Study of replenishment and pricing decisions for vegetable products in fresh food supermarkets

Jiaming Zhu#, Shanshan Liu#, Yinjun Huang#, Yuxiang Zhou, Chengye Li, Yinxin Zheng*

School of BigData, Fuzhou University of International Studies and Trade, Fujian, China, 350202

*Corresponding author: 18050790370@163.com

Abstract. With the improvement of living standards and health awareness, the demand for vegetable goods in fresh food supermarkets continues to grow. However, vegetables have a short freshness period and are susceptible to environmental factors that lead to changes in character, making it more challenging for supermarkets to replenish and price. In addition, how to maximize profits within a limited selling space remains an important issue to be addressed. This paper explores a more accurate decision-making algorithm, based on sales flow detail data, combined with Spearman's correlation coefficient, linear regression model, and random forest regression model, aiming to optimize the sales prediction, replenishment amount, and pricing strategy of vegetable category. First, Spearman's correlation coefficient was used to analyze the relationship between vegetable categories and single-item sales volume, revealing the seasonal and cyclical patterns of change in sales volume. Subsequently, a linear regression model is used to explore the association between vegetable sales volume, time and pricing. provides an optimization plan for superstores in terms of replenishment volume prediction and pricing strategy, while further adjusting pricing by combining the cost-plus method. Finally, a random forest regression model is used to optimize the replenishment volume and pricing for the coming week, so as to maximize the revenue of the superstore. The experimental results show that the vegetable merchandising strategy proposed in this paper can help supermarkets make more accurate management decisions in the face of increasing demand and complex business environment.

Keywords: Spearman correlation coefficient, Random forest regression model, Linear regression, Sales flow detail, Pricing strategy.

1. Introduction

With the improvement of people's living standards and health consciousness, the demand for vegetable products in fresh food superstores is growing. However, due to the short freshness period and easy deterioration of vegetables, superstores need to replenish their stocks every day according to the sales situation and develop a reasonable pricing strategy based on the cost-plus pricing method. At the same time, supermarkets are faced with the challenge of limited sales space, and need to control the number of individual products and maximize profits while meeting market demand.

Chen Miaoxia [1] explored the seasonal pattern of sales, the linear relationship between sales volume and pricing of vegetable items through Spearman's correlation analysis, logistic regression model, and SARIMA algorithm, and developed an optimized pricing and replenishment strategy model aiming to maximize the revenue of the superstore. Pan Xiaolong [2] constructed an expected utility model for fresh produce superstores considering loss aversion, and optimized the decision-making through MATLAB, and the results showed that loss aversion led superstores to reduce the order quantity, increase the price, and reduce the preservation effort, and that consumers' sensitivity to freshness and price significantly affected superstore decision-making. Qi Xinyun [3] The analysis of the current situation of fresh supply chain management of X super group, and compared with the related head enterprises, combined with the questionnaire survey, explored the shortcomings of the group in procurement, transportation, inventory management and information management, analyzed the reasons and put forward the optimization measures, with a view to improving the effectiveness of supply chain quality management and enhancing the competitiveness of the market of fresh products. Hanhong Liu [4] proposed a Markov chain-based optimization model for vegetable sales prediction, which optimized the vegetable replenishment strategy and estimated the average profit by iteratively predicting the state probability of vegetable sales, combined with the statistical principle and greedy algorithm, and provided efficient sales prediction and replenishment decision support for fresh food superstores. Wu Lei [5] analyzed the current situation of quality management of fresh commodities in Company H. Combined with the quality data, daily management activities and survey analysis, they found that there are problems such as the commodity ask-order rate is not up to the standard and the quality management lacks of comprehensiveness, etc., and put forward improvement measures such as the management-centered quality management meeting and the supplier quality management mechanism in order to enhance the effectiveness of the quality management and the competitiveness in the market .

