

Automated Pricing and Replenishment Decisions for Vegetable Products Based on Data Analysis

Yefan Shen^{1,*}, Yongkang Zhao¹, Xinwei Hao²

¹ School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, China, 121001

² Department of Civil and Architectural Engineering, Liaoning University of Technology, Jinzhou, China, 121001

* Corresponding Author Email: shenyefan0793@163.com

Abstract. This study delves into the critical factors for the successful operation of vegetable products supermarkets: efficient inventory management and accurate market demand analysis. Given the perishability of vegetables, their variety, and the complexity of supply chains, a linear regression model and the method of least squares have been developed to estimate model parameters. Additionally, the ARIMA model is utilized to forecast the unit prices of commodities. In a more comprehensive approach, considering various influencing factors, a combinatorial optimization algorithm (integrating linear regression and greedy algorithm) is employed to determine the optimal replenishment strategy. Based on this, the study forecasts the sales volume and unit prices for various categories over the next seven days, and accordingly, establishes the pricing strategy and predicted replenishment volume for each category. The research underscores the necessity for retailers to precisely analyze sales data and market trends, optimize supply chain management, and formulate effective pricing strategies to cope with seasonal variations and limited sales space. These measures are essential to satisfy consumer demand and enhance profitability.

Keywords: Pearson Correlation Coefficient, Linear Regression Model, ARIMA Model, Greedy Algorithm.

1. Introduction

The operation of fresh produce sales in supermarkets poses a multitude of challenges, such as inventory management, pricing strategies, and supply chain management. To achieve success, merchants must accurately discern market demand, optimize supply chain operations, and implement rational pricing strategies to augment profits and cater to consumer exigencies.

To delve into the interrelationships among the commodities outlined in the aforementioned context, along with the pricing and replenishment strategies for each product, we conducted an extensive review of pertinent literature.

The literature [1-2] uses the dynamic programming approach to analyze the retailer's ability to maximize benefits in a given situation [3], analyzing the advantages and disadvantages of traditional pricing methods, the Q-learning mechanism is used to solve the dynamic pricing problem to adapt to the market demand and competitive environment. Article [4] describes the prediction method of ARIMA model, which also provides modeling ideas for us to solve the problem [5-6]; both use an ARIMA model to forecast commodities, which leads to predictions [7]. On the other hand, accomplishes the prediction of pricing and replenishment of goods through stochastic optimization models; Literature [8] is more accurate and informative compared to other literature single model approaches, which are based on a comprehensive model of multiple algorithms. literature [9-10] both consider dynamic pricing strategies, which include, but are not limited to the way discounts.

The aforementioned literature exhibits various strengths and weaknesses. By amalgamating these methodologies, we have formulated fresh perspectives on addressing pricing and replenishment strategies. Through the establishment of corresponding mathematical models, we have unveiled the interdependencies among diverse commodities. Furthermore, to achieve maximum profitability, we have provided price forecasts and sales projections for the upcoming seven days. Additionally,

grounded in our replenishment strategy, we have delineated pricing strategies for each individual item and predicted the replenishment quantities.

2. Model Development and Implementation

2.1. Parameter Prediction in Linear Regression

Linear regression is a type of regression analysis that models the relationship between one or more independent and dependent variables using a least square function called a linear regression equation. A multivariate linear regression model is expressed in the following form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, i = 1, \dots, n, \quad (1)$$

the variable Y_i represents the dependent variable: the sales volume we aim to predict; $X_{i1}, X_{i2}, \dots, X_{ip}$ stands for the independent variables: factors affecting sales volume, such as price; $\beta_0, \beta_1, \dots, \beta_p$ denotes the model's parameters; and ε_i signifies the error term.

In order to simplify the model and avoid overfitting, we chose several features most relevant to sales volume as independent variables.

