KALBE Nutritionals

Rakamin Academy

Creating Regression and Clustering Models

Presented by:
Shania Widianingrum Puspitasari

Background Story

The inventory team give the task of predicting daily sales quantity of Kalbe Nutritionals products. The aim of this project is to help the inventory team plan sufficient and adequate daily stock inventory.

The marketing team is tasked with creating customer segmentation (clusters). The goal of this project is to create groups of customers who share similar characteristics. This customer segmentation will be used by the Marketing Team to carry out customized promotions and sales
each segment

Company Overview

447 Customers
The company have 447 customers over 2022

14 Stores
The company have 14 stores over 2022

40 years old
The average customer is 40 years old

$162,043,000
Total income company in 2022

5,020 transaction
Transactions in companies in 2022

10,057 female and 8,239 male
Total customer male and female

Exploratory Data Analysis (SQL)

1. Customer Status

```sql
SELECT p."product_name", sum (t.totalamount) as sum_amount
FROM product as p
JOIN transaction as t
ON p.productid = t.productid
GROUP BY p."product_name"
ORDER BY sum_amount DESC
LIMIT 1;
```

	product_name character varying (50)	sum_amount bigint
1	Cheese Stick	27615000

The average age with **Married** status is **43 yeras old** and **Single** is **29 years old**

2. Customer Gender

Query | Query History

```sql
SELECT "gender", avg(age) from customer
group by "gender";
```

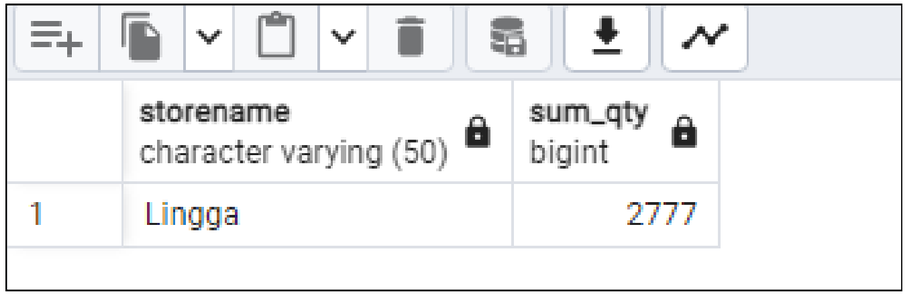
	gender integer	avg numeric
1	0	40.3264462809917355
2	1	39.1414634146341463

The average age of **women is 40 years old** and **men is 39 years old**

Exploratory Data Analysis (SQL)

3. High Quantity

```
SELECT s.storename, sum(t.qty) as sum_qty
FROM store as s
JOIN transaction as t
ON s.storeid = t.storeid
GROUP BY s.storename
ORDER BY sum_qty desc
LIMIT 1;
```

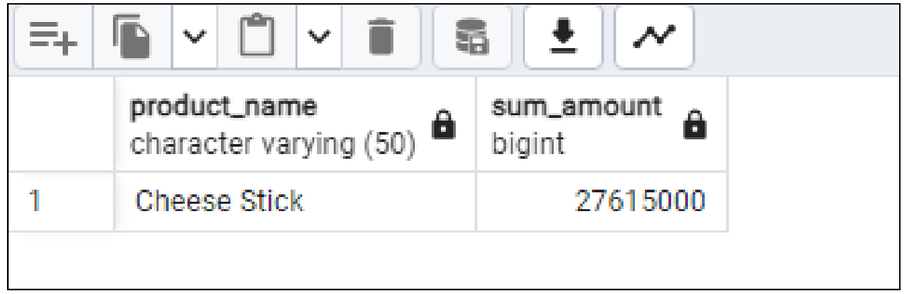


	storename character varying (50)	sum_qty bigint
1	Lingga	2777

Lingga is the store with the **highest qty** in 2022 amounting to **2,777**.

