Contents lists available at https://citedness.com/index.php/jdsi

Data Science Insights

Journal Page is available to https://citedness.com/index.php/jdsi

Clustering Coffee Sales Data using the k-Means Algorithm

Edbert Kevin

Institute of Business and Technologhy Pelita Indonesia, Pekanbaru, Riau
email: edbert.kevin@student.pelitaindonesia.ac.id

ARTICLE INFO

Article history:
Received 14 December 2025
Revised 15 January 2025
Accepted 31 January 2025
Available online: 15 February 2025

Keywords:
Coffee
Caffeine
Health
Sustainability
Global Industry

Please cite this article in IEEE style as:
E. Kevin, "Implementation of k-Means Algorithm for Coffee Sales Classification", Data Science Insights, vol. 3, no. 1, pp. 27–34.

ABSTRACT

Coffee is one of the most widely consumed beverages worldwide, with a rich history spanning centuries. Coffee is derived from the beans of the Coffea species, primarily Coffea arabica and Coffea canephora (robusta), and is prized not only for its stimulating effects but also for its complex flavor profile. This paper examines the diverse roles of coffee in human culture, its impact on health, and the global coffee industry. Coffee contains bioactive compounds, including caffeine, antioxidants, and diterpenes, which have been studied for their potential health benefits, such as improved cognitive function and reduced risk of certain chronic diseases. However, excessive consumption can lead to negative effects, including sleep disturbances and cardiovascular problems. In addition, the environmental and social impacts of coffee cultivation, including issues related to sustainability, fair trade, and climate change, are critically examined. The paper concludes with a discussion of emerging trends in coffee research, including innovations in processing methods, the rise of specialty coffees, and the growing importance of ethical sourcing in an increasingly globalized market. This comprehensive review emphasizes the need for a balanced understanding of coffee's benefits and challenges, highlighting its role as a cultural staple and a commodity in the global economy.

Correspondence:
Edbert Kevin
Institute of Business and Technologhy
Pelita Indonesia, Pekanbaru, Riau

1. Introduction

Coffee is a major tropical commodity traded worldwide, contributing half of total tropical commodity exports. Based on USDA (2016) and ICO (2018), Based on data from the International Coffee Organization (ICO), the downward trend in the price of each coffee occurred in the period 2011-2019. The decline in prices has made farmers who have large areas of land tend to store coffee beans and resell them when prices have increased. Coffee plants (Coffea L.) are plantation crops that have been cultivated for a long time. The types of coffee that are often cultivated are Arabica coffee and Robusta coffee [1][2][3][4].

The most widely known types of coffee are Arabica and Robusta coffee. Robusta coffee plants in several studies have shown that they are quite resistant to disease attacks, and have a more bitter, slightly sour taste and contain higher levels of caffeine than Arabica coffee. Robusta coffee has a higher caffeine content than Arabica coffee. Robusta coffee beans have a nutty aroma before being roasted. Arabica coffee has a predominantly sour taste rather than bitter. Arabica coffee has a citrus, fruity aroma[5][6].

Coffee is a plantation commodity whose role in the national economy is very important, the six contributions of coffee commodities to the national economy are: (1) As a source of foreign exchange for the country, (2) Farmer income, (3) Job creation, (4) Regional development, (5) Driver of agribusiness and agroindustry, (6) Supporter of environmental conservation.[7].

In America, around 100 million people consume coffee every day, while people around the world are estimated to consume more than 2.25 billion cups of coffee every day. Coffee consumption is done as a form

of hobby. In addition, consuming coffee can reduce drowsiness and eliminate fatigue, and is the main source of food ingredients that contain lots of antioxidants. The antioxidant content in coffee includes Chlorogenic acid (CGA) as the main phenolic compound in coffee with a fairly high concentration of all plant elements. Caffeine in coffee is considered to improve mood, increase concentration, reduce drowsiness, and improve cognitive function [8].

