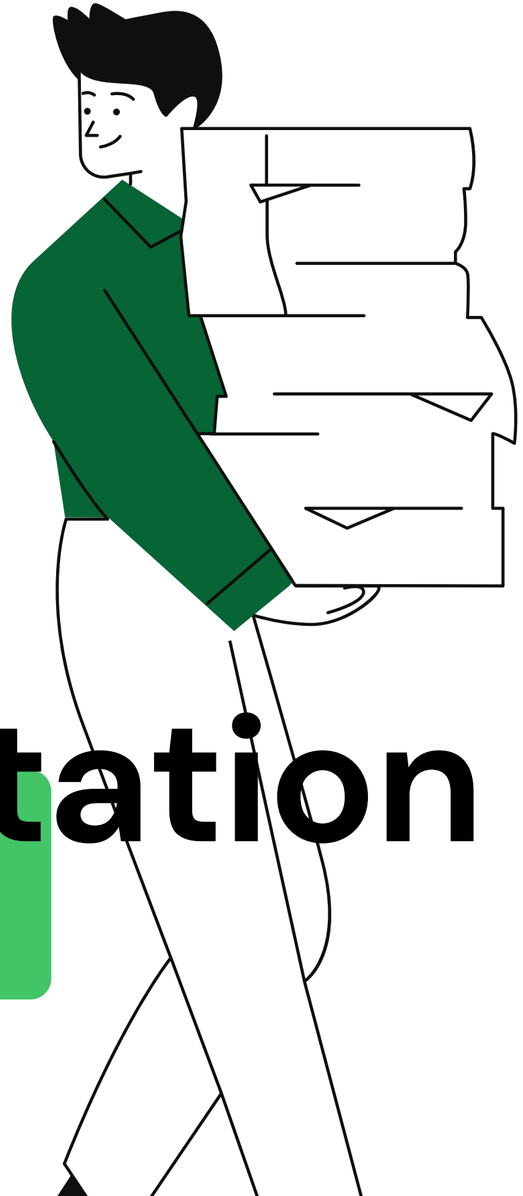
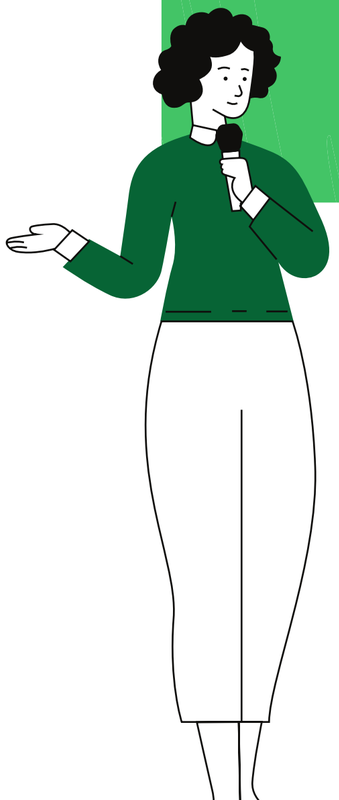
KALBE Nutritionals x Rakamin Academy

