# A Study on Pricing and Replenishment Decision of Vegetables in Fresh Superstores Based on Time Series Analysis

Jiangning Wang*

School of Pharmaceutical Engineering, Shenyang Pharmaceutical University, Benxi, China

*Corresponding author: wang-77W7@qq.com

**Abstract.** This paper focuses on an in-depth study of the vegetable pricing and replenishment decision problem in fresh produce superstores. Using data preprocessing and multiple mathematical models, the sales and correlation laws between different categories and single products of vegetable products are analyzed. Firstly, the cyclical pattern of sales is revealed through the analysis of historical sales data; secondly, the pricing and replenishment strategies are optimized by using multiple linear regression and time series models; finally, an optimization model is constructed to predict future sales and replenishment, taking into account the spatial limitation of sales.

**Keywords:** Vegetable Pricing, Replenishment Decision, Multiple Linear Regression, Time Series Analysis.

## 1. Introduction

In the fresh food superstore sector, vegetable products face the problems of shorter shelf life and deterioration of overnight products, making it difficult to re-sell them. In the absence of specific products and purchase prices, merchants need to rely on historical sales data to discount overnight items and use cost-plus pricing to make pricing and replenishment decisions [1]. The analysis of market demand and supply chain reveals the correlation between vegetable sales volume and time of day, as well as variations in supply due to time constraints, highlighting the importance of having the right sales strategy. In this study, based on the available data on vegetable categories, the correlation and sales volume distribution patterns among vegetable items were explored. Further, the relationship between total sales volume and cost-plus pricing for each vegetable category was analyzed, and total daily replenishment and pricing strategies were developed for specific time periods to maximize superstore revenue. Finally, considering the limitation of sales space, the replenishment volume and pricing strategy for individual items were developed with the aim of maximizing the revenue of the superstore while meeting the market demand [2, 3].

## 2. Vegetable marketing correlation analysis

### 2.1. Vegetable Sales Correlation Analysis Model

To more intuitively determine the number of individual products sold and the interrelationship between individual products, this paper introduces the concept of "sales". Because the data set is too large, the total sales volume of six types of vegetable products is firstly summarized in this paper. By calculating the total sales volume of each category, as shown in Fig. 1, the sales volume of leafy vegetables is much higher than that of the other five categories, close to 200,000, while the sales volume of the other five vegetable categories does not exceed 100,000, of which the sales volume of cauliflower, aquatic roots and eggplant does not exceed 50,000, which is a big gap between the sales volume of the other vegetable categories.

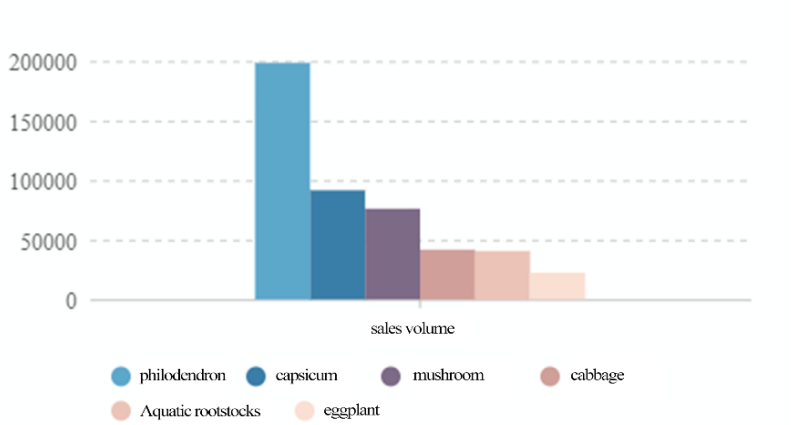



**Fig. 1** Distribution of sales volume by vegetable category

With the above data, this paper hypothesizes that the total amount of sales may be closely related to time changes and that market sales volume is a typical variable that changes over time. To analyze the relationship more intuitively between sales and time among various vegetable categories, this paper takes quarter as the time unit and draws a line graph. As shown in Fig. 2, the six categories of vegetable products basically show the cyclical change of "peak - trough - peak", with a period of about half a year, showing obvious seasonality. Moreover, all categories of vegetable products reached the highest sales value in the third quarter of 2022 and returned to the trough in the fourth quarter of 2022 (October-December) with a sudden change.

