**Journal of Computing and Electronic Information Management**
ISSN: 2413-1660 | Vol. 13, No. 2, 2024

# Research on Vegetable Pricing and Replenishment Strategies Based on Time Series and Linear Programming Models

**RuiBing Wang *, ZhiYi Niu, HuiKang Gao**

Department of Big Data and Intelligent Engineering, Shanxi Institute of Technology, Yangquan, China
* **Corresponding author**: RuiBing Wang (Email: 281286500@qq.com)

**Abstract:** In this paper, an optimization model for pricing and replenishment of vegetable commodities was established based on historical sales data, wholesale price and wastage rate using linear regression, time series forecasting and linear programming. The model is analyzed and solved by Python, MATLAB, SPSS and other software, and the optimization strategy for commodity pricing and replenishment is proposed. First, bar charts and line graphs of sales volume changes were plotted using the matplotlib library in Python, and heat maps of correlation coefficients were plotted by the Pearson correlation coefficient method. Secondly, the relationship between sales volume and pricing is analyzed by scatter plot and linear regression model, and the time series model is constructed by using SPSS for sales volume forecasting. Finally, the restocking volume and pricing strategy for the vegetable category in the coming week are given based on the linear programming model to maximize the superstore revenue.

**Keywords:** Time Series Model; Linear Programming; Linear Regression Model.

## 1. Introduction

In today's competitive market environment, merchants face many uncertainties when making replenishment and pricing decisions in the vegetable category [1]. The purpose of this paper is to establish an optimization model for pricing and replenishment of vegetables by using past sales data, wholesale prices and wastage rates, and combining mathematical methods such as linear regression, time series forecasting and linear programming to provide merchants with scientific decision support. First, the correlation between different vegetable categories is revealed by analyzing historical sales data and drawing bar charts and line graphs of sales volume changes, as well as correlation coefficient heat maps. Second, the relationship between sales volume and pricing was explored through scatter plots and linear regression models, and a time series model was constructed using SPSS to forecast future sales volume. Finally, the linear programming model, combined with specific data and constraints, provides the restocking volume and pricing strategy of vegetable categories for the superstore in the coming week to maximize the superstore's revenue. The comprehensive application of the methods in this paper will help merchants to respond more effectively to market demand fluctuations and improve operational efficiency and profitability.

## 2. Correlation analysis

### 2.1. Data preprocessing

The pre-processing of the data consists of the following parts.

First of all, we do outlier detection and processing on the given data, we use PYTHON3 to find out if there are any missing values. We use PYTHON3 to find out whether there is any missing value. After searching, we conclude that there is no missing value in the data. Therefore, in data preprocessing, we only need to detect and process outliers.

Data standardization, generally the data is scaled to zero mean and variance of the interval, to ensure that the data of multiple features of the dimensionless.

According to the field single product code, merge, get each category and single product sales data.

### 2.2. Data visualization

Distribution pattern of sales volume of vegetable products by category, as shown in Figure 1.

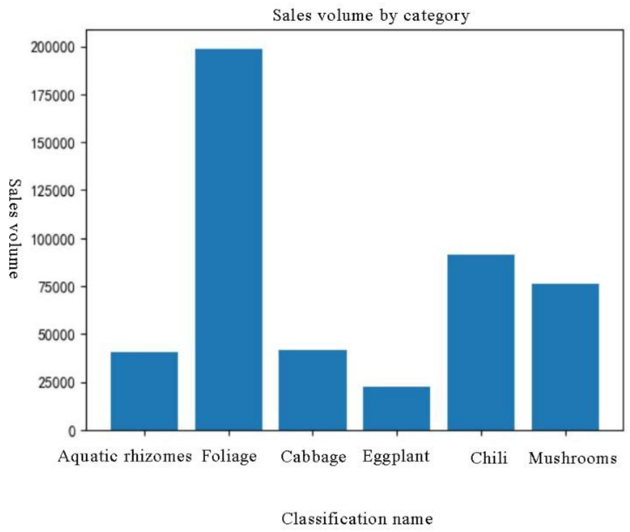


**Figure 1**. Bar chart of sales volume by category

In order to observe more intuitively the distribution of the sales volume of different categories of vegetable products, we use python's matplotlib library to draw a line graph, as shown in Figure 2 below.

