Rakamin Academy X KALBE Nutritionals

Task 5 - Final Project

Kalbe Nutritionals Data Scientist Virtual Internship

Presented by
Muhammad Randa Yandika

Background Project

For the final project, participants will have the opportunity to apply their acquired knowledge in Data Science during their internship at Kalbe Nutritionals. Project task will be to develop a predictive data model to enhance the company's business, such as optimizing competitive business strategies or conducting regression and clustering analysis using the available data. Additionally, participants need to prepare visual media to present the solutions to the clients.

The goal of this project is to demonstrate your proficiency in applying various Data Science techniques and methodologies to solve real-world business problems. By developing a predictive data model, participants will contribute to improving the company's decision-making processes and identifying opportunities for growth.

Challenge 1

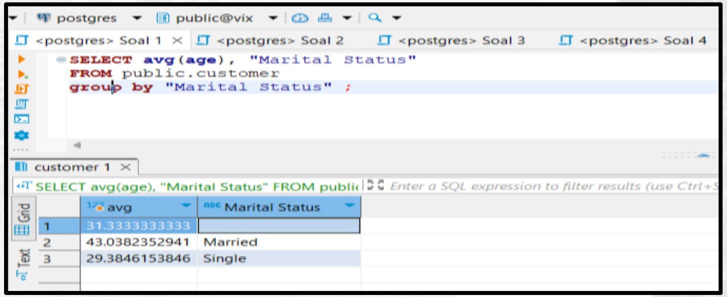
SQL, Dbeaver Connection with PostgreSQL

In this first challenge, I answered 4 questions that I wanted to know by doing exploratory data analysis on dbeaver.

Question 1: What is the average age of the customer in terms of their marital status?

Query:

```
SELECT avg(age), "Marital Status"
FROM public.customer
group by "Marital Status" ;
```



```
SELECT avg(age), "Marital Status"
FROM public.customer
group by "Marital Status" ;
```

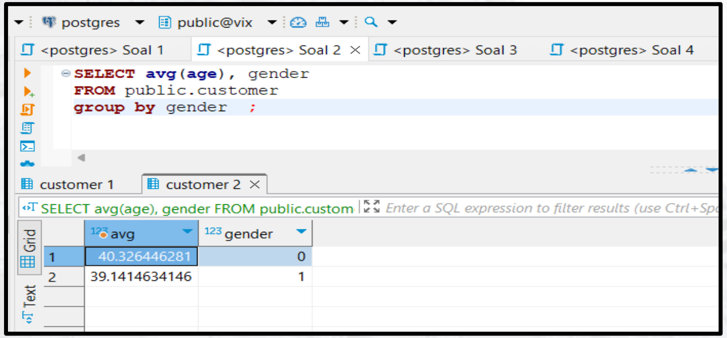
SELECT avg(age), "Marital Status" FROM publi

	avg	Marital Status
1	31.3333333333	
2	43.0382352941	Married
3	29.3846153846	Single

Question 2: What is the average age of the customer in terms of gender?

Query:

```
SELECT avg(age), gender
FROM public.customer
group by gender  ;
```



```
SELECT avg(age), gender
FROM public.customer
group by gender  ;
```

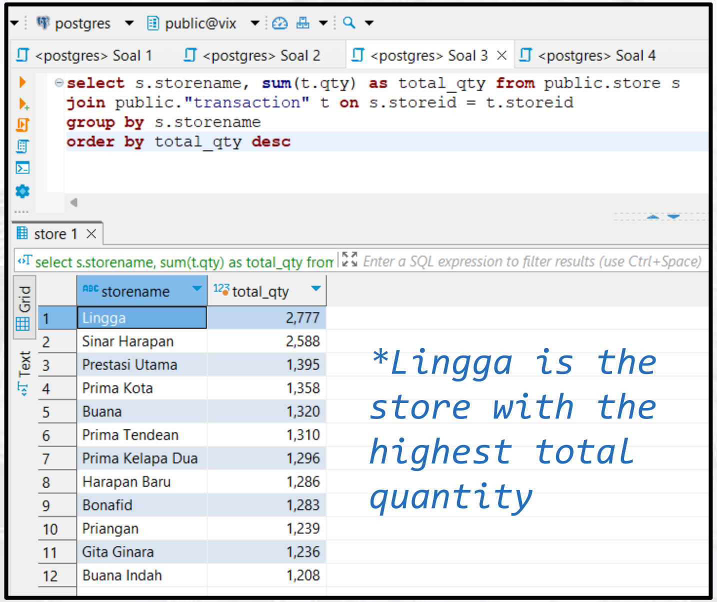
SELECT avg(age), gender FROM public.custom

	avg	gender
1	40.326446281	0
2	39.1414634146	1

Question 3: Determine the name of the store with the largest total quantity!

Query:

```
select s.storename, sum(t.qty) as total_qty
from public.store s
join public."transaction" t on s.storeid = t.storeid
group by s.storename
order by total_qty desc
```



```
select s.storename, sum(t.qty) as total_qty from public.store s
join public."transaction" t on s.storeid = t.storeid
group by s.storename
order by total_qty desc
```

