ISSN 2616-7433 Vol. 6, Issue 4: 91-96, DOI: 10.25236/FSST.2024.060414

Automatic Pricing and Replenishment Decision of Vegetable Commodities Based on ARIMA—Nonlinear Modeling

Yanzhi Hua[1,#], Xuan Yang[2,#], Suo Liang[1,#]

[1]*School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, 541004, China*
[2]*School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China*
[#]*These authors contributed equally.*

Abstract: *Vegetables have a short freshness period and are easy to deteriorate, and fresh food supermarkets need to develop reasonable replenishment and pricing strategies in order to improve profitability. In order to analyze the time distribution pattern and interrelationship of various types of vegetable sales and give the optimal replenishment and pricing strategy, this paper firstly analyzes the correlation between various types of vegetable sales by using the spearman correlation coefficient, and the results show that the correlation between sales of leafy and eggplant vegetables is weak only. Then using the elbow rule to determine the vegetable single product sales can be divided into four categories, and k-mean clustering to analyze the time series relationship between the total sales of each cluster. Finally, this paper establishes ARIMA to predict the total daily replenishment of each type of vegetables in the coming week based on the time-order relationship and establishes Nonlinear optimization model based on the prediction results, taking the profit maximization of the superstore as the goal, introduces the demand elasticity improvement model, and ultimately gives the optimal daily replenishment of each vegetable category in the coming week and the pricing strategy, and solves for the maximum total profit of ￥10335.23 .*

Keywords: *Elbow Rule, K-means Clustering, Arima, Demand Elasticity, Nonlinear Optimization*

1. Introduction

Generally, vegetables in fresh produce superstores have a short shelf life and deteriorate in character as sales time increases, and most unsold vegetables are usually not available for re-sale on the following day. For this reason, supermarkets restock daily based on historical sales and demand for the items. Without accurate knowledge of specific items and purchase prices, merchants need to determine the restocking program for that day's vegetables. Vegetables are priced on a "cost-plus" basis, often at a discount for damaged vegetables and poor quality items. Reliable market demand analysis is critical to replenishment and pricing decisions. At the demand level, the volume of vegetables sold often correlates with the time of year. At the supply level, vegetables are relatively abundantly available from April through October. Due to the limitation of selling space, a rational sales mix becomes especially critical in supermarkets. Currently, scholars have conducted fewer studies on vegetable replenishment strategies and their pricing problems, and the development of vegetable pricing and their replenishment strategies is made more difficult due to the interference of a variety of external factors such as, for example, the seasonality of vegetable sales, the choice of suppliers, and social contingencies. Yang[1] conducted a correlation analysis of vegetable prices and sales, but failed to explain the seasonality of vegetable prices; therefore, in this paper, we introduce the ARIMA model to predict the seasonal changes of vegetable prices and establish a nonlinear optimization model with the method proposed by Xu[2] as a constraint to give the pricing of vegetables and their replenishment strategies in the coming days.

2. Vegetable commodity sales analysis

Data source: http://www.mcm.edu.cn/?key=%B9%FB%C8%D5.The data contains date of sale,

ISSN 2616-7433 Vol. 6, Issue 4: 91-96, DOI: 10.25236/FSST.2024.060414

vegetable category, unit price of sale, discount situation, name of individual item, and vegetable classification. For reasons of space, some of the data are shown in Table 1 below.

Table 1: Selected data integration results

Date	Sales	Price	Style	Discount	Name	Form
2020-07-01	0.539	8.00	Yes	No	Flowering cabbage	Philodendron
2020-07-01	0.849	3.20	Yes	No	Cabbages	Philodendron

2.1 Spearman correlation analysis

In this paper, Spearman's correlation analysis[3] was carried out to analyze the sales volume of vegetables in six categories, namely, aquatic roots and tubers, foliage, cauliflower, eggplant, chili and edible mushrooms based on the data provided in the annexure. For sequences of size n $X=(x_1,x_2,\cdots,x_n)$, $Y=(y_1,y_2,\cdots,y_n)$, the Spearman correlation coefficient between them is:

$$\rho(X,Y)=1-\frac{6\sum d_i^{\,2}}{n(n^2-1)} \tag{1}$$

Substituting the preprocessed data into the above equation, Spearman correlation analysis was performed on the sales volume of vegetables in six categories and heat map of correlation coefficients of the sales volume of six vegetable categories was plotted, as shown in Fig. 1.

