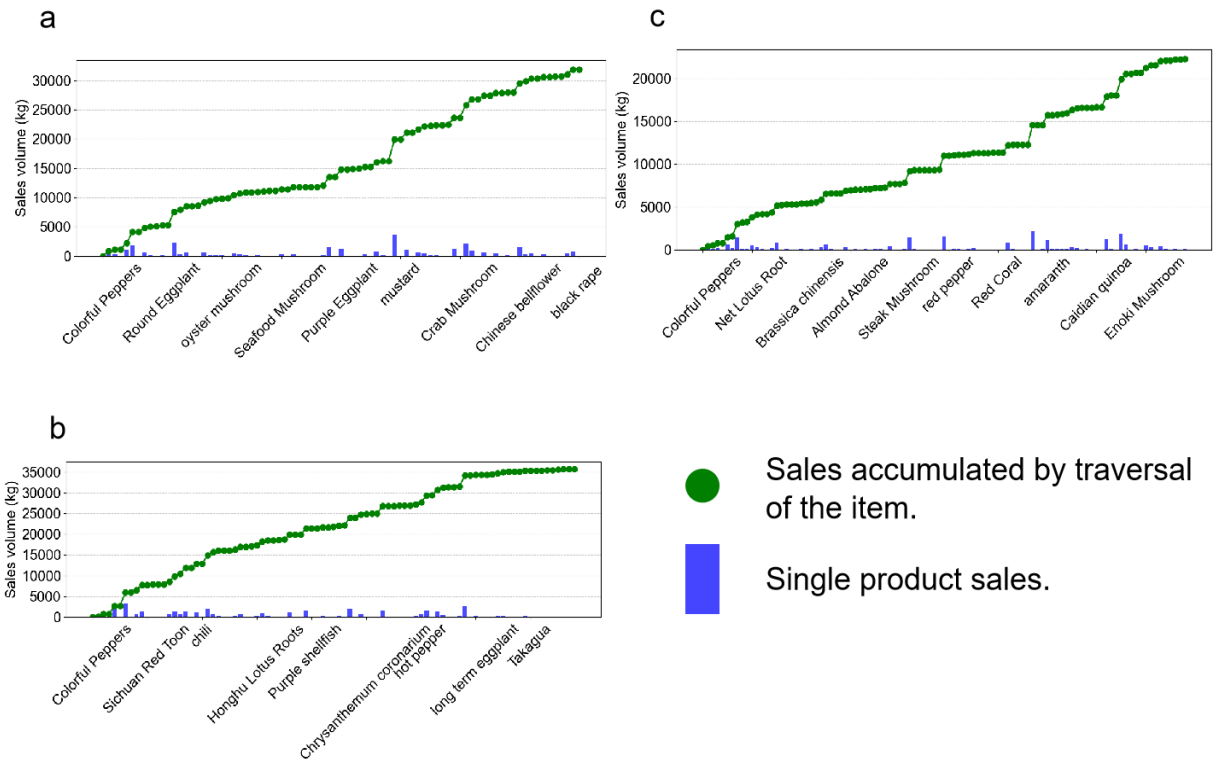

Figure 5. Q2 Vegetable Sales Volume Waterfall by Individual Product. a: Second quarter of 2021; b: Second quarter of 2022; c: Second quarter of 2023.

As shown in Figure 5, radish chilli and Yunnan lettuce in Q2 2021, bamboo leaf lettuce and screw pepper in Q2 2022 and amaranth and spinach in Q2 2023 were the best sellers.