Caffeine in coffee is known to have benefits when consumed by humans and also has adverse effects on the body if consumed at certain body conditions and in high levels of caffeine. Consuming caffeine is useful for increasing alertness, eliminating drowsiness and improving mood. Caffeine also helps physical performance by increasing endurance and increasing muscle contractions. Excessive caffeine consumption can cause tooth discoloration, bad breath, increased stress and blood pressure if consumed in the morning, insomnia, heart attacks, strokes, male infertility, digestive disorders, addiction and even premature aging [9].

2. Research Methodology

The research method used is clustering analysis with the k-means algorithm to determine the best-selling product types in coffee sales.

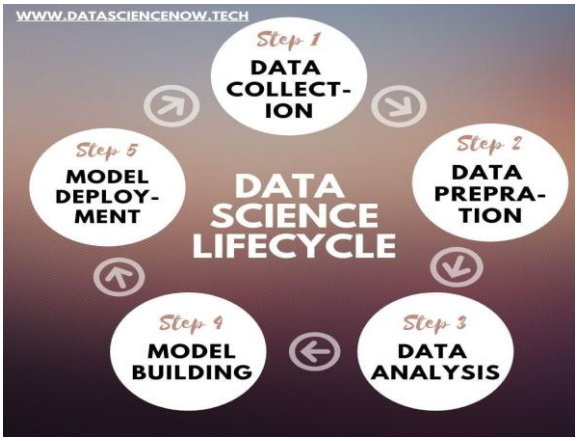

Figure 1. Process of Data Science

Cluster analysis is a data mining method for grouping data or objects based on existing information to describe the relationships between these objects. Clustering is a data analysis method to solve data grouping problems. One of the clustering methods is K-Means. K-Means is one of the algorithms in data mining that can be used to group/cluster data. K-Means has the ability to group data in large quantities with relatively fast and efficient computing time. The K-Means++ algorithm is used in the initialization process of the initial cluster center value (centroids), this is to reduce the instability that occurs in K-Means so that it provides stable and better clustering results. However, clustering results with K-Means are very dependent on the initial cluster center. Clustering results with the K-Means method are good if the determination of the cluster center is correct [10][11][12][13][14].
The stages in conducting Data Science Research are (Figure 1):

2.1 Data Collection

Data Collection is a step where relevant data is collected according to the case to be studied from reliable sources. This data can come from various sources, such as internal databases, external websites, or even social media. It is important to choose reliable and high-quality data sources to ensure the accuracy of the analysis results. Therefore, this study will use data obtained from Kaggle.com[15].

2.2 Data Preparation

Data Preparation is a step where data cleaning is carried out to obtain data that is suitable for processing. This cleaning aims to ensure that the data obtained is relevant and valid data without any errors/loss of data so that the results of the next step will have good accuracy.

2.3 Data Analysis

After the data is cleaned, the next step is to explore the data to understand its characteristics and patterns. This can be done using various statistical techniques and data visualization. The goal of data exploration is to gain a better understanding of the data and identify potential true and precise relationships.

2.4 Model Building

Model Building is the process of developing statistical or machine learning models that can be used to predict or explain outcomes. The model is trained on a subset of data known as the training set, and then evaluated on another subset of data known as the test set. It can typically be used in applications such as Rapid Miner, Tableau, Gogglecolab and so on.

2.5 Model Deployment

The final step in the data science process is interpreting the model results and communicating them to stakeholders. This involves understanding what was learned from the model and how these findings can be used to make better decisions.

3. Results of Research Methods

Here are the steps:

3.1 Data Collection

In this initial step, data will be collected by searching for an available dataset, such as the following example (Figure 2-3), which is taken from a website called Kaggle.com.

