

Research on pricing and stocking of fresh goods based on nonlinear programming

Langkun Jiang^{1, #}, Keqiang Wang^{2, #}, Jiaxu Liao^{2, #}, Xiaoping Liu^{2, *}

¹Department of Computer Technology and Applications, Qinghai University, Xining, China, 810016

²School of Finance and Economics, Qinghai University, Xining, China, 810016

*Corresponding author: qdcylxp@163.com

#These authors contributed equally.

Abstract. In the fresh supermarket, due to the particularity of fresh goods, the fresh-keeping time of vegetable goods is relatively short, generally the wholesale goods of the same day need to be sold on the same day, if the day is not sold, it will produce waste phenomenon. Therefore, fresh supermarkets generally choose to price and replenish goods every day according to their experience and sales status. By establishing the relevant mathematical model, this paper gives the replenishment and pricing decision of vegetable commodities after the supermarket optimization, so as to maximize the profit of the supermarket. First, the box chart is used to pre-process the data, and then the distribution law of the sales volume of each vegetable category and single product over time is analyzed by data analysis and visualization technology. Then the Pearson correlation coefficient of the sales volume of each vegetable category and each single product is obtained by heat map analysis, and the correlation relationship between them is inferred. Then, by calculating the average daily sales pricing of various vegetable varieties as the cost plus pricing of dishes, using linear function and logarithmic function to fit the sales volume and sales price of each vegetable category to get the functional relationship between the two. Finally, ARIMA model is used to predict the wholesale price of each dish within seven days, and then the equation model is established. Finally, SLSQP method is used to predict the total amount of replenishment and pricing strategy of six vegetable categories in the next week.

Keywords: Data analysis and visualization, ARIMA model, Nonlinear programming.

1. Introduction

The freshness period of vegetable commodities in fresh supermarkets is relatively short, and its quality is constantly declining with the increase of sales time. If most of the commodities are not sold on the same day, they can not be sold again the next day, so the supermarket will take the previous sales volume as a reference for daily replenishment. In addition, the purchase transaction of the supermarket is usually from 3:00 to 4:00 in the morning, and the merchant needs to make pricing decisions without exactly knowing the specific item and purchase price. The price of vegetables is generally "cost plus pricing" method, and the merchant will also discount the goods that are damaged and the product phase change is different. Therefore, according to the reliable market demand analysis, it is of great practical significance to work out the replenishment and pricing strategy of fresh supermarket vegetable commodities.

2. Data processing and simple analysis

2.1. Data preprocessing

In this paper, the data processing method of box diagram[1] is used to visualize the data, screen out the outliers and clean them. The median and quartile in the data are extracted from the box diagram, the upper quartile is Q3, the median is Q2, the lower quartile is Q1, where $Q3-Q1=IQR$. The quartiles are arranged in the order of Q1Q2Q3, and the bias of the data can be seen by looking at the skew of the box. In this paper, values greater than $Q3+1.5IQR$ and less than $Q1-1.5IQR$ in the data

are defined as outliers, and these outliers are deleted in this paper. Through the processing of the box diagram, we can not only see the distribution of the data, but also automatically process the outliers. The following figure shows the distribution of the six categories according to the daily sum of the box diagram.

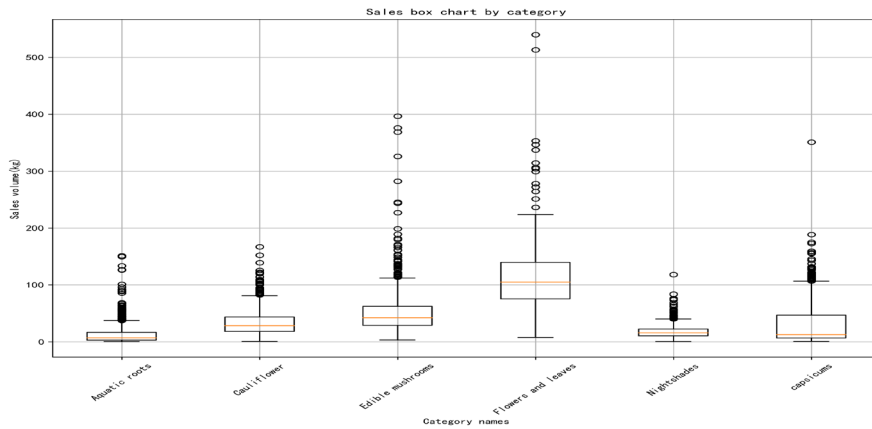


Fig 1. Box chart of sales volume of the six categories

As shown in Fig 1, the extreme outliers outside the upper and lower edges are deleted. This paper believes that data beyond the normal value is not conducive to the subsequent data analysis to obtain the general rule, and the data will be distorted if other values are used instead, so this paper directly deletes the outliers.