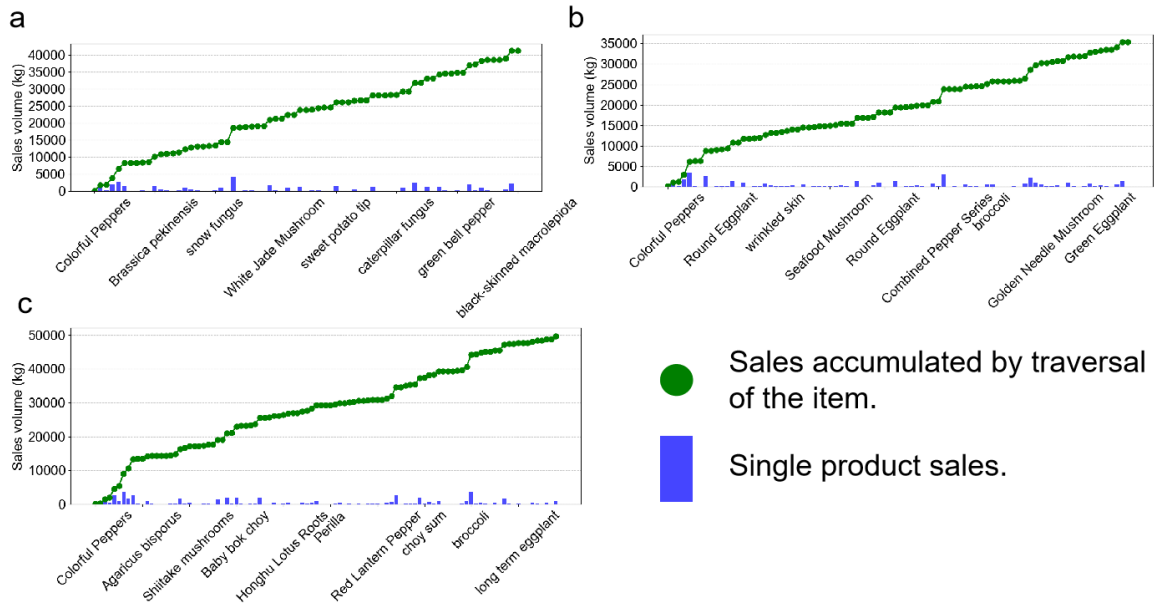

Figure 6. Q3 Vegetable Sales Volume Waterfall by Individual Product.
a: Third quarter of 2020; b: Third quarter of 2021; c: Third quarter of 2022.

As shown in Figure 6, broccoli and yellow kale in the third quarter of 2020, green stemmed loose flowers and Yunnan oilseed rape in the third quarter of 2021 and sweet kale and Chinese cabbage, rosettes in the third quarter of 2022 are the best sellers.

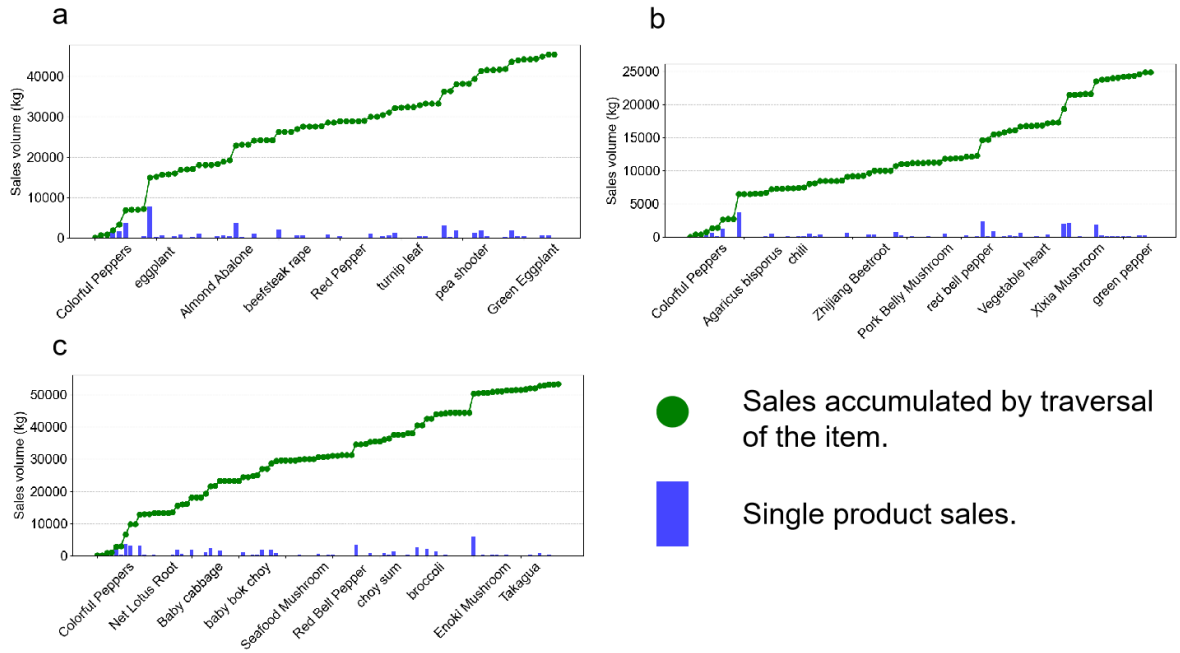

Figure 7. Q4 Vegetable Sales Volume Waterfall by Individual Product.
a: Fourth quarter of 2020; b: Fourth quarter of 2021; c: Fourth quarter of 2022.

As can be seen from Figure 7, Niushou oilseed rape, green stalks and loose flowers in the fourth quarter of 20220, golden needle mushrooms, Xixia shiitake mushrooms and coriander in the fourth quarter of 2021 and Baokang alpine Chinese cabbage in the fourth quarter of 2022 were the best sellers.

3.2. Analysis of experimental results

According to the waterfall diagram and the above conclusion, the replenishment quantity of Wuhu green pepper and Yunnan lettuce can be increased in the first and second quarters, the replenishment

quantity of broccoli can be increased in the first and third quarters, and the replenishment quantity of green stems and loose flowers can be increased in the third and fourth quarters; at the same time, vegetables with high sales volume can be bundled with those with low sales volume in each quarter, such as broccoli and bamboo leaf lettuce bundled with each other in the first quarter and broccoli and Niushou oilseed cabbage bundled with each other in the third quarter, to promote the sales volume of every single product of vegetables and to prevent the phenomenon of stagnation of sales.