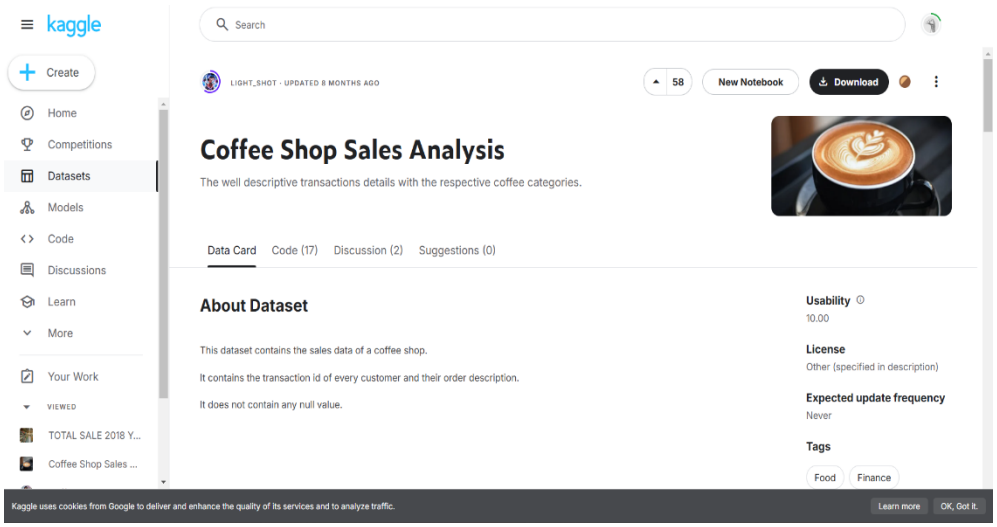

Figure 2. Data from Web Kaggle

So the data taken comes from Kaggle.com, resulting in the following data:

114301	1/6/2023	11:33:29	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Thursday	11	6	3
115405	2/6/2023	11:18:24	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Friday	11	6	4
115478	2/6/2023	12:02:45	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Friday	12	6	4
116288	2/6/2023	19:39:47	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Friday	19	6	4
116714	3/6/2023	12:24:57	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Saturday	12	6	5
116739	3/6/2023	12:44:17	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Saturday	12	6	5
116853	3/6/2023	13:58:14	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Saturday	13	6	5
117389	3/6/2023	18:57:29	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Saturday	18	6	5
118056	4/6/2023	14:31:43	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Sunday	14	6	6
118281	4/6/2023	16:30:33	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Sunday	16	6	6
118428	4/6/2023	17:53:30	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Sunday	17	6	6
118913	5/6/2023	12:01:03	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Monday	12	6	0
119196	5/6/2023	14:19:59	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Monday	14	6	0
119240	5/6/2023	14:42:16	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Monday	14	6	0
119351	5/6/2023	15:51:35	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Monday	15	6	0
119444	5/6/2023	16:52:40	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Monday	16	6	0
119692	5/6/2023	19:51:09	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Monday	19	6	0
121279	7/6/2023	10:16:40	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Wednesd	10	6	2
121416	7/6/2023	10:59:19	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Wednesd	10	6	2
121493	7/6/2023	12:01:18	3	Astoria	45	1	3	3	Tea	Brewed herbal tea	Peppermint	Large	June	Wednesd	12	6	2

Picture 3. Data in file csv

For the labeled section, the product_category section was selected because it shows what type of coffee drink will be classified, as listed in Figure 4:

A	B	C	D	E
tore_location	transaction_qt	product_category	product_type	product_detail
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
.ower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend

Figure 4. Data that would be selected as Label

3.2 Data Preparation

From this step, it is continued by preparing the data that has been taken from the previous step which can be seen from the following image