July 2023

Predict Daily Stock *and* Customer Segmentation

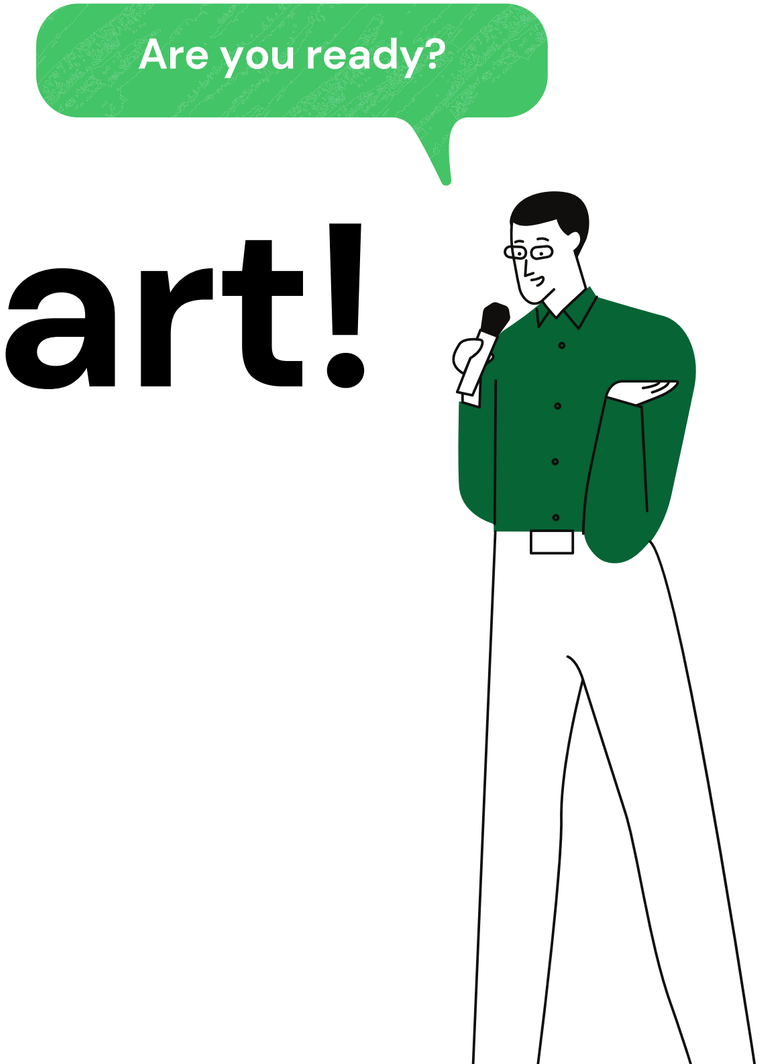


Today's Agenda



1. Introduction
2. Exploratory Data Analysis
3. Daily Stock Prediction
4. Customer Clustering

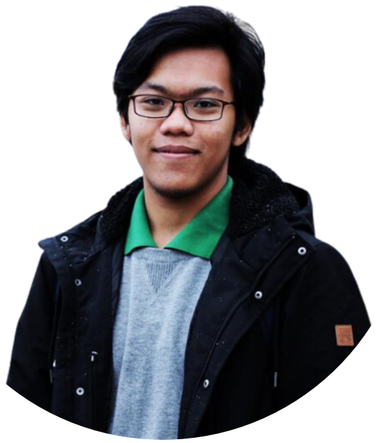
Let's Start!



Introduction

I am **Vendi**, as a Data Scientist at Kalbe Nutritionals.

This project helps the **inventory team in predicting daily stock needs** and also helps the **marketing team group customers** into several segments/clusters to increase the effectiveness of marketing activities.



Vendi

Dataset

Data contains 10 products, 14 stores, 447 customers and 5020 transactions made in 2022.

The dataset consists of 4 tables.

Tables

Customer Table

1. customerid : No Unik Customer
2. age : Usia Customer
3. gender : 0 – Wanita, 1 – Pria
4. maritalstatus : Married, Single (Blm menikah/Pernah menikah)
5. income : Pendapatan per bulan (juta rupiah)

Store Table

1. storeid : Kode Unik Store
2. storename : Nama Toko
3. groupstore : Nama group
4. typestore : Modern Trade, General Trade
5. latitude : Kode Latitude
6. longitude : Kode Longitude

Product Table

1. productid : Kode Unik Product
2. productname : Nama Product
3. price : Harga dlm rupiah

Transaction Table

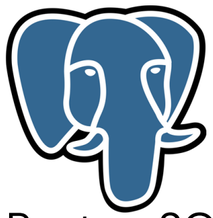
1. transactionid : Kode Unik Transaksi
2. datetransaction : Tanggal transaksi
3. qty : Jumlah item yang dibeli
4. totalamount : Price x Qty

Kalbe Nutritionals x Rakamin Academy

July 2023

Exploratory Data Analysis!

Exploratory Data Analysis with SQL



PostgreSQL

x

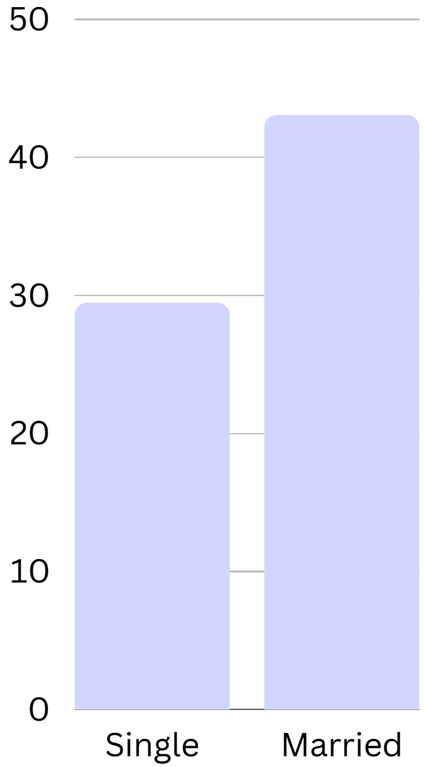
DBeaver

Click here for SQL Query Details

1. Average age of the customers per marital status.
2. Average age of the customers per gender.
3. Store name with the highest total quantity.
4. Best-selling product with the highest total amount.

