Rakamin Academy X KALBE Nutritionals

Machine Learning Project

Quantity Prediction
Customer Segmentation

Presented by
Salsabila Mardhiyah

Salsabila Mardhiyah

About Me

Hello! Welcome to my portfolio, I'm Sals.

I went to Universitas Indonesia and majored in civil engineering at the time. Interested in Data Analytics and Data Science, I also attended Data Science Bootcamp at Rakamin Academy and graduated with Excellent Grades. In my 5-year-old plan, I aim to continue growing professionally in Data Science Field and taking steps forward.

I invite you to explore my portfolio and review my work. As I believe in continuous learning and growth, I am open to any thoughts or recommendations you may have.

Feel free to connect and reach me on Linked In!

Rakamin Academy

Experiences

Project-Based Virtual Intern : Big Data Analyst
Kimia Farma

- Carry out the process of creating a Data Mart from raw data performed by SQL queries into base and aggregate tables.
- Successfully created Sales Performance Dashboard.

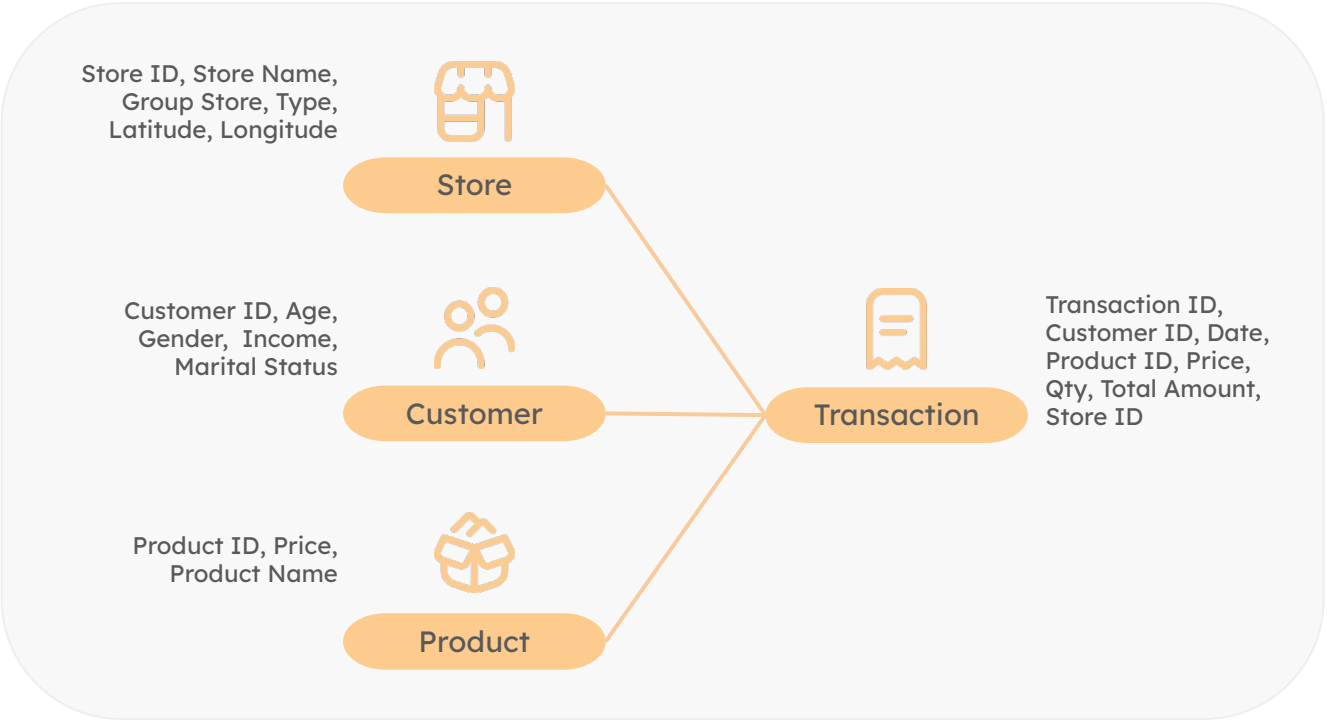
Program Planning Consultant
Ministry of Religious Affairs

- Strategically segmenting and prioritizing development planning at the district/city level to maximize impact.
- Collected related data and information then analyzed and evaluated proposals from numerous project candidates, providing expert recommendations for the stakeholder in decision-making and follow-up
- Developed a project guideline that encompassed all stages of project initiation, monitoring, and evaluation.