The initial objective of regression analysis is to estimate the parameters of the model in order to achieve the best fit for the data. Among the different criteria for determining the best fit, the least squares method is superior, so we use the least squares method to estimate the parameters of the model, which represent the strength of the relationship between the factors affecting sales volume and sales volume.

$$\hat{\beta} = (X^T X)^{-1} X^T y. \quad (2)$$

2.2. Establishment of the ARIMA Model

ARIMA model (Autoregressive Integrated Moving Average model), which is a common model for time series forecasting. The ARIMA model is an extension of the ARMA model with the addition of a differencing process in order to make non-stationary data smooth.

The basic idea of ARIMA is to consider the time series as a process that contains both deterministic and stochastic components. By performing differencing, autoregressive, and moving average operations on it, the model is able to capture the intrinsic structure and patterns of the time series. Once modeled, we can predict the future values of a time series based on its past values. Here, we attempt to predict the future unit price of a commodity based only on the unit price data up to the current month.

The ARIMA (p, d, q) model can be expressed as:

$$\hat{p}^{\{t\}} = p_0 + \sum_{j=1}^p \gamma_j p^{\{t-j\}} + \sum_{j=1}^q \theta_j \varepsilon^{\{t-j\}}, \quad (3)$$

here p is the order of the autoregressive model (AR), q is the order of the moving average model (AM), $\{t\}$ is the error term between times ranging between t and $t - 1$, γ and j are the fit coefficients, and p_0 is a constant term.

The choice of p , q within the ACF and PACF modules respectively, ACF (Autocorrelation Function) and PACF (Partial Correlation Function) are both functions that evaluate the linear relationship between historical data and current values. The ACF is given by:

$$\rho(q) = \frac{\text{Cov}(X_j, X_{j-q})}{\text{Var}(X_0)} = \frac{\frac{1}{n-q} \sum_{j=q+1}^n (x_j - \bar{x})(x_{j-q} - \bar{x})}{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2}, \quad (4)$$

using the ARIMA model, we can accurately predict prices for the day ahead based on historical data. To improve our predictions, we have adjusted our grid strategy to ensure that it adjusts the grid position based on the ARIMA predictions. In the model we have used, the displacement of the grid is determined based on the weighted sum of the long-term indices of the MA and the short-term indices of the ARIMA model. The displacement of the grid can be calculated using the following equation:

$$\tilde{p}_i^{\{t\}} = p_i + \omega (MA^{\{t\}}(N) - MA^{\{t-1\}}(N)) + \mu (ARIMA^{\{t+1\}} - ARIMA^{\{t\}}), \quad (5)$$

during the lattice moving process, the parameters ω and μ are assigned corresponding weights to the two metrics. In the backtesting stage, these two parameters will be adjusted adaptively. At first, we set the initial value of ω to be 0.3, and the initial value of μ to be 0.3 as well.

Therefore, the decision-making is based on the adjusted lattice model with the aim of predicting the price of the commodity more accurately.

2.3. Combinatorial optimization methods

We develop a specific single-item replenishment plan and pricing strategy for a future date based on constraints. We can think of this as a combinatorial optimization problem in which the objective is to maximize revenue while satisfying a set of constraints. Objective function:

$$\max W_i = \Sigma(S_i \cdot (-C_i) \cdot (1 - L_i)), \quad (6)$$

wherein, W_i represents the revenue, S_i denotes the projected sales volume of the commodities, P_i signifies the selling price of the products, C_i indicates the cost price of the merchandise, L_i represents the loss rate of the items, and i represents the individual sale items, ranging from 1 to 251.

The total number of saleable items X_i needs to be controlled at 27-33, and 251 vegetable items are sold in random combinations on a daily basis, so a Boolean quantity λ_i ($\lambda_i = 0$ or $\lambda_i = 1$) is introduced, and λ_i is a 0-1 variable, where i is each saleable item, taking the values of 1,2,3,...,251. the formula is:

$$\lambda_i = \begin{cases} 0, & \text{Unsold items of the day } i \\ 1, & \text{Selling items of the day } i \end{cases} \quad (7)$$

when λ_i is 0, it means that single item i is sold on that day; when λ_i is 1, it means that single item i is not sold on that day. Satisfy the constraints given in the question: the total number of sellable items x_i is required to be controlled at 27-33 items.