4. Total Amount

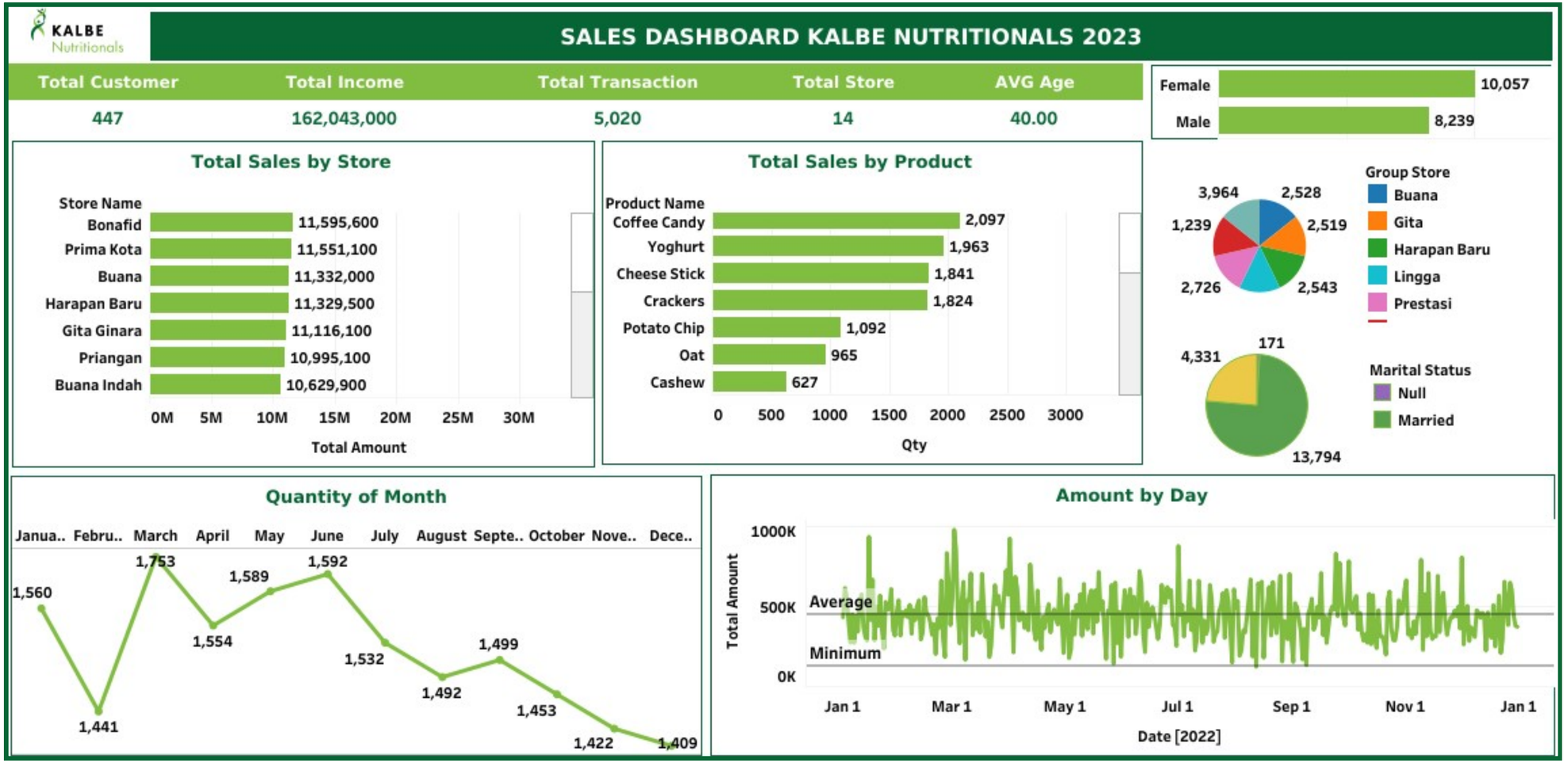
```
SELECT p."product_name", sum (t.totalamount) as sum_amount
FROM product as p
JOIN transaction as t
ON p.productid = t.productid
GROUP BY p."product_name"
ORDER BY sum_amount DESC
LIMIT 1;
```



