Research on Vegetable Pricing and Replenishment Strategies Based on ARMA Models

Zimeng Zhao[1, *, #], Xiaohua Fu[2, #]

[1]School of Foreign Languages, Xidian University, Xi'an, China, 710126

[2]Xi'an Aerospace City NO.1 Primary School, Xi'an, China, 710100

*Corresponding author: zhaozimeng777@163.com

#These authors contributed equally.

Abstract. This paper explores the importance of pricing and replenishment issues of vegetable items for merchant operations. First, data related to vegetable stocking were systematically collected to assess the correlation between different categories. Through Q-Q plot analysis, it was determined that the data did not conform to a normal distribution, so the Spearman correlation coefficient was chosen for further quantitative analysis. Considering the large number of individual vegetable categories, in order to deal with the correlation between different individual items, a hierarchical clustering method was introduced in this paper, which classified them into three categories: low volume, medium volume and high volume. In order to predict the future vegetable intake, this study adopted the seasonal ARMA model and applied the Pearson's correlation coefficient for correlation analysis by combining the cost utilization rate as well as the total sales volume. In addition, the relationship between sales and cost margin was fitted with the help of support vector machine. Finally, this study constructs a single optimization model with the objective of maximizing the revenue of the superstore. The innovation of this paper is the combination of various statistical and machine learning methods in order to solve the pricing and replenishment problem of vegetable items, aiming to improve the efficiency of merchant operations.

Keywords: Spielman Correlation Coefficient, Hierarchical Clustering, Low-dimensional Visualization, Seasonal ARMA Model, Pearson Correlation Coefficient, SVR Regression.

1. Introduction

In the marketing of vegetable commodities, they are characterized by deterioration in character with increasing time due to their short shelf life [1]. Therefore, superstores must employ precise replenishment and pricing strategies to maximize revenue [2]. The purpose of this paper is to delve into the field of vegetable sales in order to analyze the sales patterns and their interrelationships across categories and individual items, as well as to explore the correlation between total sales volume and cost-plus pricing. These findings will assist merchants to more effectively develop replenishment plans and pricing strategies for the week ahead. In this context, many merchants have adopted the "cost-plus pricing" approach, i.e., taking cost into account when calculating the price of goods and offering discounts when appropriate. The success of this strategy is directly related to the profitability of supermarkets. The literature review shows that some progress has been made in the field of vegetable marketing. Past studies have usually focused on sales patterns, seasonal demand changes, quality management and inventory control. However, few studies have delved into the sales relationships between categories and individual products and the interactions between these relationships and cost-plus pricing strategies. This is the central focus of the research in this paper. In this study, firstly, the sales data of vegetable items are comprehensively collected and organized; then the sales distribution patterns between different categories and individual items are analyzed to understand their sales trends and seasonal changes. By digging deeper into these data, this paper will explore the relationship between total sales volume and cost-plus pricing for each vegetable category, which will help merchants better determine their daily replenishment totals and pricing strategies for the coming week. The ultimate goal is to provide practical data analysis and decision support tools to help merchants achieve greater profitability in a competitive market. This paper expects to provide a

strong guidance for the operation of superstores by thoroughly studying the factors related to vegetable sales, as well as contributing new insights and methods to the field of sales management of vegetable items.

2. Materials and methods

2.1. Data acquisition and pre-processing

In this paper, the data related to the sales flow details and wholesale prices of each product from July 1, 2020 to June 30, 2023 of six vegetable categories distributed by a superstore were obtained from the network, and the data preprocessing was first carried out to summarize the sales flow details, deal with the wholesale prices, and arrive at the total sales volume of each category of vegetables.

(1) Logical problems to remove outliers

For dumpling leaves, the number is 102900011034354, the sales quantity is 160kg on June 9, 2022, which is seriously inconsistent with the facts, and we consider that it belongs to the outliers and exclude it.

(2)3σ-criteria for eliminating outliers

By referring to previous literature, we use the -criterion with the following equation

$$\mu = \bar{x} \tag{1}$$

$$\sigma = \sqrt{\overline{(x_1 - \mu)^2}} \tag{2}$$

Where μ represents the mean and σ represents the variance.

If the following equation is satisfied, the value is determined to be an outlier and rejected.

$$\begin{cases} x_i > \mu + 3\sigma \\ \text{or } x_i < \mu - 3\sigma \end{cases} \tag{3}$$

Where x refers to the sequence of category daily sales against time.

