

# Research on automatic pricing and replenishment strategy of vegetable commodities based on ARIMA algorithm

Zhang'ang Li\*, Tang Li and Kaili Zhang

Institute for Innovation in Information Technology Applications, Yuncheng Vocational and Technical University, Yuncheng, China

\*Corresponding author: 1129864580@qq.com

**Abstract.** In just a few years' time, supermarkets have transformed from a novelty into a mainstream channel that unifies the retail world. Since superstores sell vegetables with a wide variety of types, different origins, and shorter freshness periods, the incoming transaction time of vegetables is often in the early morning, so superstores usually replenish their stocks on a daily basis according to the historical sales and demand of each commodity. First, this paper visualizes the historical data of vegetable commodity purchasing and pricing in superstores in order to observe the distribution characteristics and trends, and then analyzes the correlation between categories or individual products using statistical methods such as correlation factors and covariance. To predict replenishment and pricing strategies for the coming week, this paper analyzes the relationship between overall sales and cost-plus pricing for each vegetable category, using given sales data to establish the relationship between sales volume and price for each category. Based on the ARIMA predicted sales model and given costs, the optimal replenishment quantities and pricing strategies were determined using an optimization algorithm.

**Keywords:** ARIMA; vegetable commodities; correlation analysis; replenishment strategy.

## 1. Introduction

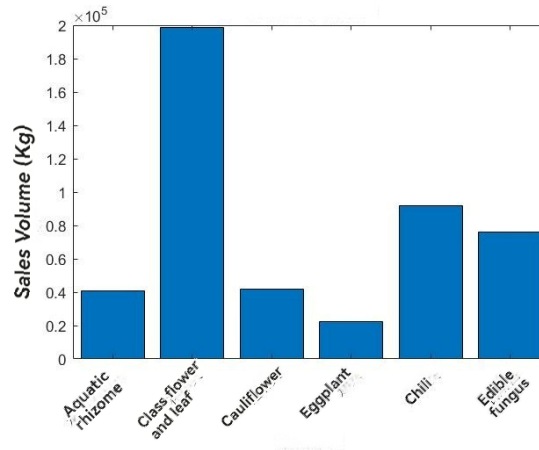
With the development of the retail industry, the competition among superstores has become more and more intense. In order to gain an edge in the competition, supermarkets need to formulate precise pricing strategies to satisfy consumers' demand for price reasonableness, and consumers' expectations for commodity prices and quality are constantly rising. Through automatic pricing, hypermarkets can better adjust prices according to market demand and competitive conditions to attract consumers and maximize revenue. Vegetable products have a shorter shelf life, and the degree of decay will accelerate with the growth of sales time, the combination of automatic pricing and replenishment decisions can effectively help supermarkets accurately predict the market demand, reduce losses, and ensure that sales commodities are always maintained in the best sales condition. Supermarkets replenish goods on the following day in response to the historical sales and demand for goods, without knowing the specific single product and purchase price, make replenishment decisions for each vegetable category on that day, give replenishment amount as well as pricing strategy, on the basis of which to achieve the maximization of the revenue of the superstore [1].

With the improvement of people's living standards, consumers' requirements on the quality of goods, taste and other aspects are also increasing, and superstores need to make measures to cope with this. Through the comprehensive analysis of historical sales data, real-time inventory data, market demand, seasonal factors, purchasing costs and other diversified factors, we can provide supermarkets with more accurate and real-time pricing and replenishment decisions, and improve the operational efficiency of supermarkets and customer satisfaction [2].

## 2. Correlation between vegetable categories

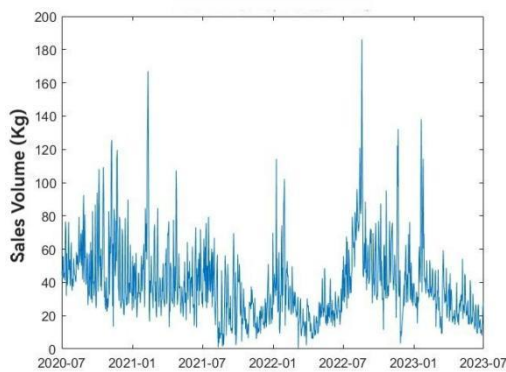
### 2.1. Data visualization

Based on the statistical sales volume of each vegetable category, it was visualized using MATLAB software, and the results are shown in Figure 1: the flower and leaf category has the highest sales volume, followed by the pepper category and edibles, and the eggplant category has the least.

