

An Investigation into the Pricing and Replenishment Strategy of Vegetable Commodities in Fresh Food Shops

Jianan Huang *

School of Mathematical Sciences, South China Normal University, Guangzhou, China, 510631

* Corresponding Author Email: 20212233006@m.scnu.edu.cn

Abstract. Fresh supermarkets face challenges in marketing and inventory management due to the short shelf life and high loss rate of fresh vegetables. This necessitates the immediate development of a practical replenishment plan and pricing strategy. The paper commences by analyzing the sales distribution patterns of various vegetable categories, identifying leaves and flowers as having notably higher sales compared to other categories. Subsequent examination of the sales trend over time reveals variations in the sales of different categories or individual vegetables across months or quarterly periods, indicating a degree of seasonality. This study calculates the Spearman correlation coefficient for each category and individual product sales, presenting the results in a heat map. Findings indicate a significant positive correlation among all vegetable categories except for solanaceae, implying bundled sales or a consumer tendency to purchase specific vegetable categories concurrently. Consequently, when formulating replenishment and pricing strategies, considering a rational sales mix is crucial. The paper constructs a linear regression model, solving the equation using the least squares approach, and identifies a negative association between sales volume and cost-plus pricing. To accommodate seasonal patterns, the Prophet time series model forecasts wholesale prices for various vegetable categories during July 1-7, 2023. The objective is to maximize profit by accurately predicting vegetable costs. A nonlinear programming model is formulated to achieve this objective, determining the maximum revenue from days 1 to 7 as $\text{¥}(108953.65 - K)$, where K represents the fixed indirect cost. The model optimizes replenishment and pricing strategies for different categories.

Keywords: Vegetable Commodities, Pricing Strategy, Spearman's Correlation Coefficient, Prophet's Time Series Model, Nonlinear Programming Model.

1. Introduction

Vegetable commodities are essential for people's livelihoods, characterized by being seasonal, regional, perishable, and having a short shelf life. With the advancement of the national economy, people's consumption level has also increased, leading to a growing need for fresh veggies. Addressing inventory issues such as vegetable stagnation, depletion, and stock shortages is currently a pressing concern in order to optimize revenue. In their study, Cui Ligang et al [1] developed a model that combines replenishment and pricing for fresh produce, taking into account the investment in preservation technology. They also suggested an adaptive differential evolutionary algorithm to solve the model and successfully demonstrated the effectiveness of the method. Fang Juhan et al [2] integrated the sales data of holiday season-sensitive items in the e-commerce market and employed Prophet's model to forecast the need for courier boxes. They subsequently devised a multi-stage dynamic ordering strategy for businesses. The authors Yuan Yuan et al [3] employed a multiple and multi-criteria ABC classification method to categorize apparel and devised an apparel ordering strategy using a gray prediction model based on the classification outcomes. Similarly, Zhang Liping et al [4] utilized a cross-programming model to depict the real supply and demand relationship and developed a hybrid algorithm incorporating the relaxation method and barrier function method to solve the model. Currently, there is limited research on the pricing and replenishment strategy of vegetable commodities. This paper aims to address this gap by utilizing sales data from a supermarket over recent years. The objective is to develop a robust market demand estimation and create an automated decision model for pricing and replenishment of vegetable commodities. Additionally, the

study aims to formulate a scientifically sound and efficient replenishment plan and pricing strategy that maximizes supermarket profits while attempting to meet market demand.

2. Analyze sales statistics for merchandise in the vegetable category

The information presented in this document is sourced from the website https://dxs.moe.gov.cn/zx/a/hd_sxjm_sthb/230523/1840580.shtml. It includes statistics on six different kinds of vegetables, as well as detailed information on sales flow, wholesale prices, and current wastage rates of vegetable goods.

2.1. Sales distribution pattern of vegetable commodities

Given the extensive range of vegetable product categories included in the appendix, traditional visual analysis methods like bar charts, line graphs, and pie charts are inadequate for examining the percentage link between the sales of each product and determining the distribution pattern. Upon examination, it was discovered that the rectangular tree diagram is more space-efficient than other charts when comparing the length and proportion of chart size. The rectangular tree diagram effectively displays the distribution of sales for a single product through its area. Figure 1 illustrates the precise scenario.

