Rakamin Academy X KALBE Nutritionals

Final Project Data Modeling

Kalbe Nutritionals Data Scientist

Presented by
Michael Fernandito Sanfia

Michael Fernandito Sanfia

About You

Lulus dari Universitas Telkom dengan jurusan Teknologi Informasi. Aktif berpartisipasi dalam Organisasi Telkom Purwakarta (TETA) sebagai Tim Minat dan Bakat serta mengorganisir acara TETA Roadshow. Juga, berpartisipasi dalam kompetisi internasional APICTA Awards di bidang ICT. Mahir dalam keterampilan terkait data seperti Analisis Data, Data Analytics, dan Pembuatan Model Machine Learning, serta memiliki keahlian dalam aplikasi seperti Python, Tableau, SQL, dan Microsoft Excel. Berminat untuk mengejar karier sebagai Data Scientist, Data Analyst, dan Analyst.

Rakamin Academy

Pengalaman Pelatihan

- Project-Based Virtual Intern : Data Scientist Kalbe Nutritionals x Rakamin Academy
- Big Data menggunakan Python by Digital Talent Scholarship
- Associate Data Science by Digital Talent Scholarship

Outline

Background	Data Analyst	Data Forcasting and Clustering
Business overview and Project Overview	Data Analysis using Postgre SQL and Tableau	Getting the total daily sales of goods and performing clustering of customers.

Rakamin Academy X KALBE Nutritionals

Background

Business Overview

PT Kalbe Farma Tbk is one of the leading pharmaceutical companies in Indonesia. They are the largest pharmaceutical company in Indonesia in terms of sales and revenue. The company was founded in 1966 and has grown rapidly since then. The company's mission is to improve the quality of life through innovative health products and services.

Project Overview

Rakamin Academy

Inventory Team

- Predicting sales quantities for entire kalbe product In this project
- The purpose of this project is to predict the total sales of products and facilitate the inventory team in managing stock levels.
- Predicting daily product sales ensures that the inventory team has efficiently supplied stock.

Marketing Team

- Clustering customer based on category In this project
- This project aim to establish targeted customer segment.
- These segments will be provided to the marketing team to deliver personalized promotions and sales strategies based on customer segmentation.

Tools

DBeaver

tableau

ANACONDA

Rakamin Academy X KALBE Nutritionals

Data Analyst

Query 1

```
select case
   when "Marital Status"='Married' then 'Menikah'
   when "Marital Status"='Single' then 'Belum Menikah'
   else 'Tidak diketahui'
End as "Status Pernikahan", avg(age) as "Rata Rata usia"
from customer
group by "Marital Status"
```

Result:

	Status Pernikahan	Rata Rata usia
1	Tidak diketahui	31.3333333333
2	Menikah	43.0382352941
3	Belum Menikah	29.3846153846

Conclusion :

Based on the results of query 1, it was found that the average age based on marital status is as follows:

- Married status has the oldest age compared to other categories, with an average age of 43 years.
- Unmarried status has the youngest age compared to other categories, with an average age of 29 years.
- Status that is unknown has an average age of 31 years.

Query 2

```
select case
   WHEN gender = 0 THEN 'Wanita'
   WHEN gender = 1 THEN 'Pria'
end as "Jenis Kelamin" ,avg(age) as "Rata Rata usia"
from customer
group by gender
```

Result:

	Jenis Kelamin	Rata Rata usia
1	Wanita	40.326446281
2	Pria	39.1414634146

Conclusion :

Based on the results of query 2, it was found that the average age based on gender is as follows:

- Woman category has the oldest age compared to other categories, with an average age of 40 years.
- Men category have younger age compared to woman categories, with an average age of 39 years.

Query 3

```
select store.storename as "Nama Toko",
sum(transaction.qty ) as "Jumlah Quantity"
from store
inner join transaction
on store.storeid=transaction.storeid
group by "Nama Toko"
order by "Jumlah Quantity" desc
```

Result:

Nama Toko	Jumlah Quantity
Lingga	2,777
Sinar Harapan	2,588
Prestasi Utama	1,395
Prima Kota	1,358
Buana	1,320
Prima Tendean	1,310
Prima Kelapa Dua	1,296
Harapan Baru	1,286
Bonafid	1,283
Priangan	1,239
Gita Ginara	1,236
Buana Indah	1,208

Conclusion :