2.2. Distribution law of each vegetable category and single product

First of all, this paper analyzes the sales volume of each category of vegetables, and obtains the distribution of the sales volume of each category of vegetables as shown in the figure below.

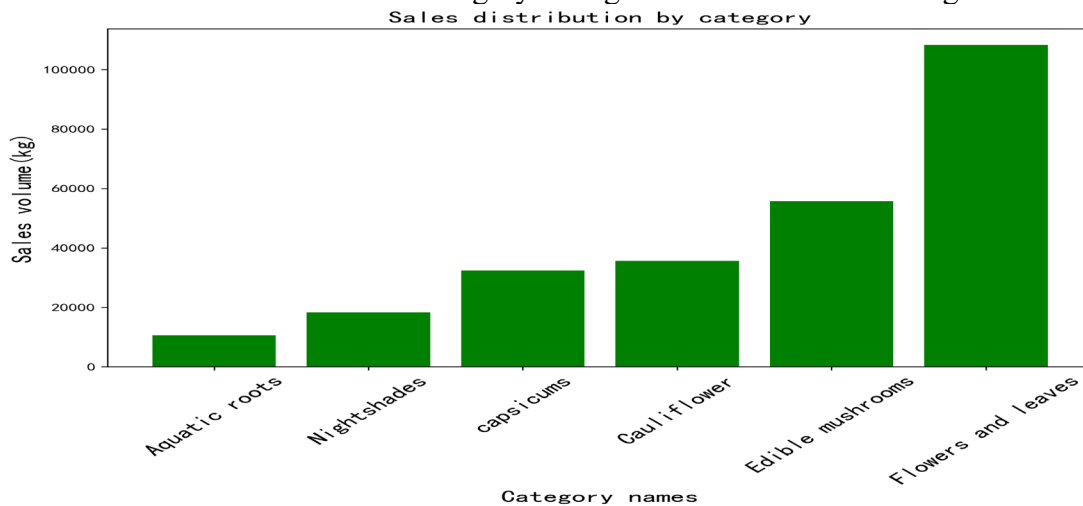


Fig 2. Distribution of sales volume of various categories of vegetables

It can be clearly seen from Fig 2 that among all vegetable categories, the sales volume of flowers and leaves is the highest, followed by chili and edible fungi, and the sales volume of aquatic roots and cauliflower is at the same level. The lowest sales volume of dishes is nightshade.

Then, this paper takes the four seasons of spring, summer, autumn and winter as the four quarters of each year, and draws a line chart of the sales volume of each vegetable category in each quarter. The figure below shows the line chart of the sales volume change of each vegetable category in 11 quarters from September 30, 2020 to June 30, 2023.

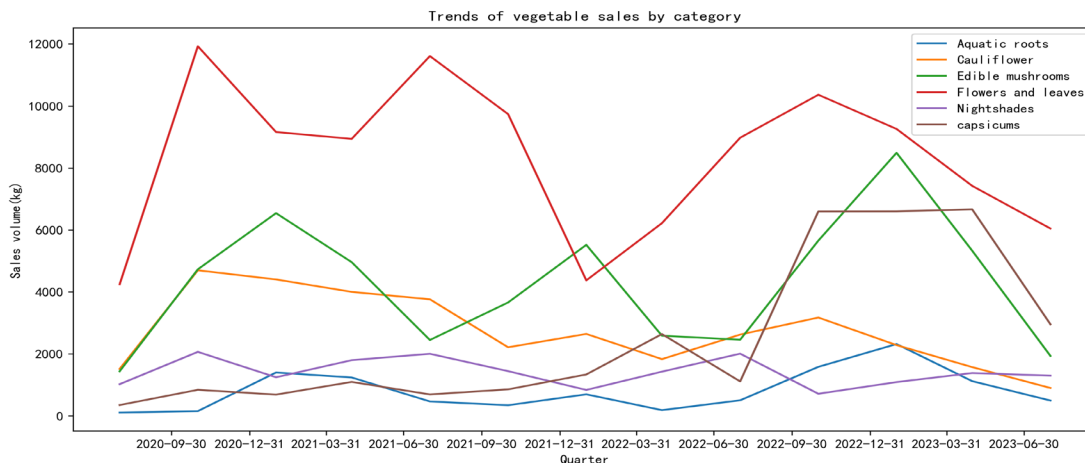


Fig 3. Quarterly changes in sales volume of vegetables by category

As can be seen from Fig 3, the sales volume of each vegetable category changes with the change of the quarter, that is, its sales volume is affected by the change of time. For example, it can be analyzed from the above figure that the sales volume of chili dishes shows an obvious cyclical trend, increasing in the first and third quarters and decreasing in the second and fourth quarters. The sales volume of Mosaic vegetables generally shows an increasing trend in the third quarter and reaches the maximum value before the Chinese New Year every year. Therefore, the quarterly variation factor has a very important reference significance for the supermarket in the formulation of replenishment and pricing strategy.

