Automatic Pricing and Replenishment Decisions for Vegetables Based on Time Series Modeling

Yi Zhang †, Jingjing Fan †, Yimeng Zheng †

School of Mathematics and Statistics, Henan University, Kaifeng, China

† These authors also contributed equally to this work

Abstract. The market is highly competitive. Reasonable pricing and timely replenishment decisions can help enterprises to grasp the market demand and competitive situation, maximize profits. Through data-driven decision-making and technical support, supply chain efficiency and stability can be optimized, customer satisfaction can be improved, and personalized products and services can be provided. This paper uses statistics to analyze the distribution law of sales volume among categories and single products, and solves the problem of finding the correlation between different categories and different single products of vegetable commodities and the distribution law of sales volume. According to the distribution law and relationship of the sales volume of each category and single product of vegetables, the replenishment plan is made by taking the category as a unit, the relationship between the total sales volume of each vegetable category and the cost-plus pricing is analyzed, and the daily replenishment volume and pricing strategy of each vegetable category in the coming week are given, and finally the optimal daily replenishment and pricing strategy is determined, which provides a reference for supermarket operation and supply chain management.

Keywords: Pearson Correlation Coefficient, Grey Relational Analysis, Time Series Prediction, Optimization Model.

1. Introduction

In fresh food supermarkets, the shelf life of general vegetable products is relatively short, and the appearance of products becomes worse with the increase of sales time. If most varieties are not sold on the same day, they cannot be resold until the next day. Therefore, the problem of food loss has gradually become an important issue in the super-green development of fresh food merchants.

The automatic pricing and replenishment decision of vegetables is a common problem in retail and supply chain management. Because the price and demand of vegetables may be affected by seasonal fluctuations, weather changes, market supply and demand and other factors, how to reasonably price and make replenishment decisions has become an important issue for enterprises and retailers.

Based on the sales flow details and wholesale prices of six vegetable categories in a supermarket from July 1, 2020 to June 30, 2023, combined with the recent loss rate of vegetable commodities, this paper uses Pearson correlation coefficient analysis, grey correlation analysis, time series prediction and knapsack model to find out the sales volume of vegetable categories and single products. And obtain that total quantity and the pricing strategy of the sup daily replenishment of the supplier. At the same time, considering the limited sales space of vegetable commodities in supermarkets, this paper obtains the optimal strategy of daily replenishment and pricing of supermarkets on the premise of meeting the market demand for various types of vegetable commodities, which maximizes the profits of supermarkets and provides a reference for supply chain management of enterprises and retailers [1, 2].

2. Data Processing and Model Building

2.1. Pearson analysis

Correlation analysis was used to observe the correlation between the data. Pearson correlation coefficient is a statistical index used to measure the degree of linear correlation between two variables X and Y [3]. Its value range is between -1 and 1. When the correlation coefficient is positive, the two variables are positively correlated; When the correlation coefficient is negative, the two variables are negatively correlated.

2.2. Grey correlation analysis model

A grey correlation analysis model can be constructed between vegetable categories and single product sales volume [4]. The grey correlation coefficient is observed by taking six vegetable categories as the parent sequence and single product as the sub-sequence. The establishment of grey correlation analysis model:

Mother sequence

$$m_0 = \left(m_0(1), m_0(2) \dots m_0(6)\right)^T \tag{1}$$

Subsequence

$$\begin{gathered} n_1 = \left(n_1(1), n_1(2) \dots n_1(6)\right)^T \\ n_2 = \left(n_2(1), n_2(2) \dots n_2(6)\right)^T \\ \cdot \\ \cdot \\ \cdot \\ n_{251} = \left(n_{251}(1), n_{251}(2) \dots n_{251}(6)\right)^T \end{gathered} \tag{2}$$

Assumptions a Is the minimum difference between the two poles, β for the maximum difference between the two poles, there is

$$a = min(i)\, min(k) |m_0(k) - n_i(k)| \tag{3}$$

$$\beta = max(i)\, max(k) |m_0(k) - n_i(k)| \tag{4}$$

The formula for the correlation coefficient G (Gamma value)

$$G\left(m_0(k), n_i(k)\right) = \frac{a+\beta^*\rho}{|m_0(k), n_i(k)| + \beta^*\rho} \, (\rho T\ aken\ as\ 0.5) \tag{5}$$

