ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

Supermarket Business Strategy Based on Time Series Forecasting Model and Genetic Algorithm

Aoyuan Zhang[1,*], Yi Tian[2], Qiang Dai[3]

[1]College of Life Sciences, Northwest A&F University, Yangling, China
[2]College of Information Engineering, Northwest A&F University, Yangling, China
[3]College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, China
[*]Corresponding author: zay020923@nwafu.edu.cn

Abstract: *This paper stands for the perspective of vegetable superstores and provides an in-depth analysis of historical operational data. Through data preprocessing, data visualization and Spearman's correlation analysis, the sales distribution patterns of vegetable categories and individual items were revealed, and the vegetable replenishment and pricing strategies for the coming week and day were further formulated by using nonlinear fitting and seasonal ARIMA models. The results of the study showed that leafy and flowering vegetables contributed the most to the sales of the superstore, followed by chili peppers and edible mushrooms; meanwhile, the optimal replenishment and pricing strategies for each individual item were obtained through genetic algorithms. These analyses provide decision support for the superstore and help maximize revenue while meeting market demand.*

Keywords: *Vegetable Superstores, Data Analysis, Time Series Modeling, Pricing Strategies*

1. Introduction

Faced with the problem of replenishment and pricing decisions in fresh produce superstores in the complex context of perishability and seasonal supply fluctuations, appropriate strategic choices have become a key factor in optimizing operations and improving profits. This study employs a series of quantitative methods to comprehensively analyze the sales dynamics, sales patterns, and interrelationships among categories of six major vegetable commodities based on historical sales and wholesale data from July 1, 2020, to June 30, 2023, respectively. This paper will also delve into the effectiveness of the cost-plus pricing strategy applied to vegetable sales and its potential correlation with total sales volume. Through these analyses, this study aims to develop a more scientific and rational daily replenishment and pricing plan with a view to maximizing the income from supermarket operations [1,2].

2. Sales patterns and data analysis

2.1 Distributional characteristics of merchandise sales volume

The content of this section focuses on exploring the sales patterns and interconnectivity of six major categories of vegetables (leafy flowers, cauliflower, aquatic roots and tubers, eggplants, peppers, and edible mushrooms) and their related individual products in fresh produce supermarkets. This study provides an in-depth analysis in both vertical and horizontal dimensions. The study focuses on exploring the sales patterns and interconnectivity of the six major categories of vegetables (leafy vegetables, cauliflower, aquatic roots and tubers, eggplants, chili peppers and edible mushrooms) and their related individual items in fresh produce supermarkets. The study provides an in-depth analysis in both vertical and horizontal dimensions [3].

In this paper, when studying the distribution pattern of the six vegetable categories, we first made a cross-sectional comparison, and visualized the three-year sales totals and shares of the six vegetable categories based on the preprocessed data, as shown in Figures 1 and 2.

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

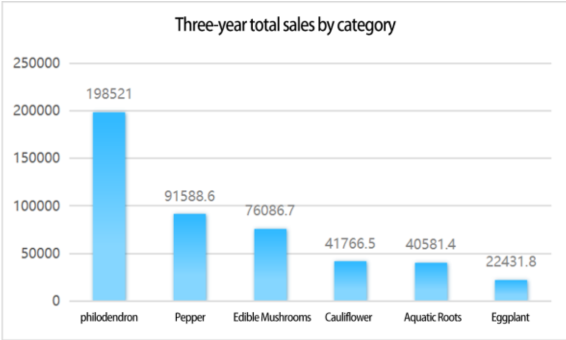

Figure 1: Three-year total sales by category

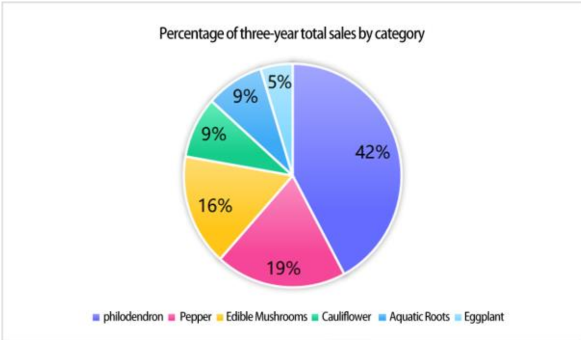

Figure 2: Three-Year Total Sales Share by Category

From the figure, the horizontal distribution flower and leafy vegetable items contribute the most to the sales volume of this superstore, followed by chili and edible mushroom vegetables, and eggplant vegetables have the least total sales volume. It indicates that in the last three years, the target customers of this superstore have the strongest desire to consume flowers and leafy vegetables and the least demand for eggplant vegetables.