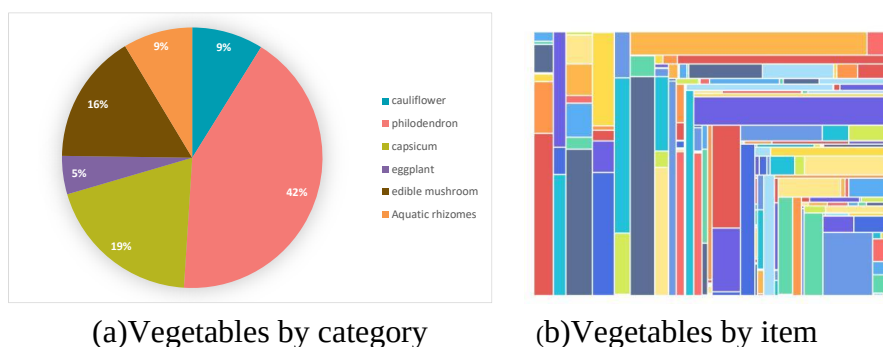


Figure 1. Share of market volume

Given the inherent nature of life, sales typically exhibit temporal fluctuations. Consequently, the month and quarter are chosen as temporal units, and graphs are created to examine the correlation between the sales of different categories and specific items over time, as depicted in Figure 2 and Figure 3. The sales data reveals distinct seasonal patterns in the sales of various categories of vegetables or individual vegetable products. For instance, sales of leaves and flowers and peppers are highest during the summer months, with peak sales occurring in August. Conversely, sales of static lotus root and Chinese cabbage are lower during the summer. Furthermore, leaves and flowers vegetables consistently outperform other categories of vegetables in terms of sales volume. This can be attributed to their vibrant colors, as indicated by relevant information. Leaves and flowers vegetables are more appealing to customers than other types of vegetables due to their cost-effectiveness and versatile applications in cooking.

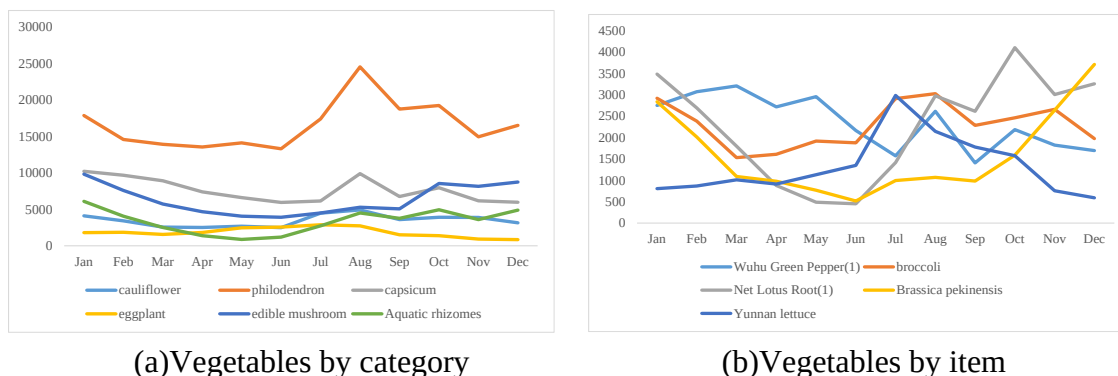


Figure 2. Sales patterns over a period of months

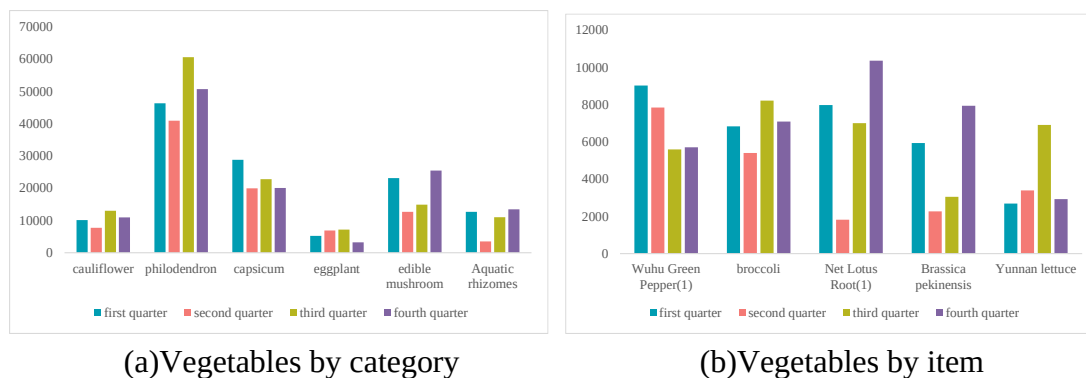


Figure 3. Sales patterns over a period of seasons

To more accurately assess the method of analyzing the correlation between sales of each category and individual product, it may be beneficial to complement the test of normality of sales with the initial assumption that sales of each category roughly follow a normal distribution, solely from a qualitative perspective.