The grey correlation degree is assumed $G\ m\ 0\ k,\ n\ i\ k$, then there is

$$G\left(m_0(k), n_i(k)\right) = \frac{1}{6} \sum_{k=1}^{6} G\left(m_0(k), n_i(k)\right) \tag{6}$$

The arithmetic mean root of each column is the degree of correlation between each subsequence and the parent sequence.

2.3. Prediction model

The time series analysis of the daily sales of vegetables over the last three years can be used to predict the daily demand for the week ahead. 70% of the data was used for training and 30% for testing [5]. The data was then substituted into a time series analysis (long and short period neural network prediction) to derive the daily vegetable replenishment for the following week. Time series analysis was performed using Matlab to derive the daily sales volume for the coming week.

2.4. Knapsack model

Knapsack model is an optimization model, which is often used to solve resource allocation problems. Its basic concept comes from how to choose some items to put into the backpack in order to maximize the benefit of a certain goal under the limited capacity of the backpack, which can be solved by greedy algorithm and dynamic programming. This topic knapsack model establishment:

The unit price, the sales volume, the total replenishment quantity are the model main factors, has s goods and a capacity is R knapsack, each goods has the weight $g\ i$ and value $q\ i$ Two attributes, requiring the selection of a number of items into the backpack, so that the total value of the backpack is the largest, and does not exceed the capacity of the backpack.

3. Results

3.1. Distribution rule of sales volume of each category

The sales data of a supermarket is recorded from July 1, 2020 to June 30, 2023, a total of 36 months. It includes six categories of flowers and leaves, cauliflower, aquatic roots, eggplants, peppers and edible fungi, and a total of 251 single products. Examining the relationship between the sales volume of each category and summing the sales volume of all single products included in different categories, we can calculate the total sales proportion of each category, in which the sales of flowers and leaves account for the majority, while the sales of eggplant, cauliflower and aquatic roots account for a relatively small proportion (Fig. 1).