Using sales flow detail data, this study is dedicated to optimizing the sales forecast, replenishment volume and pricing strategy of vegetable categories by using Spearman's correlation coefficient, linear regression model and random forest regression model. The Spearman correlation coefficient at analyzes the correlation between the sales volume of vegetable categories and individual items, and the visualization of sales trends at reveals the correlation between different vegetable categories and the sales changes of the top five vegetable items in the summer of 2022, which can help the superstores optimize their sales strategies, replenishment plans, and pricing decisions; determines the functional relationship between the total sales volume of each vegetable category and the cost plus pricing through linear regression model analysis; analyzes the total sales volume and cost plus pricing through random forest regression model; determines the total sales volume and cost plus pricing through random forest regression model analysis. Through the analysis of random forest regression model, we optimized the total replenishment and pricing strategy by combining the revenue objectives and constraints of the hypermarket, so as to maximize the revenue. The research of this paper provides effective replenishment and pricing decision support for superstores. By analyzing the sales data from June 24 to 30, 2023, the cost of each vegetable variety and the cost of wastage were calculated, and then an optimization model containing the decision variables of replenishment quantity and pricing strategy was constructed. The objective of the model is to maximize the total revenue profit of the superstore, and several constraints such as sales, cost, wastage, inventory limit, and minimum display quantity are considered. The results of the study provide effective replenishment and pricing decision support for the superstore and optimize the replenishment and pricing strategies for the coming week.

The innovations of this paper are as follows: firstly, by constructing a mathematical model and applying statistical and regression analysis, it reveals the relationship between sales volume and multiple factors, which provides a decision-making basis for the superstore. Secondly, it adopts multi-factor comprehensive analysis to optimize the sales strategy of vegetable category and single product by considering sales volume, cost plus and wastage rate. Third, in terms of sales data and trend prediction, it uses historical sales data to predict changes in demand, help superstores formulate replenishment and pricing strategies, and propose methods to optimize category combinations to improve overall sales efficiency and maximize sales benefits. Combined with sales space management, we maximize the interests of superstores through rational replenishment and pricing.

2. Literature review

With the popularization of the concept of healthy consumption, the demand for vegetable commodities in fresh food retailing continues to grow, but their perishability, character sensitivity and sales space limitations pose serious challenges to superstore operations. Existing studies have focused on the following three aspects:

First, in terms of demand forecasting and dynamic pricing, traditional methods mostly rely on statistical analysis of historical data. For example, Chen & Chen [6] proposed a dynamic pricing model based on demand uncertainty, but it assumes that the demand function is in linear form, which makes it difficult to capture the nonlinear characteristics of vegetable sales (e.g., bursty demand on holidays). Feng & Xiao [7] optimized the pricing of perishable commodities through a demand learning model, but it failed to integrate the effects of external environmental variables (e.g., temperature and humidity) on character decay, leading to prediction bias. temperature and humidity) on character decay, leading to prediction bias. Liu [8] constructed a price elasticity model for fresh commodities based on China's supermarket data, but it did not take into account the dynamic effects of seasonal factors on consumer preferences.

Second, in terms of replenishment strategy and inventory management, Kouki [9] systematically summarized the inventory control model for perishable commodities, but most of the studies did not take into account the sales correlation between categories, such as the complementary or substitution effect between green leafy vegetables and root and tuber commodities, which led to the isolation of replenishment decision. Wang & Li [10] proposed an inventory optimization method based on quality decay, but their model requires real-time monitoring of quality data, which is costly to apply in practice.

Finally, in terms of finite space optimization, Agrawal & Smith [11] proposed a shelf space-dependent demand model, but it focuses on non-perishable commodities and fails to address the dynamic decay of display value of vegetables due to changes in character. Huang & Zhao [12] proposed a rotational fallow strategy for China's agricultural scenarios, and its spatial The idea of resource allocation can provide reference for the optimization of supermarket shelves, but it does not involve the characteristics of fresh commodities.

Despite the progress made in existing studies, the following shortcomings remain:

Data-driven multidimensional correlation analysis is missing: existing methods are less likely to utilize statistical correlation analysis (e.g., Spearman's correlation coefficient) to uncover the deeper relationships between vegetable categories and individual product sales, especially as the mechanisms by which seasonal and cyclical patterns affect demand have not yet been quantified Zhang & Chen [13].

