Rakamin Academy X KALBE Nutritionals

Machine Learning Project

Quantity Prediction
Customer Segmentation

Data Science - Project Based Internship

Presented by
Oktavian Dwi Putra

Oktavian Dwi Putra

About Me

Result-oriented professional with background in digital marketing, especially SEO and a strong desire to transition into the field of Data Science. Possessing a solid foundation in statistics, machine learning, and data analysis. Eager to apply my analytical mindset, problem-solving abilities, and passion for data-driven insights to drive meaningful outcomes as a Data Scientist.

KALBE Nutritionals

Rakamin Academy

Experiences

SEO Specialist - cmlabs
Apr 2022 - Mar 2023

Improve website visibility and performance in search engines for several clients, including:

- Do keyword research for new content weekly.
- Perform on-page optimization for existing content to improve their performance.
- Monitor the ranking of targeted keywords from existing content.

Background Story

As a data scientist at Kalbe Nutritionals, I was asked to help the inventory team to predict the daily number of sales (quantity) of all Kalbe products, so that the inventory team can make sufficient daily inventory.

From the marketing team, I was asked to create customer clusters/segments based on several criteria that would be used by the marketing team to provide personalized promotions and sales treatment.

KALBE Nutritionals

Rakamin Academy

1. Exploratory Data Analysis

Tools: PostgreSQL, DBeaver

1. Exploratory Data Analysis

Query 1:
Average customer age based on marital status.

Marital Status	avg_age
	31.33
Married	43.04
Single	29.38

Query 2:
Average customer age based on gender.

gender	avg_age
0	40.33
1	39.14

Notes: 0 = female, 1 = male

Query 3:
Top three stores based on product sold.

storename	total_qty
Lingga	2,777
Sinar Harapan	2,588
Prestasi Utama	1,395

Query 4:
Top three products based on total amount.

Product Name	total_amount
Cheese Stick	27,615,000
Choco Bar	21,190,400
Coffee Candy	19,711,800

