

Coffee Buyer Segmentation Using PCA Model and K-Means Algorithm on Coffee Customers

Suharmanto¹, Muhammad Faisal*², Wiranti Sri Utami³, Janu Ilham Saputro⁴

¹Graha Ganesha University, South Tangerang, Indonesia

^{2,4}University of Raharja, Tangerang, Indonesia

³Cendekia Abditama University, Tangerang Regency, Indonesia

*Corresponding author: muhammad.faisal@raharja.info

Article Information:

Received: 19 February 2026

Revised: 08 July 2026

Accepted: 14 July 2026

DOI: 10.33050/icit.v12i2.4330

ABSTRACT

Coffee is a plant that originates from the agricultural industry which is produced and made into a drink through coffee beans and a series of processes such as roasting and grinding it into coffee powder which will be brewed. Currently, many countries have cultivated coffee plants and there are two generally well-known coffees, namely Robusta coffee and Arabica coffee. Each buyer has different characteristics in enjoying coffee drinks, including the time of purchase, the product purchased from the coffee shop, and the location of the coffee shop. Therefore, research is needed to develop a marketing strategy that suits the needs of the buyer segment. To support this research the author uses the PCA (Principal Component Analysis) model and the K-Means method to group customer characteristics. Standard deviation is used to determine the average value of each cluster in coffee shop outlets, with a PCA value on product_id of 0.99986001 and unit_price of 0.01673209 from a total of 96% of clusters. This research can be used as a good reference for students, researchers, and can also be used as a marketing strategy medium.

Keywords—Coffee, PCA, K-Means, Customer Segmentation

1. INTRODUCTION

Coffee is a popular beverage enjoyed by many people. Coffee's arrival in Indonesia dates back to the Dutch colonial era. Coffee seeds were first brought to Java by the Dutch in 1696 from Malabar, India, and the seeds themselves originated in Yemen [1]. A survey was conducted to observe popular beverage trends in 2022, 60% of respondents consumed coffee in the last 6 months and 40% in the last 3 months.

[2]. As time goes by, there are many coffee making techniques and many business people are using this popular drink for business purposes.

The challenge facing the coffee beverage business is customer preferences regarding coffee types. Coffee is divided into two types: Robusta and Arabica, with many other variations, including Aceh Gayo, Vietnam Drip, and others. Furthermore, when do customers typically purchase coffee? To manage this coffee beverage business, research can be conducted to determine the types of drinks preferred by customers, the time of day they purchase coffee, and the outlets with the highest purchase volume. The data collected will be easier to understand when visualized in graphical form.

Several factors influence someone's coffee purchase, including environmental influences or peer influence, and the spread of information about coffee through media. Coffee consumption is currently on the rise, as evidenced by the increasing number of cafes and coffee shops filled with

young people and even adults [3]. Coffee is also believed to improve concentration due to its caffeine content. Caffeine is readily absorbed by the human body within 30-120 minutes. Low doses of caffeine, between 50-200 mg, have positive effects on alertness, energy, relaxation, and a positive mood [4]. Thus, it can be concluded that coffee is the most popular drink among young people and adults.

To maintain coffee product quality and competitiveness among competitors, coffee shop owners must have a strategy to increase purchasing power and sales. With increasingly fierce competition in the business world, entrepreneurs are required to focus not only on their products but also on their customers [5]. In an effort to support the business strategy, the author tried to conduct research using segmentation analysis using the K-Means method. Segmentation analysis is able to provide an overview of the characteristics of each different cluster [6]. The K-Means method is very popular because it is simple and easy to implement. However, in some cases, this algorithm has limitations in its application. Therefore, an extraction technique using PCA (Principal Component Analysis) is applied. PCA is a method used to simplify data sets and transform them linearly [7].

PCA can be used in several studies, including for pattern recognition, semantic classification, text classification, and so on. It is usually used for analyzing data that has high dimensions [8]. The PCA method is used to reduce the dimensions of a dataset before it is grouped into K-Means. This reduction is carried out to reduce the workload of data storage memory, reduce computing time, and improve data quality [9].

Based on the explanation above regarding segmentation and PCA, the author will conduct research on customer segmentation using the PCA model. The model used by the author is to provide more accurate information regarding frequently visited coffee shops, the best time for customers to buy coffee, and their favorite products.

Customers are assets that must be managed in order to maintain the business. The more loyal customers there are, the better it is for the company's business interests [10]. Companies can carry out customer segmentation processes to support their business. The aim of the customer segmentation process is to understand customer behavior and implement appropriate customer strategies so that they can provide benefits to customers [11]. To support research on customer segmentation, the author also used data mining methods to group or cluster customers. Clustering is one of the data mining methods. Data mining has many definitions, one of which is extracting information and patterns from large data sets [12].

The cluster analysis process is a technique used to identify similar objects or individuals by paying attention to criteria. This process also groups similar elements as research objects into groups or clusters [13]. In the application of clustering techniques, there are several algorithms for the clustering process, one of which is the K-Means algorithm. K-Means is a partitional algorithm because it is based on determining the initial number of groups with the initial center point value or centroid [14].