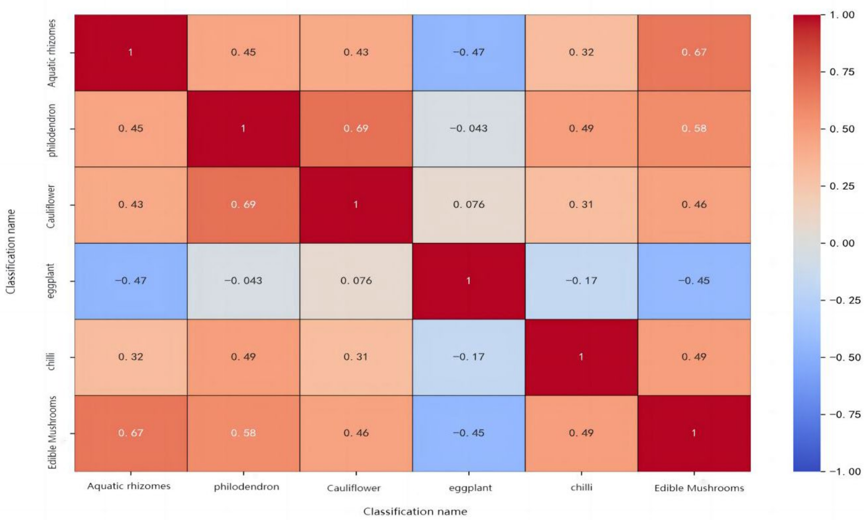

Figure 1: Heat map of correlation coefficients of sales volume of six vegetable categories

Analysis of Figure 1 shows that the correlation coefficient between the sales volume of leafy vegetables and cauliflower vegetables is 0.69, indicating that the correlation between the two is strong, followed by the correlation coefficient between the sales volume of aquatic root vegetables and edible vegetables of 0.67, indicating that the correlation between the two is strong; the correlation coefficients between the sales volume of eggplant vegetables and aquatic root vegetables, chili peppers and eggplant vegetables, and edible mushrooms and eggplant vegetables are all less than 0, indicating that each of them is negatively correlated; the correlation coefficient of sales volume of leafy and flowering vegetables and eggplant vegetables is -0.043, indicating that the correlation between the two is weak;

On the whole, except for the weak correlation between the sales volume of leafy vegetables and eggplant vegetables, the correlation between the sales volume of different vegetable categories was strong.

2.2 Cluster analysis

In order to analyze the distribution pattern and interrelationship of vegetable single product sales. In this paper, we firstly use the elbow rule to determine the number of clusters and perform k-mean clustering, then calculate the total monthly sales volume of vegetable single products of each cluster

ISSN 2616-7433 Vol. 6, Issue 4: 91-96, DOI: 10.25236/FSST.2024.060414

and draw the time series graph of clustering, further analyze the trend of the total monthly sales volume of vegetable single products of each cluster over time, and secondly, we introduce descriptive statistic to analyze the distribution law of the total sales volume of all vegetable single products of each cluster.

