Academic Journal of Mathematical Sciences
ISSN 2616-5805 Vol. 5, Issue 1: 41-47, DOI: 10.25236/AJMS.2024.050107

Research on Replenishment and Pricing Strategies for Fresh Commodities in Supermarket

Yinhao He

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

Abstract: *In order to enhance returns for supermarkets, we make a replenishment plan and a pricing scheme for vegetables. Firstly, based on the results of data exploration, the correlation between the sales volume of 6 categories of vegetables is calculated by Spearman correlation coefficient. Then, 246 vegetables are divided into high-sales category, medium-sales category and low-sales category by K-means Clustering. Next, based on past sales volume, ARIMA for analyzing time series is used to make replenishment plan for vegetables. Finally, according to the markup price of vegetables, the pricing scheme for vegetables is also constructed by ARIMA.*

Keywords: *Replenishment and Pricing of Vegetables, ARIMA, Spearman Correlation Coefficient, K-means Clustering*

1. Introduction

Due to various factors such as diverse varieties, different origins, and short shelf life, the sales volume and price of vegetables are fluctuating every day. Therefore, supermarkets need to make a reasonable replenishment plan and a pricing scheme for vegetables to enhance returns.

In the past, many scholars have conducted relevant research on the replenishment and pricing of vegetables. Mao Lisha (2023) suggests that important factors affecting vegetable price fluctuations include planting temperature, shelf life, transportation conditions, etc. Then used ARIMA to analyze and predict the price trends of different categories of vegetables [1]. Zeng Minmin (2022) established dynamic pricing models for different types of products in fresh supermarkets and found that dynamic pricing strategies can reduce losses and increase profits for fresh supermarkets.[2] Zhao Yu (2019) used kernel density estimation to fit the marginal distribution of vegetable yield and price fluctuations and used the semi parametric Copula method to estimate the joint distribution function of vegetable yield and price fluctuations [3].Wang Qingyi (2019) established a multi-stage dynamic pricing model for seasonal goods using the idea of dynamic programming, and ultimately believed that under dynamic pricing strategies, moderately increasing the number of price reductions would help increase retailers' profits [4].

After reviewing relevant literature, we found that existing research mainly analyzes various factors that affect the price of vegetables, but does not provide a feasible replenishment and pricing strategy for vegetable sellers. Therefore, we plans to build on existing research and combine the sales records of vegetables in a supermarket over the past three years to construct a feasible replenishment plan and a pricing scheme for vegetables for vegetable sellers.

2. Analyze the correlation between sales of each vegetables category and sales of individual vegetables

2.1 Characteristics of sales of each vegetables category

The sales records of vegetable products in the supermarket (www.mcm.edu.cn/) include 6 types of vegetables, namely Foliage Vegetable, Flower Vegetable, Solanum, Capsicum, Edible Fungi and Aquatic Rhizomes. Firstly, we merged the daily total sales volume of each vegetable by category to obtain a total of 1085 days of sales volume for each category of vegetables, and calculated the standard deviation (S.D.) of daily sales for each category of vegetables. The results are shown in Table 1.

From Table 1, we can see the characteristics of sales volume for six categories of vegetables. For example, Solanum has the most stable sales volume, with no significant difference in daily sales volume. What's more, the standard deviation of daily sales of Foliage Vegetable is the largest, indicating

ISSN 2616-5805 Vol. 5, Issue 1: 41-47, DOI: 10.25236/AJMS.2024.050107

significant changes in daily sales of Foliage Vegetable.

Table 1: S.D. of daily sales of 6 category of vegetables

Category	S.D. of daily sales	Category	S.D. of daily sales
Foliage Vegetable	86.30	Capsicum	53.46
Flower Vegetable	22.71	Edible Fungi	48.52
Solanum	13.49	Aquatic Rhizomes	31.37

Considering that the sales of vegetables have seasonal patterns, the sales volume of vegetables in the same category will not change significantly in the same season. Therefore, we then sorted the daily sales data into monthly sales data and obtained the changes in the monthly total sales volume of six categories of vegetables as shown in Figure 1.