Based on a literature review conducted using consistent data analysis methods and techniques, the author will analyze frequently visited coffee shops, peak times for coffee purchases, and frequently purchased products using the K-Means method and PCA analysis model.

2. LITERATURE REVIEW

The stages in conducting this research are formulated in the form of a research sequence. The sequence of research steps can be seen in the image below.

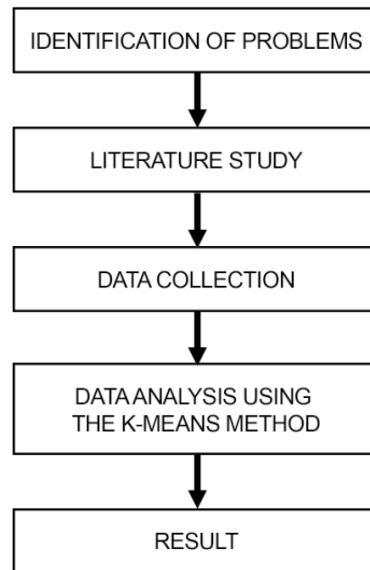


Figure 1. Sequence of research steps

The data used in this study was obtained from the kaggle.com dataset. The data contains coffee sales from a coffee shop, including date, time, coffee shop location, product, and revenue. This data represents sales transactions in 2023. This data can be seen in the table below.

Table 1. Dataset coffee shop

transacti	transacti	transacti	tra	sto	store_loc	pro	uni	product	product	product	Reven	Mont	Mont	Week	Week	Hour
1	01/01/2023	07:06:11	2	5	Lower Mani	32	3	Coffee	Gourmet br	Ethiopia Rg	\$6,00	1 Jan	7 Sun	7		
2	01/01/2023	07:08:56	2	5	Lower Mani	57	3,1	Tea	Brewed Cha	Spicy Eye O	\$6,20	1 Jan	7 Sun	7		
3	01/01/2023	07:14:04	2	5	Lower Mani	59	4,5	Drinking Ch	Hot chocola	Dark chocol	\$9,00	1 Jan	7 Sun	7		
4	01/01/2023	07:20:24	1	5	Lower Mani	22	2	Coffee	Drip coffee	Our Old Tim	\$2,00	1 Jan	7 Sun	7		
5	01/01/2023	07:22:41	2	5	Lower Mani	57	3,1	Tea	Brewed Cha	Spicy Eye O	\$6,20	1 Jan	7 Sun	7		
6	01/01/2023	07:22:41	1	5	Lower Mani	77	3	Bakery	Scone	Oatmeal Scc	\$3,00	1 Jan	7 Sun	7		
7	01/01/2023	07:25:49	1	5	Lower Mani	22	2	Coffee	Drip coffee	Our Old Tim	\$2,00	1 Jan	7 Sun	7		
8	01/01/2023	07:33:34	2	5	Lower Mani	28	2	Coffee	Gourmet br	Columbian I	\$4,00	1 Jan	7 Sun	7		
9	01/01/2023	07:39:13	1	5	Lower Mani	39	4,25	Coffee	Barista Espr	Latte Rg	\$4,25	1 Jan	7 Sun	7		
10	01/01/2023	07:39:34	2	5	Lower Mani	58	3,5	Drinking Ch	Hot chocola	Dark chocol	\$7,00	1 Jan	7 Sun	7		
11	01/01/2023	07:43:05	1	5	Lower Mani	56	2,55	Tea	Brewed Cha	Spicy Eye O	\$2,55	1 Jan	7 Sun	7		
12	01/01/2023	07:44:35	2	5	Lower Mani	33	3,5	Coffee	Gourmet br	Ethiopia Lg	\$7,00	1 Jan	7 Sun	7		
13	01/01/2023	07:45:51	1	5	Lower Mani	51	3	Tea	Brewed Blac	Earl Grey Lg	\$3,00	1 Jan	7 Sun	7		
14	01/01/2023	07:48:19	1	5	Lower Mani	57	3,1	Tea	Brewed Cha	Spicy Eye O	\$3,10	1 Jan	7 Sun	7		

3. METHODOLOGY

Based on the author's discussion, data mining with the K-Means algorithm will be used to identify coffee customer segmentation based on clusters such as store or coffee shop purchase rates, transaction times, and product characteristics.

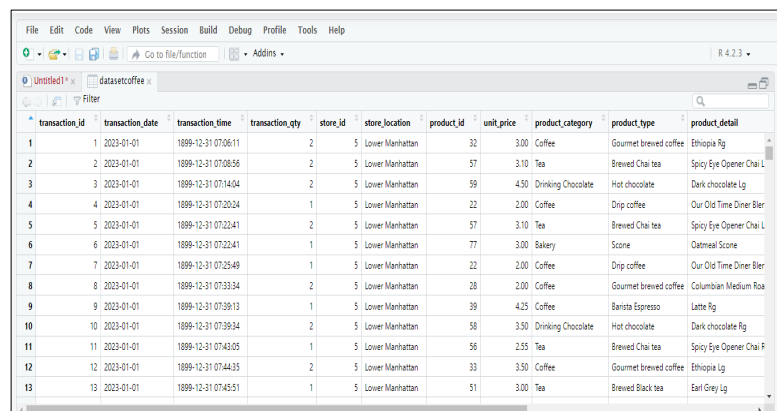
3.1. Data Processing

The data obtained by the author is coffee sales transaction data at each store or coffee shop. The data includes several transaction ID records, transaction time, store location, product, and revenue.

