Study On Replenishment and Pricing Strategy of Fresh Merchants' Super Vegetables Based on Time Series Prediction and Optimization

Jin Li [#, 1, *], Anlan Zhou [#, 1], Yuhang Wang [#, 2], Yihan Yan [3], Yueran Si [4], Yu Zheng [1]

[1] Eurasian International College, Henan University, Zhengzhou, China, 475001

[2] Software College, Henan University, Zhengzhou, China, 475001

[3] International Business College, Henan University, Zhengzhou, China, 475001

[4] Business College, Henan University, Zhengzhou, China, 475001

* Corresponding Author Email: 15939979600@163.com

[#]These authors contributed equally.

Abstract. Aiming at the problem of vegetable replenishment and pricing strategy in fresh merchant supermarket, this paper makes a replenishment plan based on categories by analyzing the distribution law and interrelationship of vegetable categories and individual product sales. and the total daily replenishment and pricing strategy in the coming week are obtained by using the time series prediction model, so as to maximize the profit. The research methods include data preprocessing, sales law analysis, interrelation analysis, linear regression model, time series prediction model and optimization model. The results show that mastering the distribution law and interrelationship of the sales volume of various categories and individual products of vegetables, and using the time series prediction model and optimization model to formulate the total replenishment and pricing strategy can maximize the profit.

Keywords: Fresh merchant super, Vegetable replenishment, Pricing strategy, Time series prediction, Optimization model.

1. Introduction

The appearance of vegetable commodities in the supermarket will deteriorate with time [1-3]. The pricing of vegetables generally adopts the method of "cost plus pricing", and Shang Chao will also sell goods with discounts on goods that are damaged or deteriorated. Because the variety of vegetables sold by Shang Chao is rich, the origin is diverse, and the purchase transaction time of vegetables is early, therefore, merchants must make replenishment decisions for each vegetable category on the same day without knowing the specific product and purchase price [4]. Therefore, reliable market demand analysis is very important for replenishment decision and pricing decision. From the demand side, the sales volume of vegetable goods is often related to time; from the supply side, the supply variety of vegetables is more abundant from April to October, and the reasonable sales combination becomes extremely important because of the limitation of over-sales space [5-10].

To sum up, this study solves the following problems:

(1) The distribution law and interrelationship of the sales volume of various categories and individual products of vegetables were analyzed.

(2) Taking the category as a unit to make the replenishment plan, this paper analyzes the relationship between the total sales volume of each vegetable category and the cost-plus pricing and gives the daily replenishment quantity and pricing strategy of each vegetable category in the coming week (July 1-7, 2023).

2. Materials and Methods

2.1. Data acquisition and preprocessing

At present, there are a number of vegetable data, including commodity information of all kinds of vegetables, sales details and wholesale prices of each product from July 1, 2020, to June 30, 2023, and recent loss rate data of each commodity. In this paper, data preprocessing is carried out. Due to the large amount of data, it is necessary to preprocess the given data to prevent the existence of error, repetition and abnormal data. It makes the calculation have an impact on the results, resulting in a large error. Python is used for data preprocessing, firstly, duplicate data recognition is carried out. The results show that there is data duplication, and then edit the code to delete the duplicate data. Then the missing data is identified, and the results show that there is no missing data.

In this study, the data in each attachment are processed, and the box chart is used to check whether there are any outliers. The method of normal distribution is also used to identify outliers. Then delete the data in the row of the abnormal value, integrate the data and remove the invalid data to get the final data and complete the data preprocessing.

2.2. Research methods.

The main contents are as follows:

(1) Analyze the distribution law and interrelationship of the sales volume of each category and single product, which is divided into four parts: the sales law of each category, the interrelation of each category, the sales rule of each item and the interrelation of each item. First of all, the data are processed, the monthly sales volume of each category is calculated on a monthly basis, and a bar chart is drawn, so as to intuitively observe the sales changes of each category in a year. Secondly, the average monthly sales of each category is calculated and the pie chart of their respective proportion is drawn. Thirdly, the Spelman coefficient of each category is calculated to express the relationship between them. Then, take the month as the unit to calculate the monthly sales volume of each product, and draw a bar chart, so as to intuitively observe the sales changes of each product in different time. Finally, the correlation coefficient between each item is calculated by excel, which indicates the relationship between each item.