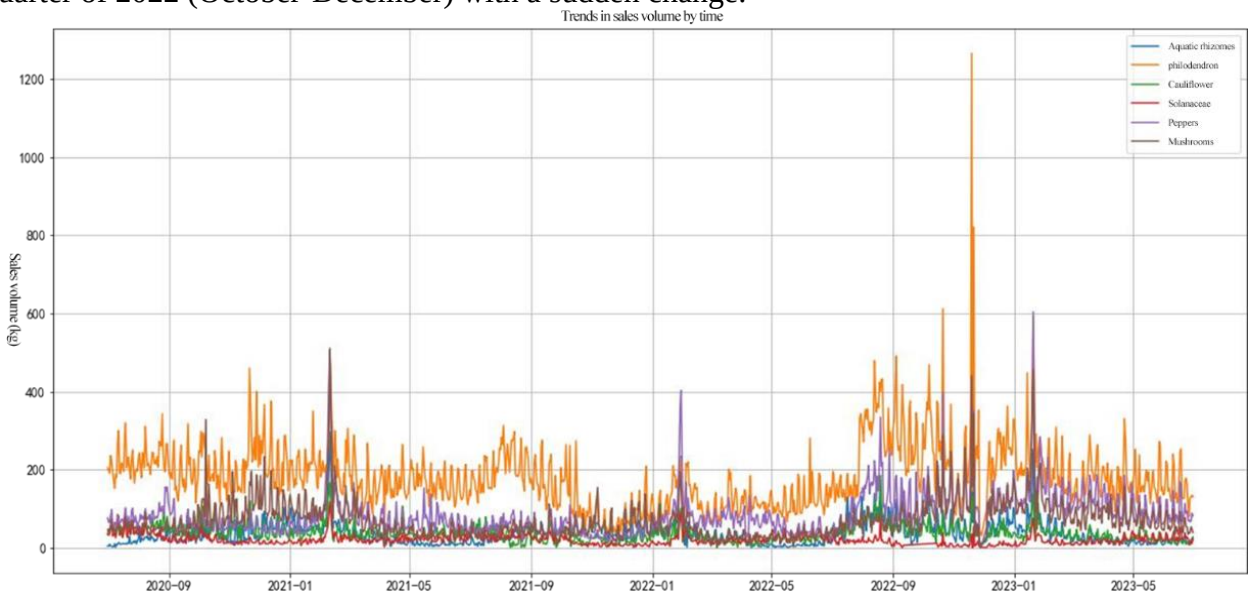



**Fig. 2** Change in sales by vegetable category

## 2.2. Vegetable marketing correlation solving

In statistics, the Pearson correlation coefficient (abbreviated as PPMCC or PCCS), which is used to measure the correlation between two variables X and Y, measures the magnitude of the correlation between the two variables, and takes a value between -1 and 1. The closer the correlation coefficient is to -1 or 1, the larger the absolute value of the correlation coefficient is, and the stronger the correlation between the two; the closer the correlation coefficient is to 0, the smaller the absolute value of the correlation coefficient is, and the weaker the correlation between the two. Through the final integration of sales month and sales, this paper utilizes the Pearson coefficient to find out the linear relationship between different columns [4]. In statistics, the formulas for sample covariance, sample standard deviation and Pearson's correlation coefficient are key tools to quantify the linear dependence between columns.

Afterwards, the absolute mean of each column is calculated, sorted in descending order of the mean, and a heat map is programmed to visualize the correlation data. The color of the heat map indicates the strength of the correlation, the darker the color the stronger the correlation. As shown in Fig. 3:

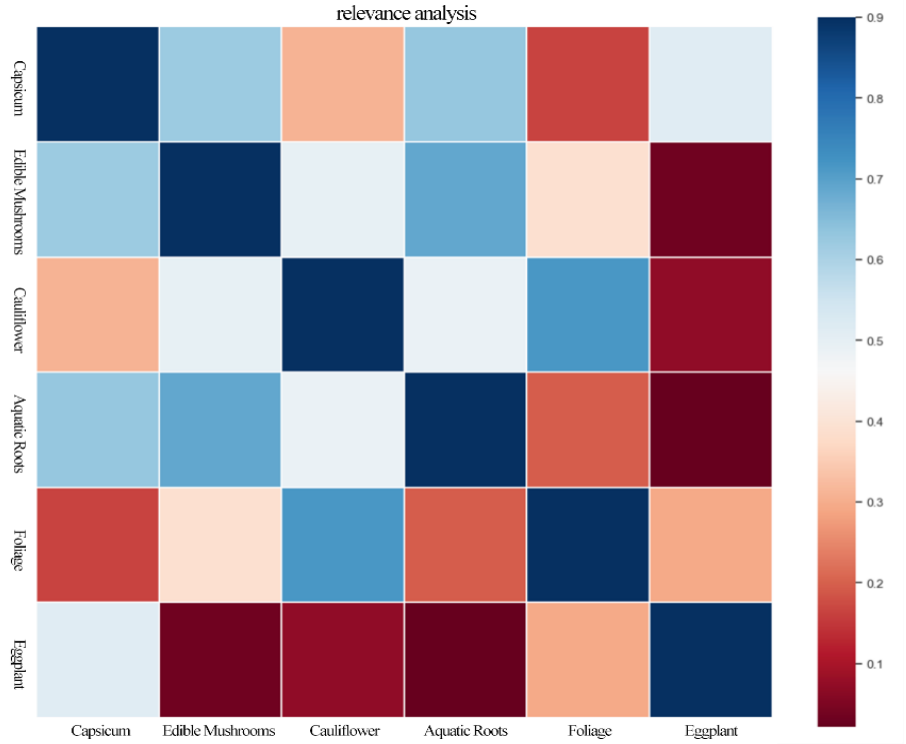



**Fig. 3** Heat map between different categories of vegetable goods

According to the heat map shown above, the analysis yielded that the floribundas showed strong correlation with peppers, cauliflowers, and edibles; and the eggplant showed weak correlation with edibles, cauliflowers, and aquatic roots. All the five categories possess more than moderate correlation except for the eggplant category, which is the only one with weak correlation.

This paper utilizes programming to process the "Item Name" column, removing the contents of the parentheses to prevent runtime errors. After grouping by "month of sale" and "item name," the sales of each group are summed to obtain the total sales of each individual item. The index is then reset, representing each row as a month and each column as a single product sale. Null values are filled with 0 to ensure all cells contain numerical values. For analyzing the correlation between different individual items in the vegetable category, the Pearson coefficient is used to find the linear relationship between the columns. The mean absolute value of each column is calculated, and the top 50 columns with the highest correlation, in descending order, are selected. Programming is employed to draw a heatmap visualizing the correlation data, with the color depth indicating the strength of the correlation.

As shown in Fig. 4, it is possible to observe the correlation more specifically between different individual products in vegetable products, for example: the golden mushroom has a strong correlation directly with bamboo leaf lettuce, yellow cabbage, and Xixia shiitake mushroom; and there is a strong correlation between baby bok choy and sweet potato tips and bamboo leaf lettuce. By analyzing the various types of individual products, it is easy to formulate replenishment strategies and pricing strategies in a more precise and targeted manner.