Based on the results of query 3, it was found that the most quantity sale are lingga Store with an amount of 2,777 has been sold on that store

Query 4

```
select product."Product Name" as "Nama Produk" ,
sum(transaction.totalamount ) as "Total Amount"
from product
inner join transaction
on product.productid=transaction.productid
group by "Nama Produk"
order by "Total Amount" desc
```

Result:

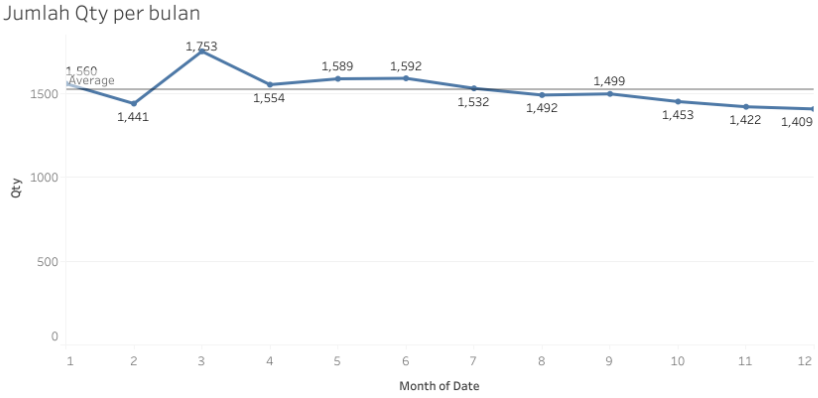
Nama Produk	Total Amount
Cheese Stick	27,615,000
Choco Bar	21,190,400
Coffee Candy	19,711,800
Yoghurt	19,630,000
Oat	15,440,000
Crackers	13,680,000
Potato Chip	13,104,000
Thai Tea	11,982,600
Cashew	11,286,000
Ginger Candy	8,403,200

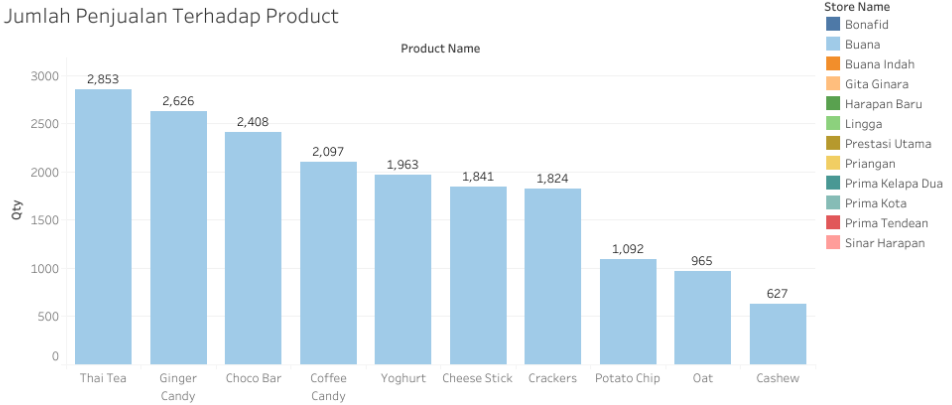
Conclusion :

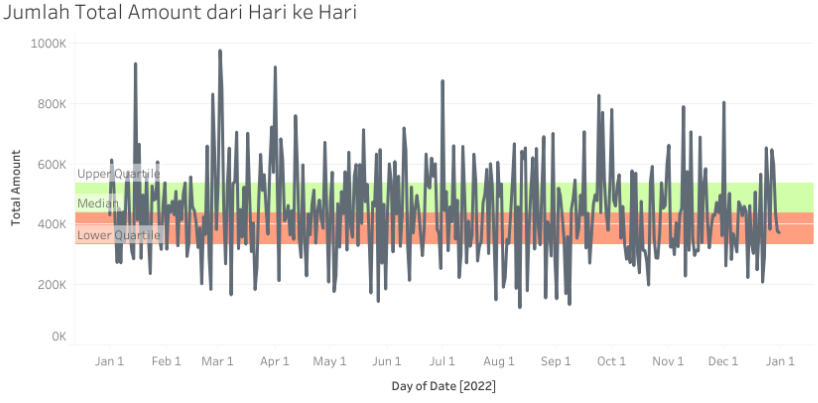
Based on the results of query 4, it was found that the most amount sale on the product are Cheese Stick with an total amount of 27,615,000 has been sold on that product