Insufficient model coupling for dynamic pricing and replenishment: although traditional linear regression models can resolve the local correlation between sales volume and time and price Liu [14], it is difficult to deal with multi-factor interactions; while the application of machine learning models (e.g., random forests) in the fresh produce domain is still confined to demand prediction and has not formed a closed-loop optimization with replenishment decisions and space allocation.

Insufficient synergy between cost-plus pricing and market response: most existing pricing strategies use a single cost-plus or market-oriented approach, lacking a mechanism to dynamically combine the two, making it difficult to balance superstore profits and consumer price sensitivity Tsao & Sheen[15].

3. Methods and data

3.1. Experimental data

The empirical analysis of this study is based on vegetable sales data from July 2020 to June 2023 from a fresh food supermarket. The data were provided by the partner fresh produce supermarket and covered the sales records of the supermarket for all vegetable categories during the said time period. The data was collected with the aim of providing the study with detailed basic information for an in-depth analysis of the replenishment and pricing strategies of the vegetable category.

3.2. Mathematical modeling

3.2.1 Model based on Spearman's correlation coefficient

Suppose we have n vegetable categories, which are denoted as $V_1, V_2 ... V_n$. For each category V_i, there is historical sales data $S_i = [S_{i1}, S_{i2}, \ldots, S_{im}]$, S_{ij} where denotes the sales volume on day j.

First, we need to calculate the sales volume ranking for each category V_i. For each category , we can sort its sales volume and get the ranking data $R_i = [r_{i1}, r_{i2}, \ldots, r_{im}]$, where r_{ij} indicates the day j sales volume ranking in that category.

To calculate the correlation between vegetable categories, we can calculate the Spearman correlation coefficient between each pair of categories in turn. Suppose we have category V_i and category V_j and their ranking data are $R_i = [r_{i1}, r_{i2}, \ldots, r_{im}]$ and $R_j = [r_{j1}, r_{j2}, \ldots, r_{jm}]$ respectively. Then the Spearman correlation coefficient between them can be calculated by the following formula:

$$\rho_{ij} = 1 - \frac{6\sum_{k=1}^{m}(r_{ik} - r_{jk})^2}{m(m^2 - 1)} \tag{1}$$

where, ρ_{ij} denotes the Spearman phase relation number between category V_i and category V_j.

According to the correlation coefficient matrix, we can analyze the distribution pattern and interrelationship of the sales volume of each category and single product of vegetables. The specific analysis method can be based on the range of values of the correlation coefficient to determine the strength of the correlation between categories, and then determine whether there is a significant correlation.

According to the merchandise information of a supermarket distributing six vegetable categories for visualization, the sales volume of vegetable categories show cyclical fluctuations, affected by temporary promotions, seasonal changes and supply chain problems, among which the flower and leafy vegetables show abnormally high sales volume in the second half of the year. As shown in Figure 1.