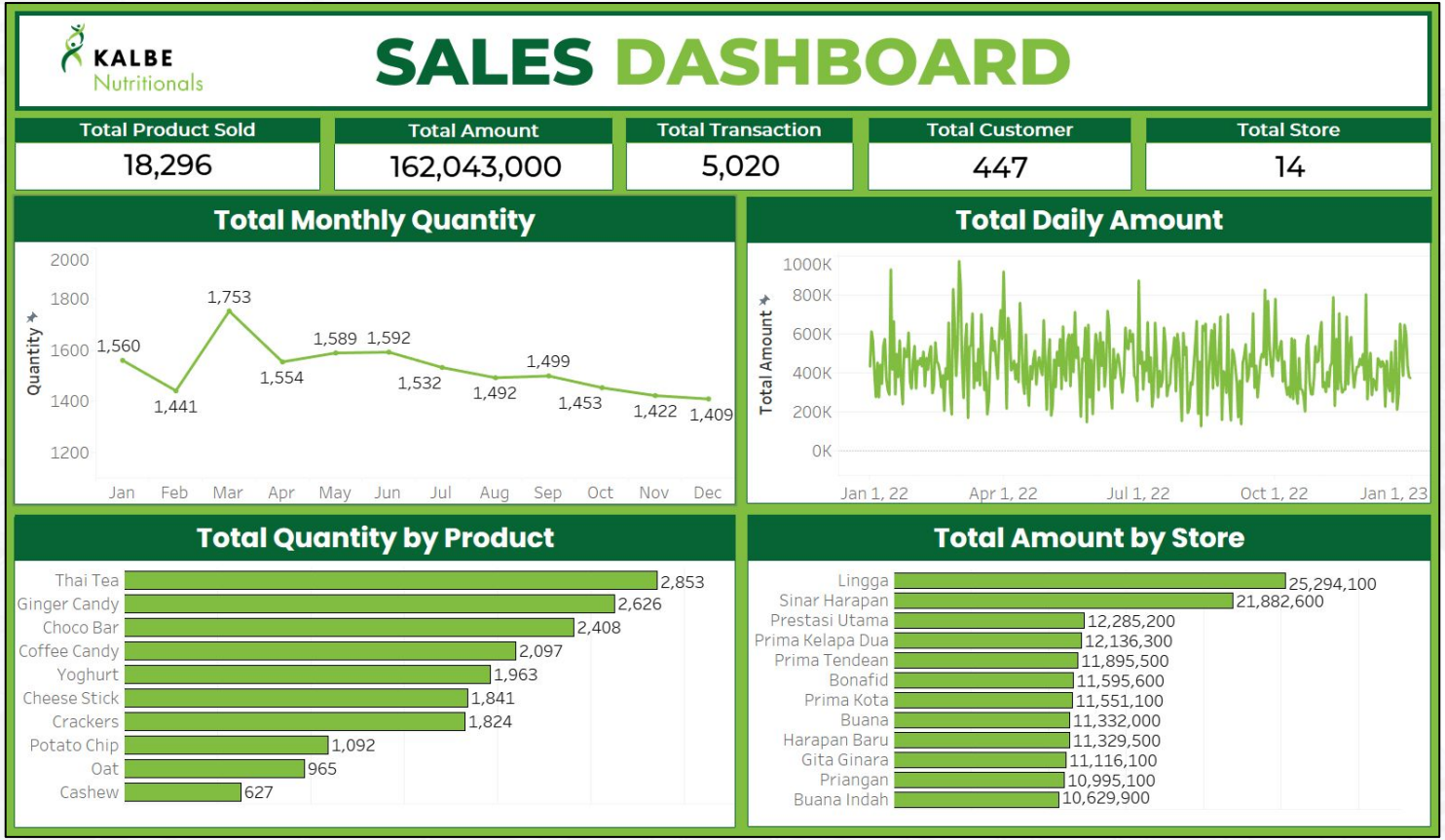
KALBE Nutritionals

Rakamin Academy

2. Data Visualization & Dashboard

Tools: Tableau Public

2. Data Visualization & Dashboard



3. Product Quantity Predictions

Tools: Jupyter Lab, ARIMA

3. Product Quantity Predictions

There are several steps that need to be done when we use ARIMA model for machine learning, including:

DATA PREPROCESSING

- Change data type
- Handle null value
- Handle duplicate data

MERGE DATA

- Data is combined into one based on the primary key and foreign key in each dataset.
- The master data has 4976 rows and 18 columns.

CREATE NEW DATA

- Create new data for regression by grouping the data using the Date column and aggregating the Qty column by summing it.
- The new data will consist of 365 rows.

3. Product Quantity Predictions

STATIONARY CHECK

For the stationary check, we use Augmented Dickey-Fuller test (ADF test) to check whether the data is stationary or not. Because the p-value is less than **α**, it indicates that there is very strong evidence to reject the null hypothesis (H0) and accept H1 which means the data is stationary and the existing data does not need to be differentiated or **d = 0**.