select s.storename, sum(t.qty) as total_qty from

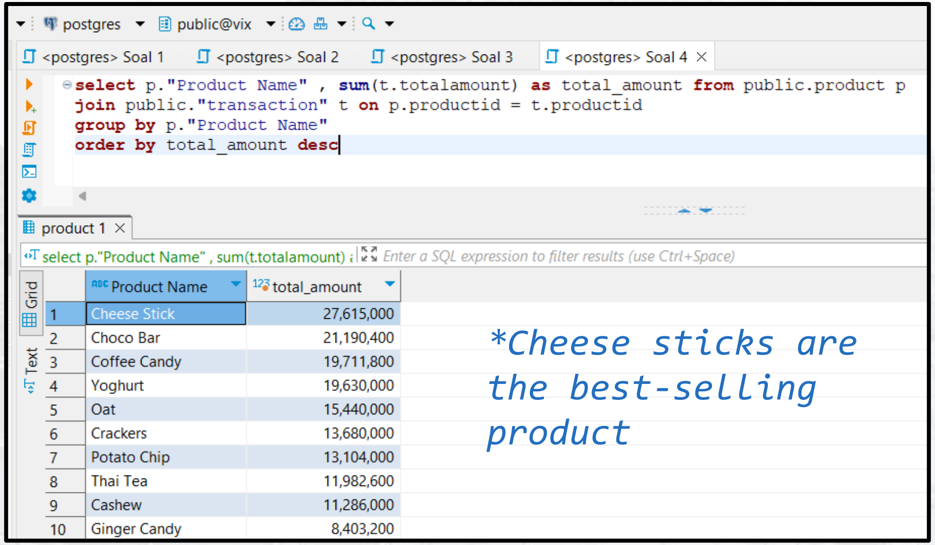
	storename	total_qty
1	Lingga	2,777
2	Sinar Harapan	2,588
3	Prestasi Utama	1,395
4	Prima Kota	1,358
5	Buana	1,320
6	Prima Tendean	1,310
7	Prima Kelapa Dua	1,296
8	Harapan Baru	1,286
9	Bonafid	1,283
10	Priangan	1,239
11	Gita Ginara	1,236
12	Buana Indah	1,208

*Lingga is the store with the highest total quantity

Question 4 : Determine the best-selling product name with the highest total amount!

Query:

```
select p."Product Name",sum(t.totalamount) as total_amount
from public.product p
join public."transaction" t on p.productid = t.productid
group by p."Product Name"
order by total_amount desc
```


postgres
public@vix
<postgres> Soal 1
<postgres> Soal 2
<postgres> Soal 3
<postgres> Soal 4
product 1
select p."Product Name" , sum(t.totalamount) a
Enter a SQL expression to filter results (use Ctrl+Space)
Grid
Text

```
select p."Product Name" , sum(t.totalamount) as total_amount from public.product p
join public."transaction" t on p.productid = t.productid
group by p."Product Name"
order by total_amount desc
```

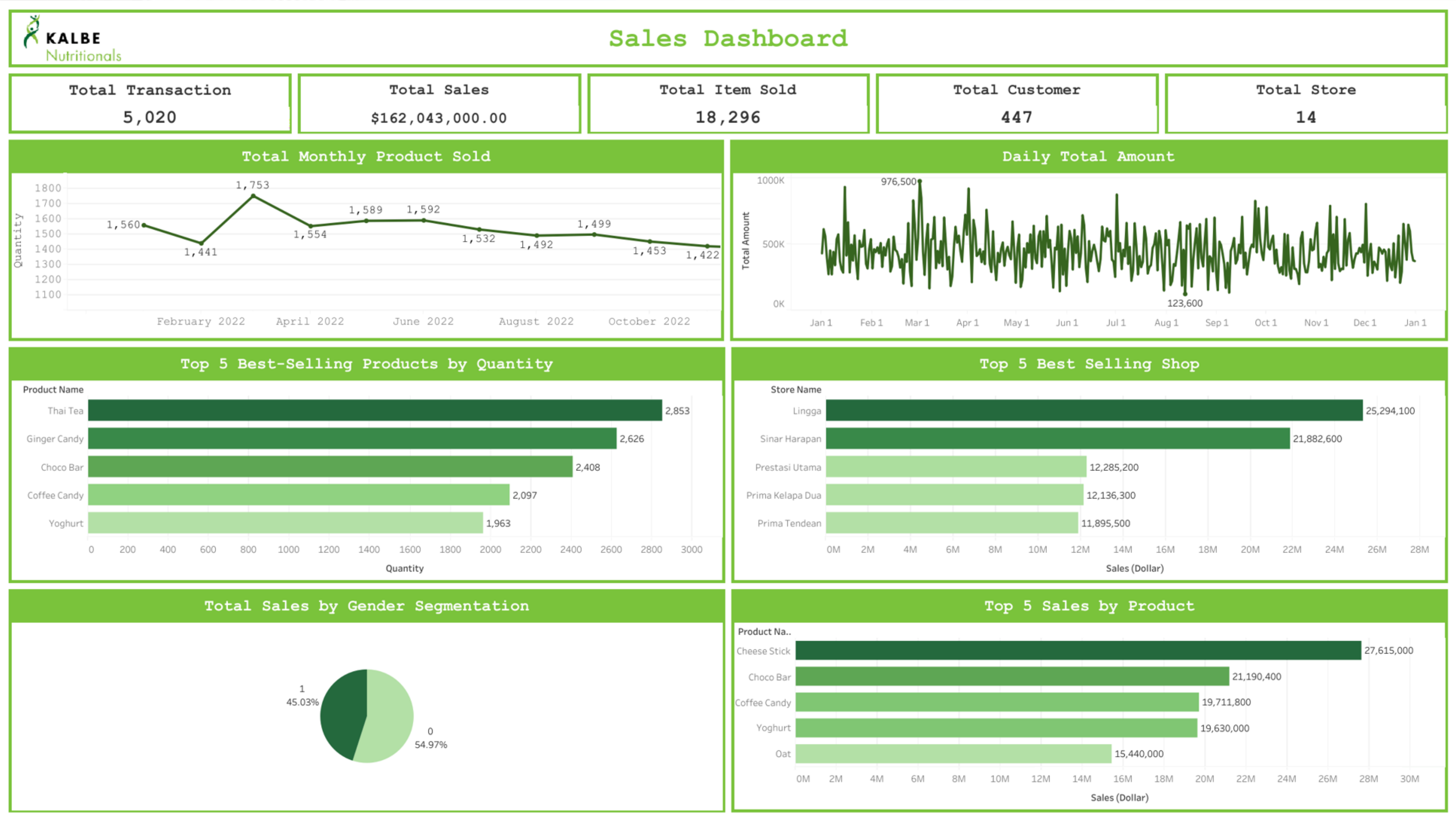
	Product Name	total_amount
1	Cheese Stick	27,615,000
2	Choco Bar	21,190,400
3	Coffee Candy	19,711,800
4	Yoghurt	19,630,000
5	Oat	15,440,000
6	Crackers	13,680,000
7	Potato Chip	13,104,000
8	Thai Tea	11,982,600
9	Cashew	11,286,000
10	Ginger Candy	8,403,200