2.2. Introduction to the methodology

This paper first establishes the distribution model between different categories of vegetable commodities as well as different individual products. Considering that the sales of vegetables are related to the seasons, the average monthly sales and average daily sales of each category of vegetables are obtained by using the processing of the data, and then the data of each single product are obtained according to the same principle; for the degree of correlation between different categories [3], the Q-Q diagram is used to determine that the data do not conform to the positive distribution, so the Spearman's correlation coefficient is used to analyze and to draw heat maps; and for the correlation degree among the different single products As there are too many types of single products, grey correlation and clustering analysis are considered to deal with the degree of correlation between them, and the criterion of clustering is to take their four-year average daily sales, transform it into the overall value, and then through low-dimensional visualization, after initially judging that it can be divided into a few categories, clustering is performed on the single products. In order to help merchants determine the total daily replenishment and pricing strategy for the coming week, the seasonal ARMA model is considered to be used to fit the curves of sales and cost margin of each image category using SVR; finally, a single-objective optimization model with the goal of maximizing the revenue of the superstore is established, which helps merchants determine the total daily replenishment and pricing strategy for the coming week [4].

3. Model Solving and Establishment

3.1. Distribution law and interrelationship of sales volume of each category and single product of vegetables

3.1.1 Distribution law

(1) Sales volume in terms of months

To find the distribution of each class corresponding to the month, each class with the corresponding single product of the monthly distribution of the superposition, and then use MATLAB to derive the distribution of sales volume in terms of the month, as shown in Figure 1.

