Vegetable Stocking and Pricing Model Based on Time Series Forecasting and Non-linear Programming

Ruyi Ren*

School of mathematics and statistics, HaiNan Normal University, Haikou, China, 571199

*Corresponding author: m17832431555@163.com

Abstract. This study aims to help fresh produce superstores develop vegetable replenishment and pricing strategies to maximize their benefits. By analyzing historical sales records and customer demand, the study used an ARIMA model to predict future demand and a planning model to optimize superstore revenues. The study also analyzed the sales data of 6 major categories and 246 individual types of vegetables and found that flower and leafy vegetables were the most popular in the market. By performing Spearman correlation analysis on the sales volume data, the study obtained the two pairs of vegetable categories with the highest degree of correlation. In addition, a positive correlation between sales volume and total price was fitted by linear regression and converted to a negative correlation between cost-plus pricing and sales volume. Finally, past costs, damage rates, price factors, and demand were taken into account to optimize each day's data individually to maximize the superstore's revenue, where the maximum profit on the first day was $1,520.6.

Keywords: Vegetables, Pricing replenishment strategy, ARIMA model, Spearman correlation analysis.

1. Introduction

In fresh food superstores, the sales management of vegetables is a core aspect of their daily operations. As vegetables are perishable and their quality and appearance will gradually deteriorate over time, supermarkets need to have an efficient replenishment strategy and pricing mechanism to ensure that customers' needs are met while maximizing the supermarket's profit. Historical sales records and customer demand are two key factors to consider when developing a replenishment plan for the day.

Historical sales records can provide valuable information about which vegetables are popular, which time periods are peak sales periods and which vegetables are perishable. By analyzing this data, supermarkets can predict future customer demand and thus determine more accurately the quantity and frequency of replenishment[1]. In addition, historical sales records can help supermarkets develop more targeted pricing strategies, such as raising prices during periods of high demand or offering discounts when sales are weak. In addition to historical sales records, customer buying demand is another important factor in developing a same-day replenishment plan. By observing customers' purchasing behavior and demand, supermarkets can better understand customers' needs and preferences and thus more accurately predict which vegetables will be popular. In addition, by interacting with customers and understanding their needs, supermarkets can better adjust their replenishment strategies and pricing methods to maximize benefits[2]. This is also essential for improving customer satisfaction and maintaining the competitiveness of supermarkets.

2. Materials and methods

2.1. Data acquisition and pre-processing

2.1.1 Source of data

Market research is the source of the data presented in this study.

2.1.2 Data preprocessing

In this study, the number of items in each category is summarized and compared through category analysis to derive the number of items in each category, after which the invalid data is deleted and processed and daily sales data and outliers are processed as shown in the box plot Figure 1.

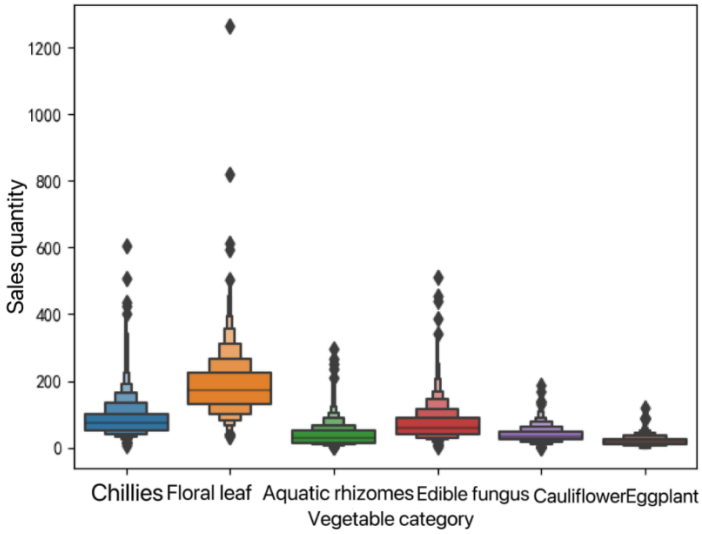

Figure 1. Boxplot of outlier processing results

By analyzing the statistical data of individual products, it was found that the sales volume of most of the individual products fluctuated greatly, which may be due to the seasonal season or the interaction of vegetables of the same type or different categories. Therefore, it is considered that the fluctuation of the sales volume of a single product is a normal phenomenon, and the data of a single product will not be treated as outliers. Afterward, a comparison of the total sales volume of vegetables by category can be made, and it can be found that the sales volume of the flower and leaf category is much higher than that of other types.