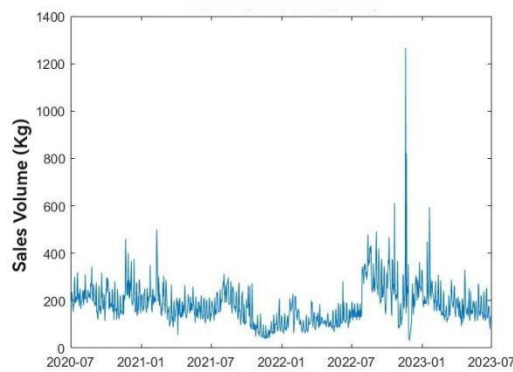


**Fig. 1** Sales Volume Comparison of 6 Vegetable Categories

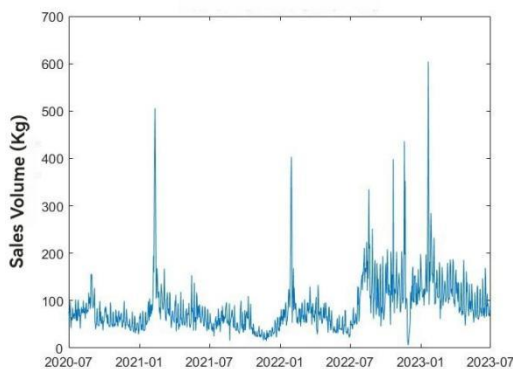
The distribution pattern of sales volume by month for each vegetable category is shown in Figures 2-7. It was concluded that cauliflower had the highest sales volume in September 2022, leafy cauliflower had the highest sales volume in November 2022, chili peppers had the highest sales volume in February 2023, eggplant had the highest sales volume in February 2021, edible mushrooms had the highest sales volume in March 2021, and aquatic roots and tubers had the highest sales volume in February 2021..



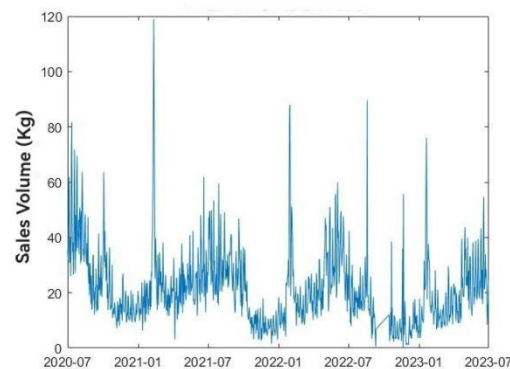
**Fig. 2** Trends in monthly sales of cauliflower