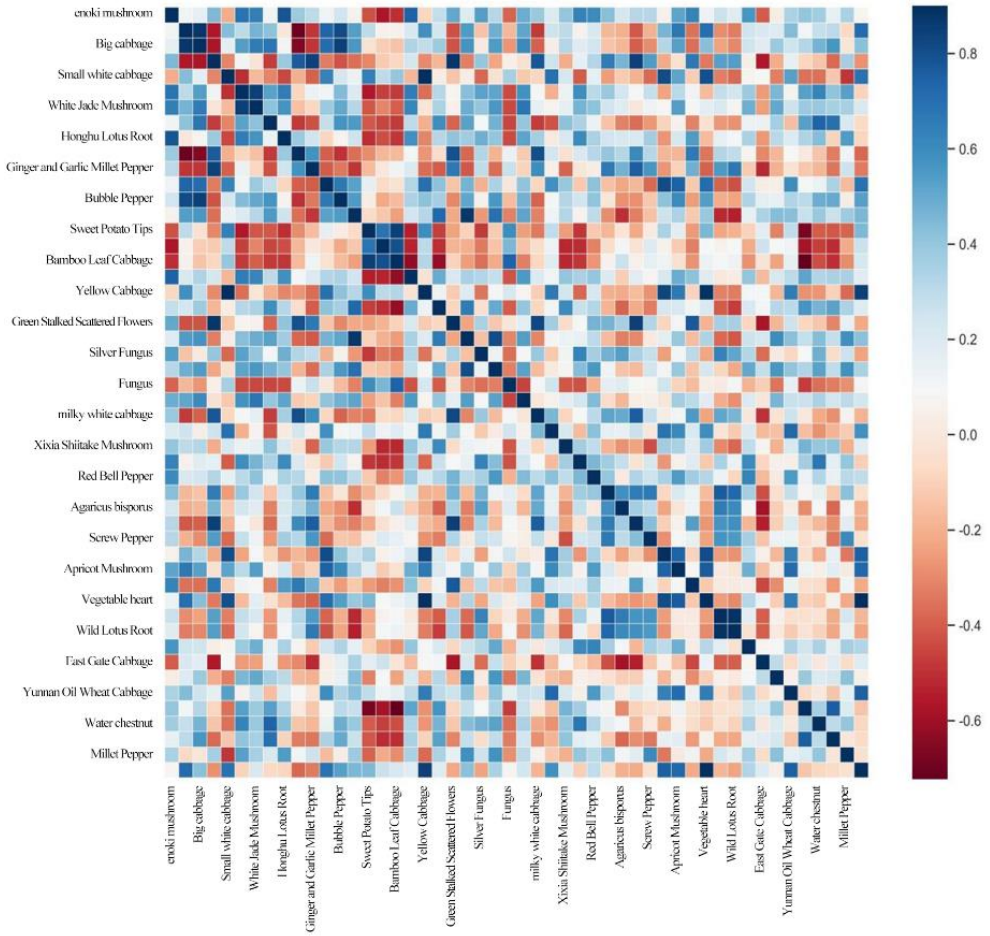



**Fig. 4** Heat map of different individual products in the vegetable category

## 3. Category sales and pricing strategy optimization

### 3.1. Multiple linear regression models

In the study of this paper, multiple linear regression model was used to describe and predict the complex relationship between sales volume and each of the influencing factors such as total sales volume and cost-plus pricing. The model is based on the least squares method of parameter estimation, which aims to minimize the sum of squares of prediction errors. This method can handle multiple independent variables and is therefore ideal for solving this research problem. Compared to the time series approach, the multiple linear regression model can consider multiple factors simultaneously, providing a more comprehensive perspective. However, it is worth noting that the multiple linear regression model assumes that the data conforms to a linear relationship and that multicollinearity should be avoided as much as possible between independent variables. Considering these factors, this paper uses the model to not only predict the sales volume in the coming week, but furthermore to develop replenishment and pricing strategies based on the predicted data. This integrated approach will help to meet the market demand more accurately and thus achieve the business objectives. Its expression is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon \quad (1)$$

The difference between the final prediction and the actual value is shown in Fig. 5.

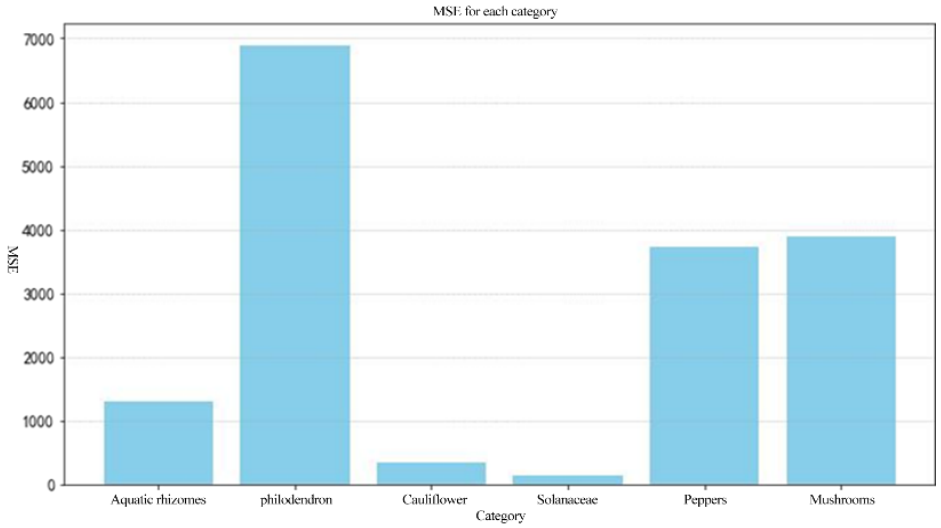



**Fig. 5** Mean error between predicted and actual values

According to Fig. 5 it can be analyzed that there are some differences between the actual values and the predicted values, and the prediction results of the linear model are not very satisfactory. Further analysis of the time series data for each category of goods, the sales graph is shown in 6.

