Information Systems and Economics (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2024.050318
ISSN 2523-6407 Vol. 5 Num. 3

# *Research on replenishment strategy of superstore fresh goods based on STL-ARIMA*

**Yujing Zhang[1,*], Hongling Qin[2], Zhihan Wei[1], Yu Dai[3], Linwen Zhang[4]**

*[1]School of Mathematics and Computer Science, Chongqing College of International Business and Economics, Chongqing, 401520, China*
*[2]Chongqing Science City Xi Yong First Primary School, Chongqing, 400030, China*
*[3]School of Bioinformatics, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*
*[4]Faculty of Electronics and Information, Chengdu Jincheng College, Chengdu, 611731, China*
*[*]Corresponding author: 18723107388@163.com*

***Keywords:*** ARIMA Model, Correlation, Replenishment Strategy

***Abstract:*** This article provides a comprehensive analysis and modelling of vegetable sales data. Firstly, the statistical analysis of the data was carried out by counting the monthly sales by category and individual item respectively and plotting the combination of sales volume and sales for each category of vegetables as well as the circle plot for individual items of vegetables so as to study the relationship and trend between them. The correlation coefficients between different vegetable categories and individual items were further calculated by Pearson's test, and strong correlations between edibles and aquatic roots and tubers, as well as correlation characteristics between chilli, cauliflower and foliar categories were found. In modelling cost-plus pricing, the article predicts the total replenishment and pricing strategy for the coming week by analysing the relationship between sales volume and cost-plus pricing and obtaining a fitted equation. For data processing, the article adopts the method of nearest neighbour interpolation to fill in missing values and data smoothing, decomposes the data into seasonal, trend and residual terms by STL decomposition, and establishes an ARIMA model to predict the sales volume in the coming week. Overall, this paper digs deeper into the relationships and trends in vegetable sales data through statistical analysis and modelling methods, which provides strong support and reference for the development of future sales strategies.

## 1. Introduction

With the development of social economy and the improvement of people's living standard, the demand for vegetables and their sales management is increasing. As one of the essential foods in people's daily life, vegetables have a wide range of sales and applications in the market. However, the sales of vegetables are cyclical, seasonal and corrosive in nature, requiring targeted sales strategies and management models to meet consumer demand and maximise profits[1].

Aiming at the practical problems in vegetable sales management, this study aims to analyse the sales patterns and interrelationships of different types of vegetables[2-3], explore the relationship between sales volume and cost-plus pricing, and forecast sales volume based on a time-series model, so as to provide a scientific basis and decision-making support for the supermarket's vegetable purchasing, price-setting and sales strategies[4].

First, we statistically analyse the sales data of different types of vegetables, including the distribution pattern of sales volume and the correlation analysis between different categories, in order to reveal the overall trend and characteristics of vegetable sales. Then, we use the cost-plus pricing theory to establish the relationship model between sales volume and cost-plus pricing, and explore the influence of price on sales volume through linear regression analysis. Finally, we forecast the sales volume of different types of vegetables based on the STL-ARIMA model, taking into account the cyclicality and seasonality of the data, to provide scientific and accurate sales volume forecasting results, and to provide effective reference and decision-making support for vegetable sales management[5-6].

Through the above modelling ideas, we aim to gain a comprehensive understanding of the characteristics and laws of vegetable sales, to provide a scientific basis for the sales strategy and management of superstores, and to achieve the optimization of vegetable sales and profit maximization[7].