In the following, by analyzing the changes in the unit price of different dishes and combining with the sales volume of different dishes in different quarters in the above figure, the line chart of the changes in the unit price of sales and the line chart of the changes in the total price of six categories of dishes are obtained respectively.

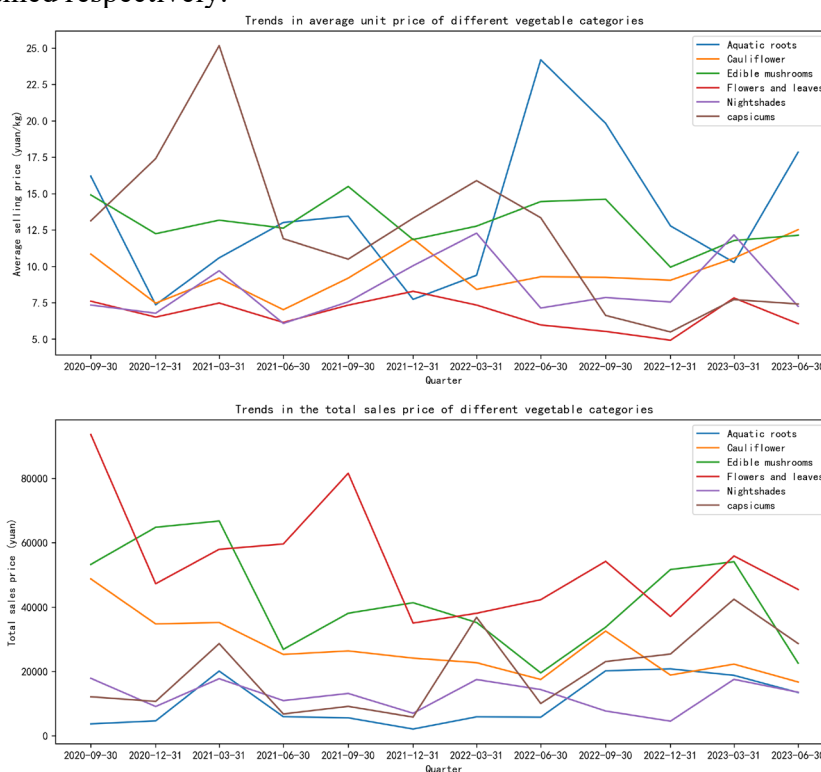


Fig 4. The change trend of the unit sales price and total sales price of each category of vegetables

The comparison of Fig 4 shows that the total sales price of vegetables changes periodically, and the vegetables with the highest sales volume in different cycles are different. For example, in the third quarter, when the total sales price of Mosaic vegetables reaches the highest, it will also become the

dishes with the highest total sales price among the six categories. At the same time, it can be observed that there is a very close relationship between the unit price of dishes and the total sales price, and the two always maintain a positive correlation. Therefore, different pricing schemes formulated in different cycles can have different impacts on the total sales price.

This paper also analyzes the 15 items with the highest sales volume in the same way. The figure 5 is the distribution of the sales volume of each vegetable item.

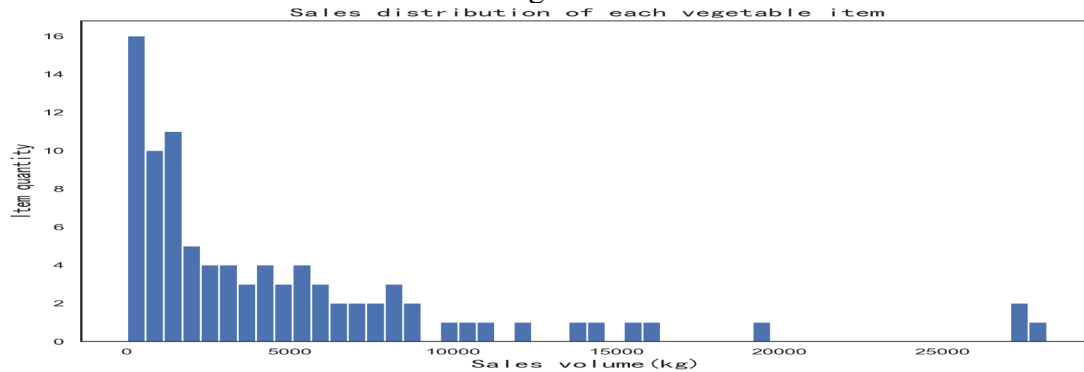


Fig 5. Distribution of sales volume of each vegetable item

The following is a summary calculation of the top 15 items and their sales volume.