*Cheese sticks are the best-selling product

Challenge 2

Data Visualization & Dashboard creation Using Tableau

Dashboard Link: https://bit.ly/kalbe_dashboard



KALBE Nutritionals

Sales Dashboard

Total Transaction	Total Sales	Total Item Sold	Total Customer	Total Store
5,020	$162,043,000.00	18,296	447	14

Total Monthly Product Sold

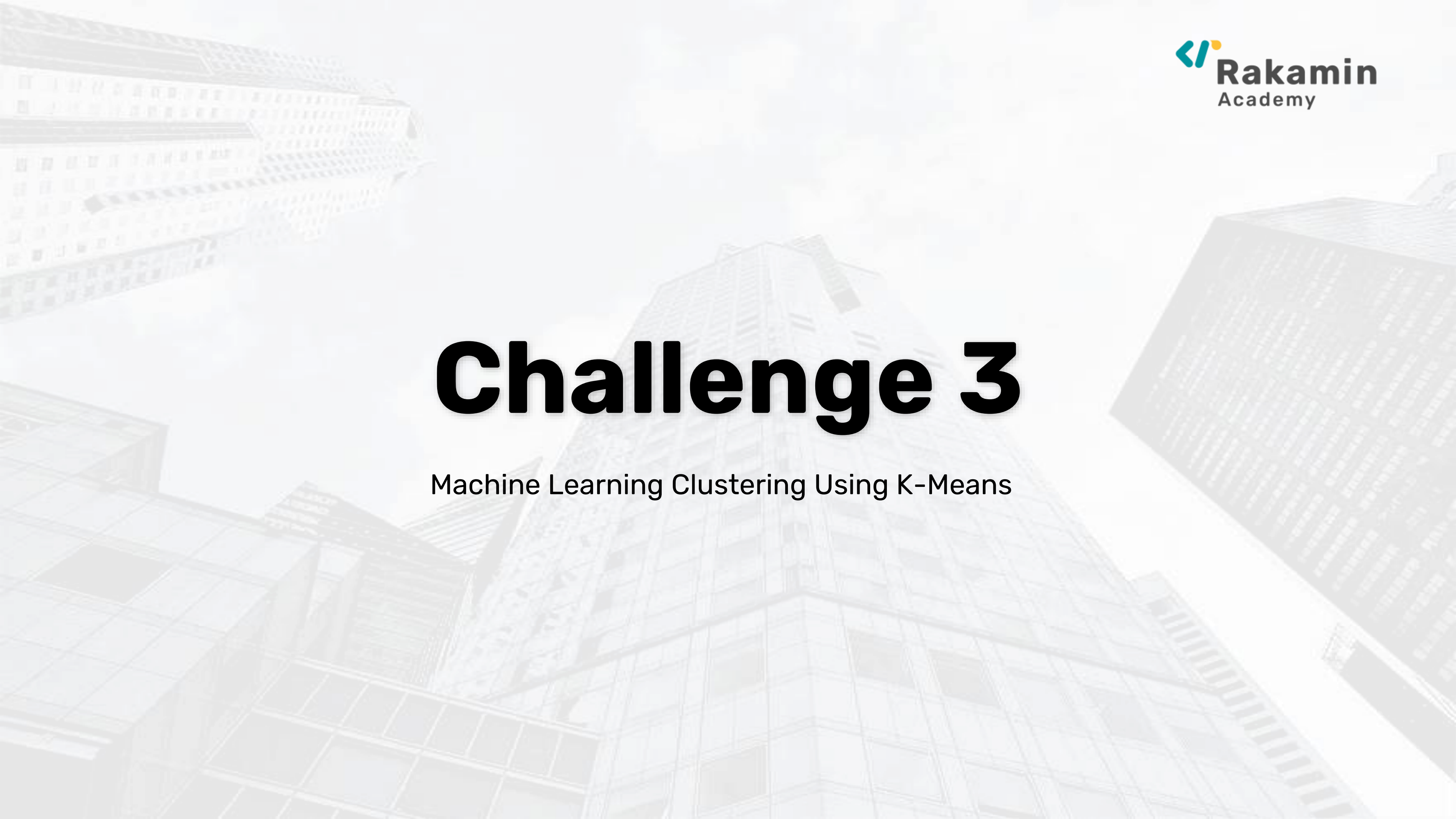
Daily Total Amount

Top 5 Best-Selling Products by Quantity

Top 5 Best Selling Shop

Total Sales by Gender Segmentation

Top 5 Sales by Product



Challenge 3

Machine Learning Clustering Using K-Means

Project Outline, the stages of the clustering process will be divided into several steps

DATA PREPARATION

- Load Dataset
- Data Cleansing (Null Value Imputation, Handling Duplicated Value, Change Dtype)

