

Sales Prediction and Customer Segmentation

VIX Data Scientist Kalbe Nutritionals

Presented by
Giselle Halim

[LinkedIn](#)

Case Study

About Kalbe Nutritionals

KALBE Nutritionals is a company engaged in the health food and beverage business that has a mission of **Providing Best Nutrition Solution for a Better Life**. This mission is a firm statement of KALBE Nutritionals in ensuring that all products will be developed with full responsibility to bring the goodness of life to more people, in Indonesia and around the world. KALBE Nutritionals has various nutritional products at every stage of life, starting from pregnancy preparation, pregnancy, nutrition for babies, children, adults, to the elderly.

Case Study

- Create SQL queries to see the average age of customers when viewed from marital status and gender, the store with the highest total quantity, and the best-selling product with the highest total amount.
- Create a dashboard with Tableau containing monthly quantity, daily total amount, sales (quantity) per product, and sales (total amount) per store.
- Create a model to predict the quantity of products sold so that the inventory team can make sufficient daily inventory stock.
- Create customer segmentation that will be used by the marketing team to provide personalized promotion and sales treatment.

Tools

- Python
- SQL
- Google Colab
- Tableau
- MySQL Workbench
- Microsoft Excel

Data Dictionary

This dataset consists of 4 csv files namely customer, store, product and transaction. It is a dummy data for FMCG case study within 1 year taken through membership program.

1. Customer

- **CustomerID** : Customer unique ID
- **Age** : Customer age
- **Gender** : “0” Female, “1” Male
- **Marital Status** : Married, Single
- **Income** : Income per month in million IDR

2. Store

- **StoreID** : Store unique ID
- **StoreName** : Store name
- **GroupStore** : Group name
- **Type** : Modern Trade, General Trade
- **Latitude** : Latitude code
- **Longitude** : Longitude code

3. Product

- **ProductID** : Product unique ID
- **Product Name** : Product name
- **Price** : Price in IDR

4. Transaction

- **TransactionID** : Transaction unique ID
- **Date** : Transaction date
- **Qty** : Quantity of items purchased
- **Total Amount** : Price x Qty

Results

SQL Query

- Average age of customers when viewed from marital status and Gender

MaritalStatus	Average_Age
Married	43.0382
	31.3333
Single	29.3846

	Gender	Average_Age
▶	Male	39
	Female	40

- Stores with the highest total quantity

StoreID	StoreName	Total_Qty
9	Lingga	1439
12	Prestasi Utama	1395
3	Prima Kota	1358
6	Lingga	1338
11	Sinar Harapan	1331
13	Buana	1320

- Product with the highest total amount

ProductID	ProductName	Total_Amount	Total_Qty
P10	Cheese Stick	27615000	1841
P1	Choco Bar	21190400	2408
P7	Coffee Candy	19711800	2097
P9	Yoghurt	19630000	1963
P8	Oat	15440000	965
P3	Crackers	13680000	1824

SQL Query Insight

- Tailor promotions and product recommendations based on marital status. Married couples might be interested in family-oriented bundles, while singles might prefer individual-sized products.
- Leverage the age average difference between married and single customers. For example, promote age-appropriate health products to each segment.

Dashboard Tableau

[Link Dashboard](#)



KALBE NUTRITIONALS 2022 SALES DASHBOARD

Total Amount
Rp162,043,000

Total Qty
18,296

Unique Transactions
4,908

Avg Total Amount
Rp32,279

Date
1/1/2022 to 12/31/2022

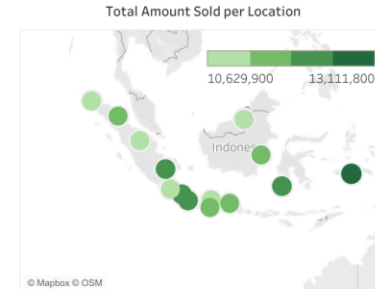
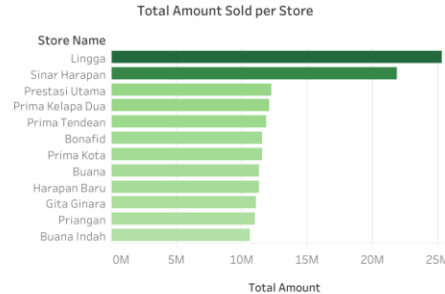
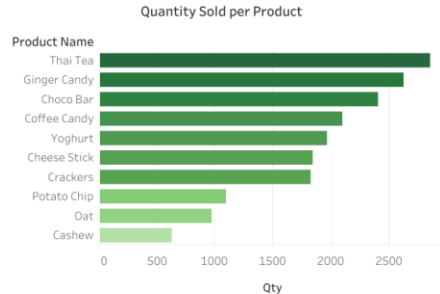


