Automated pricing and replenishment forecasting for vegetable items

Mingyang Cai

Shandong University of Science and Technology, Qingdao, China

13176001935@163.com

Abstract. Firstly, the thesis extensively analysed and processed the vegetable sales data of a fresh food superstore. It integrated the product information and sales data of each vegetable category, and carried out data cleaning and analysis through Excel tools, including sorting out the outliers, listing the sales volume of each of the six categories, using pivot tables for statistical analysis, and establishing line graphs, heat maps, scatter plots, and other visual displays of the data characteristics and correlations. Secondly, the thesis investigated the relationship between sales volume and pricing, fitted the correlation between total sales volume and pricing using a random forest regression model, and predicted the daily replenishment volume for the coming week using an LSTM time series model. It also fitted quantitative fitting formulas for metrics for each vegetable category, classified profits into normal sales volume profits and discounted sales volume profits, and ultimately maximised profits through dynamic programming models and particle swarm optimisation algorithms. Finally, the thesis considered the impact of other factors such as weather, holidays, seasons and market environment on sales volume with specific data collection and analyses that help guide better pricing strategies. The findings will contribute to a better understanding of vegetable sales behaviour and optimise superstore operational strategies.

Keywords: random forest regression, LSTM time series model, dynamic programming models, particle swarm optimisation algorithms

1. Introduction

In the rapid development of modern cities, fresh produce superstores have become an important part of the daily life of urban residents. As the main channel for providing fresh vegetables, fruits and other essential foods, fresh produce superstores play a crucial role in safeguarding the dietary health of urban residents. As people continue to pursue food quality and quality of life, the market demand for fresh produce continues to grow. However, there are quite a few challenges hidden behind this trend, especially in the supply chain management of vegetables. As one of the most common fresh products, vegetables are characterised by a high diversity of categories, a short shelf life and a fast sales cycle. These characteristics complicate vegetable supply chain management and make it a difficult task for fresh produce superstores to sell and replenish vegetables effectively.

Firstly, vegetables have a short shelf-life and are susceptible to environmental influences, and any small negligence in the process can lead to significant losses. This requires supermarkets to handle each batch of vegetables with great care, from procurement, transport, storage to display, all requiring strict quality control. Secondly, seasonal changes in vegetable sales bring another layer of complexity to

supermarkets' purchasing decisions. Different vegetable categories have significant differences in their adaptability to climate, which directly affects their growth cycle and market supply. During different seasons, certain vegetables may be in abundant supply and inexpensive, while other varieties may be scarce and expensive. Supermarkets need to adjust their replenishment strategies according to these seasonal variations in order to maintain the diversity and freshness of the items on their shelves. Further, the inter-relationship between vegetable categories is an important factor to consider. There may be alternative or complementary selling relationships between certain vegetables. For example, if there is a shortage of one type of leafy vegetable, consumers may switch to other types of leafy vegetables. Understanding the interactions between these vegetables can help supermarkets better predict consumer behaviour and thus make more accurate replenishment decisions. In addition, the uncertainty of the price of incoming vegetables adds to the difficulty of supply chain management. Due to a variety of factors such as weather, transport costs, and market supply and demand, the purchase price of vegetables may fluctuate significantly. This poses a challenge to the cost control and pricing strategies of supermarkets and directly affects their profitability.

In view of the above challenges, the research objective is to analyse vegetable sales data in depth and explore the sales volume distribution patterns, seasonal variations and their interrelationships among different categories. Through a detailed study of historical sales data and wastage rates, the thesis hope to provide fresh produce superstores with insights that will help them improve their replenishment decisions and pricing strategies. Specifically, the thesis plan to apply statistical and data analysis methods to seasonally disaggregate the sales records of various vegetables in order to identify and predict sales trends within different seasons or time periods [1]. In addition, by exploring the correlation between vegetable categories, the thesis expect to identify potential substitution or complementary effects, which will provide a basis for superstores to optimise their vegetable assortment mix [2].

Through the above research, the thesis aim to provide scientific, data-driven management advice to fresh food superstores to help them cope with changing market demands and supply chain challenges. the thesis believe that by improving sales efficiency, reducing wastage, and ultimately providing more quality food to consumers, it can add competitiveness to the fresh produce superstore business and achieve sustainable growth. Ensuring the success of this study requires not only deep industry understanding and efficient data analysis techniques, but also the ability to closely integrate with actual marketing strategies. The thesis plan to work closely with the industry partners to ensure the practical application of the findings [3]. Ultimately, the thesis expect this study to help fresh produce superstores navigate the complex and dynamic market environment, and assist them in making solid strides in safeguarding food safety and improving consumer satisfaction [4].

2. Preliminary Analysis of Data

2.1. Descriptive statistical analysis of data

The thesis have six vegetable categories of product information and sales flow detail data, through the two tables of data merging and data preprocessing, eliminating outliers (sales of negative rows), missing values interpolation to 0, filtering out the sales of data to remove the return data [5].

The sales volume of each category of vegetables is counted on a monthly basis and the following table is obtained:

Table 1. Monthly sales volume of vegetables by category.