2.2.1 Elbow Rule

Elbow rule cluster analysis is a method used to determine the number of clusters. In order to determine the number of clusters for all single product sales respectively, this paper firstly standardizes the data in Annex 2 after preprocessing, and secondly performs elbow law[4] analysis for all single product sales with the following formula:

$$SSE = \sum_{i}^{k} \sum_{p \in c_i} \left| p - m_i \right|^2 \tag{2}$$

According to the above equation, the standardized data is substituted into the calculation, and the elbow curve is drawn as shown in Figure 2 below:

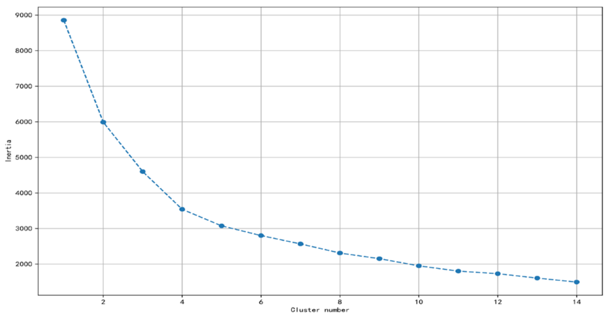

Figure 2: Elbow Curve Diagram

From the elbow rule, when the number of clusters k reaches the optimal value, if continue to increase the value of k leads to bring about a weakening of the improvement effect of the degree of clustering, and at the same time lead to a sudden decrease in the magnitude of the reduction of SSE, and ultimately gradually converge to a flat state. Therefore, it is more reasonable to choose the k value corresponding to the points where the curve changes tend to be flat. Therefore, by analyzing Figure 2, this paper determines that the number of clusters is 4 and uses Python to carry out k-mean clustering, and lists some of the clustering results as follows:

Table 2: Table of clustering results

Clustering number	0	1	2	3
ID	Yunnan Oil seed rape	Chili	Net Lotus Root(1)	Shanghai Youth
	...	...	...	...
	Baokang Alpine Cabbage	Colorful peppers (servings)	Broccoli	Yunnan lettuce

Based on the clustering results in Table 2, the total monthly sales volume of the four major categories are counted and the following clustering time series diagram is drawn.

Analysis of Figure 3 shows that the total sales volume of cluster 0 leveled off from July 2020 to November 2021, while its total sales volume showed a small range of fluctuations over time between November 2021 to April 2022 and July 2022 to February 2023. During April to July 2022, the total sales volume of Cluster 0 showed a wider range of fluctuations and an overall rapid upward trend, indicating that the total monthly sales volume of the vegetable items in this category has been increasing. Between February and June 2023, the total monthly sales volume of Aggregate #0 showed a decreasing trend over time, indicating that the total monthly sales volume of individual vegetable items in this category has been decreasing. As a whole, the monthly total sales volume of cluster #0 was the largest in September 2022, indicating that the monthly total sales volume of vegetable individual items was the best during this month.