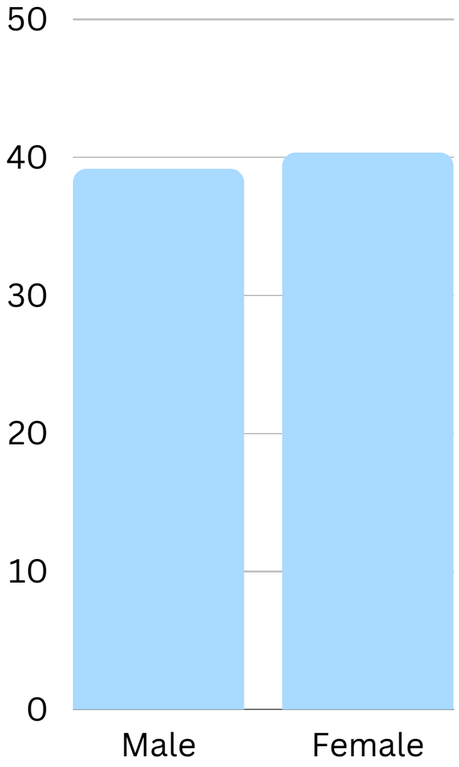
1

Average age per marital status.



2

Average age per gender.



3

Store with the highest total quantity.

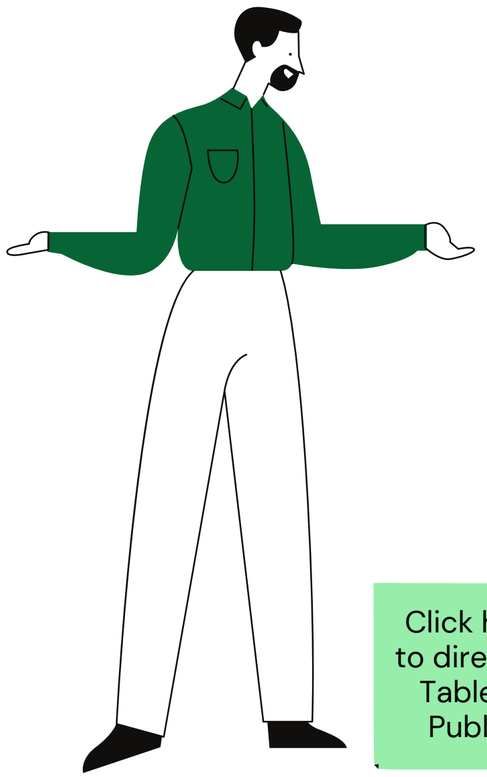
ID	Store Name	Qty
9	Lingga	370
6	Lingga	368
13	Buana	368

4

Product with the highest total amount.

ID	Product Name	Total Amount
P10	Cheese Stick	27.615.000
P1	Choco Bar	21.190.400
P7	Coffee Candy	19.711.800

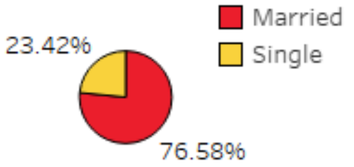
Dashboard



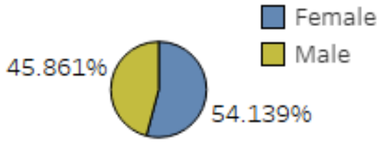
Click here to direct to Tableau Public!