H0 = Data is not stationary
H1 = Data is stationary
α = 0.05

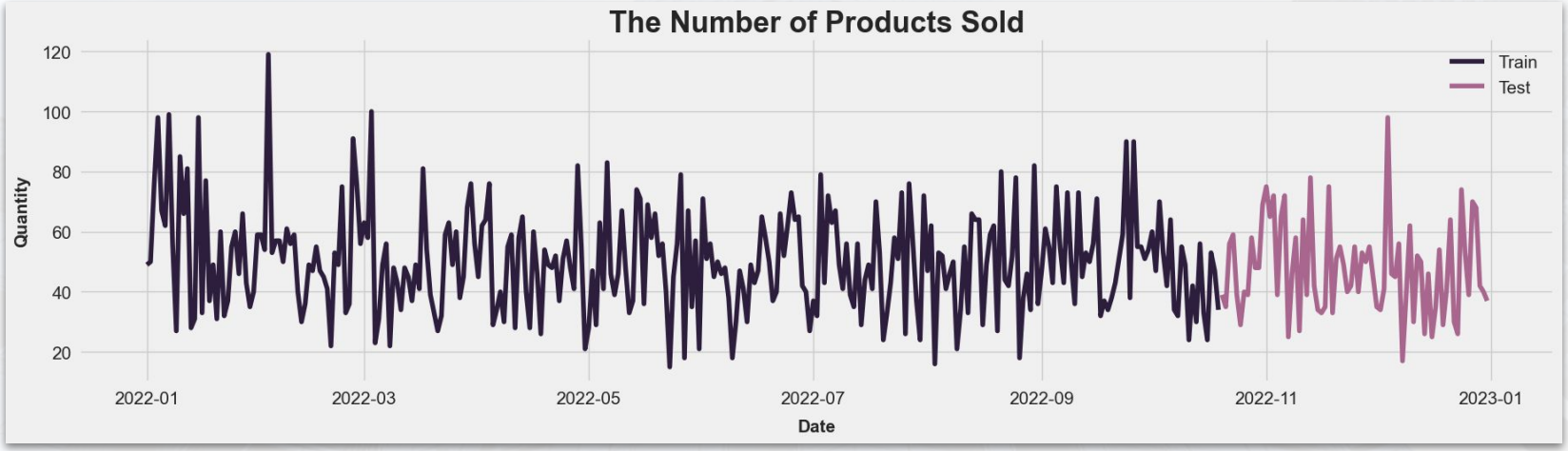
```
ADF Statistic: -19.091514
p-value: 0.000000

Conclusion:
Reject H0. Data is stationary
```

3. Product Quantity Predictions

SPLITTING DATA

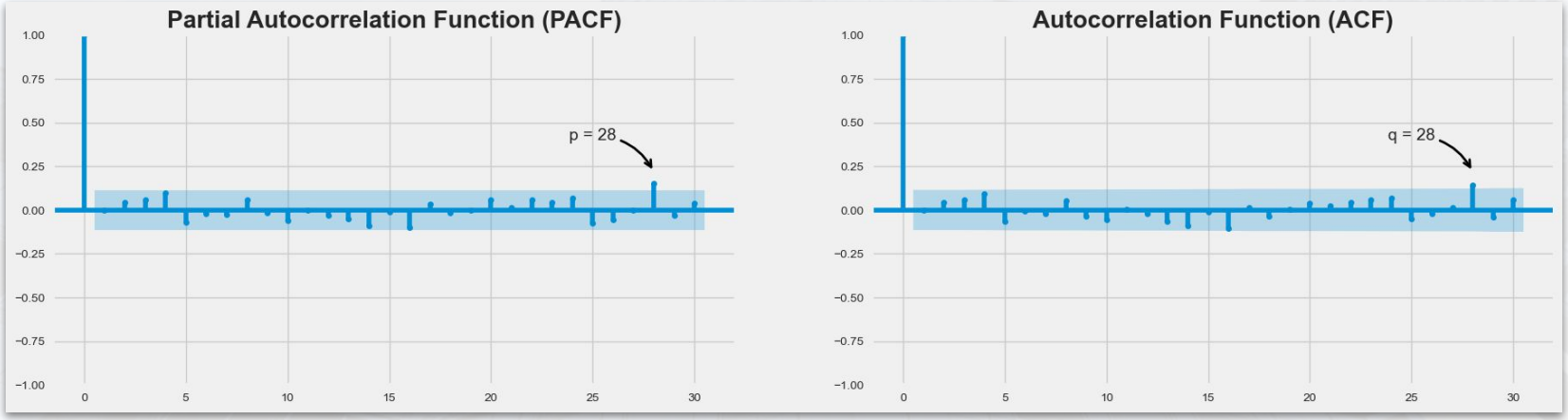
We will split data to 80:20 proportions, which is 80% data (292 rows) for training and 20% data (73 rows) for testing. We can visualize the data that have been splitted like in the image below.



