

Automatic Pricing and Replenishment Decision for Vegetable Products Based on Multiple Linear Regression

Yongyue Yu*, Qingkun Du and Xiaochen Zhang

Department of Intelligent and Information Engineering, Shandong University of Chinese Medicine University, Shandong, China

*Corresponding author: yuyongyue2022@163.com

Abstract. Combining supply-side and demand-side planning with reliable market demand analysis to develop reasonable replenishment decisions and pricing decisions is a challenge for fresh food superstores. In this paper, a model is built based on the sales data of vegetable items to optimize the original sales strategy and develop a sales plan with the maximum gain. Firstly, the original data is pre-integrated to obtain the daily sales volume of each vegetable item, and then the data is analyzed with descriptive statistics. Spearman correlation analysis is performed on the integrated data to make the correlation between each vegetable category. Pre-integrate the data to obtain the average daily selling price, total sales volume, and average cost price of each category of vegetable commodities, and then carry out multiple linear regression to determine the pricing law of each category of vegetable commodities, respectively, and the multiple linear regression equation is the relationship between the total sales volume and the cost-plus pricing. The total sales volume and cost data of the six categories of vegetable commodities are analyzed by ARIMA time series analysis to predict the total sales volume and category cost for the next seven days, which is substituted into the multiple linear regression equation to find the predicted selling price of the vegetable commodities and calculate the profit. Calculate to find the sales price of six types of vegetable goods for the next seven days. Establish an optimization model with the goal of maximizing profit and determine the replenishment strategy for the coming week by considering the daily loss rate of different categories. Based on the optimization model by expanding the constraints to develop a more detailed decision-making scheme. Finally, based on the purpose of improving the replenishment and pricing strategy of vegetable products and trying to meet the market demand to maximize the profit of the supermarket, we provide optimization advice for the data collection of the supermarket from multiple angles, determine more accurate replenishment quantity, cost price and selling price of individual products for the supermarket, and maximize the profit margin of the supermarket, as well as explaining the reasonableness of the data collection for each type of data.

Keywords: multiple linear regression ; Pricing replenishment decisions; ARIMA.

1. Introduction

In fresh store supermarkets, vegetable items have a shorter shelf life and their quality deteriorates as they are sold for longer periods of time, most of which cannot be re-sold the next day. Therefore, supermarkets usually replenish their stocks on a daily basis according to the historical sales and demand of each commodity. Due to the variety of vegetable varieties sold in supermarkets, and the purchase transaction time is usually in the early morning, so the replenishment decision of each vegetable category in the supermarket on the same day needs to be made in the specific single product and the purchase price of information gaps exist in the case. Vegetables are generally priced using the “cost-plus pricing” method, and supermarkets usually sell at a discount for goods with shipping losses and poor quality. From the demand side, there is a correlation between the sales volume of vegetables and the time of day; from the supply side, vegetables are more abundant from April to October, and due to the limitation of sales space, it is a challenge for fresh food superstores to plan a reliable analysis of market demand to make replenishment and pricing decisions.

2. Model building and solving

2.1. Correlation analysis

In order to study and analyze the distribution pattern of the sales volume of each category and single product of vegetables first pre-consolidate the collected data to get the daily sales volume of each vegetable single product. The corresponding number and sales of each vegetable single product were counted and organized to get the table of daily sales volume of each vegetable single product from July 1, 2020 to June 30, 2023, as shown in Table 1 below.

Table 1. Three Scheme comparing

Item number	The name of the item	Classification name	1	2	...	1095
102900005115762	amaranth	Mosaic and leafy	6.841	8.112	...	6.726
102900005115779	Yunnan lettuce	Mosaic and leafy	41.966	44.873	...	0
102900005115786	Takekana	Mosaic and leafy	11.352	11.556	...	12.884
102900005119944	Seafood mushrooms	Edible fungus	0.783	2.254	...	0
102900005125815	king oyster mushroom	Edible fungus	11.888	10.164	...	0

In order to clearly and intuitively observe the correlation between different categories or different single products of vegetable goods, based on the integrated data to make six classifications of vegetable single products and the enumerated two vegetable single products of the folding statistical charts respectively. The horizontal coordinate is the date in the hypothesis, and the vertical coordinate is the sales volume of the single product of the category on the same day. The statistical graphs are shown in Figures 1, 2 as well as 3.