$$X_i = \sum_{i=1}^{251} \lambda_i, \quad (8)$$

$$27 \leq X_i \leq 33, \quad (9)$$

the order quantity E for each individual item must at least meet the minimum display quantity of 2.5 kg, i.e.:

$$E_i \geq 2.5, \quad (10)$$

In summary, the optimization model constructed for the study is shown below:

$$\begin{cases} 27 \leq X_i \leq 33 \\ E_i \geq 2.5 \\ S_i \geq 0 \\ P_i \geq 0 \\ C_i \geq 0 \\ S_i, P_i, C_i \in R \end{cases} \quad (11)$$

3. Results

3.1. Vegetable commodity-related analysis

First, for the data given in the question, we visualize it to get the visual image of the vegetable singles.

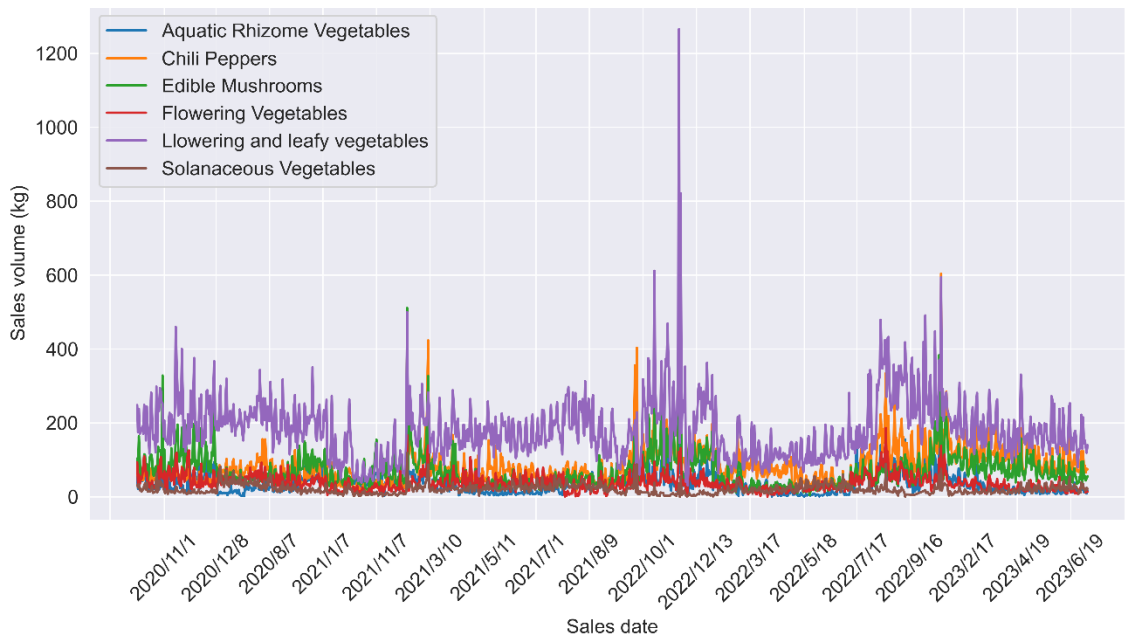


Figure 1. Changes in sales of each category of vegetables as a function of the date of sale

The overall trend of the sales of each category shown in Figure 1 shows that it is more consistent with the bell curve, therefore, for the interrelationship between them, we adopt the Pearson correlation coefficient analysis method to calculate the correlation coefficient between each two categories, so as to determine the interrelationship between these six vegetable categories.

By running the Python program, we can get the Pearson correlation coefficients between different categories.



Figure 2. Heat map of sales correlation between categories - Pearson's coefficient

From the Pearson's correlation coefficient matrix in Figure 2, it can be seen that there are indeed direct correlations between the different categories of vegetables to varying degrees, so we can conclude that there are strong linear relationships between the different categories of most of the vegetables, but the relationship between aquatic rootstocks and eggplants, and between edibles and eggplants and other vegetables is weaker. In this paper, considering the excessive number of vegetable items explored, in order to simplify the model as well as to facilitate the research, only a few vegetables with high sales volume in daily life are selected as examples to derive whether there exists a certain relationship between different items.