Vegetable categories	mean	variance	range	sum	standard deviation
Cauliflower vegetables	1160.82714	198145.519	12486.978	41789.777	438.9094
Mosaic vegetables	5518.32092	2967578.93	159.997	198659.553	1698.57181

Table 1. (continued)

Pepper vegetables	2670.69756	1117760.94	10.01	91645.112	1042.45481
Nightshade vegetables	623.392194	80742.6115	12.45	22442.119	280.178088
Edible mushroom vegetables	2114.77016	941424.044	24.989	76131.726	956.699209
Aquatic root vegetables	1127.98756	417120.542	16.95	40607.552	637.122965

The standard deviation and extreme deviation fluctuate greatly when the unit is day, reflecting a large degree of data dispersion, which is unsuitable for correlation analysis, so the following statistics are analysed on a monthly basis [6].

2.2. Distribution patterns and interrelationships of sales volume of vegetables by category
Firstly, Matlab was used to visually represent the monthly sales volume of each category of vegetables for 36 months and the following line and scatter plots were obtained:
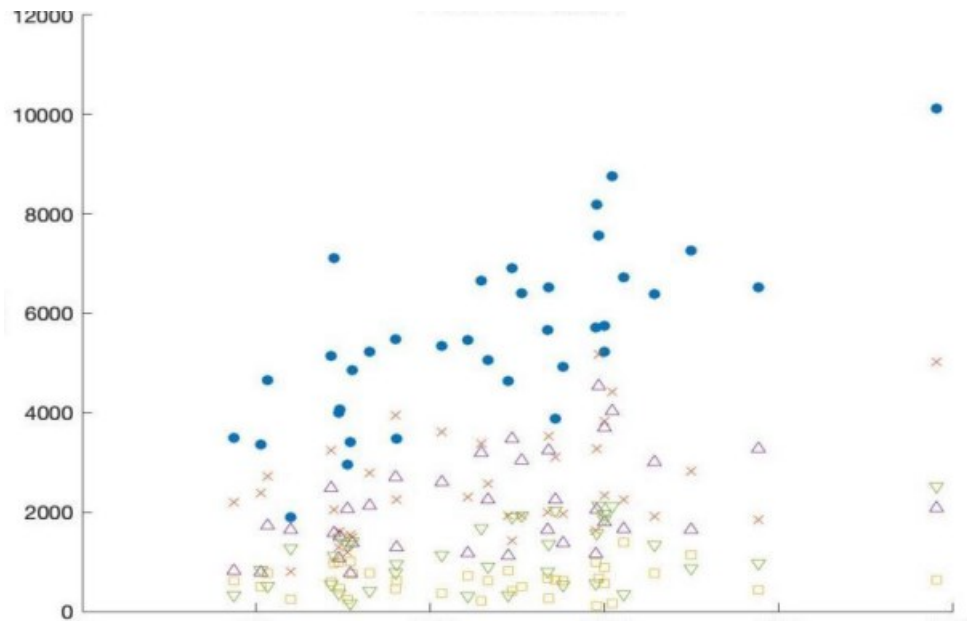

Figure 1: Scatterplot of monthly sales volume of vegetables by category for 36 months.

From the figure, it can be seen that the monthly sales of certain categories have the same fluctuation with the change of month, and the sales of certain categories are correlated with each other, and so carries out the correlation coefficient analysis for the correlation of the sales of categories with each other.

Correlation coefficient analysis:

First make a heat map for visualisation and then construct a matrix of correlation coefficients.
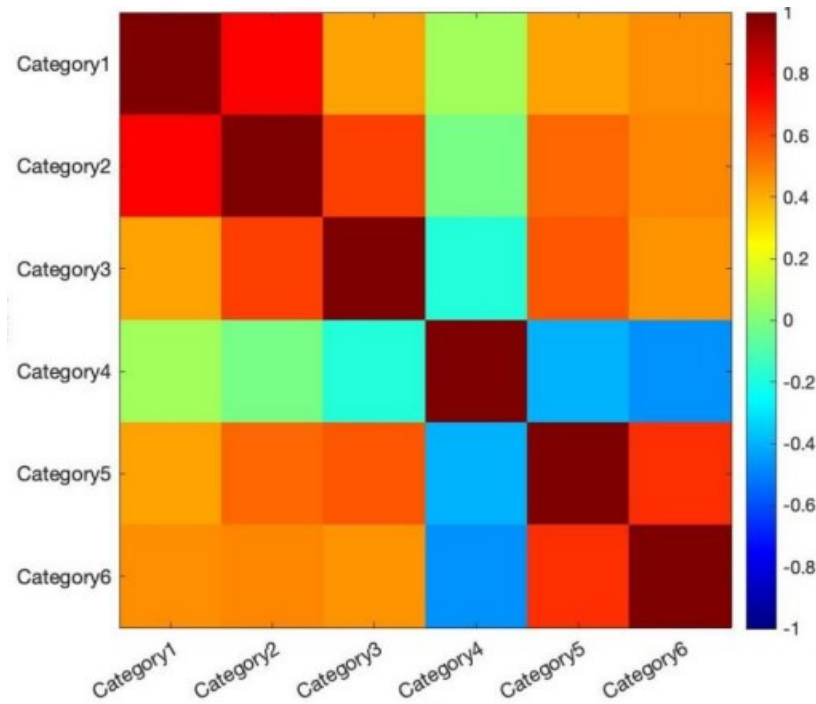
Figure 2: Heat Map of Seasonal Analysis of Sales Volume of Vegetables by Category.

Table 2. Correlation coefficient matrix.