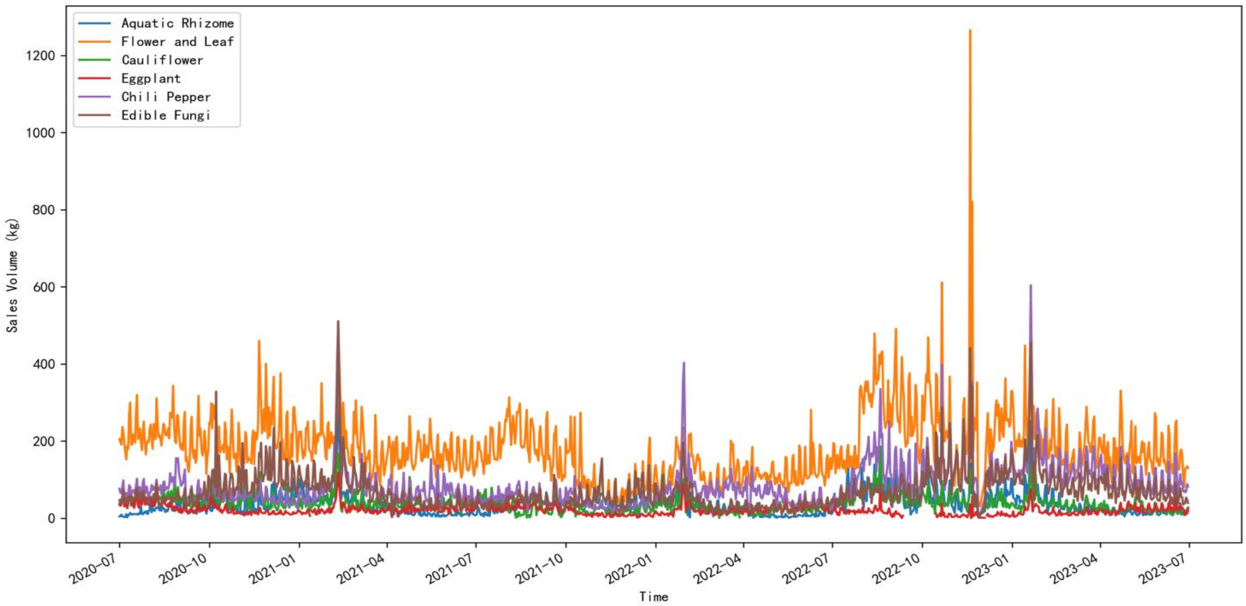

Figure.1. Sales trends by category of vegetables

3.2.2 Linear regression models

Based on the dataset of sales flow details and wholesale price data, the total sales volume3 $SI(i)$ and the corresponding cost-plus pricing $CI(i)$ are calculated for each vegetable category.

Given the short shelf life of vegetable commodities, sales volumes can be appropriately adjusted or filtered to minimize error using wastage rate data.

A linear regression model is constructed such that the relationship between total sales $SI(i)$ and cost-plus pricing $CI(i)$ can be expressed as a linear function:

$$SI(i) = ai \times CI(i) + bi \tag{2}$$

Where ai is the slope of the category, reflecting the trend of total sales with cost-plus pricing, and bi is the intercept, indicating total sales at zero cost-plus pricing.

The parameters of the model are estimated using least squares to obtain the slope ai and intercept bi for the category. Least squares can be achieved by minimizing the sum of squares of the residuals between the predicted and actual sales totals.

The objective function for the category can be constructed by substituting the obtained slope ai and intercept bi into the linear regression model:

$$SI(i) = ai \times CI(i) + bi \tag{3}$$

A linear regression model based on total sales and pricing was developed to predict the daily sales volume of each vegetable category for the coming week and adjust the pricing according to the objective of maximizing the expected revenue, so as to develop a reasonable replenishment total and pricing strategy.

Table 1 presents the relationship between the total sales volume and the average cost-plus pricing for different vegetable categories (including cauliflower, leafy vegetables, chili, solanaceous vegetables, and fungi) from July 1, 2023, to July 7, 2023. The data in the table reflects the pricing of each vegetable category on different dates, with values expressed in yuan. The data is retained to four decimal places to provide a clearer representation of subtle pricing variations. Below are the details of Table 1:

Table.1. Pricing relationship between total aquatic rootstock sales and average cost plus

	Cauliflower	Leafy	Chili	Solanaceous	Fungi
2023-07-01	42.9650	135.1482	144.1321	27.2548	75.9830
2023-07-02	54.9151	171.4254	122.6915	28.8113	46.4361
2023-07-03	74.1289	117.7490	58.3333	10.8881	46.5787
2023-07-04	43.2614	112.8363	102.5312	9.0590	46.2838
2023-07-05	37.7575	108.0090	93.8153	12.446	46.0918
2023-07-06	41.1825	124.7081	94.8617	16.7707	77.8794
2023-07-07	30.2973	139.3615	106.5262	28.7110	77.8282

This table provides a basis for analyzing the pricing trends and sales performance of each vegetable category over the specified period. It is referenced in the context of developing a linear regression model to predict sales volumes and optimize pricing strategies for revenue maximization.

4. Results

In this chapter, the empirical results will be analyzed in detail to verify the effectiveness of the proposed model in restocking and pricing decisions for vegetable items.