3. Product Quantity Predictions

PACF & ACF PLOTS

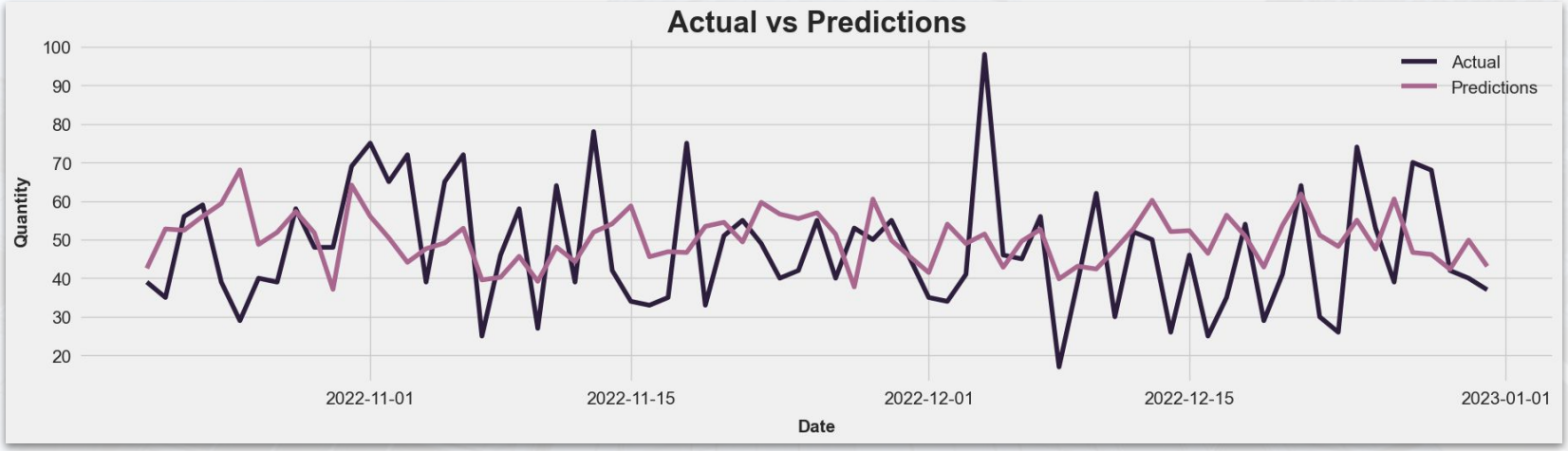
We use PACF and ACF plots to find out the p and q values. From the image below, we can see that the only lag that is out of significant limit is the 28th lag for both PACF and ACF plots. Therefore, **we will use 28 for p and q values**.



3. Product Quantity Predictions

MODELING

After we found the parameters for ARIMA model (p = 28, d = 0, q = 28), we can train the model with the data train and make prediction with the data test. We can visualize the result between the actual data and the predictions like in the image below.



3. Product Quantity Predictions

MODEL EVALUATION

```
Mean Absolute Error (MAE): 13.09
Mean Squared Error (MSE): 255.21
Root Mean Squared Error (RMSE): 15.98
Mean Absolute Percentage Error (MAPE): 31.86%
```