Average Age : 39.78

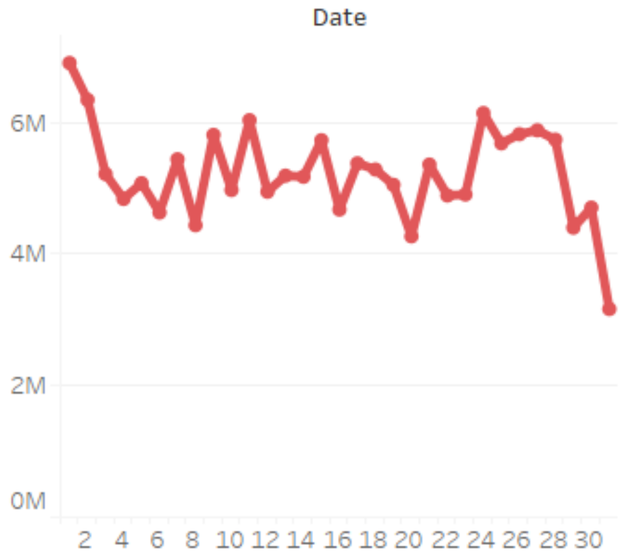
Marital Status Customer



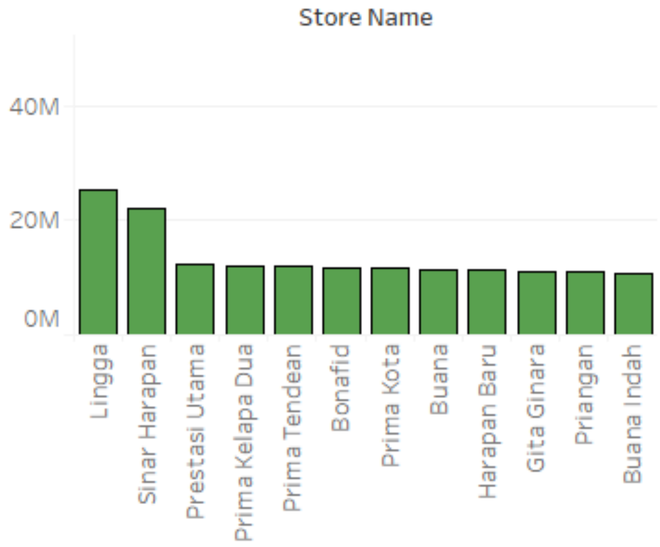
Gender Customer



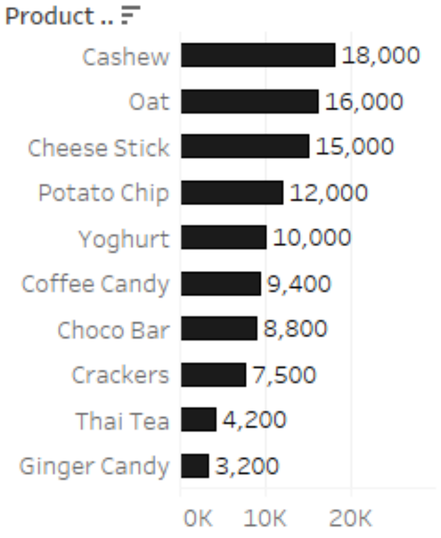
Daily Total Amount



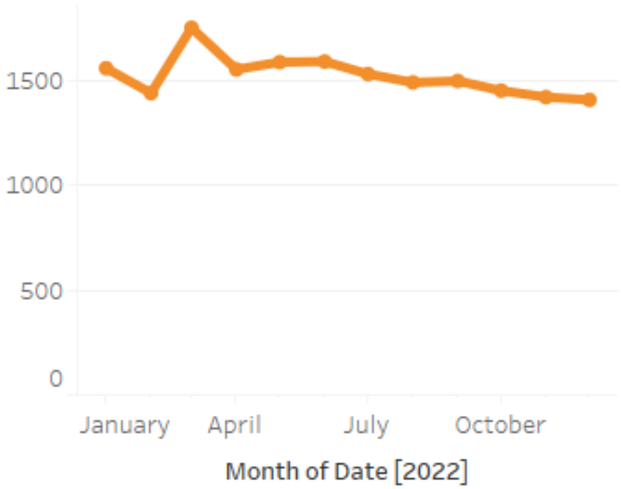
Total Amount per Store



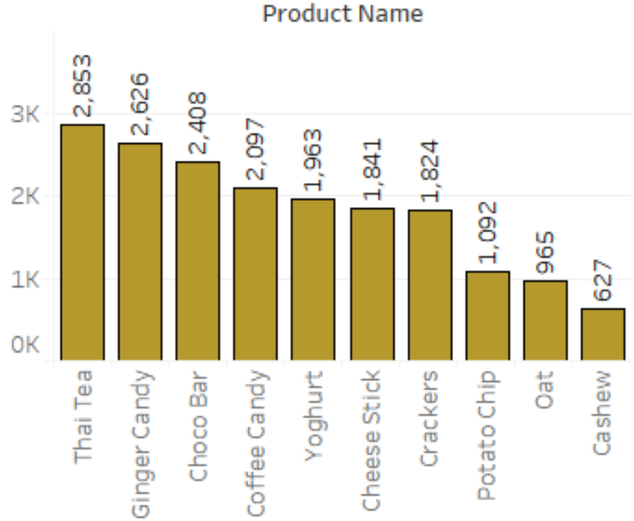
Price of Product



Transaction Quantity per Months



Total Quantity per Product



Daily Stock Prediction!

Data Cleansing

All tables were merged using SQL before.

The results of this stage will be used in the process of predicting daily stock and customer clustering.

python™ x 

The Source Code is here!

1. No **Duplicated Data**
2. No **Missing Values**
3. **Data Type Adjustment** - adjust data type of column 'datetransaction' from 'object' to 'datetime'. Then make new column, 'day', which contains information on what day in 2022 (1-365).