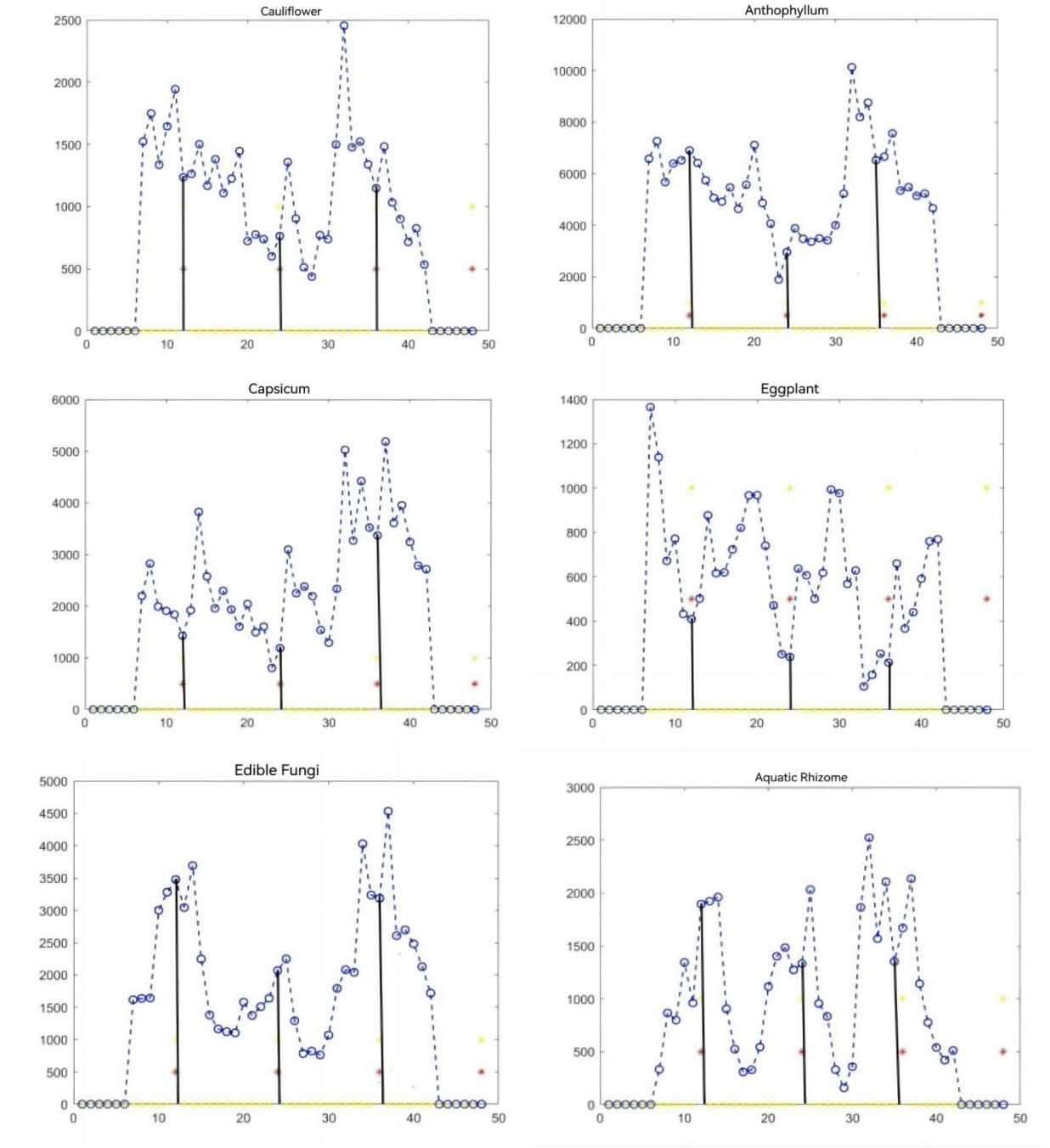

Figure 1. Distribution of sales by month

As can be seen from the figure, the sales volume of the six categories regarding the month showed seasonal changes in general, with eggplant and aquatic roots and tubers being particularly obvious.

Through analysis, we believe that this seasonal change is related to the maturity period of the vegetables, and that the inter-annual change is related to market demand, natural disasters, and so on.

(2) Sales in days

Each category is split into single items, and the sales distribution of single items is processed for 1095 days, and the missing days are replaced by 0. The single items are then superimposed, and MATLAB is utilized to derive a sales distribution graph in terms of days. As shown in Figure 2:

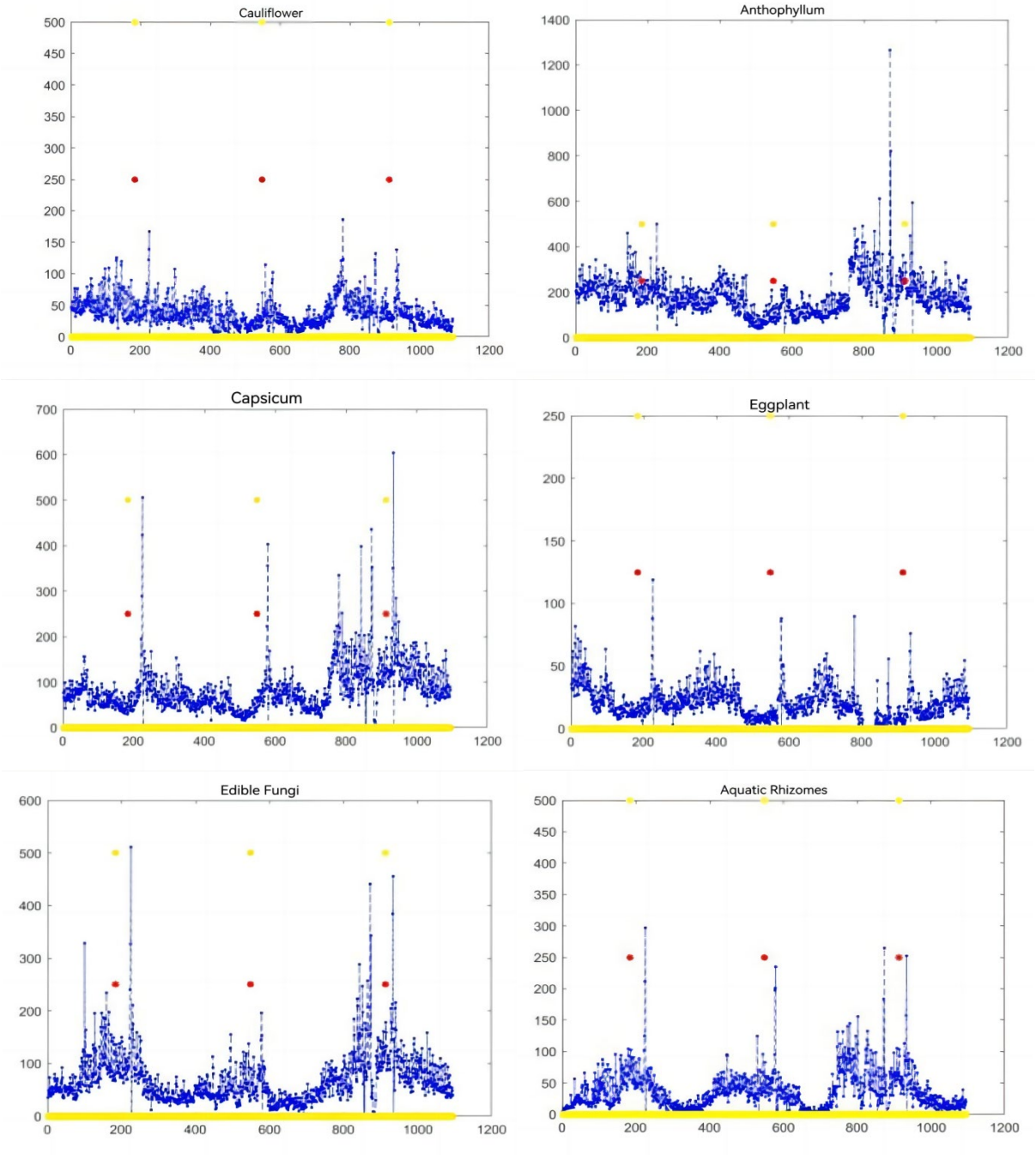

Figure 2. Distribution of sales in days

From the figure, it can be seen that the categories are similar in terms of higher sales at 200 days, 600 days and 700 days.

3.1.2 Correlation analysis

(1) Determine whether it meets the linear relationship

Calculate through SPSS software to get the matrix scatter plot. As shown in Figure 3:

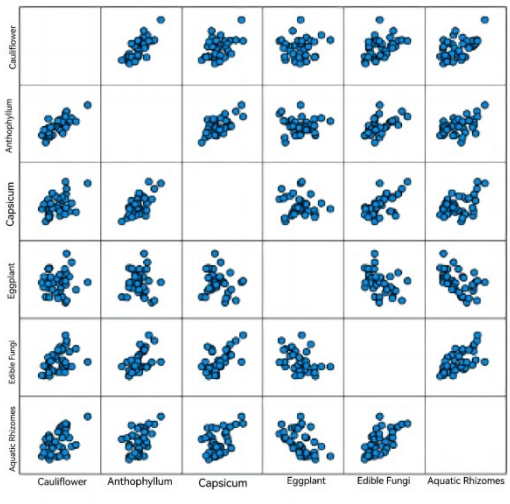

Figure 3. Matrix scatterplot of monthly sales in different categories

As can be seen from the figure, the monthly sales of different categories conform to a linear relationship.

(2) Normal distribution test

The Q-Q diagram is used to determine whether the distribution of the six categories conforms to normal distribution, and Figure 4 takes the cauliflower category as an example.

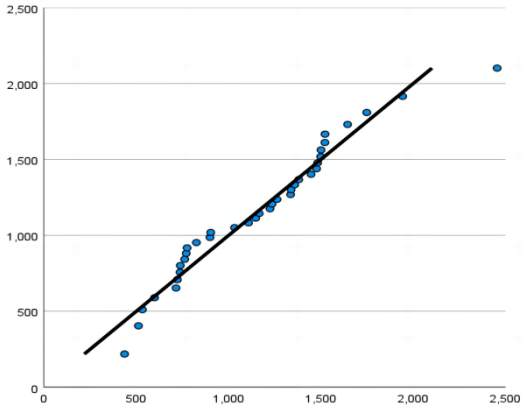

Figure 4. Normal Q-Q plot for cauliflower species

As can be seen from the figure, the distribution points of the six categories deviate from the straight line, so they do not conform to the normal distribution.

(3) Spearman correlation coefficient

For each category to find its interrelated Spearman correlation coefficient. As shown in Figure 5:

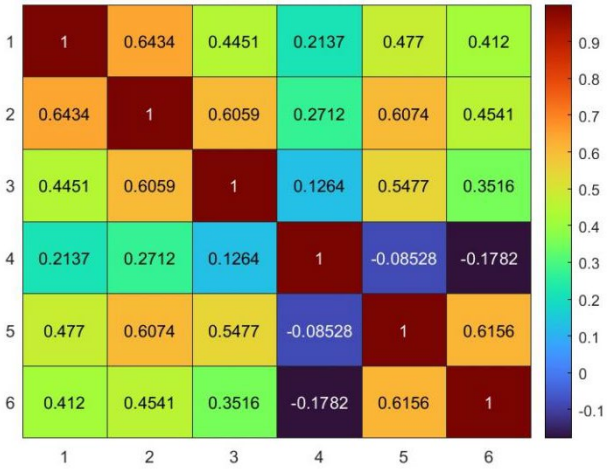

Figure 5. Heat map of correlation coefficients

In the figure serial numbers 1-6 are cauliflower, foliage, pepper, eggplant, edible fungi and aquatic roots and tubers, respectively, from the figure, it can be seen that cauliflower and foliage, edible fungi and aquatic roots and tubers are more strongly correlated, while foliage and eggplant are weakly correlated, and we initially divided these six categories into two parts according to whether the Spearman correlation coefficient is greater than 0.5.