1.0000	0.7486	0.4240	0.0687	0.4271	0.4676
0.7486	1.0000	0.6214	-0.0228	0.5432	0.4788
0.4240	0.6214	1.0000	-0.1875	0.5742	0.4567
0.0687	-0.0228	-0.1875	1.0000	-0.4047	-0.4652
0.4271	0.5432	0.5742	-0.4047	1.0000	0.6514
0.4676	0.4788	0.4567	-0.4652	0.6514	1.0000

It can be seen that there is a high degree of correlation between sales of foliar and flowering stem categories, and a weak negative trend for eggplant and aquatic rhizomes.

3. Optimal Solutions for Profit Maximisation in Each Vegetable Category

3.1. Relationship between total sales and "cost-plus pricing" for each vegetable category
First of all, the total sales of each vegetable category, from the background information, the thesis can know that "cost-plus pricing" refers to the commodity pricing, and then statistics on the pricing of each vegetable category, the use of random forest regression model, the construction of the decision tree is still based on the recursive partitioning nodes, usually use the following formula to select the best segmentation features decision tree construction based on the The decision tree is constructed based on the recursive segmentation nodes, and the following formula is usually used to select the best segmentation features[7].

$$Gini(D)=1-\sum_{K=1}^{K}(PK)^2$$

$$Entropy(D)=-\sum_{K=1}^{K} PK \, log_2 \, PK$$

With this algorithm, a plot of the relationship between total sales and pricing is obtained for each vegetable category (Y represents total sales and X represents pricing)
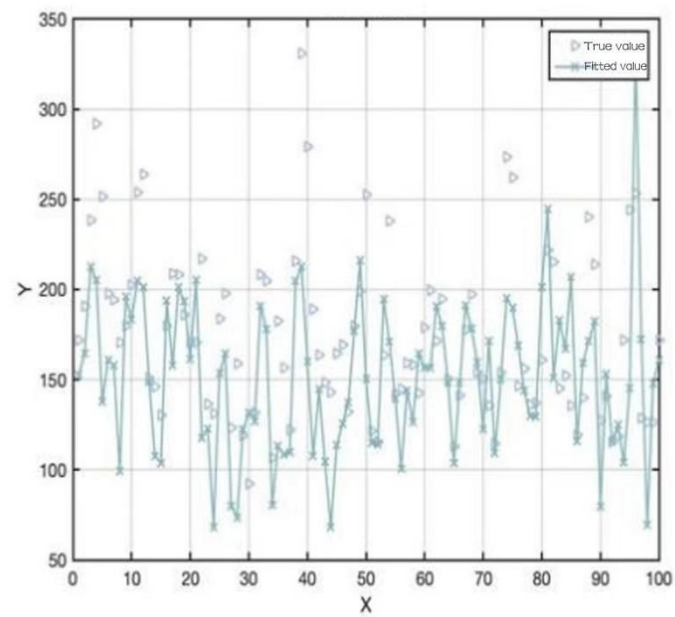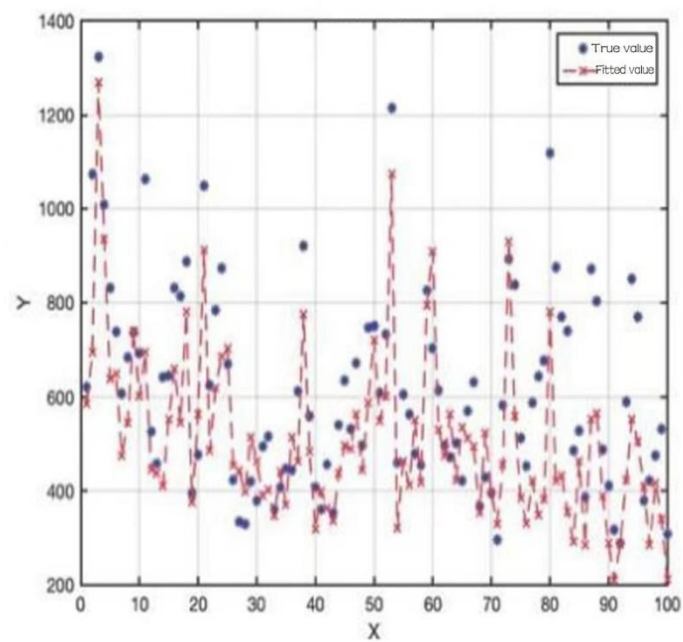
Figure 3: Total sales and pricing of mosaic vegetables.

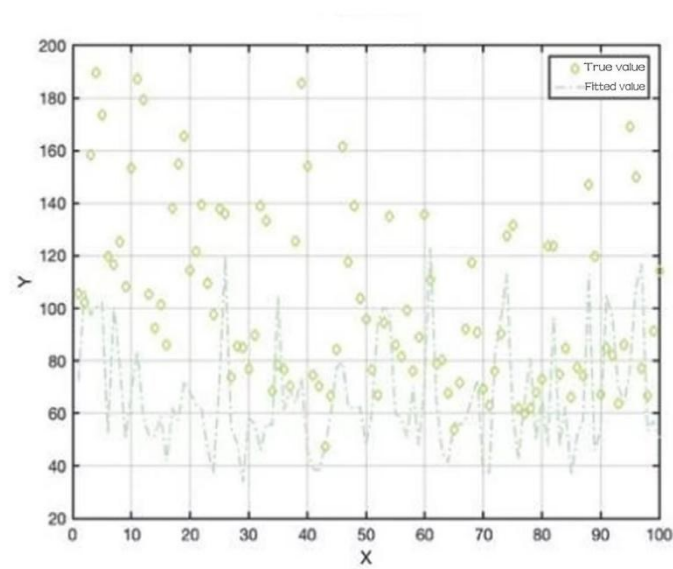**Figure 4:** Total sales and pricing of cauliflower vegetables.