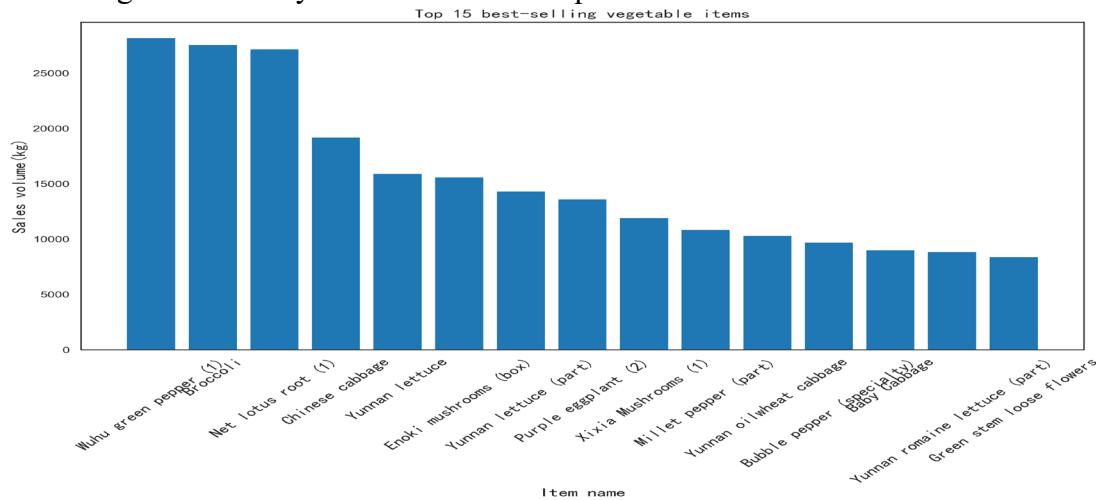


Fig 6. Top 15 best-selling vegetable items

As shown in Fig 6, it can be seen that Wuhu green pepper, broccoli and lotus root have the highest sales volume. The sales volume of these three dishes is much higher than other dishes, so these three dishes have the greatest impact on the sales volume and total sales volume of the categories they belong to.

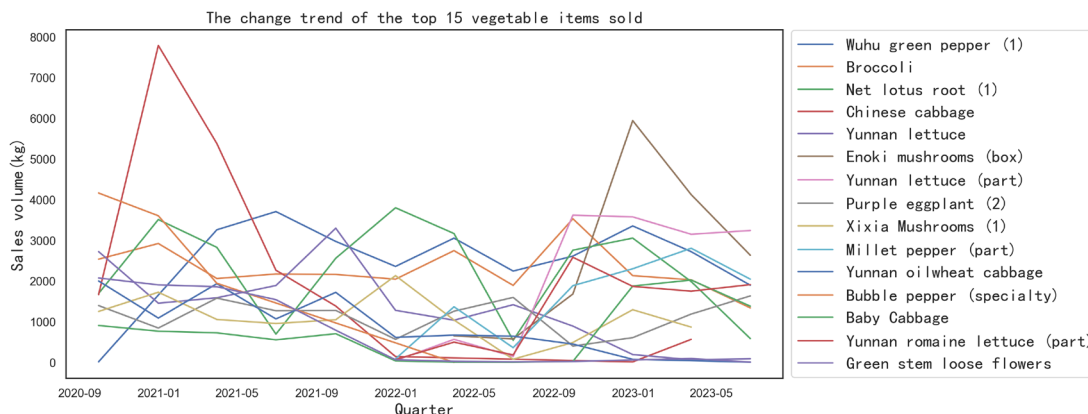


Fig 7. Sales trends of the top 15 vegetable items

As can be seen from Fig 7, the sales volume of various categories of dishes has an obvious change trend with the change of seasons, and some vegetables have abnormal sales volume in some years, which is considered by this paper to be a normal situation of market self-regulation.