MERGE DATA

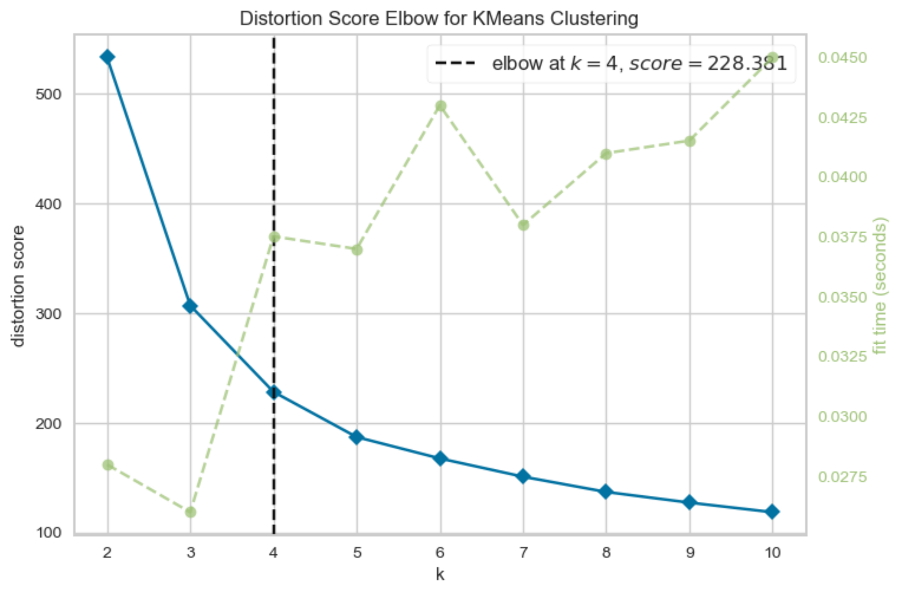
- Data is combined 1 based on the primary key and foreign key in each dataset.
- The Master dataset has 5019 entries and 18 columns.

DATA PREPROCESSING

- Create new data for clustering based on Customer ID then what is aggregated is :
    * Transaction id (count)
    * Qty (sum)
    * Total amount (sum)
- Remove Outlier using IQR and perform Data Scaling using Standardscaler on new data that was previously aggregated (Transaction id, Qty, TotalAmount)

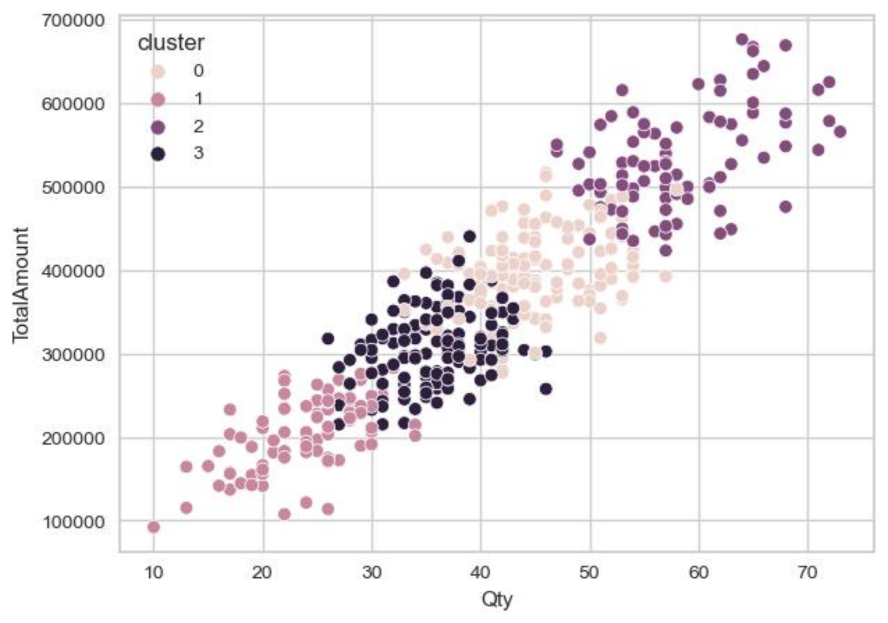
CLUSTERING

Finding the best k value using the Elbow Method. Based on the graph below, the best k value is k = 4 for cluster formation on Kmeans.



CLUSTERING

The following is the result of clustering with a total of 4 clusters formed taken from Qty and TotalAmount



CLUSTERING RESULT

- **Cluster 0** = customers with moderate to large purchases in terms of Quantity and Total Amount. These are customers who spend significant amounts of money and spend significant amounts of money.

- **Cluster 1** = customers with smaller purchases and spend relatively less money. These are customers with smaller purchases at a more affordable scale.

- **Cluster 2** = customers with large purchases and spending a significant amount of money. Represents a customer with a higher need or preference in terms of quantity and spending.

- **Cluster 3** = customers with moderate purchases and spend moderate amounts of money. These are customers with moderate needs or preferences in terms of quantity and spending.

BUSINESS RECOMMENDATION

Cluster 0: "Health Enthusiasts"
Customers with medium to large purchases in terms of Quantity and Total Amount.

- Focus on marketing campaigns that emphasize the benefits of Kalbe Nutritionals' nutritional products in supporting the health and active lifestyle of customers in Cluster 0 who have moderate to large purchases. Communicate that Kalbe Nutritionals products provide the best solutions to support a healthy and active lifestyle.
- Offer nutritional product packages that consist of products from different stages of life, such as preparation for pregnancy, nutrition for babies, and adults. Thus, customers can choose products that suit their life stages.
- Collaborate with pharmacies or large health shops that have large quantities of purchases to expand the distribution of Kalbe Nutritionals' nutritional products in Cluster 0.