We derived the same Pearson correlation coefficient plots between different individual items (higher-selling vegetable items) by running the Python program again.

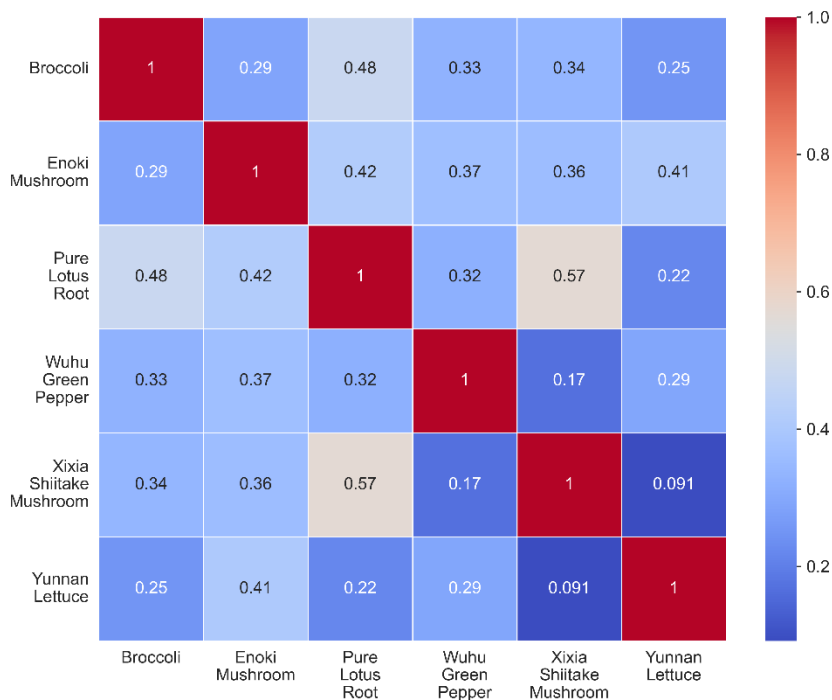


Figure 3. Pearson coefficient thermodynamic diagram between individual products

As can be seen from Figure 3, there is a certain correlation between different commodities, but the overall view of the relationship is weak, linked to the experience of daily life, but also to understand some of the influencing factors (personal preference, vegetable collocation, etc.), due to the main body of research in this paper is the prediction of the sale of commodities, and therefore do not do the specific influencing factors explored here.

3.2. Analysis of replenishment volume and pricing forecast results

For doing replenishment planning by category please analyze the relationship between total sales and cost-plus pricing for each vegetable category, we first visualize the data and make an initial guess about the relationship between sales and cost. Let's take the chili peppers category as an example for a simple analysis.