4. **Feature Selection** - *drop useless columns, such as 'storeid.1', 'latitude', 'longitude', 'customerid.1', 'productid.1' and 'price.1'*

Machine Learning

The Source Code is here!

Pre-processing

Create new data :

```
1 ml_regr = data_ml_regr.groupby(['day']).agg({'qty' : 'sum', 'transactionid' : 'count'}).reset_index()
2 ml_regr.columns = ['day', 'total_qty', 'trx_count']
```

Modelling

```
1 # adjusting the regression line with the distribution of the data (fit)
2 lr = LinearRegression()
3 lr.fit(X_train, y_train)
```

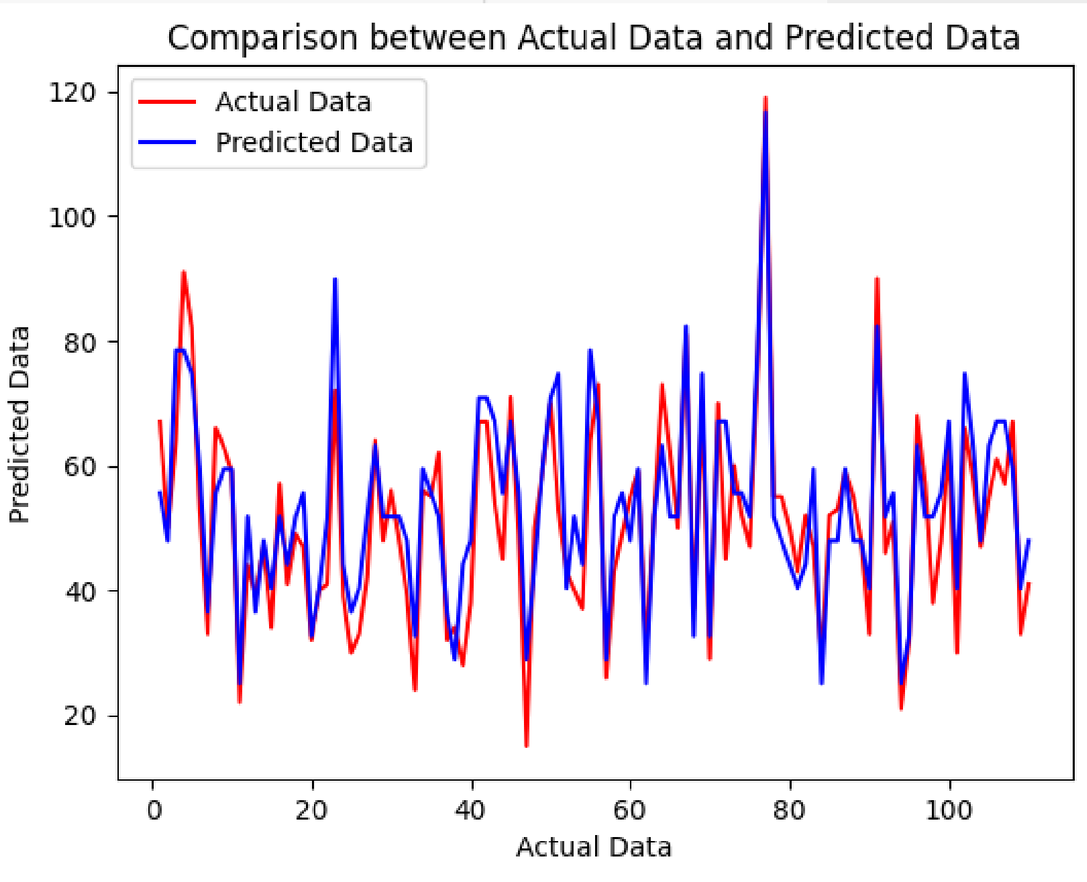
```
▾ LinearRegression
LinearRegression()
```

```
1 # the model makes predictions
2 y_pred = lr.predict(X_test)
```