As shown in Figure 1, the daily sales of cauliflower, edibles, foliage, chili peppers, eggplant, and aquatic roots and tubers in order can be seen that the peak of daily sales of cauliflower commodities is concentrated in February and August each year, and the rest of the time is relatively stable; the peak of sales of foliage commodities is concentrated in January and December each year, and the rest of the time is relatively stable; the peak of sales of chili peppers is concentrated in February and November each year, and the sales have been stable since August 2022 onwards, with sales generally rising and remaining flat for the rest of the year. The peaks of sales of eggplant commodities were concentrated in February and July of each year, and the trend of sales for the rest of the year was distributed in a W-shape; the peaks of daily sales of aquatic root commodities were concentrated in January and July of each year, and the trend of sales for the rest of the year was distributed in a W-shape. As shown in Figure 2, the peak daily sales volume of amaranth is concentrated in February to April, and the sales volume is zero from October to December, and as shown in Figure 3, the peak daily sales volume of sweet cabbage is concentrated in October to December.

Based on this, the analysis shows that the harvesting season of vegetable commodities sold in this superstore is mostly concentrated in August every year, and the supply of vegetable commodities is sufficient in this period, so the sales volume is higher; from January to February coincides with the Chinese New Year, and the consumer's consumption demand and catering demand are increased, so the sales volume of vegetable commodities is higher during this period; the daily sales volume of vegetable commodities tends to be stable in the rest of the time, and this period is not the maturity season of most vegetables, but because the greenhouse can allow the mass production of vegetables and fruits regardless of seasonal restrictions, it can provide a stable supply of market demand. Season, but because the greenhouse greenhouse can make vegetables and fruits can not be limited by the season of mass production, can be a stable supply of market demand, so the daily sales of vegetable commodities in this period tends to be stable; and Figure 2, Figure 3 shows amaranth, sweet cabbage in a few months there is a gap in the sales of this type of vegetables in several months is due to this type of vegetables itself is a lower sales price, the use of greenhouse greenhouse cultivation of lower profits, and this type of vegetables themselves have the characteristic of easy storage, consumers can

buy them in the seasonal season and then store them by themselves, so the above two kinds of vegetables can only be sold seasonally.