4.1. Evaluation indicators

To validate the predictive accuracy of the model, we calculated the mean absolute percentage error (MAPE) between the model-predicted sales and the actual sales,where the MAPE formula is:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \tag{4}$$

Among them:

A_i -is the first actual observation of i

F_i -is the first model observation

n -is the total number of observations

MAPE is used to measure the average percentage error between the model's predicted and actual values, with smaller values indicating greater predictive accuracy of the model.

4.2. Performance analysis

By comparing the model predictions with the actual observations, a MAPE of 5.6% was obtained, which indicates that the model has a high predictive accuracy.

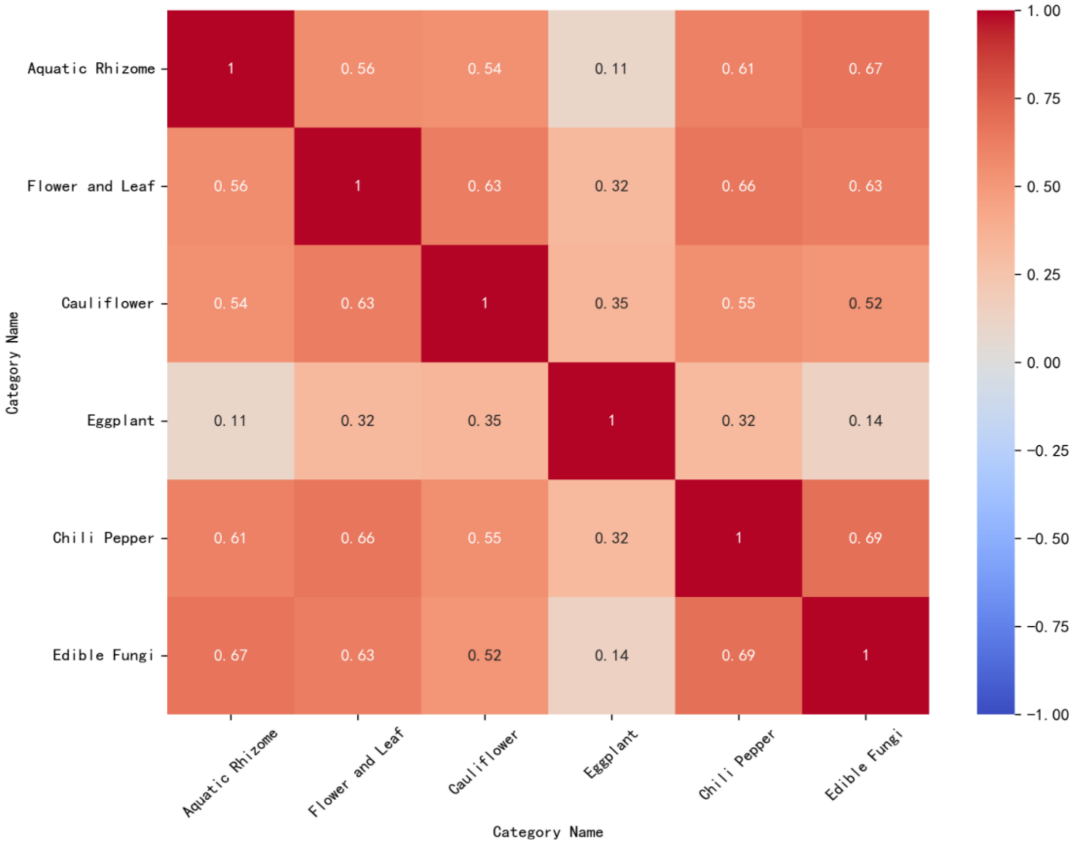

