Automated Pricing and Replenishment Decisions for Vegetable Items

Mengjia Hou[1,*]

[1]*College of Economics and Management, Qinghai Minzu University, Xining, Qinghai, China*
*Corresponding author: 2835332215@qq.com

Abstract: In the fresh food superstore, due to the variety of vegetable commodities, short shelf life, easy to deteriorate, and different origins and transportation methods, and replenishment time is set in the early morning, so the superstore needs to make replenishment decisions in advance. This paper collects the sales data of a supermarket over the past three years, and carries out relevant research on automatic pricing and replenishment decision-making for vegetable commodities. In this paper, we study the relevant distribution law of vegetable varieties and single products, and analyze which categories have a strong correlation between categories. Secondly, an optimization model is constructed to solve the total daily replenishment and the corresponding pricing to obtain the pricing and replenishment strategy of the superstore.

1. Introduction

In this paper, first of all, after drawing bar charts and line graphs between different categories and different single products, the relevant distribution law of each variety and single product of vegetables is obtained. For the super market vegetable interrelationships, when dealing with them according to categories, we use Spearman correlation coefficient analysis [1] to get the heat map, so as to conclude which categories have stronger correlation between categories, and when dealing with them according to individual items, we first use K-means clustering to cluster the individual items with stronger correlation into one category, and then we screen out the most representative several vegetables in each category, and then we get the distribution law table to get the distribution law of these vegetables. distribution law. Secondly, the data set obtained by data preprocessing is used to derive the expression of the functional relationship between sales volume and cost-plus pricing using the fitting toolbox. Taking the total daily replenishment as the decision variable and the revenue of the superstore as the objective function, an optimization model is constructed to solve for the total daily replenishment and the corresponding pricing strategy.

2. Study on the correlation and distribution pattern of vegetables in supermarkets

Firstly, the sales distribution pattern of different categories of vegetables is considered, and two different analyses are carried out according to the time division of month and hour, respectively.

When analyzing the correlation between different categories, this paper adopts Spearman correlation coefficient analysis (Spearman), the correlation coefficient of the two two data is calculated, and the heat map is obtained to describe the correlation. Then consider the sales distribution pattern of different single products of vegetables, because there are more single product categories, this paper adopts the idea of dimensionality reduction, and carries out K-means mean clustering for them [2]. The single product will be classified according to spring, summer, autumn, winter to study, using the idea of cluster analysis to determine the correlation between them, so as to draw conclusions. First of all, this paper extracts and sorts the collected data, selects the data in which the total sales volume is greater than 1000 to be analyzed, and then through analyzing the distance between each index and the center of the clustering, selects the most representative single product from these 4 categories, so that a total of 4x4=16 single products will be screened out, and finally selects some more iconic single products from within these single products to carry out the principal component analysis, so as to draw the distribution of the Regularity judgment.

2.1 Relevant formulas

Spearman's correlation coefficient definition equation:

$$\rho = 1 - \frac{6 \sum_i di^2}{n(n^2 - 1)}$$

(1)

K-means mean clustering: those that can be clustered into one category indicate a strong correlation.

$$E = \sum_{j=1}^{k} \sum_{x_i \in w} \left\| x_i - m_j \right\|^2$$

(2)

2.2 Calculus of variations

In this paper, we first collect the sales data of the business over the last three years to organize. In this paper, the data is divided into category + month, category + time period, single product + month, single product + time period of these four pieces to organize.

For category + month: this paper will be the three years in the same month of a category of sales volume to add up, and ultimately get the total sales of each single product for 12 months, made of a 6 * 12 data table, according to the distribution of data in this paper to get the content of the data distribution of Figure 1, while the correlation analysis of the data of each category, to get the heat map in Figure 3 (a) as follows [3].

For category + time period: this paper will be the three years from 9 to 21:00 a total of 13 hours of data divided into different categories to do the cumulative sum, and ultimately get a 6 * 13 data table, according to the distribution of data in this paper to get the content of Figure 2, at the same time, the correlation analysis of the data of each category, to get the heat map in Figure 3 (b) below.