## 2. Correlation modelling and solution

### 2.1. Monthly sales distribution pattern of different types of vegetables

Vegetables are classified into six categories according to their organ characteristics as well as edible characteristics: cauliflower, leafy flower, chilli, eggplant, edible fungi and aquatic roots and tubers. Considering the planting cycle of vegetables as well as the seasonal distribution, the statistical unit can be year, season, month, week, etc. However, for the sake of science, the sales from 1 July 2020 to 30 June 2023 are divided into research intervals by whole month, and the sales volume and total sales of different categories are counted every month, and the statistical results are shown in Figure 1 (limited by space, the results are only shown for one kind of commodity):

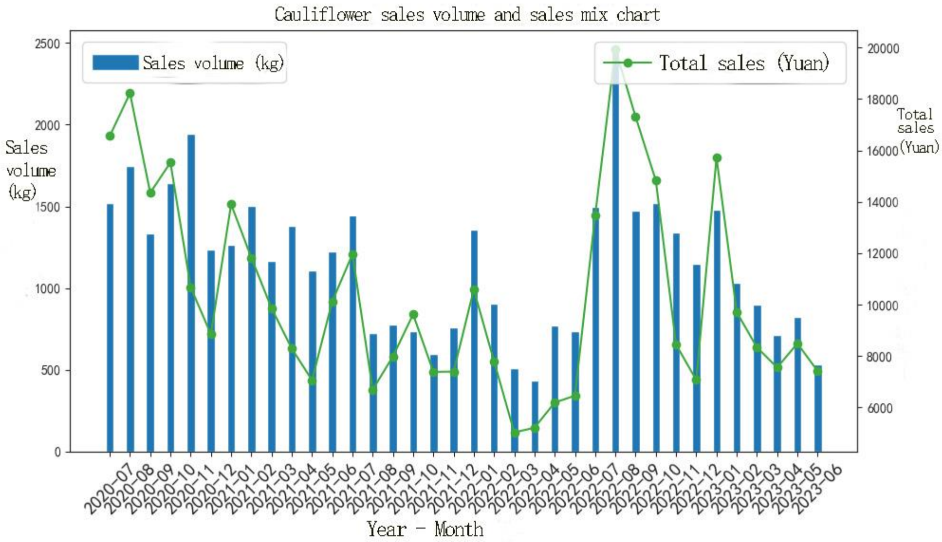


Figure 1: Sales and sales mix

As can be seen from Figure 1, according to the fact that each category of vegetables has different sales volume under different years and seasons[8], the distribution pattern of the category can be derived. The conclusions of the specific analyses for each category are as follows:

20 years and 22 years of July-October sales volume is greater than 1500 kg, indicating that the commodity in the end of the summer in the sales volume in general is better, while the total sales volume is higher, suggesting that the flower and fruit category was selling at a high and hot price during that period.

Flowers and foliage sales are basically greater than 4,000 kg, but 21 years in October -22 years in May sales are all lower than 4,000 kg, indicating that the same with the cauliflower category by the impact of the epidemic is not good for the storage of such commodities resulting in low sales and low prices; in the case of a small difference between the sales volume, each year, the sales volume of the July-September is higher than the other months, indicating that the same type of commodities with cauliflower category at the end of summer The sales volume is generally better.

Sales of chillies in the chilli category were generally below 3,000kg for two years, and were generally higher each month after July 22, but total sales were decreasing, suggesting that there was a substantial price reduction in the unit price of the commodity after July. Chilli sales are highest each year around December-February, suggesting that chilli is preferred in winter.

Eggplant sales in October - November each year is less than 400 kg, in April - July is higher than 800 kg, indicating that like to eat eggplant in the summer and do not like to eat in the autumn. Aubergines in December each year - the following February compared with other sales volume is not a big difference between the total sales of the month, indicating that the price of aubergines in winter is expensive [9-10].

Aquatic roots and tubers commodity sales every year around April is less than 500 kg, around December is higher than 1700 kg, that like to eat in the winter water roots and tubers and do not like to eat in the spring. The monthly sales volume of this category and the total amount of enjoyment are relatively growing, thus indicating that the price fluctuation is not significant.

The sales volume of edible mushrooms is lower every year around April, except for the year 23, all the other years are below 1200 kg, and higher around September-December, indicating that mushrooms are preferred in the spring and disliked in the winter, and that the price of this category does not fluctuate much.