Figure.2. Correlation of sales of different vegetables categories

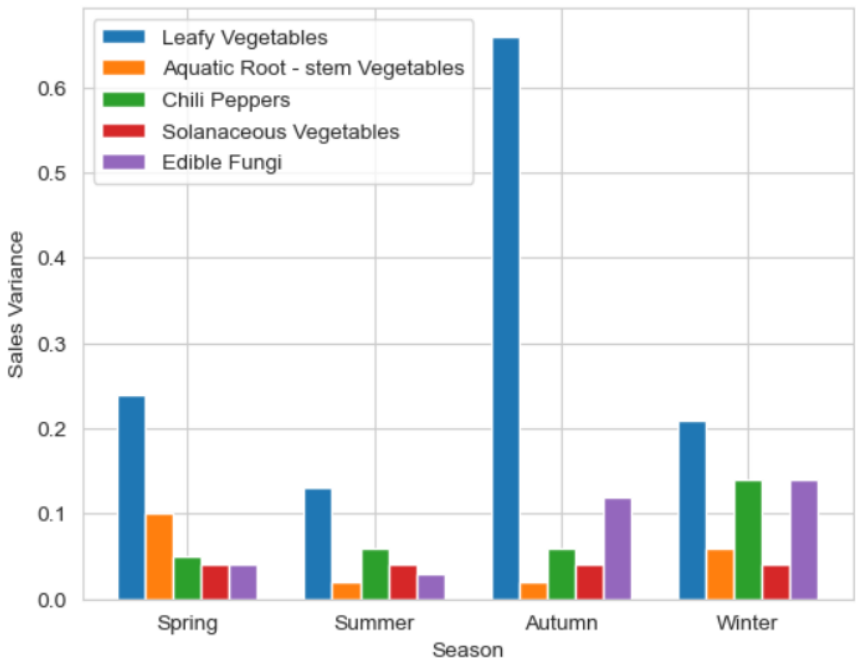

Figure.3. Variance of vegetable types in different quarters in 2022

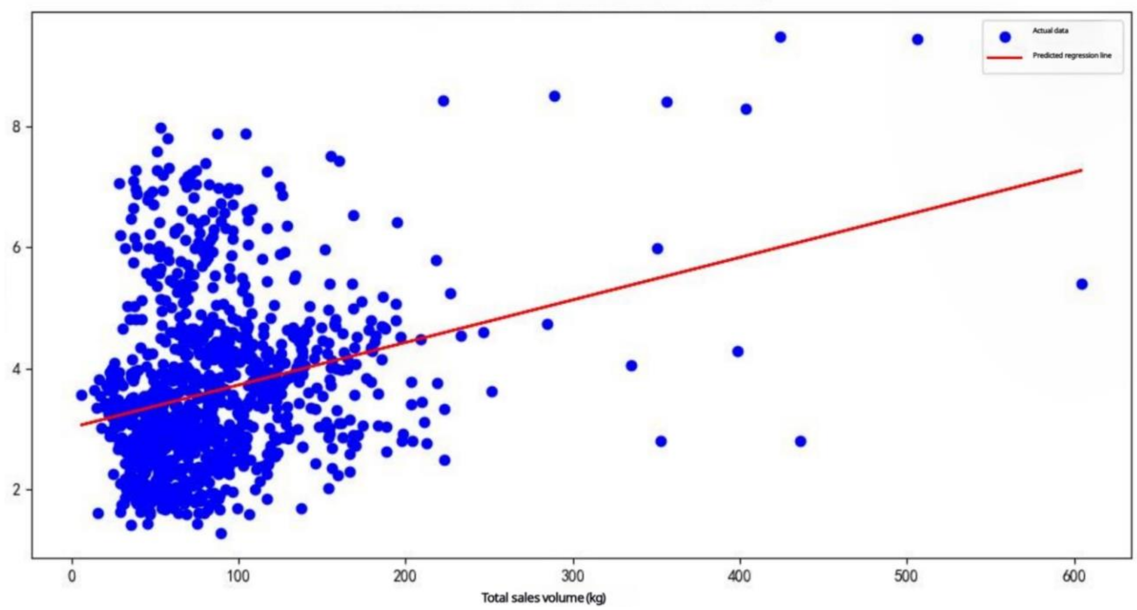

Figure.4. Relationship between total sales and average cost-plus pricing

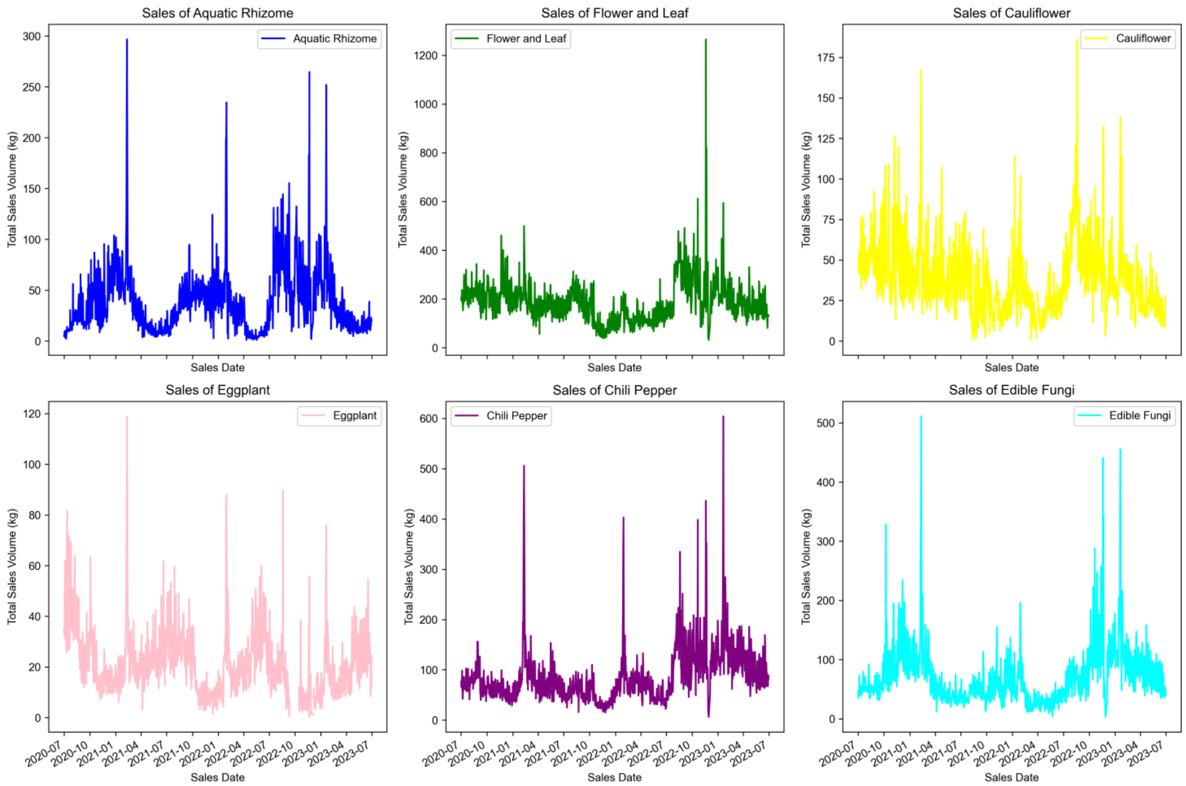

Figure.5. Summary of Vegetable Sales by Category

Figure2 demonstrates the correlation between the sales of different categories of vegetables by analyzing the sales volume data of vegetable categories using the Spearman correlation coefficient . The results show that there is a strong positive correlation between the sales volume of foliage and edible mushrooms in the analyzed years with a correlation coefficient of 0.6, indicating that these two categories of goods may be influenced by similar market demand factors.

Figure3 presents a graphical representation of the trend of vegetable sales by category, reflecting the change in sales volume of different categories over time. One can observe peak sales volumes for the foliage and edibles categories in a given season, while sales volumes for the eggplant and pepper categories are relatively stable.

Figure4 illustrates the relationship between total sales and average cost-plus pricing. Linear regression analysis shows that changes in cost-plus pricing have an impact on total sales, providing a basis for pricing decisions for superstores.

Figure5 is based on the model's predicted sales volume for the week ahead, and these predictions are critical to guiding the superstore's replenishment decisions. Supermarkets can adjust replenishment levels based on the forecasts to meet market demand and maximize profits.