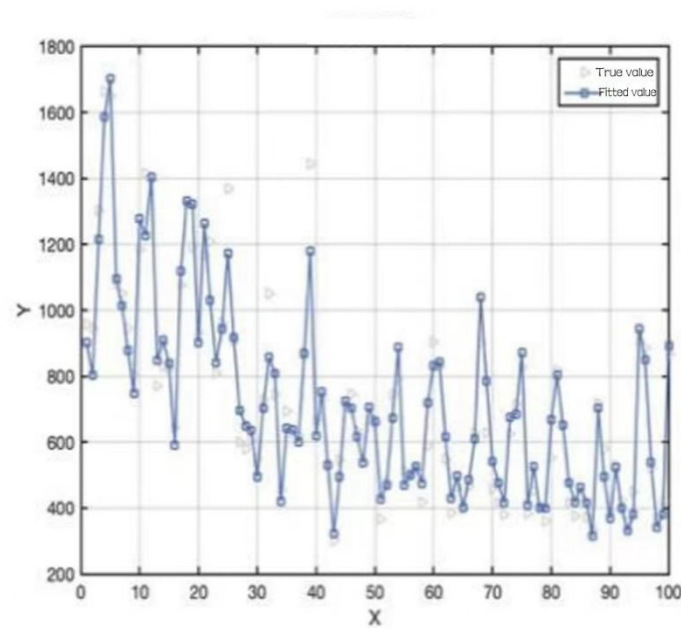**Figure 5:** Total sales and pricing of pepper vegetables.

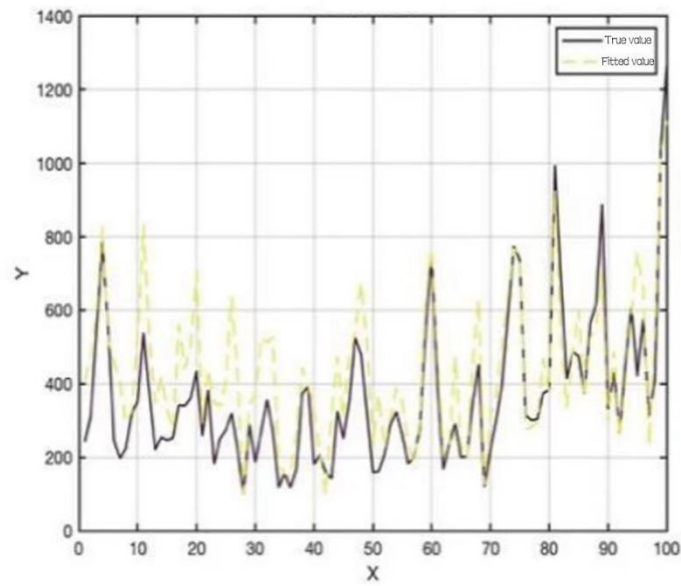**Figure 6:** Total sales and pricing of edible mushroom vegetables.

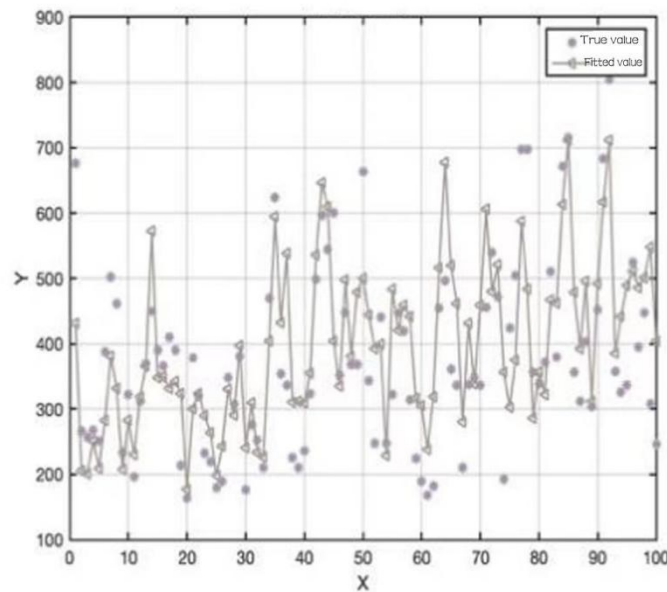**Figure 7:** Total sales and pricing of aquatic root vegetables.

Figure 8: Total sales and pricing of nightshade vegetables.

Observation of the relationship between X and Y shows that the total sales of vegetable category has a weak negative relationship with cost plus pricing, based on this pattern, it helps to formulate the sales strategy accordingly.

3.2. Forecast of total daily replenishment for each vegetable category for the coming week
To predict the total daily replenishment for the coming week firstly the total daily sales for the coming week is predicted then the corresponding wastage rate is considered. The data for July 2020-2022 and June 2023 are selected and aggregated to predict the total sales for the coming week using the LSTM

time series model, and then the LSTM time series model is optimised using the White Shark optimisation algorithm [8].