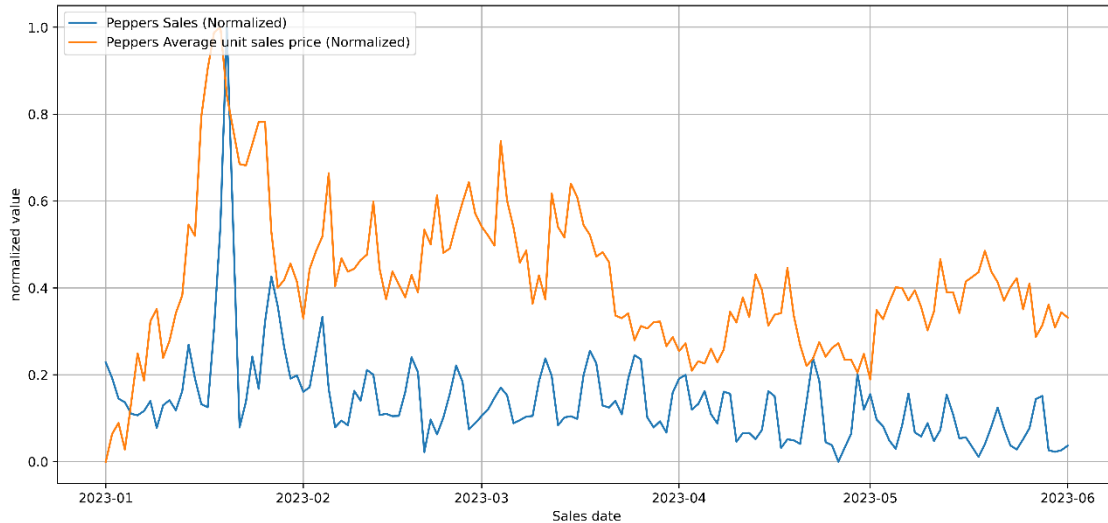


Figure 4. Line Chart of Sales Volume and Unit Price of Chili Peppers

Figure 4 reveals a clear relationship between price and sales, a relationship that suggests a linear relationship between them. Therefore, we can use past sales data to predict sales for the coming week by building a linear regression model. Linear regression is a statistical technique that allows us to predict a dependent variable (in this case sales) based on one or more independent variables (in this case price).

To validate the model, we will use two datasets: a training set and a test set. The training set is used to train the model, i.e., to allow the model to learn the relationship between the independent variable (price) and the dependent variable (sales).

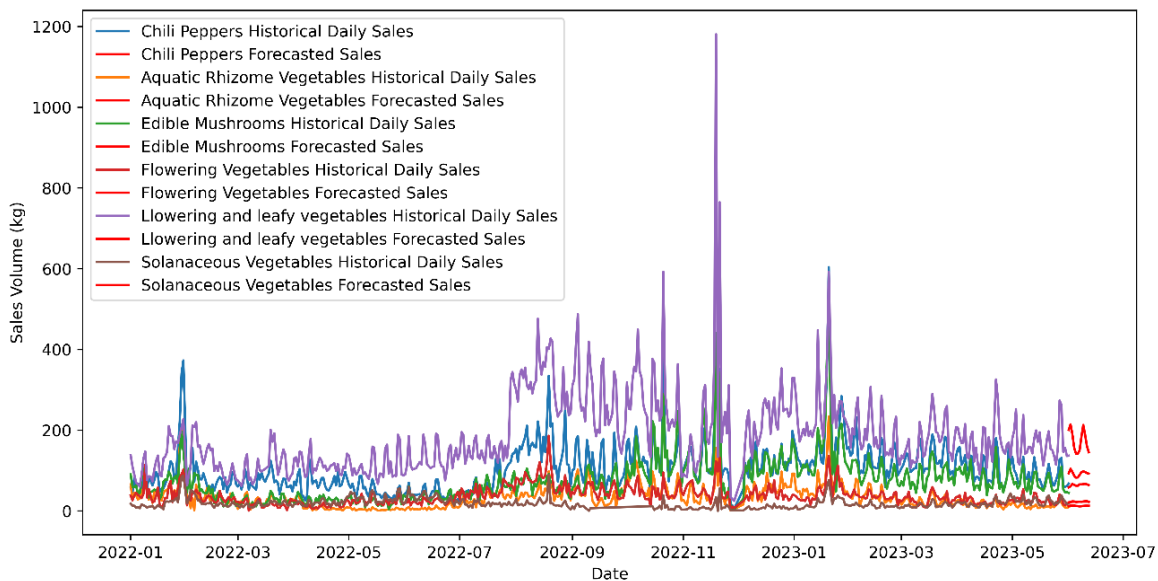


Figure 5. Daily Sales Forecast

Figure 5 shows the image of the model, we let the machine learn the known price data and sales data, explore the regular relationship, and give the prediction of the future date, we compare this prediction with the real value to validate that we found that the model has a high degree of quasi-degree of reliability.

By calculating the mean square error (MSE) between the predicted and actual values, we assessed the model's accuracy performance.