After the data is taken from the Kaggle website. Figure 5 shows the data:

transactio	transactio	transactio	store_id	store_loca	product_ic	transactio	unit_price	Total_Bill	product_c	product_ty	product_d	Size	Month Na	Day Name	Hour	Month	Day of We
114301	1/6/2023	11:33:29	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Thursday	11	6	3
115405	2/6/2023	11:18:24	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Friday	11	6	4
115478	2/6/2023	12:02:45	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Friday	12	6	4
116288	2/6/2023	19:39:47	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Friday	19	6	4
116714	3/6/2023	12:24:57	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Saturday	12	6	5
116739	3/6/2023	12:44:17	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Saturday	12	6	5
116853	3/6/2023	13:58:14	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Saturday	13	6	5
117389	3/6/2023	18:57:29	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Saturday	18	6	5
118056	4/6/2023	14:31:43	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Sunday	14	6	6
118281	4/6/2023	16:30:33	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Sunday	16	6	6
118428	4/6/2023	17:53:30	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Sunday	17	6	6
118913	5/6/2023	12:01:03	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Monday	12	6	0
119196	5/6/2023	14:19:59	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Monday	14	6	0
119240	5/6/2023	14:42:16	3	Astoria	45	1	3	3	Tea	Brewed he	Peppermin	Large	June	Monday	14	6	0

Figure 5. Data in file excel

The data consists of several components such as (1) Transaction_id, as a unique sign in the form of a code in the transaction, (2) Transaction_date, as the transaction date and time, (3) Transaction_tiem, as the transaction time, (4) Store_id, as the store identifier, (5) Store_Location, as the location of the store where it is sold, (6) Product_id, as the product ID sold, (7) Transaction_qty, as the amount in the transaction, (8) Unit_Price, as the price of the product, (9) Total_bill, as the total number of orders in the transaction, (10) Product_category, as the product name, (11) Product_detail, as the details of the product, (12) Size, as the product size, (13) Month name, as the month name, (14) Day name, as the day name, (15) Hour, as the hour sign, (16) Month, as the month sign, (17) Day of week, as the day sign. to what date in the week.

3.3 Data Analysis

After the Data Preparation Step, it is continued to Data Analysis which is used for the coffee shop data, the format is csv so it is changed into excel format first so that it will be easier to process and view the display. In this case, the attributes used are only store_location, product_category, product_type, product_detail, and quantity. Because the initial goal is to focus on determining the best-selling type of coffee drink product. Figure 6 displays the csv data that has been changed to excel and the data that has been prepared:

	A	B	C	D	E
1	store_location	transaction_qt	product_category	product_type	product_detail
2	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
3	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
4	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
5	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
6	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
7	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
8	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
9	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
10	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
11	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
12	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
13	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
14	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
15	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
16	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
17	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
18	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
19	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
20	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
21	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
22	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
23	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
24	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
25	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
26	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
27	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
28	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend
29	Lower Manhattan	1	Coffee	Drip coffee	Our Old Time Diner Blend

Figure 6. Result from Data Exploration

3.4 Data Modeling

In this step, open an application called Rapid Miner, where after opening the application, create a new file and it will be displayed on an empty layer, as Figure 7:

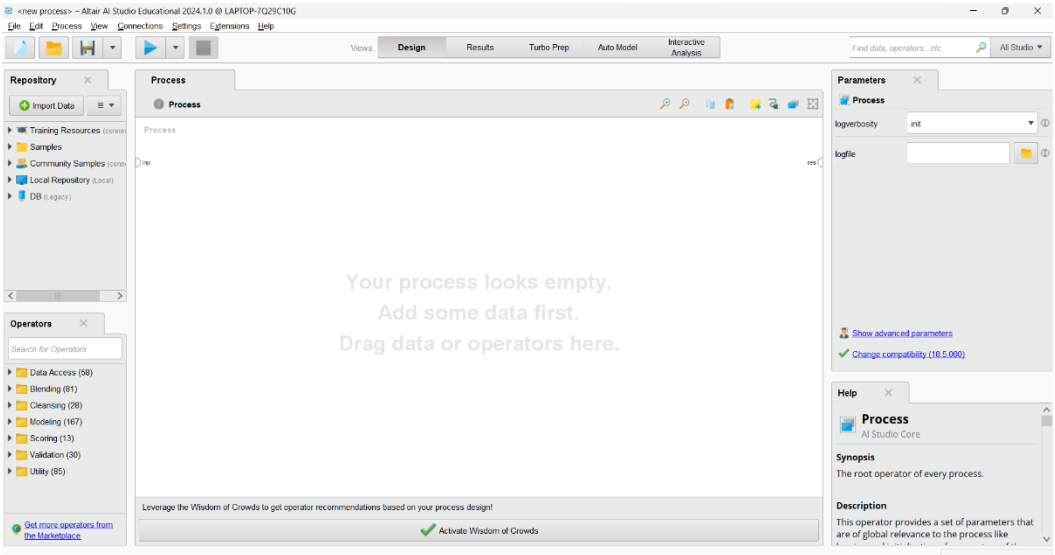

Figure 7. Early View of Rapid Miner

After that, create a processing form for the K Means Algorithm by typing in the Operators section, use the search bar and type the input read excel file, it can be csv or xlsx, here I choose xlsx, then drag the operator menu to a blank sheet, then the input section will appear and don't forget to enter the results of the analysis data that has been done in the previous step where the product category section is used as a label by changing its role, as follows(figure 8):

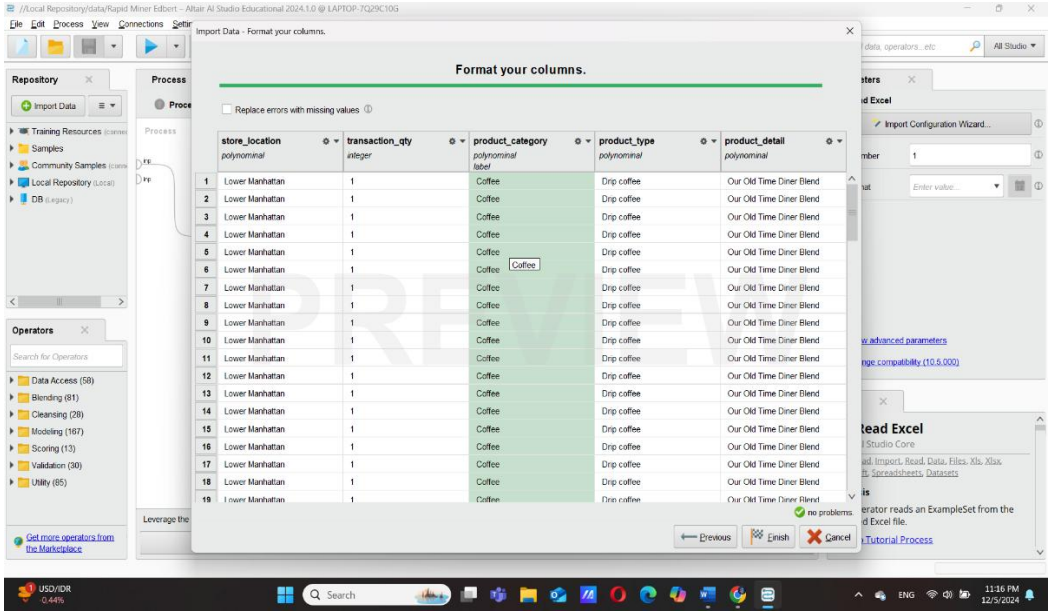

Figure 8. Part of Data that wanted to be Labeled

After that, click finish and also type Kmeans clustter to 2, then type again in the search bar in the operator section and type the name K-Means Clusttering. Then, drag and place it on the right side of the read excel button, followed by typing Performance and drag the position to the right of it K-Means Clusttering. Then, connect the read excel data to Clustering then connect it again to performance and connect it again to the last input, as in the following Figure 9:

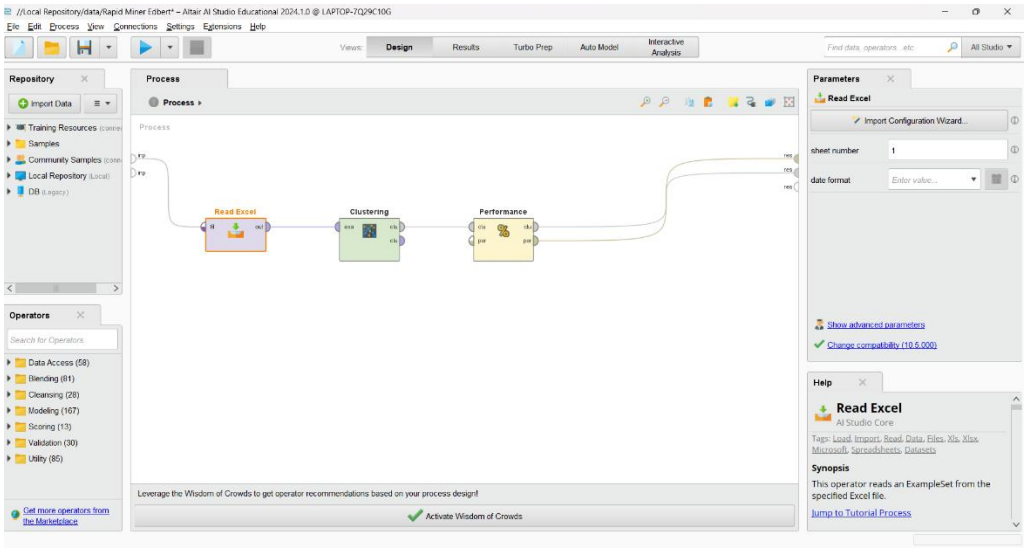

Figure 9. The Input on K-Means process

3.5 Data interperet

Based on the results of the implementation of the k-Means algorithm model using RapidMiner previously conducted, the outcomes are as follows:

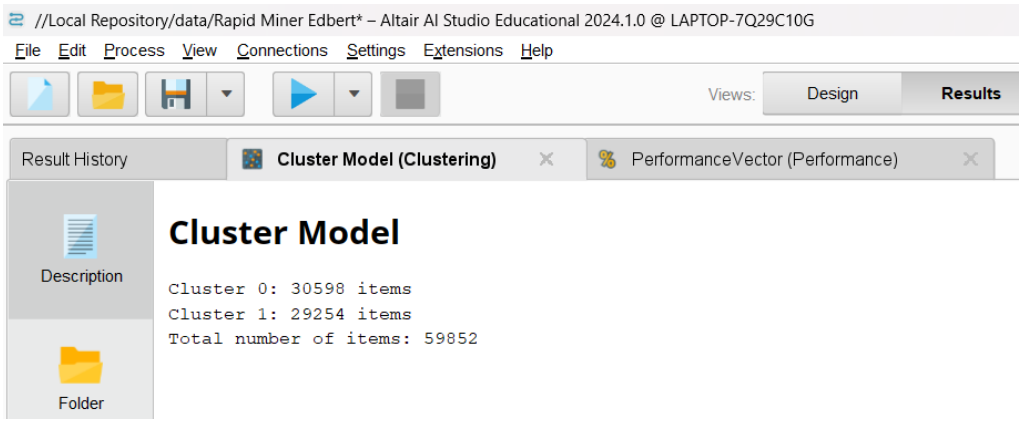

Figure 10. Result of Analysis K-Means

Based on the image (Figure 10), it can be seen that cluster 0 with a total of 30,598 products is categorized as not selling well and cluster 1 with a total of 29,254 products is categorized as selling well, with a total data of 59,852.

4. Summary

It can be concluded that Coffee is a plantation commodity whose role in the national economy is very important, six contributions of coffee commodities to the national economy where this data concerns coffee sales from 3 stores abroad where the attributes used are only store_location, product_category, product_type, product_detail, and quantity. Then, for the analysis method using K-Means which gets the results in the form of cluster 0 with a total of 30598 products categorized as not selling well and cluster 1 with a total of 29254 products categorized as selling well, with a total data of 59852.

References

[1] B. Rahardjo, R. Hasbullah, and F. M. Taqi, "Coffee Shop Business Model Analysis," *Integr. J. Bus. Econ.*, vol. 3, no. 2, p. 140, 2019, doi: 10.33019/ijbe.v3i2.153.