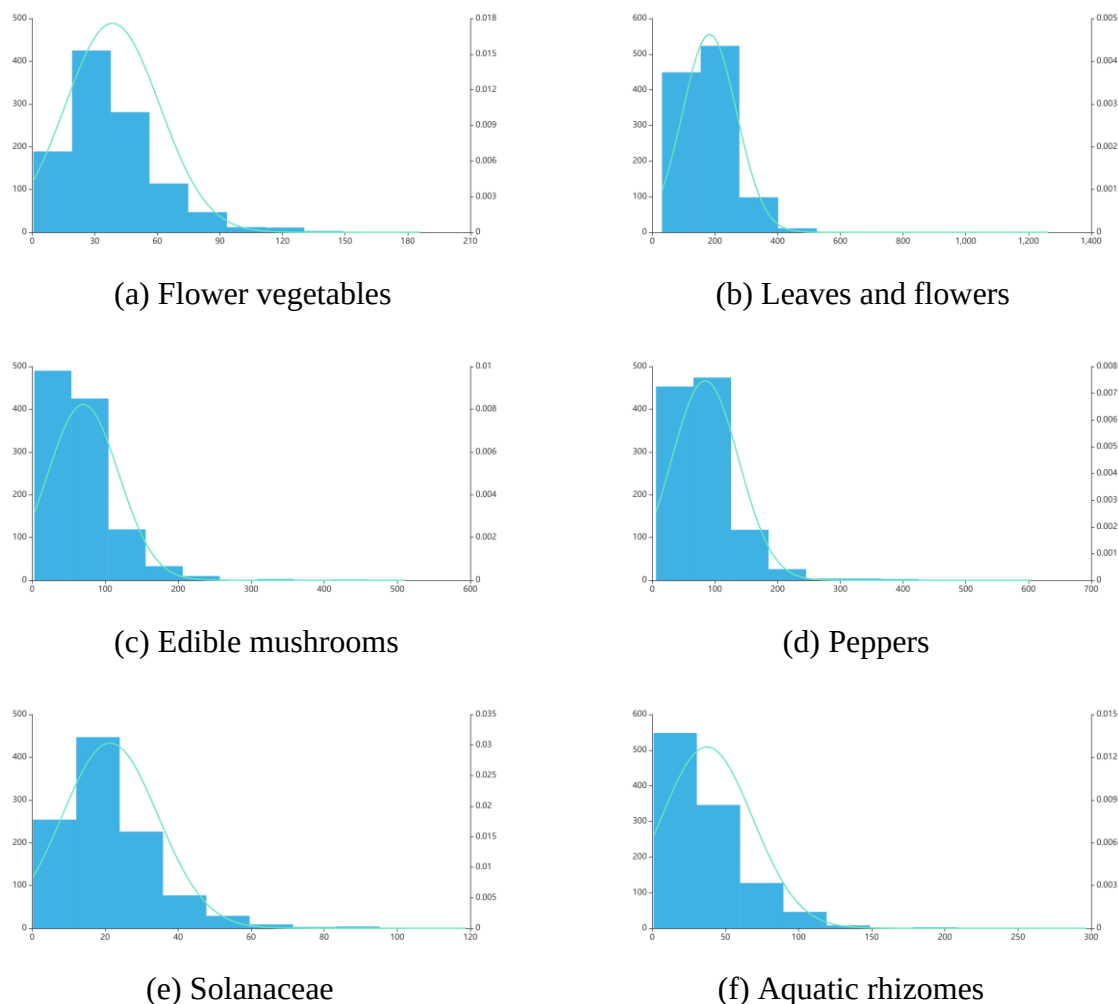


Figure 4. Normality test histogram for sales categorized by category.

Figure 4 and Table 1 display the results of descriptive statistics and normality tests for each category of this superstore. These tests include measures like as median and mean, which are utilized to assess the normality of the data. Given that the sample size for each category is less than 5000 ($N < 5000$), the S - W test is employed in this paper. The results indicate that the P-value for each category is less than 0.05, signifying statistical significance. This leads to the rejection of the initial hypothesis, suggesting that none of the data meet the criteria for a normal distribution.