(2) First of all, the total sales volume of each vegetable category is calculated on a daily basis. The cost-plus pricing method is a method to determine the product price according to the unit cost of the product plus a certain proportion of profit. Then calculate the average price addition rate of each category of vegetables on a daily basis and get the daily price addition rate of each category of vegetables. Then the relationship between total sales and cost-plus pricing of each vegetable category is obtained by SPSS linear regression. A time series prediction model is established, and the total daily sales in the coming week is deduced by using the data of merchant super-sales, and then the total daily replenishment in the coming week is calculated according to the loss rate of each commodity. Then the objective function of the maximum profit is established by using the functional relationship between the total sales and cost-plus pricing, and then the pricing strategy of each vegetable category in the coming week is obtained by particle swarm optimization algorithm.

3. Model construction and analysis.

3.1. Distribution and interrelationship of sales volume of various categories and individual products of vegetables.

3.1.1 Sales rules of each category.

In order to find out the sales rules of each category, through the analysis and processing of data, this study finally decided to draw the bar chart of monthly sales volume of each category on a monthly basis (figure 1).

From the bar chart above, it can be concluded that the sales of edible fungi and aquatic rhizomes decreased significantly from May to August, the sales of eggplant vegetables decreased sharply around November, and the sales of cauliflower, mosaic and pepper vegetables were more balanced throughout the year. Sales will increase slightly during the Spring Festival.

After careful analysis, this study also found that, except eggplant vegetables, the sales of other kinds of vegetables increased significantly at the end of 2022, while eggplant vegetables decreased significantly. This study speculates that the reason for this change may be that the output of eggplant vegetables has dropped sharply for some reason, and residents have no choice but to use other kinds of vegetables instead of eggplant vegetables.

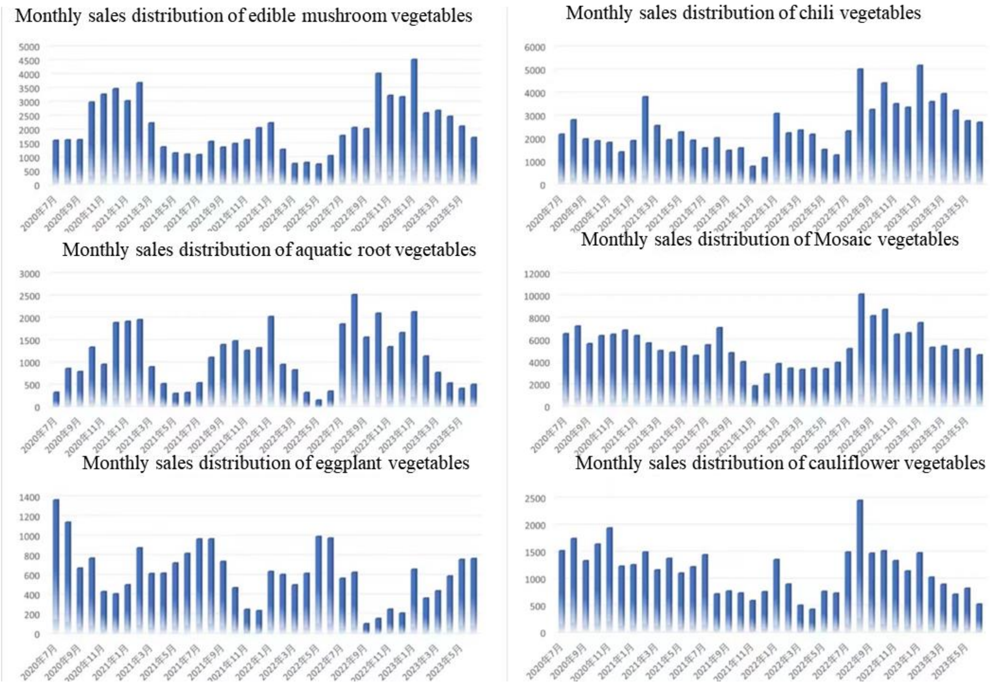

Figure. 1 Bar chart of monthly sales volume of each category.