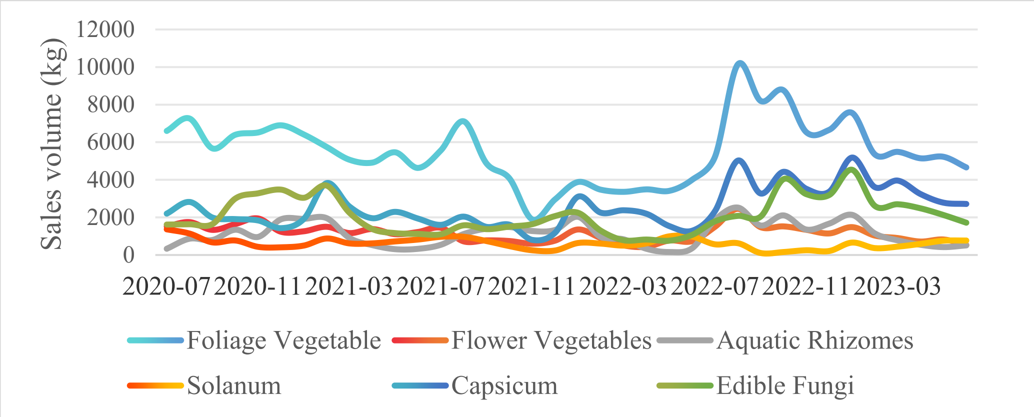

Figure 1: Monthly sales of 6 categories of vegetables

From Figure 1, it can be seen that Foliage Vegetable always have the highest sales, followed by Edible Fungi and Capsicum, while Solanum has the lowest sales. In addition, by observing Figure 1, we can also know some common characteristics of the sales volume of six categories of vegetables:

①The monthly sales volume of six categories of vegetables is in a dynamic process of change. This may be because, by comparison, the variety of vegetables available for sale is generally more abundant from April to October each year.

②The sales volume of six categories of vegetables is relatively high from August to September each year.

③Starting from the late 2022, the overall sales volume of vegetables has been higher than the sales in 2020 and 2021. We suppose that because of the pandemic, the frequency of people going out to shop has been greatly reduced, which has led to a decrease in the number of people buying vegetables, leading to a generally low sales volume of vegetables. However, in the late 2022, the pandemic basically ended, and people's daily life will gradually return to the state before the pandemic, and the frequency of shopping will increase, which will promote the increase of vegetable sales.

2.2 Correlation between sales of different categories of vegetables

Table 2: Spearman correlation coefficient results of 6 category of vegetables

	Foliage Vegetable	Flower Vegetable	Edible Fungi	Aquatic Rhizomes	Solanum	Capsicum
Foliage Vegetable	1	0.695 (0.000**)	0.578 (0.000**)	0.448 (0.006**)	-0.043 (0.802)	0.488 (0.003**)
Flower Vegetable	0.695 (0.000**)	1	0.462 (0.005**)	0.427 (0.009**)	0.076 (0.659)	0.31 (0.066)
Edible Fungi	0.578 (0.000**)	0.462 (0.005**)	1	0.669 (0.000**)	-0.447 (0.006**)	0.49 (0.002**)
Aquatic Rhizomes	0.448 (0.006**)	0.427 (0.009**)	0.669 (0.000**)	1	-0.467 (0.004**)	0.316 (0.060)
Solanum	-0.043 (0.802)	0.076 (0.659)	-0.447 (0.006**)	-0.467 (0.004**)	1	-0.17 (0.323)
Capsicum	0.488 (0.003**)	0.31 (0.066)	0.49 (0.002**)	0.316 (0.060)	-0.17 (0.323)	1
**, * represents significance levels of 1% and 5%.						

ISSN 2616-5805 Vol. 5, Issue 1: 41-47, DOI: 10.25236/AJMS.2024.050107

In order to analyze the correlation between the sales volume of six categories of vegetables, we need to calculate the correlation coefficient. First of all, Pearson correlation coefficient cannot be used because the sales volume data of these six categories of vegetables clearly does not follow a normal distribution. At the same time, the sales of vegetables are time series so the sales aren't ordinal variable, therefore Kendall rank correlation coefficient cannot be used. Finally, we use Spearman correlation coefficient to analyze the correlation between the sales volume of different categories of vegetables, as shown in Table 2.

According to Table 2, the correlation coefficient between the sales of Foliage Vegetable and Flower vegetable is $r_s = 0.695$, $p = 0.000 < 0.01$, indicate a strong positive correlation between the sales of Foliage Vegetable and Flower Vegetable. So it can be considered that as the sales of Foliage Vegetable increase, the sales of Flower Vegetable will also increase. The reverse is also true.