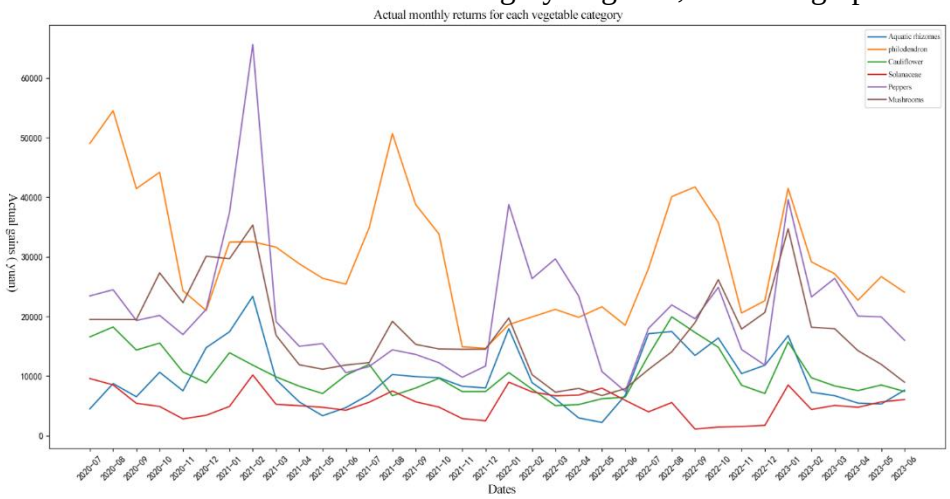



**Fig. 6** Real Returns on Different Vegetable Products

Fig. 6 clearly reveals the significant impact of the time factor on the sales volume of vegetable products, in particular, the product sales show some cyclical variations. Based on this observation, this paper forecasts the future sales volume of each vegetable category from a time series perspective. This analysis not only confirms the important role of the time factor in the variation of sales volume, but also provides strong data support for more accurate replenishment and pricing decisions.

### 3.2. Category Sales and Pricing Strategy Forecasts

In this paper, time series methodology is used for the treatment and in particular, ARIMA (Autoregressive Integral Sliding Average Model) is applied for detailed time series analysis [5]. Time series analysis is a statistical method specialized in the study of data points arranged in chronological order, which is widely used in various fields, such as stock prices, climate change, monthly sales, etc. In the scenario of a fresh produce superstore, time series analysis is of great practical value because it can accurately predict future sales trends, thus helping merchants make more scientific and reasonable pricing and replenishment decisions. Taking "aquatic rhizomes" as the research object, this paper combines the ARIMA model to predict the sales volume of this type of plants from July 1 to 7, 2023.The ARIMA model analyzes the past sales data, identifies the potential seasonal and trending patterns, and then predicts the sales volume of this type of plants in the future period. . The results are shown in Table 1.

**Table 1.** Forecast of sales of aquatic rooted plants

| Date of sale | Sales volume (kg) |
|---|---|
| 2023-07-01 | 18.95 |
| 2023-07-02 | 16.32 |
| 2023-07-03 | 17.17 |
| 2023-07-04 | 16.73 |
| 2023-07-05 | 18.61 |
| 2023-07-06 | 18.39 |
| 2023-07-07 | 18.49 |

The advantage of time series analysis over linear regression model forecasting is that it provides date-specific forecasts, which facilitates more granular and time-sensitive management decisions. For example, merchants can more accurately price and restock their products based on forecast data during specific holidays or events.

On this basis, this paper extends the analysis of individual categories to all categories for a more comprehensive and in-depth analysis. By applying the time series model, the model can predict the sales trend of each category in the coming period, so that merchants can make a more balanced and reasonable resource allocation among different categories.