- Month
- ☒ January
 - ☒ February
 - ☒ March
 - ☒ April
 - ☒ May
 - ☒ June
 - ☒ July
 - ☒ August
 - ☒ September
 - ☒ October
 - ☒ November
 - ☒ December

Group Store
All

Store Name
All

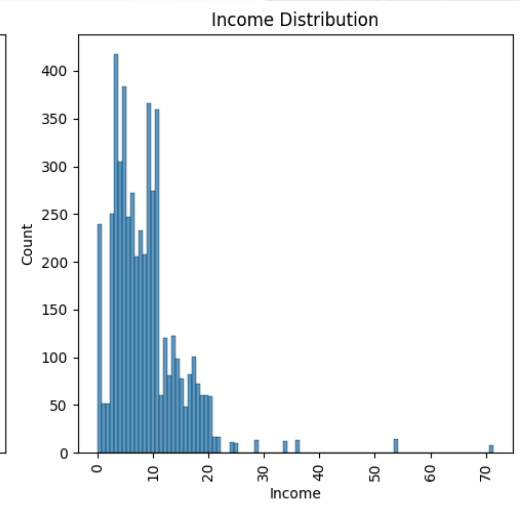
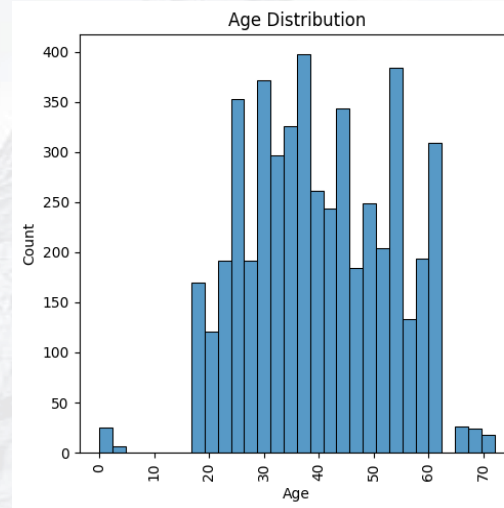
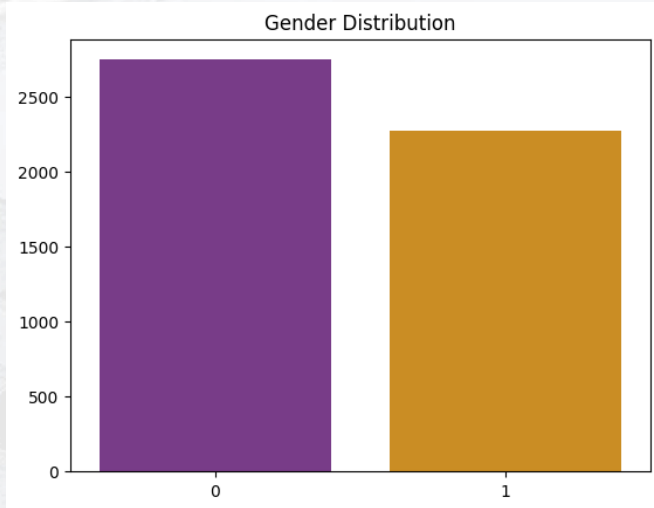
Product Name
All



Insight

- Analyze the reasons behind March's high sales and peak. For example, a specific marketing campaign, a new product launch, or a seasonal trend. Replicating successful tactics in other months could boost sales.
- While thai tea, ginger candy, etc. were top sellers in quantity, focus efforts on the most profitable products (cheese stick, choco bar, etc.) to maximize revenue.
- Analyze high-selling store (e.g. Lingga) strategies to understand their success. For example, specific product promotions, customer service techniques, wider product variety, better deals, etc. Implement these strategies in other stores to boost sales.

