Rina Fitriana

CUSTOMER SEGMENTATION WITH K-MEANS ALGORITHM AND BUSINESS STRATEGY BUSINESS INTELLIGENCE IN VE...

Teknik Industri

Document Details

Submission ID

trn:oid:::3618:127973022

Submission Date

Feb 11, 2026, 9:08 AM GMT+7

Download Date

Feb 11, 2026, 9:23 AM GMT+7

File Name

13. CUSTOMER SEGMENTATION WITH K-MEANS ALGORITHM AND BUSINESS STRATEGY BUSINESS ....pdf

File Size

2.1 MB

18 Pages

6,939 Words

38,449 Characters

3% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Small Matches (less than 10 words)

Exclusions

- 2 Excluded Sources
- 14 Excluded Matches

Match Groups

7 Not Cited or Quoted 2%
Matches with neither in-text citation nor quotation marks

5 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
2% Publications
2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **7** Not Cited or Quoted 2%
  Matches with neither in-text citation nor quotation marks
- **5** Missing Quotations 1%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
2% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	Udayana University on 2018-08-08	<1%
2	Publication	Frista Millenia Trisuciana, Deden Witarsyah, Edi Sutoyo, Jose Manuel Machado. "C...	<1%
3	Student papers	International Centre for Education on 2024-04-08	<1%
4	Publication	Desfiana Suci Rachmahwati, Rachmadita Andreswari, Faqih Hamami. "Customer ...	<1%
5	Publication	U Andayani, S Efendi, N N U Siregar, M F Syahputra. "Determination System for H...	<1%
6	Student papers	University of Reading on 2025-01-10	<1%
7	Internet	moam.info	<1%
8	Publication	Maila D.H. Rahiem. "Towards Resilient Societies: The Synergy of Religion, Educati...	<1%
9	Publication	Syahroni Hidayat, Budi Sunarko, Uswatun Hasanah. "chapter 7 K-Means Clusterin...	<1%
10	Student papers	Victoria University on 2025-09-14	<1%

11 Internet

www.knack.com **<1%**

Jurnal Teknologi Industri Pertanian 35 (2): 118-135, August 2025
DOI: https://doi.org/10.24961/j.tek.ind.pert.2025.35.2.118
ISSN: 0216-3160 EISSN: 2252-3901
Terakreditasi Peringkat 2
SK Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 158/E/KPT/2021
Tersedia online http://journal.ipb.ac.id/index.php/jurnaltin

CUSTOMER SEGMENTATION WITH K-MEANS ALGORITHM AND BUSINESS STRATEGY BUSINESS INTELLIGENCE IN VEGETABLE ONLINE RETAILING

SEGMENTASI PELANGGAN DENGAN ALGORITMA K-MEANS DAN STRATEGI BISNIS INTELIJEN BISNIS PADA RETAIL ONLINE SAYURAN

Nurochman[1)], Dedy Sugiarto[1,2)], and Rina Fitriana[1,3)]

[1)]Master of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti
[2)] Study Program of Information Systems, Faculty of Industrial Technology, Universitas Trisakti
[3)]Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti
Jl. Kyai Tapa No.1, Grogol, Jakarta, Indonesia
*Email: rinaf@trisakti.ac.id

Paper: Received December 24, 2024; Revised May 15, 2025; Accepted July 11, 2025

ABSTRAK

Sebagian besar UMKM masih memiliki kendala untuk tumbuh dan berkembang ke level bisnis. Penerapan sistem business intelligence diharapkan dapat membantu dalam pengambilan keputusan bisnis yang tepat dan cepat sehingga UMKM dapat tumbuh dan berkembang. Penelitian ini bertujuan untuk menentukan segmentasi pelanggan berdasarkan pengelompokan produk yang diminati oleh konsumen. Selain itu, penelitian ini bertujuan untuk mengetahui manfaat business intelligence dalam memberikan informasi kinerja bisnis untuk mengambil keputusan. Penelitian ini menggunakan algoritma k-means untuk clustering. Business intelligence menggunakan perangkat lunak Power BI untuk visualisasi. Berdasarkan hasil analisis pengelompokan produk dengan algoritma k-means, jumlah cluster yang optimal adalah 2 (k=2). Penentuan nilai k=2 menggunakan rata-rata jarak centroid sebesar 121.624.275.127, dan validasi nilai DBI minimum = 0,052. Dua segmen konsumen berdasarkan hasil clustering adalah cluster 0 (28%) dan cluster 1 (72%). Insight pada dashboard penjualan adalah fluktuasi penjualan harian, dominasi produk tertentu yang diminati, dan produk yang penjualannya rendah. Inisiatif strategi untuk jangka panjang adalah segmentasi pelanggan untuk promo yang lebih personal, fokus pada langganan & repeat order, mengoptimalkan pemasaran digital dan menggunakan analisis prediktif untuk memperkirakan tren penjualan. Pada dasbor produksi, pesanan dan stok ditemukan informasi seperti produksi harian cenderung melebihi pesanan, yang menyebabkan kelebihan stok, sementara pesanan berfluktuasi secara tidak konsisten. Tantangan utamanya adalah produksi dan permintaan yang tidak seimbang, kelebihan stok pada produk tertentu, pesanan yang tidak stabil, dan produk yang kurang diproduksi.

Kata Kunci: analitik data, algoritma k-means, intelihensia bisnis, usaha mikro kecil menengah (UMKM),

ABSTRACT

Most MSMEs still have obstacles to growing and developing at the business level. Applying a business intelligence system is expected to assist in making appropriate and quick business decisions so that MSMEs can grow and develop. This research aimed to determine customer segmentation based on product clustering that consumers demand. In addition, this study aims to determine the benefits of business intelligence in providing business performance information to make decisions. This research uses the k-means algorithm for clustering. Business intelligence uses Power BI software for visualisation. Based on the results of analysing product clustering with the k-means algorithm, the optimal number of clusters is 2 (k = 2). Determination of the value of k = 2 uses an average centroid distance of 121,624,275,127, and validation of the minimum DBI value = 0.052. Based on the clustering results, cluster 0 (28%) and cluster 1 (72%) are two consumer segments. Insights on the sales dashboard are daily sales fluctuations, the dominance of certain products in demand, and products with low sales. Strategy initiatives for the long term are customer segmentation for more personalized promos, focus on subscriptions and repeat orders, optimising digital marketing, and the use of predictive analytics to forecast sales trends. On the dashboard of production, order, and stock, information, such as daily production tends to exceed orders, leading to overstock, while orders fluctuate inconsistently. The key challenges are unbalanced production and demand, overstock on certain products, unstable orders, and underproduced products.

Keywords: business intelligence, data analytics, k-means algorithm, Micro Small Medium Enterprise

INTRODUCTION

Indonesian Micro Small Medium Enterprise (MSMEs) can absorb 97% of the workforce, contribute 57% of gross domestic product (GDP), and contribute 15% to national exports (Muhamad, 2023). Digital transformation can help MSMEs compete in the industrial era 4.0. Applying digital technology and data analysis can improve the operational performance (KPI) of MSMEs (Masi, 2024). Mastery

3

*Coressponding Author

of technology can help scale MSME businesses (Herliana, 2015).

The implementation of digitalization in MSMEs is difficult because they have limited financial resources, time, little knowledge and experience of digitalization (Gouveiaa and Mamede, 2022). Most MSMEs use digitalization only at the marketing stage by creating online catalogs or entering their products in the marketplace (Duque *et al.*, 2022). Many MSMEs refrain from implementing business intelligence because there are few references to this solution model. Further research needs to be done on the role of business intelligence on marketing performance, finance, and human resource management in specific business fields such as culinary or others (Fitriyana and Mahendrawathi 2019). On the other hand, research needs to be done on developing a reference model of business intelligence components for MSMEs (Shobana *et al.*, 2023). Studies on the application of business intelligence and data analysis in MSMEs need to be carried out (Llave, 2019). In the current era, MSMEs must be able to monitor business and use resources efficiently and effectively, especially information resources so that they can survive in changing economic conditions (Tutuneaa, 2012). MSMEs get the benefits of business intelligence in processing and summarizing data but are less able to utilize the knowledge of the benefits of business intelligence. Therefore, it is necessary to expand the study of the utilization of business intelligence for MSMEs (Tsiakiris, 2015).

The development of business intelligence based on data in MSMEs is important in the era of the Industrial Revolution 4.0 to increase productivity and global competitiveness in the MSME sector, which plays a role in Indonesia's economic development. Data analytics can provide deep insights into consumer behavior related to preferences, shopping habits, and market trends. With this, MSMEs can adjust marketing strategies, optimize product stock, and satisfy customer experience. The development of business intelligence in MSMEs can manage their business more precisely with valid and real-time data, thus increasing their competitive advantage (Hadi, 2020). Business intelligence and data analysis are one of the bases for making business decisions (Roy and Cortesi, 2022)

Business intelligence has been applied in various industries, such as retail, insurance, banking, and manufacturing (Fitriana *et al.*, 2012). The application of business intelligence with Power BI in MSMEs can provide information on sales transactions, sales details per unit, contribution per product, and product per category (Kumar *et al.*, 2023). Business intelligence has the function of collecting, storing, accessing, and analyzing company data. This function can be done with automation to be very efficient. The application of business intelligence can increase competitive advantage and benefit the company because it can transform raw data into future benefits (Fitriana *et al.*, 2017). Business intelligence can help businesses manage and optimize in-store inventory, manage tight profit margins, analyze detailed historical customer data such as purchase patterns and customer demographics (Malik, 2018).

Sayur Express, as MSME, uses the Accurate application for recording financial transactions, financial reporting, and inventory management. Sayur Express sells products through marketplaces such as Shopee and Tokopedia. Sayur Express sells food products with proportions of fresh vegetables 52%, side dishes 17%, fresh vegetable packages 11%, fruit 6%, seasonings 6%, the remaining fresh fish, groceries, chicken, and frozen below 5%.

Sayur Express needs to know its customers' behavior, especially those who consume vegetable products. Sayur Express needs to know daily production, stock, and sales. This information is needed for strategic business decisions related to the number of purchases and inventory (stock) that align with the business (target of sales) in Sayur Express. Therefore, a study of data analysis and intelligent business is carried out to provide that information.