Project Overview

Project User	Objective
Inventory Team	Find out the projected amount of products sold to enable the establishment of daily inventory
Marketing Team	Create customer segmentation to provide personalized promotions and sales treatment

Database


Store ID, Store Name, Group Store, Type, Latitude, Longitude
Store
Customer ID, Age, Gender, Income, Marital Status
Customer
Product ID, Price, Product Name
Product
Transaction
Transaction ID, Customer ID, Date, Product ID, Price, Qty, Total Amount, Store ID

Project Workflow

1. **Data Exploration**
   Tools: PostgreSQL, DBeaver

2. **Dashboard Creation**
   Tools: Tableau Public

3. **Machine Learning Modelling**
   Tools: Python, Google Colaboratory

Customer Age Demographics

1. Data Exploration
2. Dashboard Creation
3. Machine Learning Modelling

Marital Status

Married
(43 years average)

Single
(29 years average)

Gender

Man
(39 years average)

Woman
(40 years average)

SQL Query

Top Store by Product Sold

1 Data Exploration

	Store	Quantity
1.	Lingga	2,78K pcs
2.	Sinar Harapan	2,59K pcs
3.	Prima Kota	1,40K pcs

SQL Query

Top Product by Sales Amount

1 Data Exploration

	Product	Total Amount
1.	Cheese Stick	27,62 M
2.	Choco Bar	21,19 M
3.	Coffee Candy	19,71 M

SQL Query

Sales Performance Dashboard

1. Data Exploration
2. **Dashboard Creation**
3. Machine Learning Modelling

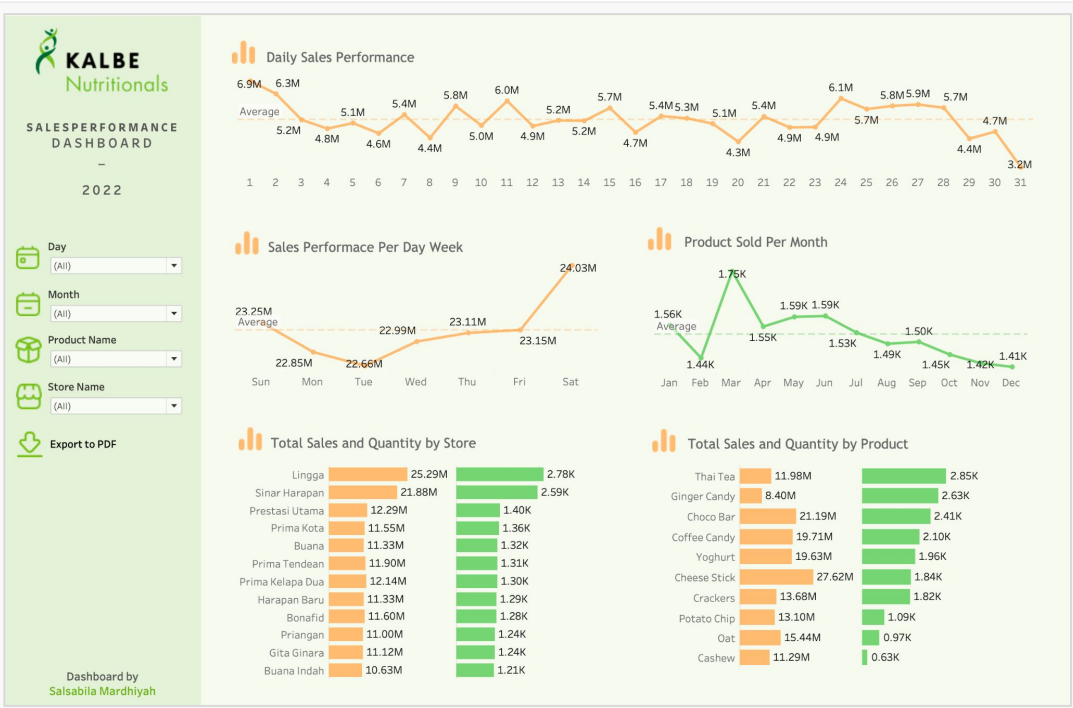
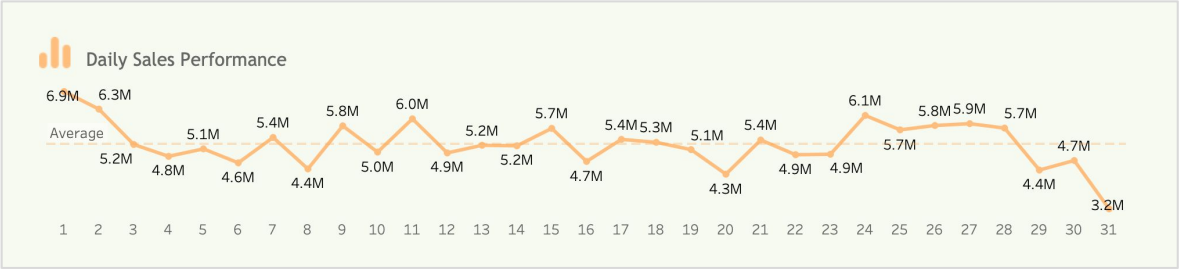