In addition, this study also draws a pie chart of the average monthly sales of all kinds of vegetables (figure 2), which can clearly observe the share of all kinds of vegetables. Residents have the greatest demand for mosaic vegetables, with an average monthly sale of 42%, followed by peppers and edible fungi, accounting for 19% and 16%, respectively. Eggplant vegetables account for only 5%, and the demand is the least. Therefore, Shang Chao should mainly focus on mosaic vegetables when formulating replenishment strategies.

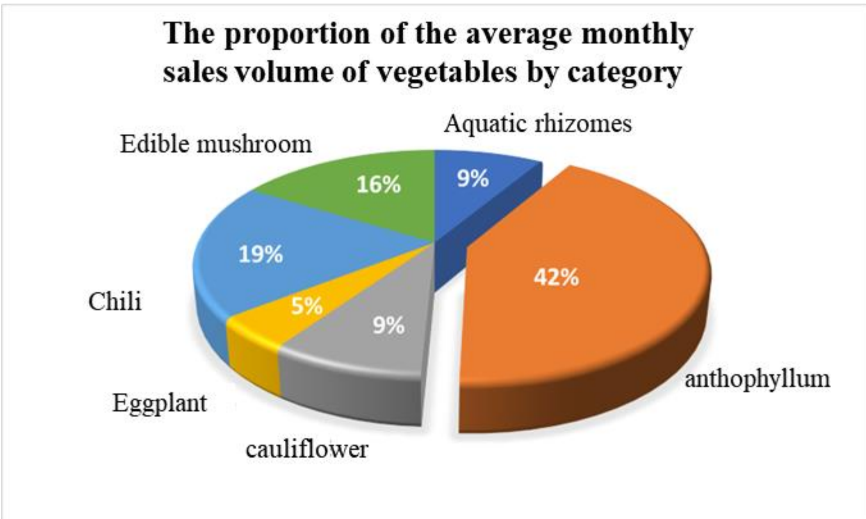

Figure. 2 Pie chart of average monthly sales of all kinds of vegetables.

3.1.2 The relationship between different categories.

In order to analyze the relationship among different categories, the Spelman coefficient among different categories of vegetables was calculated by using SPSS (Table 1).

As a result, this study shows that the correlation coefficient of mosaic, cauliflower and pepper is higher, and that of edible fungus and aquatic rhizome is higher, so they can be considered to be complementary to each other. However, eggplant and pepper, mosaic, cauliflower and aquatic rhizome have lower correlation coefficient, so they can be considered as substitutes for each other. This result is similar to the sales law of all kinds of vegetables and has reference value.

Table 1 Spelman coefficient among different types of vegetables

	Aquatic rhizomes	anthophyllum	cauliflower	solanacea	capsicum	Edible mushroom
Aquatic rhizomes	1					
anthophyllum	0.56105	1				
cauliflower	0.542009	0.626229	1			
solanacea	0.105874	0.322621	0.347625	1		
capsicum	0.614127	0.659296	0.55119	0.317654	1	
Edible mushroom	0.669914	0.630847	0.522487	0.14076	0.687464	1

3.1.3 Sales rules of each item.

Due to the large number of items, a representative product was selected from each category for analysis. in this study, red pepper (1), pure lotus root (1), broccoli, apricot abalone mushroom (1), Yunnan lettuce and purple eggplant (2) were selected to represent pepper, aquatic rhizome, cauliflower, edible fungus, mosaic and eggplant respectively. In order to find out the sales law of each item, through the analysis and processing of the data, this study finally decided to draw the bar chart of the monthly sales volume of each item on a monthly basis (figure 3). According to the chart analysis, the sales trend of pure lotus root, broccoli and purple eggplant (2) is basically consistent with that of aquatic rhizome, cauliflower and eggplant vegetables.