The data contains 14 attributes, and we will select which ones will be used for this study. Then, the data will be analyzed using PCA techniques, including other attributes. From all the data being processed, cluster modeling will be performed using Principal Component Analysis.

3.2. Data Preparation

The data obtained consists of 149,456 coffee purchase transaction data, representing sales transactions from early January 2023 to June 30, 2023. Preparation was carried out by managing the values of each attribute to generate a PCA (Principal Component Analysis). The data will then be uploaded for processing using the Posit Cloud application with the R programming language. The data upload process can be seen in the image below.



transaction_id	transaction_date	transaction_time	transaction_qty	store_id	store_location	product_id	unit_price	product_category	product_type	product_detail
1	2023-01-01	1899-12-31 07:06:11	2	5	Lower Manhattan	32	3.00	Coffee	Gourmet brewed coffee	Ethiopia Rg
2	2023-01-01	1899-12-31 07:08:56	2	5	Lower Manhattan	57	3.10	Tea	Brewed Chai tea	Spicy Eye Opener Chai L
3	2023-01-01	1899-12-31 07:14:04	2	5	Lower Manhattan	59	4.50	Drinking Chocolate	Hot chocolate	Dark chocolate Lg
4	2023-01-01	1899-12-31 07:20:24	1	5	Lower Manhattan	22	2.00	Coffee	Drip coffee	Our Old Time Diner Bler
5	2023-01-01	1899-12-31 07:22:41	2	5	Lower Manhattan	57	3.10	Tea	Brewed Chai tea	Spicy Eye Opener Chai L
6	2023-01-01	1899-12-31 07:22:41	1	5	Lower Manhattan	77	3.00	Bakery	Scone	Oatmeal Scone
7	2023-01-01	1899-12-31 07:25:49	1	5	Lower Manhattan	22	2.00	Coffee	Drip coffee	Our Old Time Diner Bler
8	2023-01-01	1899-12-31 07:33:34	2	5	Lower Manhattan	28	2.00	Coffee	Gourmet brewed coffee	Columbian Medium Roa
9	2023-01-01	1899-12-31 07:39:13	1	5	Lower Manhattan	39	4.25	Coffee	Barista Espresso	Latte Rg
10	2023-01-01	1899-12-31 07:39:34	2	5	Lower Manhattan	58	3.50	Drinking Chocolate	Hot chocolate	Dark chocolate Rg
11	2023-01-01	1899-12-31 07:43:05	1	5	Lower Manhattan	56	2.55	Tea	Brewed Chai tea	Spicy Eye Opener Chai F
12	2023-01-01	1899-12-31 07:44:35	2	5	Lower Manhattan	33	3.50	Coffee	Gourmet brewed coffee	Ethiopia Lg
13	2023-01-01	1899-12-31 07:45:51	1	5	Lower Manhattan	51	3.00	Tea	Brewed Black tea	Earl Grey Lg

Figure 1. Upload dataset to Posit Cloud application

After uploading the data, the author wanted to determine the average coffee sales by date, time, store, and product. These average values can be seen in the image below.

```
> summary(coffee_shop)
transaction_id transaction_date
Min. : 1.0 Min. :2023-01-01 00:00:00.00
1st Qu.:250.5 1st Qu.:2023-01-01 00:00:00.00
Median :500.0 Median :2023-01-01 00:00:00.00
Mean :500.0 Mean :2023-01-01 10:47:12.43
3rd Qu.:749.5 3rd Qu.:2023-01-02 00:00:00.00
Max. :999.0 Max. :2023-01-02 00:00:00.00
transaction_time transaction_qty
Min. :1899-12-31 07:01:59.00 Min. :1.000
1st Qu.:1899-12-31 11:25:27.50 1st Qu.:1.000
Median :1899-12-31 13:40:47.00 Median :1.000
Mean :1899-12-31 13:38:07.61 Mean :1.427
3rd Qu.:1899-12-31 16:00:33.00 3rd Qu.:2.000
Max. :1899-12-31 19:58:52.00 Max. :2.000
store_id store_location product_id
Min. :3.00 Length:999 Min. :22.00
1st Qu.:3.00 Class :character 1st Qu.:32.00
Median :5.00 Mode :character Median :45.00
Mean :5.34 Mean :46.22
3rd Qu.:8.00 3rd Qu.:56.00
Max. :8.00 Max. :87.00
```