BUSINESS RECOMMENDATION

Cluster 1: "Budget Shoppers“
Customers with smaller purchases and spend relatively less money.

- Focus on special promotional campaigns for Kalbe Nutritionals nutritional products that have affordable prices and can meet the nutritional needs of customers in Cluster 1 who have smaller purchases. Communicate that Kalbe Nutritionals products provide high value at an affordable price.
- Increase awareness of the benefits of Kalbe Nutritionals' nutritional products through educational programs about the importance of good nutrition for health, especially at certain stages of life. That way, customers can better understand the importance of consuming quality nutrition.
- Offer special discounts or shopping vouchers as incentives for customers in Cluster 1 who make repeated purchases or buy several products at once.

BUSINESS RECOMMENDATION

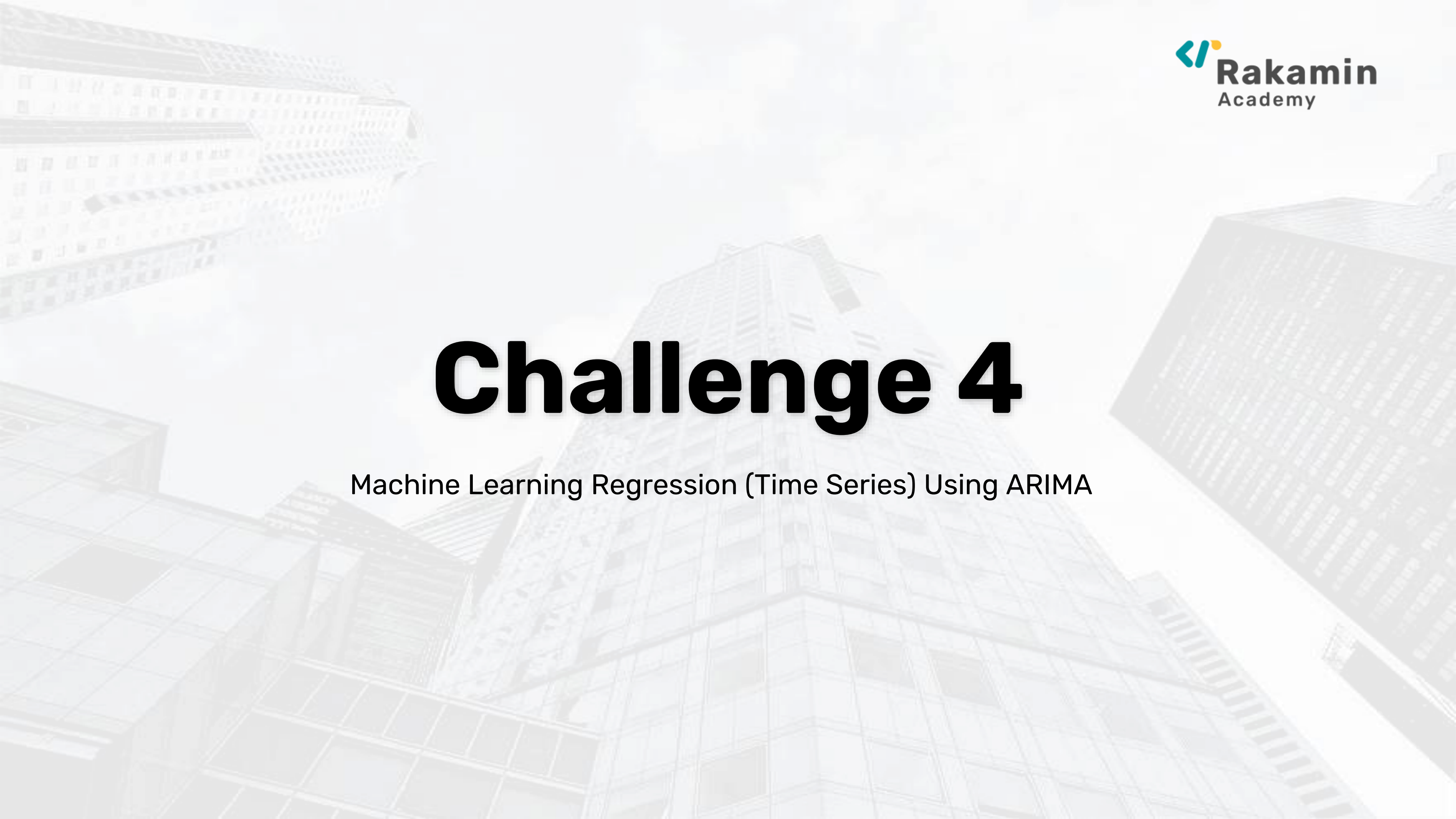
Cluster 2: "Nutrition Enthusiasts“
Customers with large purchases and spending significant amounts of money.

- Focus on developing nutritional products for Kalbe Nutritionals that are more specialized and rich in nutrients, especially for adults who have higher needs or preferences in terms of quantity and spending. Communicate that Kalbe Nutritionals products in Cluster 2 are the best choice to meet specific nutritional needs.
- Building relationships with doctors or large clinics that have clients with higher nutritional needs, so that Kalbe Nutritionals products can be recommended to patients.
- Hold a special event involving health professionals to educate customers in Cluster 2 about the benefits and advantages of Kalbe Nutritionals' nutritional products in meeting advanced nutritional needs.

BUSINESS RECOMMENDATION

Cluster 3: "Quality Seekers“
Customers with moderate purchases and spending moderate amounts of money.