Table.1. Summary of findings

category	median	average	standard deviation	skewness	kurtosis	S-W test
Solanaceae	18.88	21.38	13.17	1.74	5.89	0.89(0.000***)
Flower vegetables	33.86	38.28	22.79	1.56	4.34	0.90(0.000***)
Leaves and flowers	171.35	180.72	85.33	3.07	28.56	0.84(0.00***)
Peppers	72.19	83.55	53.50	3.25	19.20	0.76(0.000***)
Edible mushrooms	56.96	69.76	48.90	3.03	17.01	0.78(0.000***)
Aquatic rhizomes	30.02	36.93	31.22	2.59	12.91	0.80(0.000***)

Note: ***, **, * represent 1%, 5% and 10% significance levels, respectively.

2.2. Spearman's correlation analysis

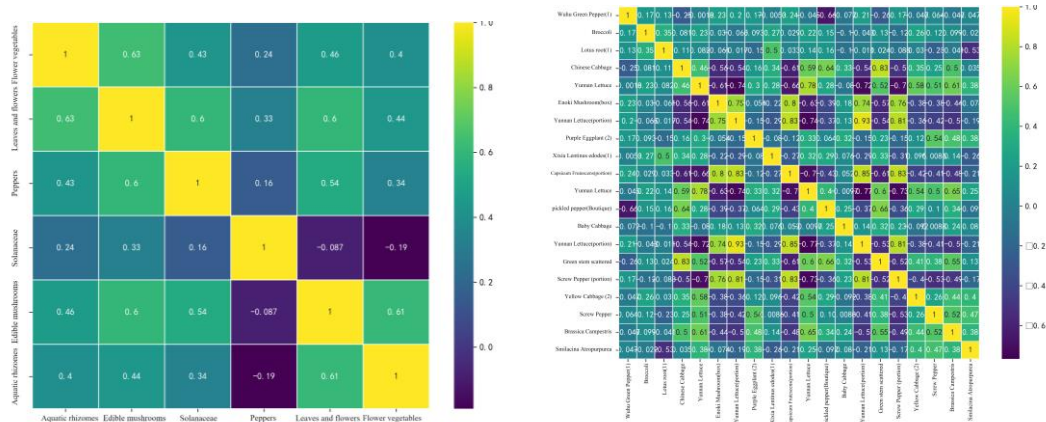
The two primary correlation coefficients frequently employed in statistics to quantify the correlation between data are Pearson's correlation coefficient and Spearman's correlation coefficient. These two distinct correlation coefficients are suitable for various categories of data. Based on the study results, it is seen that the sales volume of each category does not follow a normal distribution. Consequently, Spearman's correlation coefficient is selected as a metric to assess the relationship between the sales volume of each category and individual products.

The Spearman's correlation coefficient is a statistical measure used to quantify the degree of correlation between two variables. It provides a numerical value that ranges from -1 to 1, indicating the strength and direction of the relationship between the variables. A Spearman's correlation value near to -1 implies a robust negative correlation between the two variables. This means that if one variable increases, the other variable tends to decrease. A Spearman correlation value near to 1 signifies a robust positive correlation between the two variables, meaning that if one variable grows, the other variable also tends to increase. When the Spearman correlation value approaches 0, it indicates the absence of a linear correlation between the two variables or a very weak correlation that does not allow for a definitive conclusion [5-6].

The calculation of Spearman's correlation coefficient is determined by the formula.

$$\rho_s = \frac{\sum_{i=1}^N (R_i - \bar{R})(S_i - \bar{S})}{[\sum_{i=1}^N (R_i - \bar{R})^2 \sum_{i=1}^N (S_i - \bar{S})^2]^{\frac{1}{2}}} \quad (1)$$

Let R_i and S_i be the ranks of the values observed for i . $\bar{R}$ and $\bar{S}$ are the average ranks of x and y , respectively. N represents the total number of observations.



(a) Vegetables by category

(b) Vegetables by item

Figure 5. Heat map of Spearman's correlation coefficient

Figure 5 demonstrates the significant inter-category sales for several vegetable commodities. Among them, there is a strong and positive correlation between the sales of leaves and flowers and flower vegetables, with a Spearman correlation coefficient of 0.63. On the other hand, the

interrelationship between the sales of solanaceae and edible mushrooms, as well as aquatic rhizomes, is relatively weak and shows negative correlation, with Spearman correlation coefficients of -0.087 and -0.19 respectively. For individual products, only the highest sales of three years are considered in this study. This study focused on analyzing the relationship between different single products by selecting the top 20 products with the highest sales over the past three years. The findings revealed varying correlations among the products, with some demonstrating a strong positive correlation, as indicated by a high Spearman correlation coefficient of 0.83. Conversely, some products exhibited a negative correlation, with an absolute value of the Spearman correlation coefficient as low as 0.04. In summation, it may be preliminarily determined that: in the process of later development of replenishment and price strategies, an acceptable sales mix has become vitally crucial.