Evaluation

```
r2 score : 0.7801872762442322
mean squared error : 58.095885233062006
```

```
Intercept of the model :  -1.6236995752724042
Coefficient of the model :  [3.81464395]
```



Conclution and Recomendation

1. **Model regresi linear: y = 3.815x - 1.624.**
   The linear equation can be used in the future to build a model that can predict daily stock requirements.

2. To predict with bigger number of products (or transactions), we can do **modeling for each product**. More detailed prediction will be more effective in preparing stocks of each product for daily needs.

Customer Clustering!

Machine Learning

The Source Code is here!

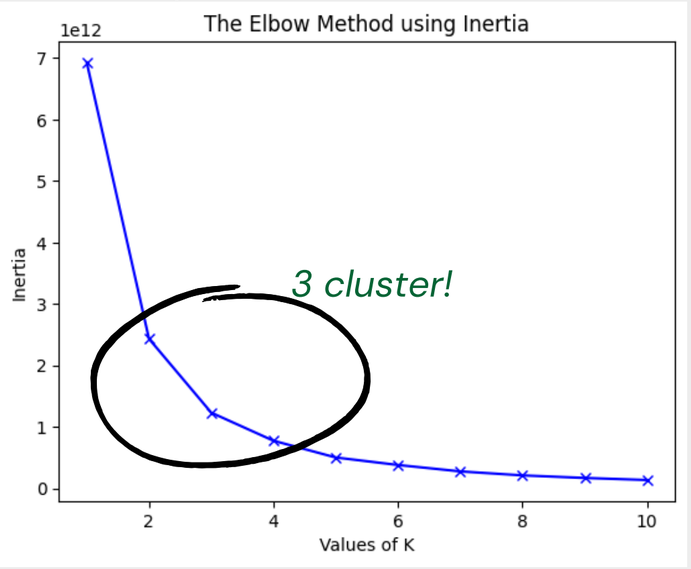
Pre-processing

Clustering with 3 criteria: total transaction, total quantity and total spending amount

```
ml_clus = data_ml_clus.groupby(['customerid']).agg({'transactionid' : 'count',
                                                    'qty' : 'sum',
                                                    'totalamount' : 'sum'
                                                    }).reset_index()
ml_clus.columns = ['customer_id', 'trx_count', 'total_qty', 'total_spent']
```

Modelling

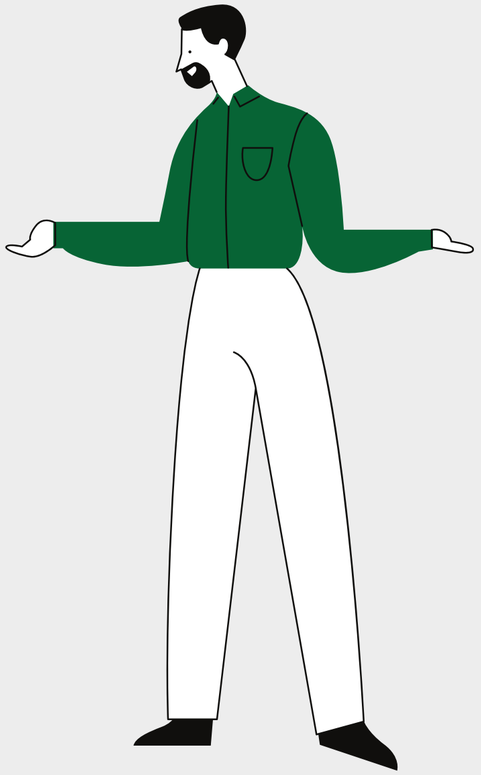
Find the optimal number of clusters :



Modelling with 3 cluster :

```
# modeling with 3 cluster
kmeans = KMeans (n_clusters = 3, init='k-means++',
                 n_init=10, max_iter=300,
                 tol=0.0001, random_state=100)
kmeans.fit(X.values)
X['cluster'] = kmeans.labels_
```

cluster	total_customers
0	191
2	162
1	94



Cluster Analysis

Cluster

High Spender

Cluster 1 has highest average transactions, highest average product purchase quantity and largest average total purchases.