- Increase the availability of the most popular Kalbe Nutritionals nutritional products and are sought after by customers in Cluster 3. Ensure that these products are always available in sufficient stock.
- Increase communication about quality and the latest technology used in the production of Kalbe Nutritionals nutritional products, so that customers in Cluster 3 are more confident about the quality of the products they buy.
- Focus on developing nutritional products for Kalbe Nutritionals that help strengthen the immune system and health in general, considering that customers in Cluster 3 are looking for reliable products for their health.



Rakamin
Academy

Challenge 4

Machine Learning Regression (Time Series) Using ARIMA

The stages of the Regression process will be divided into several steps

DATA PREPARATION

- Load Dataset
- Data Cleansing (Null Value Imputation, Handling Duplicated Value, Change Dtype)

MERGE DATA

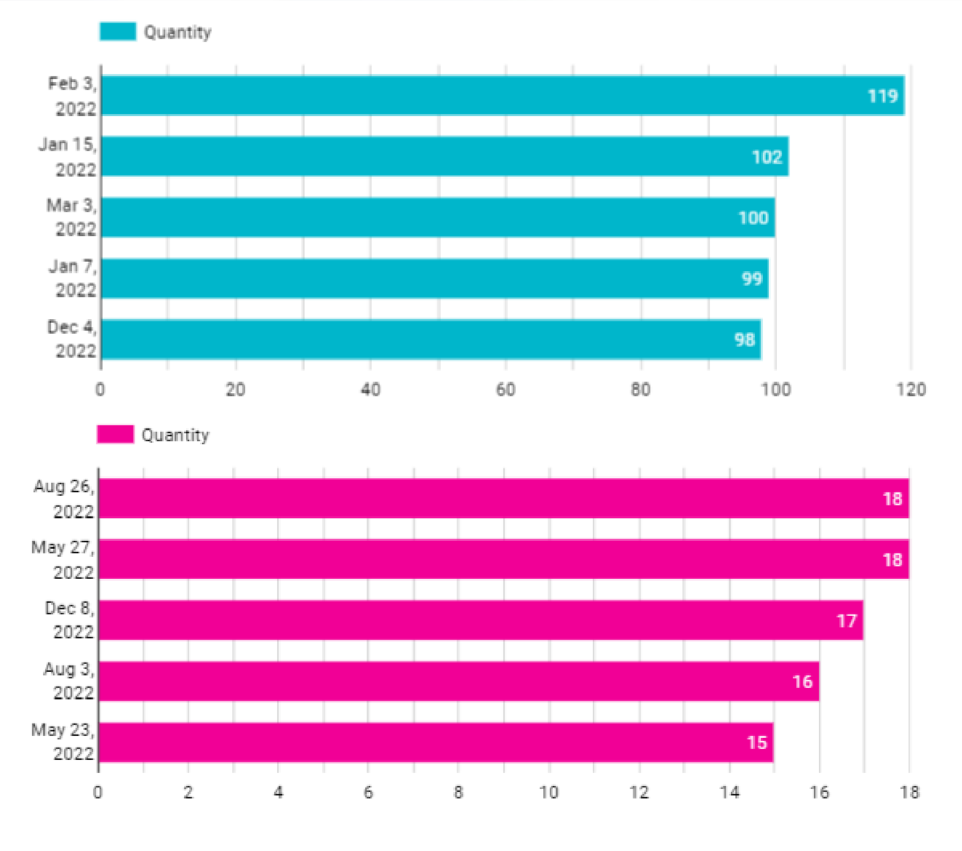
- Data is combined 1 based on the primary key and foreign key in each dataset.
- The Master dataset has 5019 entries and 18 columns.

DATA PREPROCESSING

- Create new data for clustering based on Date then what is aggregated is :
  * Qty sum
- The new data will consist of 365 Rows.

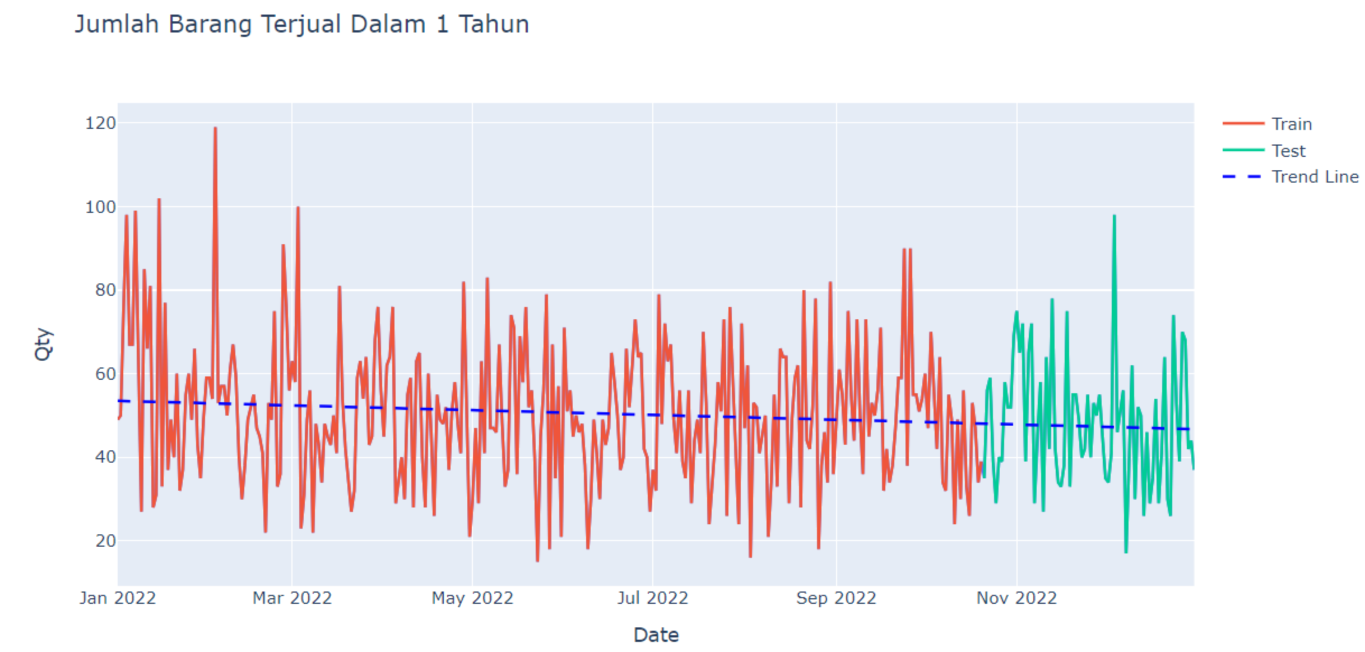
Day with the highest and the lowest Quantity of item sold