In the longitudinal comparison, this paper is based on the monthly total sales data of each category obtained by preprocessing to get the change of the sales volume share of each category in these three years with the month, as shown in Figure 3.

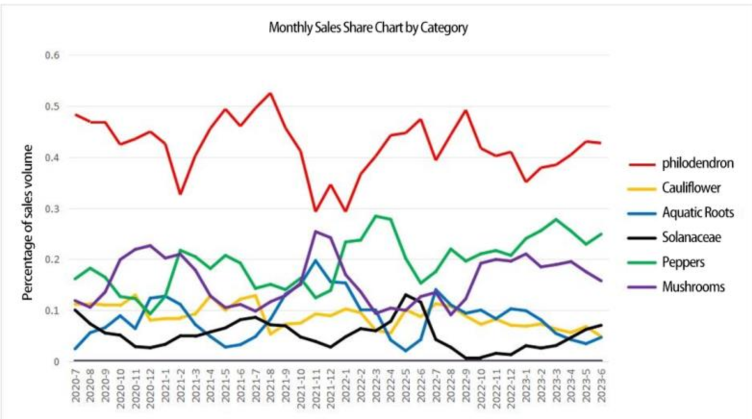

Figure 3: Monthly Sales Share by Category

Next, this paper takes vegetable single product as the research object, analyzes the distribution of three-year total sales volume of each single product among all vegetable single products operated by superstores, and organizes the order to get the vegetable single products with total sales volume ranked in the top 20, as shown in Figure 4.

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

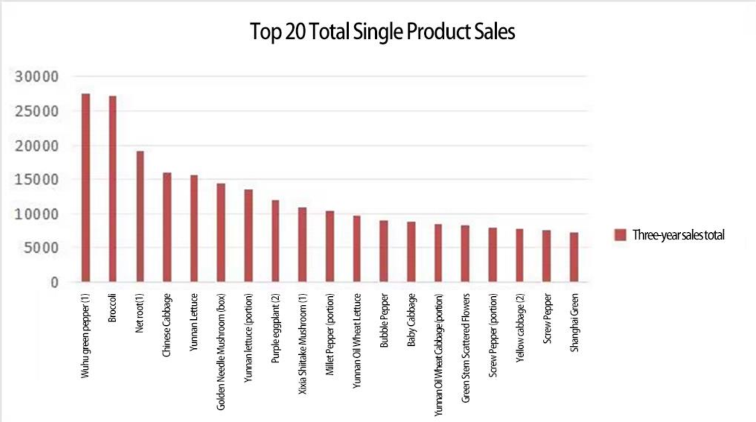

Figure 4: Top 20 three-year total sales volume of individual vegetable products

As can be seen from the figure, among all the vegetable items, Wuhu green pepper (1), broccoli and other vegetable items have excellent sales, followed by net lotus root (1), Chinese cabbage, Yunnan lettuce, enoki mushrooms (box) and other vegetable items also have top sales, and Yunnan lettuce (portions), purple eggplant (2), Xixia shiitake mushrooms (1), millet peppers (portions) and other vegetable items have good sales as well. In relation to the reality of life, combined with the processing of the data, this paper analyzes that the sales of vegetables with high popularity and easy to mix and match with good taste and relatively simple practices tend to be stable and high, on the contrary, the sales of vegetables with low popularity and more troublesome to consume are unstable and low. This provides a reliable basis for supermarkets to formulate relevant business strategies.

To make the comparison of sales volume of vegetable commodities more objective, this paper investigates the total sales volume share of each individual item in the same category. Due to the large number of individual items of leafy flowers, edible mushrooms and other vegetables, in order to ensure the intuitiveness of the results, this paper defines the individual items in leafy flowers, chili peppers, eggplants and edible mushrooms with a sales volume share of less than 4% as "others", and the individual items in aquatic roots and tubers with a sales volume share of less than 2% as "other ", resulting in the following six pie charts.