The formula for calculating the amount of incoming goods:

$$U=S\times(1+K\%)$$

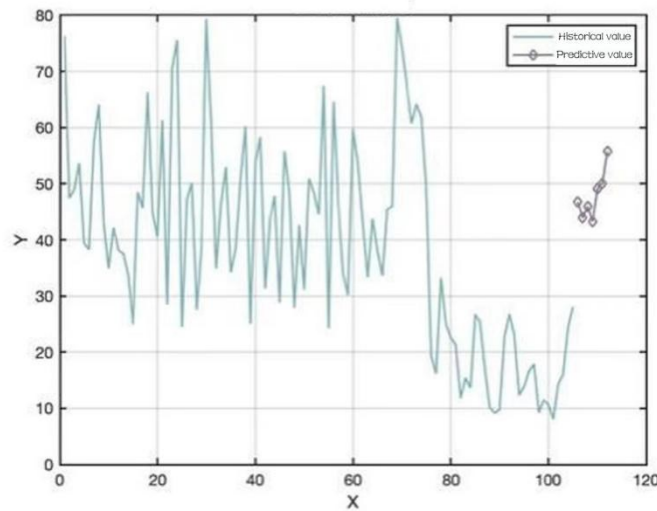It is possible to obtain a graph of the predicted sales of each type of vegetable dish:

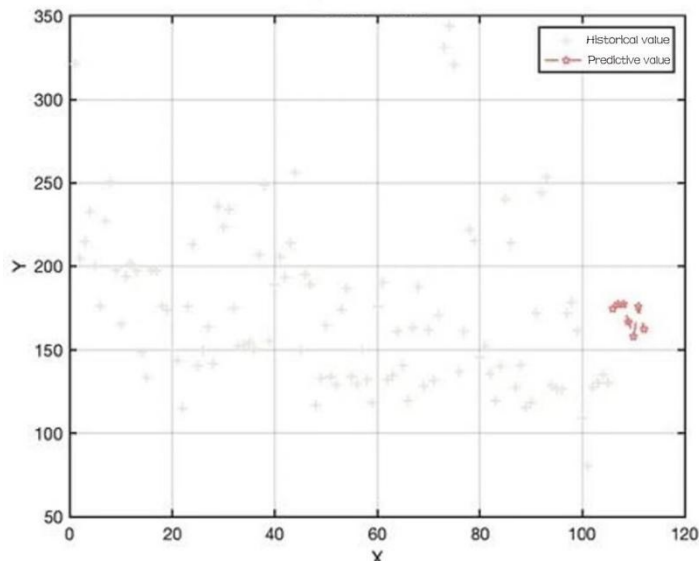**Figure 9:** Sales of mosaic vegetables for the week ahead.

Figure 10: Sales of cauliflower vegetables for the week ahead.

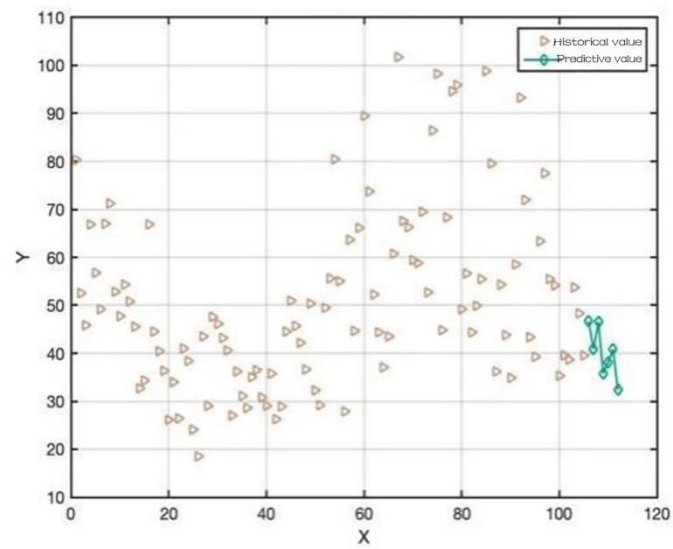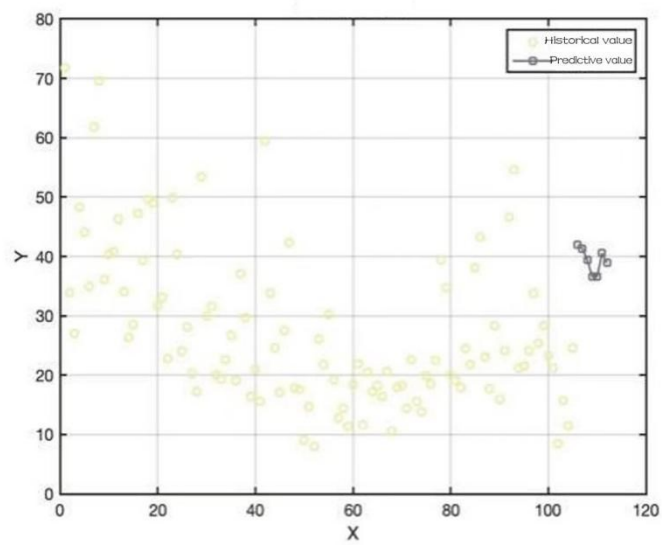**Figure 11:** Sales of edible mushroom vegetables for the week ahead.