2.2. Methodology

This study involves analyzing the distribution patterns and correlations of vegetable sales data in order to develop effective replenishment strategies and pricing methods, as well as to forecast future demand and build optimization models in order to maximize the benefits for the superstore. Firstly, the problem of the distribution pattern of sales data can be studied by data visualization. In turn, we can display the monthly and daily sales volume of each category and single item, so as to observe its sales pattern. Through these graphs, we can visualize the trends and changes in sales data and better understand the sales situation.

Next, Spearman correlation analysis will be used to examine the degree of correlation between the sales data for each category[3]. To analyze the correlation more accurately, the effect of seasonal factors on the sales data needs to be removed first. By removing the seasonal factors, we can reflect the real correlation between sales data more accurately. Next, this study will develop a non-linear programming model to maximize the profit of superstores. The model will be based on historical sales data and future demand forecasts, and optimize the replenishment quantity and pricing strategy to maximize the profit of hypermarkets. The advantage of this model is that it can take into account various factors, including historical sales data, future demand forecasts, costs, prices, etc., to more accurately predict future sales and profits.

Finally, the ARIMA time series model will be used to forecast the demand data for each category for the coming week. The ARIMA model is a commonly used time series forecasting method that automatically identifies trends and cyclical variations in a time series to more accurately predict future demand.

In summary, this study will use data visualization, Spearman's correlation analysis, non-linear programming models, and ARIMA time series models to conduct in-depth analyses of vegetable sales

data to formulate effective replenishment strategies and pricing methods and to maximize the benefits of hypermarkets. The use of these methods will help to improve the management efficiency and profitability of hypermarkets.

3. Modelling and solving

3.1. Study of the distribution patterns of sales of large categories of goods

3.1.1 Monthly Sales Distribution of Major Categories

In order to understand the sales distribution pattern of broad categories, we first separated and aggregated the collected sales data by month. We then visualized the average monthly sales on a line graph. Upon analyzing the data, we found that the sales of all types of vegetables were significantly higher during the period from April to October as compared to other months. This could be attributed to the fact that during this period, a wider variety of vegetables are available to meet the different needs of consumers. Additionally, we observed some disparities in the sales quantities among the various types of vegetables due to differences in their popularity and ways of use. For instance, leafy and chilli vegetables were more popular among consumers.

3.1.2 Distribution of sales of major categories of goods by time of day

To further explore the sales distribution patterns of the major categories, the sales data were separated and aggregated by different hours of the day. The average hourly sales of each major category were visualized through a line graph. It was observed that vegetable sales increased sharply until mid-dinner and then declined, which is consistent with common sense. In the afternoon from 1pm to 2pm, the number of vegetable sales after 8pm was very small, probably because people rested at home after dinner and did not need to go out to buy vegetables again. Meanwhile, the number of vegetable sales in the morning was slightly higher than that in the afternoon before dinner, probably because people were more inclined to buy all the vegetables they needed for the whole day in one go in the morning.

3.1.3 Seasonality of broad commodity groups

Seasonal decomposition of aquatic root vegetables was carried out in order to verify whether vegetable sales are affected by seasonal factors[4]. The seasonal decomposition was carried out by taking one year (365 days) as a cycle and plotting the image is shown in Figure 2.