Figure 2. Average transaction value Date, time, and store

```

unit_price    product_category    product_type
Min.   :2.000    Length:999    Length:999
1st Qu.:2.500    Class :character    Class :character
Median :3.000    Mode  :character    Mode  :character
Mean   :3.113
3rd Qu.:3.500
Max.   :4.750
product_detail Revenue    Month...13
Length:999    Min.   :2.000    Min.   :1
1st Qu.:3.000    1st Qu.:3.000    1st Qu.:1
Median :4.000    Median :4.000    Median :1
Mean   :4.398    Mean   :4.398    Mean   :1
3rd Qu.:6.000    3rd Qu.:6.000    3rd Qu.:1
Max.   :9.500    Max.   :9.500    Max.   :1
Month...14    Weekday...15    Weekday...16
Length:999    Min.   :1.000    Length:999
Class :character    1st Qu.:1.000    Class :character
Mode  :character    Median :7.000    Mode  :character
Mean   :4.303
3rd Qu.:7.000
Max.   :7.000
Hour
Min.   : 7.00
1st Qu.:11.00
Median :13.00
Mean   :13.14
3rd Qu.:16.00
Max.   :19.00

```

Figure 3. Average transaction value of Products, months, and hours

4. RESULT AND DISCUSSION

The next step is clustering modeling using the K-Means method. This process uses the open-source software Posit Cloud to determine PCA values using the K-Means algorithm. The cluster selection process is determined by the product_id and unit_price columns. The clustering results can be seen in the image below.

```

Cluster means:
product_id unit_price
1    26.16740    2.602643
2    75.60544    3.510204
3    36.54455    3.500495
4    45.93264    2.715026
5    55.96522    3.357609

```

Figure 4. Cluster process

The experiment was conducted with the aim of finding out which products are frequently purchased by customers, which stores are most frequently visited by customers, and which times customers most frequently purchase coffee.

4.1 Coffee Purchase Time Attribute Modelling

The dataset obtained from kaggle.com contains coffee purchase transaction data, including purchase times. The author wanted to analyze the highest coffee purchase times at each time. The author visualized the highest coffee purchase times between 12:00 PM and 3:00 PM using the Posit Cloud application, as shown in the image below.

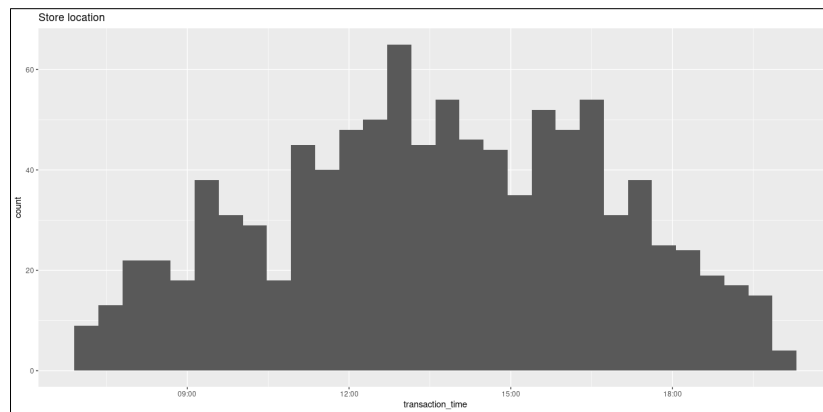


Figure 5. Visualization of coffee purchase levels based on time/hour

Then the author displays a visualization to find the average coffee purchase based on time/hour.

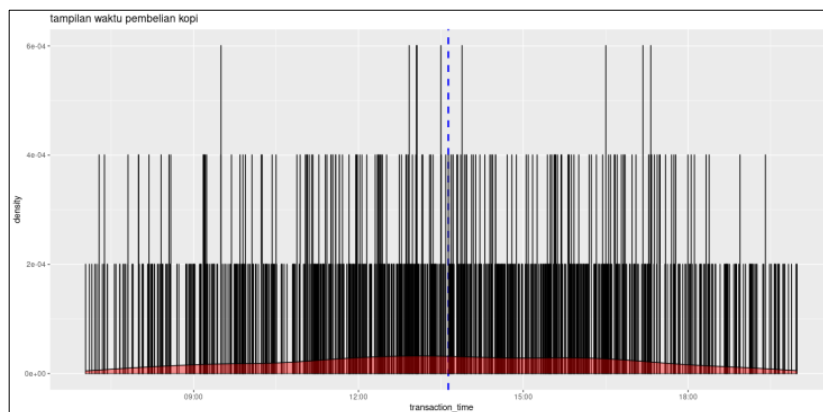
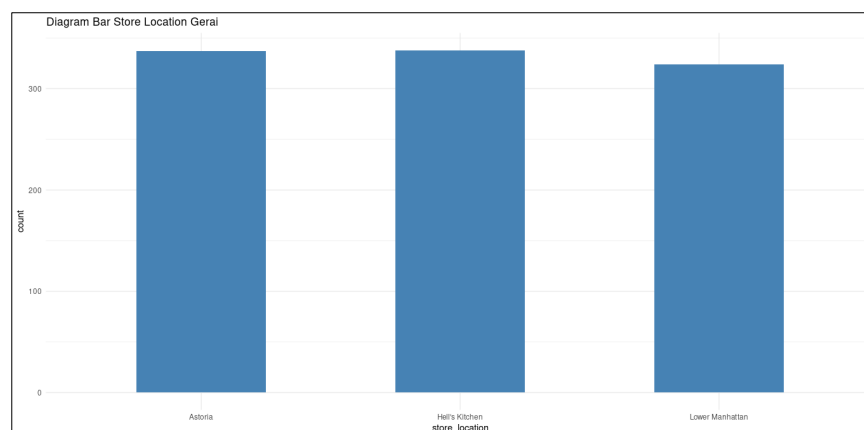