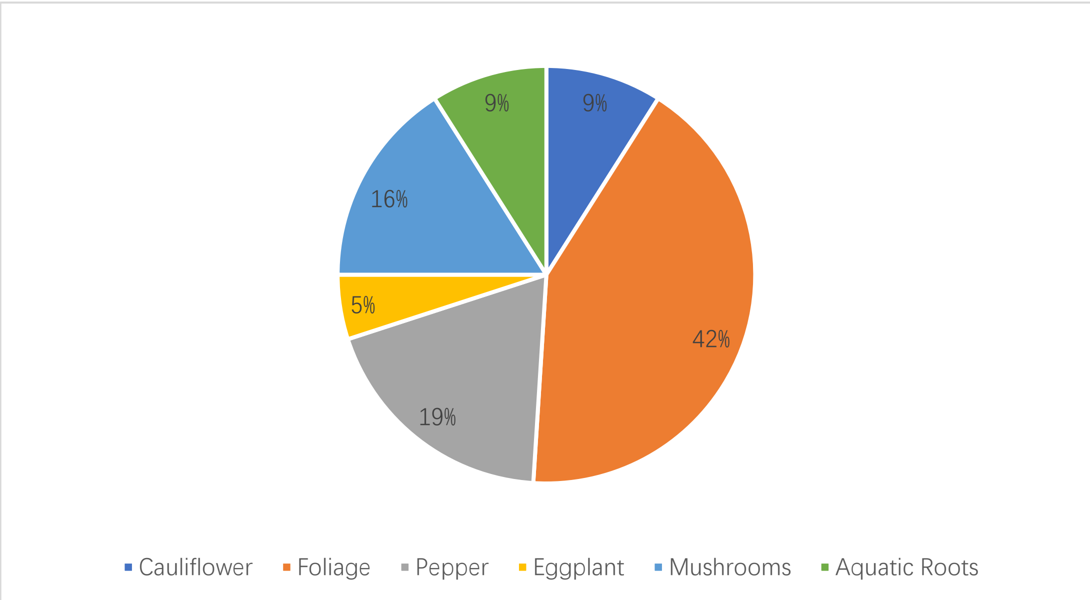

Figure 1. Total sales volume of each category

Considering that the sales volume is a typical continuous variable that changes over time, we select a month as the unit, calculate the total sales volume of the same category in three years by month, and draw a picture of the fluctuation of each category's sales volume over time (Fig. 2). From the picture, we find that the distribution of each category's vegetable sales volume is roughly normal distribution, which shows that the sales volume in July, August and September is the highest in the whole year. Sales at other times were lower than those in July, August and September. In addition, we also found that the sales of flower and leaf were significantly more than other categories, the sales of eggplant and cauliflower were the least, the sales of flower and leaf, pepper and edible fungi were on the rise, and the sales were seasonal.

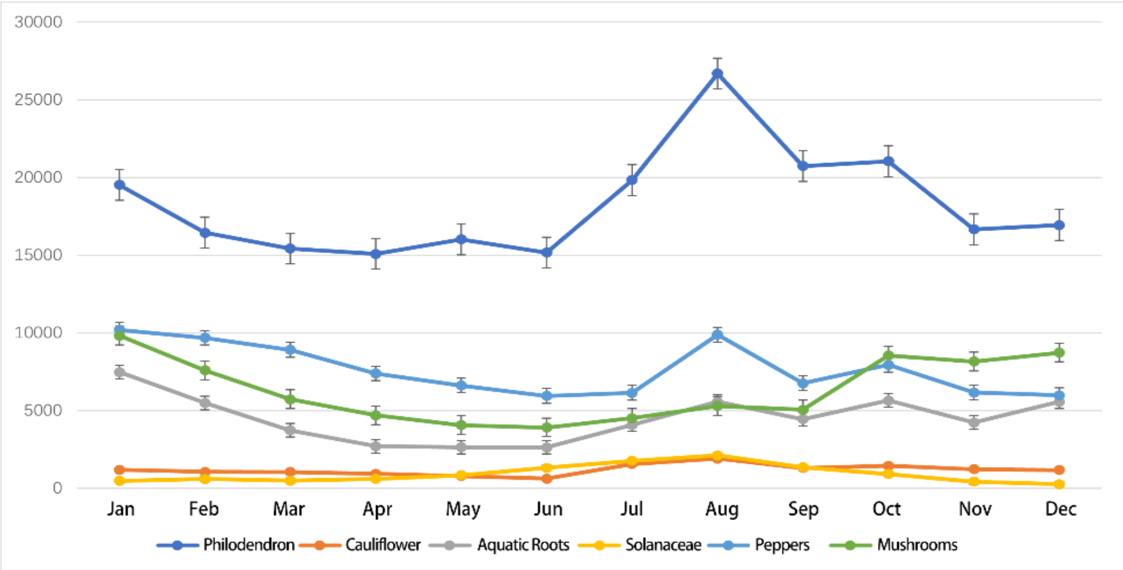

Figure 2. Monthly sales volume of each category

The Pearson correlation coefficient between each category data was then calculated (Fig. 3). It can be seen from the Fig. that there is a strong correlation between flowers and leaves and peppers, and between water roots and stems and edible fungi, but there is a weak correlation between water roots and stems and eggplants. These six categories can be initially divided into two parts, of which the five categories of mosaic, cauliflower, pepper, edible fungi and aquatic plants are the first category, which have a moderate degree of correlation, while the correlation between eggplant and the former is weak.