3.1.4 Cluster analysis

Taking the average daily sales of vegetable individual products as the classified data, the first low-dimensional visualization was carried out, as shown in Figure 6:

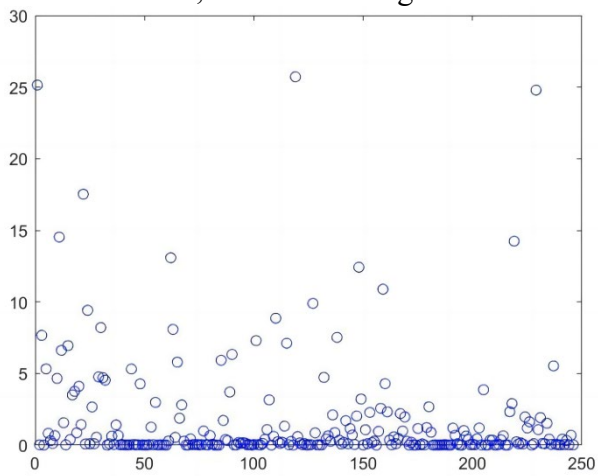

Figure 6. Low-dimensional visualization of average daily sales of a single product

According to the low-dimensional visualization it can be seen that these individual products can be divided into three categories[5].

Hierarchical clustering and grouping with vegetable category, single item and sales volume as indicators, the same group has strong correlation, and get the spectrogram using average linkage, from which we separate vegetables into three categories: low volume, medium volume and high volume.

3.2. Determination of total daily replenishment and pricing strategy for the coming week

3.2.1 ARMA time series forecasting model

In order to predict the daily replenishment quantity and wholesale price of each category of vegetables in the coming week based on the daily sales and wholesale price for three years, the ARMA time series forecasting model can be used to realize the overall process as shown in Figure 7:

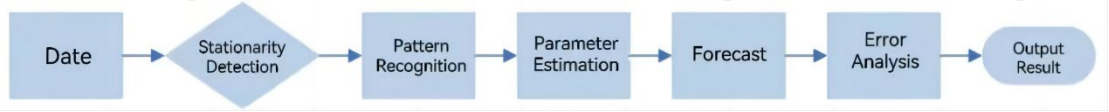

Figure 7. Flowchart of ARMA time series forecasting model

(1) Smoothness test

In order to test whether the time series is smooth or not, we use MATLAB's adftest () to analyze the data for smoothness, and the test found that all six types of data are smooth.

(2) Model identification

Taking the cauliflower category as an example, the autocorrelation function of the time series was analyzed, as shown in Figure 8:

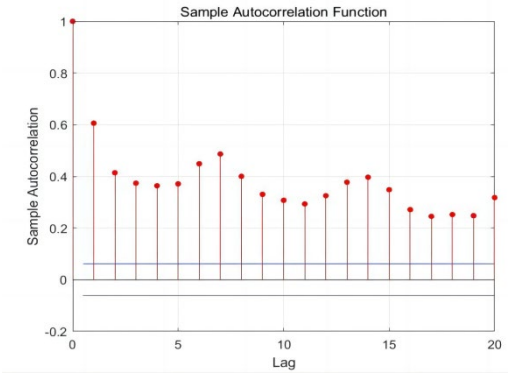

Figure 8. Cauliflower autocorrelation

As shown above, the autocorrelation function has a trailing nature. As an example, the partial autocorrelation function of the analyzed time series for cauliflower species is as follows:

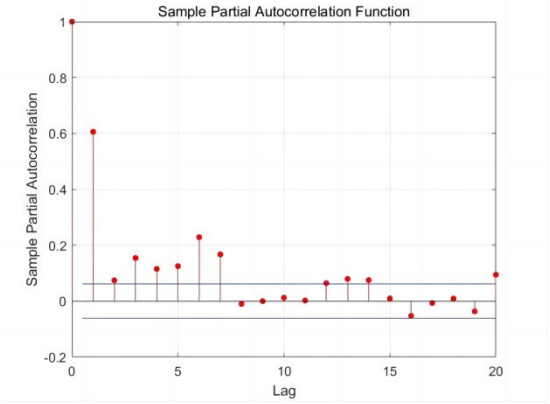

Figure 9. Cauliflower class biased autocorrelation

As shown in the figure above, the partial autocorrelation function has a truncated tail. Combined with the above two points so the ARMA model was selected to predict the daily restocking volume and wholesale price of each category of vegetables in the coming week.