For single product + month: with category + month is basically the same, according to different single product to do the cumulative sum, and finally get a data table.

For single product + time period: basically the same as category + time period, according to different single product to do the cumulative sum, and finally get a data table.
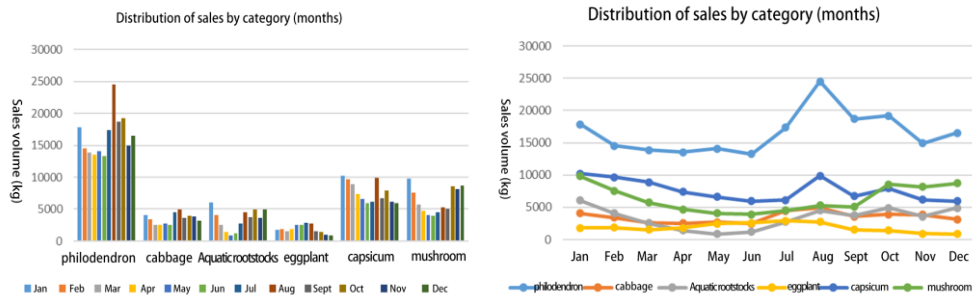
Figure 1: Monthly Sales Distribution by Category
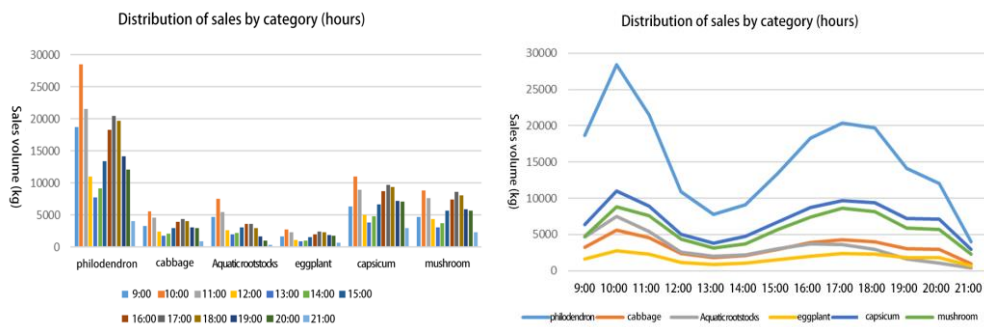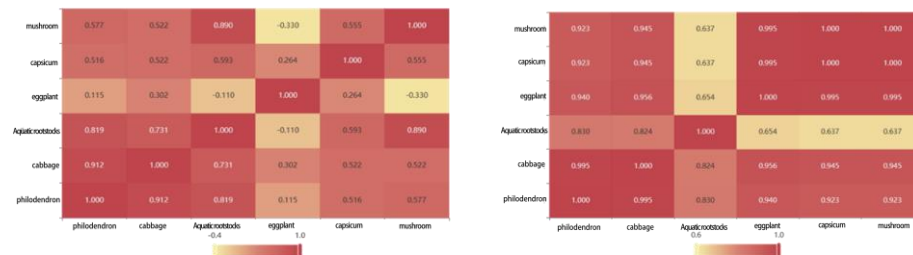

Figure 2: Category Time Period Sales Distribution

(a) Heat map of category month relevance   (b) Heat map of category time period correlation