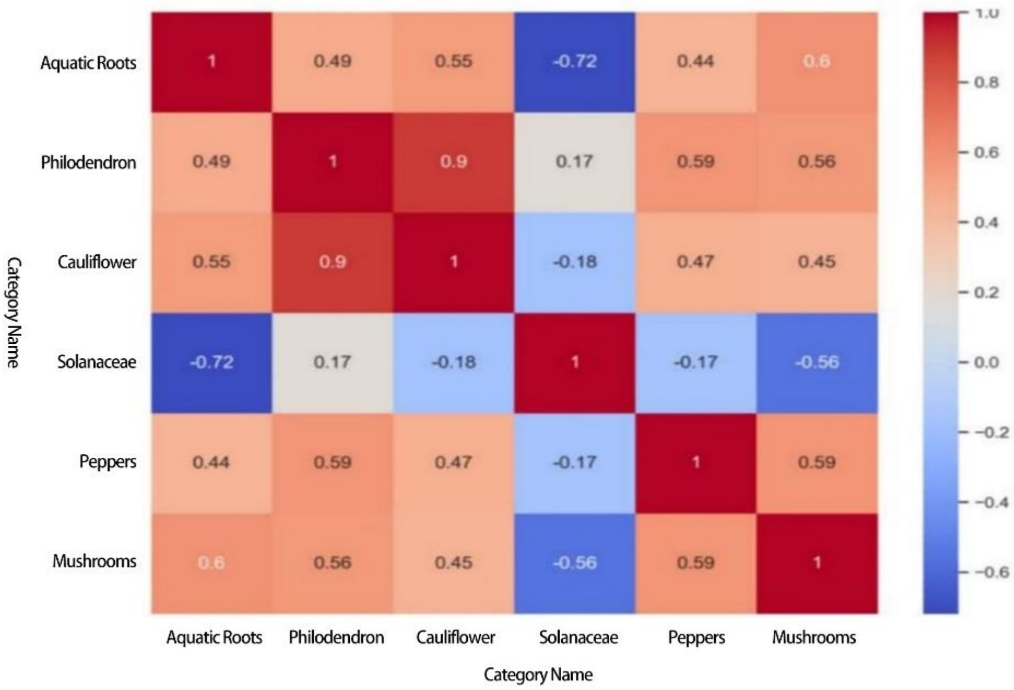

Figure 3. Matrix Thermodynamic Diagram between Categories

Analyze the change rule of the sales of each single product. Considering that there are more data in the single product category, which cannot be analyzed like the six major categories of vegetables, we can perform descriptive statistical analysis (maximum value, minimum value, etc.), and then calculate the total sales volume of 251 single products, of which the top 20 vegetable products are shown in Fig. 4. It was found that Wuhu Green Pepper (1) had the highest sales volume, followed by Broccoli and Lotus Root (1), while Screw Pepper, Shanghai Green and Bamboo Leaf Lettuce had lower sales volumes.