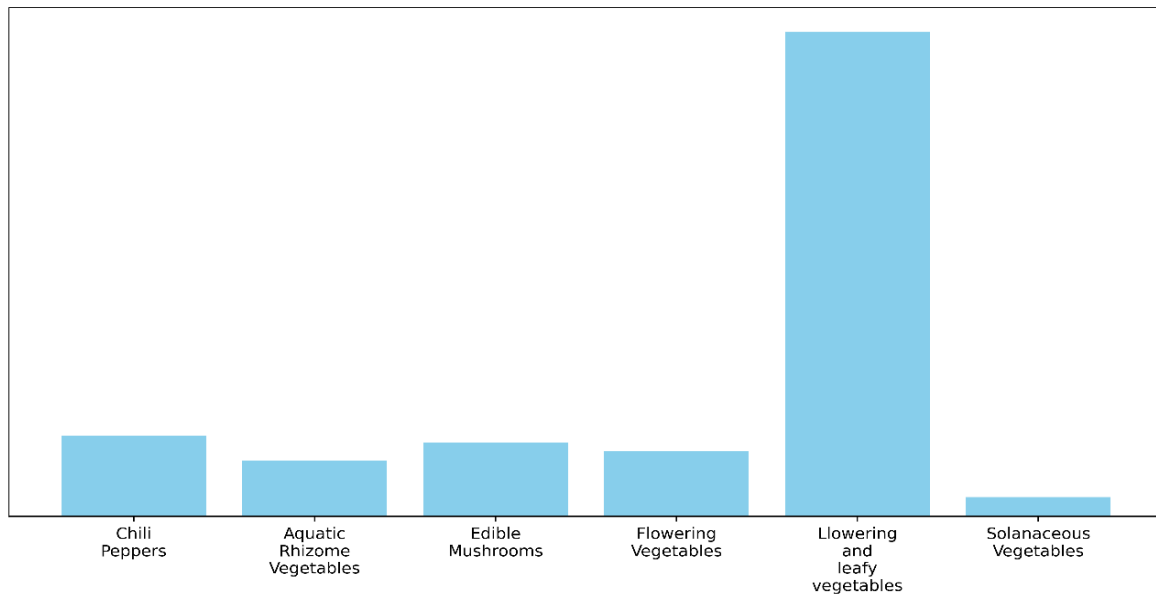


Figure 6. Mean square error for each category.

Figure 6 shows the MSE calculated for all categories, as can be seen from the figure, most of the categories have smaller MSE, which means that our prediction results are more accurate, but inevitably there is one category that has a larger MSE, in which it may be that the model is overly sensitive to outliers, which leads to a reduction in the robustness of the model.

Implement a cost-plus pricing strategy: To maximize profits, we will adjust pricing based on the cost of the item and the forecasted sales volume.

By combining the wholesale price, churn rate, and forecasted sales, we will calculate the total daily replenishment and pricing strategy for each category.

First, we need to know that the total replenishment is based on the expected demand and under different pricing strategies, the demand value will be different, so we need to know the price demand curve and as the curve ends, we can determine a certain price to maximize our revenue. In this paper, we take an example of the chili pepper category and calculate it to get its price demand curve.