	product_name character varying (50)	sum_amount bigint
1	Cheese Stick	27615000

Cheese sticks had the **highest** number of sales in a year at **$27,615,000**

Sales Dashboard

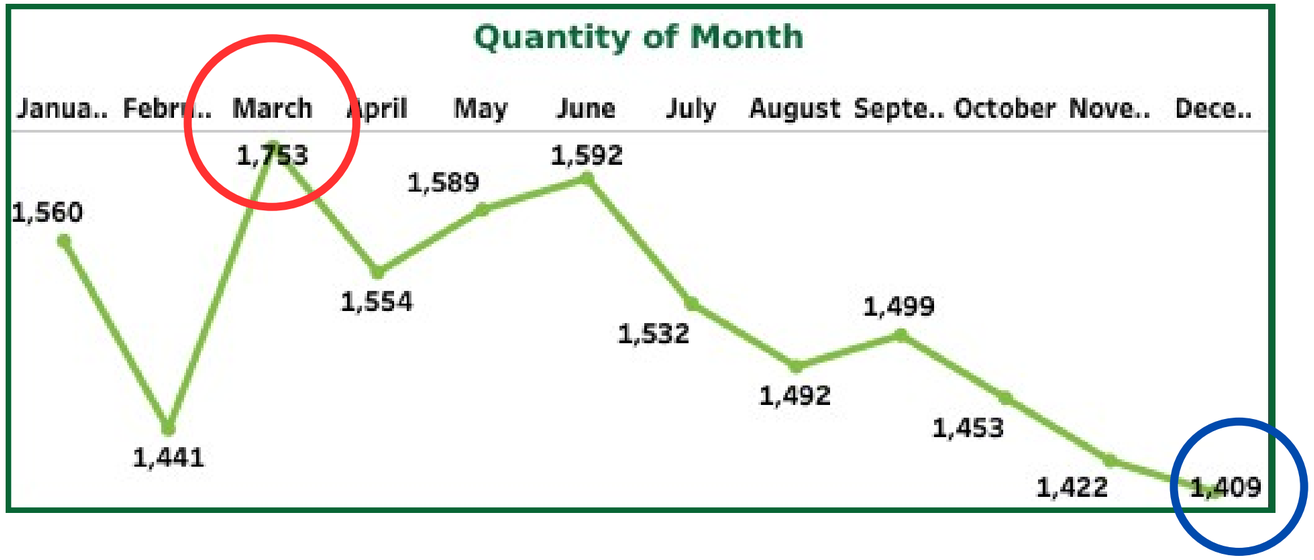


Link Dashboard

Sales Dashboard

Total Customer	Total Income	Total Transaction	Total Store
447	162,043,000	5,020	14

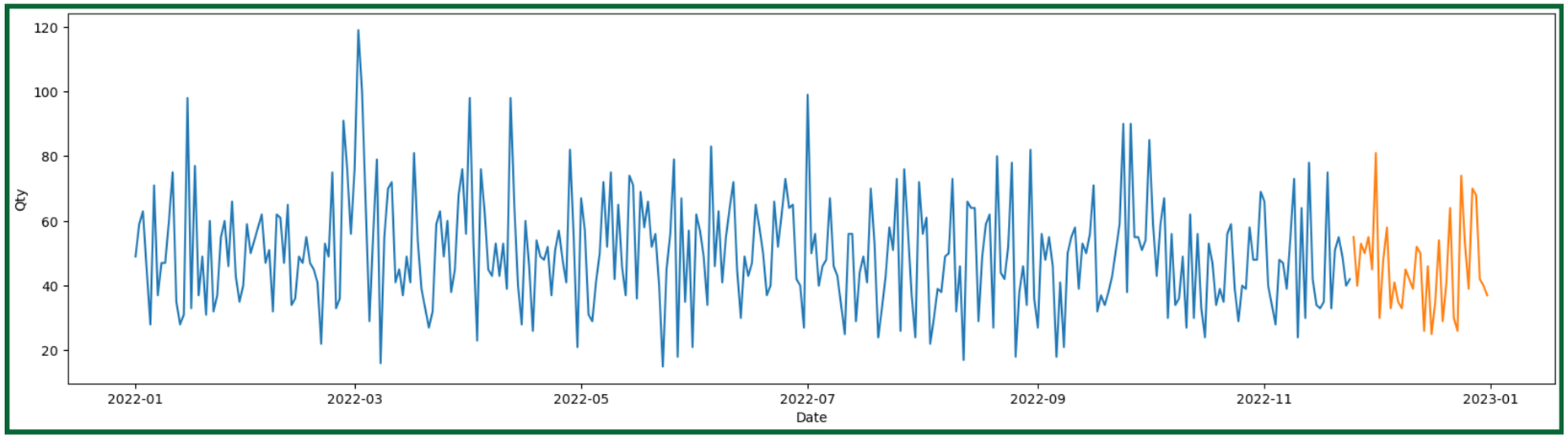
- The summary total of important information in 2022 is **447 customers, 5020 transactions, 14 stores,** and give the **sales amount with Rp 162M**.



- From this result, the **highest** quantity item in 2022 at **March** with **1.753 items**.
- The **lowest** quantity item in 2022 at **December** with **1.409 items**.