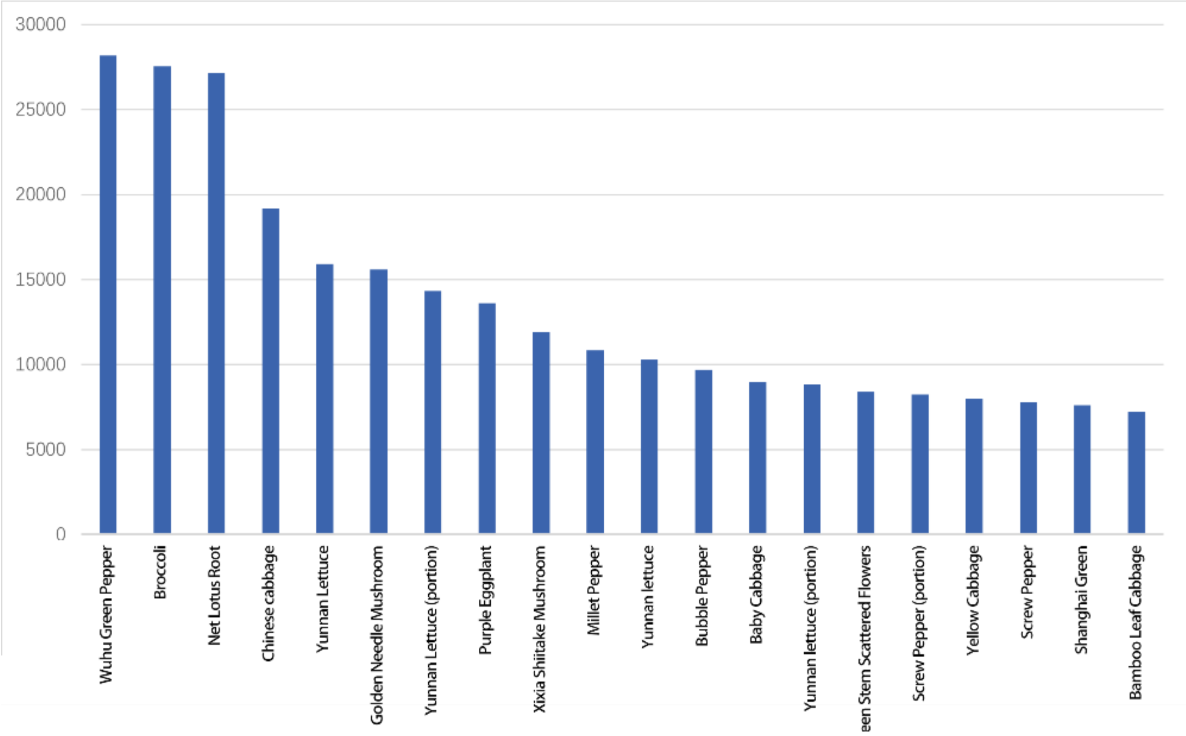

Figure 4. Total sales volume of top 20 single products

A grey correlation analysis model can be constructed between vegetable categories and single product sales, and the grey correlation coefficient can be observed by taking six vegetable categories as the parent sequence and single product as the sub-sequence (Fig. 5).

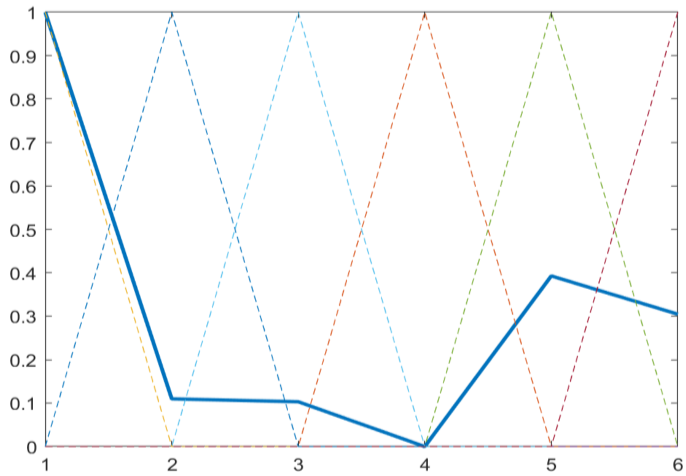

Figure 5. Degree of grey correlation between each category and single product

3.2. Total replenishment quantity and pricing strategy for the next week

In the case of the maximum total revenue, to solve the total daily replenishment and pricing strategy of each vegetable category in the coming week, we can first solve the mosaic category, and so on.

Set the decision variable as $x_{\text{Flowers and leaves}}$ (i =1, 2,… 7), indicating that th i Sales price of smallpox leaves.

Cost-plus pricing is a common pricing strategy because it is based on the cost of the product and determines the final price by adding a suitable markup. The following are the general steps for using cost-plus pricing:

$$z_{i\ Flowers\ and\ leaves} = C * (1 + h) \tag{7}$$

Among $z_{\text{Flowers and leaves}}$ indicates pricing, h Represents the addition rate, C Represents the purchase price (the purchase price of the category of vegetables is represented by the average purchase price of all single products contained in the category of vegetables). Therefore, the daily profit can be expressed as

$$w_{i\ Flowers\ and\ leaves} = y_{i\ Flowers\ and\ leaves} * x_{i\ Flowers\ and\ leaves}(1-l) - C * y_{i\ Flowers\ and\ leaves} \tag{8}$$

Among l Is the loss rate. Therefore, the total profit of the supermarket in a week can be expressed as

$$W = \sum_{i=1}^{7}(w_{iFlowers\ and\ leaves} + w_{iCauliflower} + w_{iChili} + w_{iEggplant} + w_{iEdible\ fungi} + w_{iAquatic\ rhizome}) \tag{9}$$

Wherein, the addition ratio h and $y_{\text{iFlowers and leaves}}$ as an independent variable, the optimization class is solved.