What's more, the correlation coefficient between the sales of Solanum and Aquatic Rhizomes is $r_s = -0.467$, $p = 0.004 < 0.01$, indicate a strong negative correlation between the sales of Aquatic Rhizomes and Solanum. So it can be considered that as the sales of Aquatic Rhizomes increase, the sales of Solanum will decrease. The reverse is also true.

Similarly, we can observe a correlation between the sales of other different categories of vegetables. Summarize all related relationships as follows:

The sales volume of Foliage Vegetable is positively correlated with the sales volume of Flower Vegetable, Edible Fungi, Aquatic Rhizomes, and Capsicum; The sales volume of Flower Vegetable is positively correlated with the sales volume of Edible Fungi, Aquatic Rhizomes, and Capsicum; The sales volume of Edible Fungi is positively correlated with the sales volume of Aquatic Rhizomes and Capsicum; The sales volume of Capsicum is positively correlated with the sales volume of Aquatic Rhizomes; The sales volume of Solanum is negatively correlated with the sales volume of Edible Fungi and Aquatic Rhizomes.

2.3 Relationship between sales of each vegetable

Based on the data provided by the supermarket, we can conduct cluster analysis on the monthly sales data of various vegetables to discover the relationship between the sales volume of different vegetables. At present, common clustering algorithms include K-means clustering algorithm, hierarchical cluster method, DBSCAN, etc. Therefore, we need to conduct an analysis of these three clustering algorithms based on the characteristics of transaction data, and ultimately select the most suitable algorithm for clustering analysis.

DBSCAN is a density based clustering method that can handle clusters of any shape and size. However, DBSCAN is generally suitable for data with two indicators, which is not consistent with the characteristics of the monthly sales data of each individual vegetable we have compiled. Therefore, we won't use DBSCAN.

The hierarchical cluster method aggregates the closest two types of data by calculating the distance between them, and iterates this process repeatedly until all data points are grouped together. After the calculation is completed, the number of clusters can be determined according to the lineage diagram. The principle of hierarchical cluster is relatively simple, but is difficult to calculate. Now we have monthly total sales data for 246 different products, which is very large in volume, using hierarchical cluster method for clustering analysis may greatly increase computational complexity and make it difficult to classify through lineage diagrams. Therefore, we don't use hierarchical cluster method.

Based on the analysis, we decide to use K-means clustering algorithm. The principle and process of K-means clustering algorithm are relatively simple, and its efficiency is much higher than hierarchical cluster method when processing large sample data. However, we recognize that the prerequisite for using the K-means clustering algorithm is to determine the number of clusters. For the sales data from the supermarket, we still do not know the number of clusters. Therefore, in order to determine the number of clusters, we will select the optimal number of clusters based on the Elbow Method and the line chart of clustering coefficient, and then perform cluster analysis. In summary, the flowchart of K-means clustering algorithm is shown in Figure 2.

ISSN 2616-5805 Vol. 5, Issue 1: 41-47, DOI: 10.25236/AJMS.2024.050107

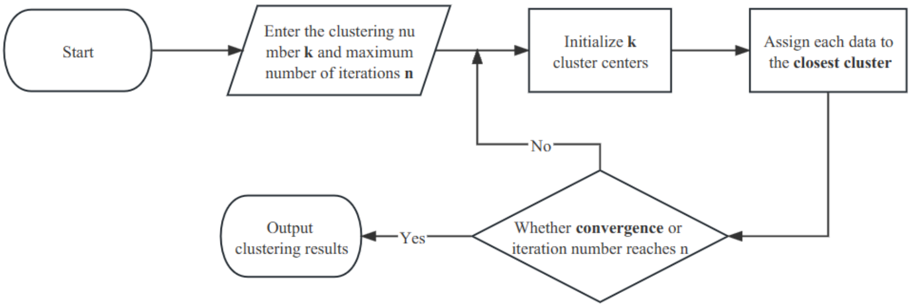

Figure 2: Flowcharts of K-means clustering analysis

The basic idea of using the Elbow Method to determine the optimal number of clusters is to measure the effectiveness of the K-means clustering algorithm only through intra cluster distance. The sum of squares of the distance between samples within the same cluster and the center of its cluster can be recorded as the degree of distortion of that cluster, and the sum of the degree of distortion of each cluster is recorded as the intra cluster distance. The smaller the value, the closer the data within the cluster is, the better the clustering effect. Conversely, the clustering effect is worse. Usually, the relationship between the number of different clusters and their corresponding intra cluster distances is presented in a two-dimensional image, commonly referred to as the line chart of clustering coefficient. Observing the line chart of clustering coefficient, it can be seen that as the number of clusters gradually increases, the intra cluster distance first decreases rapidly. However, when the number of clusters takes a certain value, the change in intra cluster distance gradually becomes gentle. Even if the number of clusters increases again, the value of intra cluster distance will not change significantly [5]. This number of clusters is the optimal number of clusters we require.