3. Development of pricing and replenishment strategies for vegetable products

This article involves the daily calculation of cost-plus pricing for several categories during the data preparation stage. Cost-plus pricing is a pricing strategy that takes into account the cost of production; however, it is important for merchants to also consider market demand when developing pricing strategies. Regression analysis facilitates merchants in comprehending the correlation between cost-plus pricing and sales volume, and subsequently ascertaining the magnitude of influence that various pricing levels have on sales volume. Consequently, merchants can enhance their comprehension of market demand and formulate a rational pricing strategy by considering both market demand and cost.

Consider the foliage as an illustration. Let's designate the selling price as the independent variable x and the sales volume under the selling price as the dependent variable y . The objective equation can be expressed as $y = a * x + b$, and we can proceed to construct a linear regression model. To illustrate the correlation between the selling price and sales of foliage vegetables, a scatter plot was created as depicted in Figure 6. Each point on the scatter plot (x_i, y_i) represents the distribution of a sample from the leaves and flowers vegetables dataset on a Cartesian coordinate system. It shows the relationship between the dependent variable (sales) and the independent variable (selling price), as well as the overall trend of change.

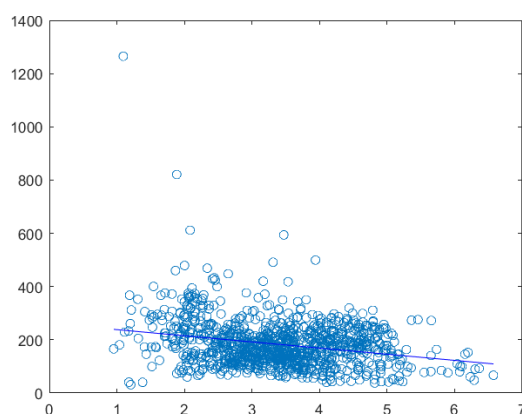


Figure 6. Correlation between Sales and Costs

In the analysis of variance (ANOVA), the F-test statistic has a value of 17.59 and the associated P-value is 0.000***. This P-value is considered statistically significant, indicating strong evidence against the null hypothesis that the regression coefficient is equal to zero. The model adequately fulfills the specified requirements. The covariate covariance performance indicates that all VIF values are below 10, indicating that the model does not have any multicollinearity issue. Based on the aforementioned test findings, the regression model may be established as follows.

$$y = 215.886 - 5.53 * P \quad (2)$$

Let P represent the selling price. Regression model formulations for other categories of vegetable commodities were derived using a similar approach:

Table.2. Formulas for regression models specific to each category

Category	Formula
Solanaceae	$y = 21.543 - 0.018 * P$
Flower vegetables	$y = 39.523 - 0.101 * P$
Leaves and flowers	$y = 215.886 - 5.53 * P$
Peppers	$y = 93.492 - 0.994 * P$
Edible mushrooms	$y = 112.758 - 4.787 * P$
Aquatic rhizomes	$y = 40.52 - 0.302 * P$

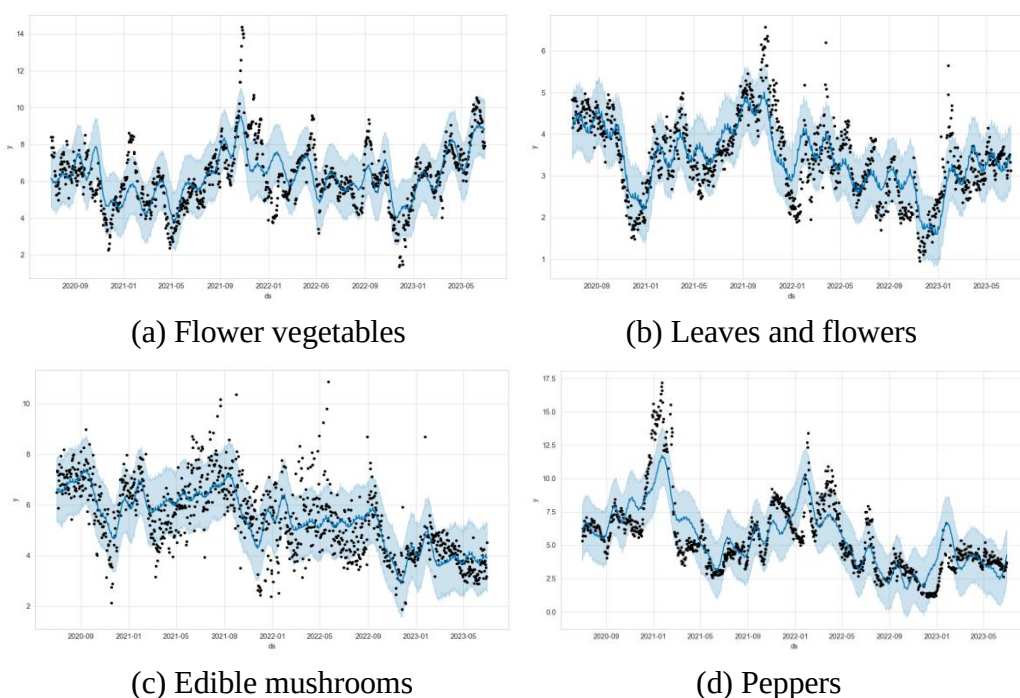
The regression model formulas presented in Table 2 establish the correlation between sales volume and unit pricing across several types of commodities. These regression models can be utilized to forecast the sales volume at various pricing levels and to formulate a logical pricing strategy accordingly.