4. Conclusions

By dividing the sales volume of vegetable categories into seasons, we find that different categories show different sales patterns under seasonal changes. The sales volume of aquatic root and cauliflower categories showed stable performance, while the sales volume of flower and leaf categories fluctuated more. Kendall's correlation analysis revealed the correlation between the sales volume of different categories of vegetables, and there were obvious positive or negative correlations between some categories, which provided a reference for supermarkets to formulate purchasing strategies. In addition, the waterfall diagram shows the sales of individual vegetable items in different seasons, which guides supermarkets to optimize their product mix. It is recommended that supermarkets focus on seasonal variations and correlation between categories when purchasing vegetables. It is recommended to increase purchasing efforts in categories with higher sales volume, and at the same time, rationalize the product mix according to the seasonal sales characteristics of individual products to enhance sales efficiency and profitability.

The Kendall correlation analysis method applied in this paper has demonstrated its utility and breadth in numerous fields, especially in critical areas such as sales forecasting and medical diagnostic analysis. The method is favored because of its unique ability to handle nonlinear relationships, deal with categorical data, and quantify correlations by considering multiple factors simultaneously. These characteristics make Kendall Correlation Analysis a powerful tool in the hands of decision-makers, helping them adjust their strategies based on real-time data. In particular, Kendall Correlation Analysis demonstrates its unique research feasibility and usefulness when studying the distribution pattern of vegetable sales.

References

[1] Zhang Hao, Dong Chunyu, Yang Haichan, Sun Sijing, Han Yu, Huang Zuzhi... Bao Li. (2024). Evaluation and correlation analysis of heavy metal pollution in farmland soil and vegetables in Zhaotong City. Environmental Science (02), 1090-1097. doi: 10.13227/j.hjkx.202303178.

[2] Feng Bihong, Gu Yang. Research on the influence of weather factors on vegetable prices in production areas [J]. Agricultural Development and Equipment, 2023(1):89-90.

[3] Han Zhe, Bi Hongwen, Zheng Yanyan, Li Yang, Liu Kebao, Lv Zhiqun, Wu Guorui.Analysis of vegetable price trend and industrial development countermeasures in Heilongjiang Province in 2021[J]. Modern Agricultural Science and Technology, 2023(7):210-214.

[4] Ji Peng, Chen Fangfang, Xu Tianqi & Qi Qi. (2022). Fault routing based on Kendall correlation coefficient and CEEMD decomposition of zero sequence current phase method. Journal of Yunnan University for Nationalities (Natural Science Edition) (05), 595-601.

[5] Yu Yi, Li Zixiang, Mai Xiaoming, Xu Yuming. (2020). Kendall correlation analysis of physico-mechanical indexes of soft soil in Meizhou river bank phase. Journal of Guangzhou Nautical College (02), 65-69.

[6] Zhou Hui, Wei Xiaojun. (2006). Nonparametric test of financial development based on Kendall's correlation coefficient. Financial Economics (20), 142-143.

[7] Ying Dongqin. Selection and optimization countermeasures of farmers' vegetable marketing methods [J]. Southern Agricultural Machinery, 2019, 50(24):62-62.

[8] Li Weiwei. (2015). Exploring the correlation and fluctuation characteristics of different kinds of vegetable prices--Take Beijing as an example. Price Monthly (12), 36-41. doi: 10.14076/j.issn.1006-2025.2015.12.08.

[9] Ma Zhiqin,Wang Changqin. (2022). Diagnostic value of electrocardiographic waterfall map combined with electrocardiographic scattergram in the rapid determination of atrioventricular relationship. Chinese Journal of Cardiac Pacing and Electrophysiology (05), 462-465. doi: 10.13333/j.cnki.cjcpe.2022.05.015.

[10] Li, Yumei. (2015). Method of normality test for data. Journal of Huaihua College (11), 81-82. doi: 10.16074/j.cnki.cn43-1394/z.2015.11.020.

[11] Liu Jindong, Wang Tiancheng, Chen Yonghong, Zhang Yu. Improved method and application based on Anderson-Darling normality test [J]. Practice and Understanding of Mathematics, 2023, 53(4):184-193.

[12] Hu Tianyi, Yang Guang, Chen Bo, Yu Yang, Tao Yuan. (2016). ARIMA slope displacement monitoring model based on Pearson correlation test. Water Conservancy and Hydropower Technology (01), 71-75. doi: 10.13928/j.cnki.wrahe.2016.01.016.

[13] Zhao Zhigang, Li Ning, Yifeng. Research on budget reform of county and district departments in Liaoning province--analysis based on Person correlation model [J]. Local Finance Research, 2013(10):47-5259.

[14] He Shuang. Analysis of factors affecting institutional evaluation results based on Spearman rank correlation and Topsis-fuzzy comprehensive judgment algorithm model [J]. Journal of Changchun Normal University, 2022, 41(10):9-16.