First of all, set any number of clusters and cluster the monthly total sales of each vegetable using the K-means clustering algorithm. The line chart of clustering coefficient can be obtained as shown in Figure 3, where the horizontal axis represents the number of clusters.

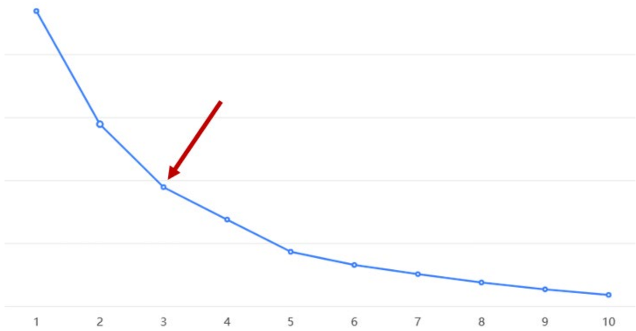

Figure 3: Line chart of clustering coefficient

According to Figure 3, when the number of clusters is 3, the downward trend of the intra cluster distance curve gradually slows down, that is, the change in intra cluster distance significantly decreases. Therefore, we can assume that the optimal number of clusters is 3.

Next, we set the number of clusters to 3 and cluster the monthly total sales data of each individual vegetable again. The summary of clustering results is shown in Table 3, and the evaluation of clustering effectiveness are shown in Table 4.

Table 3: Summary of clustering results

Category	**Frequency**	**Proportion (%)**
Category 1	19	7.72
Category 2	218	88.62
Category 3	9	3.66
Total	246	100.00

According to Table 3 and the specific clustering results, Category 1 has 19 vegetables, including Shanghai Greens, Chinese Cabbage, and so on; Category 2 has 218 vegetables, including Lettuce from Niushou, Cabbage Moss, etc; Category 3 has 9 vegetables, including Green Peppers from Wuhu, broccoli, and more.

ISSN 2616-5805 Vol. 5, Issue 1: 41-47, DOI: 10.25236/AJMS.2024.050107

Table 4: Evaluation of clustering effectiveness

Silhouette Coefficient	DBI	CH
0.742	1.152	117.688

The DBI is used to measure the ratio of the sum of intra cluster distances to inter cluster distances between any two clusters. A smaller DBI value indicates better clustering performance. The CH measures the compactness within a cluster by calculating the sum of the squares of the distances between each data within the cluster and the center of the cluster, and measures the separation of the data set by calculating the sum of the squares of the distances between the center points between clusters and the center points of the data set. The CH is the ratio of separation to compactness, so a larger CH indicates a better clustering effect. From Table 4, it can be seen that DBI=1.152, CH=117.688, and Silhouette Coefficient = 0.742 is close to 1, so we suppose that the clustering effect is good.

Finally, we calculated average sales volume of 3 categories of vegetables, as shown in Table 5.

Table 5: Average sales volume of 3 categories

Category	Representative Vegetables	Average sales volume(kg)
Category 1	Chinese Cabbage	8033.94
Category 2	Cabbage Moss	711.29
Category 3	Broccoli	18174.43

Based on the category characteristics presented in Table 5, we can conclude that Category 3 is high-sales category, Category 1 is medium-sales category, and Category 2 is low-sales category. Therefore, for 246 vegetables, 9 were included in high-sales category, with an average sales volume of 18174 kg in high-sales category; 19 vegetables were included in medium-sales category; 218 vegetables were included in low-sales category, accounting for 88.6%.

3. Make future replenishment plan and pricing scheme for various vegetable categories

3.1 Make future replenishment plan for each category of vegetables

In order to make future replenishment plan for each category of vegetables based on past sales data, it is not comprehensive to only analyze the sales data of each vegetable category in the past. Each vegetable product will have a certain loss rate, and the loss of vegetables will have a certain negative impact on sales. Therefore, we suppose that when formulating a future replenishment plan, in addition to referring to past sales, it is also necessary to consider the impact of loss rate, namely.