ISSN 2616-7433 Vol. 6, Issue 4: 91-96, DOI: 10.25236/FSST.2024.060414

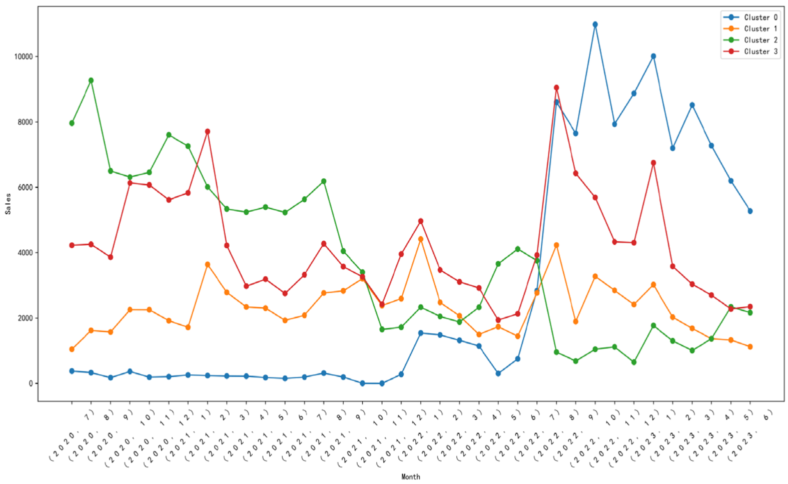

Figure 3: Clustering timing diagram

3. Automatic replenishment and pricing strategy development for vegetable commodities

3.1 Sales forecast based on ARIMA model

ARIMA is a commonly used time series forecasting model[5]. It predicts future trends by analyzing the auto correlation and moving average terms of time series data. In this paper, based on the 3-year time series of the total daily sales of six types of vegetables with cyclical and stable changes over time, using ACF (auto correlation function) and PACF (partial auto correlation function) to determine the p, d and q values of the ARIMA model, this paper establishes the ARIMA (1, 0, 1) model and uses the model to predict the daily sales of six types of vegetables in the coming week, i.e., from July 1 to 7, 2023 The total amount, and its prediction results are shown in the following Table 3:

Table 3: Forecast results of total daily replenishment of six types of vegetables

Date	Projected total daily sales of six types of vegetables(kg)
2023/7/1	356.15
2023/7/2	373.19
2023/7/3	386.49
2023/7/4	396.89
2023/7/5	405.02
2023/7/6	411.37
2023/7/7	416.33

3.2 Nonlinear Optimization Model Building and Solving

This paper establishes a nonlinear optimization model with the objective of maximizing the superstore's profit in the coming week. Considering the impact that the commodity loss rate, selling price range, and total daily sales forecast results will have on the superstore's revenue, this paper uses this as a constraint. In addition, this paper defines the demand elasticity coefficients for aquatic roots and tubers, edible mushrooms, and chili peppers, respectively, and estimates the elasticity coefficients, whose estimated values are respectively $E_{\text{Aquatic rhizomes}} = -0.5$, $E_{\text{Edible fungi}} = -0.5$, $E_{\text{Capsicum}} = 0.5$, The demand elasticity coefficient can describe the effect of price change on sales volume, which is used in this paper to improve the optimization model, so the final expression of its objective function is as follows:

$$max\ Z = \sum_{d=1}^{7} \sum_{i=1}^{n} (p_{i,d} - c_{i,d}) \times R_{i,d} \times (1 - L_{i,d}) \tag{3}$$

The constraints are:

ISSN 2616-7433 Vol. 6, Issue 4: 91-96, DOI: 10.25236/FSST.2024.060414

1) Adjustment of sales volume taking into account elasticity of demand:

$$R_{i,d} \times (1 - L_i) = \frac{S_d}{n} \times (1 + E_i \times \frac{p_{i,d} - p_{i,initial}}{p_{i,initial}}) \tag{4}$$

2) Price range constraints:

$$p_{i,\min} \leq p_{i,d} \leq p_{i,\max} \tag{5}$$

In the above equation, $p_{i,d}$ is the selling price of the goods i on the d day; $R_{i,d}$ is the replenishment quantity for commodity i on day d ; S_d is the total sales volume forecasted for day d ; $c_{i,d}$ is the wholesale price of good i on day d ; E_i is the coefficient of elasticity of demand for good i ; L_i is the wastage rate for commodity i ; $p_{i,\min}$ and $p_{i,\max}$ is the minimum and maximum selling price of good i ; $p_{i,initial}$ is the initial selling price of good i .