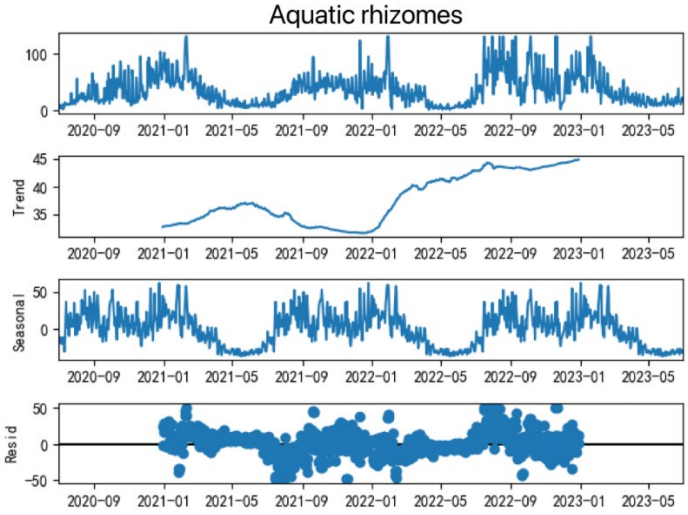

Figure 2. Seasonal decomposition of aquatic root vegetables

The seasonality of sales was found to be significant and showed a gradual increase. Next, the residuals were subjected to the Ljung-Box white noise test and the results are shown in Table 1[5][6].

Table 1. Residual test table for aquatic rhizomes

Lag order	Ljung-Box statistic	P-value
1	258.0724467	4.51526E-58
2	389.6613543	2.43282E-85
3	450.6366696	2.37E-97
4	505.9378215	3.4816E-108
5	574.2677188	7.33E-122

The results show that the residuals belong to the white noise sequence, which indicates that the seasonal decomposition is good. Meanwhile, observing the seasonality results shows that aquatic root vegetables have a low sales period from April to June every year, which is consistent with its monthly sales distribution image. This indicates that seasonality has a certain effect on vegetable sales.

From the above analyses, it can be concluded that the sales of broad categories of goods are affected by seasonal factors and the sales of different categories of goods vary greatly.

3.1.4 Correlation analysis of broad commodity groups

In order to examine the sales relationship between the major categories, the data collected was subjected to Spearman correlation analysis. The correlation coefficients between the major categories were aggregated and used to create a daily sales correlation matrix of the major categories, which is plotted in Figure 3 to provide a more intuitive understanding of the correlation between the major categories and vegetables.

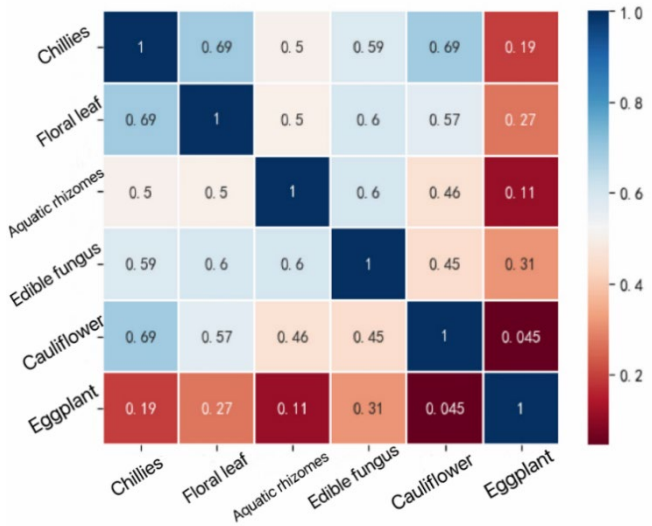

Figure 3. Daily sales correlation matrix for each category

The correlation matrix diagram shows that the relationship between the sales of the main categories is quite complex. According to the analysis, there's a higher correlation between leafy vegetables and fruits and flowers, roots and aquatic roots and tubers, and lycopersicons and pulses. This could be due to the similarities in cultivation and consumption of vegetables between these categories. On the other hand, the correlation between leafy vegetables and lycopersicons and rhizomes and flowers and fruits were found to be low, which could be due to the differences in cultivation and consumption of vegetables between these categories. It was also noted that some of the broad categories showed negative correlations, which could be due to the substitution or complementarity of these commodities in home cooking.