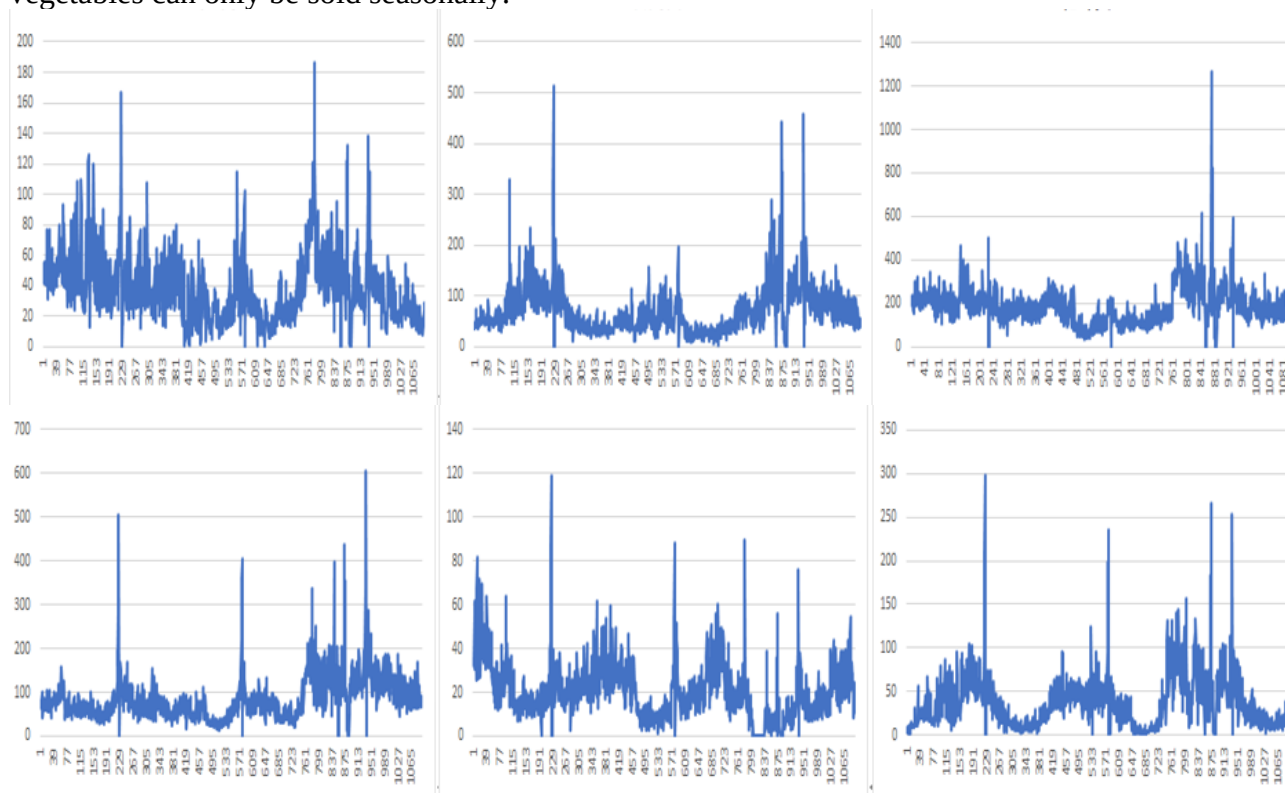


Fig. 1 Daily Sales Statistics of Six Types of Vegetable Commodities

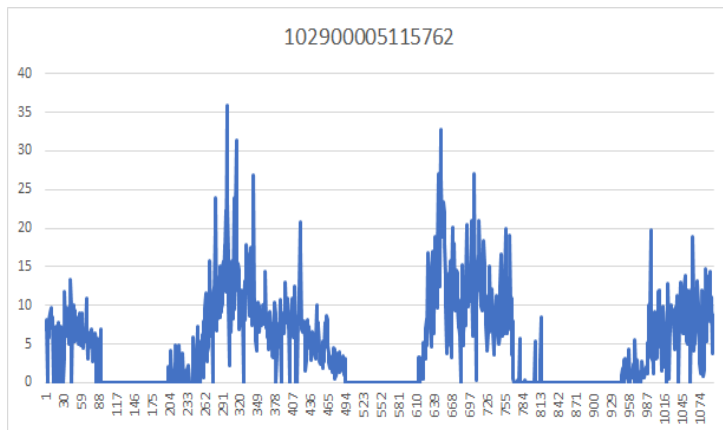


Fig. 2 Amaranth Daily Sales Statistics

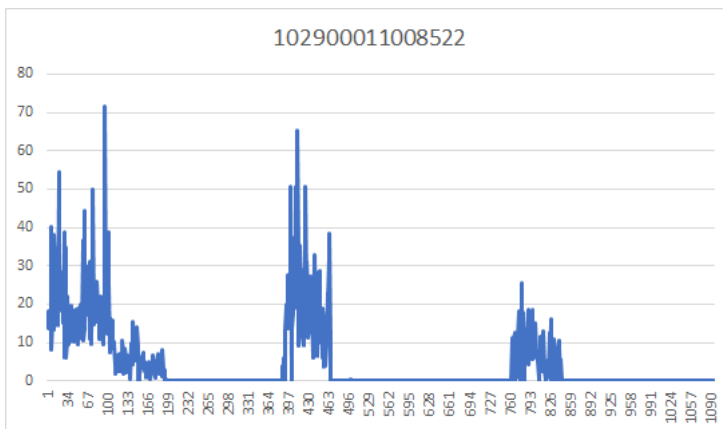


Fig. 3 Daily Sales Chart of Sweet Cabbage

To summarize, it can be obtained that the sales of vegetable commodities are higher in January to February and August every year; the daily sales of vegetable commodities tend to be stable or change regularly at times other than the peak; the sales of vegetable commodities are affected by the harvest season and holiday factors; the sales of vegetable commodities are affected by whether the vegetables themselves are easy to store or not.