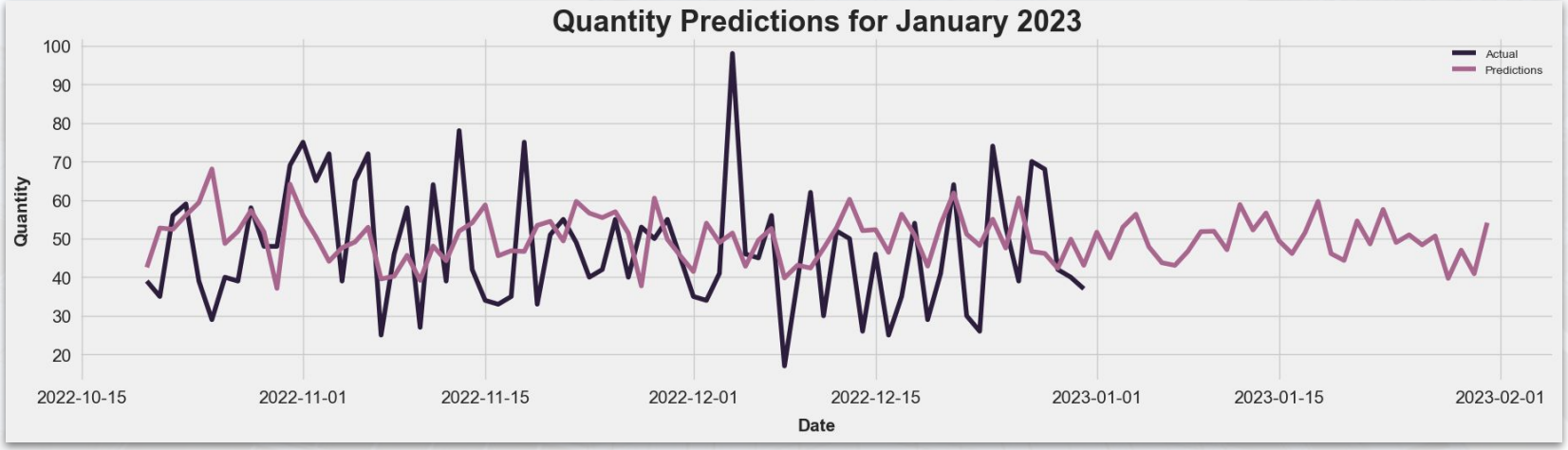
Generally, we can see that the predictions line has similar characteristics to the actual line with the evaluation metrics in the above. Based on the evaluation metrics, we can conclude that the performance from model that we used is good enough because it has low error values.

3. Product Quantity Predictions

FUTURE PREDICTIONS

We can also use the model to predict the quantity of product needed in January 2023. Based on the prediction result, the number of quantities needed in January 2023 has integer statistic values:

Mean: 50, Median: 50, Min: 32, Max: 60, Total: 1545



KALBE Nutritionals

Rakamin Academy

4. Customer Segmentation

Tools: Jupyter Lab, K-Means Clustering

4. Customer Segmentation

There are several steps that need to be done when we use K-Mean Clustering model for machine learning, including:

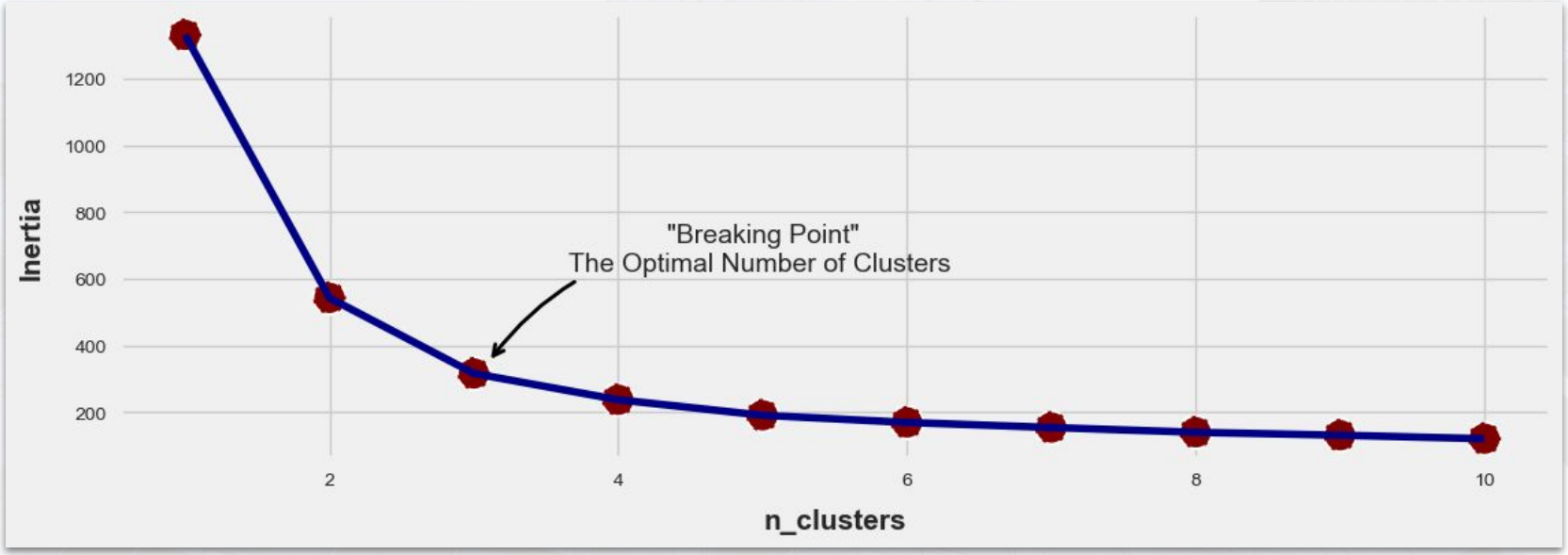
CREATE NEW DATA

- Create new data for clustering by grouping the data using the CustomerID column and aggregating the TransactionID column by counting it, the Qty column by summing it, and TotalAmount column also by summing it.

4. Customer Segmentation

ELBOW METHOD

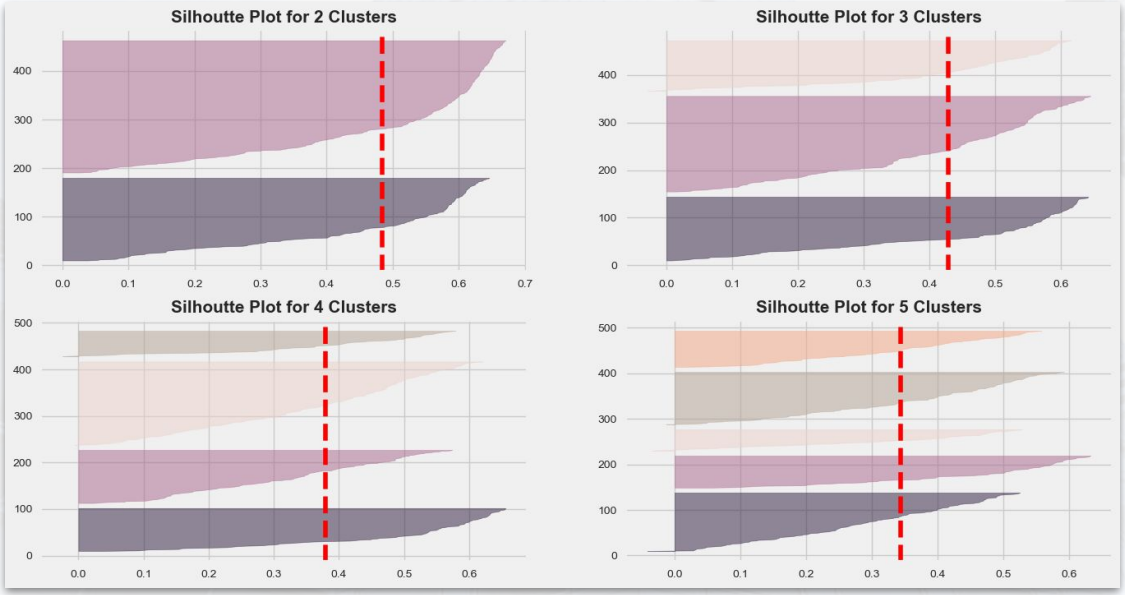
From the inertia value plot below, we can see the optimal number of clusters is at **the breaking point**. After that point, the inertia value starts to decrease insignificantly. Therefore we can conclude that **the optimal number of clusters is 3**.