Feb, 3rd 2022 with 119 items
And
May, 23th 2022 with 15 items



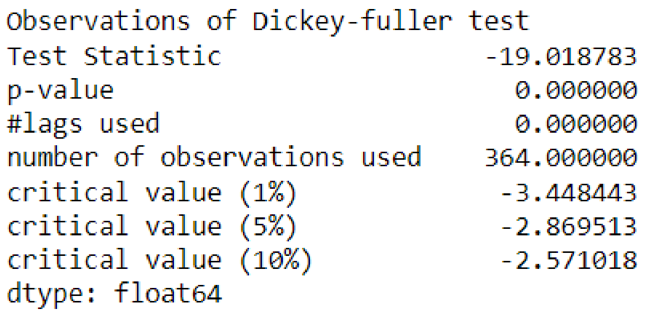
Quantity Trends

There was a decrease in the quantity of items sold in this 1 year



Stationary Test

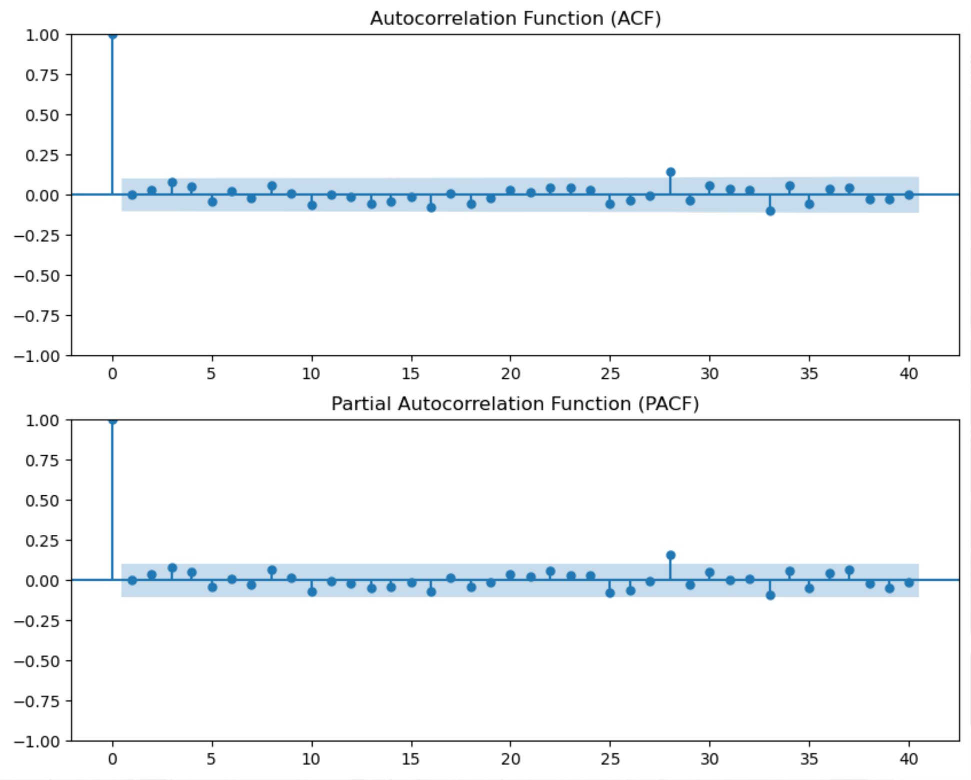
If the p-value of the Augmented Dickey-Fuller test (ADF test) is 0, it indicates that there is very strong evidence to reject the null hypothesis, namely the hypothesis that the data has a unit root and is not stationary. Existing data does not need to be transformed or differentiated because it is stationary



```
Observations of Dickey-fuller test
Test Statistic                 -19.018783
p-value                          0.000000
#lags used                       0.000000
number of observations used    364.000000
critical value (1%)             -3.448443
critical value (5%)             -2.869513
critical value (10%)            -2.571018
dtype: float64
```