Figure 3: Heat Map of Correlation of Vegetable Category

Due to the vegetable single species category is too much, the amount of data is too large, it is not easy to single to statistics, at the same time its relationship with each other is more difficult to analyze, if then use the analysis of the relationship between the categories - Spearman correlation coefficient method, first of all, whether there is a statistically significant relationship between the XY ($P < 0.05$) test, and then analyze the correlation coefficient for the positive and negative direction and the degree of correlation, then this paper may have to consider $251 * 251 = 63001$ groups of data relationships, which is too complex and does not have a prominent representative. Therefore, this paper chooses to use cluster analysis to analyze these data. In order to be more accurate to provide some reference value to the superstore, this paper chooses to use four seasons to segment the data for clustering, so this paper will be divided into four data: single product sales data in March to May, single product sales data in June to August, single product sales data in September to November, single product sales data in December, January and February. Then due to the large number of 251 data, and some single products have almost no sales within the selected months, so this paper sums up the sales data of the three months, and finally selects the data with sales volume greater than 100kg for K-means clustering analysis. The related graphs are finally obtained as shown in Figure 4 and Figure 5.
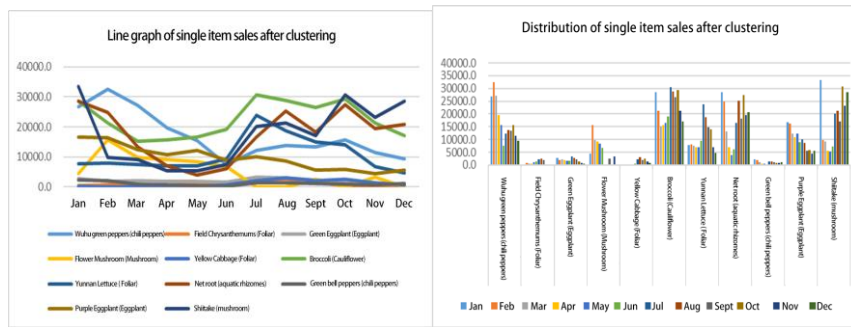
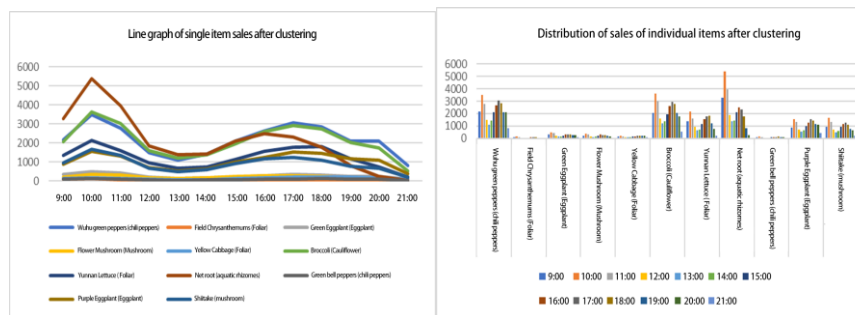Figure 4: Line Chart and Bar Chart of Single Product Sales (Months)

Figure 5: Line and Bar Charts of Sales of Individual Items (Time Period)

2.3 Correlations and Distribution Patterns of Vegetable Categories

From the heat map, we can learn that the correlation coefficient between the three categories of flowers and leaves, cauliflower and aquatic roots and tubers is relatively large and the correlation is relatively strong, which means that when the sales of flowers and leaves increase, the sales of cauliflower and aquatic roots and tubers will also increase. The correlation between the eggplant category and the other categories is relatively low, which means that he may not be affected by the sales of the other categories and is an independent sales pattern.

At the same time every month the sales of the flower and leaf category is located in the first place, and each time period flower and leaf volume sales are also high in the bar chart, all of them are located in the top. This can be from a certain point of view that the foliage category is a category that people often buy, so supermarkets can choose to buy more foliage products than other categories in the purchase of goods. From July to September, sales of leaves, peppers, aquatic roots and tubers peak during the year. This shows that in the summer, people tend to buy this dish. It can also be said that July to September is the harvest time for some dishes in these categories. In the period from July to September, the sales line of edible fungi did not show a concave shape, indicating that this time may not be the harvest period of fungal plants. So, when the supermarket purchases vegetables in the summer, it can pay more attention to the dishes of leaves, peppers, aquatic roots and tubers in this area. Because at this time it may reduce the wholesale amount of such dishes, but also cater to the choice of the public.