RESEARCH AND METHOD

This research is done focusing on vegetable products as objects in Sayur Express. Business intelligence applications use Power BI to create dashboard reporting for visualization. Data analytics with the k-means algorithm is used for consumer clustering.

Data Analytics

K-means is one of the non-hierarchical clustering methods that groups data in the form of one or more clusters/groups (Indraputra and Fitriana, 2020). This algorithm groups data in the same object through a series of iterative clustering processes until a perfect cluster is formed. It begins by determining the number of clusters, and then each data point will be allocated to a cluster based on the closest distance of the centroid to a particular cluster (Naik. 2023).

The data processed is online order. The first stage, data preparation, is determining the dataset used for processing according to the object of research, namely vegetables purchased by or sent to customers. The second stage, Pre-processing, is filtering, replacement, deleting, renaming, transforming, or outlier handling the attribute of the dataset before processed modeling by k-means. Google Collab and RapidMiner were used for pre-processing and cluster modeling. The third stage, implementation, makes the clustering model using the k-means algorithm.

The elbow method and Davies Bouldin Index (DBI) can be used to validity test for clustering results. The elbow method is a method used to determine the best number of clusters by looking at the percentage of comparison results between the number of clusters that will form an elbow at a point. If the value of the first cluster and the value of the second cluster give an angle in the graph or the value decreases the most, then the cluster value is the best. Davies Bouldin Index (BDI) is one of the approaches to determine the optimal cluster value. This approach aims to maximize the distance between clusters (high heterogeneity) and, at the same time, try to minimize the distance between objects in a cluster (high homogeneity) at the same time trying to minimize the distance between objects in a cluster. The best number of clusters has the minimum DBI value (Satriawan *et al.*, 2021).

Business Intelligence

As an indication of system performance, the BI system can help to make decisions on how to increase sales and the number of customers. Designing a business intelligence model can be done through several stages: system requirements analysis, identification of data and information requirements, data warehouse design, data warehouse development, data mining process, data visualization, and system performance evaluation (Hidayat and Fitriana, 2022). According to Vercellis, designing a business intelligence model consists of 4 stages: analysis, design, planning, implementation and control (Vercellis, 2009). Figure 1 shows flow research methodology that consists of analysis, design, planning, and implementation.

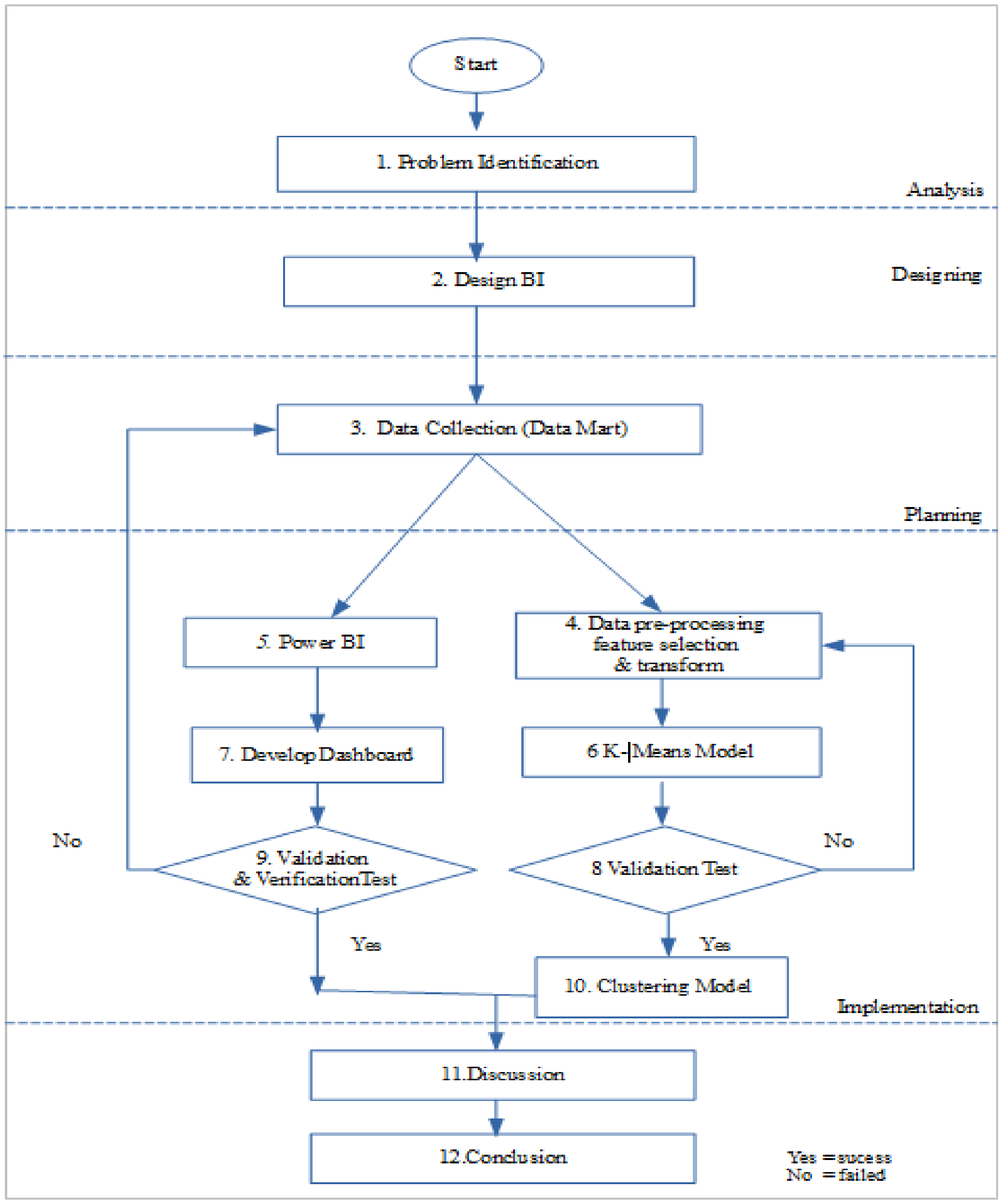


Figure 1. Research stage

Analysis Stage

The analysis stage is problem identification. The organization's need for a business intelligence system was carefully identified. Information was extracted and collected by interviewing employees across departments. Business problems were formulated to identify the business requirements. The objectives, benefits, and priorities of the business intelligence system were clearly defined. Business intelligence design should be based on analyzing business/user needs (Sulaiman and Yahaya, 2013)

Design Stage

In the design stage, the existing information systems infrastructure should be assessed. Critical decision-making processes supported by business intelligence systems should be examined and tested to determine adequate information needs. Figure 2. is the design of the business intelligence model that will be developed at Sayur Express.

Planning Stage

In the planning stage, the functions of the business intelligence system must be defined and described in detail. The data and data sources used are assessed. The architecture of the business intelligence system includes the data warehouse and data marts. A data warehouse is a place to store information from various sources, in a certain scheme, and usually in one place (Sulaiman and Yahaya, 2013). Define data and mathematical models and ensure data is available for each model. This stage defines the requirements, and data is collected and analyzed. The algorithms used are verified for their efficiency in solving the problem.

An integrated, clean, and consistent data warehouse is used for data analysis. The form of data representation in the data warehouse is a multidimensional model (cube) that allows viewing data from different perspectives with an online analytical processing (OLAP) approach (Kovacic *et al.*, 2022). If the data elements have been designed, the ETL process is carried out to convert the data from the initial form to another pre-designed form. Kettle ETL (Pentaho) is used for ETL processes that are capable of manipulating data from multiple sources (Sugiarto *et al.*, 2021).

Implementation Stage

In the next stage, implementation, a prototype of the business intelligence system is created to determine the gap between the actual requirements and the project specifications. The data is processed with the k- k-means algorithm to create a customer clustering model. Primary data is processed with Power BI to create a dashboard for sales analysis. Power BI was used to create the sales analysis dashboard. The dashboard that has been created will then be tested for validity and verification, and the cluster model is tested with a validation test.

Figure 3 is the architecture of the business intelligence system to be developed. Primary data from the marketplace, namely product sales and customers, is processed with Power BI and the k-mean algorithm.

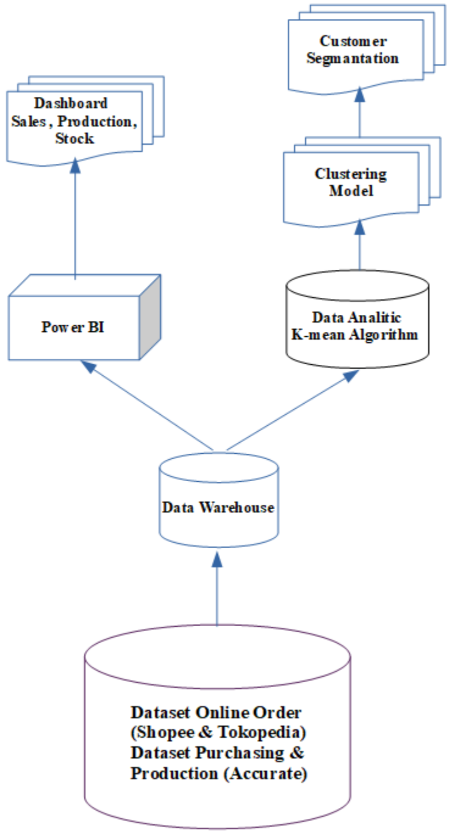


Figure 3. Architecture business intelligence system

RESULT AND DISCUSSION

The flow of business processes at Sayur Express is explained in Figure 4 which starts from preparing a list of vegetable raw material purchases (1) to suppliers with PO and to traditional markets in cash (2). Raw materials received/purchased from suppliers will be received by the Pre-packing team for product processing: filter, sort, and select, then packaged (3). Products will be placed in storage, and production data will be entered into the Accurate system (4). Proof of purchase will be entered into the Accurate system by the Inventory Admin (5). Online orders through the Marketplace will be managed by Admin Data Entry (6) and then forwarded to the Packing and Food prep team to be followed by preparing the products (7). Admin Data Entry will prepare incoming order reports (8). Products that have been packaged will be picked up and delivered by courier (9). The Finance Admin will prepare the daily sales report (10).