Further, this paper also introduces the average markup percentage and wholesale price based on historical sales to infer the expected sales price and wholesale price in the coming month. These forecasts not only consider seasonal, trend and cyclical factors, but also incorporate a variety of factors such as cost and market demand to generate more accurate and reliable forecast data. The specific forecast results are shown in Table 2, and merchants can use this data to adjust their pricing strategies and supply chain management to maximize profits and improve customer satisfaction.

**Table 2.** Projected selling and wholesale prices of different vegetable categories

| Name | Radio | Price | Expected sales price |
|---|---|---|---|
| Aquatic Roots | 0.729818 | 7.285789 | 12.603088 |
| Foliage | 0.836976 | 8.203400 | 15.069448 |
| Cauliflower | 0.803050 | 8.326000 | 15.012193 |
| Eggplant | 0.712397 | 4.089000 | 7.001992 |
| Peppers | 0.766107 | 6.291556 | 11.111561 |
| Mushrooms | 0.617496 | 8.066944 | 13.048250 |

In superstore operations, it is critical to accurately predict replenishment volume, especially for vegetable categories with short shelf life and easy to damage. To more accurately forecast replenishment volume, this paper proposes a definition formula for replenishment volume: replenishment volume = forecasted sales volume / (1 - wastage rate). This definition considers two key factors, the forecasted sales volume, and the attrition rate of the merchandise, to provide a more comprehensive and accurate replenishment volume forecast.

Based on the above definition, combined with the forecast selling price and wastage rate, this paper further calculates the expected replenishment volume for each vegetable category. First, a time-series model was used to forecast the sales volume of each vegetable category over a future period. Then, based on historical data and market dynamics, the wastage rate of the corresponding category is estimated. Finally, these two factors were substituted into the definition formula of replenishment volume to obtain the expected replenishment volume of each vegetable category.

The specific forecast results are shown in Table 3:

**Table 3.** Projected selling prices and replenishment of different vegetable categories

| Classification name | Expected sales price ($/kg) | Replenishment |
|---|---|---|
| Aquatic Roots | 12.603088 | 135.0783 |
| Foliage | 15.069448 | 6145.5928 |
| Cauliflower | 15.012193 | 1018.2792 |
| Eggplant | 7.001992 | 482.4585 |
| Peppers | 11.111561 | 3992.3506 |
| Mushrooms | 13.048250 | 2623.3588 |

The selection of the full week of July 1-7, 2023, as the forecasting cycle in this paper is reasonable. Since there are no scheduled holidays during this cycle, the forecast data more accurately reflects daily sales and replenishment patterns. Further, the stability of the 30-day forecast data increases the accuracy of the forecast for that week. In this case, the above forecasts are not only highly feasible, which further confirms the value of the time series model as well as the replenishment definition formula proposed in this paper in practical applications.

## 4. Optimization of replenishment and pricing strategies for individual products

### 4.1. A model for single-item replenishment and pricing optimization

Based on the previous analysis, new constraints are added to the model, aiming to develop a replenishment plan and pricing strategy for a single product for the coming day, and the scenario can be considered as a combinatorial optimization problem to be solved. The constraints involved include:

(1) The total number of products available for sale needs to be kept between 27 and 33.

(2) The order quantity of each product needs to satisfy the minimum display quantity of 2.5 kg.

(3) The exact price is determined based on cost-plus pricing.

Based on the above constraints, this paper uses linear programming for modeling and solving.

### 4.2. Solving for single-item replenishment and pricing optimization

Sales data from June 24-30, 2023, was selected for analysis in this paper. If there is missing data in this time, the corresponding data for May will be substituted. The sales volume for July 1, 2023, is predicted through a linear programming model based on the sales pattern of the past week. With all constraints met, the model outputs seven replenishment and pricing scenarios involving 27 to 33 products.

After careful comparison, the model predicts the maximum total revenue when the total number of products available for sale is 33, which is approximately $60.06. This result suggests that by appropriately increasing the assortment of products, it helps to increase the total return of the superstore. Then, this paper further calculates the predicted replenishment of each individual item under this scenario and integrates the results in Fig. 7. This provides a comprehensive set of operational recommendations for superstore decision makers, which helps to make replenishment and pricing decisions in a more refined and scientific manner.