It can also be observed that for almost all categories, the 9am-12am period is the time when they sell the most, followed by the 15am-19pm period. We believe that this data is reasonable, as these two time periods correspond to the pre-lunch and pre-dinner time periods, where purchases are made in time for the preparation of lunch and dinner, and where people want to eat fresher vegetables and avoid stocking up on them to avoid mold and mildew. 12:00 to 15:00 is the time for lunch and lunch breaks, and the decrease in the number of sales of vegetables at this time is also justified. This also proves the validity of our model. At the same time, this set of data can also be adjusted to some of

the supermarket's working hours or stocking program, you can choose to sell in the two small peaks of the front goods, to avoid hoarding vegetables lead to its mold makes the final profit is low.

2.4 Correlations and Distribution Patterns of Individual Vegetable Items

Based on the results obtained from clustering, this paper can draw the following conclusions:

In spring, people may choose to buy purple eggplant while buying broccoli, and may choose to buy single items such as spinach and enoki mushrooms (box) while buying Yunnan lettuce (portion);

In summer, people may choose to buy single items such as choy sum and baby bok choy while buying yellow cabbage, and may choose to buy bubble peppers while buying shiitake mushrooms in Xixia and green peppers in Wuhu;

In the fall, people may choose to buy single items such as choy sum, small bok choy, screw peppers, and red peppers while purchasing yellow cabbage, and may choose to buy single items such as Wuhu green peppers, bubble peppers, and Yunnan lettuce while purchasing Xixia shiitake mushrooms.

In winter, while purchasing yellow cabbage, people may buy single items such as cordyceps, small bok choy, and choy sum, and while purchasing Yunnan lettuce, they may choose to buy single items such as baby lettuce, bubble peppers, and Honghu lotus root.

Supermarkets can estimate the sales volume of another single product based on the sales volume of a single product, and then get the optimal stocking program.

Because of too much data, so this paper chooses the representative 11 single product indicators screened out in the previous section, and analyzes their distribution patterns. This paper analyzes the relationship between the sales volume of single product and month and time period. As can be seen from Figure 6 (a), each item has its peak season and low season. For example, Wuhu green pepper, the first half of the single product sales are high, the second half of the sales are low, and the net lotus root is roughly the opposite. Sales during July-October are extremely high, which is consistent with common sense, and the net lotus root production is the largest in this period. The sales of broccoli fluctuated somewhat throughout the year, but its sales volume was consistently high, indicating that broccoli basically belongs to the range of home-cooked food, and supermarkets can choose to purchase this type of single item year-round; in contrast, the sales volume of white mushrooms was lower in all seasons of the year, suggesting that fewer people are willing to purchase this type of single item, and supermarkets can reduce the purchase of this type of single product accordingly. The sales volume of yellow cabbage starts to rise from June, reaches its peak in August, and then falls back to its lowest point around November. Supermarkets can choose to increase the purchasing volume of yellow cabbage during the period from June to November.

See Figure 6(b) shows that the purchasing volume of these single items peaks around 10 a.m. and reaches another small peak around 17 p.m. The distribution law of these single items over time is similar to the distribution law of categories over time, which can also verify the correctness of the model constructed in this paper from the other side.
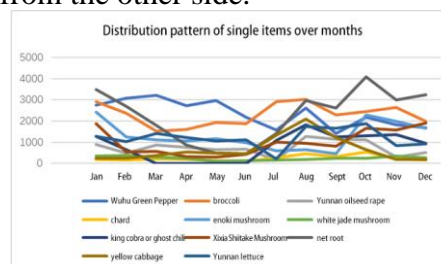
(a) The pattern of change of the single product with the month
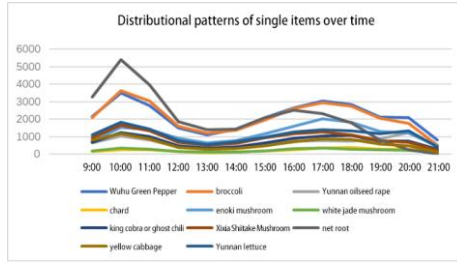
(b) The pattern of change of single products over time

Figure 6: Vegetable Single Product Distribution Pattern

3. A Model of the Relationship Between Total Daily Replenishment, Superstore Revenue and Pricing Strategy

In this paper, we first analyze the relationship between total sales volume and cost-plus pricing for the vegetable category to establish a suitable function model describing the relationship between total sales volume and cost-plus pricing for the category [4]. It is also necessary to establish a relevant model for calculating the maximum value of the profit obtained and the value of the replenishment volume and pricing at this maximum value. However, the determination of the replenishment quantity must be based on the prediction of the inventory quantity, so this paper should collect relevant data to predict the future partial value.