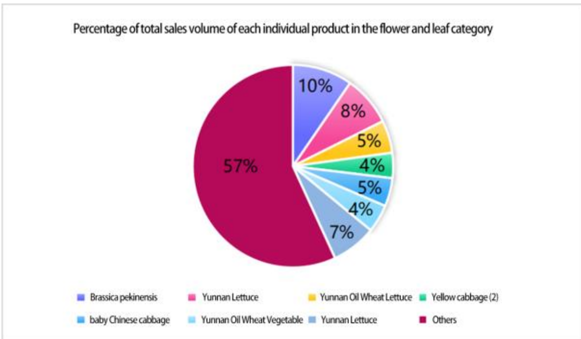

Figure 5: Percentage of single product sales in the flower and leaf category

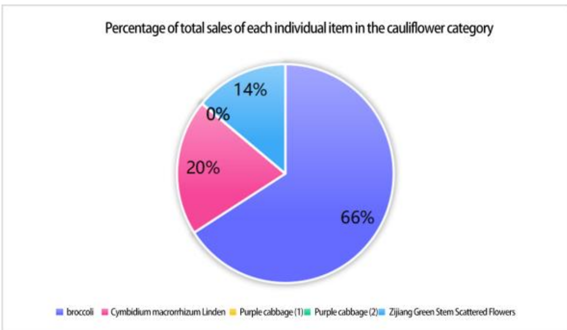

Figure 6: Cauliflower Category Single Product Sales Percentage

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

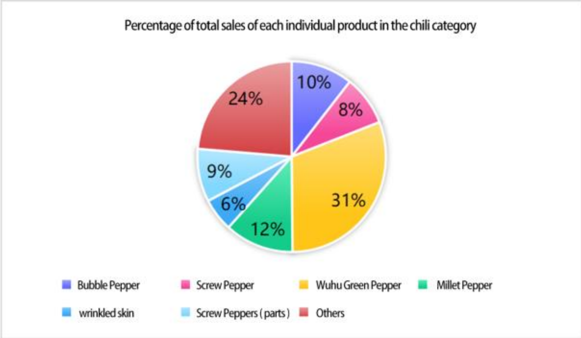

Figure 7: Percentage of Sales of Individual Products in the Chili Category

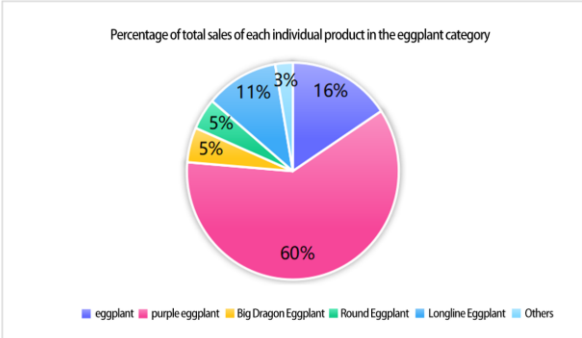

Figure 8: Sales share of single product in the tomato category

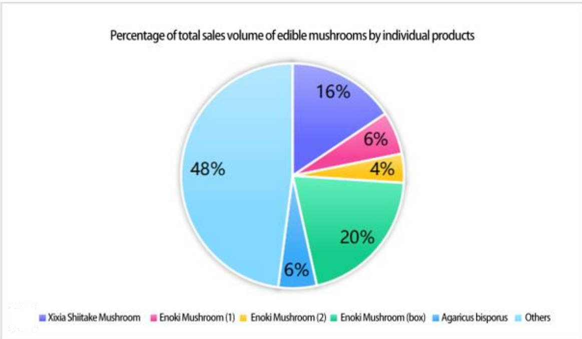

Figure 9: Edible Mushroom Single Product Sales Percentage

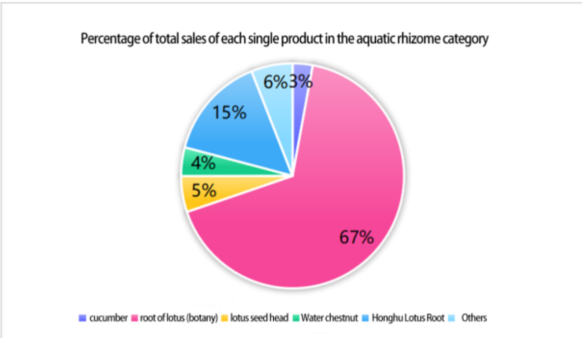

Figure 10: Aquatic Roots and Tubers Single Product Sales Percentage

As can be seen from Figures 5 to 10, among the flowering and leafy vegetables, the sales of each individual item were relatively average overall, with relatively more Chinese cabbage and Yunnan lettuce; among the cauliflower vegetables, broccoli led the sales breaks, and the two types of individual items, purple cabbage (1) and purple cabbage (2), were poorly sold; among the chili peppers, the sales of Wuhu peppers (1) were good, with the remaining various types of individual item sales varying; and among the eggplant vegetables, the purple eggplants (2) accounted for the majority of the sales contribution in this category.

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

2.2 Sales volume interdependencies

In the cross-sectional dimension, the study first quantified the monthly sales volume shares of six vegetable categories to observe their sales distribution patterns. Subsequently, through correlation tests, the study further elucidated the sales correlations among different vegetable categories.

To explore the interrelationships among the vegetable categories and their individual products, Spearman's rank correlation analysis was used. This is a non-parametric method of analysis based on the hierarchy of statistical data, which is usually used to investigate the correlation between two or more variables and whether there is some degree of association [4].

The Spearman correlation coefficient is calculated as follows:

$$\left\{ \rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \right. \tag{1}$$