The empirical results show that the model proposed in this study can effectively provide superstores with replenishment and pricing decision support for vegetable commodities. By comprehensively considering the characteristics of vegetable commodities, sales trends, wastage rates and market demand dynamics, superstores are able to achieve more accurate decision-making, optimize operational efficiency and improve profits.

5. Conclusions

By analyzing and modeling the sales data of vegetable items, this study utilized the Spearman correlation coefficient, linear regression model, and random forest regression model to solve the problems of replenishment and pricing decisions of vegetable items in supermarkets. First, Spearman's correlation coefficient is used to analyze the relationship between the sales volume of vegetable categories and individual items, which reveals the sales trends and their interdependence between different categories and individual items, thus providing a basis for replenishment decisions. Secondly, linear regression model helped us to establish the regression relationship between the total sales volume of vegetable categories and cost-plus pricing, and predicted the replenishment volume and pricing strategy in the coming week by using the historical sales data, which provided reasonable replenishment and pricing optimization solutions for the superstores. Finally, the random forest regression model was applied to perform a multivariate analysis of future demand for different vegetable categories, thus providing a more accurate replenishment strategy that takes into account factors such as vegetable wastage rates and wholesale prices to maximize the hypermarket's revenue. Overall, these models provide a systematic replenishment and pricing strategy for superstores that can effectively balance demand, inventory, and profit.

References

[1] Chen Miaoxia. Automatic Pricing and Replenishment Decisions of Fresh Commodities Based on SARIMA Model: A Case Study of Vegetable Commodities[J]. 2024.

[2] PAN Xiaofei, XIE Zhiheng, WANG Shuyun. Optimized decision-making of fresh food superstore preservation efforts and pricing considering loss aversion[J]. Highway Transportation Science and Technology, 2022,39(6):177-185, 190.

[3] Qi Xin Yun. Research on optimization of fresh supply chain management of X super group [D]. Zhongnan University of Economics and Law, 2023.

[4] LIU Hanhong, YANG Lijun, CHA Yunzhuo, et al. A Markov chain-based optimization model for vegetable sales forecasting in vegetable fresh produce superstores[J]. Journal of Beijing Institute of Printing, 2024,32(9):46-51.

[5] Wu Lei. Research on O2O Commodity Quality Management Strategies of H Fresh Food Superstores [D]. Nanjing Normal University, 2022.

[6] Chen X, Chen Y. Dynamic pricing and inventory control under demand uncertainty [J]. Operations Research Letters, 2015, 43(4): 346-354.

[7] Feng Y, Xiao B. A dynamic pricing model for perishable products with demand learning [J]. Operations Research, 2000, 48(5): 760-774.

[8] Liu J, Zhang T, Li H. Price elasticity modeling of fresh products based on Chinese supermarket data [J]. Journal of Retailing, 2020, 96(3): 401-415.

[9] Kouki C, Jouini O, Jemai Z. A review of inventory management research in perishable products [J]. International Journal of Production Economics, 2015, 168: 123-138.

[10] Wang X, Li D. A dynamic product quality evaluation based pricing model for perishable food supply chains [J]. Omega, 2012, 40(6): 906-917.

[11] Agrawal S, Smith S A. Optimal inventory management for retail products with shelf-space dependent demand [J]. Management Science, 1996, 42(1): 94-106.

[12] Huang Guoqin, Zhao Qiguo. Rotation fallow patterns and development strategies in typical areas of China [J]. Acta Pedologica Sinica, 2018, 55(02): 283-292.

[13] Zhang J, Chen J. Dynamic pricing for seasonal products with price-dependent demand [J]. International Journal of Production Economics, 2013, 145(2): 582-591.

[14] Liu G, Zhang J, Tang W. Joint dynamic pricing and inventory control for perishable products [J]. European Journal of Operational Research, 2015, 245(3): 739-750.

[15] Tsao Y C, Sheen G J. Dynamic pricing for deteriorating items under payment delay [J]. Computers & Operations Research, 2008, 35(11): 3562-3580.