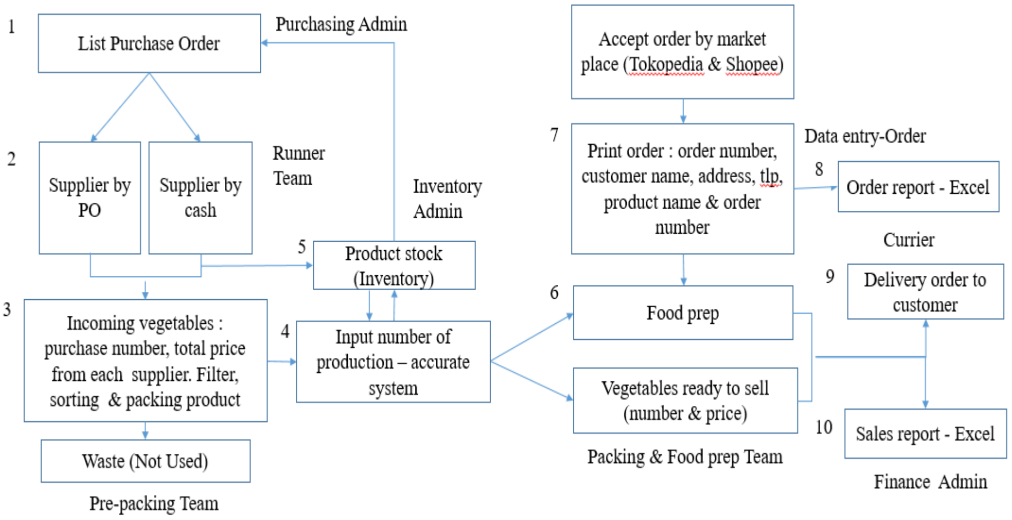


Figure 4. Business process flow in sayur express

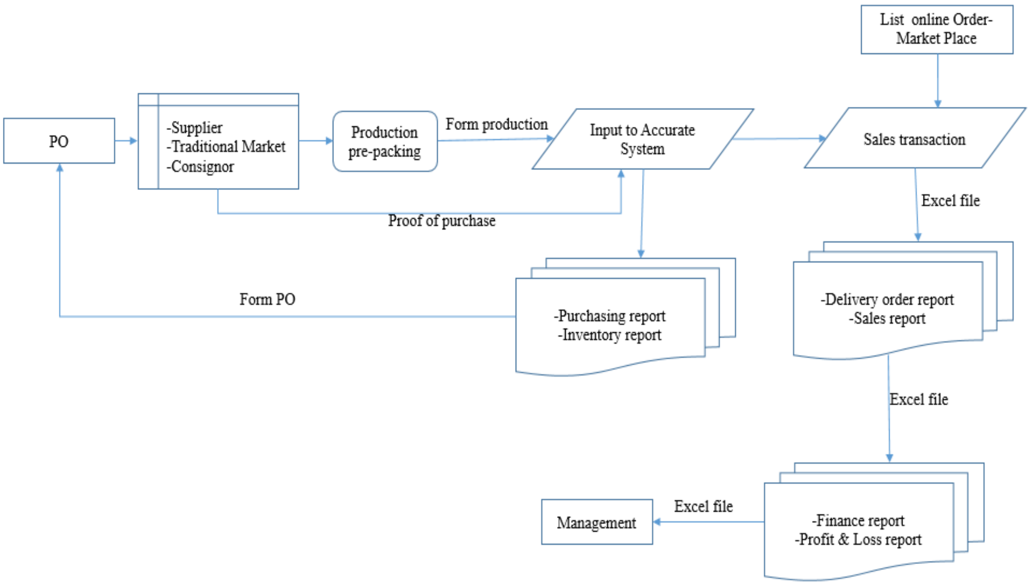


Figure 5. Transaction data flow in sayur express

Figure 5 explains the flow of data processed at Sayur Express starting from the purchase form of vegetable raw materials will be received by the supplier who will then send the purchased raw materials. Proof of purchase from suppliers will be entered in the Accurate system along with production data. The final product inventory report and daily raw material purchase plan are obtained by the Accurate system. Sales transactions will occur according to the online order list from the Marketplace. Reports on incoming orders and sales results will be generated in Excel. Financial and profit-loss reports are generated in Excel.

Data Analytics

Data Preparation

The research used online order data on May 2024. The data has 72533 rows of data with 50 attributes, as shown in Figure 6. The data consists of object, integer, and float types.

In the “Order Status” attribute, orders with the status “Finished” and “On Delivery” are selected as Figure 7. For order status “Canceled” and “Need to be Sent” are removed from the data.

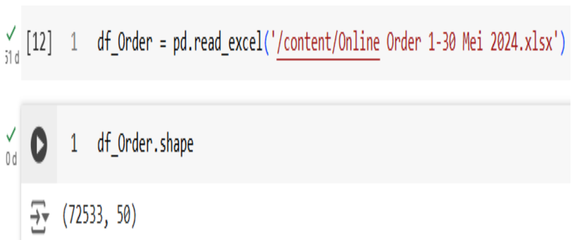
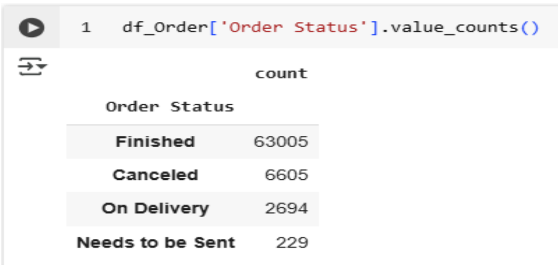


Figure 6. Data Online Order



```
1  df_order['Order Status'].value_counts()
```

Order Status	count
Finished	63005
Canceled	6605
On Delivery	2694
Needs to be Sent	229

Figure 7. Attribute Order Status

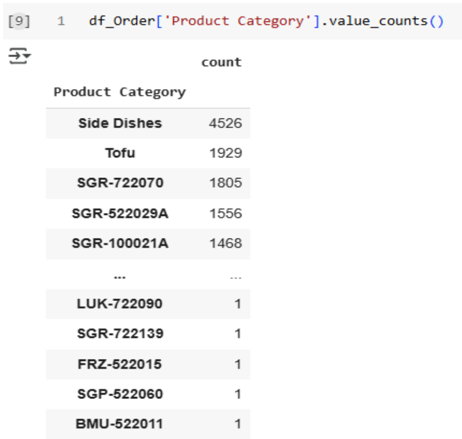
In the "Product Category" attribute, the fresh vegetable category "SGR" and fresh vegetable package "SGP" are selected. Other product categories, such as side dishes, fruits, and seasonings, are removed from the data because they are not the object of research, as shown in Figure 8. The file was saved as df_order2.xlsx.

The results of this data selection obtained data as much as 40916, as shown in Figure 9. Selected data is as dataset.

Data Pre-Processing

In the first pre-processed dataset will be filtered, splitted and replaced for somes attributes and values by RapidMiner as show in Figure 10. In the 2nd stage dataset (40916) is pre-processed with the some replace operators in RapidMiner as shown in Figure 11. K-means is an algorithm that processes numeric type data. Therefore, we selected 11 attributes that are numeric type as Figure 12. The outlier are erased in dataset as Figure 13. In the 2nd stage outlier values will be removed by Google Colab.

The dataset that has been cleaned of outliers has 21,178 rows of integer type data as shown in Figure 14. Next, the dataset will be processed to make clustering model by k-mean algorithm in RapidMine



```
[9]  1  df_order['Product Category'].value_counts()
```

Product Category	count
Side Dishes	4526
Tofu	1929
SGR-722070	1805
SGR-522029A	1556
SGR-100021A	1468
...	...
LUK-722090	1
SGR-722139	1
FRZ-522015	1
SGP-522060	1
BMU-522011	1

Figure 8. Attribute product category
.

```
[15]  1  df_sales = pd.read_excel('/content/df_order2.xlsx')
      1  df_sales.shape
(40916, 12)
```