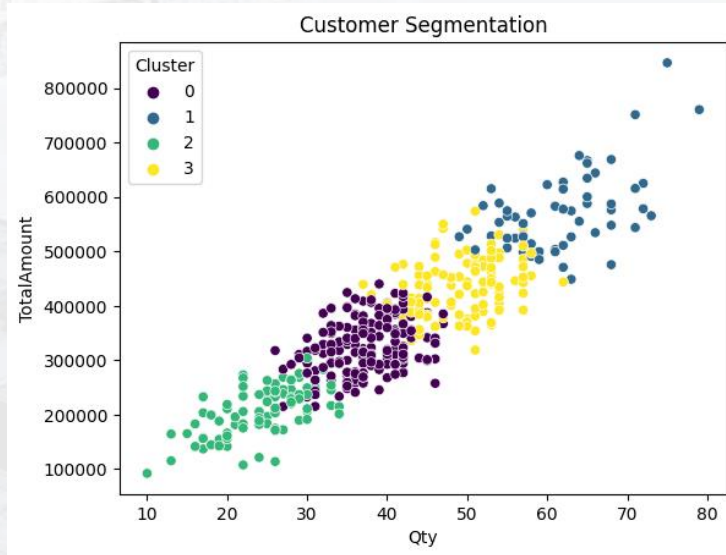
Customer Distribution



Customer Distribution

- There are more male customers than female. Explore ways to attract more females. Consider targeted promotions for female-oriented products or campaigns addressing specific female health needs.
- The age of customers is in the range of 20 - 60 years with some outliers (age below 10 years and above 65 years). The majority are in the age of 25 - 55 years. The 25-55 age group is the core demographic. Develop targeted marketing campaigns and product offerings suited to their preferences.
- Customer income is in the range of 0 - 20 million rupiah with some outliers (income above 20 million IDR). Target marketing efforts towards the majority income range (0-20 million IDR). Explore premium offerings for the higher-income outliers (above 20 million IDR).
- Utilize customer relationship management tools to personalize recommendations and promotions based on purchase history and demographics.

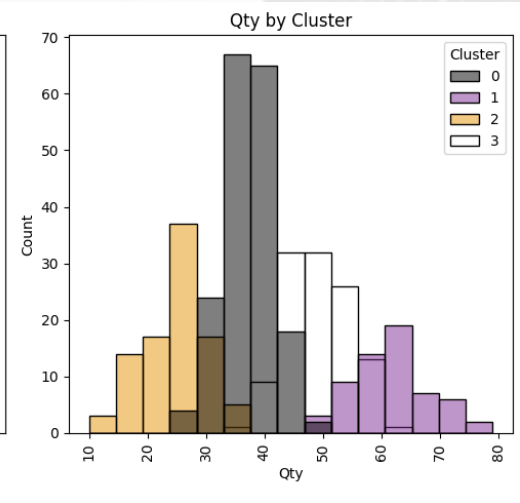
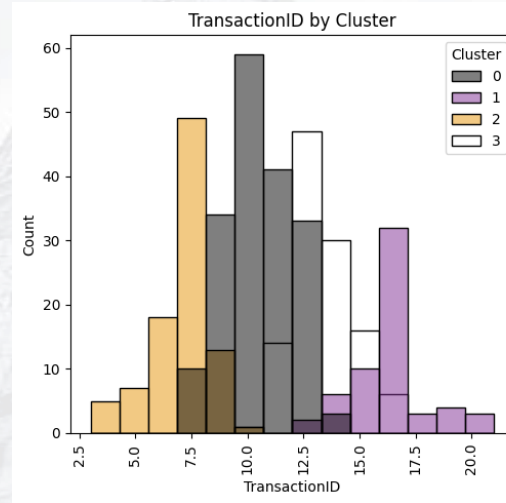
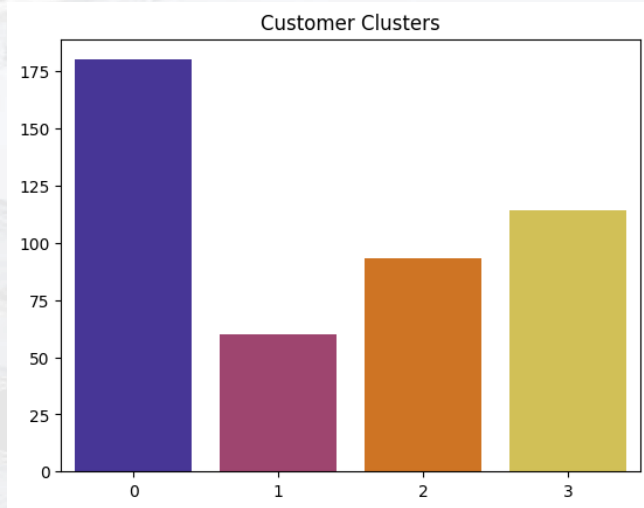
Clustering (KMeans)



There are 4 clusters for customer segmentation:

- Cluster 0 (Moderate Health Shoppers)
- Cluster 1 (High-Value Health Enthusiasts)
- Cluster 2 (Occasional Health Shoppers)
- Cluster 3 (Health Enthusiasts)

Customer Segmentation



Many customers are in cluster 0 (Moderate Health Shopper), and the least in cluster 1 (High-Value Health Enthusiasts).

Customer Segmentation

Cluster	Average Quantity of Goods Purchased (pcs)	Average Amount of Purchase (IDR)
0	37	Rp325,663
1	62	Rp572,100
2	25	Rp208,284
3	49	Rp436,204

Cluster 2 has the lowest average number of purchases. While Cluster 1 has the highest average number of purchases.

Customer Segmentation

Cluster 0 (Moderate Health Shoppers)

This group has a moderate frequency and amount of purchases. They prioritize effectiveness and efficiency in purchasing. They balance quality with affordable prices.