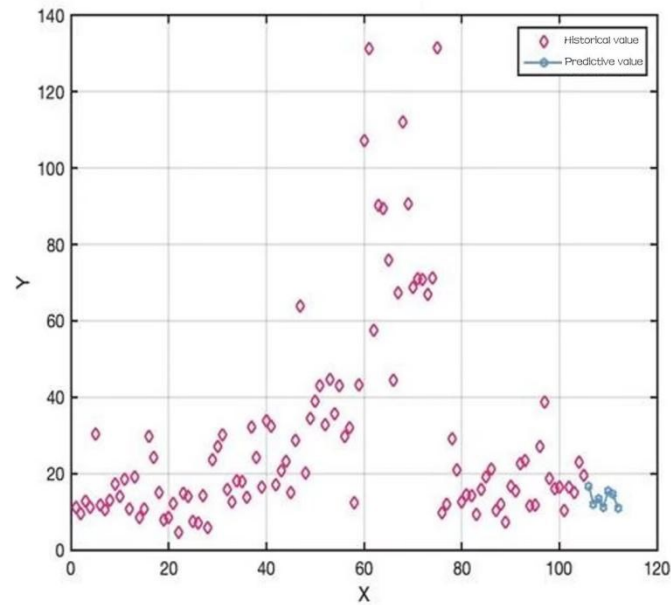**Figure 12:** Sales of nightshade vegetables for the week ahead.

Figure 13: Sales of aquatic root vegetables for the week ahead.
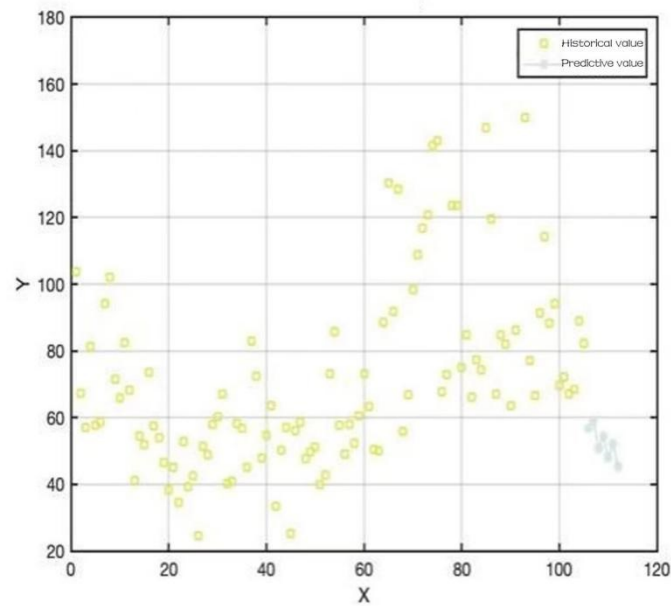

Figure 14: Sales of pepper vegetables in the week ahead.

The total daily replenishment for each vegetable category for the coming week can be calculated as shown in the table below:

Table 3. Total daily replenishment by vegetable category for the coming week.

dates / kind	2023/7/1	2023/7/2	2023/7/3	2023/7/4	2023/7/5	2023/7/6	2023/7/7
Cauliflower vegetables	51.882	48.754	51.105	48.047	54.463	55.539	61.910
Mosaic vegetables	196.053	198.909	199.154	187.246	177.234	197.718	182.410
Pepper vegetables	61.262	63.775	54.583	58.6210	51.603	51.603	48.911
Nightshade vegetables	44.722	43.956	41.901	38.943	38.996	43.140	41.399
Edible mushroom vegetables	50.504	44.183	50.281	38.707	41.224	44.1478	34.975
Aquatic root vegetables	19.017	13.696	15.242	12.734	17.455	16.513	12.510

3.3. Quantitative fitting of indicators for the vegetable category

Table 4. Monthly distribution of sales volume by category.

Kind	Trend of sales volume distribution by category by month	R^2
Cauliflower vegetables	$y = 1.0081x^2 - 51.148x + 1652.6$	$R^2 = 0.1563$
Mosaic vegetables	$y = 6.4729x^2 - 242.68x + 7090.1$	$R^2 = 0.1356$
Pepper vegetables	$y = 3.0175x^2 - 60.953x + 2313.4$	$R^2 = 0.3333$
Nightshade vegetables	$y = 0.6592x^2 - 34.921x + 972.39$	$R^2 = 0.2038$
Edible mushroom vegetables	$y = 3.6404x^2 - 123.06x + 2751.3$	$R^2 = 0.1508$
Aquatic root vegetables	$y = -1.1353x^2 + 46.048x + 786.43$	$R^2 = 0.0339$

Table 5. Trend of profit distribution by category by month.

Kind	Trend of profit distribution by category by month	R2
Cauliflower vegetables	$y = 10.21x^2 - 342.64x + 4973$	$R^2 = 0.1664$
Mosaic vegetables	$y = 11.094x^2 - 521.06x + 901.3$	$R^2 = 0.0525$
Pepper vegetables	$y = 6.8613x^2 - 538.17x + 12354$	$R^2 = 0.1362$
Nightshade vegetables	$y = 0.7064x^2 + 2.3076x + 962.46$	$R^2 = 0.0492$
Edible mushroom vegetables	$y = 8.5661x^2 - 488.31x + 9483.4$	$R^2 = 0.3398$
Aquatic root vegetables	$y = 4.7643x^2 - 151x + 3337.4$	$R^2 = 0.0643$

Table 6. Monthly trends in post-discount profits by category.

Kind	Trend in distribution of discount margins by month	R^2
Cauliflower vegetables	$y = -0.2413x^2+1.6115x + 141.79$	$R^2 = 0.146$
Mosaic vegetables	$y = 0.0024x^2-65.077x + 356.53$	$R^2 = 0.347$
Pepper vegetables	$y = 1.0697x^2-68.662x + 563.85$	$R^2 = 0.4521$
Nightshade vegetables	$y = 0.76x^2-15.324x + 56.426$	$R^2 = 0.0791$
Edible mushroom vegetables	$y = -3.9066x^2+105.07x - 85.911$	$R^2 = 0.2848$
Aquatic root vegetables	$y = 0.5545x^2-7.365x + 35.9$	$R^2 = 0.0492$

Table 7. Monthly distribution of post-discount sales volume by category.