Tableau Viz

Sales Performance Dashboard

2 Dashboard Creation



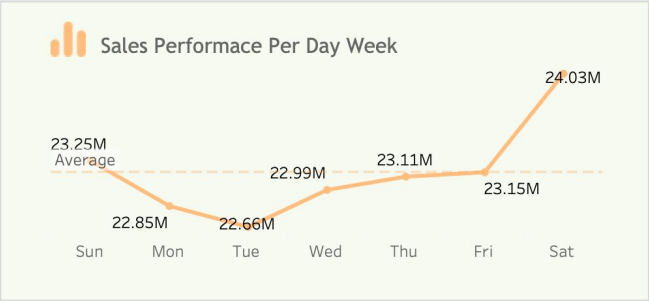
- This graph shows the total amount of sales day by day for year 2022. To see the trend, it could be filtered by month.
- Dashed ‘Average’ line shows the average sales amount from the beginning to the end of the month.

Tableau Viz

Sales Performance Dashboard

2

Dashboard Creation



- Graph beside shows total amount of sales per day week for the year.
- It shows that most sales occur on Saturdays and the least amount of sales occur on Mondays and Tuesdays.

- This graph shows total of product sold by month for the year.
- It shows that most product sold in March and the least quantity of product sold in December.

Tableau Viz

Sales Performance Dashboard

2

Dashboard Creation

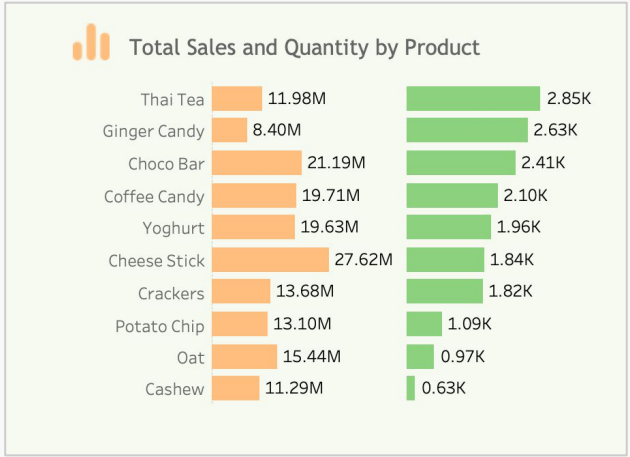
- This graph shows total amount of sales and quantity of product sold on each store for the year.
- Lingga Store has the most sales also quantity of product sold in 2022.
- On the contrary, Buana Indah Store has both the least amount of sales and quantity of product sold.

Tableau Viz

Sales Performance Dashboard

2

Dashboard Creation

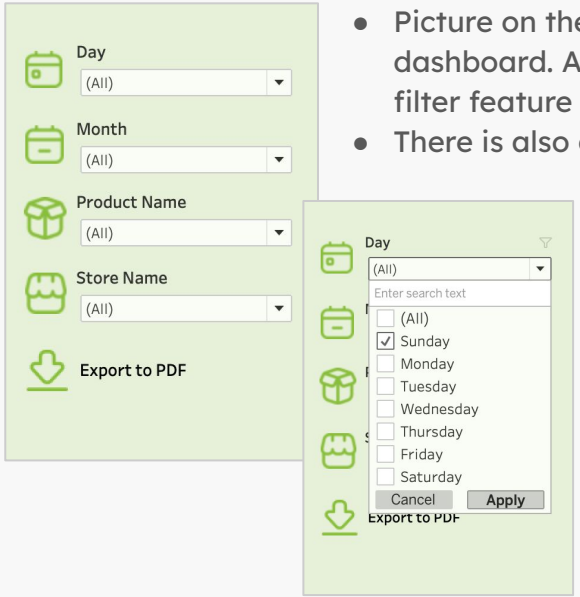


- This graph shows total amount of sales and quantity of product sold by product for the year.
- Cheese sticks rank first in terms of total sales, but not necessarily in terms of number of products sold.
- Meanwhile, Thai Tea is in first place in terms of the number of products sold, but its total sales do not show significant figures.