Figure 6. Visualization of average coffee purchase rate by time/hour

4.2 Store Attribute Modelling (Coffee Shop)

The next step is to visualize the store or coffee shop model. This process aims to determine the purchasing levels at each store in the data set obtained from kaggle.com. This visualization can be seen in the image below.



Gambar 7. Proses visualisasi *store*

4.3 Product Attribute Modelling

The product transactions in the coffee sales transaction dataset will be processed into a visual display. From the data records obtained, the author wants to determine the products with the highest purchase rate. With the visual display, the author can easily read which products are most frequently purchased by customers. This display can be seen in the image below.

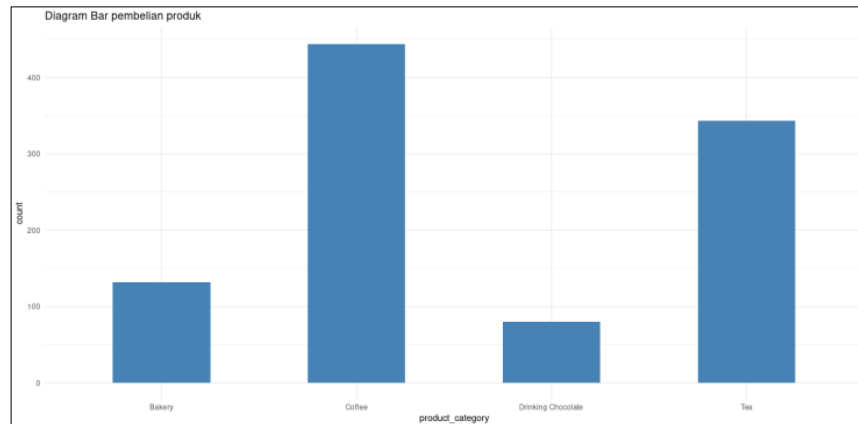


Figure 8. Visualization of product purchase rate

Based on the visualization above, the product with the highest sales rate is coffee. The author will then create a single visualization display showcasing product transactions and time. This can be seen in the image below.

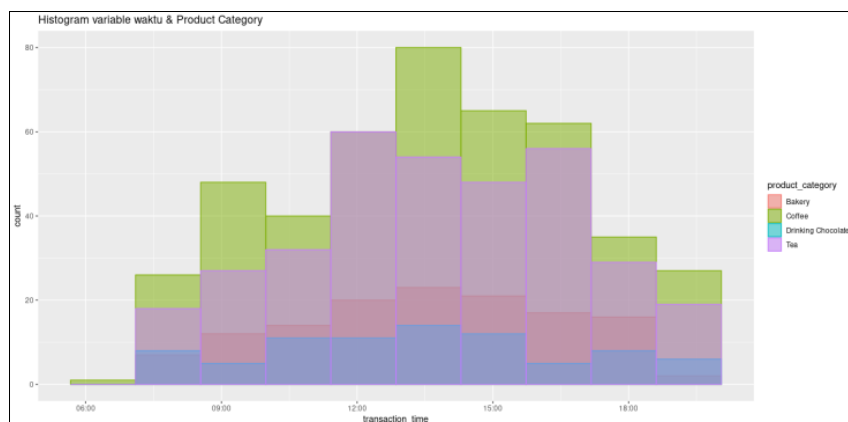


Figure 9. Visualization of product transactions and time

Based on the image above, coffee is the product with the highest sales level, followed by bakery, drinking chocolate, and tea.

4.4 Principal Component Analysis

The PCA method is a multivariate analysis method. It is used to reduce large amounts of data into smaller pieces without losing important information. The next step is to perform the PCA analysis on the product_id and unit_price data. The results of this analysis can be seen in the image below.

Importance of components:		
	PC1	PC2
Standard deviation	16.4373	0.60132
Proportion of Variance	0.9987	0.00134
Cumulative Proportion	0.9987	1.00000

Figure 11. PCA analysis

The variable model above from the product_id and unit_price columns produces 2 main components of the standard deviation value, proportion of variance, and cumulative proportion.

```
> ppclust$rotation[,1:2]
      PC1      PC2
product_id 0.99986001 -0.01673209
unit_price 0.01673209  0.99986001
>
```

Figure 12. PCA analysis of Product_id and Unit_price

From the image above, the PC model is as follows:

$$PC1 = 0.99986001 + 0.01673209$$

$$PC2 = -0.01673209 + 0.99986001$$

4.5 K-Means

In this study, the authors used the Gap method to determine the number of clusters. This method is one of the K-Means methods for determining the number of clusters. The image below is a line diagram of the Gap method. In the image below, the graph line is highest at number 5. After number 5, there is a decrease. Therefore, the authors determined that there are five clusters in this study according to the Gap method to determine the number of clusters.