Therefore, the expression of this nonlinear optimization model is as follows:

$$max\ Z = \sum_{d=1}^{7} \sum_{i=1}^{n} (p_{i,d} - c_{i,d}) \times R_{i,d} \times (1 - L_{i,d}) \tag{6}$$

$$s.t. \begin{cases} R_{i,d} \times (1 - L_i) = \dfrac{S_d}{n} \times (1 + E_i \times \dfrac{p_{i,d} - p_{i,initial}}{p_{i,initial}}) \\ p_{i,\min} \leq p_{i,d} \leq p_{i,\max} \end{cases} \tag{7}$$

Based on the constraints of the objective function, this paper utilizes python programming to solve for its maximum profit of ￥10,335.23 and gives the total amount of replenishment and pricing for each category of the superstore in a week, the results are shown in Table 4 and Table 5:

Table 4: Results of total daily replenishment

Date	Aquatic rhizomes	Philodendron	Cauliflower	Eggplant	Capsicum	Edible fungi
7.1	14.4226895	67.4332926	66.1597545	64.8833684	96.0853111	22.9747647
7.2	14.7188991	70.6596391	69.3251686	67.9877137	100.439878	24.3329194
7.3	43.2454594	73.1778556	71.7958263	70.4107063	104.236932	25.1972759
7.4	20.5404864	75.1469873	73.7277691	72.3053772	106.913339	24.7311863
7.5	45.6054383	76.6863181	75.2380283	73.7864997	109.042277	31.2848523
7.6	47.1174492	77.8886244	76.4176280	74.9433420	110.894130	29.8742818
7.7	46.8097526	78.8277488	77.3390162	75.8469543	112.293367	28.4841240

Table 5: Pricing strategy results table

Date	Aquatic rhizomes	Philodendron	Cauliflower	Eggplant	Capsicum	Edible fungi
7.1	20.6879893	7.73062364	12.4859334	7.42772622	12.5758393	11.4963790
7.2	20.7750834	7.75942458	12.6177093	7.37313996	12.5301817	11.4574053
7.3	14.7958529	7.75979711	12.6426565	7.37570103	12.5697028	11.4578176
7.4	19.7589456	7.75737016	12.6428761	7.41376003	12.5469683	11.6197470
7.5	14.7374545	7.75429783	12.6975572	7.43069168	12.5363738	10.7810736
7.6	14.5775653	7.75900628	12.6999808	7.43653461	12.5606594	11.0406638
7.7	14.7511710	7.77568176	12.6586184	7.42254020	12.5711435	11.2768281

4. Conclusions

This paper first analyzes the correlation between the sales of various types of vegetables, uses the elbow rule to determine that the sales of vegetable single products can be divided into four categories,

ISSN 2616-7433 Vol. 6, Issue 4: 91-96, DOI: 10.25236/FSST.2024.060414

and carries out k-mean clustering to analyze the temporal relationship between the total sales of each cluster. In order to formulate the optimal replenishment pricing strategy for the coming week, this paper establishes a time series model to predict the total daily replenishment of each type of vegetables in the coming week, and based on the prediction results, establishes a nonlinear optimization model with the objective of maximizing the profit of the superstore, and introduces a model for improving the elasticity of demand, which finally gives the optimal daily replenishment and pricing strategy of each type of vegetables in the coming week. The nonlinear optimization method used in this paper may fall into the local optimal solution, which can be improved by using modern heuristic algorithms and obtain more accurate results.

References

[1] Yang Haoxu. Correlation Analysis of Vegetable Price and Sales—Taking Oil Wheat Vegetable as an Example [J]. Food Safety Guide, 2018, No. 212(21): 179-181.
[2] Xu Guangming, Liu Yihan, Wu Runfa, et al. Optimization method of train stop scheme based on elastic demand [J]. China Railway Science, 2023, 44(06): 184-195.
[3] Gao Mingyu, Cai Linhui, Sun Changcheng. A short-circuit fault detection algorithm for battery pack based on Spearman rank correlation and neural network [J]. Journal of Electronics and Information, 2022, 44(11): 3734-3747.
[4] Wang Chaowei. Study on Typical Day Selection Based on Improved K-means Clustering [D]. Shanghai: shanghai dianji university, 2022(03).
[5] Wu Hui, Wang Jiapeng, Wu Wenjing et al. Prediction and analysis of global gas meter temperature based on ARIMA model [J]. Modern Information Technology, 2023, 7 (16): 147-150.