The distribution pattern between different items, under the premise of huge data dimensions, we take the top 20 selling items for illustration and comparison, as shown in Figure 3.

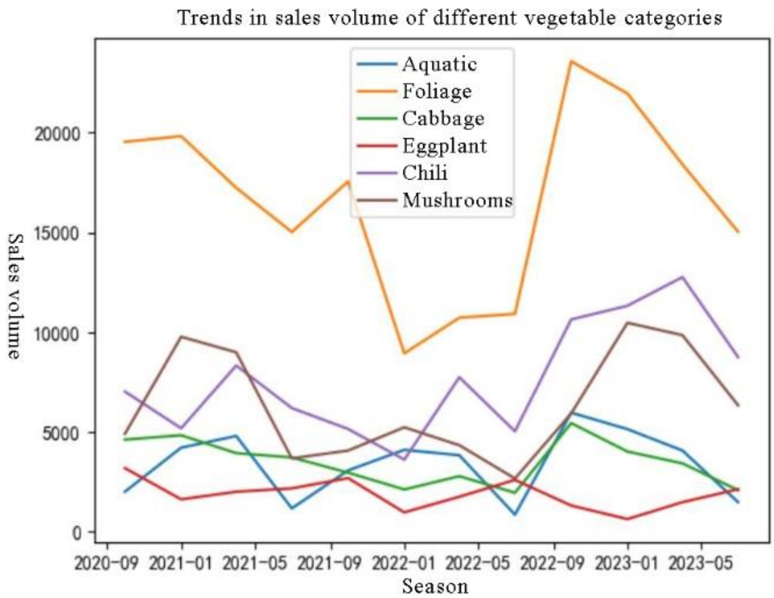


**Figure 2.** Schematic diagram of the sales volume of different vegetable categories

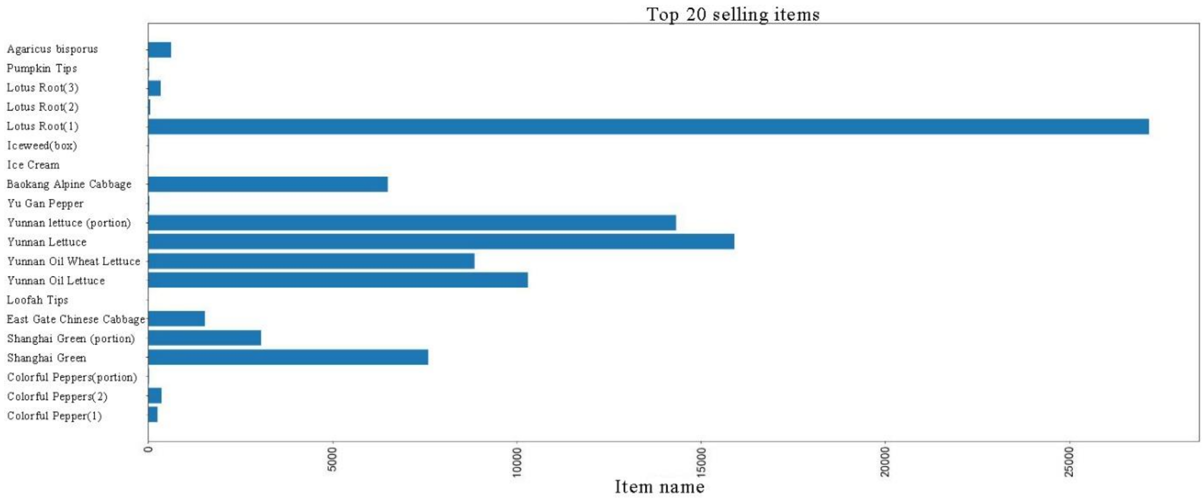


Figure 3. Distribution pattern among the top 20 selling items

## 2.3. Correlation analysis

The data were processed and tested for correlation using SPSS. Here, we analyze the correlation by introducing Pearson correlation coefficient [2].

