Rakamin Academy

KALBE Nutritionals

Final Project Data Scientist VIX

Product & Customer Analysis

Time Series Forecasting - Customer Segmentation

Wanda Listathea Putri

- July 2023 -

Wanda Listathea Putri

LINKS

LinkedIn – Github - Medium

Previous Projects

- Credit Scoring Modeling - Python (2023)
  Link: Credit Scoring
- Salary Prediction Modeling and Deployment – Python, Tableau (2022)
  Link: Salary Prediction
- Sentiment Analysis - Python (2021)
  Link: Sentiment Analysis

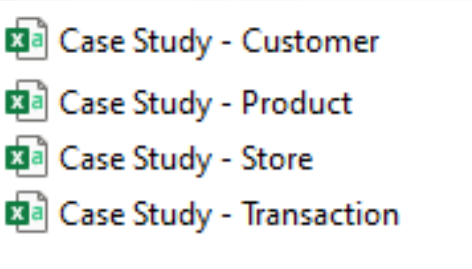
Case Study

Getting a new project from Inventory Team and Marketing Team

1. Inventory Team
   - Project: Make a prediction of the total quantity of all products
   - Goal: To find out the estimated number of products sold so the team can make a sufficient daily stock

2. Marketing Team
   - Project: Make a customer segmentation based on some criteria
   - Goal: To give personalized promotion and sales treatment

Dataset




```
customer: (447, 5)
store: (14, 6)
product: (10, 3)
transaction: (5020, 8)
```

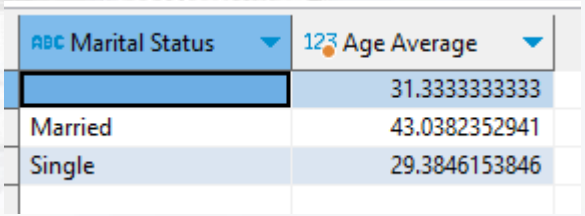
Transaction data is just for a year, 2022

Exploratory Data Analysis

Using Dbeaver (PostgreSQL) to perform the query

- Query 1: What is the average age of the customer based on marital status?
- Query 2: What is the average age of the customer based on gender?
- Query 3: Determine the store name with the highest total quantity!
- Query 4: Determine the product name with the highest total amount!

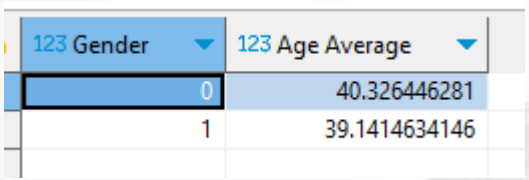
QUERY 1



Marital Status	Age Average
	31.3333333333
Married	43.0382352941
Single	29.3846153846

The average age of the customer based on marital status are 43 years old for the married customer and 29 years old for the single customer (3 data are null)

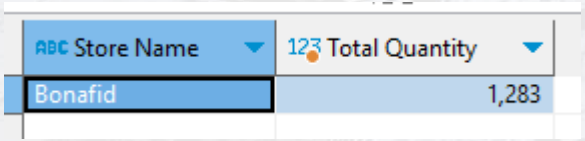
QUERY 2



Gender	Age Average
0	40.326446281
1	39.1414634146

The average age of the customer based on their gender are 40 years old for women and 39 years old for men

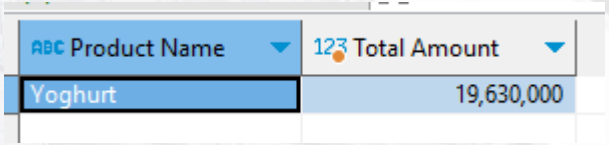
QUERY 3



Store Name	Total Quantity
Bonafid	1,283

The store with the highest total quantity is Bonafid store which sold 1,283 products in 2022

QUERY 4



Product Name	Total Amount
Yoghurt	19,630,000

The product with the highest total amount is yoghurt which earns 19,6 million in 2022