Kind	Discounted sales volume trends by month	R^2
Cauliflower vegetables	$y=-0.3666x^2+10.901x+21.208$	$R^2 = 0.1219$
Mosaic vegetables	$y =0.4542x^2+4.9233x+76.852$	$R^2 = 0.2782$
Pepper vegetables	$y =0.0627x^2+9.871x-13.137$	$R^2 = 0.3928$
Nightshade vegetables	$y =-0.7796x^2+12.41x+4.9038$	$R^2 = 0.0707$
Edible mushroom vegetables	$y =0.3687x^2-1.0047x+28.983$	$R^2 = 0.6077$
Aquatic root vegetables	$y = -1.0246x^2+30.202x-23.543$	$R^2 = 0.1222$

3.4. Daily pricing strategy for each vegetable category for the week ahead

Based on the above data, a dynamic programming model is constructed to maximise the seven-day total profit W of the superstore, where the profit can be divided into normal sales profit and discount sales profit [9][10]. The specific modelling process is as follows:

(1) Decision variable: Quantitative polynomial fitting formula for each indicator in the vegetable category

(2) Objective function: Maximise profit of superstores

$$W=P\times N\times(1-K\%)+P\times N\times(1-K\%)$$

(3) restrictive condition

Constraint 1: Discount margin P usually has an upper and lower limit to ensure that the price is reasonable.

Constraint 2: Non-discounted profit P-runs usually have an upper and lower limit to ensure that they are reasonably priced.

Constraint 3: positive loss rate k per cent

Constraint 4: Non-discounted sales in each category N is restricted

Constraint 5: Discounted sales in each category N is restricted

Global optimal solution using particle swarm algorithm (PSO)

The optimal solution for pricing of each dish category is calculated using Matlab as shown in the table below:

Table 8. Optimal solution for pricing each dish.

Times	Cauliflower vegetables	Cauliflower vegetables	Cauliflower vegetables	Cauliflower vegetables	Cauliflower vegetables	Cauliflower vegetables
7.1	9.9654	6.5276	9.9488	7.2366	12.8322	14.2328
7.2	9.9835	6.7623	10.2310	7.0261	12.8921	13.9786
7.3	11.0013	6.8576	9.5433	6.9261	12.4355	13.6723
7.4	10.0469	5.9645	9.5288	6.2199	11.9235	15.0816
7.5	8.9789	7.1524	8.9544	7.6722	12.6732	14.2355
7.6	10.0325	6.5438	9.9773	7.0599	13.0278	14.8603
7.7	10.2699	5.8235	10.0542	6.5383	13.0613	14.6273

4. Impact Of Other Fators On Sales Volume

4.1. Holiday factors

By filtering out the holidays: New Year's Day (1 day off on 1 January), Spring Festival (3 days off from the first to the third day of the first month of the lunar calendar), Qingming Festival (1 day off on the same day as Qingming on the lunar calendar), Labour Day (1 day off on 1 May), Dragon Boat Festival (1 day off on the same day as Dragon Boat Festival on the lunar calendar), Mid-Autumn Festival (1 day off on the same day as Mid-Autumn Festival on the lunar calendar), and National Day (3 days off on the 1st to the 3rd day of the 10th October), and comparing them with the average of the usual sales, as shown below:
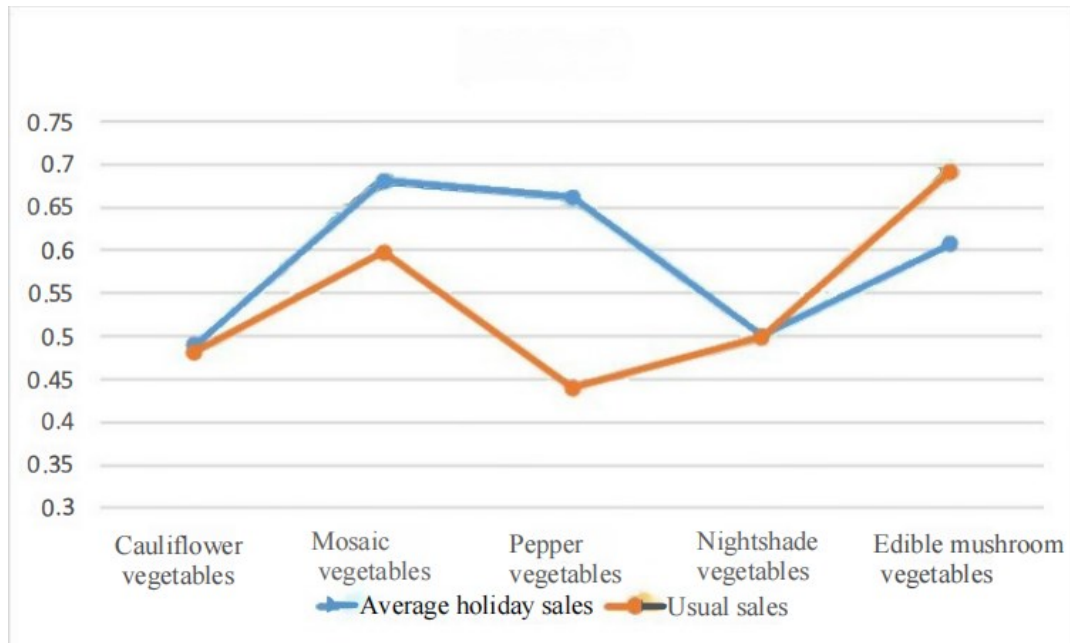

Figure 15: Comparison of holiday sales with usual sales.