Sample covariance:

$$Cov(x,y) = \frac{\sum_{i=1}^{n}(X_i - X)^2}{n-1} \tag{1}$$

Sample standard deviation:

$$\sigma(r) = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}} \tag{2}$$

Pearson's correlation coefficient:

$$r(x,y) = \frac{\mathrm{cov}(x,y)}{\sqrt{\sigma(x)}\sqrt{\sigma(y)}} \tag{3}$$

The results of the correlation analysis of the sales volume of different categories of vegetable goods are shown in Figure 4 below.

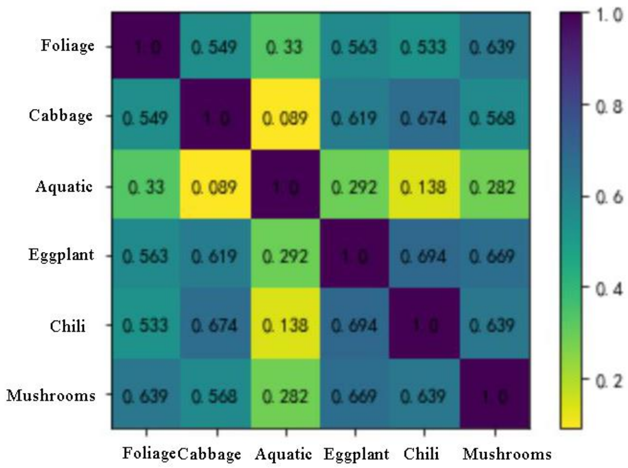


**Figure 4.** Heat map of correlation coefficients of different categories of vegetable goods

# 3. Replenishment and pricing strategies

## 3.1. Relationship between sales volume and cost

We use sales volume and pricing as two variables. By plotting a scatter plot of the mean of sales volume and pricing for each day, as shown in Figure 5, it helps us to observe and analyze the relationship between sales volume and pricing, as well as possible trends or patterns.

This relationship can then be analyzed and quantified even further using a fitted linear regression model, as exemplified below for aquatic rhizomes only [3].

Linear regression (least squares) analysis results are shown in Table.1, under linear regression analysis: Significance P-value for F-test is $0.000^{***}$, showing significance at the level.

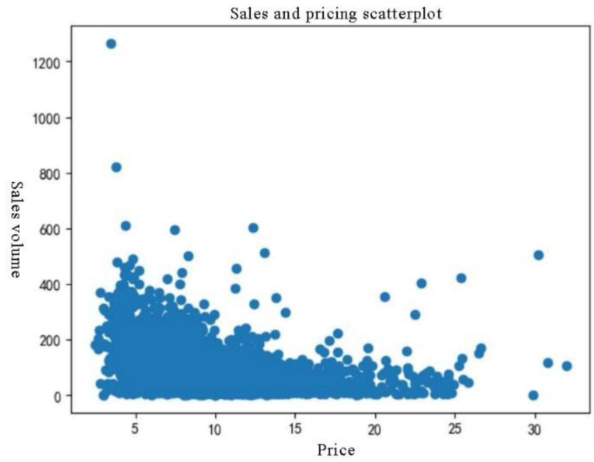


**Figure 5.** Scatter plot of sales and pricing

**Table 1.** Results of linear regression analysis

| Linear regression analysis results n=1085 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Non-std coefficient | Std coefficient | | $t$ | $P$ | VIF | $R^2$ | Align $R^2$ | $F$ |
| | B | Standard error | Beta | | | | | | |
| Constant | 63.663 | 2.656 | - | 23.973 | 0.000*** | - | 0.093 | 0.092 | $F$=110.707 |
| Price | -2.532 | 0.241 | -0.305 | -10.522 | 0.000*** | 1 | | | $P$=0.000*** |
| Dependent variable: sales volume (kg) | | | | | | | | | |

## 3.2. Time series modeling for forecasting vegetables by category

Time series analysis (ARIMA) is based on historical data to forecast future periods [4,5], the model's goodness-of-fit $R^2$ is 0.764 and the model performs well. The forecast results for the next 7 periods are shown in Figure 6 and Figure 7, Table.2 and Table.3.