3.1. Prediction of Future Wholesale Prices with Prophet Time Series Modeling

To optimize the advantages for the merchants, it is important to acquire the wholesale prices for various vegetable categories for the upcoming seven days in order to construct the model. Commonly employed techniques for predicting future market conditions encompass time series forecasting, gray forecasting, and Markov forecasting. Due to the inherent uncertainty in predicting wholesale prices for various kinds of vegetables, the time series forecasting algorithm offers a notable benefit as it can more accurately model the past data of these time series.

This study employs the Prophet algorithm for the purpose of predicting. In contrast to conventional time series forecasting techniques [7-9], the Prophet algorithm possesses the capability to automatically identify and represent many seasonal patterns, including various time scales such as daily, weekly, monthly, and yearly. This has the ability to more accurately reflect the fluctuating patterns of vegetable prices over various time intervals. Additionally, the Prophet algorithm exhibits superior capabilities in modeling nonlinear trends and can provide more precise predictions of price movements [10]. The Prophet algorithm utilizes past wholesale pricing data to train the model, enabling it to forecast the wholesale prices of several vegetable groups over the upcoming seven days. The valuable insights acquired regarding future market conditions might assist merchants in formulating rational pricing plans to optimize gains.

3.2. Application of the Prophet Model to Wholesale Prices



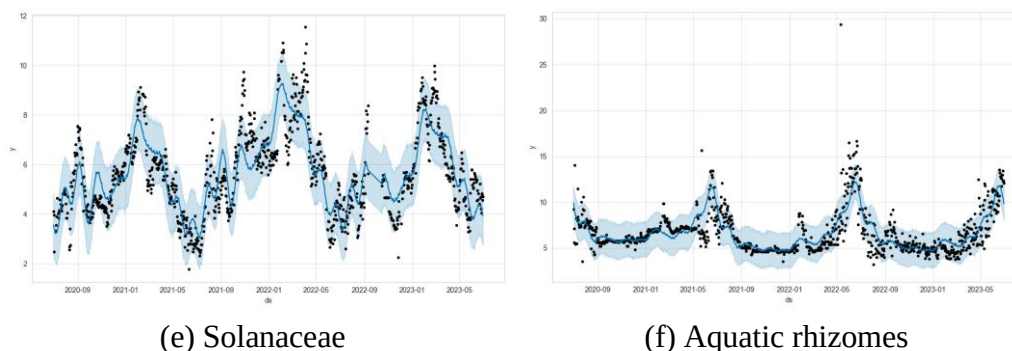


Figure 7. Prophet model fit by category

The black dots in Figure 7 represent real data on historical wholesale prices, the dark blue line represents the Prophet model's fit to this data, and the light blue intervals represent the range of predicted intervals, where each timestamp has a predicted upper limit point and a predicted lower limit point, and the range in between is the predicted interval at that timestamp. The trend of the wholesale price data can be understood by looking at the fitted confidence values. For each timestamp, the confidence value (i.e., the dark blue point) represents the mean of the prediction interval (i.e., the light blue interval). In addition, this paper calculated the goodness of fit for each category, using MAPE (Mean Absolute Percentage Error) as an assessment metric [11]. The results are shown in Table 3.