Spearman correlation analysis is performed on the categories of vegetable commodities, and the Spearman correlation analysis is calculated as follows.

$$\rho = \frac{\frac{1}{n} \sum_{i=1}^n (R(x_i) - \overline{R(x)}) \cdot (R(y_i) - \overline{R(y)})}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n (R(x_i) - \overline{R(x)})^2 \right) \cdot \left(\frac{1}{n} \sum_{i=1}^n (R(y_i) - \overline{R(y)})^2 \right)}} \quad (1)$$

In this paper, only the heat map of correlation coefficient of daily sales of “cauliflower” and “foliage” are selected to show, and the heat map of correlation coefficient is shown in Fig. 4, Fig. 5 and Fig. 6.

From the results of Spearman's correlation analysis, Figure 4-Figure 6, it can be seen that the correlation between the daily sales of each single product of foliage, chili, eggplant, and edible fungus commodities is high; the correlation between the daily sales of each single product of cauliflower commodities is low; and the correlation between the daily sales of each single product of aquatic root and tuber commodities is average.

In summary, it can be concluded that the natural output frequency of cauliflower, pepper, eggplant, and edible fungus commodities is high, storage resistance is average, and the difficulty of greenhouse cultivation is low, so the supply chain is stable, and therefore the consumption frequency and single consumption variability are small. The aquatic roots and tubers sold by the supermarket are grown in water, which makes greenhouse cultivation more difficult, and the supply chain is mainly based on natural cultivation. As there are differences in the frequency of output and storage resistance of each product, there are also differences in the frequency of consumption and single consumption.

Among the cauliflower commodities sold in the supermarket, purple cabbage is easy to be stored, so consumers can purchase it in seasonal seasons and then store it for a long period of time, with a lower consumption frequency and a higher single consumption; broccoli, green stalks, loose flowers and other commodities with a warm and humid climate, with the production areas mainly located in the south, whose supply chain includes natural output and greenhouse cultivation, and whose storage resistance is general, so the consumption frequency is higher and the single consumption is lower. Consumption is low.

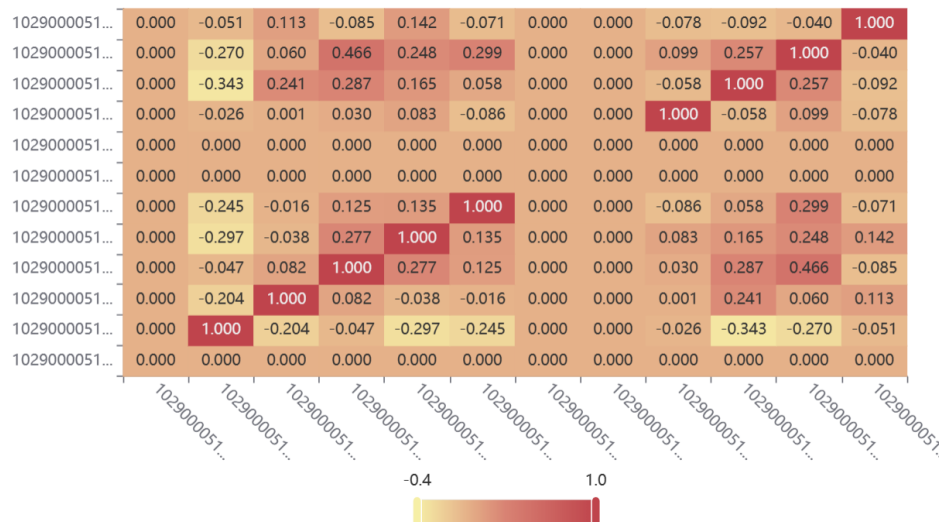


Fig. 4 Heat map of Spearman's correlation coefficient for daily sales of each individual item in the flower and foliage category