**Table.2** Time series sales forecast table

| Order (time) | Predicted results |
|---|---|
| 1 | 20.13586414894382 |
| 2 | 21.081809125932146 |
| 3 | 21.466163205272245 |
| 4 | 21.840572671006864 |
| 5 | 22.205294825829213 |
| 6 | 22.56058031509231 |
| 7 | 22.90667329905815 |

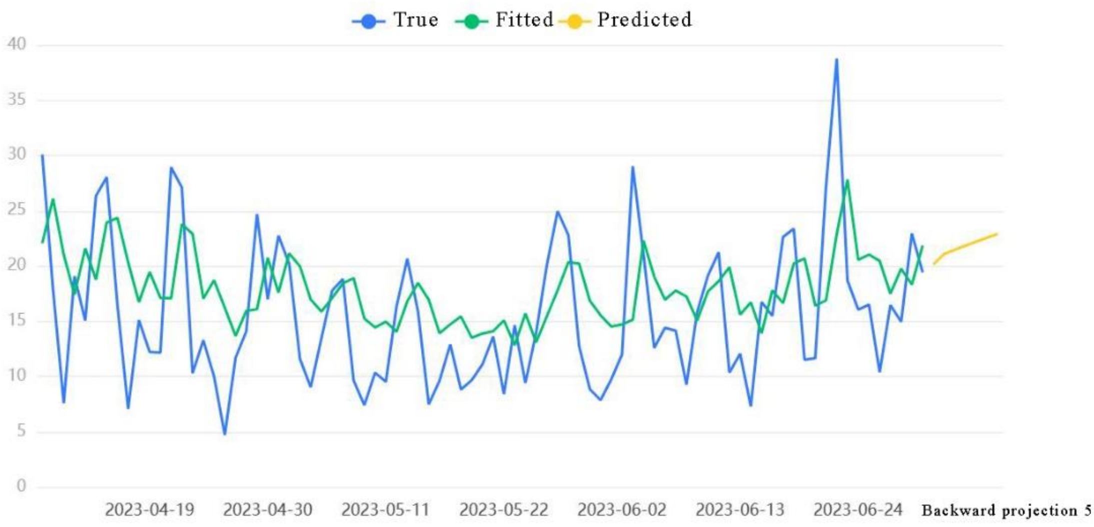


**Figure 6.** Time series volume forecasting graph

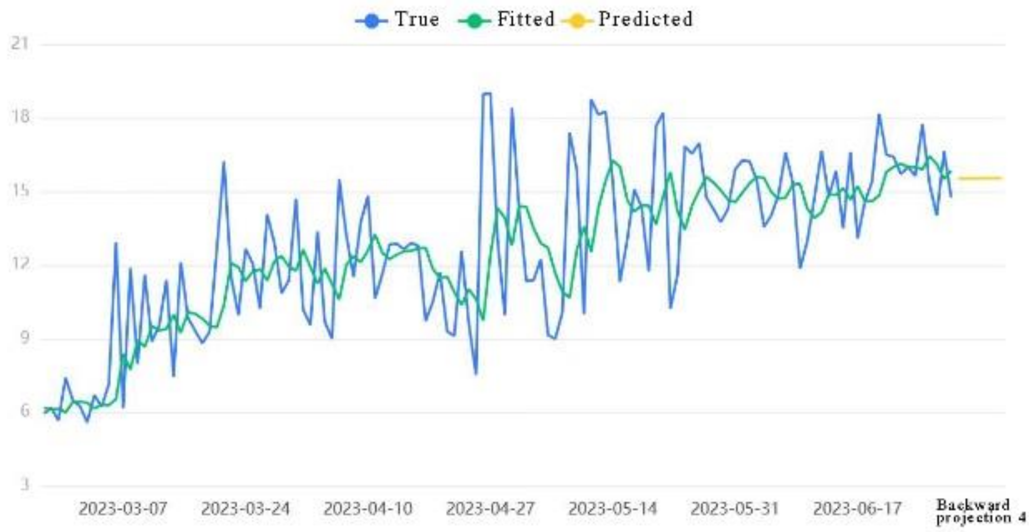


**Figure 7.** Time series pricing forecasting charts

**Table.3** Time series pricing forecasting table

| Order (time) | Predicted results |
|---|---|
| 1 | 15.533289447003675 |
| 2 | 15.535757522790242 |
| 3 | 15.538225598576808 |
| 4 | 15.540693674363375 |
| 5 | 15.543161750149942 |
| 6 | 15.545629825936508 |
| 7 | 15.548097901723075 |

## 3.3. Optimization modeling

According to the linear programming model [6], the define decision variables: Determine the variables to be optimized and set the decision variables as $x_{i\ aquatic\ rhizomes}$, $i = 1,2,3,\cdots,7$, indicating the total number of aquatic rhizomes and tubers sold on day $i$.

Let the rate per day be $w_{i\ aquatic\ rhizomes}$ aquatic rhizomes, the final pricing $y_{i\ aquatic\ rhizomes}$ can be expressed as,

$$y_{i\ \text{aquatic rhizomes}} = (1 + w_{i\ aquatic\ rhizomes})c_{i\ aquatic\ rhizomes} \tag{4}$$

Create an objective function: Translate the objective into a mathematical expression where $c_{i\ aquatic\ rhizomes}$ denotes the cost of entry of aquatic rhizomes on that day. The total profit of the superstore in a week can be expressed as,