3.2. Study the distribution pattern of single product sales

3.2.1 Comparison of total sales volume of individual products

Comparing the total sales of individual items, we found that the sales of Wuhu green pepper (1), broccoli and net lotus root (1) were much higher than the other individual items, indicating that there was a high demand for them in the market.

3.2.2 Monthly sales distribution of individual products

Due to the large number of individual products in this question, we choose the monthly sales of the top ten representative individual products to analyze the monthly sales distribution pattern of individual products. By analyzing the images, we can see that the sales volume of different individual vegetables have obvious seasonality and different seasonal trends. This may be because the planting and growing of these individual vegetables is affected by different climates and seasons, resulting in certain seasonal differences in their sales.

3.2.3 Individual product sales distribution relationships

The correlation analysis of single items is treated in the same way as that of large categories of goods, i.e., the seasonal decomposition is used to remove the effects of data trends and seasonality on the results of the correlation analysis.

Due to the large number of single items, this paper only shows the analysis results of the top ten single items in terms of sales, which are as follows: Index([' Wuhu Green Pepper (1)', 'Broccoli', 'Net Lotus Root (1) ', 'Chinese cabbage', 'Yunnan lettuce', 'enoki mushrooms (box)', 'Yunnan lettuce (portion) ','Purple Eggplant (2)', 'Xixia Mushroom (1)','Millet Pepper (portion)'], dtype= 'object') and then Spearman correlation analysis of related data.

There is a positive correlation between the sales of some vegetables because they are often added in combinations in home cooking, but there is a negative correlation because some vegetables cooked together make the dish taste strange or cannot be cooked together.

3.3. ARIMA model building and solving

In order to better predict the total amount of replenishment and pricing strategy for the coming week, we first consider building an ARIMA model to predict the vegetable sales for the coming week. ARIMA model is a time series forecasting model that extracts useful information from historical data through difference, autoregressive, and moving average operations, and uses this information to predict future trends[7]. The following are the exact steps of ARIMA model for forecasting vegetable sales:

(1) Data preparation: collect data on vegetable sales to be analyzed and forecasted, usually denoted as Y_t, where t is the index of the time step.

(2) Difference operation: If the time series is not smooth, a difference operation is needed to make it smooth. The difference operation is denoted as(where B is the lag operator and a is the differenced time series):

$$Y_{\mathrm{t}}{}' = (1 - B)^d Y_t \tag{1}$$

(3) Model selection: the AR (AutoRegressive) and MA (MovingAverage) models are selected for the order p and q. These parameters determine the number of autoregressive and moving average components.

(4) Autoregressive part (AR part): the autoregressive part represents the linear relationship between the current time step Y_t and the previous p time steps AR part is expressed as:

$$Yt = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \mathrm{L} + \varphi_p Y_{t-p} + \varepsilon_t \tag{2}$$

(5) Moving average part (MA part): the moving average part represents the linear relationship between the white noise error of the current time step Y_t and the previous q time steps The MA part can be expressed as(where c is the constant term, $\theta_1, \theta_2, \mathrm{L}, \theta_q$ is the moving average coefficient, and e_t is the white noise error at the current time step.

$$Yt = c + \mathrm{e}_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \mathrm{L} + \theta_q e_{t-q} \tag{3}$$

(6) Model fitting: estimation methods (e.g., least squares) are used to estimate the parameters φ and θ of the AR and MA parts, as well as possible constant terms , so that a fit to the ARIMA model is obtained.

(7) Model diagnosis: the fitted ARIMA model is diagnosed to check whether the residuals satisfy the white noise nature to ensure the quality of the model.

(8) Model prediction: a fitted ARIMA model is used to make predictions of vegetable sales data for the coming week.