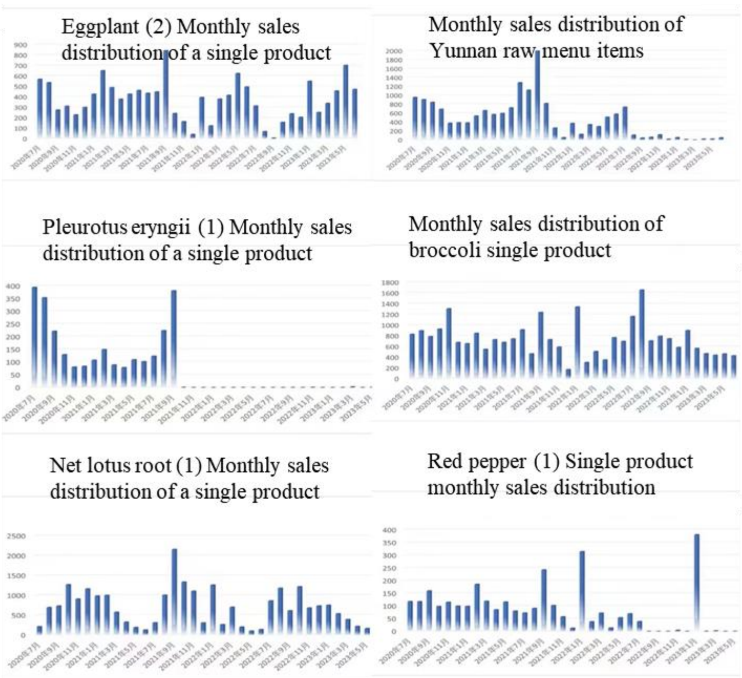

Figure 3. Bar chart of monthly sales volume of each item.

3.1.4 The relationship between individual items.

In order to analyze the interrelation of each item, this study calculated the interrelation of each item by SPSS. After analysis, there was a high correlation between pakchoi and millet pepper, green cabbage and colorful pepper, fungus and millet pepper, and between pakchoi and pure lotus root, broccoli and millet pepper. Therefore, this study speculates that when customers buy cauliflower and mosaic vegetables, they will also buy chili vegetables for seasoning. The correlation between pakchoi and Pleurotus ostreatus, broccoli and Pleurotus ostreatus was poor, which was consistent with that of the previous categories.

3.2. Vegetable replenishment and pricing model of Shang Chao.

3.2.1 Establish a linear regression model to analyze the relationship between total sales and cost-plus pricing.

Linear regression is a statistical analysis method that uses regression analysis in mathematical statistics to determine the quantitative relationship of interdependence between two or more variables. In this study, the variables of linear regression analysis are the total sales volume of each vegetable category and the pricing addition rate of each vegetable category. Therefore, this study calculates the total sales amount of each vegetable category in daily units and the pricing addition rate of each vegetable category by using excel, and then calculates the average daily pricing addition rate of each vegetable category to get the daily pricing addition rate of each vegetable category. Finally, using Python to establish a linear regression model to find out the quantitative relationship between the total sales volume of each vegetable category and the cost addition rate, and analyze the relationship between the total sales volume and cost addition pricing of each vegetable category (figure 4).