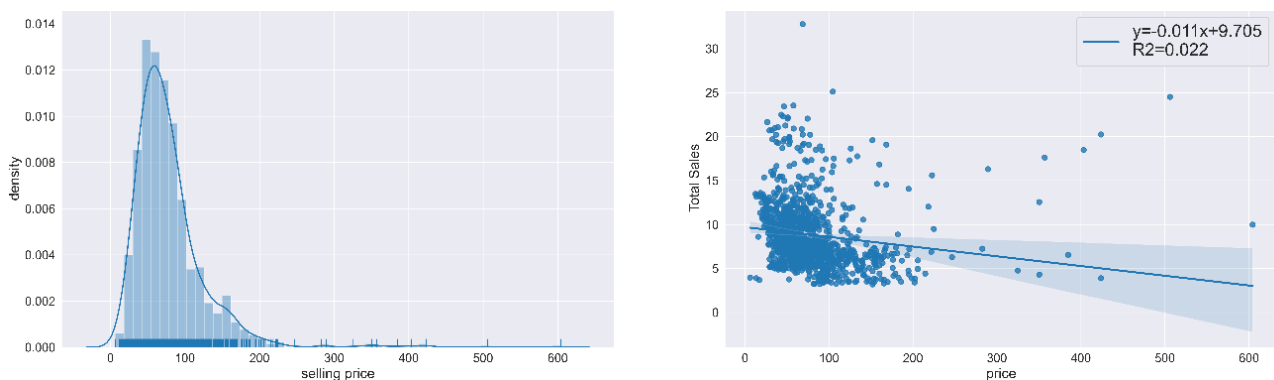


Figure 7. Chili Peppers Price distribution and Chili Peppers Demand price curve

Through Figure 7, we can see the trend of price demand for chili peppers, based on which we get the price demand curves of other classes through the same method, not to be expanded here one by one.

After obtaining the price-demand curves for all categories, we can get the desired forecast results by applying the ARIMA model.

Table.1. Sales unit price and sales volume forecast for six types of vegetable products.

Date	Classification Name	Replenishment volume	Pricing strategy	Classification Name	Replenishment volume	Pricing strategy
7/1	Solanaeous Vegetables	35.16	7.96	Edible Mushrooms	67.80	8.32
7/2		36.55	7.98		74.50	8.26
7/3		29.85	7.91		69.22	8.24
7/4		29.85	7.91		70.33	8.18
7/5		32.43	7.87		71.85	8.16
7/6		32.43	7.87		70.12	8.19
7/7		31.86	7.88		72.09	8.17
7/1	Llowering and leafy vegetables	251.02	6.70	Chil Peppers	76.41	6.55
7/2		244.85	6.80		76.41	6.56
7/3		280.78	6.80		64.58	6.56
7/4		280.78	6.80		64.58	6.56
7/5		280.78	6.80		64.58	6.56
7/6		280.78	6.80		64.58	6.56
7/7		280.78	6.80		64.58	6.56
7/1	Flowering Vegetables	54.60	13.09	Aquatic Rhizome Vegetables	18.94	24.82
7/2		53.79	13.03		18.94	24.65
7/3		35.98	12.82		14.25	24.67
7/4		35.79	12.80		14.25	24.64
7/5		34.29	13.03		14.25	24.67
7/6		35.69	12.81		14.25	24.65
7/7		35.98	12.82		14.25	24.66

From the predicted results (Table 1), there is basically no major fluctuation in the daily sales unit price, while the combination of multi-directional influences will produce different sales volume every day, and therefore also verifies the correctness of this model.

3.3. Maximizing Return Results

Based on our forecasting analysis, and to maximize expected profits, we have selected the 33 items that are expected to be the most profitable. 33 items that are expected to be the most profitable. Below is the recommended replenishment and pricing strategy for the 33 items for the July 1, 2023, forecast.

Table.2. July 1 Product Forecast

Item Name	Sales date	Weighted wholesale price	Replenishment volume	Pricing strategy	Expected income
Xixia Flower Mushroom(1)	2023-07-01	7.80	32.77	14.50	219.58
Yunnan Lettuce	2023-07-01	3.50	54.25	6.96	187.27
Choy Sum	2023-07-01	6.46	34.98	10.69	147.98
Yunnan Lettuce Greens	2023-07-01	3.34	36.84	6.62	120.54
Purple Eggplant(2)	2023-07-01	15.28	6.94	30.52	105.74
Broccoli	2023-07-01	4.06	26.10	8.10	105.23
Spinach	2023-07-01	5.92	28.88	9.48	102.84
Helical Pepper	2023-07-01	4.46	22.50	8.89	99.73
Wuhu Green Pepper(1)	2023-07-01	18.01	5.42	36.02	97.51
Long Line Eggplant	2023-07-01	2.87	26.50	5.71	75.15
Yunnan Lettuce(portion)	2023-07-01	3.45	21.86	6.88	74.93
Spinach(portion)	2023-07-01	3.80	16.80	7.10	55.46
Helical Pepper(portion)	2023-07-01	7.36	7.30	14.65	53.19
Button Mushroom(box)	2023-07-01	7.02	7.58	14.03	53.16
Honghu Lotus Root with Skin	2023-07-01	7.64	10.56	12.46	50.90