After building the ARIMA model, the total revenue of the superstore is maximized by optimizing the model parameters [8]. Specifically, we assume a linear relationship between total sales and cost-plus pricing, i.e., total sales = pricing parameter × cost. We then use linear regression to fit this relationship and determine the objective function to maximize total revenue. Finally, an optimization algorithm is used to solve for the optimal total daily replenishment and pricing strategy. The resulting pricing replenishment strategy is shown in Table 2:

Table 2. Pricing replenishment strategy

Notes: T and P are the total replenishment and pricing of commodities for the period July 2023, respectively.

Dates	Pepper		Aquatic		Flower-leaf		Cauliflower		Fungus		Eggplant	
	T	P	T	P	T	P	T	P	T	P	T	P
7/1	78.5	15.2	27.17	13.08	154.6	10.88	21.91	13.89	44.72	9.03	33.6	5.7
7/2	78.5	15.2	36.32	12.76	154.6	10.88	23.48	13.44	46.67	7.98	23.2	9.5
7/3	78.5	15.2	32.82	12.88	154.6	10.88	21.91	13.88	48.15	7.19	22.4	9.8
7/4	78.5	15.2	36.03	12.77	154.6	10.88	21.17	14.09	48.14	7.19	26.4	8.3
7/5	78.5	15.2	34.84	12.81	154.6	10.88	21.91	13.88	47.64	7.46	29.6	7.1
7/6	72.1	16.8	33.38	12.86	154.6	10.88	21.91	13.88	44.94	8.91	28.8	7.4
7/7	69.4	17.5	28.51	13.03	147.3	12.02	12.85	16.41	47.97	7.28	26.4	8.3

3.4. Sensitivity analysis

3.4.1 Patterns of sales distribution by broad categories and by individual products

We will perform a sensitivity analysis of the total daily replenishment and the optimization model used for the pricing strategy for each vegetable category for the coming week[9].

Here we take the first day as an example to test the model sensitivity by varying the quantity of vegetables demanded and the results obtained are shown in Table 3.

Table 3. Linear programming sensitivity analysis

Notes: C is the change in demand, P is the pricing of the products and T is the total replenishment of the products.

C	Pepper		Aquatic		Flower-leaf		Cauliflower		Fungus		Eggplant		Total profit
	T	P	T	P	T	P	T	P	T	P	T	P	
15%	60.62	8.9	18.48	11.6	159.45	8.32	35.7	9.0	47.51	8.6	19.28	8.5	1396
10%	64.18	8.7	19.56	11.4	168.86	8.23	37.8	8.9	50.31	8.5	20.41	8.5	1437
-5%	67.75	8.5	20.65	11.3	178.24	8.16	39.9	8.8	53.11	8.5	21.55	8.5	1479
5%	74.87	8.1	22.82	11.0	197.00	8.02	44.1	8.7	58.69	8.3	23.81	8.5	1562
10%	78.44	8.0	23.9	10.8	206.38	7.97	46.2	8.6	61.49	8.3	24.94	8.5	1603
15%	8.48	7.8	24.99	10.7	215.76	7.91	48.3	8.5	64.28	8.2	26.08	8.5	1645

Through the test, it is learnt that as the demand for vegetables rises, the profit of the supermarket increases gradually, so the model is able to determine the corresponding purchase quantity and pricing of vegetables according to the different demand for vegetables, and the model works well[10].

3.4.2 ARIMA model

The multi-objective optimization problem is constructed to maximize the profit sum and minimize the demand difference, which will be analyzed for model sensitivity in the following:

Due to the large number of individual vegetable categories, we verified the sensitivity of the model by varying the demand for vegetables only to observe the overall profit, and the results obtained are shown in Table 4.

Table 4. Linear programming sensitivity analysis

Changes in demand	15%	-10%	-5%	+5%	+10%	+15%
Total profit	594.5407	936.7701	960.0074	1173.7	1246.6	1412.4

4. Discussion

This study uses a variety of modelling approaches to help supermarkets solve replenishment strategies and pricing problems to improve sales. These methods include data visualisation, Spearman correlation analysis, linear regression, ARIMA models and planning models. These methods enable supermarkets to visualise sales distribution patterns, identify vegetable categories or individual items with highly correlated sales volumes, establish a positive correlation between sales volume and total price, predict future sales trends, and develop more refined replenishment strategies and pricing approaches.