The study calculated the Spearman correlation coefficients between different single items using Spearman rank correlation analysis. Significant correlations were found between a variety of items, and some of the items with correlation coefficients of 1 are listed in Table 1. From the perspective of political economy and daily life practice, two highly correlated items may have a complementary (strong positive correlation) or substitute (strong negative correlation) relationship.

Table 1: Correlation coefficients between individual items of selected vegetables

Vegetable Item 1	Correlation Coefficient	Vegetable Item 2
Green Vegetable Tails	1	Kale Leaves
Milk White Cabbage Bud	1	Large Mustard Greens
Chrysanthemum Rapeseed	1	Large Mustard Greens
Hongshan Vegetable and Lotus Root Gift Box	1	Hongshan Vegetable Premium Tote Bag
Red Oak Leaf	1	Red Coral (Thick Leaf)
Green Avocado	1	Red Coral (Thick Leaf)
Green Avocado	1	Red Oak Leaf
Mustard Greens	1	Large Mustard Greens
Mustard Greens	1	Milk White Cabbage Bud
Mustard Greens	1	Chrysanthemum Rapeseed

To demonstrate the relationships more visually between different vegetable categories and individual products, heat map analysis was used in this study. Heat maps are an efficient and intuitive data analysis tool that can quickly identify data patterns and generate easy-to-understand visualizations to express the association between a large amount of information and influencing factors. In a heat map, different colors represent different levels of values, i.e., similarities or differences between data points.

By analyzing the Spearman correlation between the sales of six vegetable categories and 245 individual items, two corresponding heat maps were generated for this study. These heat maps help to provide a clearer understanding of the sales correlations between individual categories and individual items.

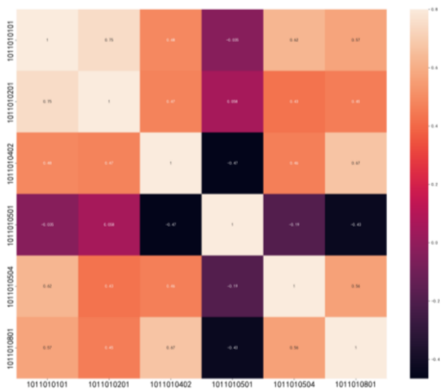

Figure 11: Heat map of correlations by category

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

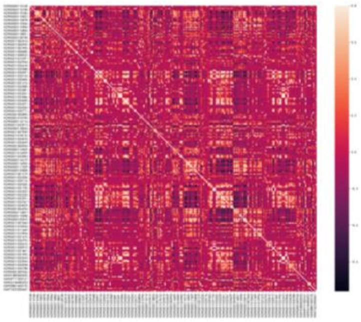

Figure 12: Heat map of correlation between individual products

As can be seen in Figures 11 and 12, the correlation coefficient between leafy and cauliflower vegetables reaches 0.75, indicating that there is a higher probability of successful sales of leafy and cauliflower vegetables along with the sale of cauliflower vegetables, which may be related to the sales mix, the situation of the purchase of goods, the sales strategy, and the consumer's purchasing habits, among other reasons.