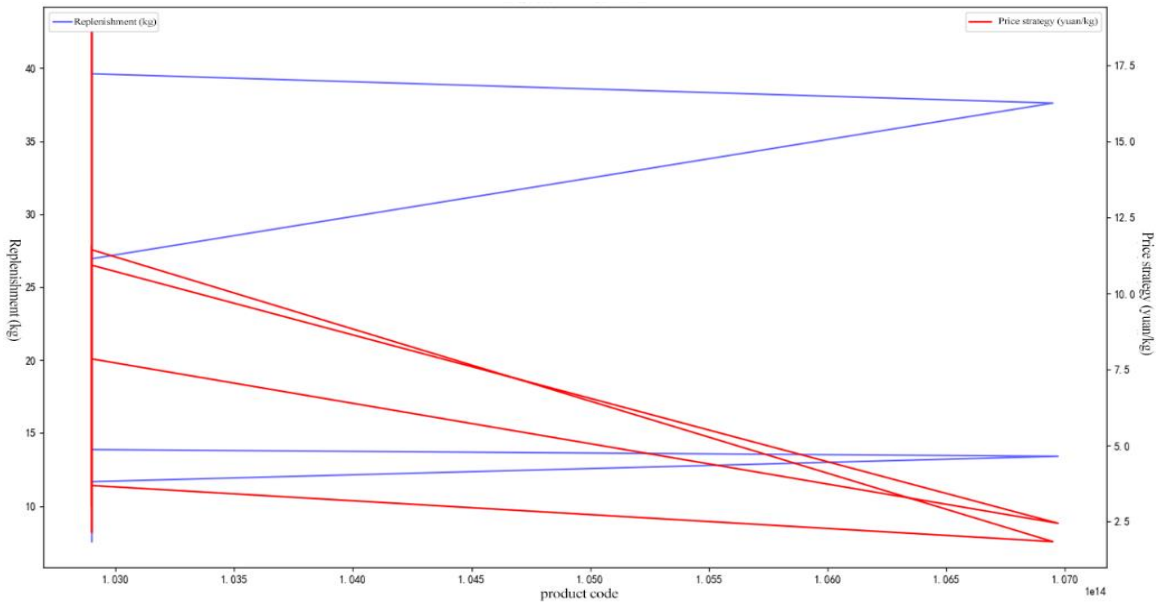



**Fig. 7** Forecasted sales and replenishment of different vegetable items

## 5. Summary

This study explores the sales and replenishment strategies of vegetable items in supermarkets through meticulous data preprocessing and multivariate statistical analysis and constructs an optimization model to seek maximum benefits. Data preprocessing before model construction ensures the accuracy of the model. Through multivariate linear analysis, Pearson correlation coefficient and other methods, the distribution laws and interrelationships of the sales data were deeply explored. The optimization model considers relevant constraints for the total daily replenishment and pricing strategy, aiming to achieve optimal pricing and replenishment decisions to maximize profits. Consideration of multifaceted data factors provides background information for developing comprehensive merchandise pricing and replenishment decisions. However, this study also reveals shortcomings such as the complexity of data processing, the complexity of model building, and the difficulty of model implementation, which may affect the application of the model in actual superstores. Nevertheless, this study provides useful references for merchandise pricing and replenishment decisions in superstores and draws lessons for future research in related fields.

## References

[1] MENG Yi, ZHANG Yuhua, XU Aojie et al. Current situation and strategy of supply chain management in fresh food e-commerce [J]. China Fruit and Vegetable, 2023, 43(09): 88-92. DOI: 10.19590/j.cnki.1008-1038.2023.09.015.

[2] Gu, S. H.. A study on dynamic pricing of fresh products in H-retailers considering freshness variation [D]. Donghua University, 2023. DOI: 10.27012/d.cnki.gdhuu.2023.001318.

[3] Yang Shuai, Huang Xiangmeng, Wang Junbin. Research on joint optimization strategy of shelf allocation and pricing for fresh food [J]. Supply Chain Management, 2022, 3(08): 49-59. DOI: 10.19868/j.cnki.gylgl.2022.08.005.

[4] Zhang YH. Research on time-lag Pearson correlation analysis and key variable prediction method for air separation equipment [D]. Hangzhou University of Electronic Science and Technology, 2023.

[5] WU Hui-Hui, WANG Jia-Peng, WU Wen-Jing et al. Analysis of global air surface temperature prediction based on ARIMA model [J]. Modern Information Technology, 2023, 7(16): 147-150. DOI: 10.19850/j.cnki.2096-4706.2023.16.032.