Although the analysis of individual categories can meticulously analyse the distribution pattern of vegetables, but there is a lack of contrast and correlation between the categories, this paper will be 6 categories of sales changes in graphical processing, the results are as follows Figure 2:

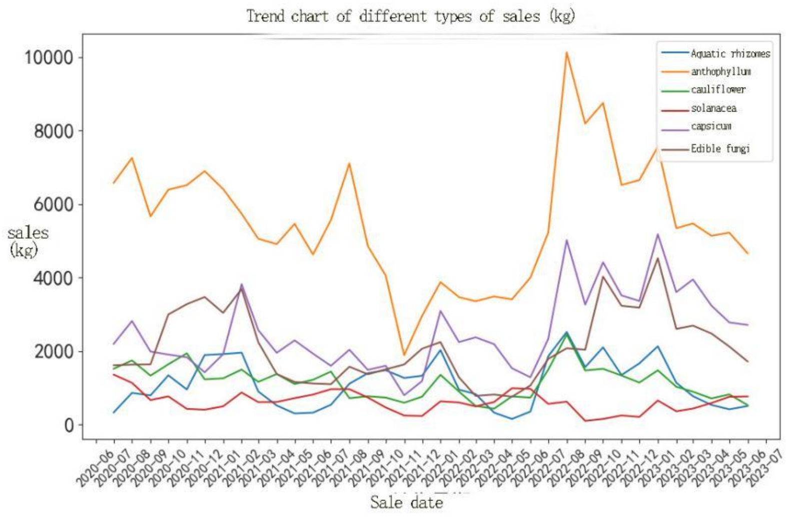


Figure 2: Trends in sales volume by category

Overall, the distribution of the 6 categories of vegetables has a high degree of similarity, and the sales in October 2021-May 2022 are not high.

After July 22 there is a significant increase in the sales of all categories of vegetables except eggplant and the demand for all categories of vegetables rises in December every year. Comparison of sales between them: foliage > chilli > edible mushrooms > aquatic roots and tubers > tomatoes, people buy foliage class demand most of the time much higher than the other 5 categories, vegetable super in the purchase of goods in the consideration of foliage class low timeliness of the case of a reasonable amount of reserves.

### 2.2. Correlation analysis between different categories

When people buy vegetables, they tend to buy not only one category, but more often a combination of multiple categories, so there is a certain correlation between the categories. Usually, the calculation of correlation coefficient mainly includes three kinds of correlation coefficients: Pearson, Spearman, and Kendall's correlation coefficient. Combined with the characteristics of the data of vegetable categories extracted in this paper and the objective of consideration, Pearson is selected for correlation analysis.

Pearson correlation test will be carried out for 6 types of vegetable species, and the test results are shown in Figure 3:

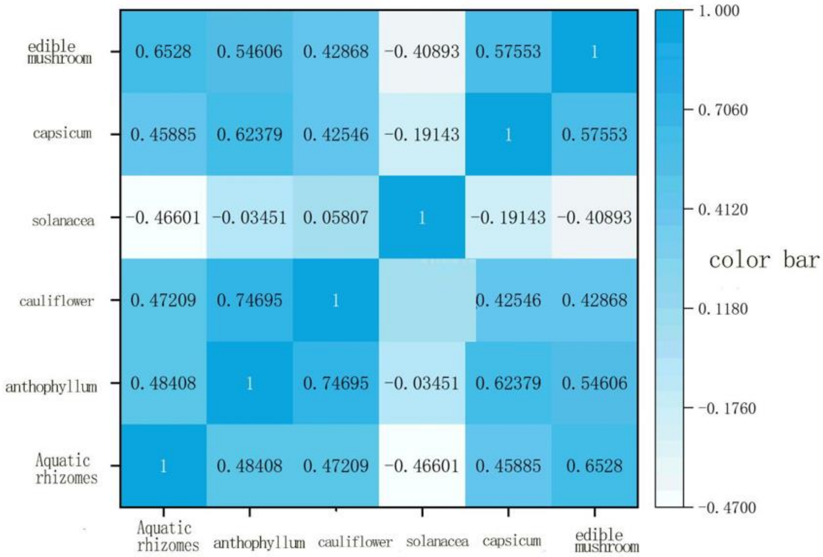


Figure 3: Sales correlation

From the test results, the correlation coefficient between edible mushrooms and aquatic rhizomes 0.6<0.6528<0.8, the correlation is strong correlation; the correlation coefficient between foliage and chilli 0.6<0.62379<0.8, the correlation is strong; the correlation coefficient between foliage and cauliflower tested to be 0.74695, the correlation is strong. The correlation coefficients of aquatic rhizomes with phloem, cauliflower, eggplant and chilli were 0.4-0.6, moderately correlated; the correlation coefficients of phloem and chilli and edible fungus were 0.6-0.4, moderately correlated; the correlation coefficients of chilli and edible fungus were .4<0.57553<0.6, moderately correlated, and the rest of the correlations were weak and non-correlated.

### 2.3. Monthly Sales Distribution Pattern of Different Individual Products

In order to simplify the research data, this paper will be different origins and different packaging methods of the same single product after the merger of processing, to obtain a total of 152 varieties of vegetables.

Among them, the cumulative share of less varieties reached more than 40%, due to the sample varieties, the remaining sales of each variety accounted for a very small value, the research significance is not very large, the merger of its processing as ‘other’, each quarter to select the first 8 categories of large distinction for analysis. As can be seen from the figure, no matter in which quarter, Wuhu green pepper, enoki mushroom, broccoli, Yunnan oleaginous vegetables, these four single products are chosen by people with high frequency, which indicates that vegetable superstores should make reasonable preparations when purchasing goods. For the four quarters of the proportion of large single product extracted 13, the distribution pattern of its comparative analysis, August 2020 - March 2021, cabbage sales are significantly higher than other categories, and then reduced; in June 2022, after most of the single product sales have risen, the sales are roughly the same trend (figure 4).

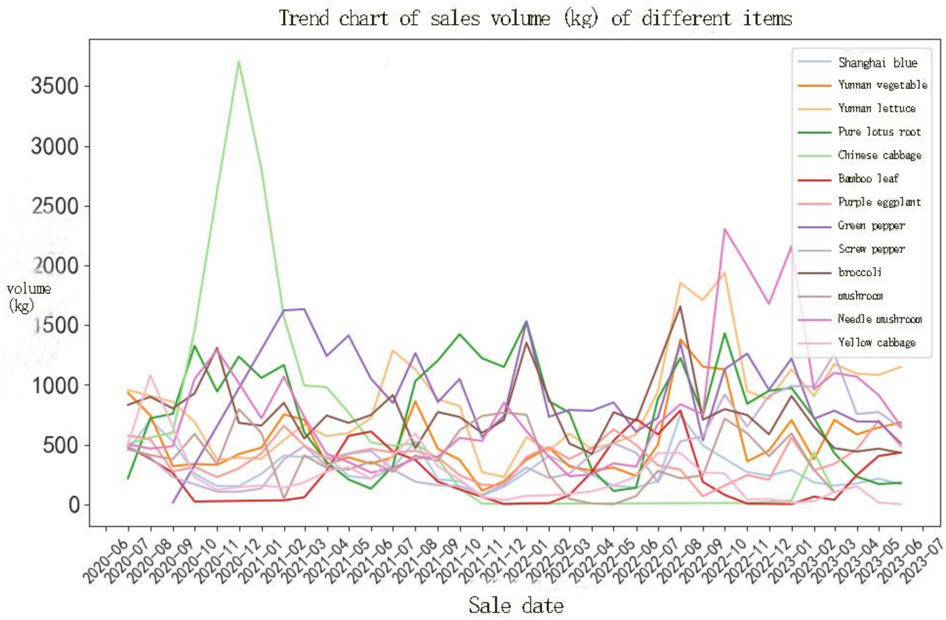


Figure 4: Trends

## 2.4. Correlation analysis between different categories

23% of the different single products are related to each other, and 1% of the single products are related to each other up to more than 0.9 or even 1. In this paper, we select some of the extremely strong correlation between the single products to show and analyse:

A correlation coefficient of 1 indicates that purchasing both varieties together is absolute; in the case of line 1, for example, if a customer purchases red oak leaf then he or she will definitely purchase red coral, and if he or she purchases red coral he or she will also definitely purchase red oak leaf. The correlation coefficient is less than 1, indicating that the correlation between single product 1 and single product 2 purchased together is not absolute, but as the correlation coefficient increases, indicating that the relationship closeness increases, the possibility of purchasing together is high.

## 3. Replenishment modelling and solving

Vegetables belong to the typical perishable products, with cyclical, seasonal, corrosive characteristics, with the passage of time, customer purchases will be affected, so to develop some strategies to sell the product, but also timely replenishment to maximise the benefits of the problem based on the development of the replenishment strategy based on the category as a unit.

### 3.1. Relationship between total sales of vegetable category and cost-plus pricing

Cost-plus pricing theory is a cost-based pricing theory. The cost of a vegetable product from purchase to sale is the basis of the price, and the price of the product also depends on the price mark-up added to the cost.

$$x = C \times (1 + r_i) \tag{1}$$

where x is the unit price of the commodity, C is the total cost of the commodity, and ri is the commodity markup rate.

$$r_i = 1 + \frac{m_i + n_i - p_i}{p_i} \tag{2}$$

According to the cost-plus pricing formula, the cost-plus pricing of each type of vegetables is solved, in order to explore the relationship between the total sales of vegetable categories and cost-plus pricing, this paper adopts linear regression to fit a function of each category and the total sales to portray the relationship between them (Figure 5).

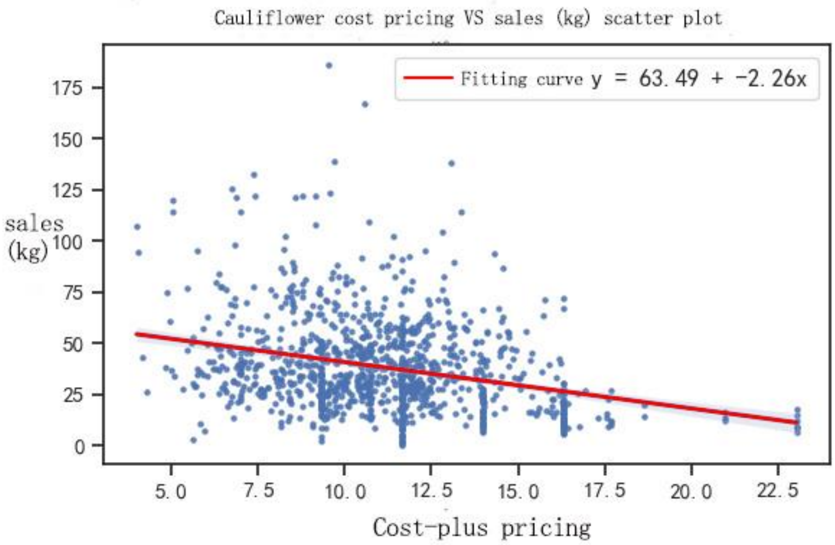


Figure 5: Fitting results

From the fitting results, it can be seen that the cost-plus pricing of all vegetable categories and the total volume of sales can be fitted into a linear equation, with a strong negative linear correlation. It can be obtained that when the price of vegetables is high, the sales volume decreases; when the price is low, the sales volume increases. At present, this law is of great practical significance to the price control and auxiliary operation of supermarkets.

### 3.2. STL-ARIMA-based sales forecast for different categories

The vegetable sales have a certain degree of periodicity, so the STL is selected to decompose the sales volume of each type of vegetables, which can obtain a better decomposition effect for the data with periodicity. The time series number YV at a certain moment is decomposed into three components, of which, the trend component TV represents the trend and direction of the data; the cycle component SV represents the pattern of change that repeats over time in the data; the residual EV is the remaining component of the original sequence after subtracting the trend component and the cycle component.