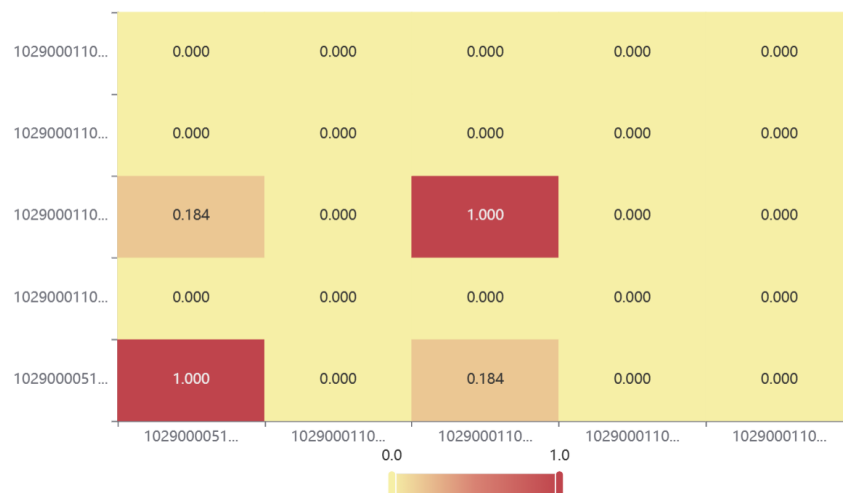


Fig. 5 Heat map of daily sales of cauliflower products by single item

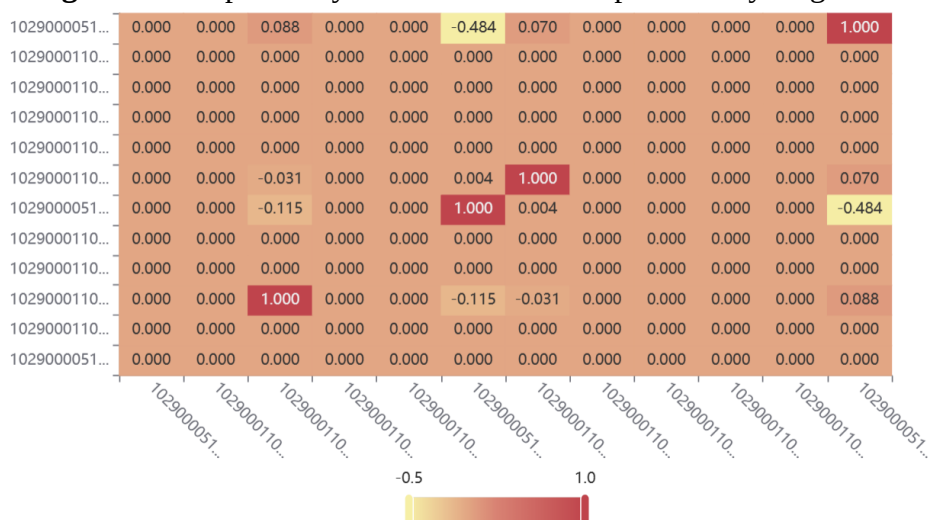


Fig. 6 Aquatic Roots & Tubers daily sales heat map by item

2.2. Regression model

Data integration, pre-statistics for each category of the daily wholesale price and pricing, respectively, to find the total unit price per day and the number of sales on the first day of each category of vegetable commodities, with each category of vegetable commodities per day of the total unit price divided by the number of sales on each day, to get the average unit price per day of each category of vegetable commodities; with each category of vegetable commodities per day of the total cost of the price of the daily divided by the number of sales on a daily basis, to find the vegetable commodities for each category of cost price per day for each type of vegetable commodity. Some of the data on the average selling price, total sales volume, and average cost price for each type of vegetable commodity are shown in Table 2.

Table 2. The average daily selling price, total sales volume, and average cost price

Days	Average selling price	Total sales volume	Average cost price
1	11.86667	35.365	7.894
2	13.63875	48.51	7.836
3	14.32349	42.442	7.824
4	14.09123	47.262	7.984
...	...	...	...
1093	9.529167	53.742	4.748888889
1094	9.854545	48.314	4.837777778
1095	13.98696	39.572	4.623

Let the average selling price be y , the total sales volume be x_1 , and the average cost price be x_2 . Multiple linear regressions are performed on the average selling price, total sales volume, and average cost price to determine the pricing law for each category of goods, and the resulting multiple linear regression equations are the relationship between the total sales volume and cost-plus pricing. The multiple linear regression equations are shown in Table 3:

Table 3. the relationship between total sales volume and cost-plus pricing

Category name	Multiple linear regression equations
Mosaic and leafy	$y = 3.409714 - 0.00233a_1 + 0.729134a_2$
Cauliflower	$y = 2.652122 - 0.00822a_1 + 1.240567a_2$
Aquatic rhizomes	$y = 3.272124 - 0.01036a_1 + 0.955214a_2$
Nightshades	$y = 1.988957 - 0.00026a_1 + 1.197951a_2$
Chili peppers	$y = 1.299738 - 0.005888a_1 + 1.083158a_2$
Edible fungus	$y = 7.625866 - 0.01538a_1 + 0.919447a_2$

2.3. ARIMA and Optimization Models

ARIMA time series analysis of total sales and cost of six categories of vegetable commodities were conducted separately, and time series plots and time series prediction tables of total sales and category cost were derived for each category of vegetable commodities respectively, and the time series prediction tables obtained from the analysis were integrated.