$$M_{\text{total}} = \sum_{6}^{i=1} (M_{i\ \text{cabbage}} + M_{i\ \text{foliage}} + M_{i\ \text{chili}} + M_{i\ \text{eggplant}} + M_{i\ \text{mold}} + M_{i\ aquatic\ rhizomes}) \tag{5}$$

Constraints are linear combinations of decision variables. In the case of aquatic rhizomes.

$$y_{i\ aquatic\ rhizomes} = a_{10} - a_{11}x_{i\ aquatic\ rhizomes} \tag{6}$$

The inequality constraint is a non-negative constraint on the variables of interest.

In summary, the optimization model is:

$$maxM_{\text{total}} = \sum_{7}^{i=1} \sum_{6}^{i=1} M_{ij} \tag{7}$$

$$\text{s.t.} \begin{cases} y_{ij} = a_{j0} + a_{j1}x_{ij} \\ y_{ij} = (1 + m_{ij}) + c_{j1} \\ M_{ij} = m_{ij}x_{ij} \\ x_{ij},\ y_{ij},\ M_{ij} >= 0 \\ M_{ij} \in R \\ i = 1,2,\ldots,7 \\ J = 1,2,\ldots\ldots,6, \end{cases} \tag{8}$$

# 4. Replenishment program

Based on section 3, the total number of vegetable items that can be sold is limited to 29, and the order quantity of each vegetable item is less than 2.5 kg, introducing the switch variable $q_i = 0or1, i = 1,2,3,\cdots,251$. $q_i$ as 0-1 variable, the value of 1 means that the vegetable item was sold on that day, and the value of 0 means that the vegetable item was not sold on that day. Summarized as,

$$q_i = \begin{cases} 0, \text{the i-th item was not sold on that day} \\ 1, \text{the i-th item is sold on the same day} \end{cases} \tag{9}$$