Table.3. Assessing the adequacy of fit for each category

Category	MAPE
Solanaceae	0.14
Leaves and flowers	0.15
Flower vegetables	0.16
Peppers	0.24
Edible mushrooms	0.15
Aquatic rhizomes	0.13

The goodness-of-fit results evaluate the precision of the Prophet model in forecasting various categories. A lower Mean Absolute Percentage Error (MAPE) number signifies a more accurate forecast by the model, whereas a higher MAPE value suggests the presence of potential errors in the model's prediction. Based on the values, it is evident that the Prophet model accurately forecasts for each category, with minimal discrepancy between the predicted outcomes and the actual data. Table 4 displays the projected outcomes of the wholesale pricing for several categories in the market from July 1 to July 7, 2023.

Table.4. Wholesale Price Trends for Various Market Categories from July 1-7, 2023

Category	July 1	July 2	July 3	July 4	July 5	July 6	July 7
Solanaceae	3.85	3.88	3.73	3.62	3.62	3.56	3.54
Peppers	4.50	4.51	4.52	4.74	4.87	4.95	5.05
Edible mushrooms	3.91	3.89	3.88	3.98	3.82	3.80	3.82
Leaves and flowers	3.64	3.63	3.58	3.60	3.62	3.65	3.68
Flower vegetables	8.86	8.85	8.69	8.69	8.77	8.74	8.75
Aquatic rhizomes	9.49	9.35	9.02	8.83	8.70	8.39	8.15

3.3. Optimization Modeling

The aim of this study is to optimize merchant benefits through an optimization model that involves making judgments on the mix of replenishment volume and pricing for different categories on a daily basis. During the previous discussion, we have established the correlation between sales volume and pricing for various categories from July 1st to July 7th. Additionally, we have obtained the wholesale prices for these seven days. This paper presents a linear programming model that uses the cost-plus pricing method to predict the pricing and replenishment volume of different categories for each day,

building upon the prior discussion. Table 5 provides a precise definition of the variables that are utilized in the modeling procedure.

Table.5. Assessing the adequacy of fit for each category

Notation	Definition
K	Fixed overhead expenses
x_i	Price of category i
c_i	Wholesale price of category i
s_i	Sales volume of category i
b_i	Incoming quantity of category i
q_i	The slope coefficient represents the relationship between category i and the selling price in the regression equation.
p_i	The intercept coefficient of the regression equation for category i and selling price.

3.3.1 Determination of the objective function

This model is an optimization model with the purpose of maximizing profit, denoted as I. Profit is the result of subtracting costs from sales proceeds, and it may be mathematically expressed as:

$$\text{Max} \cdot I = \sum_{i=1}^6 x_i \times (p_i - q_i \times x_i) - c_i \times b_i - K \quad (3)$$

The number of sales can be estimated by utilizing historical sales data and a forecasting model, denoted by " i " which represents 6 categories (i = 1, 2, ..., 6). The wholesale price can be calculated using the previously solved regression model. The cost, referred to as K' , includes expenses such as labor and energy, in addition to the wholesale cost. As an illustration, we will focus on maximizing the profit on the first day, and the calculations for the remaining 6 days can be done in a similar manner.

3.3.2 Determination of constraints

The subsequent limitations can be derived within the framework of practical existence:

- (1) The quantity of sales is determined by the price.
- (2) To effectively manage the balance between profit and wastage, it is important to ensure that the difference between the selling price and the product of the volume of incoming shipments and the price of arriving shipments does not exceed the product of the volume of incoming shipments and (1 - wastage rate / 100).

3.3.3 Summary of optimization models

Based on the analysis provided above, the following summary is presented:

$$\text{Max} \cdot I = \sum_{i=1}^6 x_i \times (p_i - q_i \times x_i) - c_i \times b_i - K \quad (4)$$

st.

$$\begin{cases} s_i = p_i - q_i \cdot x_i \\ p_i - q_i \cdot x_i \leq b_i * (1 - \text{wastage rate}/100) \end{cases}$$

The optimization model in this work is solved using the Lingo program. The maximum return (in dollars) for the period of July 1-7, 2023, is calculated as follows:

$$\text{Max } I = 15409.06 + 15411.24 + 15432.10 + 15427.61 + 15429.80 + 15433.00 + 15430.84 - K = 108953.65 - K$$

Tables 6 and 7 provide the most favorable amounts for restocking from July 1-7, 2023, along with corresponding prices:

Table.6. The most favorable amounts for restocking from July 1-7, 2023

Category	July 1	July 2	July 3	July 4	July 5	July 6	July 7
Leaves and flowers	108.20	108.24	108.43	108.35	108.28	108.17	108.09
Flower vegetables	22.06	22.06	22.07	22.07	22.06	22.06	22.06
Solanaceae	11.58	11.58	11.58	11.58	11.58	11.59	11.59
Peppers	47.76	47.75	47.83	47.89	47.89	47.93	47.94
Aquatic rhizomes	21.42	21.41	21.41	21.37	21.35	21.33	21.32
Edible mushrooms	50.62	50.68	50.70	50.43	50.87	50.94	50.89