Daily replenishment volume = daily sales volume + possible losses

Next, we calculate the past replenishment volume of each vegetable based on its past sales volume and loss rate, and accumulates it by category to obtain the past replenishment volume of each vegetable category. For convenience, taking the next week as an example, we establish ARIMA to analyze the past replenishment volume of each vegetable category, predicts the daily replenishment volume of each vegetable category in the next week, and make replenishment plans based on this. Taking Foliage Vegetable as an example, the best fitting model is ARIMA(1,0,9), and the evaluation of ARIMA(1,0,9) are shown in Table 6.

Table 6: Evaluation of ARIMA(1,0,9)

Index	Average value	Index	Average value
R^2	.744	MaxAPE	131.942
RMSE	51.946	MAE	37.006
MAPE	18.635	Normalized BIC	8.023

According to Table 6, $R^2 = 0.744$, is relatively close to 1. In addition, RMSE and MAPE are at lower levels. Considering that the transaction data of supermarkets are all measured data, which may have high volatility, it can be supposed that the model has a good fitting effect.

Table 7: Future daily replenishment

Date	Daily replenishment (kg)	Date	Daily replenishment (kg)
July 1st	144.76	July 5th	159.30
July 2nd	139.73	July 6th	152.37
July 3rd	138.71	July 7th	155.79
July 4th	153.29	/	/

According to ARIMA(1,0,9), the daily replenishment volume of Foliage Vegetable in the next week can be predicted, as shown in Table 7.

The same method can be used to analyze and predict the past replenishment volume of each vegetable category. Based on the prediction results, a daily replenishment plan for each vegetable category in the next week can be formulated, as shown in Table 8.

Table 8: Daily replenishment plan for the next week of 6 categories of vegetables (kg)

Date	Foliage Vegetable	Flower Vegetables	Aquatic Rhizomes	Solanum	Capsicum	Edible Fungi
July 1st	144.76	25.40	25.08	26.47	87.22	44.66
July 2nd	139.73	23.22	22.53	24.34	82.14	47.44
July 3rd	138.71	21.71	24.55	22.59	80.63	49.94
July 4th	153.29	23.15	23.14	21.44	83.07	53.40
July 5th	159.30	24.24	24.73	20.00	87.52	52.31
July 6th	152.37	25.36	24.05	20.65	91.09	50.92
July 7th	155.79	27.22	24.61	20.30	91.99	48.86

In summary, if we want to make future replenishment plans for each category of vegetables, we can construct ARIMA based on the same method and make predictions, ultimately making a future replenishment plan.

3.2 Make future pricing scheme for each category of vegetables

After making a replenishment plan for each vegetable category in the future, we also want to make a pricing scheme for each vegetable category in the future. The pricing of vegetables follows the principle of "markup pricing", which adds "markup pricing" to the wholesale price, and ultimately obtains pricing. In order to make pricing scheme for various vegetable categories in the future and maximize profits, we will analyze the past supply and demand situation and pricing of each vegetable category. The vegetables sold in supermarkets vary from category to category, and the sales of different items in the same category also have their own characteristics. In order to make the average markup pricing of each vegetable category more representative and accurate, we first calculate the proportion of daily sales of each vegetable in the same category to the total sales of that category on that day, and use this as the weight of each vegetable. Then, combined with the daily markup pricing of each individual vegetable, calculate the daily "weighted average markup pricing" for each vegetable category.

The daily weighted average markup pricing for a vegetable category is defined as follows:

$$X = \sum x_i q_i \,, \tag{1}$$

x_i is the markup pricing of each individual item under the category, which is the difference between the pricing of the vegetable on that day and the wholesale price of the vegetable on that day, and q_i is the sales weight of each vegetable under the category, which is the proportion of the sales volume of the vegetable item on that day to the total sales volume of the entire category on that day.

After calculating the daily weighted average markup pricing for each vegetable category, we establishes ARIMA to analyze the weighted average markup pricing for each vegetable category, and automatically finds the optimal parameters for ARIMA based on AIC information criteria. Taking Foliage Vegetable as an example, the best ARIMA for this category is ARIMA(1,1,1), and the formula is

$$y(t) = -0.001 + 0.498 * y(t-1) - 0.788 * \varepsilon(t-1), \tag{2}$$

The evaluation of fitting effect of $\mathrm{ARIMA}(1,1,1)$ are shown in Table 9.