Therefore, the control constraint on the total number of items can be expressed by summing the $q_i$ switching variables. That is,

$$27 \le \sum_{251}^{i=1} q_i \le 33 \tag{10}$$

For each individual product ordering quantity to satisfy the minimum display quantity constraint, which can be expressed as the total amount of each individual product $e_i, i = 1,2,3,\cdots,251$ need to be greater than 2.5 kg, can be expressed as follows,

$$e_i \ge 2.5,\ i = 1,2,3,\cdots,251 \tag{11}$$

The optimization model constructed is shown below,

$$maxM_{\text{total}} = \sum_{251}^{i=1} q_i m_i x_i \tag{12}$$

$$\text{s.t.} \begin{cases} y_i = (1 + m_i)c_i \\ y = kx + b \\ e_i \ge 2.5 \\ c_i, x_i \ge 0 \\ q_i = 0 \text{ or } 1 \\ m, y \in R \\ 27 \le \sum q_i \le 33 \end{cases} \tag{13}$$

The final result is obtained as shown in Table.4.

**Table.4** Result

| Single product code | Total daily replenishment | Pricing strategy | Single product code | Total daily replenishment | Pricing strategy |
|---|---|---|---|---|---|
| 102900005115625 | 3.172 | 111.79% | 102900011035078 | 2.762 | 102.07% |
| 102900005115779 | 10.443 | 134.23% | 102900011035740 | 1 | 85.37% |
| 102900005115878 | 9.014 | 142.23% | 102900011021842 | 7 | 71.14% |
| 102900005115861 | 18.215 | 94.02% | 102900011023464 | 5.317 | 127.20% |
| 102900005116257 | 2.299 | 46.21% | 102900011030059 | 13 | 98.47% |
| 102900005116530 | 17.72 | 47.46% | 102900011030097 | 9 | 94.31% |
| 102900005116714 | 18.037 | 60.12% | 102900011030110 | 15 | 137.31% |
| 102900005122654 | 2.129 | 59.75% | 102900011031100 | 19 | 118.59% |
| 102900011013274 | 2 | 125.27% | 102900011031216 | 13 | 118.18% |
| 102900011016701 | 26.919 | 58.49% | 106949711300259 | 20 | 74.51% |
| 102900011033944 | 3.151 | 138.72% | 102900051010455 | 7.795 | 96.43% |
| 102900011034231 | 13 | 70.94% | 102900051000463 | 4.015 | 38.87% |
| 102900011034439 | 5 | 59.37% | | | |

## 5. Conclusions

In this study, based on historical sales data and wholesale prices, we developed an optimization model for pricing and replenishment of vegetable commodities through linear regression, time series forecasting and linear programming. Our analysis shows that the use of correlation coefficient heat map and linear regression model can effectively reveal the relationship between sales volume and pricing, which provides an important basis for future sales volume prediction. Through the time series model and linear programming, we successfully gave the restocking volume and pricing strategy of vegetable category in the coming week to maximize the revenue of the superstore. In addition, we realized more specific replenishment quantity and pricing strategy for individual items on the existing basis, which provides more accurate decision support for the operation of hypermarkets. By combining these methods, this study provides merchants with comprehensive market demand analysis and optimized decision-making solutions, which help improve operational efficiency, reduce costs, and ultimately enhance the competitiveness and profitability of superstores.

## References


[1] Li Y, Xu Q, Wang Y, et al. Genetic Algorithms Application for Pricing Optimization in Commodity Markets[J]. Mathematics, 2024, 12(9): 1289.

[2] Pan H, You X, Liu S, et al. Pearson correlation coefficient-based pheromone refactoring mechanism for multi-colony ant colony optimization[J]. Applied Intelligence, 2021, 51: 752-774.

[3] Osgood D W. Poisson-based regression analysis of aggregate crime rates[M]//Quantitative methods in criminology. Routledge, 2017: 577-599.

[4] Dimri T, Ahmad S, Sharif M. Time series analysis of climate variables using seasonal ARIMA approach[J]. Journal of Earth System Science, 2020, 129: 1-16.

[5] Anwar M Y, Lewnard J A, Parikh S, et al. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence[J]. Malaria journal, 2016, 15: 1-10.

[6] Bellingeri A, Gallo A, Liang D, et al. Development of a linear programming model for the optimal allocation of nutritional resources in a dairy herd[J]. Journal of dairy science, 2020, 103(11): 10898-10916.