3. Pricing strategy and sales impact

3.1 Cost Pricing and Sales Relationship Model

In this study, cost-plus pricing is a pricing strategy based on multiplying the unit cost of a product by a specific scale factor to determine its selling price. To conduct a more accurate analysis, this study first excludes discounted sales and order data with only the purchase price but no sales record, thus obtaining a valid data set that meets the requirements of the study. Further, this study independently extracted the sales flow data of discounted items and non-discounted but reduced-price items and analyzed both types of data in depth. This step helps to understand the relationship between cost-plus pricing and sales volume more precisely as shown in Figures 13 to 18.

Transition from the cost-plus price of a single product to the cost-plus price of the category, this paper adopts a weighted average approach, according to the same single product in a month of different time periods at different prices corresponding to the contribution of sales volume to calculate the single product in the month's weighted average price, and then according to the contribution of the sales volume of each single product in the same category to calculate the category's weighted average price as shown in Table 2.

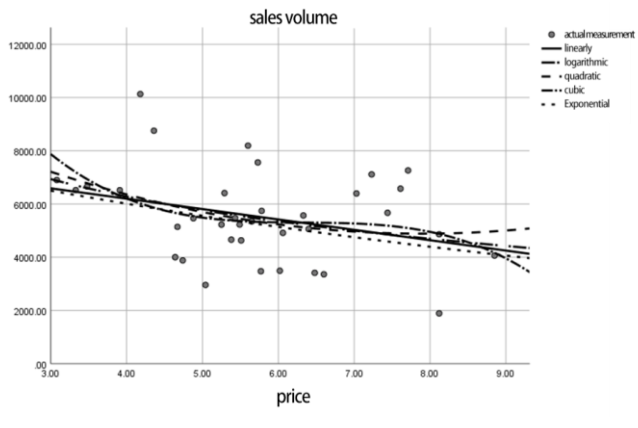

Figure 13: Curve fitting for foliar classes

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

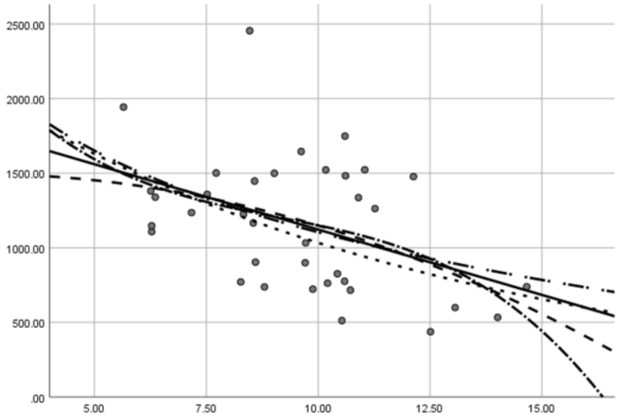

Figure 14: Curve fitting for cauliflower species

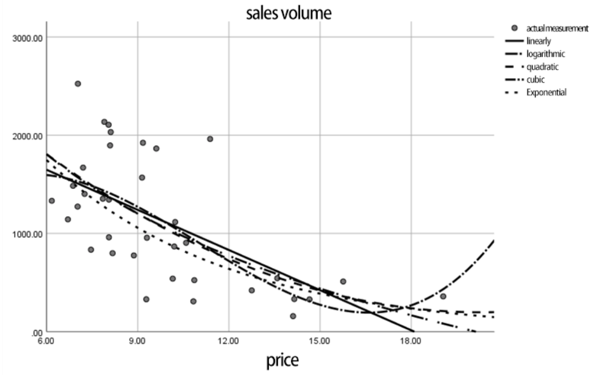

Figure 15: Curve fitting for aquatic rhizomes

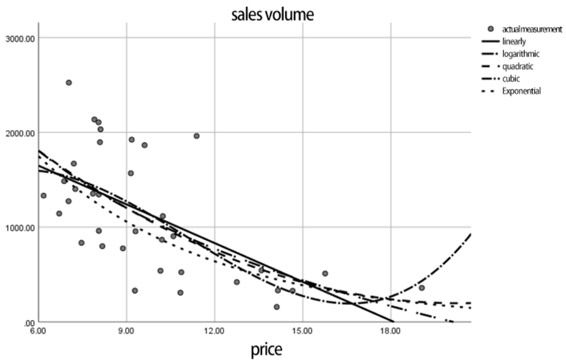

Figure 16: Solanum curve fitting

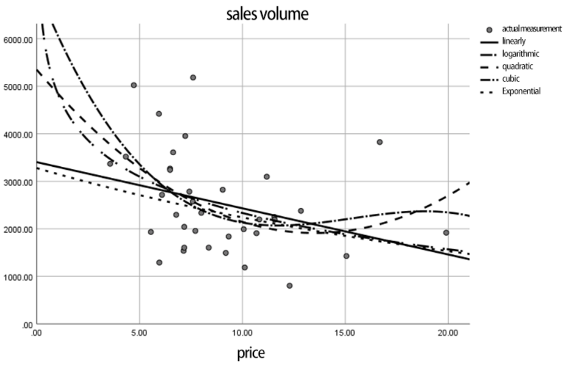