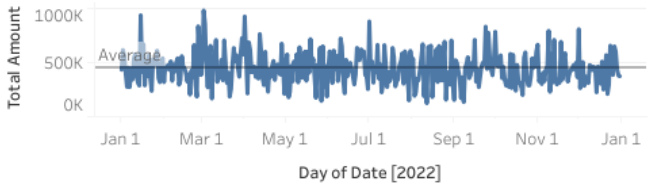
Dashboard

Using Tableau to make a dashboard

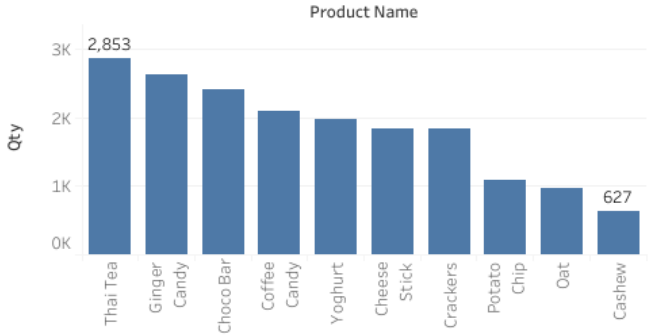
- Worksheet 1: The month-to-month total quantity
- Worksheet 2: The day-to-day total amount
- Worksheet 3: The quantity by product
- Worksheet 4: The total amount by store name