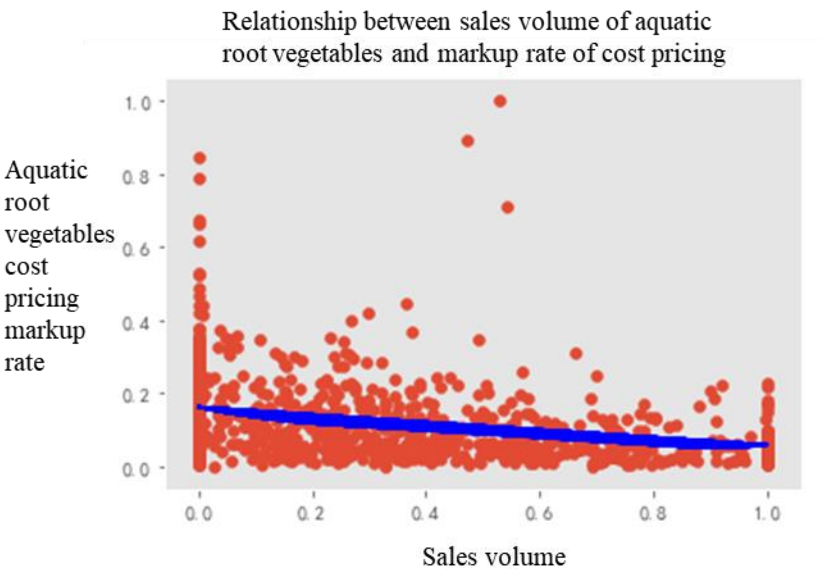

Figure 4 Relationship between total sales volume of each vegetable category and cost-plus pricing.

3.2.2 Using time series forecasting model to formulate replenishment strategy.

Time series prediction model is a mathematical expression that observes and measures a certain variable or a group of variables in production and scientific research. It will be arranged in time order at a series of times (taking time as variables) and is used to explain the variables and their interrelations. In this study, the total sales volume of each vegetable category in a single day was taken as a group of variables. There are six vegetable categories, so this study needs to establish six groups of time series forecasting models to predict the sales of each vegetable category in the coming week. First of all, this study reads the data and processes the data, and uses the moving average graph to test the stability of the original data. Then stabilize the unstable data until it passes the stability test.

The stability of six groups of original data and the processed stable data are tested (figs. 5, 6 and 7). Then this study creates the ARIMA model and fits the model. First-order difference reduction, again first-order difference reduction, moving average reduction and logarithmic reduction are carried out, and the order determination results of the model are obtained.

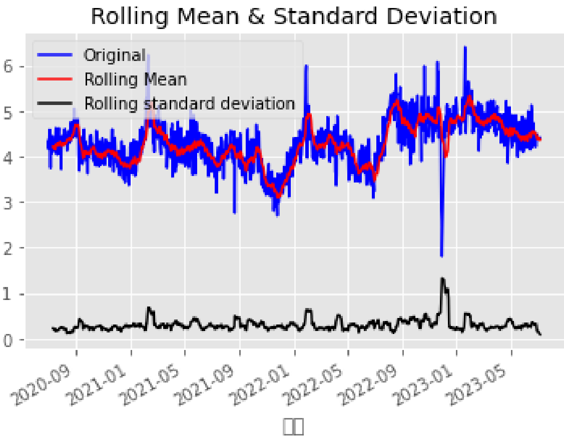

Figure 5 Stability test of raw data of chili peppers

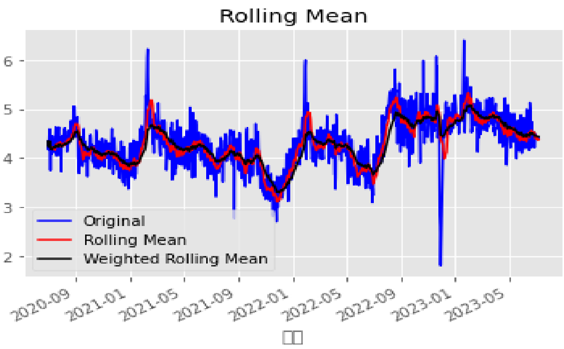

Figure 6 Stability test of processed data of chili peppers

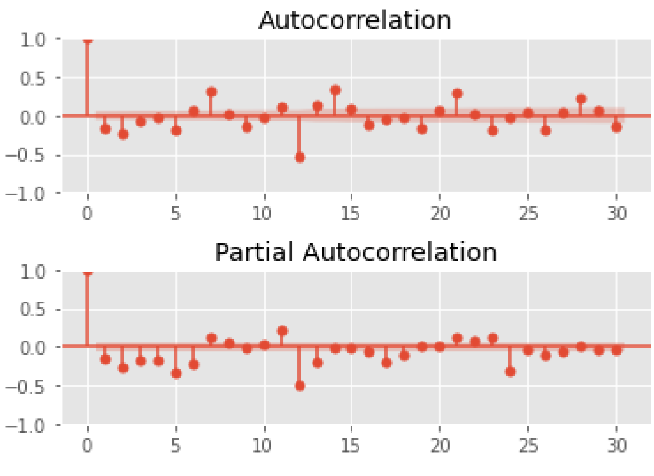

Figure 7 result of order determination of pepper model.