Figure 9. Number of Dataset

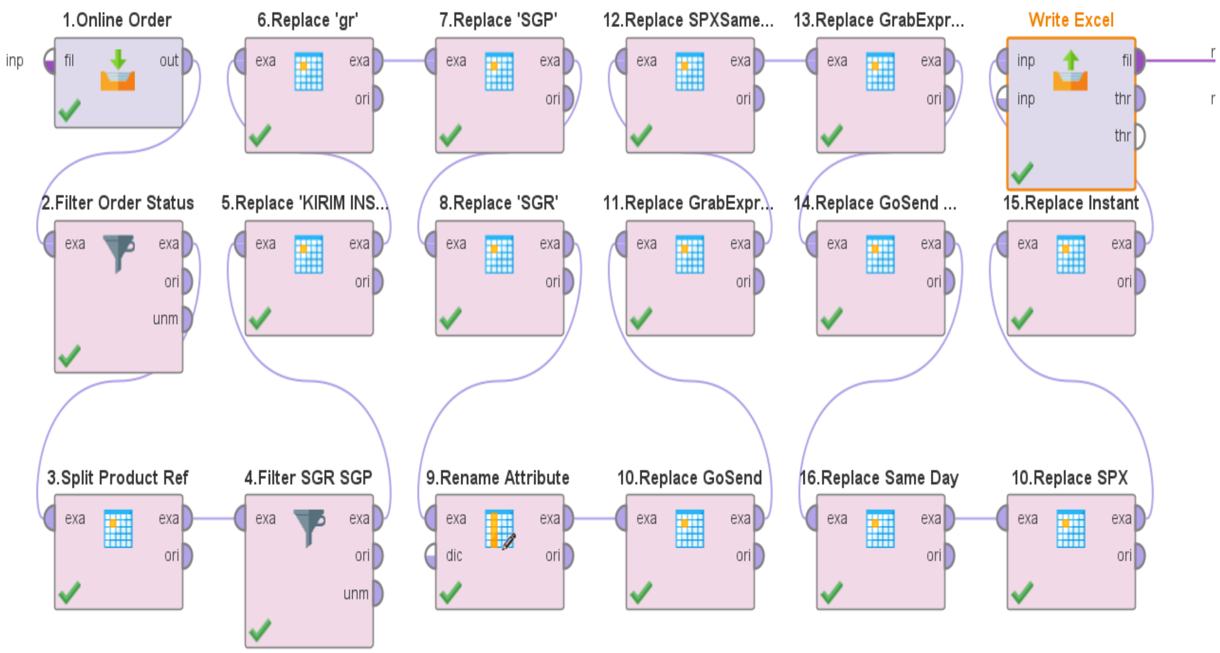


Figure 10. 1st data preprocessing

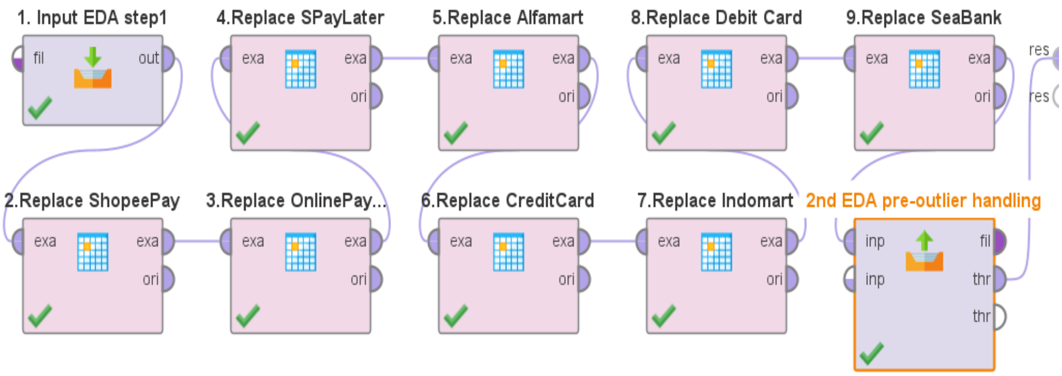


Figure 11. 2nd data preprocessing

```
[15] 1 df_sales = pd.read_excel('/content/df_order2.xlsx')

1 df_sales.shape
(40916, 12)

[ ] 1 df_sales.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40916 entries, 0 to 40915
Data columns (total 12 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Payment Method              40916 non-null  object
 1   Delivery Metodh             40916 non-null  int64
 2   Category                    40916 non-null  int64
 3   Product Unit                40916 non-null  int64
 4   Order Volume (Unit)         40916 non-null  int64
 5   Product Price               40916 non-null  int64
 6   Dicount Price               40916 non-null  int64
 7   Number                      40916 non-null  int64
 8   Total Product Price         40916 non-null  int64
 9   Order Quantity              40916 non-null  int64
 10  Total Payment               40916 non-null  int64
 11  Estimation - Delivery Cost  40916 non-null  int64
dtypes: int64(11), object(1)
memory usage: 3.7+ MB

[ ] 1 feature_columns = ['Delivery Metodh','Category','Number','Product Unit','Order Volume (Unit)','

[ ] 1 df_sales1 = df_sales[feature_columns]
```

Figure 12. Selected atribute for handling outlier

```
outlier_index =[] # defines an empty list to store the index of outliers
for i in (df_sales1.columns): # creates a loop for each column in the dataframe
 if (df_sales1[i].dtypes in ['float64']):
  print(i,':',df_sales1[i].dtypes) # print the column name and data type

 Q1 = df_sales1[i].quantile(0.25)
 print('Q1', Q1)

 Q3 = df_sales1[i].quantile(0.75)
 print('Q3', Q3)

 IQR = Q3-Q1
 print('IQR', IQR)

 nilai_min = df_sales1[i].min()
 nilai_max = df_sales1[i].max()

 min_IQR = Q1 - 1.5 * IQR
 max_IQR = Q3 + 1.5 * IQR

 # Finding outlier
 if (nilai_min < min_IQR):
  print('Low outlier is found <', min_IQR)
  print('Low Outlier Index :', list(df_sales1[df_sales1[i] < min_IQR].index))
 # insert the outliers index into the outlier_index variable
 outlier_index.append(list(df_sales1[df_sales1[i] < min_IQR].index))

 if (nilai_max > max_IQR):
  print('High outlier is found >', max_IQR)
  print('High outlier Index : ', list(df_sales1[df_sales1[i] > max_IQR].index))
 # memasukkan indeks outliers ke variabel outlier_index
 outlier_index.append(list(df_sales1[df_sales1[i] > max_IQR].index))

 print('\n')
```

Figure 13. Code of outlier handling

```
[ ] 1  df_clean = df_sales1.drop(sorted(unique_out_ind), axis=0)
    2  # mereset indeks dataframe setelah menghapus outliers
    3  df_clean.reset_index(drop=True, inplace=True)
    4  df_clean.head()
```

	Delivery Metodh	Category	Number	Product Unit	Order Volume (Unit)
0	7	1	1	100	1250
1	7	1	1	250	1250
2	7	1	1	250	2665
3	7	1	1	500	2665
4	7	1	1	175	2665

```
[ ] 1  df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21178 entries, 0 to 21177
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Delivery Metodh               21178 non-null  int64
 1   Category                      21178 non-null  int64
 2   Number                        21178 non-null  int64
 3   Product Unit                  21178 non-null  int64
 4   Order Volume (Unit)           21178 non-null  int64
 5   Order Quantity                21178 non-null  int64
 6   Product Price                 21178 non-null  int64
 7   Dicount Price                 21178 non-null  int64
 8   Total Product Price           21178 non-null  int64
 9   Total Payment                 21178 non-null  int64
 10  Estimation - Delivery Cost    21178 non-null  int64
dtypes: int64(11)
memory usage: 1.8 MB
```

Figure 14. Dataset for clustering

```
1  correlation = df_sales1.corr()
2  plt.figure(figsize=(10,8))
3  sns.heatmap(correlation, annot=True, cmap='coolwarm')
4  plt.show()
```

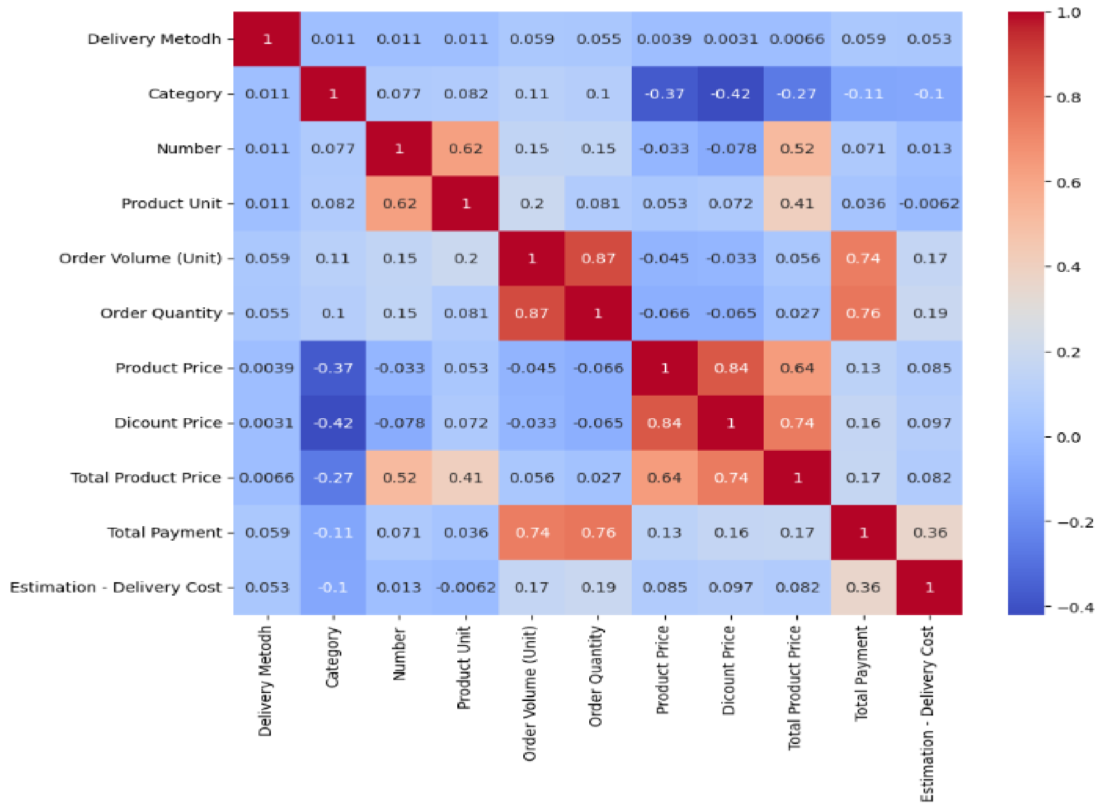


Figure 15. Correlation analysis

Correlation analysis between our attributes we perform. Based on the results of the correlation analysis of the attributes shown in Figure 15, there is a strong relationship between the attributes of total payment with order volume and order quantity, order volume with order quantity, product price with discount price, total price of products purchased with discount price, product price with total price of products purchased.

In this study, we explore clustering analysis related to total payment, product price, discount price, product volume, order volume, and order quantity with a correlation value of more than 0.60, as shown in Table 1.

Clustering Implementation

We have performed product clustering of the data sets with and without outliers. Comparison of 2 research datasets to find the best dataset for product clustering based on the smallest BDI value. Based on the result clustering model by the k-mean algorithm in RapidMiner, the average centroid distance and DBI are obtained as in Table 2 and Tabel 3. For each k = 2, 3, 4, 5, 6, 7, 8, 9, 10. The minimum DBI is 0.052 on k = 2 in Table 4. So, the best k value is 2 from the dataset without an outlier. It means BDI = 0.052 in K=2 has high homogenates in the same cluster.