Machine Learning

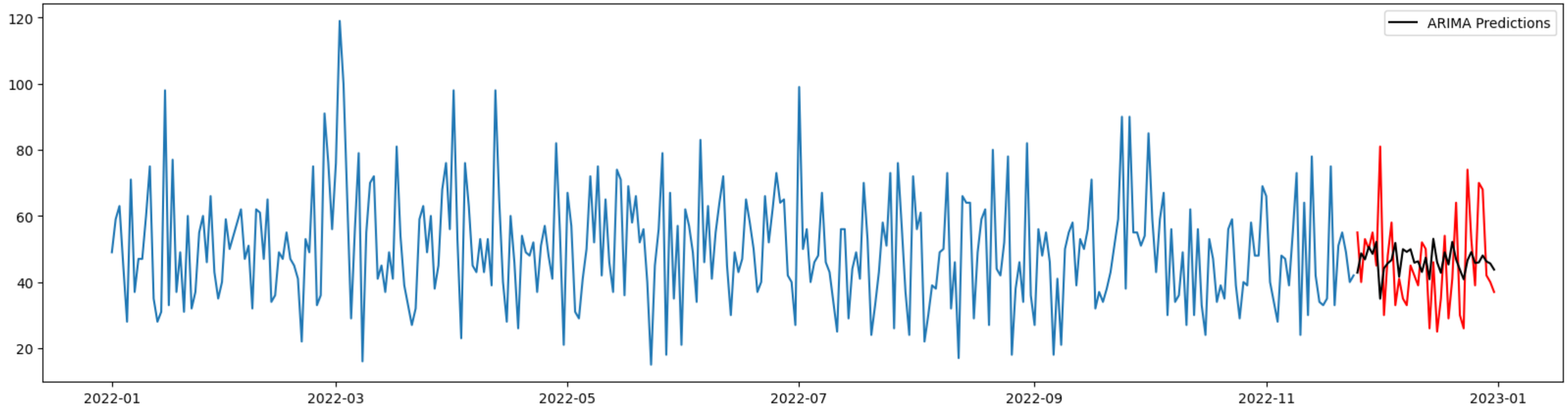
Regression with ARIMA Models



- The figure is the result of graphic daily quantity sold item in 2022.
- The blue line is data train which is the daily quantity sold item from 1 january - 24 November 2022.
- The orange line is data test which is daily quantity sold item from 25 November - 31 Desember 2022.

Machine Learning

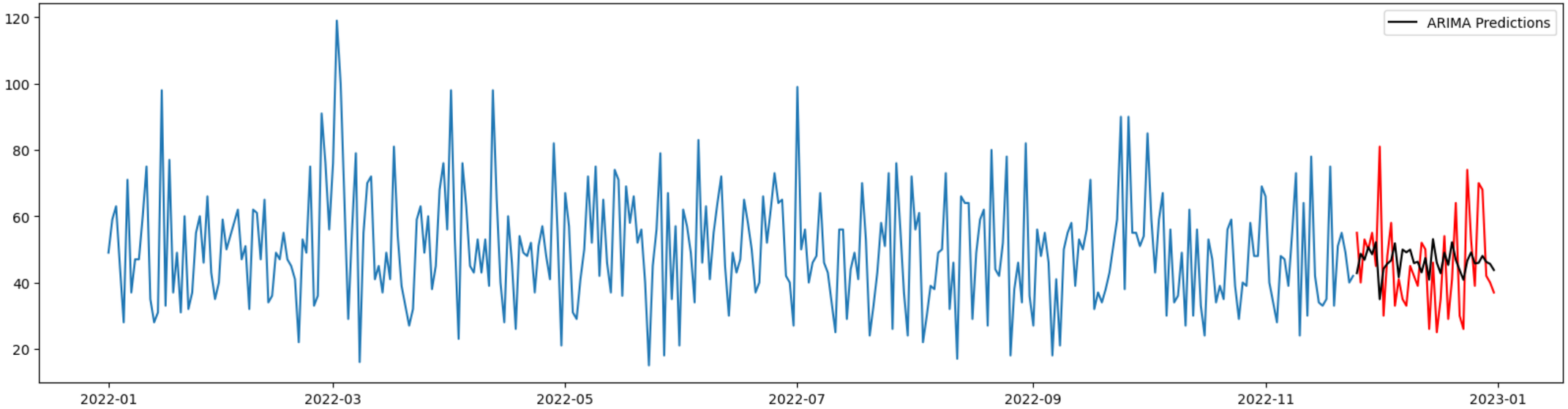
Regression with ARIMA Models



- The figure is the result of graphic daily quantity sold item in 2022 with addition of ARIMA models.
- The black line is the predicted data of quantity sold item where using data test from 25 November - 31 December 2022.
- The result from ARIMA models have differ slightly from data test line and give less quantity than reality.

Machine Learning

Customer Segmentatation with K-Means



- The figure is the result of graphic daily quantity sold item in 2022 with addition of ARIMA models.
- The black line is the predicted data of quantity sold item where using data test from 25 November - 31 December 2022.
- The result from ARIMA models have differ slightly from data test line and give less quantity than reality.

Clustering

The optimal clustering range using the elbow method is in the range of 4.

cluster_label	CustomerID	TransactionID	Qty	TotalAmount
0	165	11.066667	41.551515	346976.363636
1	194	11.587629	40.613402	409496.391753
2	81	10.777778	40.395062	287344.444444
3	4	7.250000	29.500000	150100.000000

Cluster 0

- Provide regular promotions to increase transactions
- Upsell products at high prices.

Cluster 1

- Build good relationships with customers
- Conduct surveys to develop customer interest.

Cluster 2

- Provide significant discounts to increase customer transactions
- Conduct surveys to identify potential product development

Cluster 3

- Offer loyalty promotions to maintain transactions
- Conduct customer satisfaction surveys, and upsell products at higher prices

THANK YOU

Project link on Github

KALBE Nutritionals

Rakamin Academy