Table 9: Fitting effect of ARIMA (1,1,1)

Index	Character	Value
Sample quantity	N	1085
Information criterion	AIC	20.957
	BIC	40.91
Goodness of fit	R²	0.853

According to Table 10, $R^2 = 0.853$, is relatively close to 1, indicating that the fitting effect of $\mathrm{ARIMA}(1,1,1)$ is very good.

For convenience, taking the next week as an example. Based on $\mathrm{ARIMA}(1,1,1)$, we predict the daily weighted average markup pricing of Foliage Vegetable for the next week, shown in Table 10.

ISSN 2616-5805 Vol. 5, Issue 1: 41-47, DOI: 10.25236/AJMS.2024.050107

Table 10: Future weighted average markup pricing of Foliage Vegetable

Date	Weighted average markup pricing	Date	Weighted average markup pricing
July 1st	2.01	July 5th	2.02
July 2nd	1.98	July 6th	1.94
July 3rd	1.93	July 7th	1.91
July 4th	2.12	/	/

Similarly, ARIMA can be constructed to analyze the daily weighted average markup pricing of each vegetable category, and predict the weighted average markup pricing of each vegetable category for the next week, in order to make pricing schemes for each vegetable category in the next week. The daily weighted average markup pricing prediction results for each vegetable category in the next week are shown in Table 11. Pricing schemes for each vegetable category can be made.

Table 11: Daily pricing scheme for the next week of 6 categories of vegetables (CNY)

Date	Foliage Vegetable	Flower Vegetables	Aquatic Rhizomes	Solanum	Capsicum	Edible Fungi
July 1st	2.01	4.19	5.40	3.25	2.66	1.83
July 2nd	1.98	4.36	5.27	3.01	2.58	1.96
July 3rd	1.93	4.01	5.48	2.92	2.19	1.99
July 4th	2.12	4.31	5.77	2.83	2.44	2.21
July 5th	2.02	4.63	5.89	2.98	2.80	2.15
July 6th	1.94	5.01	5.52	3.11	2.52	2.19
July 7th	1.91	4.49	5.23	2.88	2.24	2.01

In summary, if we want to make future pricing schemes for each category of vegetables, we can construct ARIMA based on the same method and make predictions, ultimately making a future pricing scheme.

4. Conclusions

By analyzing the past sales records of each category of vegetables, it can be seen that the sales volume of Foliage Vegetable is positively correlated with the sales volume of Flower Vegetable, Edible Fungi, Aquatic Rhizomes, and Capsicum; The sales volume of Flower Vegetable is positively correlated with the sales volume of Edible Fungi, Aquatic Rhizomes, and Capsicum; The sales volume of Edible Fungi is positively correlated with the sales volume of Aquatic Rhizomes and Capsicum; The sales volume of Capsicum is positively correlated with the sales volume of Aquatic Rhizomes; The sales volume of Solanum is negatively correlated with the sales volume of Edible Fungi and Aquatic Rhizomes.

By analyzing the past transaction status of each vegetable sold by supermarkets, 246 vegetables can be divided into high-sales category, medium-sales category, and low-sales category. 9 vegetables are included in high-sales category with an average sales volume of 18174kg; 19 vegetables were included in medium-sales category, with an average sales volume of 8034kg; 218 vegetables were included in low-sales category, accounting for 88.6%, with an average sales volume of 711kg.

Based on the past sales volume and the impact of loss rate, ARIMA is used to analyze and predict the past replenishment volume of various categories of vegetables sold by the supermarket, providing a method for making future replenishment plans for each category of vegetables. Moreover, based on the definition of weighted average markup pricing, ARIMA is used to analyze the past pricing and sales volume of six categories of vegetables, providing a method for making future pricing scheme for each category of vegetables.

References

[1] Mao Lisha. Research on pricing strategy and production and marketing model of vegetable wholesale market from the perspective of supply chain [D]. Central South University of Forestry and Technology, 2023.
[2] Zeng Minmin. Research on the dynamic pricing strategy of a fresh supermarket based on time situation [D]. Southwestern University of Finance and Economics, 2022.
[3] Zhao Yu. Vegetable income insurance pricing: based on semi-parametric Copula method [J]. Research of Agricultural Modernization, 2019, 40(2): 308-315.
[4] Wang Qingyi, Xu ling. Dynamic pricing model of retailer with consideration of customer perceived value [J]. Logistics Technology, 2019, 38(04):23-28+117.
[5] Zeng Ruming. Improvement and application of K-means clustering algorithm [D]. China West Normal University, 2023.