Finally, the prediction result of time series is obtained, that is, the sales volume is predicted.

After using the time prediction model to predict the sales of each vegetable category in the coming week, this study is based on the recent loss rate data of each product and the following formula: (Yyogi is the total daily sales of each category of vegetables, Ichimi is the total daily replenishment of each category of vegetables, r is the price addition rate.)

$$I_i = (1 + r) * Y_i \tag{1}$$

Calculate the daily replenishment strategy for each vegetable category in the coming week.

3.2.3 Using single-objective income optimization model to formulate pricing strategy.

Optimization model refers to the mathematical model used to find the extreme value or maximum and minimum value of a function in practical work. The optimization model involves decision variables, objective functions and constraints. Decision variables refer to the quantities to be determined related to constraints and objective functions involved in the optimization problem. The objective function refers to the function which is related to the extreme value of the variable in the optimization problem. The constraint condition refers to the restriction that the variable must satisfy when finding the extreme value of the objective function in the optimization problem.

(1). Objective function.

In order to maximize the profit of Shang Chao, the income is understood as the income of accounting in this study. Accounting income refers to the difference between the realized income and the corresponding expenses from the transactions during the enterprise period. Therefore, the objective function of this study is listed as follows: (W is the merchant's super income, yfanti is the total daily sales of each single vegetable, zambii is the commodity pricing, ionomi is the daily replenishment of each single vegetable, and xonomi is the purchase price of the commodity.)

$$MaxW = \sum_{i=1}^{n} \mathrm{y}_i * z_i - \sum_{i=1}^{n} i_i * x_i \tag{2}$$

(2). Restriction condition.

This study assumes that the total sales of each category does not exceed the historical upper and lower limit of the total sales of this category, and that the addition rate of the cost-plus pricing in July 2023 does not exceed the historical upper and lower limit of the cost-plus pricing.

$$r_{min} \le r_i \le r_{max} \tag{3}$$

$$Y_{min} \le Y_i \le Y_{max} \tag{4}$$

The summary is as follows:

$$MaxW = \sum_{i=1}^{n} \mathrm{y}_i * z_i - \sum_{i=1}^{n} i_i * x_i \tag{5}$$

$$r_{min} \le r_i \le r_{max} \tag{6}$$

$$Y_{min} \le Y_i \le Y_{max} \tag{7}$$

(3). Decision variable.

In this study, using the functional relationship between the total sales and the cost addition rate, the cost is expressed as the total sales. Here, this study assumes that all the goods have been sold, that is, the total amount of sales is the same as that of replenishment after consumption. Therefore, the decision variable is only the total sales volume.

(4). Particle Swarm Optimization algorithm for solving Optimization Model.

Brief introduction of Particle Swarm Optimization algorithm. Particle Swarm Optimization (POS) is a random search algorithm based on group cooperation, which is developed by simulating the foraging behavior of birds. The particles in POS can also be said to be the solution of POS optimization problem. POS simulates the predation behavior of birds, and the solution of each optimization problem in POS is a particle (bird) in the search space. POS is initialized as a group of random particles, and the optimal solution is found continuously through iteration. when the maximum number of iterations or the global optimal position satisfies the minimum bound, the algorithm ends, and the optimal solution is found.

(5) Solution result.

The relevant forecast results are shown in Table 2. In order to better formulate the replenishment strategy and pricing strategy of vegetable commodities, this study believes that it is also necessary to consider the fresh-keeping time of each single vegetable, the correlation of each single vegetable and the fresh industry supply chain and other factors.