Flowchart of the implementation of STL decomposition algorithm (Figure 6):

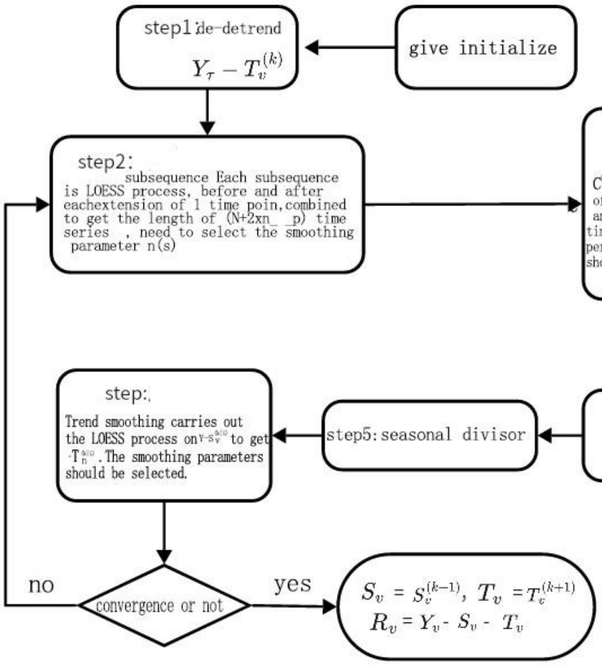


Figure 6: STL decomposition algorithm

There are six types of vegetables, the process of using the STL decomposition algorithm for each type of sales is the same, in order to avoid the process of cumbersome, selected one of the aquatic rhizomes to solve the process of display. Because the total sales = sales before discount + sales after discount, this question is also selected only one of the aquatic rhizome class sales before discount for the solution process demonstration.

To make full use of the multiple periodicities in the data set, the multiple periodicities in the data can be decomposed into different components by the STL algorithm procedure above, and then the prediction can be carried out separately for the periodicities contained in each component. The STL decomposition of the aquatic rhizomatous data yields the trend component as shown in Figure 7:

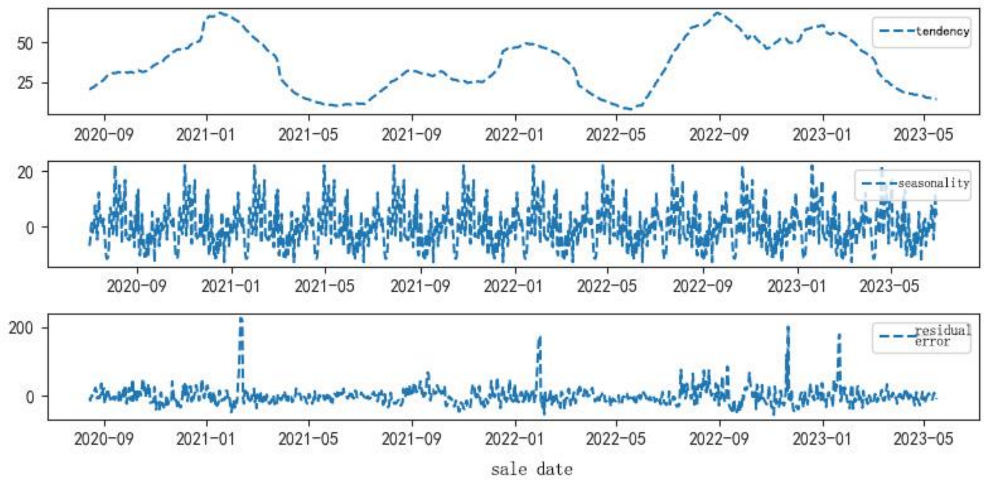


Figure 7: Algorithm Decomposition

Based on the following issues in forecasting: the effect of the observed values of vegetable category sales in the past periods on the current values; the trend and seasonality factors in the time series; and the effect of the past forecast errors on the current values. From this, the time series patterns hidden behind the trend and seasonal factors data are extracted and then used to forecast future data. The mathematical expression for the model is.