Tableau Viz

Sales Performance Dashboard

2 Dashboard Creation



- Picture on the side shows Filter section on the dashboard. All the figures have been linked with the filter feature to ease the process of analysis.
- There is also a button to export dashboard to pdf.

- Filter dropdown is also equipped with 'Apply' button in order to make the dashboard more efficient. So that they don't make continuous adjustments while selecting the desired filters.

Tableau Viz

Machine Learning

1 Data Exploration

2 Dashboard Creation

3 **Machine Learning Modelling**

Project User	Modelling Task
Inventory Team	Quantity Prediction Using Time Series Regression
Marketing Team	Customer Segmentation Using K-Means Clustering

Coding Script

Time Series Data

3 Machine Learning Modelling

Quantity Prediction

Customer Segmentation

Date

Train Data (80%)
Jan 1st - Oct 19th, 2022

Test Data (20%)
Oct 20th - Dec 31st, 2022

Quantity

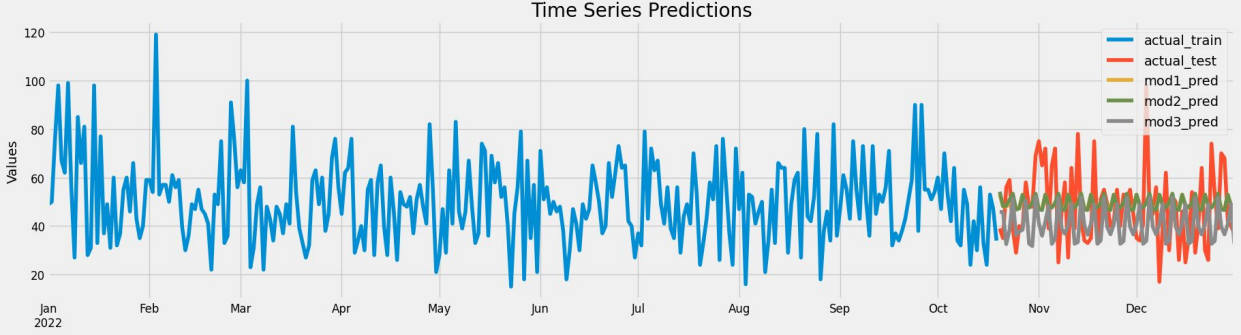
Quantity of product sold each day

Coding Script

Time Series Forecasting

3 Machine Learning Modelling

Quantity Prediction



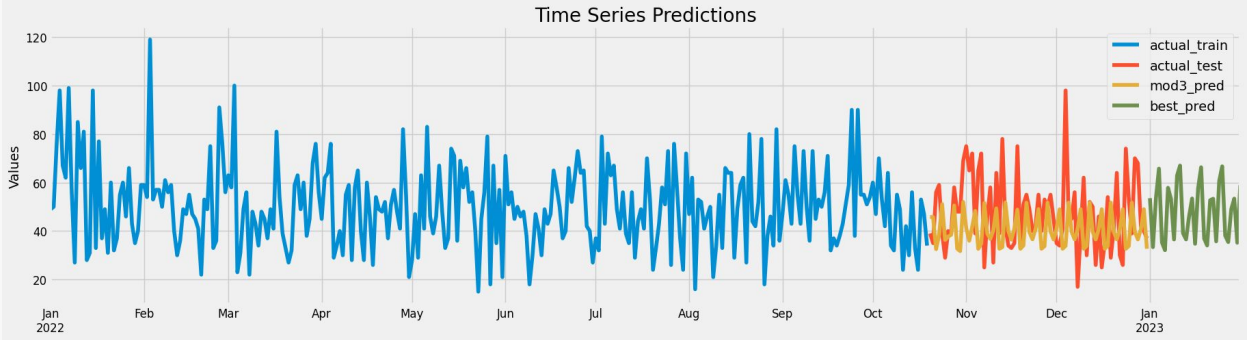
- Figure above shows plot of train data, test data, and three time series modelling scenario model1, model2, and model3.
- It's generally seen that the model3 line has more similar characteristics to the actual line than others. Also, only model3 passed the residual diagnostics with evaluation metrics below:
  - MSE: 370, MAE: 14.8, MAPE: 0.3