Table 1. The correlation analysis of the attributes

Attributes	**Total Payment**	**Total Product Price**	**Product Price**	**Discount Price**	**Product Unit (Size)**	**Order Quantity**	**Order VOlume**
Total Payment						0.76	0.74
Total Product Price			0.64	0.74			
Product Price		0.64		0.84			
Discount Price		0.74	0.84				
Product Unit (Size)						0.62	
Order Quantity	0.76				0.62		0.87
Order Volume	0.74					0.87	

Table 2. Average centroid distance & DBI (with outlier)

k	**Average within centroid distance**	**Validity - Davies Bouldin Index (DBI)**
2	348.446.582.833	0.054
3	192.715.581.977	0.056
4	118.296.261.462	0.054
5	85.324.220.977	0.057
6	67.639.293.919	0.060
7	56.315.424.930	0.064
8	49.229.371.993	0.067
9	42.767.199.457	0.067
10	38.558.626.680	0.069

Table 3. Average Centroid Distance & DBI (without outlier)

k	**Average Within Centroid Distance**	**Validity - Davies Bouldin Index (DBI)**
2	121.624.275.127	0.052
3	65.305.713.555	0.055
4	39.455.117.619	0.056
5	27.725.679.474	0.058
6	21.724.594.862	0.061
7	17.545.513.336	0.065
8	15.242.235.603	0.070
9	13.455.931.837	0.072
10	12.356.731.881	0.077

9

Figures 16 and 17 present the results of the elbow method applied to identify the optimal number of clusters (k) for customer segmentation in an online vegetable retail business. This method is different with DBI. The method was executed under two conditions: with and without the presence of outliers in the dataset. The total within-cluster sum of squares (WCSS) is used as the primary metric to evaluate the compactness of clusters. The plot in Figure 16 indicates a sharp decline in WCSS from k=2 (348,446,582,833) to k=4 (85,324,220,977), suggesting that increasing the number of clusters significantly reduces the intra-cluster variance.

However, beyond k=4, the rate of decrease becomes less pronounced, with WCSS values gradually tapering off until k=10 (38,558,626,680). After outlier removal, the WCSS values become significantly lower and more proportionate, as illustrated in Figure 17. The curve shows a marked decrease from k=2 (121,624,275,127) to k=4 (27,725,679,474), after which the reduction in WCSS begins to plateau. The elbow method, particularly when applied after removing outliers, proves to be an effective approach in determining the optimal number of clusters in customer segmentation.

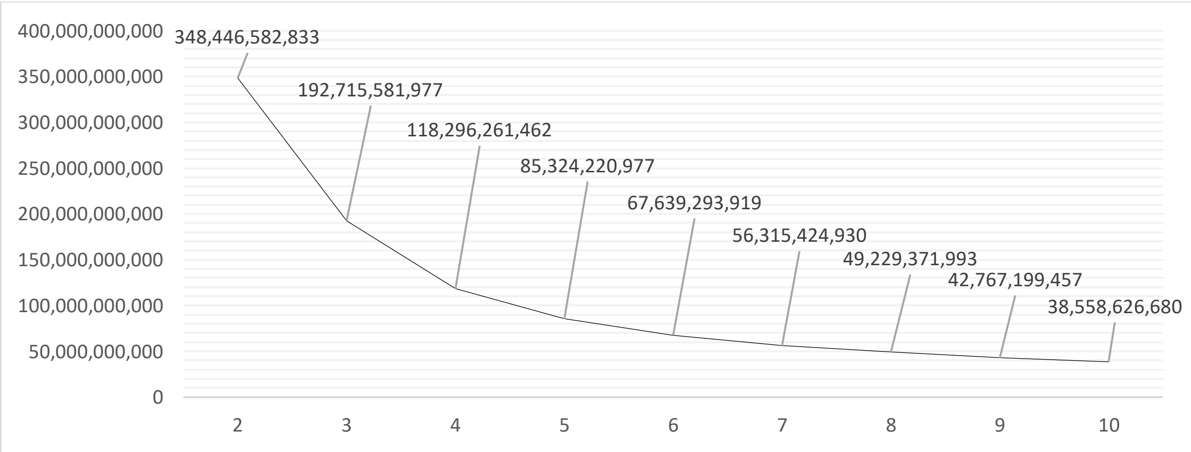


Figure 16. Graph of elbow centroid distance (with outlier)

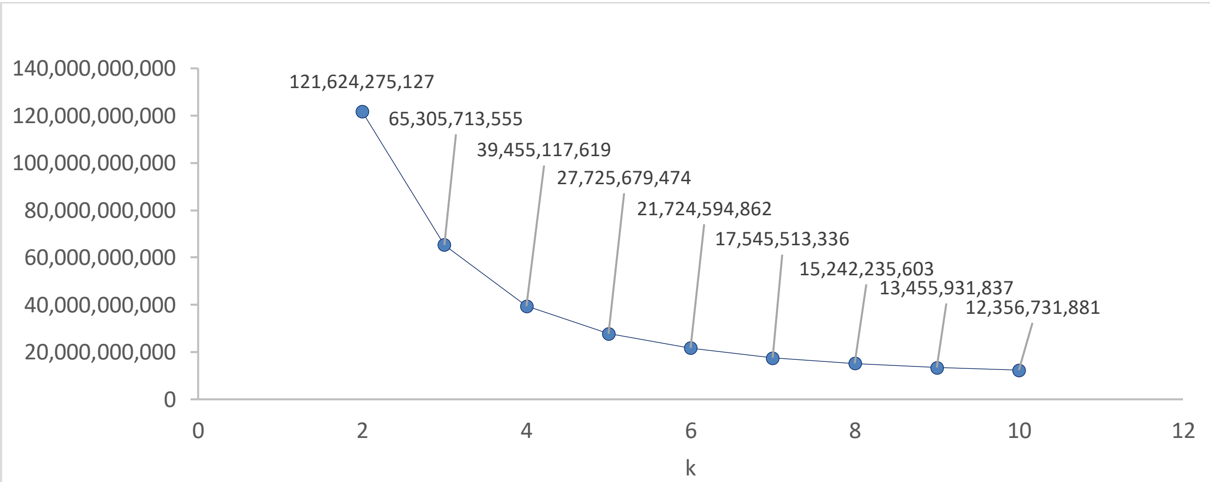


Figure 17. Graph of elbow centroid distance (without outlier)

In k = 2, the product is divided into 2 clusters, as shown in Table 4. Most products are in Cluster 1 (72%), and the rest are in Cluster 0 (28%). By k-mean it will be explored the behavior of each cluster related to consumption patterns in terms of the product price, product size, and number or volume of products purchased.

Table 5. Cluster distribution in k=2

Cluster	Number	%
0	5886	28%
1	15292	72%
Total	21178	

Analysis

Customer segmentation based on the clustering product is shown in Figure 18 -23. Finding the consumer behavior in this clustering:

1. In cluster 0, consumers (28%) shop for vegetables at prices below IDR 20,000 with a spending budget of less than IDR 150,000 per day, as shown in Figure 18.
2. In cluster 1, consumers (72%) shop for vegetables at prices above IDR 20,000 with a spending budget of more than IDR 150,000 per day, as shown in Figure 18
3. In cluster 0 discount price below IDR 18.000 has contribution to attract consumers (28%) shopped product with spending budget above IDR 150,000 as Figure 19.
4. In cluster 1 discount price below IDR 15.000 has contribution to attract consumers (72%) shopped product with spending budget under IDR 150,000 as Figure 19.
5. In cluster 0 number of daily consumption (28%) is about 5 – 30 units with spanding budget above IDR 150,000 as Figure 20.
6. In cluster 1 number of daily consumption (72%) is mostly below 20 units with spending budget below IDR 150,000 as Figure 20.
7. In cluster 0 volume of daily consumption (28%) is about 2 – 8kg with spanding budget above IDR 150,000 as Figure 21.
8. In cluster 1 volume of daily consumption (72%) is mostly below 6 kg with spanding budget below IDR 150,000 as Figure 21.