Tableau Result

Rakamin Academy







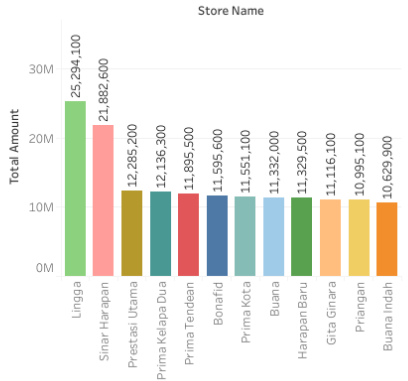


Tableau Result Explanation

- From the first graph, it can be concluded that from January to December 2022, the total sales of goods did not show significant changes each month, with an average of 1500 items sold per month.
- From the second graph, it can be concluded that Thai tea is the best-selling product compared to other products, with a total sales of 2853, so the stock of Thai tea products should be closely monitored to ensure availability.
- From the third graph, it is found that the total amount per day fluctuates compared to other graphs, with the upper quartile around 580-430 thousand, the median at 430 thousand, while the lower quartile is around 430-350 thousand.
- Furthermore, from the last graph, a map is used to see the distribution of stores, and Lingga store has the highest total amount with 25,294,100.

Tableau Dasboard Link:
https://public.tableau.com/app/profile/michael.sanfia/viz/StudiKasus_16964437908120/Dashboard1?publish=yes

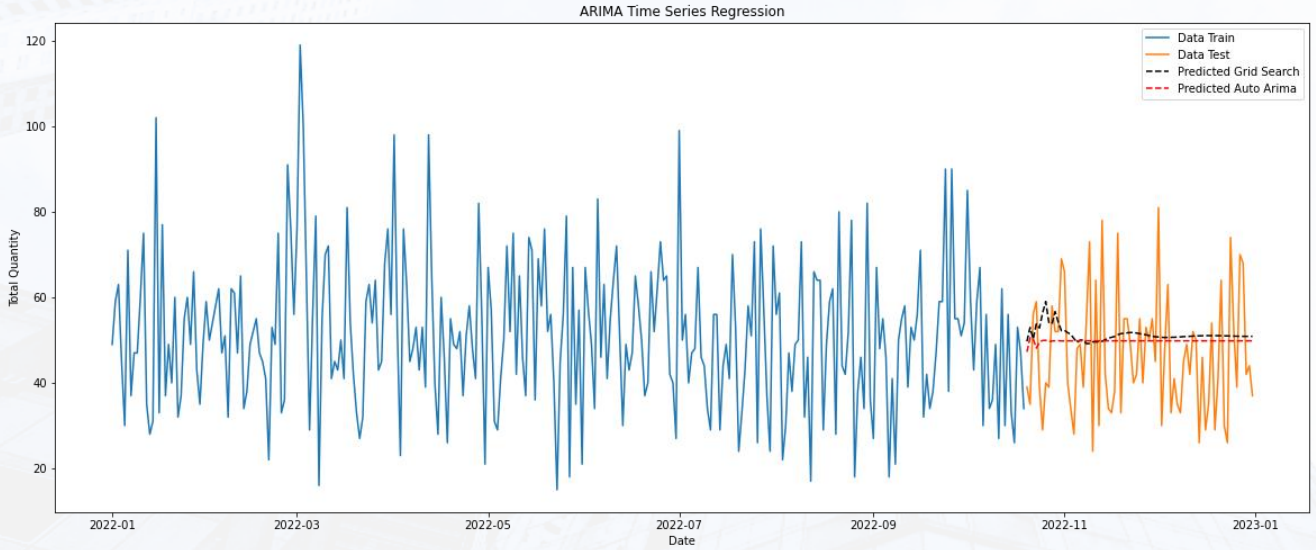
Rakamin Academy X KALBE Nutritionals

Data Forcasting and Clustering

Proces Data Forcasting

Data Undestanding	Data Preprocesing	Data Forcasting
Understanding the data characteristics, trends, seasonal patterns, and other factors that may influence data behavior.	• Addressing missing values. • Handling outliers. • Performing imputation or interpolation processes if necessary.	• Training the model using the training data to understand patterns and trends in the data. • Using the testing data to validate the model's performance and ensure that the model generalizes well to new data.