Table 2 Forecasting results

		July 1, 2023	July 2, 2023	July 3, 2023	July 4, 2023	July 5, 2023	July 6, 2023	July 7, 2023	Pricing (CNY / kg)
Aquatic rhizomes	Forecast sales volume	8.952413	24.41754	55.943295	24.430835	7.159353	17.085716	21.700448	9.14
	Loss rate	11.97473684	11.97473684	11.97473684	11.97473684	11.97473684	11.97473684	11.97473684	
	Purchase quantity	10.17027689	27.73924113	63.55368106	27.75434475	8.133293492	19.41001411	24.65252272	
Mosaic class	Forecast sales volume	48.17028	128.444443	132.214779	179.95264	118.605943	101.197647	79.236761	6.26
	Loss rate	10.2803	10.2803	10.2803	10.2803	10.2803	10.2803	10.2803	
	Purchase quantity	53.68974707	143.1619176	147.3642678	200.5720483	132.1960985	112.7931179	88.31590052	
Cauliflower	Forecast sales volume	16.012348	42.459578	43.3181	36.22927	17.05101	52.27622	35.851342	9.39
	Loss rate	14.142	14.142	14.142	14.142	14.142	14.142	14.142	
	Purchase quantity	18.64980316	49.4532577	50.45319015	42.19673181	19.85954716	60.8868364	41.75655384	
Eggplant	Forecast sales volume	7.096658	26.189217	28.552876	34.465614	16.276087	26.908486	19.018138	8.7
	Loss rate	7.122	7.122	7.122	7.122	7.122	7.122	7.122	
	Purchase quantity	7.640838519	28.19743858	30.74234587	37.10847994	17.5241575	28.97186201	20.47647236	
Capsicum	Forecast sales volume	31.146982	73.699851	125.726625	97.651848	53.805324	84.651149	56.829068	11.94
	Loss rate	8.515333333	8.515333333	8.515333333	8.515333333	8.515333333	8.515333333	8.515333333	
	Purchase quantity	34.04612285	80.55978525	137.4291776	106.7412186	58.8134886	92.53042295	62.11868073	
Edible fungi	Forecast sales volume	17.685928	34.298182	64.218462	38.331284	20.294293	32.734581	22.038155	8.99
	Loss rate	8.130972222	8.130972222	8.130972222	8.130972222	8.130972222	8.130972222	8.130972222	
	Purchase quantity	19.25124106	37.33378139	69.90218962	41.72383765	22.09046236	35.63179212	23.98866684	

4. Discussion

4.1. Fresh-keeping time of individual vegetables.

As in the supermarket of fresh food, the general vegetable products are kept fresh for a short time, and the appearance will get worse with the increase of sales time, so most of the varieties that have not been sold out on the same day cannot be sold again the next day. Therefore, to master the fresh-keeping time of each single vegetable is particularly important for replenishment strategy and pricing strategy. This study can be divided into two cases according to the relationship between vegetable fresh-keeping time and replenishment time. The first situation is that the fresh-keeping time of individual vegetables is longer than that of replenishment. In this question, the variety that has not been sold out on the same day can continue to be sold the next day. And this study can also work out the discount rate and pricing strategy according to the fresh-keeping time of individual vegetables and converting the residual value of individual vegetables. On the other hand, the fresh-keeping time of a single vegetable is lower than that of replenishment. This single vegetable cannot be sold the next day. Therefore, in the replenishment strategy, this kind of individual vegetables can be purchased a little less. In this study, it is considered that the fresh-keeping time of this single product is longer than that of replenishment. And in this study, according to its discount strength, speculate its fresh-keeping time, according to its fresh-keeping time to determine its replenishment frequency. According to the analysis, the value loss rate of mosaic vegetables is lower in all kinds of vegetables. Therefore, this study suggests that more mosaic vegetables can be added in the replenishment strategy.

4.2. Correlation of individual vegetables.

Correlation refers to the degree of correlation between two variables. Generally speaking, it can be observed from the scatter diagram that the two variables have one of the following three relationships: positive correlation, negative correlation and irrelevance. In economics, this study usually refers to two types of goods with positive correlation as complements and two types of goods with negative correlation as substitutes. Complementarity means that there is a dependent relationship between the consumption of two types of goods, and consumers usually use it together to meet the needs of consumers. Usually when the demand for a commodity increases, so does the demand for its complements. Substitutes are goods that can be substituted by both of them, which can meet the same type of needs of consumers. Usually when the demand for a commodity increases, the demand for its substitutes will decrease. In question 1, the correlation of individual vegetables was obtained in this study. This study believes that the replenishment strategy of each single vegetable can be formulated according to the correlation of each single vegetable. When the replenishment of a positively related single vegetable increases, so does the replenishment of the commodity. When the replenishment of a single vegetable negatively related to a single vegetable increase, the replenishment of the commodity decreases. And vice versa.