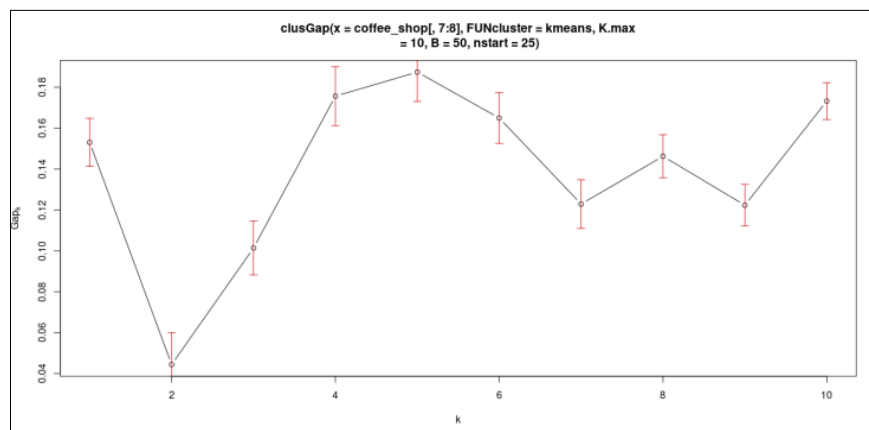


Figure 13. Line diagram of the Gap method

Next, the author entered the R language program into the posit cloud to display the cluster results. The image below shows the cluster results for the coffee shop dataset processed from data obtained from kaggle.com. The processed data consists of the product_id and unit_price columns.

```
K-means clustering with 5 clusters of sizes 227, 147, 202, 193, 230

Cluster means:
  product_id unit_price
1  26.16740   2.602643
2  75.60544   3.510204
3   36.54455   3.500495
4  45.93264   2.715026
5  55.96522   3.357609

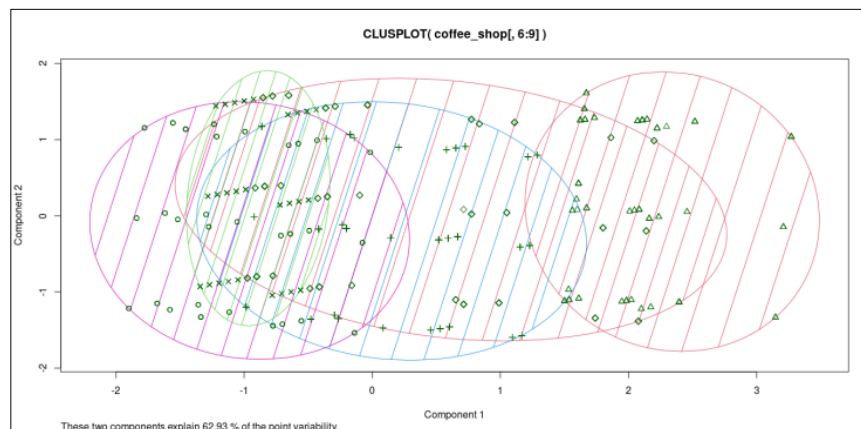
Clustering vector:
[1] 3 5 5 1 5 2 1 1 3 5 5 3 5 5 2 4 2 4 5 5 3 5 2 5 3 4 2 4 2 3 3 1 1 1 5 1 5 3 4 4 1 5 2 5 2 4 5 4 1
[50] 4 3 4 4 1 5 3 4 3 3 2 5 3 3 4 2 5 1 3 3 3 2 4 4 3 2 3 1 4 1 3 3 1 2 3 5 3 1 1 5 5 5 2 4 3 2 5 5 3
[99] 5 5 4 1 2 3 2 1 3 2 4 2 3 5 5 5 4 2 4 4 4 5 5 5 1 5 2 4 3 4 3 5 3 2 1 1 5 2 5 2 5 2 5 1 3 1 4 5 2
[148] 1 3 4 3 5 3 4 2 5 1 4 3 2 3 4 4 1 5 4 1 4 3 1 4 3 3 4 1 4 5 1 3 4 3 5 1 1 5 4 5 5 5 5 1 5 1 3 2 3
[197] 1 5 3 5 5 2 5 2 1 3 4 5 2 4 1 5 4 4 3 1 1 4 2 1 4 1 2 4 5 1 5 1 3 1 5 5 5 4 5 2 1 1 3 3 1 4 5 5 1
[246] 1 2 1 1 3 5 5 4 4 3 5 2 5 2 1 5 3 1 5 5 3 3 3 1 4 4 4 5 1 2 4 3 3 1 2 3 3 1 1 2 5 5 1 3 4 4 3 5 2
[295] 3 2 5 1 5 4 4 2 4 5 4 1 3 3 5 1 5 4 2 3 5 4 4 4 1 2 5 4 3 1 4 3 3 1 4 1 5 3 2 1 5 1 3 5 1 1 3 1 4
[344] 5 5 4 5 2 4 1 3 1 1 2 3 2 3 4 3 4 3 3 3 2 3 1 1 4 3 3 3 4 4 1 1 4 4 2 3 1 1 5 3 3 1 1 4 5 5 4 1 3
[393] 3 1 3 5 3 4 1 4 5 4 2 4 5 2 5 4 4 1 5 2 3 4 4 3 4 1 5 3 1 3 5 1 1 3 2 3 1 2 2 4 2 3 4 3 5 5 2 3 2
[442] 4 2 5 4 2 4 5 5 2 5 1 2 3 1 3 2 5 4 3 5 5 1 3 5 2 4 5 4 1 2 1 1 2 1 3 1 1 5 1 3 5 5 2 4 5 3 3 2 5
[491] 5 4 5 4 5 5 1 5 1 4 5 5 5 1 3 1 1 1 1 5 5 5 3 2 4 3 3 1 5 1 5 3 1 5 3 3 5 4 3 4 2 4 3 5 4 5 3 1 5
[540] 5 4 3 5 5 1 3 3 5 3 2 5 1 5 2 5 4 1 5 3 2 1 4 2 3 1 1 5 4 5 5 2 1 5 3 1 4 4 3 1 1 1 1 1 1 4 3
[589] 3 1 3 5 3 3 5 2 5 3 2 1 2 4 5 3 1 1 1 2 5 1 2 2 4 1 4 3 1 1 1 3 3 5 3 5 1 2 5 5 1 4 5 1 2 1 5 4 1