**Fig. 3** Trends in monthly sales of foliage



**Fig. 4** Monthly Sales Trend of Chili Peppers



**Fig. 5** Monthly sales volume trend of eggplant

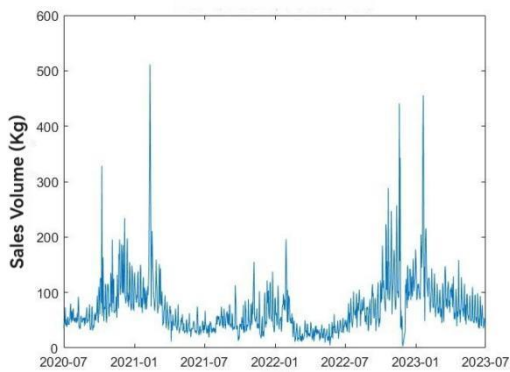


Fig. 6 Monthly Sales Trend of edible mushrooms

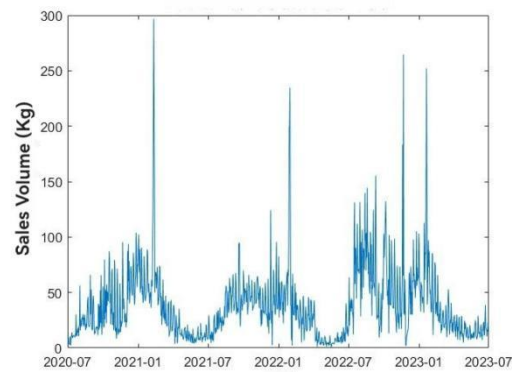


Fig. 7 Monthly Sales Trend of aquatic roots and tubers

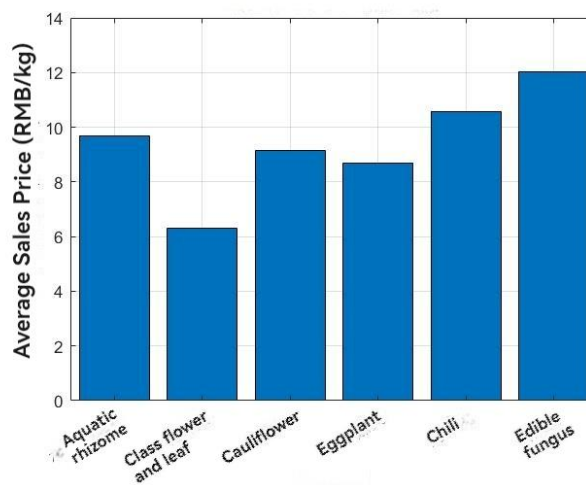


Fig. 8 Average selling price of each vegetable category

As shown in Figure 8, the average sales price of each vegetable category was tallied: edibles were the highest and foliage the lowest.

## 2.2. Correlation analysis of vegetable categories

The resulting inter-category sales correlation is shown in Figure 9 below [3]:

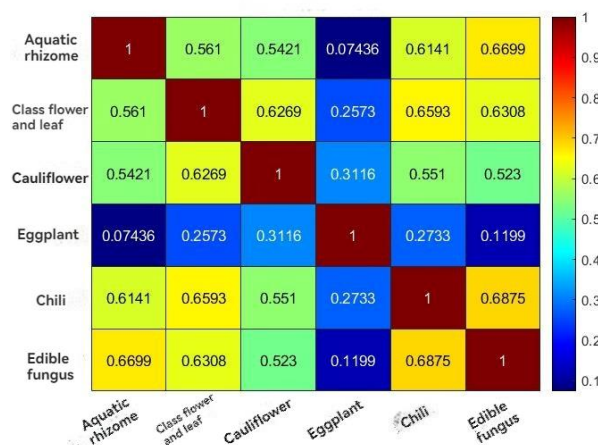


Fig. 9 Correlation of sales volume between categories

The correlation between the categories: chili peppers and edible mushrooms had the highest positive correlation of 0.6875, while eggplant and aquatic roots and tubers had the lowest correlation of 0.07436.

### 3. A study of total replenishment forecasting and pricing decisions in the vegetable category

#### 3.1. Relationship between total sales volume and average pricing

Suppose demand is linear and negatively correlated with price. Then:

$$Q_i = a - bP_i \tag{1}$$

Where  $a$  and  $b$  are derived from a linear programming equation that plots total sales against average pricing for the vegetable category.

The revenue function is [4]:

$$W_i = P_i \times Q_i \times (1 - M_i) - C_i \times Q_i \tag{2}$$

Our goal is to maximize total revenue across all vegetable categories:

$$\max \sum_{i=1}^6 W_i \tag{3}$$

The values of  $a$  and  $b$  are derived by plotting a linear regression equation based on a graph of total sales versus average pricing for the vegetable category (shown below). The model can be solved using optimization techniques such as linear or non-linear programming.

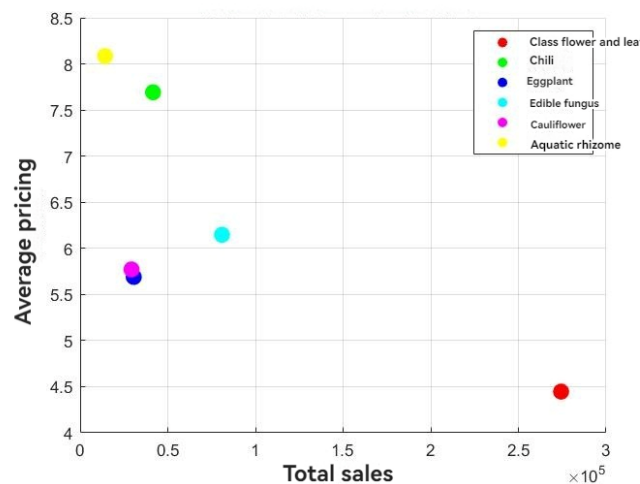
Price constraints: Prices cannot be set too low or too high due to practical considerations.

$$0.5C_i \leq P_i \leq 2C_i \tag{4}$$

Sales volume constraint: the sales volume cannot be negative.

$$Q_i \geq 0 \tag{5}$$

The relationship between total sales and average pricing for the vegetable category is shown in Figure 10.



**Fig. 10** Relationship between total sales and average pricing in the vegetable category

#### 3.2. Total daily replenishment forecast

Using time series analysis: For each vegetable category, a time series analysis method, such as the Autoregressive Moving Average Model (ARIMA), was used to predict the total daily replenishment for the coming week.