Mid Spender

Customers in cluster 0 make transactions in a fairly rare amount, although not as many as cluster 1 but not the smallest.

Low Spender

Cluster has the smallest transactions, quantities and purchases compared to other groups.

Recomendation

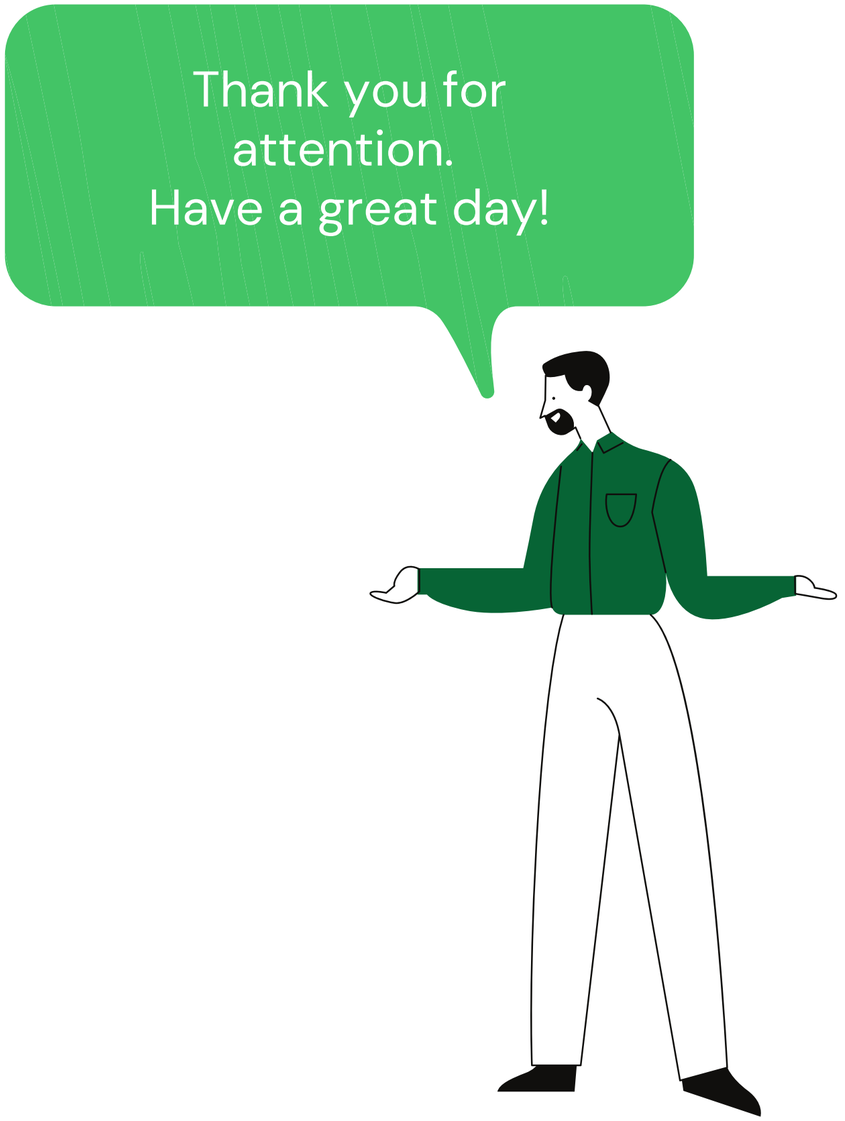
1. Monitoring transactions and retention from the **High Spender** group, can **improve service** so that these groups do not churn in the future.

2. For the **Mid Spender** group, further analysis can be carried out on how to increase transactions by **providing more personal recommendations**, as well as deeper analysis on how to optimize promos in this segment and keep shopping.

3. For the **Low Spender** group, further analysis can also be carried out on how to increase the **desire to make transactions**. This can be caused by products or prices that do not match.

KALBE Nutritionals x Rakamin Academy

July 2023



That's all!

Final Result is here!