[638] 1 1 4 4 4 5 5 5 5 4 3 4 1 2 1 2 1 1 1 3 4 1 4 3 3 1 4 1 2 3 5 1 3 1 3 5 5 4 4 4 4 2 1 2 5 1 4 2
[687] 4 4 4 2 1 4 3 3 4 5 4 1 5 1 3 2 3 2 5 1 3 1 4 5 5 1 2 4 2 4 4 5 1 5 2 1 4 2 3 2 1 5 1 2 1 5 5 5 1
[736] 4 5 3 1 4 1 2 5 2 2 5 5 1 4 4 1 5 2 5 5 1 5 1 3 4 1 1 3 2 1 1 4 1 5 5 2 3 2 4 1 1 4 4 3 4 1 5 4 4
[785] 3 5 1 2 4 1 3 1 1 3 1 1 3 4 4 2 3 1 5 1 2 5 2 5 1 2 1 2 3 3 3 4 1 1 5 5 2 3 5 3 1 4 5 5 2 5 1 5
[834] 1 5 2 4 3 1 1 3 4 2 1 3 4 2 5 1 3 4 2 5 1 1 3 3 3 4 2 3 2 5 3 2 4 1 1 2 1 1 2 1 4 5 3 2 5 5 4 1 4
[883] 5 4 4 4 2 5 3 2 4 3 4 3 5 2 5 2 4 1 3 2 4 5 2 4 2 1 4 4 2 4 2 5 2 4 4 3 3 5 1 5 2 5 4 1 5 4 5 2
[932] 3 4 5 3 3 5 3 4 3 1 4 1 2 5 3 5 2 4 1 1 5 1 2 5 3 1 1 2 5 3 5 5 4 5 2 4 1 5 4 3 1 5 3 2 3 3 4 4 1
[981] 3 2 1 3 1 1 4 1 2 5 5 3 4 4 1 5 5 2 3

Within cluster sum of squares by cluster:
[1] 1814.957 3629.725 1733.614 1265.951 2421.731
(between_SS / total_SS = 96.0 %)

Available components:
[1] "cluster"      "centers"      "totss"      "withinss"      "tot.withinss" "betweenss"
[7] "size"        "iter"        "ifault"
```

Figure 14. K-Means Cluster

The clusters analyzed in this study were to determine the K value for product_id and unit_price. Based on the analysis results using Posit cloud software with the K-Means clustering method, there were 5 clusters with values of 227, 147, 202, 193, and 230. Then the author displayed a visualization of the cluster plot display. This display can be seen in the image below.



Gambar 15. Cluster plot

4.6 Customer Segmentation

After using the PCA process, the clusters that will be formed are made simpler, this process can be seen in the image below.

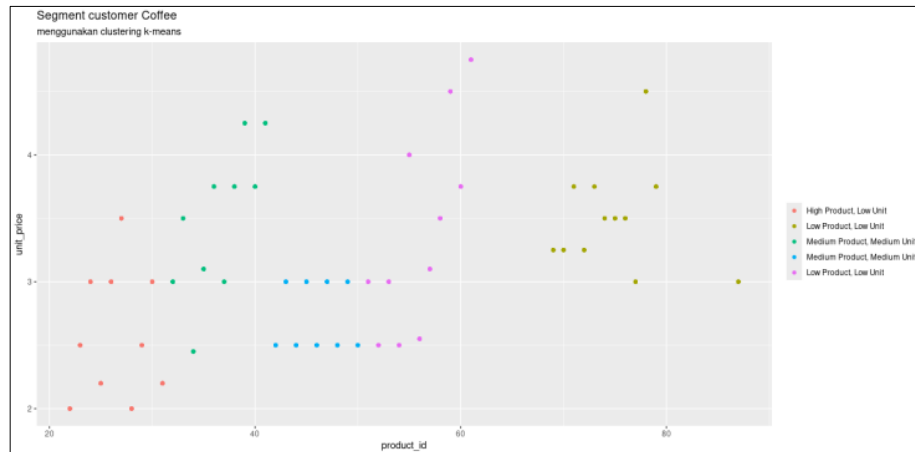


Figure 16. Coffee Customer Segmentation

From the image above, it shows cluster 1 Low product low unit, cluster 4 Low product low unit, cluster 2 Medium product medium unit, cluster 3 Medium product medium unit, cluster 5 High product low unit.