Table 2 shows the projected future information for 15 different vegetable products, including unit sales price, sales weight, etc, with the pricing strategy and replenishment plan obtained from the prediction we can get a benefit maximization plan from which we can clearly see the detailed strategy for each product and apply it to the Vegetable and Fruit Supermarket in order to get as much benefit as possible.

4. Conclusions

In this paper, the replenishment and pricing strategy of vegetable commodities is studied by taking the vegetable price data of a supermarket as an example. By analyzing the relevant data of sales flow details and wholesale price of each commodity, combining with efficient inventory management and accurate market demand, the Pearson correlation coefficient is calculated, and the analysis shows that there is a strong linear relationship between most of the vegetables, based on which the relationship between total sales and cost-plus pricing of each type of vegetables is analyzed, and the regression type is established based on the measured data. The established regression analysis shows that there is a regression relationship between total sales and cost-plus pricing, and the established ARIMA model can be used to predict the pricing and sales of each individual product. Finally, based on the replenishment strategy, the pricing strategy and predicted replenishment quantity for each individual product were determined.

The model used in this study provides a feasible framework for predicting sales volume and developing pricing strategies. However, the model is limited by historical sales data in terms of forecasting and may not have sufficient forecasting accuracy when faced with fluctuating market conditions or consumer behavior. Pricing strategies, although based on cost and expected sales volume, do not cover key factors such as competitors' strategies, consumers' willingness to pay, and product substitution and complementary relationships. In addition, factors such as seasonality, promotional activities, and inventory levels are not adequately considered in this paper.

References

- [1] Wang X. Joint Decision Making of Replenishment, Pricing, and Fresh Keeping Input in Fruit and Vegetable Cold Chain: Based on Markov Process[J]. *Mobile Information Systems*, 2022, 2022.
- [2] Kaya O, Bayer H. Pricing and lot-sizing decisions for perishable products when demand changes by freshness[J]. *Journal of Industrial and Management Optimization*, 2020, 17(6): 3113-3129.
- [3] Chen W, Liu H, Xu D. Dynamic pricing strategies for perishable product in a competitive multi-agent retailers' market[J]. *Journal of Artificial Societies and Social Simulation*, 2018, 21(2).
- [4] Fattah J, Ezzine L, Aman Z, et al. Forecasting of demand using ARIMA model[J]. *International Journal of Engineering Business Management*, 2018, 10: 1847979018808673.
- [5] KumarMahto A, Biswas R, Alam M A. Short term forecasting of agriculture commodity price by using ARIMA: based on Indian market[C]//*Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part I 3*. Springer Singapore, 2019: 452-461.
- [6] Matyjaszek M, Fernández P R, Krzemień A, et al. Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory[J]. *Resources Policy*, 2019, 61: 283-292.
- [7] Azadi Z, Eksioğlu S D, Eksioğlu B, et al. Stochastic optimization models for joint pricing and inventory replenishment of perishable products[J]. *Computers & industrial engineering*, 2019, 127: 625-642.
- [8] Du W, Zuo Y. How to Set Reasonable Prices for Vegetable Products? A Comprehensive Model Study Based on Multiple Algorithms[J]. *Highlights in Science, Engineering and Technology*, 2023, 70: 456-462.
- [9] Galda I. Literature Review-Dynamic Pricing in Food Supply Chains[J]. 2022.
- [10] Khan M A A, Shaikh A A, Panda G C, et al. Inventory system with expiration date: Pricing and replenishment decisions[J]. *Computers & Industrial Engineering*, 2019, 132: 232-247.