Figure 17: Curve fitting for chili peppers

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

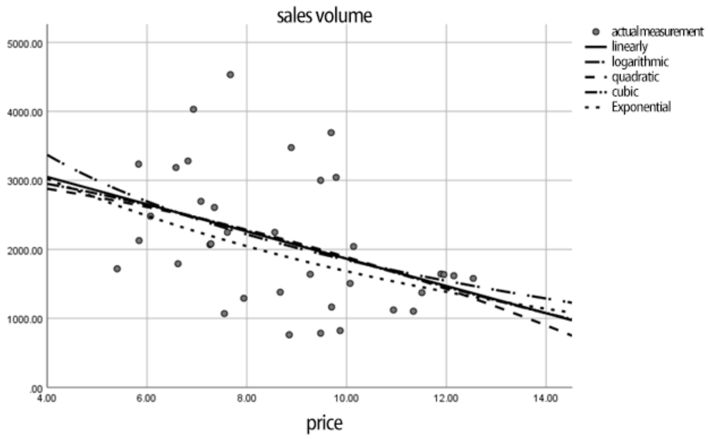

Figure 18: Edible Mushroom Curve Fitting

Table 2: Relationship between total sales and cost-plus pricing for each vegetable category

Vegetable Category	Relationship
Leafy Flowers	$Y = 10.867 - 0.002 * x + 3.893 * 10^{-7} * x^2 - 2.121 * 10^{-11} * x^3$
Cauliflower	$Y = 22.468 - 0.028 * x + 1.804 * 10^{-5} * x^2 - 3.728 * 10^{-9} * x^3$
Aquatic Root Vegetables	$Y = 18.575 - 0.19 * x + 1.067 * 10^{-5} * x^2 - 1.971 * 10^{-9} * x^3$
Eggplant Category	$Y = 7.829 + 0.014 * x - 2.571 * 10^{-5} * x^2 + 1.126 * 10^{-8} * x^3$
Chili Peppers	Y=11.629 * exp (-0.000134 * x)
Edible Mushrooms	$Y = 9.757 + 0.001 * x - 1.194 * 10^{-6} * x^2 + 1.953 * 10^{-10} * x^3$

3.2 Mechanisms of pricing impact on sales volume

The seasonal ARIMA model of time series forecasting model can predict the time series containing seasonality and trend, which is suitable for solving the sales forecast of time-sensitive just-demanded products. ARIMA model includes moving average (MA) model and autoregressive (AR) model. The former utilizes a linear combination of past residual terms to observe future residuals, while the latter describes the relationship between current and lagged values and predicts future values using historical data. The specific formulas of ARIMA forecasting model are as follows [5]:

$$\hat{p}^{(t)} = p_0 + \sum_{j=1}^{p} \gamma_j p^{(t-j)} + \sum_{j=1}^{q} \theta_j \varepsilon^{(t-j)}. \tag{2}$$

Here p is the order of the autoregressive model (AR) and q is the order of the moving average model (AM). ε,t is the error term between times ranging between t and t - 1. γ and $\theta$is the fitting coefficient. p_0 is a constant term.