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (3)$$

We consider the establishment of ARIMA model for pre-discount and post-discount sales were predicted, and the data for smoothness test and white noise test, if it does not pass the test on it.

(1) smoothness test

The smoothness of the time series is the first time series modelling, smooth time series in the bar chart generally shows the characteristics of fluctuations around its mean up and down, and non-smooth sequences generally at different moments have different trends, different moments have different mean values, if the time series is not smooth time series can be preprocessed into a smooth time series. The data of aquatic rhizomes before discounting are smooth and not processed, while the unstable data of other classes are processed by t-order differencing:

$$\Delta y_t = y_t - y_{t-1} \quad (4)$$

(2) Model fixed order

The final choice of ARIMA (2,0,1) as a model, the other classes of the process is consistent, the use of ARIMA model on the STL decomposition of six categories of vegetables before and after the discount sales prediction, the prediction model will be analysed for accuracy. RMSE and MSE are commonly used and goodness-of-fit metrics to assess the difference between the predicted values of the model and the actual observed values, and smaller RMSE and MSE values indicate a better fit of the model. The accuracy of sales prediction for different vegetable categories before and after discounting is shown in Table 1.

Table 1: Model Accuracy

| VegetableTypes | Salesbeforeandafterdiscount | MAE | RMSE | ARIMA(p,d,q) |
|---|---|---|---|---|
| AquaticRootsandTubers | BeforeSale | 2.39 | 4.15 | (2,0,1) |
| | Aftersale | 0.50 | 1.02 | (2,0,2) |
| FloweringandLeafy | Beforesale | 6.49 | 12.46 | (1,1,1) |
| | Aftersale | 1.30 | 2.35 | (1,1,1) |
| Cauliflower | Beforesale | 2.05 | 2.94 | (2,0,1) |
| | Aftersale | 0.94 | 1.22 | (3,0,2) |
| Eggplant | BeforeSale | 1.10 | 1.78 | (2,0,1) |
| | Aftersale | 0.12 | 0.45 | (2,0,2) |
| Chilli | BeforeSale | 3.56 | 6.58 | (2,0,2) |
| | Aftersale | 0.66 | 1.21 | (1,0,3) |
| Mushrooms | BeforeSale | 3.37 | 6.27 | (2,0,1) |
| | Afterdiscount | 0.65 | 1.19 | (1,1,2) |

The MAE of each category after discounting was small, mostly less than 1, and the RMSE was mostly small. Overall, the sales volume of each vegetable category was fitted well,especially the aquatic roots and tubers, eggplant, chilli and edible mushrooms fitted best after discounting, with a mean square error of about 0.5.

The daily replenishment of different vegetable categories was predicted for 1-7 July 2023 and the results are shown in Table 2:

Table 2: Projected results

| VegetableTypes | Beforeandafterdiscount | 07-01 | 07-02 | 07-03 | 07-04 | 07-05 | 07-06 | 07-07 |
|---|---|---|---|---|---|---|---|---|
| AquaticRootsandTubers | BeforeSale | 11.3 | 12.3 | 13.7 | 15.2 | 16.9 | 18.6 | 20.3 |
| | AfterSale | 5.9 | 5.9 | 5.8 | 5.7 | 5.5 | 5.4 | 5.2 |
| FloweringandLeafy | BeforeSale | 105.0 | 103.5 | 102.4 | 101.6 | 101.0 | 100.7 | 100.4 |
| | Aftersale | 16.4 | 15.7 | 15.2 | 14.8 | 14.6 | 14.5 | 14.4 |
| Cauliflower | BeforeSale | 17.3 | 18.9 | 20.5 | 21.9 | 23.2 | 24.4 | 25.5 |
| | Aftersale | 0.9 | 1.2 | 1.5 | 1.7 | 1.7 | 1.7 | 1.6 |
| Eggplant | BeforeSale | 18.5 | 18.2 | 18.0 | 17.9 | 17.9 | 17.9 | 18.0 |
| | Aftersale | 0.0 | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 |
| Chilli | BeforeSale | 75.8 | 78.2 | 77.9 | 76.1 | 74.3 | 73.5 | 74.0 |
| | Aftersale | 4.6 | 4.6 | 4.6 | 4.6 | 4.7 | 4.5 | 4.7 |
| Mushrooms | BeforeSale | 35.3 | 36.1 | 37.4 | 39.1 | 40.9 | 42.9 | 44.8 |
| | Afterdiscount | 8.6 | 8.4 | 8.3 | 8.2 | 8.2 | 8.2 | 8.2 |