- To increase sales, promotion of quality products at affordable prices can be done.
- Discounts/promotions can be given to increase purchases and customer loyalty.
- Organize campaigns to raise awareness of products that can help their lives become better and healthier.

Customer Segmentation

Cluster 1 (High-Value Health Enthusiasts)

This group has the highest frequency and number of purchases. This group is a health conscious customer and does not think too much about price for purchasing the best quality products. This group is also a brand loyal customer.

- This group is the biggest contributor to sales so they can be given exclusive promotions and benefits as well as a reward system to maintain loyalty.
- Campaigns can also be carried out by informing the best products or the latest products owned by the brand for their every need.
- Provide personalized recommendations for customers in this segment.
- Ask for feedback from them regarding a product to be used as one of the references for the future.

Customer Segmentation

Cluster 2 (Occasional Health Shoppers)

This group is a customer who has the least frequency and amount of purchases. This group has an emphasis on choosing products that can meet their needs at the most affordable price. They also tend to buy products only when needed and do not pay much attention to brands.

- As a step to increase sales and loyalty, provide discounts to encourage purchases.
- Conduct campaigns to realize the importance of health at vital times.
- Provide product samples to introduce the brand and get them interested in buying the product.

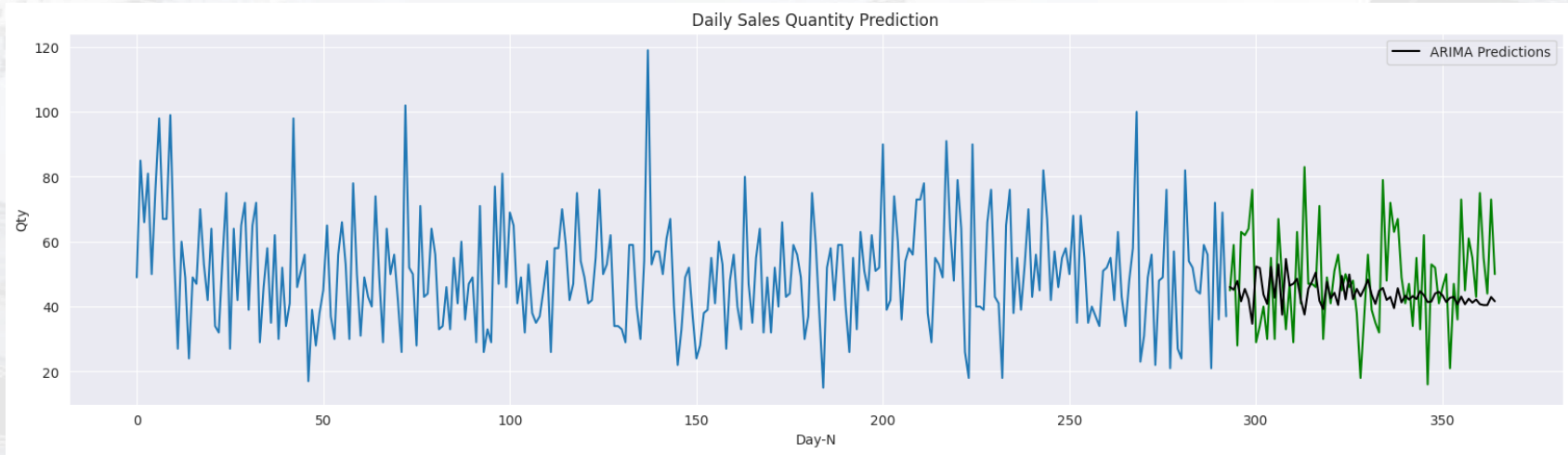
Customer Segmentation

Cluster 3 (Health Enthusiasts)

This group has a considerable frequency and number of purchases, just less than the top group. This group contains customers who are willing to spend money on high-quality health products and are willing to spend more money to get them as long as they are worth it.

- Do upselling to increase interest in buying premium products.
- In the campaign, emphasize the quality and efficacy of the product in everyday life. In addition, introduction to relevant premium products can be done.
- Promote bundling with premium products to generate interest in buying.

Regression (ARIMA)



Using the ARIMA (40,2,1) model to create a regression machine learning model that can predict the number of products sold so that inventory management can be optimized.

Regression (ARIMA)

The metric measurement results show that the ARIMA model created has not been able to predict well enough. This can also be seen from the visualization of the prediction results on the previous page, which shows that the ARIMA prediction is still far from the actual value.

To improve model performance, hyperparameter tuning or experimentation with other models is required.

```
Mean Absolute Error (MAE): 12.85  
Mean Squared Error (MSE): 271.13  
Root Mean Squared Error (RMSE): 16.47  
Mean Absolute Percentage Error (MAPE): 28.85%
```

Thank You