The category cost is the weighted sum of the costs of all individual items in the category, with the weight being the sales of the individual items divided by the total sales of the category. Based on the total sales volume and predicted category cost derived from the ARIMA time series analysis prediction, substituting into the multiple linear regression equations in Table 3 for calculation, the sales price of the six types of vegetable goods for the next seven days (July 1-7, 2023) is found.

An optimization model is established, and the decision variable is x_{ij} , which indicates that the i th vegetable category ordered x_{ij} quantities on the j th day; the objective function is W ; and the constraints are to be satisfied with the sales volume of each day (taking into account the rate of wastage).

$$\begin{aligned} \max W &= \sum_i \sum_j (y_{ij} * m_{ij} - x_{ij} * N_{ij}) \\ \text{st. } \sum_j x_{ij} * (1 - a_{ij}) &\geq m_{ij} \end{aligned} \quad (2)$$

Through the optimization model, the total replenishment of the six types of vegetable goods in the coming week, together with the predicted sales, costs, selling prices, here only lists the relevant results of the foliage category, as shown in Table 4.

Table 4. Table of predicted results for foliar species

Time	Forecast sales	Forecast the cost price	Predicted selling price
1	131.2365368	3.686268343	5.791716451
2	128.482718	3.675140431	5.79001911
3	136.9198245	3.677805327	5.772303719
4	143.7668317	3.674796176	5.754156117
5	143.1832322	3.674121199	5.755023755
6	138.8227195	3.672485995	5.763991467
7	136.345668	3.671245807	5.768858734

3. Literature References

In recent years, fresh food superstores are facing increasingly severe challenges as consumer demand for fresh food increases and market competition intensifies. Among them, how to effectively make replenishment and pricing decisions to maximize profits has become a key issue in fresh food superstore management. To this end, many scholars and researchers have conducted extensive studies. Liu and Liu, in their study, proposed a merchandise pricing and replenishment decision-making strategy based on a seasonal ARIMA model. The study emphasized the important influence of seasonal factors on merchandising and predicted future sales trends through ARIMA model, thus providing strong support for replenishment and pricing decisions [1]. Fu, Li and Wan, on the other hand, optimized the replenishment and sales strategies for vegetable commodities by using multivariate linear regression and nonlinear programming methods. They analyzed the various factors affecting the sales volume of vegetables through multiple linear regression model and developed a nonlinear planning model to develop optimal replenishment and sales strategies [2]. Wang also studied the pricing and replenishment decisions of vegetable commodities based on ARIMA model. He predicted the future sales trends and cost changes through ARIMA model, and formulated the pricing and replenishment strategies by taking into account the market situation and the actual situation of the superstore[3].

4. Summary

The model can be extended to all urban superstores to forecast the sales volume, cost price, selling price and the choice of stocking items from the next few days to the next few years, so as to better grasp the market changes, seize the influence of the season and other factors, and firmly grasp the fluctuation of market information to maximize the profit. This paper can process a large amount of data in a short period of time to establish an optimization model for pricing and replenishment. However, the optimization model is prone to higher error rates for remote areas or areas with harsh natural environments.

References

- [1] Liu J, Liu B. Commodity Pricing and Replenishment Decision Strategy Based on the Seasonal ARIMA Model[J]. Mathematics, 2023, 11(24): 4921.
- [2] Fu X, Li X, Wang W. Optimization Of Replenishment and Sales Strategies of Vegetable Commodities Based on Multiple Linear Regression and Non-Linear Programming[J]. Highlights in Business, Economics and Management, 2024, 25: 211-227.
- [3] Wang Z. An ARIMA-based study of pricing and replenishment decisions for vegetable commodities[J]. Highlights in Business, Economics and Management, 2023, 21: 1020-1024.