It can be seen that the average holiday sales of each category are generally higher than the usual sales, and the average holiday sales of the chilli category are significantly higher than the usual sales, indicating that people generally love to eat heavy vegetables such as chilli during holidays. From this,

it is a good way to go to restaurants, food markets and other places to collect data related to the sales volume of chilli.

4.2. Weather factors

Vegetable price volatility affects supply chains and consumer choice. Adverse weather can lead to reduced yields, increase storage and transport difficulties, and affect consumer demand. Traders have to make pricing decisions taking into account costs, changes in market demand and the impact of seasonality.

1. Availability: Weather conditions (e.g., droughts, floods, storms) affect vegetable production and may lead to price increases. Changes in production costs and changes in market demand need to be factored into pricing considerations.

2. Storage and transport: Certain vegetables require special handling to maintain quality and freshness. Adverse weather can damage the supply chain and make storage and transport more difficult, leading to unstable supplies and higher costs.

3. Consumer demand: Different weather conditions affect consumer demand for different types of vegetables. Summer may favour cool, water-rich vegetables, while winter requires high-energy, high-calorie vegetables. Traders should adjust their replenishment and pricing strategies according to seasonal demand.

5. Summary

This paper provides scientific management suggestions for fresh produce superstores through the study of vegetable sales and replenishment forecasting, using data integration, descriptive statistics and correlation analysis. The research results show that vegetable sales volume by month is more suitable for correlation analysis, and the sales volume of each category has a weak negative correlation with pricing. By building a prediction model, the daily replenishment volume in the coming week can be effectively predicted and profit can be improved with the help of optimisation algorithms. These findings are crucial for fresh produce superstores to navigate through the market environment and improve consumer satisfaction [11].

This study aims to provide scientific, data-driven management recommendations for fresh produce superstores to help address market demand and supply chain challenges. By improving sales efficiency, reducing wastage, and ultimately providing more quality food, fresh produce superstores can enhance their competitiveness and achieve sustainable growth. In addition, successful implementation of this study requires not only deep industry understanding and efficient data analysis techniques, but also close integration with actual marketing strategies. The team plan to work closely with the industry partners to ensure that the results of the study can be effectively applied. The team look forward to helping fresh produce superstores navigate the complex market environment through this study, helping them to safeguard food safety and enhance consumer satisfaction[12].

Referances

[1]   Tang Guosheng. Exploration on the Construction of Innovative Practice Base of Mathematical Modelling for College Students Based on SWOT Analysis--Taking Jiangsu University of Science and Technology as an Example[J]. Journal of Higher Education, 2023, 9(11): 53-56. DOI: 10. 19980/j.CN23- 1593/G4.2023.11.013.

[2]   Xie Yujing,Han Huili. A multidimensional analysis of mathematical modelling in the junior high school mathematics textbook of Bei Shi Da edition[J]. Teaching and Management, 2023(09):73-76.

[3]   LIU Zhimei. The deep integration of mathematical modelling and higher vocational mathematics teaching[J]. Journal of Jiamusi Vocational College,2023,39(03):152-154.

[4]   Xu Yatao,Wu Libao. Research on the evaluation index system of teaching mathematical modelling in high school based on Delphi-AHP[J]. Journal of Neijiang Normal College,2023,38(02):113-119.DOI:10.13603/j.cnki.51- 1621/z.2023.02.018.

[5]   YANG Benzhao, SHI Yanan, DUAN Qianheng, LI Guangsong, YU Gang. Exploration of full-cycle teaching practice for university students' mathematical modelling competition[J]. University Education,2023(04):44-46.

[6]   Huang Jian,Xu Binyan. Development trend of teaching and learning research on mathematical modelling in international perspective - an analysis based on the 14th International Congress on Mathematics Education[J]. Journal of Mathematics Education,2023,32(01):93-98.

[7]   Li Shuai. Research on commodity pricing and ordering strategy considering strategic buying behaviour in e-commerce environment[D]. Yanshan University,2018.

[8]   Hong Qihan. Optimal design of antenna based on improved simulated annealing and white shark algorithm[D]. Donghua University,2023.DOI:10.27012/d.cnki.gdhuu.2023.001164.

[9]   Braik M,Hammouri A,Atwan J,et al. White Shark Optimizer:A novel bio-inspired meta-heuristic algorithm for global optimisation problems[J]. Knowledge-Based Systems, 2022,243:108457.

[10]   Zeng, Minmin. Research on Dynamic Pricing Strategy of Fresh Community Supermarket Based on Time Situation A[D]. Southwest University of Finance and Economics, 2021. DOI:10.27412/d.cnki.gxncu.2021.002858.

[11]   Zhang He. Vegetable sales prediction based on integrated learning[D]. Yunnan Normal University,2021.DOI:10.27459/d.cnki.gynfc.2021.000144.

[12]   MENG Fanrong,ZHANG Peng,YAN Qiuyan. A data flow prediction model based on weighted moving average[J]. Computer Application Research,2009,26(10):3680-3682+3686.