First calculate the actual inventory:

$$ai = \sum_i \frac{sj}{1 - \partial j} \tag{3}$$

The actual inventory is then used to calculate the average wastage rate:

$$\partial i = 1 - \frac{ai}{Si} \tag{4}$$

Calculate the total sales of a category using summation:

$$m_i = \sum_j sjpj \tag{5}$$

Calculate the total cost using the wholesale price of each individual item * the actual inventory of the individual items:

$$Ci = \sum_j wjaj \tag{6}$$

Use the total cost/actual inventory to figure out the cost per item:

$$ci = \frac{Ci}{ai} \tag{7}$$

Calculate the profit, if the inventory is too high on the day, it will result in having unsellable products, so subtract $(ai + n - Si)ci$, discounting is not considered here:

$$Y = (x - ci)Si - (ai + n - Si)ci$$

(8)

Here ci is the cost per unit price derived from the data collection and the current day's forecast using time series analysis.

It is now sufficient to find the value of x that maximizes Y. Once x is obtained, Si can be obtained by $Si = f(x)$ (this function is obtained by fitting x to Si).

Equation (9) can also be used to find the optimal inventory K at the optimal sales volume Si:

$$K = \frac{Si}{1 - ai}$$

(9)

After finding the value of K, equation (10) is used to calculate the replenishment quantity:

$$n = K - ai$$

(10)

Here ai is the forecasted inventory for the day.

Existing objective function: $Y = (x - ci)Si - (ai + n - Si)ci$

In this paper, the constraints are formulated as follows by taking the maximum value of the cost of the superstore in the past as xMax and the minimum value as xMin, and by taking the maximum value of the actual inventory as nMax and the minimum value as nMin:

$$\begin{cases} xMin \leq x \leq xMax \\ nMin \leq n \leq nMax \end{cases}$$

(11)

The C language greedy algorithm is used to calculate the maximum value of Y, and the value of x and n under the maximum value can be calculated.

For solving the function of $Si = f(x)$, this paper uses polynomial fitting [5], where the cost-plus pricing is fitted to the sales volume s data to obtain the model and thus the pricing strategy.

4. Conclusions

This paper processes and analyzes the sales data of superstore vegetables, obtains the distribution law and correlation of superstore vegetables in each category and single product, and further obtains the daily replenishment quantity and pricing strategy of superstore vegetables by establishing fitting and prediction models to maximize superstore profitability. It provides a good technical support to improve the profit of all kinds of superstores and reduce the loss rate of vegetables.

References

[1] Deng Rongjie, Fang Zhaoben. Yield analysis of financial products based on Spearman correlation analysis [J]. Statistics and Decision Making, 2019, 35(16):164-167. DOI:10.13546/j.cnki.tjyjc. 2019.16.036.

[2] Sun L, Liu MH, Xue ZZ. Feature selection combining artificial bee colony and K-means clustering [J/OL]. Computer Science and Exploration:1-18[2023-09-16].http://kns.cnki.net/kcms/detail/11.5602.TP.20230407.1412.004.html.

[3] Weichao Xu. A review of correlation coefficient research[J]. Journal of Guangdong University of Technology, 2012, 29(3): 12-17.

[4] An Qi. Research on Pricing of Scientific and Technological Research Services Based on Cost-plus Pricing Method[J]. Library Research and Work, 2021(10):25-31+24.

[5] Hasan M R , Mashud A H M , Daryanto Y ,et al.A non-instantaneous inventory model of agricultural products considering deteriorating impacts and pricing policies[J].Kybernetes: The International Journal of Systems & Cybernetics, 2021(8):50.