The Month-to-Month Total Quantity

The Day-to-Day Total Amount

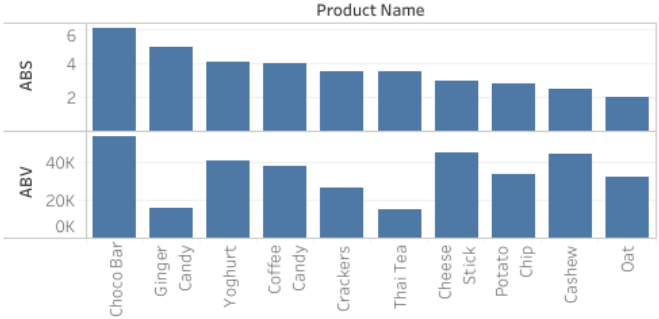


The Quantity by Product

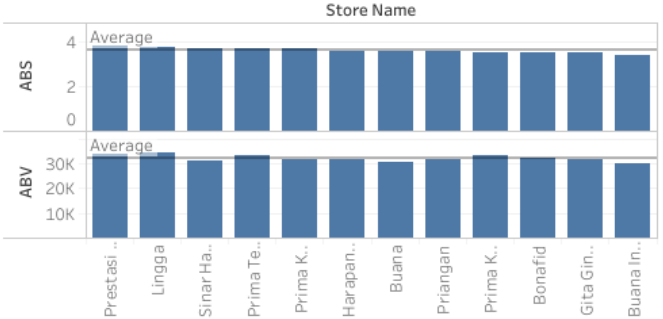


The Total Amount by Store Name

ABS vs ABV by Product



ABS vs ABV by Store

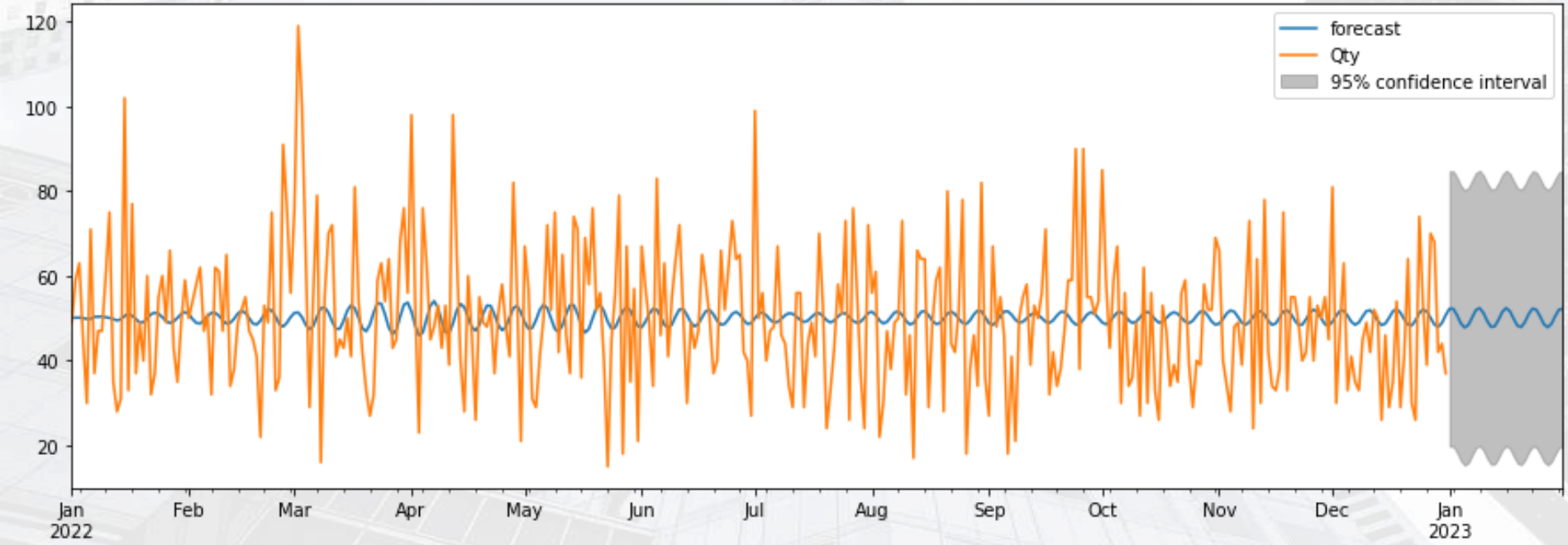


Time Series Forecasting

- Goal: To predict the daily total quantity of products sold
- Create new data for regression (group by Date and aggregated is Qty in sum)
- Method: ARIMA

Use ARIMA(2,0,2)

- AIC Score: 2474.86
- RMSE: 15.05



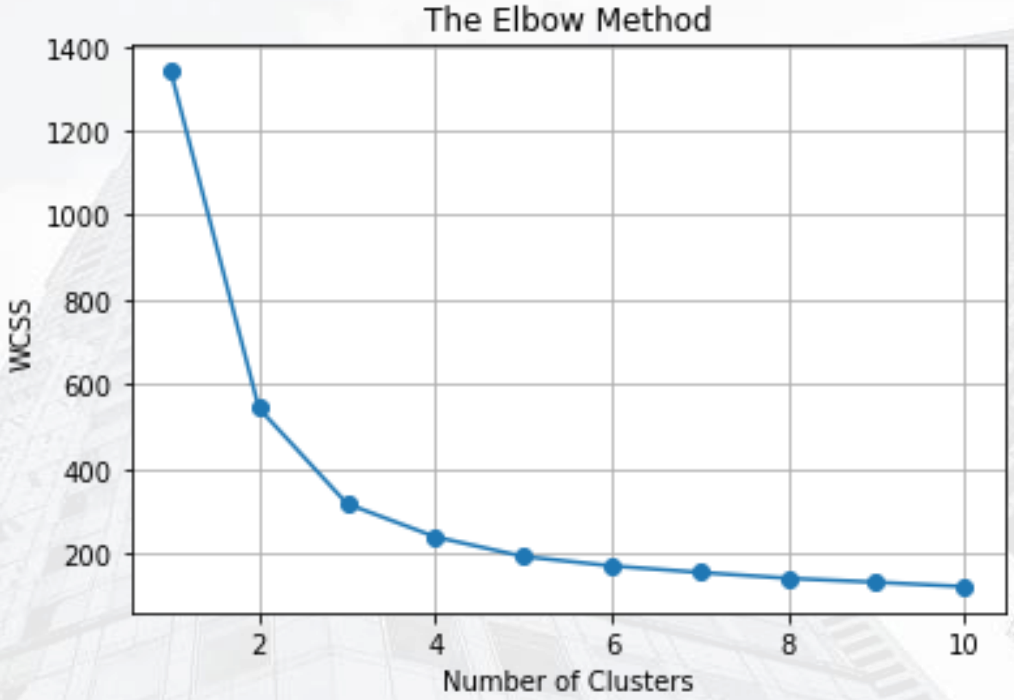
The prediction of quantity products for the next 30 days