2.3. The internal relationship between the sales volume of each vegetable category and single item

First of all, this paper takes a quarter as a cycle and uses heat map to analyze the correlation between various categories. Since the data is normally distributed, this paper chooses Pearson coefficient [2] as the basis for heat map analysis, and the calculation formula is as follows:

$$\rho_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

Where X_i and Y_i are the observed values of variables X and Y , and \bar{X} and \bar{Y} are the average values of variables X and Y , and n is the number of observed values.

The figure 8 shows the correlation coefficient [3] matrix of the sales volume of each category of vegetables.

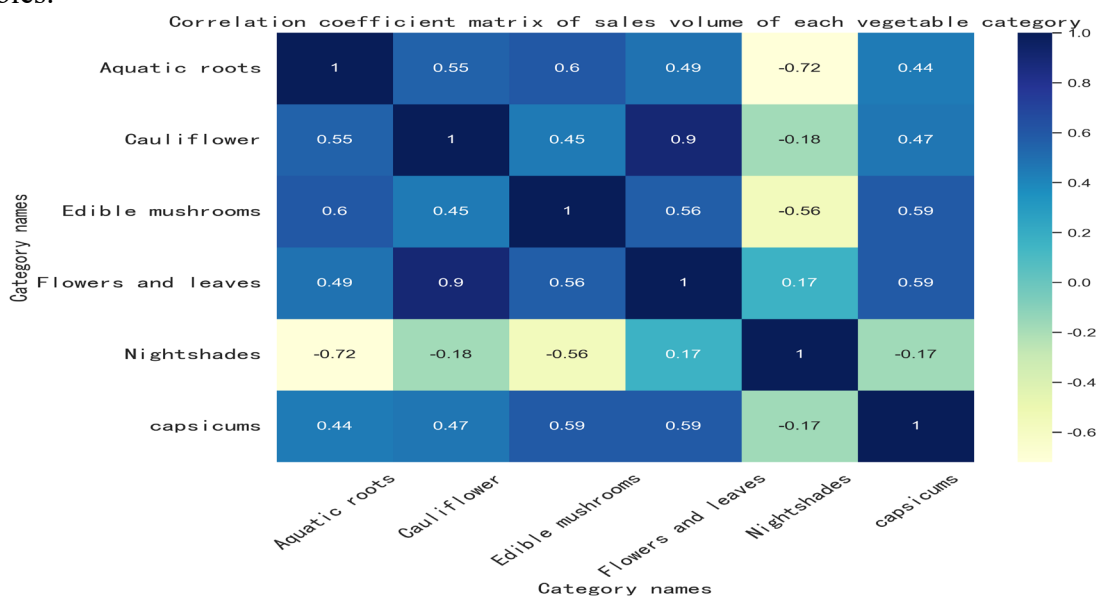


Fig 8. Correlation coefficient matrix of sales volume of each vegetable category

In the thermal map coefficient matrix[4], the greater the absolute value of the coefficient, the greater the correlation between the variables, and when the coefficient is positive, the variables are positively correlated, and when the coefficient is negative, the variables are negatively correlated. The color of the coefficient matrix is related to the absolute value of the coefficient, the greater the absolute value, the stronger the correlation, the darker the color of the matrix. It can be seen from the figure above that there is a strong correlation between the following dishes:

In this paper, it is found that there is a strong positive correlation between cauliflower and flowering leaves, and the existence of both is conducive to the sales of commodities. The correlation coefficient between nightshade and aquatic rhiza is -0.72, indicating that there is a certain negative correlation between the two. When the sales of nightshade increase, the sales of aquatic rhiza will decrease.

Then, this paper also used a similar method to analyze the relationship between the top 20 sales of different products. The following figure is the thermal map correlation coefficient matrix of the top 20 sales of products.