4. Customer Segmentation

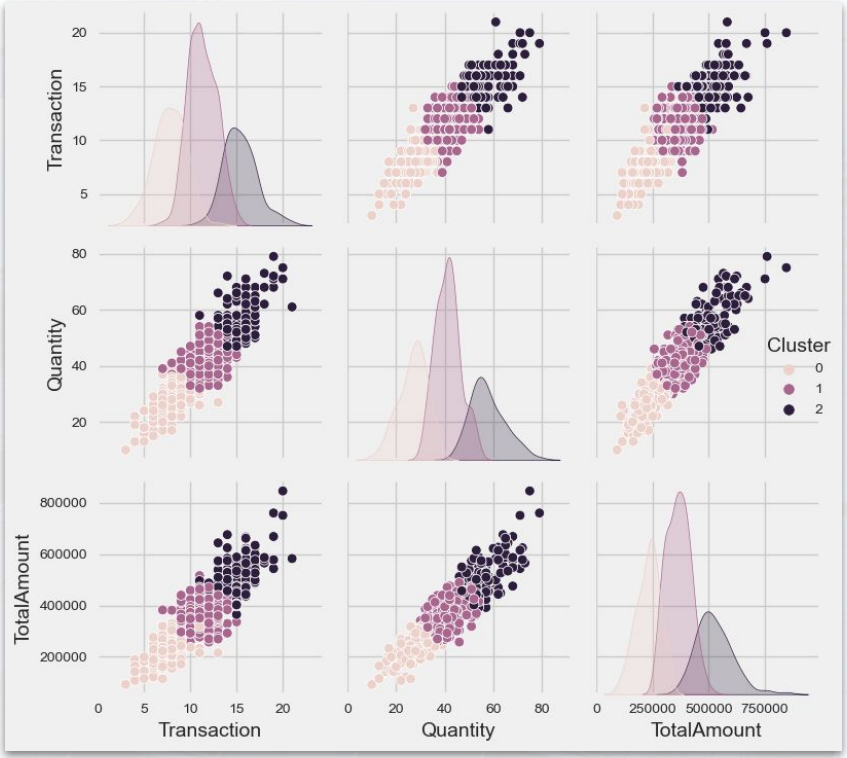
SILHOUTTE SCORE

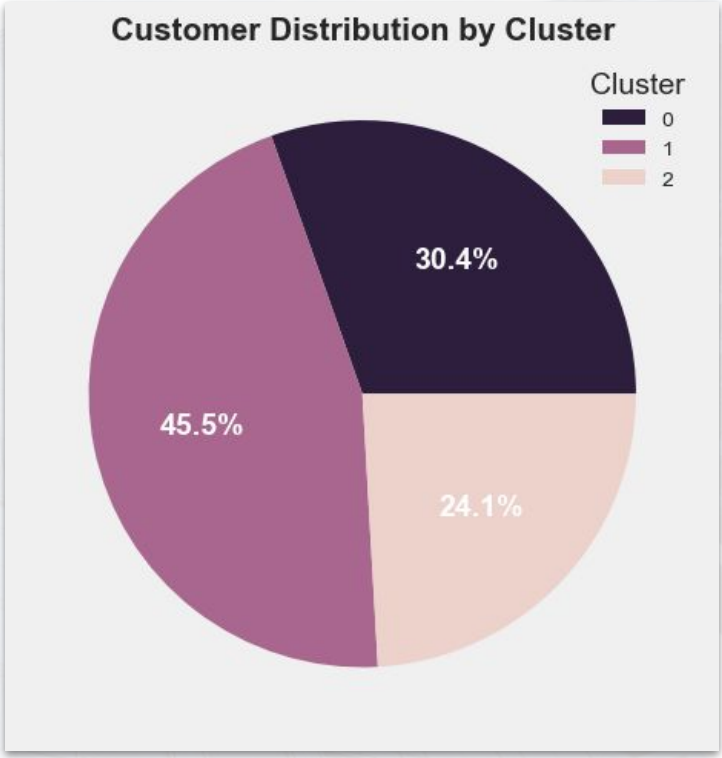
From silhouette plot below, we should choose the number of clusters with a large coefficient average value while considering the proportional distribution of the clusters formed. Therefore, we can conclude that **the optimal number of clusters is 3**.



4. Customer Segmentation

CLUSTER VISUALIZATION





4. Customer Segmentation

INTERPRETATION

Cluster 0 : Low-Level Customer
Customers who make infrequent transactions, buy small quantities, and generate low total amounts.

Cluster 1 : Mid-Level Customer
Customers who have moderate transaction frequency, purchase medium quantities, and generate intermediate total amounts.

Cluster 2 : High-Level Customer
Customers who make frequent transactions, buy large quantities, and generate high total amounts.