Constraints:

$$\begin{cases} 0 \le l_{Flowers\ and\ leaves} \le 1 \\ 0 \le h_{iFlowers\ and\ leaves} \le 1 \\ y_{iFlowers\ and\ leaves} = -0.0004{z_{iFlowers\ and\ leaves}}^2 + 0.0465z_{iFlowers\ and\ leaves} + 8.9781 \end{cases} \tag{10}$$

Finally, the total daily replenishment volume and pricing strategy of each category in the coming week are as follows (Table 1, Table 2):

Table 1. Cost-plus pricing by category

Time	Flowers and leaves	Cauliflower	Water-rooted stems	Eggplant	Capsicum	Edible fungi
Mon	6.66564	10.60229	23.53495	5.913563	7.084585	8.322393
Tue	6.159046	10.85461	28.15568	5.934582	7.593853	8.296983
Tue	6.102741	10.56074	21.55044	6.081179	7.348434	8.631897
Thu	6.081502	10.48394	23.26036	5.97666	7.175581	8.43569
Fri	5.331541	10.67992	24.08393	6.171796	7.111595	8.021036
Sat	5.397235	10.03241	22.00264	6.242599	7.432774	8.542184
Sun	4.934426	10.2497	21.58263	6.103283	7.846836	8.43885

Table 2. Total Daily Replenishment of Each Category

Time	Flowers and leaves	Cauliflower	Water-rooted stems	Eggplant	Capsicum	Edible fungi
Mon	86.87206	55.69	165.8426	40.38234	80.23569	42.85476
Tue	88.61506	55.686	190.0412	40.38807	10.37197	42.21051
Tue	90.36806	55.682	155.4508	40.37532	79.00897	41.463
Thu	92.12606	55.678	164.4074	40.4021	78.46099	40.90626
Fri	93.88406	55.674	168.7129	40.38784	77.94229	40.28431
Sat	95.63806	55.671	157.8183	40.32629	77.43652	39.59775
Sun	97.382	55.668	155.6184	40.42104	76.883	38.95198

3.3. Optimal Daily Replenishment and Pricing Policy

In order to obtain the optimal daily replenishment quantity and pricing strategy, it is necessary to add the restriction requirements of the total number of saleable single products and the minimum display quantity of single products (the total number of saleable single products is 27-33, and the minimum display quantity of each single product is 2.5 kg), and the research object becomes the saleable vegetable single products on June 24-30, 2023. Then the time series prediction model is used to measure the daily replenishment quantity of six major vegetable categories on July 1. Under the condition of meeting the market demand for each category of vegetable commodities, the total income is maximized, and then the backpack model is established to select as many single products with higher average daily income as possible, and finally the daily replenishment quantity and pricing strategy of single products on July 1 are obtained.

Average daily income of single product w, is solved as follows,

$$w = \left(P - C_{Average} - l * P\right) y_{Average} \tag{11}$$

Among P for the average price of the predicted single product on July 1, C_{Average} is the average purchase price of a single product on June 24-30. y_{Forecast} is the total sales volume of the forecasted single product on July 1.

Similar to the method in Section 3.2, the data is sorted and filtered through excel, and then substituted into the time series analysis to obtain the daily replenishment quantity and average pricing of 36 types of single products on July 1, and then the average daily revenue is calculated, as shown in the Fig. below (Fig. 6).