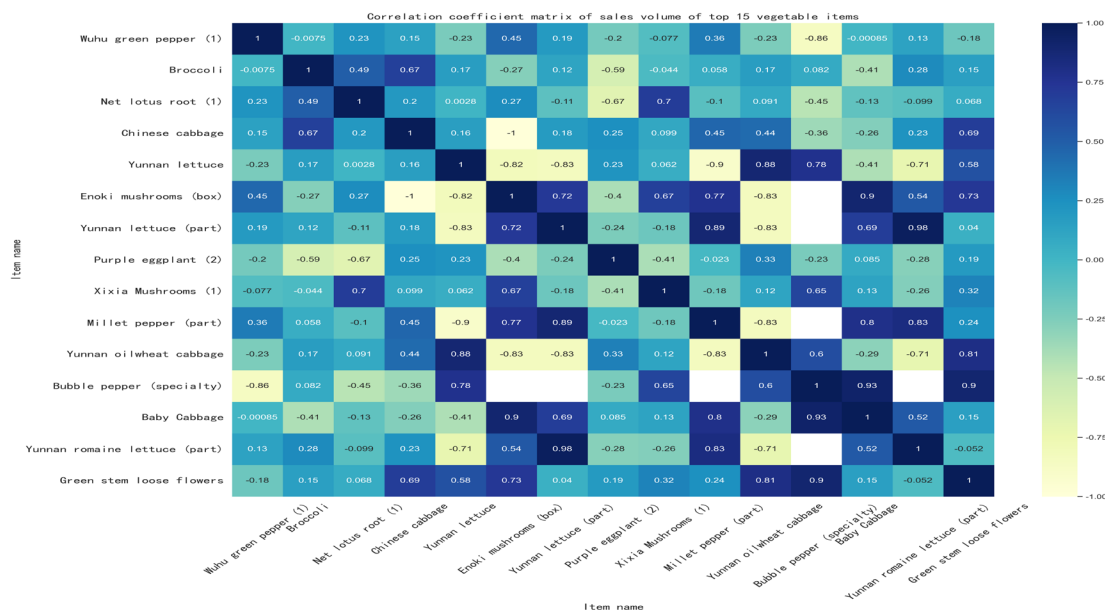


Fig 9. The correlation coefficient matrix of the top 20 vegetable products

As can be seen from Fig 9, the correlation coefficient between Yunnan romaine lettuce and Yunnan lettuce is 0.98, indicating a high degree of correlation between the two, indicating that there may be some binding relationship between the two. Secondly, the correlation coefficient between screw pepper and baby cabbage is 0.9, indicating that the two are complementary and can be combined. At the same time, the correlation coefficient between screw pepper and millet pepper is -0.9, indicating that the two have a negative correlation. Combined with reality, it can be inferred that there is a certain substitution relationship between the two, and the existence of each other's sales will be affected.

3. The purchase quantity and pricing forecast of each category of vegetables in the next seven days

First of all, this paper analyzes the relationship between the total sales volume of various vegetable categories and cost plus pricing[5], and then determines the impact of sales pricing on the total sales volume. Thus, the average daily sales pricing of various vegetable varieties is obtained, and it is used to replace the cost plus pricing, as shown in the formula:

$$\bar{X} = \frac{\sum_{j=1}^6 \sum_{i=1}^n \frac{B_{ji} Q_{ji}}{1 - \mu}}{\sum_{i=1}^n \frac{Q_{ji}}{1 - \mu}} \quad (2)$$

\bar{X} Average selling pricing; B_{ji} Purchase price; Q_{ji} Sales volume; μ commodity loss rate; j different categories; i different items

This paper considers using a variety of functions to fit the relationship between average sales pricing and total sales of various vegetable varieties, and finally decides to use linear function and logarithmic function to regression the relationship between the two. The result with the best fitting effect is selected as our cost-plus pricing - total sales model. The following Table.1 shows the selected fitting model parameter table for each category of vegetables.

Table.1. fitting model parameters for each vegetable category

Category names	Model name	Model parameter
Aquatic roots	Logarithmic function	[-63.59114255-18.58488965 241.6957458]
Cauliflower	Linear function	[-2.86674326 63.68050329]
Edible mushrooms	Linear function	[-3.09794596 90.42585915]
Flowers and leaves	Logarithmic function	[-22.5485502 2.49503751 184.04972139]
Nightshades	Logarithmic function	[-7.95216963 1.58427143 35.71925612]
capsicums	Logarithmic function	[-17.20757631 3.10649698 105.12660057]

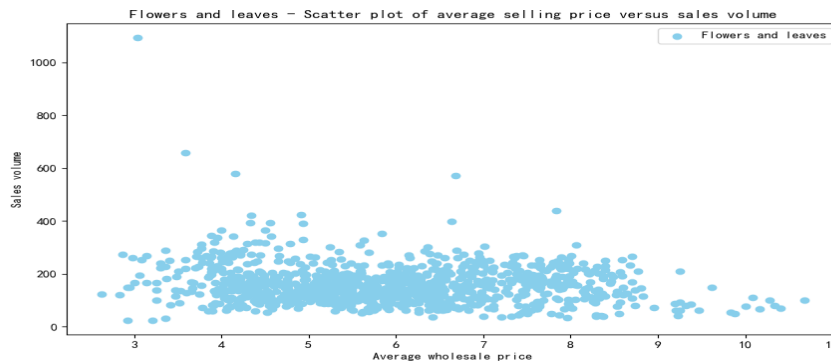


Fig 10. Fitting graph of the relationship between average selling price and sales volume of flowers and leaves