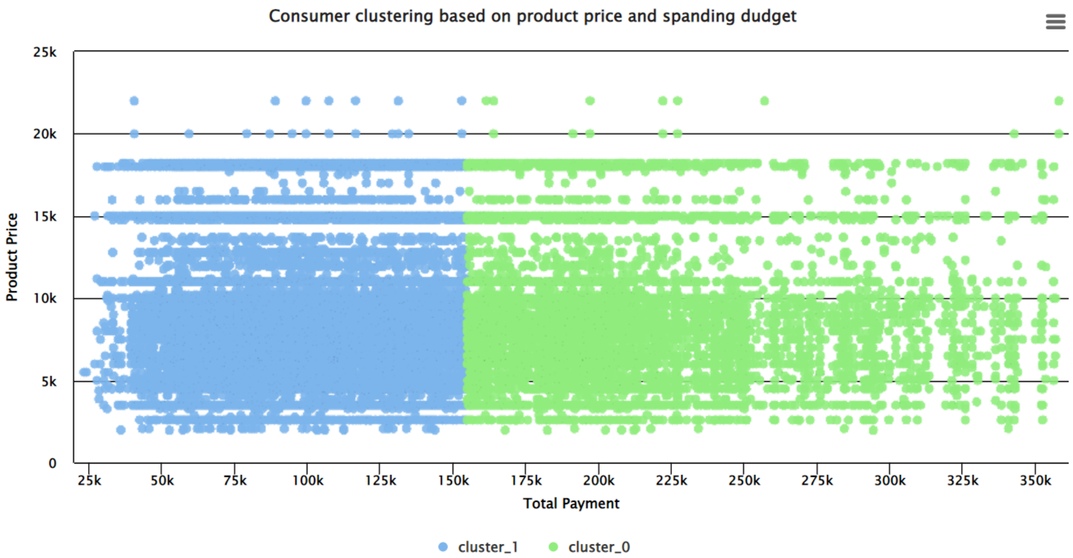


Figure 18. Customer clustering: Product Unit – Total Payment

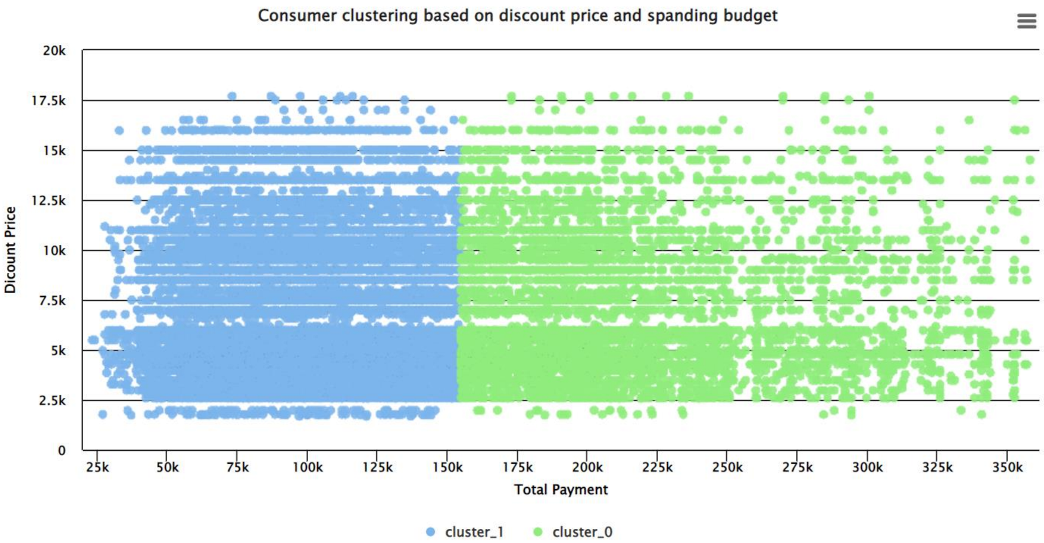


Figure 19. Customer clustering: Discount Price – Total Payment

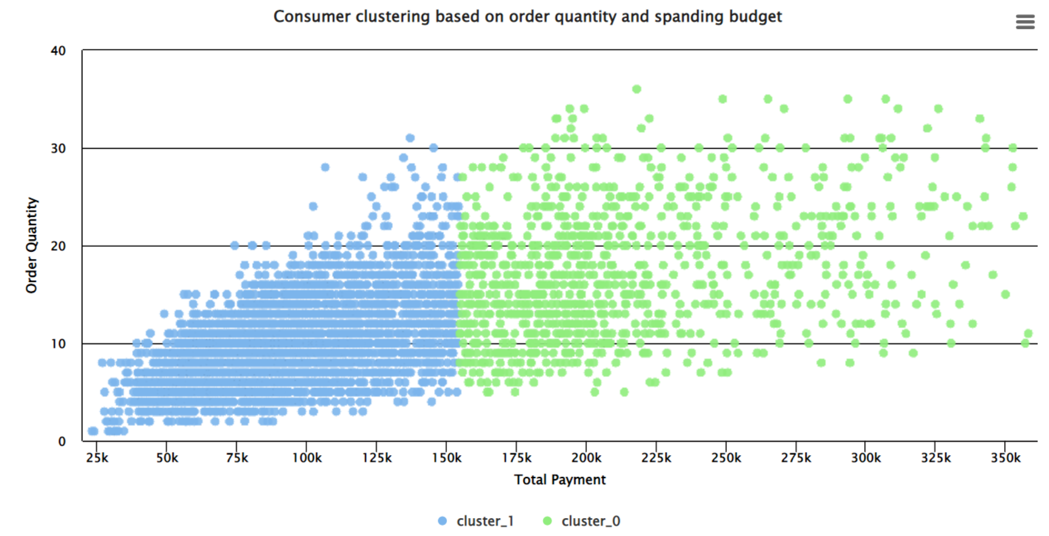


Figure 20. Customer clustering: Order Quantity – Total Payment

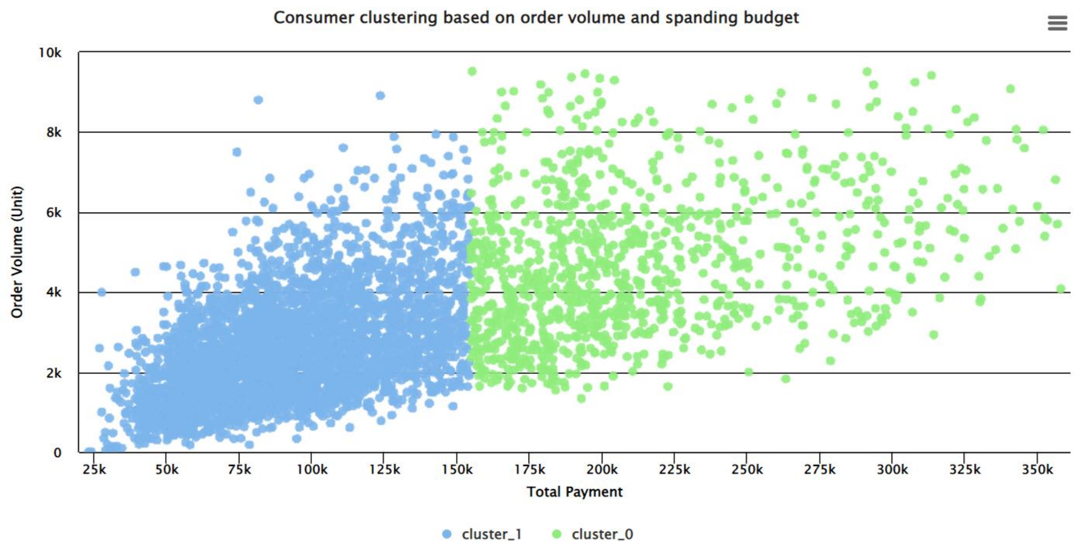


Figure 21. Customer clustering: Order Volume – Total Payment

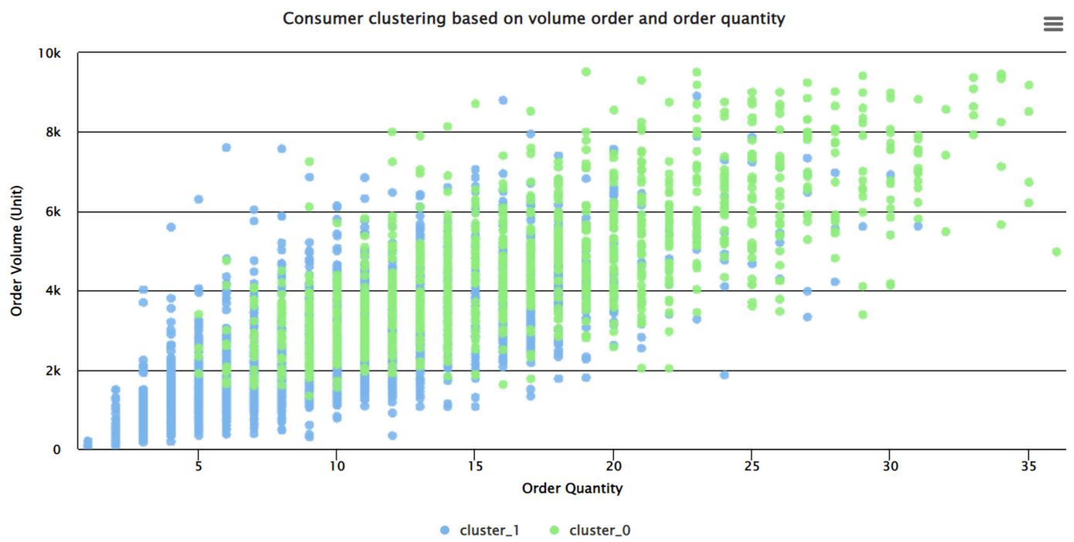


Figure 22. Customer clustering: Order Quantity – Order Volume

9. In cluster 0 daily consumption (28%) is mostly between 5- 30 units and 2-9 kg as Figure 22
10. In cluster 1 daily consumption (72%) is mostly below 20 units and 7 kg as Figure 22
11. In cluster 0 discount price between IDR 1.000 – 16.000 has contribution to attract consumers (28%) shopped product with volume order 2 kg – 10 kg as Figure 23.
12. In cluster 1 discount price between IDR 1.000 – 16.000 has contribution to attract consumers (72%) shopped product with volume order 0.5 kg – 6 kg as Figure 23.
13. In cluster 0 discount price between IDR 2.000 – 16.000 has contribution to attract consumers (28%) shopped product with number of order between 6 -35 units as Figure 24.
14. In cluster 1 discount price between IDR 2.000 – 16.000 has contribution to attract consumers (72%) shopped product with number of order below10 units as Figure 24.

Strategic Initiative for Marketing

Cluster 0 has 28% customer contribution. Customers in this cluster have a larger spending budget and consume more vegetables than in cluster 1. Therefore, it is necessary to acquire new customers to increase sales volume and revenue. An attractive, effective, and digital-based marketing strategy is needed to expand and add new consumers to the online retail vegetable MSME business. Here are some strategies that can be applied.

Increase Brand Awareness

- Optimize social media (Instagram, Facebook, TikTok, WhatsApp Business): Create interesting content such as healthy recipes, tips on storing vegetables, or the benefits of certain vegetables. Use Reels, Stories, and Live Selling features to increase interaction with potential customers. Post customer testimonials to build trust.

Figure 23. Customer clustering: Discount Price – Order Volume

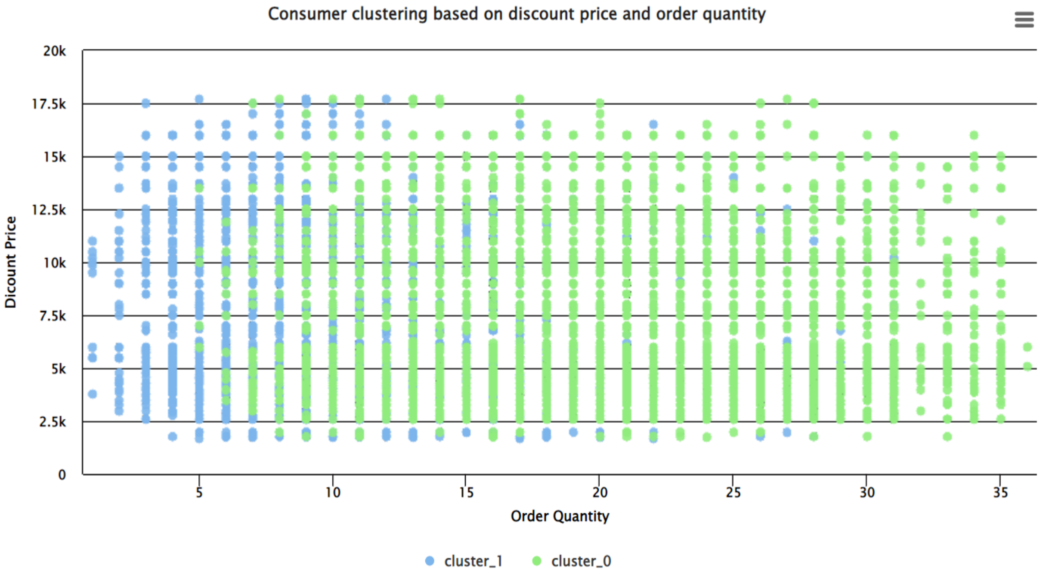


Figure 24. Customer clustering: Discount Price – Order Quantity

- Join Bazaars & Digital Markets: Join offline events such as healthy food bazaars or organic markets. Take advantage of marketplaces such as Shopee, Tokopedia, and Grab Mart to reach more customers.