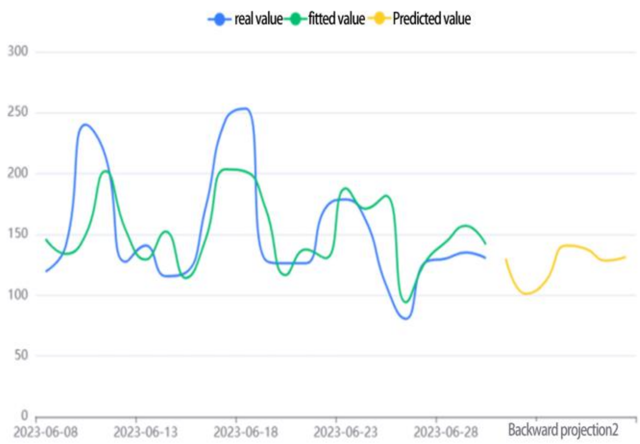

Figure 19: Fitting curve of predicted data for foliar species

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

To accurately predict the total appropriate daily replenishment for each vegetable category from July 1 to 7, 2023, a seasonal ARIMA (Autoregressive Integral Sliding Average Model) was constructed in this study. Taking leafy and flowering vegetables as the empirical object, this study generated fitted curves and the forecast data of sales volume for the coming week. To increase the precision of the forecasts, all forecasts were retained to three decimal places. This approach not only provides scientific forecasts about future demand, but also provides strong data support for daily replenishment decisions in Table 3 and Figure 19.

Table 3: Time series prediction results

Order (time)	Projected results
1	130.077
2	101.160
3	111.624
4	140.740
5	138.121
6	128.377
7	131.354

The predicted sales volume of leafy and flowering vegetables for the coming week can be known from the chart, which serves as an important reference for the daily replenishment volume. In developing the pricing strategy, this paper develops the replenishment strategy based on the predicted sales volume of each category for the next seven days as the replenishment volume and develops the pricing strategy for each category for 6 each day for the next seven days based on the relationship between the sales volume and the cost-plus pricing obtained in the previous section, and the results are shown in Table 4.

Table 4: Pricing strategy by category for the next 7 days

Order (Time)	Leafy Flowers	Cauliflower	Aquatic Root Vegetables	Eggplant Category	Chili Peppers	Edible Mushrooms
1	10.62	21.75	16.8951	8.41	11.50	9.80
2	10.68	21.87	17.5577	8.47	11.49	9.80
3	10.66	22.01	15.7204	8.12	11.49	9.81
4	10.60	22.06	14.6295	8.15	11.49	9.81
5	10.61	21.98	15.1440	8.19	11.49	9.82
6	10.63	22.03	14.2953	8.22	11.49	9.81
7	10.62	21.97	15.7063	8.15	11.49	9.82

After conducting an analysis of the sales flow data for discounted items and non-discounted but price-reduced items, it was determined that between 1 to 2 items are sold at a reduced price without an additional discount after 19:00 each day. This finding is significant as it helps the superstore avoid a situation where a considerable number of goods cannot be sold again. Given the relatively random nature of discounting, the pricing strategies for both discounted and non-discounted but price-reduced sales are formulated based on the pricing strategies provided in the previous section.

An analysis of three years of non-discounted reduced-price sales revealed that floral and foliage items accounted for 41.5% of total sales. Whereas, during July 1-7, 2021, and 2022, edibles items participated in discounted sales at 74%. Therefore, consideration can be given to selling any of the floral and foliage items at a reduced price after 19:00 for the next seven days, as well as discounted sales of edibles that have declined in sales, in order to stimulate the desire of customers to buy, increase the overall sales of each category, and reduce the inventory buildup on that day, so that the superstore's revenues can be maximized.