Fig 10 shows the fitting effect of the fitting function[6] selected in this paper on the flower and leaf class. Secondly, factors affecting profits also include the average wholesale price of various types of vegetables, so we also need to forecast the average wholesale price. The daily average wholesale price of vegetables can be calculated by the following formula:

$$\bar{X} = \frac{\sum_{i=1}^n \frac{B_{ji} Q_{ji}}{1 - \mu}}{\sum_{i=1}^n \frac{Q}{1 - \mu}} \tag{3}$$

The chart below shows the change of the average wholesale price of various types of vegetables.

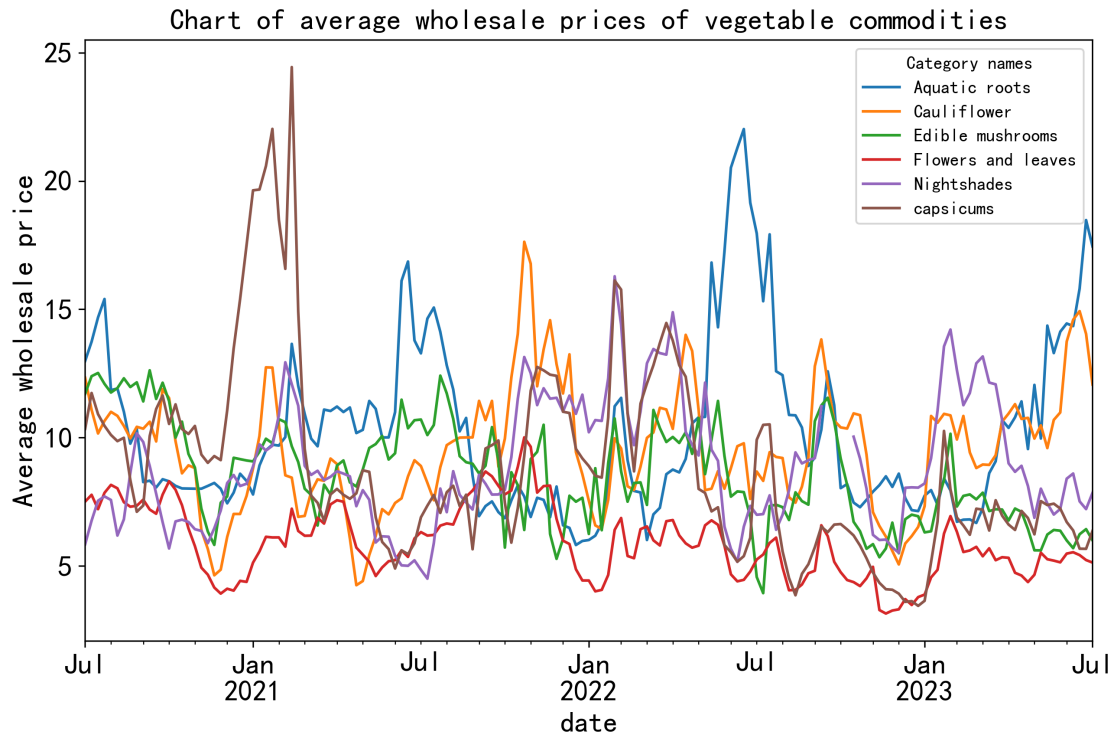


Fig 11. Chart of average wholesale prices of various vegetable commodities

As can be seen from Fig 11, the average wholesale prices of various vegetable commodities will change with time. For example, we can clearly see from the figure that the average wholesale price of chili will increase significantly in winter, which is probably due to the difficult production of chili in winter, which will increase the average wholesale price.

In this paper, ARIMA model[7] is used to forecast the average wholesale price of various vegetables in the next 7 days. The parameters of the ARIMA model are usually expressed as p, d and q, corresponding to the order of autoregression, difference and moving average respectively. Choosing the appropriate p, d and q values is a key step in building an ARIMA model.

In this paper, the Akaike Information Criterion[8] (AIC) of the model is used to select the best model parameter (2,2,2), and forecast the average wholesale price of vegetable categories in the next seven days.