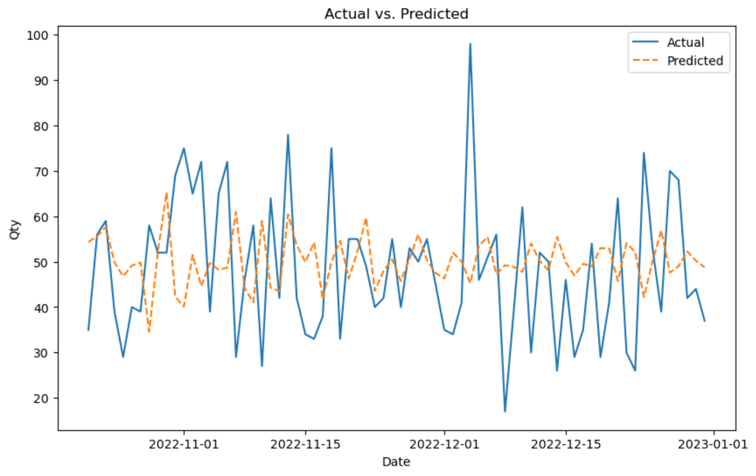
Graph Analysis

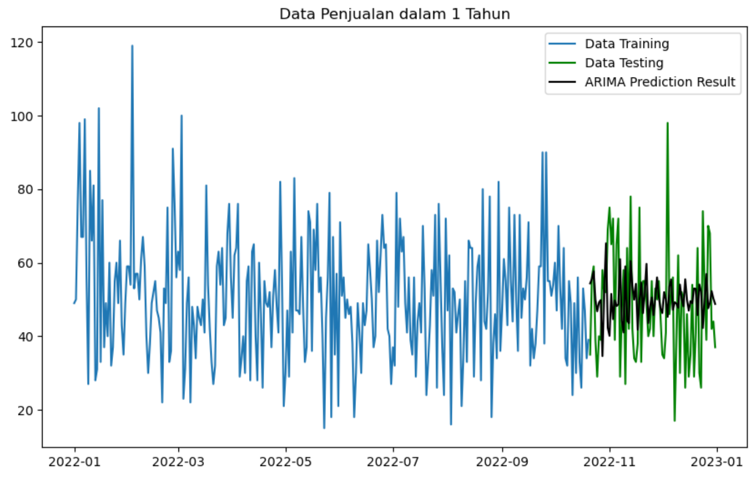
ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots from time series data. The ACF chart will help you identify the Q values, while the PACF chart will help you identify the P values.



Forecasting Result

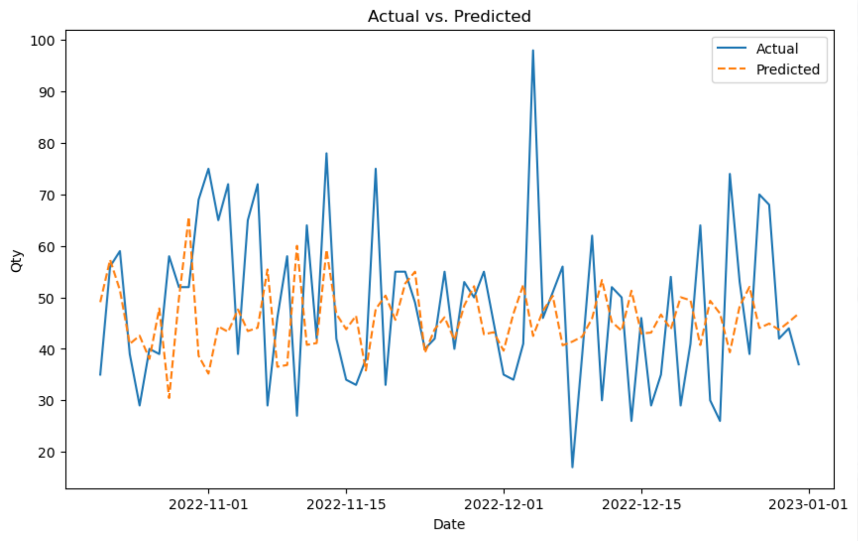
The resulting model for forecasting the daily total quantity of products sold is still far from what was expected. Because the forecasting results do not approach the original data with a Mean Absolute Error (MAE): 14.33.

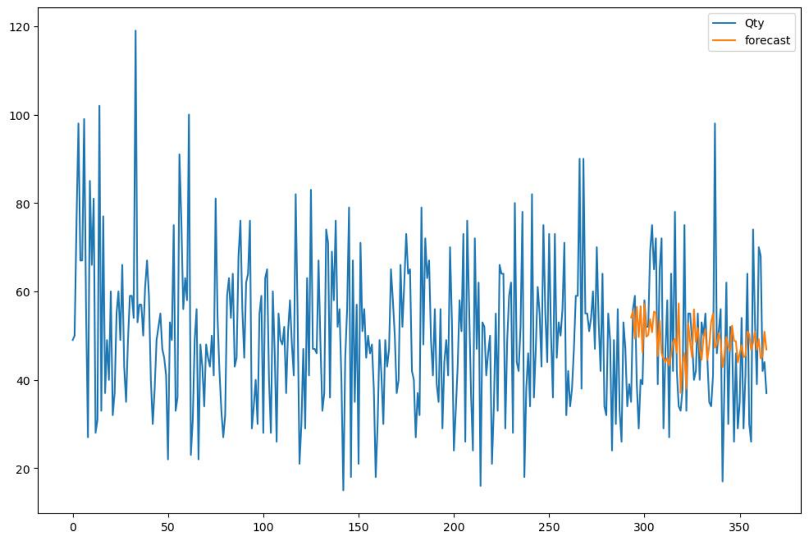




Forecasting Result – Hyperparameter Tuning

Even though hyperparameter tuning has been used, the resulting model still does not have significant changes. However, there was a decrease in the MAE value to 13.66





Conclusion

- Must set the order value in the ARIMA method correctly so that it can produce forecasting values that are close to the original data.
- Identify Seasonality Patterns: If there are seasonal patterns in the data, be sure to pay attention to these patterns in selecting the appropriate model and order values.
- Trying to use other time series methods that might be more suitable for the data used.

Repository

RESULT FOLDER:
https://drive.google.com/drive/folders/1SIqBDEsSpNrHNQbLsSjx66F0_JPYZcsc?usp=sharing

GITHUB:
https://github.com/randayandika/portofolio/tree/main/VIX%20-Kalbe

MY PORTOFOLIO:
https://randayandika.github.io/

Thank You

Rakamin Academy X KALBE Nutritionals