Table.7. The most favorable amounts for corresponding prices from July 1-7, 2023

Category	July 1	July 2	July 3	July 4	July 5	July 6	July 7
Leaves and flowers	21.59	21.55	21.52	21.53	21.55	21.56	21.58
Flower vegetables	3.95	3.95	3.95	3.95	3.95	3.95	3.95
Solanaceae	6.03	6.03	6.03	6.03	6.03	6.03	6.03
Peppers	9.14	9.90	9.90	9.90	9.90	11.45	11.45
Aquatic rhizomes	6.96	6.96	6.96	6.97	6.98	6.98	6.99
Edible mushrooms	13.91	13.91	13.90	13.96	13.87	13.86	13.87

4. Conclusions

This work presents the development of a linear regression model to address the issues of rapid depletion and limited shelf life of vegetable products. Additionally, a negative association between sales volume and cost-plus pricing was identified. A Prophet time series model was implemented to forecast the wholesale prices of various vegetable categories for the period of July 1-7, 2023. A nonlinear programming model was created to determine the maximum income, optimal replenishment volume, and pricing for different categories ranging from 1 to 7 days. The revenue is calculated as $\text{¥}(108953.65 - K)$, where K represents the fixed indirect cost.

Given the presence of significant correlations among certain items in the data, which can be bundled together and sold as a sales mix, it is reasonable to take into account the additional value of this sales mix when making future price decisions. This aspect was overlooked during the previous discussion on pricing. During the linear regression analysis, it was seen that varied sales volumes were associated with the same price, indicating that the price alone was insufficient to account for the variation in sales volume. To enhance the model's accuracy, the inclusion of a temporal component should be contemplated. Temporal considerations can significantly influence sales volume, including seasonal fluctuations, market trends, and other relevant aspects. The inclusion of a temporal variable allows for a more comprehensive representation of these potential influences, hence enhancing the precision of the model.

References

- [1] Cui Ligang, Li Yali, Liu Jinxing et al. Joint decision making for multi-product replenishment and pricing considering investment in preservation technology[J]. Industrial Engineering and Management, 2023, 28(03): 17-26.
- [2] FANG Juhan, DING Jingzhi. Ordering strategy of express box in e-commerce warehouse of company A [J]. China Logistics and Purchasing, 2023(03): 78-80.
- [3] YUAN Yuan, GAO Bo. Research on data-driven apparel logistics classification and ordering strategy[J]. Logistics Engineering and Management, 2022, 44(04): 29-32.
- [4] ZHANG Liping, ZHANG Meng, BAI Yu. Cross-planning based ordering strategy optimization model and solution algorithm [J]. Journal of Beijing Architectural Engineering Institute, 2012, 28(02): 64-67.
- [5] Yu Qun, Huo Xiaodong, He Jian et al. Trend prediction of power outages in China based on Spearman's correlation coefficient and system inertia [J]. Chinese Journal of Electrical Engineering, 2023,43(14): 5372-5381.
- [6] Yuan Li, Yutian Liu, Li-Wei Feng. Nonlinear dynamic process feature extraction and fault detection based on Spearman correlation analysis [J]. Journal of Shandong University of Science and Technology (Natural Science Edition), 2023, 42(02): 98-107.
- [7] MEHMET G U, RESAT K. Cases of Residual Types in Diagnostic Checking for ARMA Model [J]. Hacettepe Journal of Mathematics and Statistics, 2019, 43(3): 543-552.
- [8] NWOGU E C, WU H C. Arima Model of Crude Oil Exports in Nigeria [J]. Journal of the Nigerian Association of Mathematical Physics, 2010, 17: 285-292.

- [9] HEITHAM A H, HASHEM A, TIMOTHY R, et al. Forecasting the Jordanian Stock Index: Modelling Asymmetric Volatility and Distribution Effects within a GARCH Framework [J]. Copernican Journal of Finance & Accounting, 2015, 4(2): 9-25
- [10] Ma Xiaoqin, Li Na, Luo Hongjiao et al. A short-term forecasting method for user electricity consumption based on Prophet [J]. Information Technology, 2023(08): 29-34.
- [11] LIU Yang, ZHOU Haoyue, LU Jinqi et al. Prediction of influenza-like illness trend based on Prophet model [J/OL]. Disease Surveillance: 1-7 [2024-01-24].