After obtaining the average wholesale price of vegetables and the cost plus pricing -- total sales, a nonlinear programming model[9] is constructed. The model constructed is as follows:

$$W_{d \max} = \sum_{j=1}^m Q_{jd} \cdot S_{jd} - \frac{Q_{jd}}{(1-\mu) \cdot B_{jd}} \tag{4}$$

$$m = 6 \tag{5}$$

$$d = 1,2,3,4,5,6,7 \tag{6}$$

$$Q_{jd} \cdot S_{jd} > \frac{Q_{jd}}{(1-\mu) \cdot B_{jd}} \tag{7}$$

$$Q_j = f(S_j) \tag{8}$$

W Profit; S selling price; d days; m markup rate

The goal of function (4) is to maximize the profit of the supermarket, function (5) is to meet the market demand for six categories, function (6) is to restrict the number of seven days of sales, function

(7) is to limit the total sales price of dishes is greater than the total wholesale price of dishes, and function (8) is the sales-selling price function obtained from the above fitting model[10]

After solving the model group, we can get the decision of the supermarket in the next 7 days as shown in the following table 2:

Table.2. decisions made by market in the next 7days

	Aquatic rhizomes		Flowers and leaves		Cauliflower	
Date	Quantity of purchases	Pricing	Quantity of purchases	Pricing	Quantity of purchases	Pricing
1	11.56316	19.65050	161.06785	9.42027	21.91164	15.75563
2	11.62261	19.61964	161.12388	9.40530	21.92636	15.75130
3	11.52303	19.67134	161.29107	9.36081	21.91781	15.75381
4	11.51596	19.67501	161.09507	9.41299	21.92677	15.75118
5	11.50532	19.68053	161.34379	9.34684	21.91951	15.75332
6	11.50035	19.68312	161.12297	9.40554	21.92770	15.75090
7	11.49554	19.68562	161.37754	9.33791	21.92112	15.75284
	Nightshade		capsicums		Edible mushroom	
Date	Quantity of purchases	Pricing	Quantity of purchases	Pricing	Quantity of purchases	Pricing
1	16.80752	14.00508	76.59791	11.02526	55.12705	13.07586
2	16.81418	13.99538	76.44354	11.09000	56.56407	12.65584
3	16.79942	14.01688	76.60750	11.02125	57.25665	12.45340
4	16.80389	14.01036	76.47033	11.07872	57.52567	12.37477
5	16.80071	14.01500	76.62793	11.01273	57.64622	12.33953
6	16.79904	14.01744	76.49478	11.06845	57.70886	12.32122
7	16.79763	14.01950	76.64893	11.00397	57.74951	12.30934

4. Conclusion

Based on a large number of real and reliable sales data, commodity information and prices, this paper establishes a variety of models, fully considering the market demand and changes. In the process of establishing the model, multiple factors, such as total sales, cost, attrition rate and the impact of demand on sales profit, are comprehensively considered, thus providing a more comprehensive and effective replenishment and pricing decision. In terms of pricing, the cost plus pricing method is adopted, which comprehensively considers the impact of transportation loss and quality decline on pricing, and provides the optimal pricing strategy for supermarket, which is helpful to maintain a certain profit level. In addition, the use of ARIMA model to predict the future wholesale price of dishes, and the introduction of nonlinear programming constraints, is of great help to the supermarket's replenishment and pricing strategy, and can obtain optimized decisions.

References

- [1] Li Chuang. Method and System for Processing historical Performance Data of Storage Devices based on Box Diagram[J], 2018.
- [2] Cheng Juanjuan. Empirical Research on the relationship between scientific research and teaching in universities -- Based on Pearson Correlation Coefficient analysis [J]. Higher Education, 2022.
- [3] Liu Hui, Shao Fubo, Gong Xiang. Comparative Study on Classical Correlation Coefficient and Statistical Efficacy [J]. Computer Software and Computer Applications, 2022.
- [4] Zhao Wei-Zhen, and Lai Yong-quan. Thermal map mapping Method and System based on Geohash Algorithm[J], 2023.
- [5] LIU Xin. Research on Product Selling Price formulation method [J]. Enterprise Economics, 2019.

- [6] Feng Yao, Hou Yanan. Research on Test index estimation based on Fitting function [J]. China Financial Computer, 2019.
- [7] Ma Zhengda. Traffic Prediction based on trend fitting method, smoothing method and ARIMA model [J]. Design Technology of Posts and Telecommunications, 2023.
- [8] Yang ChunFeng et al. Connectivity of intracranial electrical effects based on improved Wiener Granger causal index algorithm based on universal Akaike information content criteria[J]. Journal of Biomedical Engineering ,2018.
- [9] Li XinWei. Application of Genetic Algorithm to nonlinear programming. Electronic Technology and Software Engineering[J],2(2018).
- [10] Zheng Minhui, Wang Huixin. Improve the full cycle value in response to market demand changes [J]. China Communications News,2022.