Offer Attractive Promos & Programs for New Customers

- Discount & Cashback for First Buyers: Provide discounts or cashback for new customers to be interested in trying your product.
- Bundling Packages & Free Shipping: Offer savings packages (for example a complete vegetable package for 1 week).Provide free shipping for certain areas or with a minimum purchase.
- Referral Program: Encourage existing customers to invite friends with bonus discounts if their friends buy.

Increase Ease of Shopping

- Open Various Purchasing Channels: In addition to the marketplace, use WhatsApp Business, Instagram Shop, and online store websites to facilitate transactions.
- Chatbot & Fast Customer Service: Respond quickly to new customer inquiries so they don't hesitate to buy.

Targeted Digital Marketing (Paid Ads & Email Marketing)

- Facebook & Instagram Ads: Use targeted ads to reach potential customers in certain areas.
- Offer attractive promos in ads to increase conversions.
- WhatsApp & Email Marketing: Send attractive offers or exclusive discounts to potential customers who have visited your store.

Targeted Digital Marketing

- Create a WhatsApp or Telegram Group: Create a customer community that discusses healthy living tips, recipes, and special promos.

Cluster 1, with a consumer percentage of 72%, needs to be maintained. For this reason, strategic initiatives are needed to build customer loyalty. To retain consumers of vegetable products in online retail vegetable MSMEs, the main focus should be on customer satisfaction, a comfortable shopping experience, and effective retention strategies. Here are some marketing strategies that can be applied:

1. Maintain Product and Service Quality
   - Make sure the vegetables are always fresh: Implement a good stock management system so that products are always in the best condition when received by customers.
   - Attractive and Eco-friendly Packaging: Use packaging that maintains freshness and is recyclable to attract environmentally conscious customers.
   - Fast and Secure Delivery: Work with a trusted delivery service or use your own courier for certain areas to ensure that the vegetables are not damaged when they reach the customer.
2. Maintain Product and Service Quality Vegetable Membership or Subscription: Offer a subscription program (weekly box) with benefits such as lower prices or additional products.
   - Discount or Cashback for Loyal Customers: Reward customers who shop frequently to increase their loyalty
3. Increased Interaction and Engagement with Customers
   - Educational Content: Create content on social media about the benefits of vegetables, healthy recipes, or tips on how to store vegetables to make them last longer.
   - Live Selling and Customer Testimonials: Use live streaming to introduce products in person and share testimonials from satisfied customers.
   - Responsive Customer Service: Respond quickly to customer queries or complaints to build trust.
4. Personalize Offers and Shopping Experience
   - Product Recommendations Based on Shopping History: Send product recommendations based on customers' shopping habits via WhatsApp or email.
   - Birthday or Special Moments Promo: Give discount vouchers or small gifts when customers have birthdays or on special days.
   - Survey and Feedback: Ask customers for feedback regularly to understand their needs and improve services.
5. Digital Marketing Optimization
   - SEO and Google My Business: Make sure your store is easily found on Google with SEO optimization and listing on Google My Business.
   - Paid Advertising Strategy (Facebook Ads, Instagram Ads, Google Ads): Target ads to audiences who have already purchased so that they come back to shop.
   - Email & WhatsApp Marketing: Send product updates, exclusive promos, or interesting tips to keep interacting with customers.

Business Intelligence

Based on the data visualization in the sales dashboard as Figure 25, information was found:

Obstacles & Solutions

1. Daily Sales Fluctuations

- Unstable daily sales may indicate dependence on certain days or moments. It could be due to promos, payday, or customer spending habits.

Solution:

- Use a more consistent promo strategy (for example, weekly discounts or weekly vegetable pack subscriptions).
- Analyze the factors influencing sales spikes (whether due to specific days, promo campaigns, or other factors).
- Increase social media and WhatsApp engagement to remind customers about the latest promos or stock.

2. Specific Product Domination

The best products, such as the Katuk Bonus Clear Vegetable Foodprep Package, show a positive trend, but other products are still less in demand. Customers may prefer food packs over individual raw materials.

Solution:

- Focus on the best-selling products and increase the variety of food packages according to customer preferences.
- Experiment with bundling less-selling products with more popular products.
- Conduct customer surveys to understand the product preferences they need more.

3. Products with Low Sales

Some products have very little sales contribution, which could be due to a lack of demand or lack of promotion.

Solution:

- Evaluate products with low sales (is it less in demand, too expensive, or not well known to customers?).
- Use flash sale strategies or special discounts to increase sales of these products.
- Experiment with educational content on social media that highlights the benefits and how to use products that are not selling well.

Long-term Strategy Recommendations

- Customer segment based on purchasing patterns: Use customer data to provide more personalized promos.
- Focus on subscriptions & repeat orders: Offer weekly/monthly subscription plans for repeat customers.
- Optimize digital marketing campaigns; Strengthen promotions on social media, marketplaces, and WhatsApp to increase awareness.

11 Use predictive analytics: Predict future sales trends based on historical data. With these strategies, online vegetable retail MSMEs can improve sales stability and product diversification and strengthen customer loyalty.

Figure 26. show behavior Production, Order, and Stock. Information was found :

- Imbalance between Production, Order, and Stock: Daily production tends to be higher than the number of orders, leading to excess stock on some days. Daily orders show fluctuations, with peak production on May 27-30, but orders do not always follow the same trend.
- Settled Stock for Some Products: Broccoli and Green Sawih have the highest stock (888 and 1009 pcs), indicating overstock. Green Cayenne Pepper has the lowest stock (160 pcs), which could be high demand but sub-optimal production. Total Production is greater than total order.

Obstacles & Solutions

1. Unbalanced Production with Demand

More production is done than incoming orders, causing excess stock and potential losses due to unsold or expired products.

Solution:

- Use a demand forecasting-based production system to reduce overstock. Analyze daily order trends and adjust production volumes accordingly.
- Implement a pre-order or vegetable membership system to adjust production according to customer orders.

2. Overstock on Certain Products

Products such as Broccoli and Sawih Hijau have high stock, indicating that these products may be overproduced or in less demand than other products.

Solution:

- Do bundling promos or discounts for products with excess stock to sell faster.
- Check whether products with high stock have a short shelf life and distribute them faster with flash sale promos or to business partners (restaurants, catering).
- Use slow-moving and fast-moving product analysis to avoid overproduction in the future.

3. Unstable Orders, Causing Inconsistent Production

Daily orders show significant fluctuations, making determining the optimal production quantity difficult.

Solution:

- Improve marketing strategies to increase orders consistently (for example, with weekly subscription boxes or incentives for repeat orders).
- Use data-driven marketing campaigns to identify when customers spend more and adjust production based on these patterns.
- Build customer loyalty with a fixed subscription program to make orders more predictable.

4. Some products may be underproduced

Green Cayenne Pepper has a relatively low stock (160 pcs), which may indicate that demand is higher than production. If the stock runs out quickly, customers may switch to competitors. Patterns to adjust production rhythms based on high and low demand trends.

Solution:

- Analyze whether the demand for Green Cayenne Pepper is higher than that produced and adjust the production quantity.
- Ensure smooth procurement of raw materials so that there are no obstacles in production for products with high demand.
- Apply a dynamic pricing system (flexible prices according to demand) to optimize profits when demand is high.

Long-term Strategy Recommendations

- Implement Demand Forecasting: Use historical data to predict demand and efficiently organize production.
- Optimize Inventory Management System: Use the FIFO (First In, First Out) method to sell old stock first.
- A more Aggressive Marketing Strategy for slow-moving products: Focus on high-stock products through promos, bundling, or distribution to other businesses.

Apply Subscription or Pre-Order System for More Stable Orders: Production can be more accurate according to customer needs.