5. CONCLUSION

By applying the K-Means algorithm and PCA analysis, coffee shop owners can understand customer interest in the products sold at each coffee shop outlet so they can make the right decisions regarding sales and product orders. Through this research, it can be seen that coffee is the most frequently purchased beverage by customers compared to other products sold at coffee shops, followed by tea, bakery, and drinking chocolate. Then, the best time for coffee shop visitors is from 12:00 to 17:00.

The author acknowledges that this study has limitations—specifically, that the data used is publicly available data obtained from Kaggle—but hopes that the research will provide new insights into customer segmentation and serve as a reference for future studies employing similar analysis methods and models.

ACKNOWLEDGMENT

The author would like to thank the Editorial Team of the ICIT Journal and Raharja University for giving us the opportunity to have our journal published in the ICIT Journal.

REFERENCES

- [1] M. T. YULIANDRI, " Sejarah Singkat Masuknya Kopi Ke Indonesia," Otten Coffee, 09 Mei 2023. [Online]. Available: <https://ottencoffee.co.id/majalah/sejarah-singkat-masuknya-kopi-ke-indonesia>. [Accessed 25 September 2024].
- [2] D. ANGELIA, "Apa Jenis Minuman Favorit Masyarakat Indonesia di Tahun 2022?," 2022. [Online]. Available: <https://goodstats.id/article/apa-jenis-minuman-favorit-masyarakat-indonesia-di-tahun-2022-s2gWd>. [Accessed 30 SEPTEMBER 2024].

- [3] A. E. DAMAYANTI, B. WIRJATMADI and S. SUMARMI, "Manfaat Konsumsi Kopi dalam Meningkatkan Kemampuan Mengingat (Memori): A Narrative Review," *Media Gizi Kesmas*, vol. 12, no. 1, pp. 463-468, 2023.
- [4] A. AMINI and S. AKBAR, "Pengaruh Konsumsi Kopi Terhadap Konsentrasi Belajar Pada Mahasiswa Fakultas Teknik," *Ibnu Sina: Jurnal Kedokteran dan Kesehatan-Fakultas Kedokteran Universitas Islam Sumatera Utara*, vol. 23, no. 2, pp. 202-214, 2024.
- [5] E. F. L. AWALINA and W. I. RAHAYU, "Optimalisas iStrategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *JATI*, vol. 13, no. 2, pp. 122-137, 2023.
- [6] S. A. PERDANA, S. F. FLORENTIN and A. SANTOSO, "Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift," *SEBATIK*, vol. 26, no. 2, pp. 446-457, 2022.
- [7] S. DEWI and M. A. I. PAKERANG, "Implementasi Principal Component Analysis Pada K-Means Untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang," *JUPI*, vol. 8, no. 4, pp. 1186-1195, 2023.
- [8] D. N. ZHAO, B. WU and X. YI, "Generalized PCA Method and Its Application in Uncertainty Reasoning," *IEEE*, vol. XX, no. XX, pp. 1-9, 2019.
- [9] D. HEDIYATI and I. M. SUARTANA, "Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro," *JIEET*, vol. 5, no. 2, pp. 49-54, 2021.
- [10] A. WIBOWO and A. R. HANDOKO, "Segmentasi Pelanggan Ritel Produk Farmasi Obat Menggunakan Metode Data Mining Klasterisasi Dengan Analisis Recency Frequency Monetary (RFM) TERMODIFIKASI," *JTIK*, vol. 7, no. 3, pp. 573-580, 2020.
- [11] K. Z. WIJAYA, A. DJUNAIDI and F. MAHANANTO, "Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Analisis RFM di Ova Gaming E-Sports Arena Kediri," *TEKNIK ITS*, vol. 10, no. 2, pp. 230-237, 2021.
- [12] M. ARHAMI and M. NASIR, *Data Mining Algoritma dan Implementasi*, YOGYAKARTA: ANDI OFFEST, 2020.
- [13] R. M. SARI, *Metode Clustering Untuk Penyakit Berbahaya*, SUMATERA BARAT: SERASI MEDIA TEKNOLOGI, 2024.
- [14] C. PRIANTO and S. BUNYAMIN, *Panduan Pembuatan Aplikasi Clustering Gangguan Jaringan Menggunakan Metode Kmeans Clustering*, BANDUNG: INFORMATICS RESEARCH CENTER, 2020.