Coding Script

Forecasting Result

3 Machine Learning Modelling

Quantity Prediction



- Figure above shows plot of train data, test data, best model model3 prediction, and the forecast result of quantity of product needed in January 2023.
- Based on forecast result, quantity needed in January 2023 has statistics below:
  - Mean: 48, Median: 52, Min: 32, Max: 67, Total: 1495

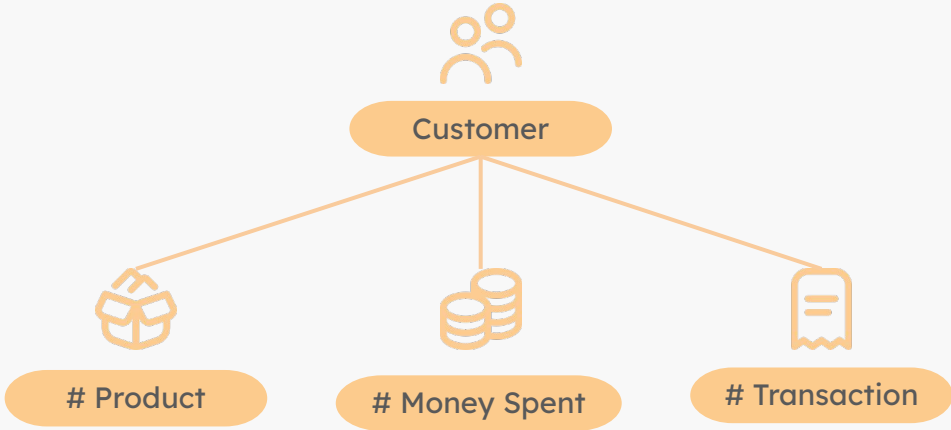
Coding Script

Clustering Data

3 Machine Learning Modelling

Quantity Prediction

Customer Segmentation

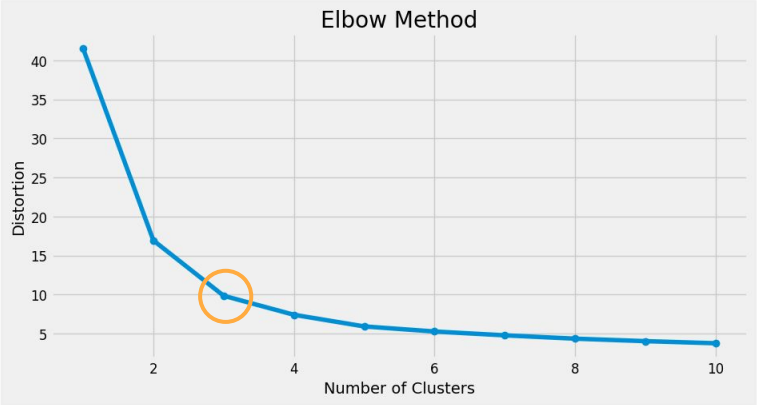


Coding Script

Number of Clusters

3 Machine Learning Modelling

Customer Segmentation



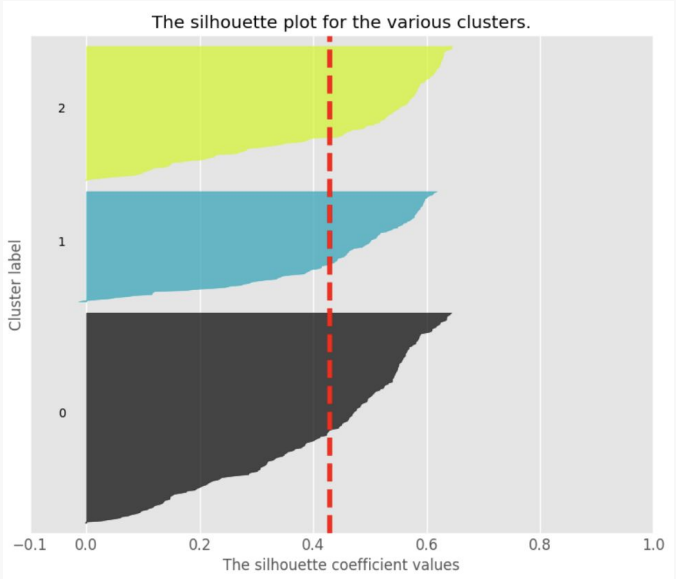
To determine the optimal number of cluster, we have to select the value of k at the elbow, ie the point after which distortion/inertia starts decreasing in a linear fashion. Thus for the given data, we can conclude that the optimal number of clusters is 3.