However, the ARIMA model has high data requirements, requiring sufficient historical data and stable sales trends. For some emerging vegetable categories or vegetables with unstable sales trends, the ARIMA model may be less applicable. To improve the reliability of the model, it needs to be adapted and optimised according to market demand, price fluctuations, inventory constraints and other practical situations.

5. Conclusions

By analyzing and processing the sales data of fresh food superstores, this study establishes a relevant mathematical model, aiming to help superstores formulate reasonable replenishment strategies and pricing methods, to improve their revenues. Firstly, by analyzing the historical sales data, it was found that flowering and leafy vegetables were the most popular in the market, and the distribution pattern of the sales data was analyzed by methods such as seasonal decomposition. Secondly, methods such as linear regression and the ARIMA model were used to predict the future sales trend. Finally, a planning model was developed to optimize the data for each day, taking into account a variety of factors to maximize the superstore's revenue. Maximizing the superstore benefits was made possible by these models. Meanwhile, some shortcomings of the models were found, such as the ARIMA model requires sufficient historical data and stable sales trends, which provides a direction for subsequent research.

References

[1] Megan R W, Kathleen M L, Darin J E, Caitlin E C, Melissa N L, et al. Longitudinal Fruit And Vegetable Sales In Small Food Retailers: Response To A Novel Local Food Policy And Variation By Neighborhood Socioeconomic Status[J], International journal of environmental research and public health, 2020, 17(15).

[2] Feng, L., Zhang, J., Tang, W.. Optimal Inventory Control and Pricing of Perishable Items Without Shortages[J], IEEE Transactions on Automation Science and Engineering, 2016, PP(99): 1-14.

[3] Zongshuang W, Jungang L, Yufei T, Min G, Yanyue G, Shu X, Yuhua Z, et al. Temporospatial variations and Spearman correlation analysis of ozone concentrations to nitrogen dioxide, sulfur dioxide, particulate matters and carbon monoxide in ambient air, China[J], Atmospheric Pollution Research, 2019, 10(4): 1203.0-1210.0.

[4] Anne C R K, Stefano F S, Laio O S, Viviana C M, Leandro D S C, et al. Structure Optimization of Ensemble Learning Methods and Seasonal Decomposition Approaches to Energy Price Forecasting in Latin America: A Case Study about Mexico[J], Energies, 2023, 16(3184).

[5] Taehyun No, Taewook Lee. Wild bootstrap Ljung-Box test for residual autocorrelation in vector autoregressive models[J], Journal of the Korean Data And Information Science Society, 2020, 31(3): 477-485.

[6] Myeongwoo Lee, Taewook Lee. Wild Bootstrap Ljung-Box Test For Autocorrelation In Vector Autoregressive And Error Correction Models[J], The Korean Journal of Applied Statistics, 2016, 29(1): 61-73.

[7] Raydonal Ospina,João A. M. Gondim,Víctor Leiva,Cecilia Castro. An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil[J]. Mathematics,2023,11(14).

[8] Surindar G W, Malik J, P. N K, Thomas F, Jyoti B, Roop R, Sathyapriya E, Rajasekhar B, et al. Minimizing the Error Gap in Smart Framing by Forecasting Production and Demand Using ARIMA Model[J], Journal of Food Quality, 2022, 2022.

[9] RAZAK, NURUL HANIM, HASHIM, HASLENDA, YUNUS, NOR ALAFIZA, et al. Integrated linear programming and analytical hierarchy process method for diesel/biodiesel/butanol in reducing diesel emissions[J]. 2022,337(Feb.20):130297.1-130297.21.

[10] Dimitrios Magos. Linear Programming Sensitivity Analysis And Related Topics[J], INTERFACES, 2012, 42(1): 89-90.