Finally, the daily replenishment volume for the next seven days is obtained, as shown in Table 1:

Table.1. Daily replenishment for the next seven days

Cauliflower prediction	Anthophyllun prediction	Capsicum prediction	Eggplant prediction	Edible Fungi prediction	Aquatic Rhizomes prediction
23.38	132.5	79.58	20.45	45.15	19.93
21.85	126.52	80.15	19.11	49.73	19.75
20.81	126.64	80.46	17.29	52.48	20.6
21.89	139.84	82.15	17.57	54.25	20.27
22.89	148.06	81.77	17.63	55.07	21.16
23.63	140.83	84.41	19.84	54.83	21.82
24.81	142.86	86.31	20.55	55.78	22.48

(3) Parameter estimation and prediction

We use the seasonal ARMA model to fit and predict the daily intake and wholesale price data for the coming week, and compare the predicted data with the real data, and the comparison graph is shown in Figure 10:

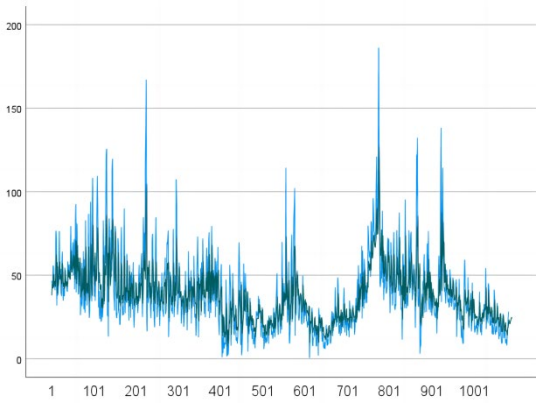

Figure 10. Comparison of predicted data with real data

(4) Error analysis

The mean square error at this point, as shown in Table 2:

Table.2. Mean Square Error

Cauliflower	Anthophyllun	Capsicum	Eggplant	Edible Fungi	Aquatic Rhizomes
16.23	61.59	36.59	9.11	33.67	21.70

This data is within the margin of error allowed for the daily business volume of the superstore, and is a guide to the daily purchase volume and wholesale prices for the coming week.

3.2.2 Relationship between vegetable sales volume and cost pricing ratio

Take the cauliflower category as an example, the daily distribution of cauliflower sales volume and the daily distribution of the markup rate are counted, where the markup rate is calculated by the formula: unit price/wholesale price-1, and the days when there is no data for the wholesale are excluded, and the markup rate is calculated again. A scatter plot with sales volume as the independent variable and markup rate as the dependent variable was plotted, and regression analysis was performed using SVR to find out the possible relationship between sales volume and markup rate, in which the correlation analysis of sales volume and markup rate is shown in Figure 11:

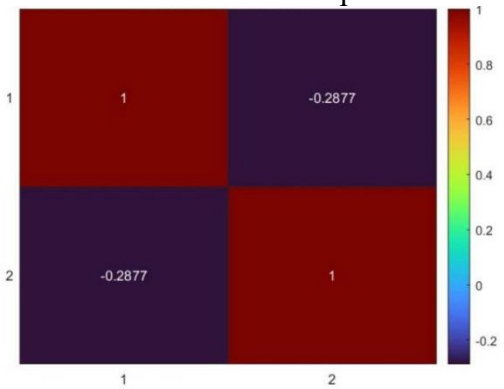

Figure 11. Heat map of Pearson's correlation coefficient between sales volume and markup rate

As can be seen from the figure, there is a weak correlation between sales volume and markup rate, the four kernel functions of SVR regression are used to analyze the sales volume and markup rate, respectively, to take the kernel function with the highest accuracy, and then use this kernel function as the basis for obtaining the highest accuracy of the regression model to make the regression model and , which, due to the reason that the distribution of sales volume over time is not homogeneous for the different categories in order to obtain a better fitting curve, to first adjust the image according to the The percentage of the model training sequence in the original sequence. The SVR regression for the cauliflower category is shown in Figure 12:

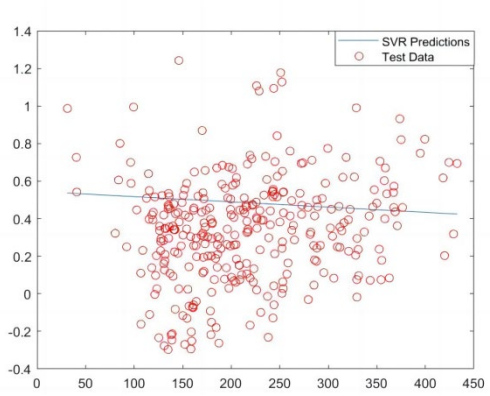

Figure 12. SVR regression plot for cauliflower species

As can be seen from the figure, the sales volume of the category and the markup rate are negatively correlated. For different categories in the SVR regression, the length of sequence intercepts is shown in Table 3:

Table 3. Sequence Intercept Length

Cauliflower	Anthophyllun	Capsicum	Eggplant	Edible Fungi	Aquatic Rhizomes
0.7 × length	0.7 × length	0.55 × length	0.55 × length	0.7 × length	0.7 × length

For the selection criteria of the intercept length is determined by Figure 13:

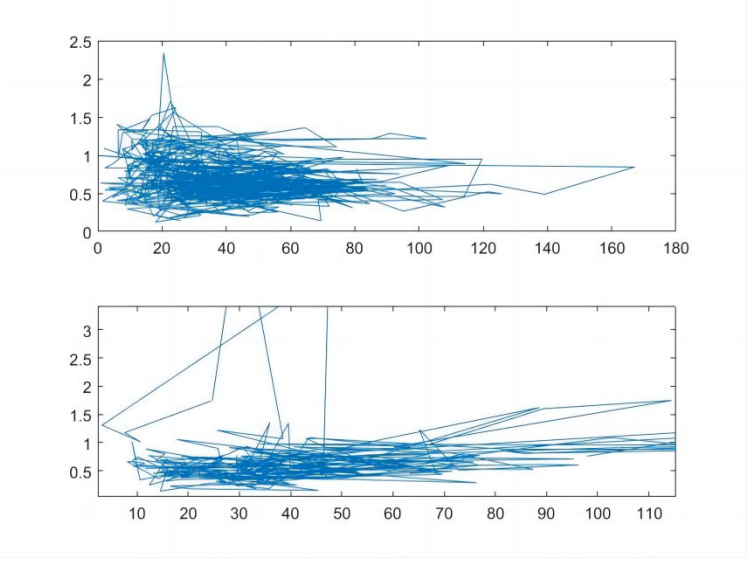

Figure 13. Distribution of training series (top) and validation sequence (bottom)

When the intercept length is appropriate, the distributions of the training and validation sequences should be approximately the same.

3.2.3 Pricing Strategy Model for Vegetable Category [6]

We take the maximum revenue of the superstore as the optimization objective, and take the process boundaries as the constraints to build the optimization model and solve it. The optimization objective (for a single category) is as follows:

$$= \frac{g}{1-H} \tag{4}$$

$$W = \max \sum_{j=1}^{6} \sum_{i=1}^{7} (1 + L_{ij}) D_{ij} \cdot g + D_{ij} \cdot \quad _{ij} \tag{5}$$

$$W = \max \sum_{j=1}^{6} \sum_{i=1}^{7} [(1 + L_{ij}) g + \quad _{ij}] D_{ij} \tag{6}$$

$$W = \max \sum_{j=1}^{6} \sum_{i=1}^{7} \{[1 + f(g)]g + \quad_{ij}\} D_{ij} \tag{7}$$

Where is a function of the SVR model with sales volume as the independent variable and markup rate as the dependent variable. For different purchase prices does not affect the total profit to the maximum.

Thus the pricing strategy for the coming week for each vegetable category is determined.

4. Discussion

In this paper, it is argued that the factors affecting the replenishment pricing decision, in addition to the cost-plus pricing, purchase volume, and supply volume mentioned above, the superstore also needs to collect the following data [7]:

① supply quantity, including the maturity of each category of vegetables, transportation mode (cold chain transportation, etc.), tranfsportation distance, transportation costs, etc;

② Demand, including the market demand for each category of vegetables;

③ loss rate, including the freshness period of each category of vegetables, discounted selling rate, etc.;

④ environment, including the zone in which it is located, the competition of the surrounding supermarkets, online shopping for vegetables and so on.

These data traders can keep an eye on the statistics in the process of operation.

In order to further improve the revenue of supermarkets, the following suggestions can be referred to

①Regulate the transportation system

Rationale: Supermarkets use cost-plus pricing to set the selling price, of which freight is an important part of the price of the markup, thus making pricing decisions, for customers, standardize the transportation system can not only reduce the freight costs, so that customers can buy cheaper vegetables, but also reduce losses in the transportation process, improve the quality of vegetables, so that customers can enjoy the fresher and higher quality of vegetables; for supermarkets, standardize the transportation system to save costs and increase revenue. For supermarkets, regulating the transportation system can save costs and increase revenue, while improving vegetable quality, obtaining repeat customers, and facilitating the scale of operation of supermarkets.