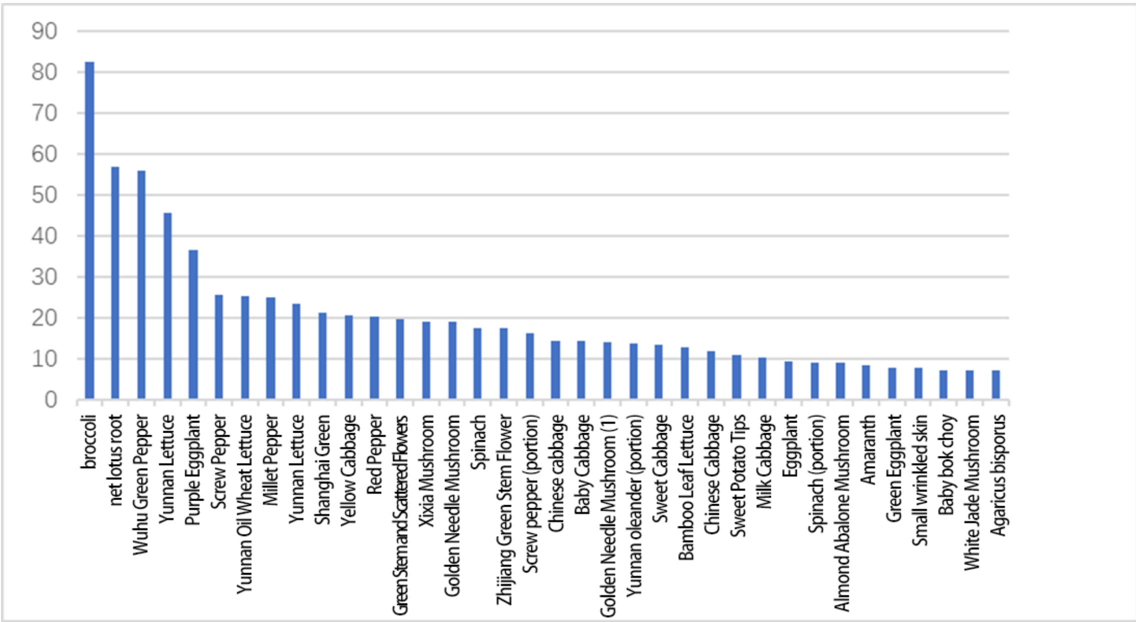

Figure 6. Average daily income of 36 vegetable products

For the optimization of daily replenishment and pricing strategy, we can establish a knapsack model. Consider taking the category as the unit and taking its total number as the knapsack, the single products in each category are calculated separately and then filled into 6 knapsacks. In this case, the optimal solution can be obtained by proving that the greedy algorithm can obtain better performance. The results are shown in the following table. There are 33 categories of single products in total, and the total profit is 716.4345 yuan.

4. Conclusion

Based on the sales flow and wholesale price data of a supermarket, this paper finds the relationship between different categories and different single products of vegetable commodities and the distribution law of sales volume, and gives the total daily replenishment volume and pricing strategy of vegetable commodities in the supermarket in the coming week, and optimizes the strategy through mathematical models. To maximize the benefits of supermarkets and provide reference for enterprise supply chain management.

References

[1] ZHOU Shanshan, MA Haibin, MEN Guoming. Carbon emission evaluation of public building projects based on gray correlation analysis [J]. Journal of North China Institute of Science and Technology, 2023, 20 (04): 89 - 95. DOI: 10.19956/j. cnki. ncist. 2023. 04. 013.

[2] Y. X. Gao, S. N. Zhang, Y. X. Zhang et al. A material production scheduling model based on time series forecasting [J]. Mathematical Modeling and its Applications, 2023, 12 (02): 109 - 114. DOI: 10. 19943/j.2095 - 3070.jmmia. 2023. 02. 11.

[3] Cheng Juanjuan. Empirical research on the relationship between research and teaching in colleges and universities--analysis based on Pearson's correlation coefficient [J]. Science and Technology in Chinese Colleges and Universities, 2022 (10): 46 - 52. DOI: 10.16209/j. cnki. cust. 2022. 10. 016.

[4] M. Zeng. Research on Dynamic Pricing Strategies of Fresh Community Supermarket Based on Time Situation A [D]. Southwest University of Finance and Economics, 2021. DOI: 10.27412/d. cnki. gxncu. 2021. 002858.

[5] Yao Chun, Sun Jianhong. Research on ordering and pricing of fresh produce in fresh supermarket [J]. SAR economy, 2016 (02): 103 - 104.