[2] Y. A. N. Aini and T. I. Noor, "Strategi Pemasaran Usaha Kedai Kopi Dalam Menghadapi Situasi Pandemi Covid-19 (Studi Kasus Pada The Celcius Coffee, Cianjur)," *J. Ilm. Mhs. AGROINFO GALUH*, vol. 9, no. 3, pp. 1021–1033, 2022.

[3] B. & Rondius, "No TitleФормирование парадигмальной теории региональной экономики," *Экономика Региона*, pp. 1–11, 2012.

[4] A. Muhlis and S. -, "Analisis Daya Saing Kopi Indonesia Di Pasar Internasional," *Agribios*, vol. 21, no. 1, p. 25, 2023, doi: 10.36841/agribios.v21i1.2807.

[5] D. Budi, W. Mushollaeni, Y. Yusianto, and A. Rahmawati, "Karakterisasi Kopi Bubuk Robusta (coffea canephora) Tulungrejo Terfermentasi dengan Ragi Saccharomyces Cerevisiae," *J. Agroindustri*, vol. 10, no. 2, pp. 129–138, 2020, doi: 10.31186/j.agroindustri.10.2.129-138.

[6] A. Kinasih, S. Winarsih, and E. A. Saati, "Karakteristik Sensori Kopi Arabica Dan Robusta Menggunakan Teknik Brewing Berbeda," *J. Teknol. Pangan dan Has. Pertan.*, vol. 16, no. 2, p. 12, 2021, doi: 10.26623/jtphp.v16i2.4545.

[7] A. Awaluddin, N. Nuraeni, and M. Ilsan, "Analisis Keberlanjutan Usahatani Kopi Arabika Bawakareng Kecamatan Sinjai Barat Kabupaten Sinjai," *AGROTEK J. Ilm. Ilmu Pertan.*, vol. 2, no. 2, pp. 73–84, 2019, doi: 10.33096/agrotek.v2i2.63.

[8] A. E. Damayanti, B. Wirjatmadi, and S. Sumarmi, "Benefits of Coffee Consumption in Improving the Ability to Remember (Memory): A Narrative Review," *Media Gizi Kesmas*, vol. 12, no. 1, pp. 463–468, 2023, doi: 10.20473/mgk.v12i1.2023.463-468.

[9] A. I. Latunra, E. Johannes, B. Mulihardianti, and O. Sumule, "Analisis kandungan kafein kopi (Coffea arabica) pada tingkat kematangan berbeda menggunakan spektrofotometer UV-Vis," *J. Ilmu dan Alama*, vol. 12, no. 1, pp. 45–50, 2021, [Online]. Available: https://journal.unhas.ac.id/index.php/jai2

[10] R. Rahmati and A. W. Wijayanto, "Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 73–80, 2021.

[11] W. Mega, "Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita," *J. Inform.*, vol. 15, no. 2, pp. 160–174, 2015.

[12] L. Rahmawati, S. Widya Sihwi, and E. Suryani, "Analisa Clustering Menggunakan Metode K-

Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)," *J. Teknol. Inf. ITSmart*, vol. 3, no. 2, p. 66, 2016, doi: 10.20961/its.v3i2.654.

[13] A. R. Jannah, D. Arifianto, and M. Kom, "Penerapan Metode Clustering dengan Algoritma K-Means untuk Prediksi Kelulusan Mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember," *J. Manaj. Sist. Inf. dan Teknol.*, vol. 1, no. 1210651237, pp. 1–10, 2015.

[14] C. Ramdani and N. Safadila, "Analisis Data Akademis dengan Menerapkan Algoritme K-Means dan K-Means++," *LEDGER J. Inform. Inf. Technol.*, vol. 1, no. 4, pp. 155–160, 2022.

[15] Y. R. Pratama and A. Siswanto, "Data Science Insights," vol. 2, no. 1, pp. 1–8, 2024.