4.3. Fresh industry supply chain.

The circulation loss of fresh vegetables refers to the physical loss caused by theft and loss of products from origin to consumer market caused by inadequate infrastructure, backward operation technology and poor management. as well as the cost input caused by the incidental displacement of impurities in the process of circulation, the price difference caused by information distortion, the devaluation of products caused by oversupply, and the seizure and invalidation of harmful products. The financial calculation of the loss rate is distorted and the corresponding valuable losses such as garbage disposal, environmental protection and market management costs. Therefore, this study suggests that the supply chain of fresh industry should be shortened, direct sales of origin should be adopted, the number of upstream suppliers should be reduced, and high-quality upstream suppliers should be selected. Improve the market system, create a mature upstream production organization, and promote the industrial intensification of the agricultural industry. Dredge traffic jams, optimize logistics management, shorten the transportation time of fresh food, and reduce the physical loss of fresh food itself.

5. Conclusion.

Through studying the problem of vegetable replenishment and pricing strategy in fresh merchant excess, this paper finds that mastering the distribution law and interrelationship of vegetable sales volume of various categories and individual products and using time series prediction model and optimization model to formulate replenishment quantity and pricing strategy, can make the merchant super get the maximum profit. In addition, the fresh-keeping time of individual vegetables is also an important factor affecting replenishment strategy and pricing strategy. In the future research, we can further explore the impact of other factors on the replenishment and pricing strategy of fresh vegetables, in order to improve the operating efficiency and revenue.

References

[1] Yiyang Q, Minseong K, Byeongil A. Analysis of Factors Affecting Vegetable Price Fluctuation: A Case Study of South Korea[J]. Agriculture,2023,13(3):577-577.

[2] Hongyu Y, Yuanxin C, Yuemeng S, et al.The Dynamic Impacts of Weather Changes on Vegetable Price Fluctuations in Shandong Province, China: An Analysis Based on VAR and TVP-VAR Models[J].Agronomy,2022,12(11):2680-2680.

[3] Chen S, Liang C, Ximeng W, et al.A spatial network analysis of vegetable prices based on a partial granger causality approach[J].Frontiers in Physics,2022,10.

[4] Kentaka A, Monirul M I, Arifa J. Effects of the State of Emergency during the COVID-19 Pandemic on Tokyo Vegetable Markets[J]. Sustainability,2022,14(15):9719-9719.

[5] Youzhu L, Miao Z, Jinsi L, et al. Research on the Disturbance Sources of Vegetable Price Fluctuation Based on Grounded Theory and LDA Topic Model[J]. Agriculture,2022,12(5):648-648.

[6] C. S H, L. K W, Ron S, et al.Comparison of Fruit and Vegetable Prices between Farmers' Markets and Supermarkets: Implications for Fruit and Vegetable Incentive Programs for Food Assistance Program Participants[J].Nutrients,2022,14(9):1842-1842.

[7] Bixuan Y, Frank A, Tao L. Consumer behavior and food prices during the COVID-19 pandemic: Evidence from Chinese cities[J]. Economic Inquiry,2022,60(3):1437-1460.

[8] Lisha M, Yin H, Xiaofan Z, et al. ARIMA model forecasting analysis of the prices of multiple vegetables under the impact of the COVID-19. [J]. PloS one,2022,17(7): e0271594-e0271594.

[9] Rosemary N, Ayalew B. Fuzzy Logic based Vegetable Price prediction in IoT[J]. Procedia Computer Science,2022,203807-812.

[10] A R, D N. Vector error correction model relationship between three vegetable oil products[J]. IOP Conference Series: Earth and Environmental Science,2021,892(1).