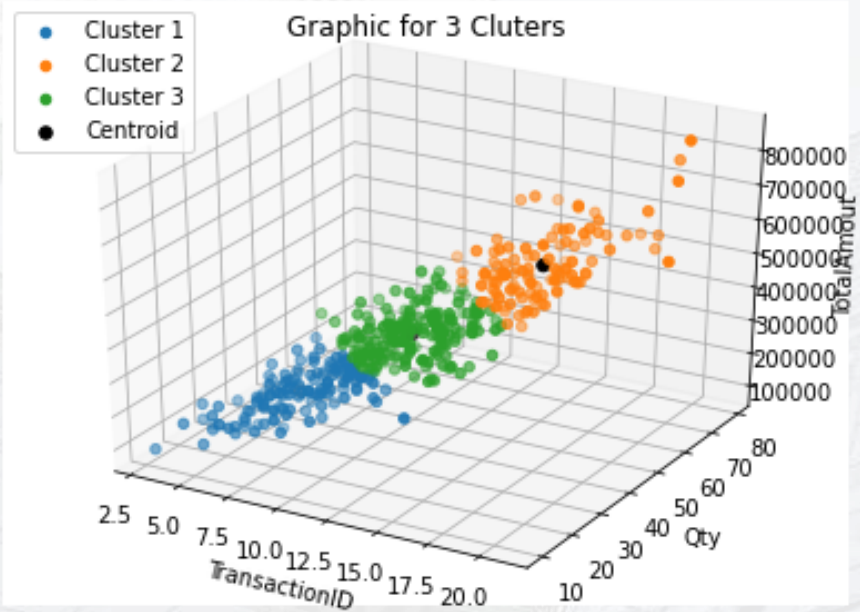
	Qty
2023-01-01	52.0
2023-01-02	52.0
2023-01-03	51.0
2023-01-04	49.0
2023-01-05	48.0
2023-01-06	48.0
2023-01-07	50.0
2023-01-08	52.0
2023-01-09	52.0
2023-01-10	51.0
2023-01-11	49.0
2023-01-12	48.0
2023-01-13	48.0
2023-01-14	50.0
2023-01-15	52.0
2023-01-16	52.0
2023-01-17	52.0
2023-01-18	50.0
2023-01-19	48.0
2023-01-20	48.0
2023-01-21	49.0
2023-01-22	51.0
2023-01-23	52.0
2023-01-24	52.0
2023-01-25	50.0
2023-01-26	49.0
2023-01-27	48.0
2023-01-28	49.0
2023-01-29	50.0
2023-01-30	52.0

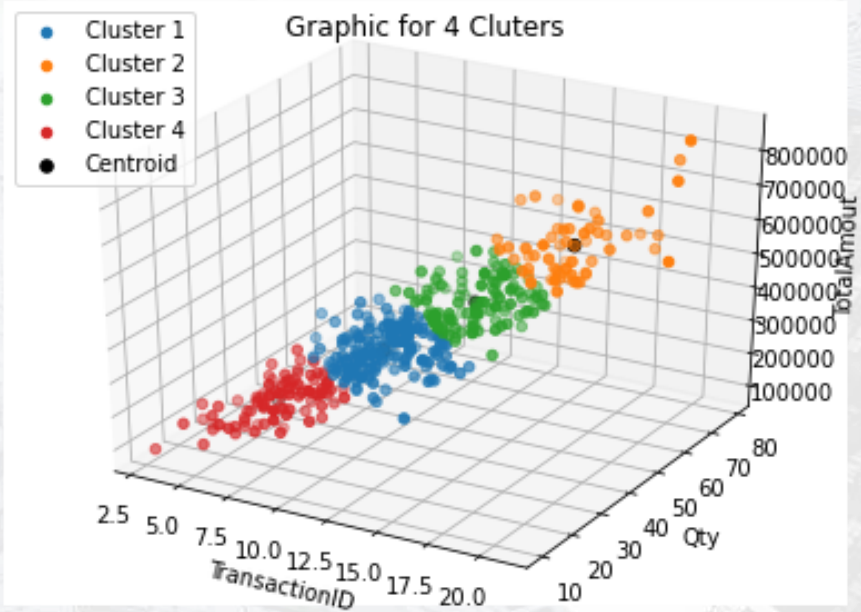
Customer Segmentation

- Goal: To create clusters of similar customers
- Create new data for clustering (group by CustomerID and aggregated are TransactionID in count, Qty in sum, and TotalAmount in sum)
- Method: Kmeans Clustering

Consider to use k=3 or k=4




Graphic for 3 Cluters
Cluster 1
Cluster 2
Cluster 3
Centroid
TransactionID
Qty
TotalAmount


Graphic for 4 Cluters
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Centroid
TransactionID
Qty
TotalAmount

Centroids for k=3

	TransactionID_centroid	Qty_centroid	TotalAmount_centroid
Cluster 1	7.785185	26.933333	229388.888889
Cluster 2	15.363636	57.636364	524504.545455
Cluster 3	11.282178	41.188119	363267.326733

Centroids for k=4

	TransactionID_centroid	Qty_centroid	TotalAmount_centroid
Cluster 1	10.427778	37.350000	325663.333333
Cluster 2	16.316667	61.650000	572100.000000
Cluster 3	13.254386	49.078947	436203.508772
Cluster 4	7.021505	24.505376	208283.870968

Link

Tableau Public
Github

Rakamin Academy

KALBE Nutritionals

Final Project Data Scientist VIX

Thank You

Wanda Listathea Putri

- July 2023 -