Based on the forecast, the daily replenishment strategy is the daily forecasted sales plus losses，Sales strategy is shown in table 3:

$$T_i = \frac{y_i}{1-\theta_i} \tag{5}$$

Table 3: Sales strategy

| Date | 2023/7/1 | 2023/7/2 | 2023/7/3 | 2023/7/4 | 2023/7/5 | 2023/7/6 | 2023/7/7 |
|---|---|---|---|---|---|---|---|
| Ordering Strategy | 47.91 | 48.26 | 49.67 | 54.54 | 53.36 | 55.54 | 57.60 |

## 4. Conclusions

This research paper systematically analyses vegetable sales data with the aim of revealing the distribution patterns of sales volumes and predicting future trends. Firstly, the relationship between different vegetable categories and individual products was explored in depth through statistical and visual analyses of their sales. Based on this, further research on cost-plus pricing was conducted to analyse the relationship between sales volume and pricing strategy through scatter plots and fitted equations, which provided a prediction model for the total replenishment and pricing in the coming week. Through statistical methods such as Pearson's test and ARIMA model, the correlation between different vegetable types and individual items and future sales trends were comprehensively demonstrated, providing important data support and suggestions for vegetable sales management and decision-making.

## References


*[1] Li Teng, Zhang Rulan, Sun Yan. Research on dynamic optimisation strategy of return position allocation for intelligent warehouse shelves [J/OL]. Frontiers of Engineering Management Science and Technology, 1-10[2024-04-28].*

*[2] Xie WJ. A review of research on joint replenishment problem [J]. China Storage and Transportation, 2024, (04): 138-139.*

*[3] Guo Y. Impact of digital transformation on supply chain management [J]. Old Brand Marketing, 2024, (06): 42-44.*

*[4] Nie Yuxuan. Research on automatic pricing and replenishment strategy of fresh commodities based on ARIMA prediction optimisation model - taking vegetable commodities as an example [J]. Commercial Exhibition Economy, 2024, (05): 19-22.*

*[5] Guan ZZ, Wang YW, Kang HWF. A study on supply chain ordering strategy preference considering returns [J]. Industrial Engineering, 2024, 27 (01): 120-127.*

*[6] Jia Mengyao, Zhang Jihui. Optimisation of site selection-inventory-distribution integration considering disruption risk [J/OL]. Complex Systems and Complexity Science, 1-10[2024-04-28].*
*[7] Ma Chao, He Juan, Huang Fuyou,et al. VMI Supply Chain Coordination Considering Loss Avoidance and Replenishment Strategies [J]. Operations Research and Management, 2024, 33 (01): 69-75.*
*[8] Huang Lixue, Yao Guanxin. Research on horizontal replenishment of fresh produce in supermarket chains [J]. Logistics Science and Technology, 2024, 47 (01): 100-104.*
*[9] Li Z, Zhang X P, Zhou Y F, et al. Performance assessment of blood supply chain based on complex adaptive system and depot ageing strategy [J]. Business Research, 2023, 30 (06): 106-117.*
*[10] Lin Qinwei. Research on dynamic optimisation method of automotive MRO parts supply chain based on CPFR [D]. Jimei University, 2023.*