The ARIMA model uses a difference method to convert non-stationary time series data into stationary time series data, and then predicts future data based on the stationarity of the time series data and the model parameters. The ARIMA model consists of three parts: the autoregressive process (AR), the differencing process (I), and the moving average process (MA). The differencing process for time series data is as follows.

$$\Delta x_t = x_t - x_{t-1} = x_t - Lx_t = (1 - L)x_t \tag{6}$$

$$\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = (1 - L)^2 x_t \tag{7}$$

$$w_t = \Delta^d x_t = (1 - L)^d x_t \tag{8}$$

Where  $x_t$  represents the original time series,  $w_t$  is the smooth series obtained after d-th order differencing,  $L$  is the lag operator, and  $t$  represents the moment. In this case, the ARIMA model can be written in the form of equation (9)

$$w_t = \delta + \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q} \tag{9}$$

Where  $u_t$  represents the white noise sequence and  $\delta$  is a constant term. In this model,  $p$  is the order of the autoregressive part and  $q$  is the order of the moving average part. The above model is denoted as ARIMA (p, d, q), and Eq. (9) can be rewritten as Eqs. (10) to (12) using the lag operator polynomials  $\Phi_p(L)$  and  $\Theta_q(L)$ .

$$\Phi_p(L)w_t = \delta + \Theta_q(L)u_t \tag{10}$$

$$\Phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p \tag{11}$$

$$\Theta_q(L) = 1 - \theta_1 L - \dots - \theta_q L^q \tag{12}$$

The ARIMA model characterizes the relationship between the current state and the historical state of the time series process, in which the p-order autoregressive process characterizes the lagging effect of the state itself and its weight, and the q-order moving average process characterizes the effect of the current and historical noises in the time series process and its weight.

According to the basic principle of ARIMA model, the non-stationary time series data can be converted into stationary time series data by difference operation, and then according to the characteristics of stationarity, the model suitable for expressing the operating characteristics of the system can be obtained by refining and analyzing the historical data, based on which the future operating state of the system can be effectively predicted [5].

The predicted total daily replenishment and pricing strategies for each vegetable category in the coming week (July 1-7, 2023) are as shown in Table 1:

**Table 1.** Three Scheme comparing

Vegetable category	Pricing	Daily replenishment
Cauliflower	8.52	10.13
Philodendron	10.27	100.52
Capsicum	9.18	50.95
Eggplant	9.41	18.74
Edible mushroom	9.14	44.06
Aquatic rhizomes	10.43	20.30

Based on the results, it can be seen that aquatic roots and tubers are priced the highest and have the third lowest daily replenishment; flowers and leaves are priced second and have the highest daily replenishment; and cauliflower is priced the lowest and has the lowest daily replenishment.

#### 4. Summary

This paper predicts the replenishment volume of each category in the coming week based on the historical sales data of vegetable goods in supermarkets and gives the pricing strategy. In the automatic pricing and replenishment decision-making process of vegetable commodities, the decision support system can help companies quickly develop the best pricing and replenishment strategies. The system can provide enterprises with scientific and reasonable suggestions based on data on

market demand, supply costs, seasonal factors, bidding strategies, replenishment strategies, etc., and improve decision-making efficiency and accuracy

In order to improve the accuracy of the model and make better replenishment and pricing decisions for vegetable commodities, the superstore can also collect data on consumer preferences, seasonal and seasonal factors, weather forecasts, inventory management, and promotions, etc., to fully consider the risks posed by price fluctuations, changes in market demand, and disruptions in the supply chain, and to formulate appropriate risk management measures.

## References

- [1] LU Xingyu, MON Yu, DONG Ming, et al. Research on automatic pricing and replenishment decision of vegetable goods[J]. Consumption Guide,2024(5):39-49.
- [2] Lu Yajie. Research on dynamic pricing of high quality fresh vegetables in China's supermarkets[D]. Beijing:Beijing Jiaotong University,2010.
- [3] Aihaiti Keranmu,Mukhtar Zaje,Wu Zhengbao. Correlation analysis of appearance quality indexes of 'gray jujube' fruit[J]. Zhejiang Forestry Science and Technology,2023,43(1):89-93.
- [4] Zhang Ping. Research on cost-plus loan pricing model based on customer contribution adjustment[D]. Beijing:University of International Business and Economics,2015.
- [5] Zhen Zhang,Wei Hu,Weige Zheng,et al. Analysis and prediction method of oxygen partial pressure of manned spacecraft based on ARIMA model[J]. Manned Spaceflight,2023,29(5):653-658.