Coding Script

Number of Clusters

3 Machine Learning Modelling

Customer Segmentation



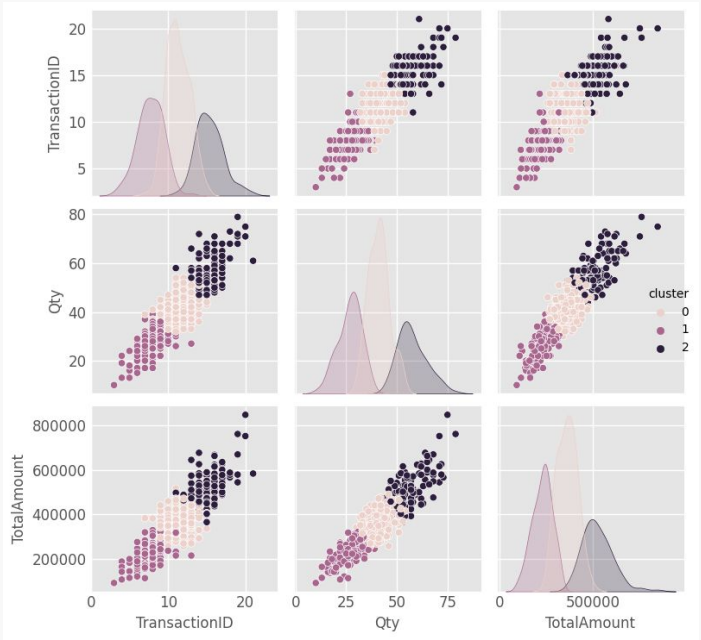
From silhouette plot we have to select the bigger value of the coefficient average and also consider proportional distribution of the clusters formed. Thus for the given data, we can conclude that the optimal number of clusters is 3.

Coding Script

Clustering Result

3 Machine Learning Modelling

Customer Segmentation



Beside is a pairplot of each cluster's parameters. This shows different average characteristics of 3 clusters formed by the model.

Coding Script

Cluster Characteristics

3 Machine Learning Modelling

Customer Segmentation

*on average	Regular Cust.	Mid-Level Cust.	High-Value Cust.
# Transactions	7	11	15
# Product	26	41	57
# Money Spent	228K	261K	524K

Coding Script

Marketing Strategy

Regular Cust.

Recommend products based on their past buying behavior to spark interest in new categories.

Encourage customers to spend more by promoting bundles of related products they've purchased before.

Mid-Level Cust.

Develop loyalty programs that reward frequent transactions, encouraging to continue purchasing regularly.

Personalized recommendations to new products or upsell complementary items.

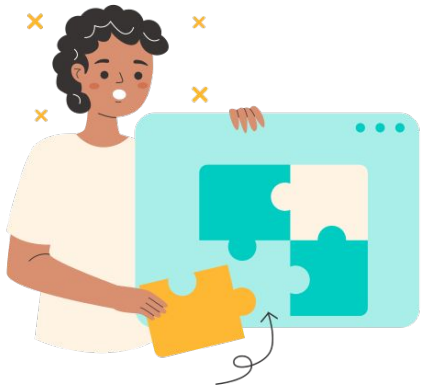
High-Value Cust.

Create exclusive programs, offering unique benefits and access to limited-edition products.

Offer early access to new product launches or exclusive events to strengthen their loyalty.

Business Recommendation

Project User	Action Item
Inventory Team	Perform an analysis obtained from the dashboard and compare the forecasting results. Then incorporate the forecasts into supply chain and inventory management processes to ensure having the right amount of products available to meet customer demand.
Marketing Team	Carry out personalized marketing strategies based on the characteristics of each customer segment. Conduct in-depth predictions to find out what products are most likely to buy.



Result Documentation

GitHub

LinkedIn

Video

Dashboard

Folder

SQL Query

.py File

.ipynb File

Thank You