4. Replenishment and pricing optimization strategies

The projected sales of the notation sheets on July 1, 2023, are X_j, the selling price is S_j, Wholesale prices are P_j, The incoming quantity is V_j, The attrition rate is L_j, The range of values of n is [27, 33].

$$Max \sum_{j}^{n} X_j * S_j - V_j * P_j * (1 + L_j) \tag{3}$$

The constraints are:

ISSN 2616-5775 Vol. 6, Issue 11: 133-143, DOI: 10.25236/AJCIS.2023.061117

$$\begin{cases} 27 \le n \le 33 \\ V_j \ge X_j \\ V_j \ge 2.5 \end{cases} \tag{4}$$

The superstore strategy is obtained by solving this planning model using genetic algorithm of SPSSPRO. The final replenishment and pricing decisions are obtained as shown below in Table 5:

Table 5: Individual Product Replenishment Strategies and Pricing Strategies, July 1, 2023

Pricing (¥/Kg)	Restocking Quantity (Kg)	Forecasted Procurement Price (¥/Kg)	Product
5.80	32.29	3.60	Yunnan Lettuce (portion)
5.80	21.43	2.14	Small Chili Pepper (portion)
4.50	21.29	2.86	Yunnan Oyster Plant (portion)
5.20	14.23	3.38	Wuhu Green Pepper (1)
6.80	7.00	4.10	Spinach (portion)
12.00	6.84	8.79	Corkscrew Pepper
5.20	6.42	2.53	Milk White Cabbage
16.00(8.00)	6.03	10.46	Clean Lotus Root (1)
6.00 (3.60)	5.93	3.21	Wood Ear Vegetable
8.00	3.86	4.12	Shanghai Green
16.00(9.60)	3.00	11.62	Gourd (1)
6.00	2.68	4.05	Green Eggplant (1)

By integrating and analyzing the historical sales data, making price reductions for the three types of vegetable items, namely, net root (1), mullein, and kohlrabi (1), after 19:00 hours on the same day (the prices after the price reductions are the values in the parentheses) will cause higher sales volume, which is conducive to the improvement of the overall operational income of the superstore.

In this paper, based on solving the July 1 replenishment strategy, the statistics of the replenishment volume of each category, and compared with the previous statistics of the six categories of three-year total sales volume of the respective share, as shown in Table 6. The results show that the July 1 replenishment volume share of each category is roughly like its three-year total sales share, indicating that the replenishment strategy developed for the superstore does not unilaterally pursue the maximization of profitability of individual categories, but rather meets the market demand for vegetable goods in various categories, and considers the social benefits while achieving the greatest possible returns.

Table 6: Historical and Projected Share and Situation of Each Vegetable Category

Vegetable Category	Three-Year Total Sales Proportion	July 1st Restocking Proportion
Leafy Flowers	42.15%	39.44%
Cauliflower	8.87%	6.83%
Aquatic Root Vegetables	8.62%	9.32%
Eggplant Category	4.76%	6.17%
Chili Peppers	19.45%	24.76%
Edible Mushrooms	16.16%	13.49%

5. Conclusions

This study effectively analyzes the sales distribution pattern, sales forecast, and sales strategy of the vegetable category through a combination of models and methods. A more relevant analysis and prediction is achieved through the seasonal ARIMA model and combined sales strategy. However, the model is still insufficient in dealing with unexpected events and scenario-specific applicability. Future improvements include more accurate data categorization and expanding the applicability of the model to meet actual operational needs more comprehensively.

References

[1] Wu Huihui, Wang Jiapeng, Wu Wenjing et al. Analysis of global air surface temperature prediction based on ARIMA model [J]. Modern Information Technology, 2023, 7(16): 147-150. DOI:10. 19850/j. cnki.2096-4706.2023.16.032.

[2] Yu ZT, Bei YJ, Zhang W et al. A study on the current situation of physical health, mental health and social adaptability of students in higher vocational colleges and the correlation of Spearman's rank [J]. Sports Vision, 2023(12): 7-11.

[3] Gu Sihong. A study on dynamic pricing of fresh products in H-retailers considering freshness change [D]. Donghua University, 2023. DOI:10.27012/d.cnki.gdhuu.2023.001318.

[4] Zhu Zhongrui. Research on the influence law and action mechanism of corn canopy on redistribution of sprinkler irrigation water and water droplet movement characteristics [D]. Northwest Agriculture and Forestry University, 2022. DOI:10.27409/d.cnki.gxbnu.2022.002201.

[5] M. Zeng. 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.