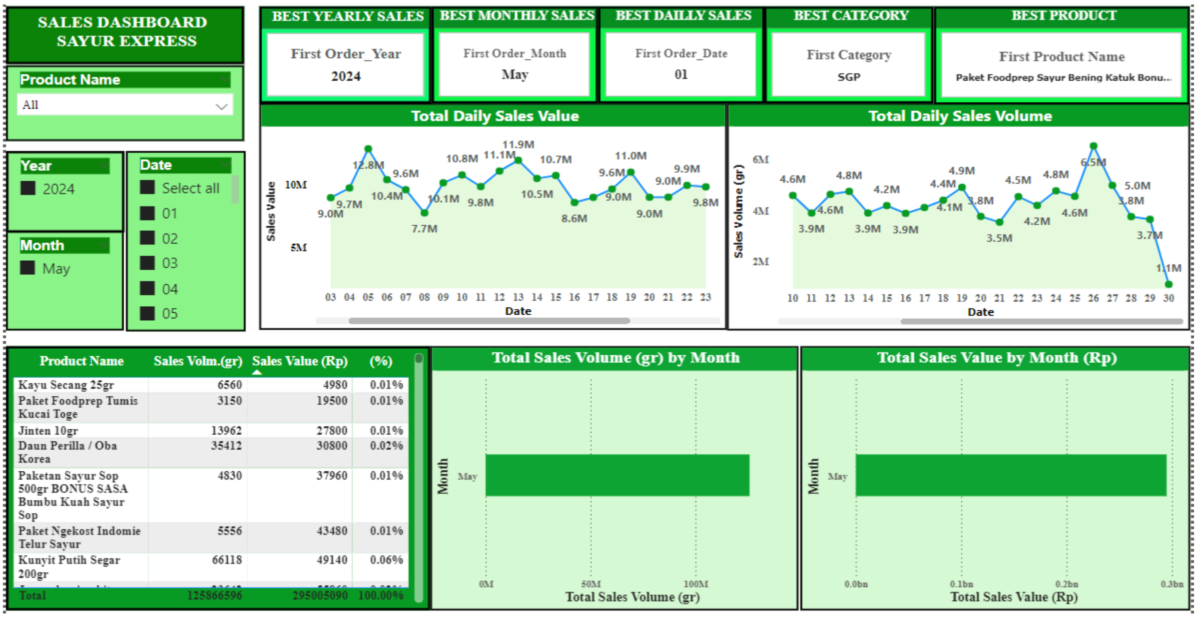


Figure 25. Sales dashboard

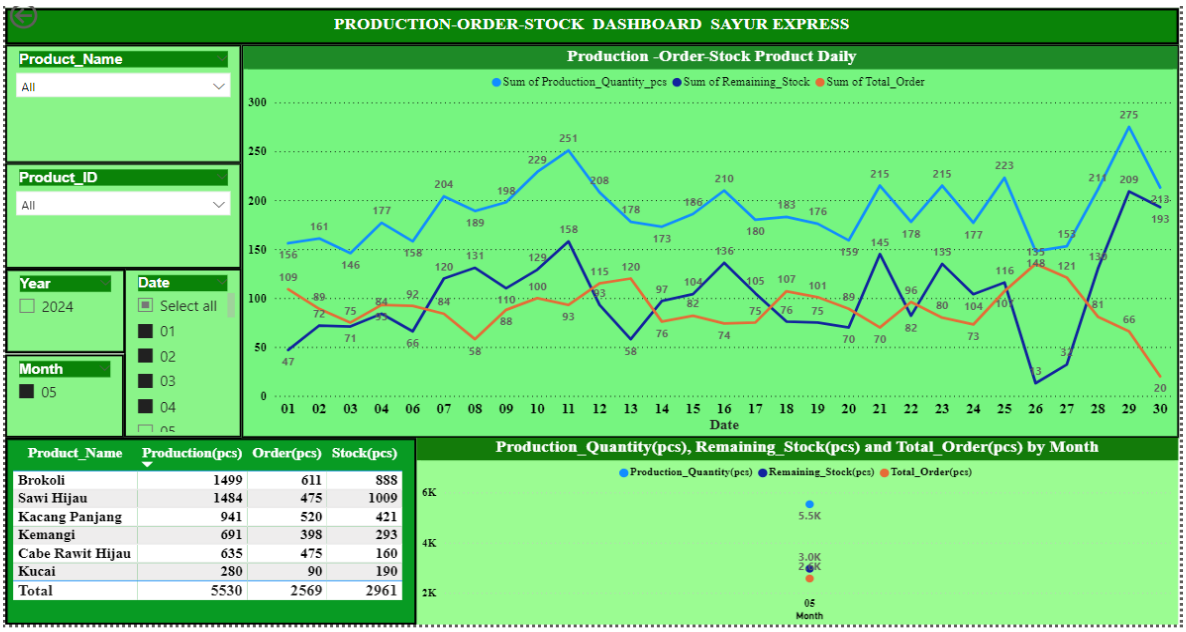


Figure 26. Production-order-stock dashboard

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The clustering results show two customer segments: Cluster 0 (28%): Consumers with bigger budgets and more consumption. Cluster 1 (72%): Consumers with smaller budgets and lower purchase volumes. Recommended strategy: Cluster 0: Focus on new customer acquisition through brand awareness, attractive promos, and digital marketing. Cluster 1: Maintain customer loyalty with quality service, personalized offers, and closer interaction. Sayur Express, as an MSME – online retail company, can apply strategic recommendations in marketing to retain and acquire customers.

Daily sales fluctuations, dominance of some products, and low-selling products were found. Production often exceeds orders, leading to overstocking of some products while other products are understocking. Strategic Recommendation: Optimize digital marketing strategy with more personalized customer segmentation. Use predictive analytics to project sales trends and adjust production accordingly. Implement subscription & repeat order system for demand stability. Improve stock & production management based on daily demand analysis. Applying business intelligence and the k-means algorithm is proven to help MSMEs understand customer behavior, optimize marketing strategies, and improve production and stock efficiency. This enables MSMEs to make smarter, data-driven business decisions for more sustainable growth.

Recommendations

For future research, it is recommended:

1. Using sales data over a longer period of time (more than one month) to identify seasonal patterns.
2. Adding customer demographic or geographic variables for broader segmentation.

REFERENCES

Abai HZ, Yahaya JH, Deraman A. 2013. User requirement analysis in data warehouse design: A Review. *Procedia Technology* 11(Iceei):801–6. doi: 10.1016/j.protcy.2013.12.261.

Duque J, Godinhob A, and Vasconceloscd J. 2022. Knowledge data extraction for business intelligence a design science research approach. *Procedia Computer Science* 204(2021):131–39. doi: 10.1016/j.procs.2022.08.016.

Fitriana R, Saragih J, and Luthfiana N. 2017. Model business intelligence system design of quality products by using data mining in r bakery company. *IOP Conference Series: Materials Science and Engineering* 277(1). doi: 10.1088/1757-899X/277/1/012005.

Fitriana R, Eriyatno, Djatna T, Kusmuljono BS. 2012. Peran sistem intelijensia bisnis dalam manajemen pengelolaan pelanggan dan mutu untuk agroindustri susu skala usaha menengah. *Jurnal Teknologi Industri Pertanian.* 22(3):131–39. doi: https://doi.org/10.24961/j.tek.ind.pert.2020.30.1.110.

Fitriyana D and Mahendrawathi ER. 2019. Business process maturity level of MSMEs in East Java, Indonesia. *Procedia Computer Science* 161:1098–1105. doi: 10.1016/j.procs.2019.11.221.

Gouveia DF and Mamede HS. 2022. Digital transformation for SMES in the retail industry. *Procedia Computer Science* 204(2021):671–81. doi: 10.1016/j.procs.2022.08.081.

Hadi MZ. 2020. Peluang Implementasi teknologi big data dan block chain untuk peningkatan kinerja perdagangan pada sektor UMKM Di Indonesia pada era industri 4.0. *Cendekia Niaga* 3(1):71–80. doi: 10.52391/jcn.v3i1.463.

Herliana S. 2015. Regional innovation cluster for small and medium enterprises (SME): a triple helix concept. *Procedia - Social and Behavioral Sciences* 169(August 2014):151–60. doi: 10.1016/j.sbspro.2015.01.297.

Hidayat MK and Fitriana R. 2022. Penerapan sistem intelijensia bisnis dan k-means clustering untuk memantau produksi tanaman obat. *Jurnal Teknologi Industri Pertanian* 32(2):204–19. doi: 10.24961/j.tek.ind.pert.2022.32.2.204.

Indraputra RA and Fitriana R. 2020. K-Means clustering data COVID-19. *Jurnal Teknik Industri* 10(3):275–82. doi: 10.25105/jti.v10i3.8428.

Kovacic I, Schuetz GC, Neumayr B, Schrefl M. 2022. OLAP patterns: a pattern-based approach to multidimensional data analysis. *data and Knowledge Engineering* 138(November 2021):101948. doi: 10.1016/j.datak.2021.101948.

Kumar S, Chuli A, Jain A, Prasanth N. 2023. Data analytics for pandemic management using mapreduce and apriori algorithm. *Procedia Computer Science* 230:455–66. doi: 10.1016/j.procs.2023.12.101.

Llave MR. 2019. A review of business intelligence and analytics in small and medium-sized enterprises. *International Journal of Business Intelligence Research* 10(1):19–42. doi: 10.4018/IJBIR.2019010102.

Malik S and Jeswani R. 2018. Literature review and techniques of machine learning algorithm used in business intelligence for inventory management. *International Journal of Engineering Sciences & Research*

Technology. doi: 10.5281/zenodo.1135987.

Masi D, Collis I. Panchal G. Clegg B. Koupaei EE. 2024. Digital transformation and business Intelligence for a SME: systems thinking action research using PrOH modelling. *Procedia Computer Science* 232(2023):1809–18. doi: 10.1016/j.procs.2024.02.003.

Muhamad N. 2023. Jumlah usaha mikro, kecil, dan menengah/UMKM Di Indonesia Berdasarkan Kelasnya (2021). *Https://Databoks.Katadata.Co.Id/*. [Retrieved March 22, 2024] https://databoks.katadata.co.id/datapublish/2023/10/13/usaha-mikro-tetap-merajai-umkm-berapa-jumlahnya#:~:text=Kementerian Koperasi dan Usaha Kecil,UMKM) di Indonesia pada 2021.).

Naik GR. 2023. AI based inventory management system using odoo. *International Journal of Scientific Research in Engineering and Management* 07(08):8–11. doi: 10.55041/ijsrem25510.

Sulaiman NS and Yahaya JH. 2013. Development of dashboard visualization for cardiovascular disease based on star scheme. *Procedia Technology* 11(Iceei):455–62. doi: 10.1016/j.protcy.2013.12.215.

Roy S, Cortesi A, and Sen S. 2022. Context-aware OLAP for textual data warehouses. *International Journal of Information Management Data Insights* 2(2):100129. doi: 10.1016/j.jjimei.2022.100129.

Satriawan MA, Andreswari R, and Pratiwi ON. 2021. Segmentasi pelanggan telkomsel menggunakan metode clustering dengan rfm model dan algoritma K-Means telkomsel customer segmentation using clustering method with RFM model and K-Means algorithm. *E-Proceeding of Engineering :* 8(2):2876–83.

Shobana J, Gangadhar C, Arora RK, Renjith PN, Bamini J, Chincholkar YD.2023. E-commerce customer churn prevention using machine learning-based business intelligence strategy. Measurement: *Sensors* 27 (2023) 100728. doi:10.1016/j.measen.2023.100728.

Sugiarto D, Mardianto I, Najih M, Adrian D, Pratama A. 2021. Perancangan dashboard untuk visualisasi harga dan pasokan beras di pasar induk beras cipinang. *Jurnal Teknologi Industri Pertanian* 31(1):12–19. doi: 10.24961/j.tek.ind.pert.2021.31.1.12.

Tsiakiris T, Tsopogloy S. and Antoniadis I. 2015. Business intelligence during times of crisis: adoption and usage of ERP systems by SMEs. *Procedia - Social and Behavioral Sciences* 175:299–307. doi: 10.1016/j.sbspro.2015.01.1204.

Tutunea MF and Rus RV. 2012. Business intelligence solutions for SME's. *Procedia Economics and Finance* 3(12):865–70. doi: 10.1016/s2212-5671(12)00242-0.