② Conduct market research[8]

Reason: market demand is an important factor affecting the replenishment decision to carry out market research is conducive to understanding customer demand, more replenishment of the customer's general favorite dishes, less replenishment of cold vegetables, so as to formulate different replenishment strategy, can effectively stabilize the source of customers, to avoid the remaining sales of vegetables, more than the freshness of the period of time and so on, to reduce the loss of the super to improve revenue.

③Improve the preservation technology[9]

Rationale: Improve the freshness technology can effectively alleviate the problem of short freshness period of vegetables, the character deteriorates with the increase of time. Supermarkets can use cold chain transportation in the transportation process, stored in cold storage before sale, and use freezers, water spraying and other methods to stop the decay of vegetables when selling, to extend the storage period, from which different pricing and replenishment strategies are developed.

④ Carry out online vegetable purchasing service

Rationale: With the development of the Internet, more and more people choose to buy vegetables online, as a brick-and-mortar superstore should also comply with the development of the times, to carry out online shopping services. Supermarkets establish an exclusive grocery shopping applet, cultivate a team of high-quality delivery staff, do a good job of after-sales service, and develop different pricing and replenishment strategies, which are conducive to attracting more potential customers and increasing sales.

⑤ Reasonable market competition

Reason: If there are other similar superstores in the vicinity of a superstore, competition will inevitably arise. We can adjust the pricing strategy by taking discount offers and giving small gifts, etc., to utilize our own advantages and grab customers.

5. Conclusions

In the field of vegetable merchandising, the issues of pricing and replenishment are crucial to the business of merchants[10]. This study delves into the key issues in the vegetable merchandising arena, aiming to provide insights into the sales patterns and their interrelationships across categories and individual items, as well as exploring the correlation between total sales volume and cost-plus pricing. These findings will help merchants plan their replenishment programs and pricing strategies more effectively for the week ahead. In conclusion, this study provides in-depth data analysis and decision support that emphasizes the interrelationships among different factors, thus providing strong support for the operation of vegetable commodities. Future research directions will include further optimizing the model to suit the needs of different markets and merchants, and verifying the validity of the model through actual operations. These efforts will help improve the efficiency and profitability of vegetable merchandising, provide traders with a sustainable business direction, and also provide a more in-depth research foundation for academia.

References

[1] Kabila A,Senior F A O,M R G. Market vegetable hygiene practices and health risk perceptions of vegetable sellers in urban Ghana [J]. International journal of environmental health research, 2019, 29(2).

[2] Sarma M, Nanere M, Trebilcock P. Pricing strategies for organic vegetables based on Indonesian consumer willingness to pay [J].Management Science Letters, 2020.

[3] Gyung M L. Classification of Vegetable Commodities by the Codex Alimentarius Commission [J]. Journal of Food Hygiene and Safety, 2019, 34(1).

[4] Mesa P C J,Mesa P J F,Leon T J J. Scheduling vegetable sales to Â supermarkets in Â Europe: The tomato case [J]. Agricultural Economics, 2022, 68(11).

[5] Abigail P,Debra Z,Jérémie L. Adult's Mental Representations of Foods and Vegetable Categorization are Predictive of Food Acceptance [J]. Appetite, 2022, 179.

[6] Chiranjit C, Kumar P G, Kumar R P, et al. Solving a Mathematical Model for Small Vegetable Sellers in India by a Stochastic Knapsack Problem: An Advanced Genetic Algorithm Based Approach [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2022, 30(05).

[7] South Sudan to create jobs with fruits and vegetables strategy [J]. M2 Presswire, 2022.

[8] Widad Z,Haneen T,Aisha S, et al. Consumption of fruits and vegetables among adolescents in Arab Countries: a systematic review [J]. The international journal of behavioral nutrition and physical activity, 2023, 20(1).

[9] Upreti S,Thapa S,Kafle S, et al. A Review on Marketing Loss and Channels of Fruits and Vegetables [J]. Food and Agribusiness Management, 2022, 3(1).

[10] Akram H,J F L V,Rodrigo V. Socioeconomic inequalities in the consumption of fruit and vegetables in Peru between 2014 and 2019 [J]. Public health nutrition, 2022, 25(12).