Result Model Forcasting



Performance Result

Grid Search

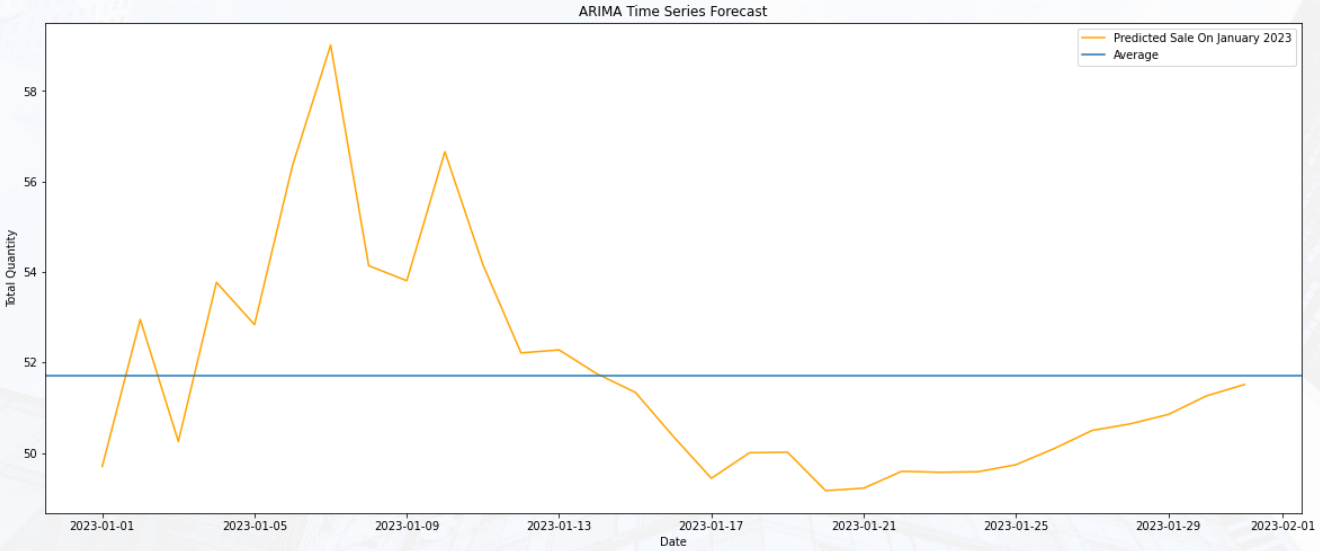
- MAE: 12.78
- RMSE: 15.21
- MAPE: 32.02%

Auto ARIMA

- MAE: 11.86
- RMSE: 14.17
- MAPE: 29.39%

The above is a chart of the ARIMA algorithm's prediction results for the total quantity each day. The data is divided into two parts: the training data and the testing data. The ARIMA prediction results are obtained with the dashed black line representing the result of ARIMA prediction using the best algorithm from grid search. Meanwhile, the dashed red line represents the result of the best ARIMA algorithm using auto ARIMA.

Forcasting The Next 1 Month

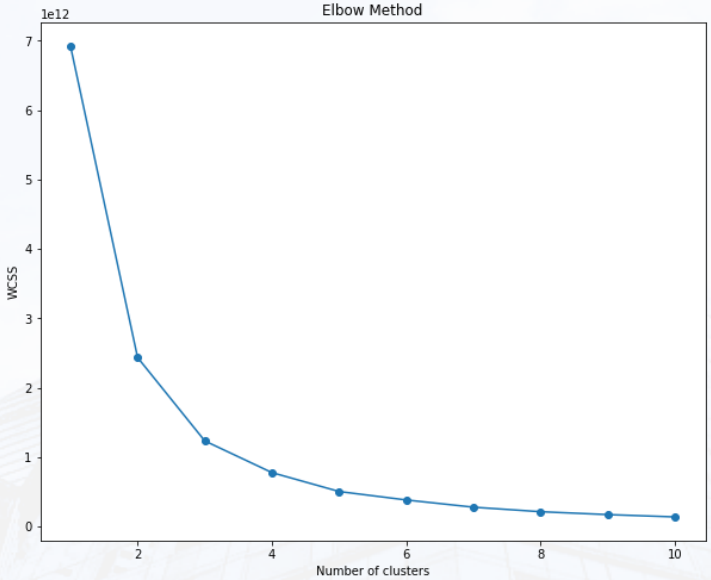


Above is the graph of the forecasted results of the product for the upcoming month, namely January 2023. As seen in the image, the predicted average for January 2023 is 51.7 product sales per day. Therefore, inventory management for the next month should align with the forecasted results.