	CustomerID	Transaction		Quantity		TotalAmount	
	count	mean	median	mean	median	mean	median
Cluster							
0	135	7.785185	8.0	26.933333	28.0	229388.888889	235300.0
1	202	11.282178	11.0	41.188119	41.0	363267.326733	362400.0
2	107	15.383178	15.0	57.654206	57.0	525431.775701	512400.0

4. Customer Segmentation

BUSINESS RECOMMENDATION

Cluster 0 : Low-Level Customer

- Use special promotions and discounts to encourage more frequent purchases. We can also use email marketing and personalized offers to re-engage them.
- Offer recommendations for complementary products or services to increase the size of their orders.

Cluster 1 : Mid-Level Customer

- Focus on retaining and growing the loyalty using personalized loyalty programs that offer incentives and rewards for continued shopping.
- Analyze their purchase history to identify products or services that are commonly bought together. Promote cross-selling and upselling opportunities to increase their average transaction value.

Cluster 2 : High-Level Customer

- Create exclusive offers, early access, and personalized experiences for this cluster to make them feel valued.
- Reward this cluster for their loyalty and encourage them to refer friends and family to our business. Consider creating a referral program that benefits both parties.
- Identify high-value products or services and promote them to this cluster. Offer bundled packages or discounts for premium offerings.

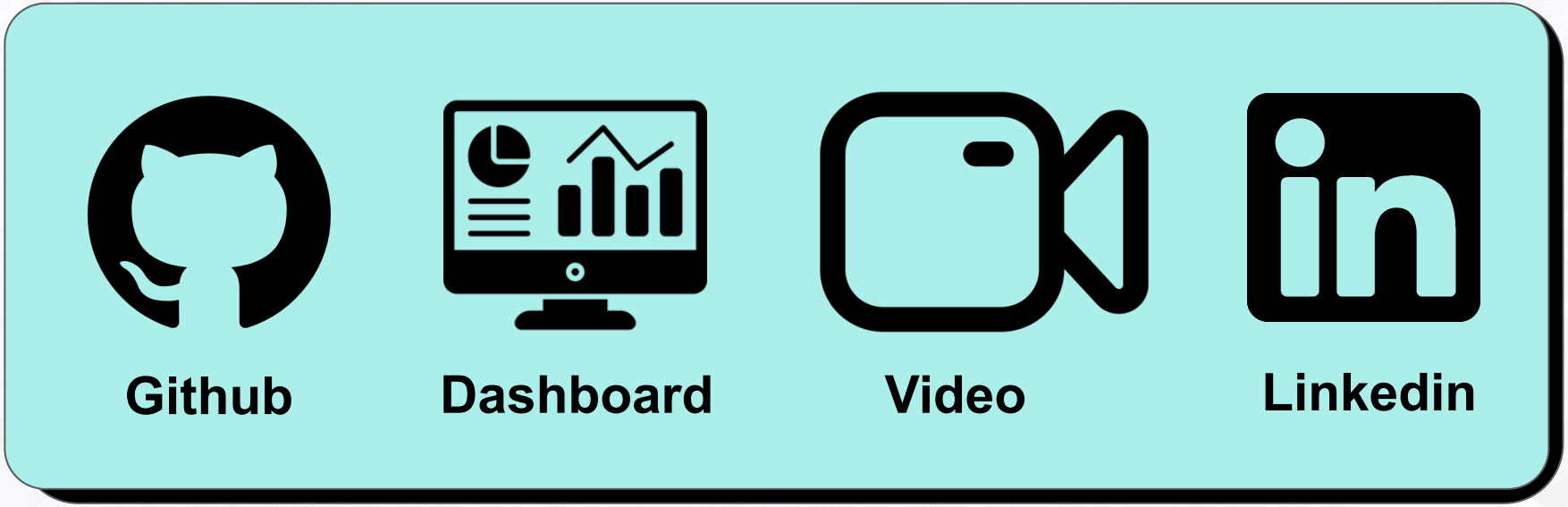
Rakamin Academy

5. Documentation

5. Documentation

KALBE Nutritionals

Rakamin Academy



Thank You