Proces Data Clustering

Data Undestanding	Data Preprocesing	Data Clustering
Understanding the data characteristics that may influence data behavior.	• Addressing missing values. • Handling outliers. • Performing imputation or interpolation processes if necessary. • Aggregating the features that will be used	In this process, the determination of the cluster count will be conducted using the elbow method. After obtaining the optimal cluster count, cluster modeling will be carried out using the determined number of clusters.

Elbow Method



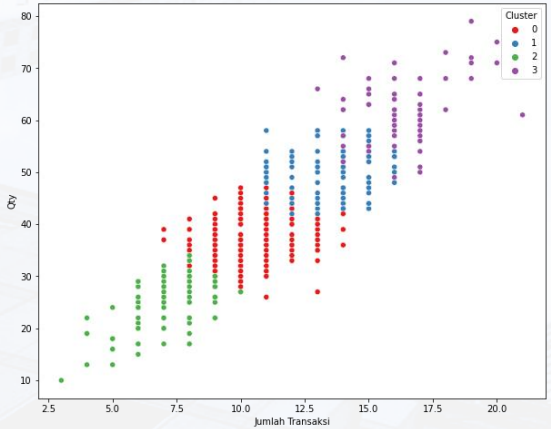
From the elbow method graph above, it can be concluded that 4 clusters are the best results because the graph after it is very sloping while the graph in cluster 3 to 4 is still quite steep.

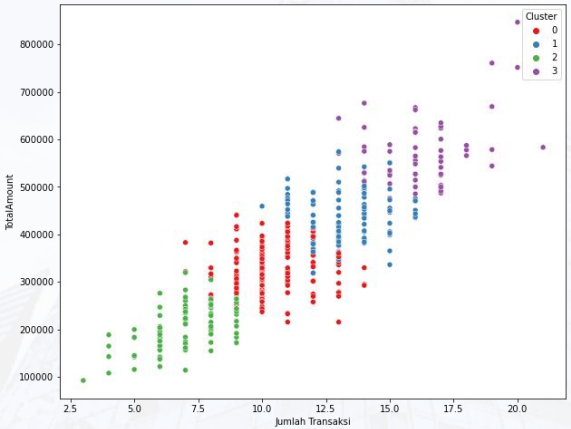
Result Clustering

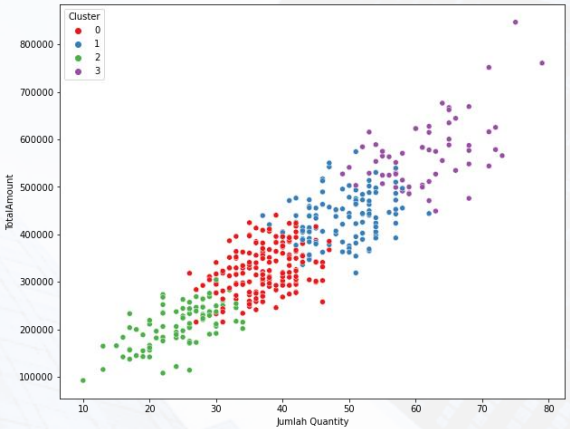
Rakamin Academy

Number Cluster	Transaction Average	Qty Average	Total Amount Average	Total Customer
0	10	37	325,663.33	180
1	13	49	436,203.50	114
2	7	25	208,283.87	93
3	16	62	572,100	60

Result Clustering







Marketing Strategies

Cluster 0 (Moderate Transactions, Moderate Quantity, Moderate Total Amount):

•Offer bundling deals or discounts with a certain quantity purchase to encourage volume.

Build brand awareness with effective and relevant advertising strategies.

Cluster 1 (High Transactions, High Quantity, High Total Amount):

•Provide incentives or loyalty programs to retain customers contributing significantly to sales.

•Consider offering premium products or services and support them with excellent customer service.

Marketing Strategies

Cluster 2 (Low Transactions, Low Quantity, Low Total Amount):

•Focus on increasing transaction frequency by providing discounts or special offers for repeat purchases.

•Offer bundling packages or additional products to boost transaction value.

Cluster 3 (Very High Transactions, Very High Quantity, Very High Total Amount):

•Develop exclusive loyalty programs to support loyal customers in this category.

•Focus on premium product or service development that aligns with customer preferences.

Thank You