



X



**KALBE**  
Nutritionals

# Machine Learning Project

Membuat model Regression dan Clustering

**Kalbe Nutritionals Data Scientist  
Project Based Internship  
Program**

Presented by  
Putri Aulia



# Putri Aulia

## About You

A fresh graduate of Andalas University, majoring in the Mathematics and Data Science Department, who loves to learn something new especially to improve data knowledge and skills, and has a strong will. She is interested in statistics, data science, and data analysis. She gained experience while studying mathematics, which has helped her be a skilled problem-solver with critical thinking abilities and an organized person.

## Insert Your Experience

### **DQLab Free Class**

Data Science Fundamentals

**Sept 2023**

### **Computer Laboratory Assistant**

Mathematics and Data Science

Computation Laboratory

**Apr 2021 - Feb 2023**

### **Data Science Course**

Class program of INDOMS School 2021  
(Kampus Merdeka)

**Juli 2021 - Aug 2021**



# Case Study - Background Story

Kamu adalah seorang Data Scientist di Kalbe Nutritionals dan sedang mendapatkan project baru dari tim inventory dan tim marketing.

**Dari tim inventory**, kamu diminta untuk dapat membantu memprediksi jumlah penjualan (quantity) dari total keseluruhan product Kalbe

- Tujuan dari project ini adalah untuk mengetahui perkiraan quantity product yang terjual sehingga tim inventory dapat membuat stock persediaan harian yang cukup.
- Prediksi yang dilakukan harus harian.

**Dari tim marketing**, kamu diminta untuk membuat cluster/segment customer berdasarkan beberapa kriteria.

- Tujuan dari project ini adalah untuk membuat segment customer.
- Segment customer ini nantinya akan digunakan oleh tim marketing untuk memberikan personalized promotion dan sales treatment.

Task 1 - Dbeaver Connection with PostgreSQL

Task 2 - Tableau Public Create Dashboard

Task 3 - Machine Learning Regression (Time Series)

Task 4 - Machine Learning Clustering

# Task 1 - Dbeaver

Query 1 : Berapa rata-rata umur customer jika dilihat dari marital statusnya ?

```
select "Marital Status" , round(avg(age)) as  
rata_rata_umur from customer group by "Marital Status"
```

|   | ABC Marital Status ▼ | 123 rata_rata_umur ▼ |
|---|----------------------|----------------------|
| 1 |                      | 31                   |
| 2 | Married              | 43                   |
| 3 | Single               | 29                   |

Query 2 : Berapa rata-rata umur customer jika dilihat dari gender nya ?

```
select gender , round(avg(age)) as  
rata_rata_umur from customer group by gender
```

|   | 123 gender ▼ | 123 rata_rata_umur ▼ |
|---|--------------|----------------------|
| 1 | 0            | 40                   |
| 2 | 1            | 39                   |

Query 3 : Tentukan nama store dengan total quantity terbanyak!

```
select storename , sum(qty) as jumlah_quantity from "transaction"  
join store on store.storeid = transaction.storeid  
group by storename  
order by jumlah_quantity desc
```

|   | ABC storename ▼ | 123 jumlah_quantity ▼ |
|---|-----------------|-----------------------|
| 1 | Lingga          | 2,777                 |

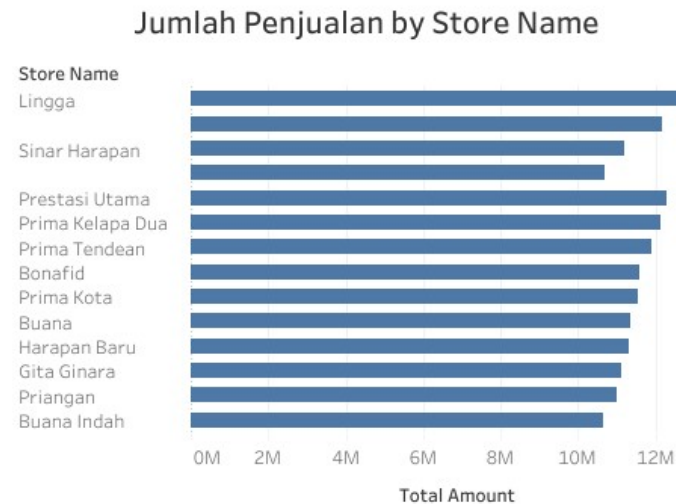
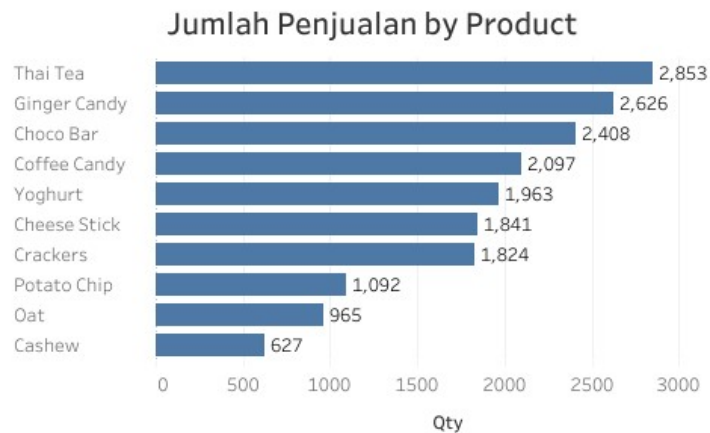
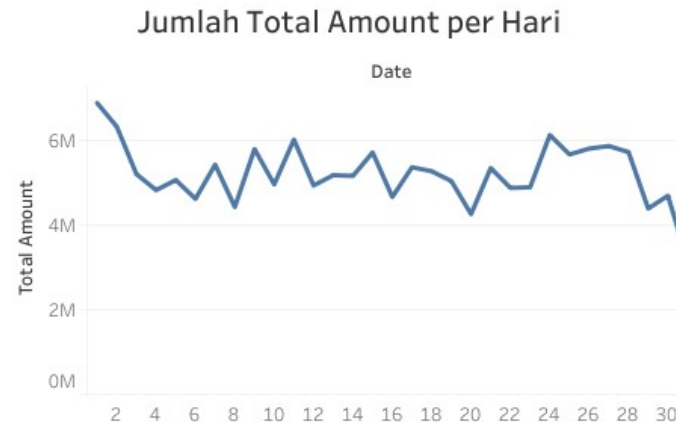
Query 4 : Tentukan nama produk terlaris dengan total amount terbanyak!

```
select "Product Name" , sum(totalamount) as total_amount from "transaction"  
join product on product.productid = transaction.productid  
group by "Product Name"  
order by total_amount desc
```

|   | ABC Product Name ▼ | 123 total_amount ▼ |
|---|--------------------|--------------------|
| 1 | Cheese Stick       | 27,615,000         |

## Task 2 - Data Visualization & Tableau Public

### Create Dashboard



### Interpretasi

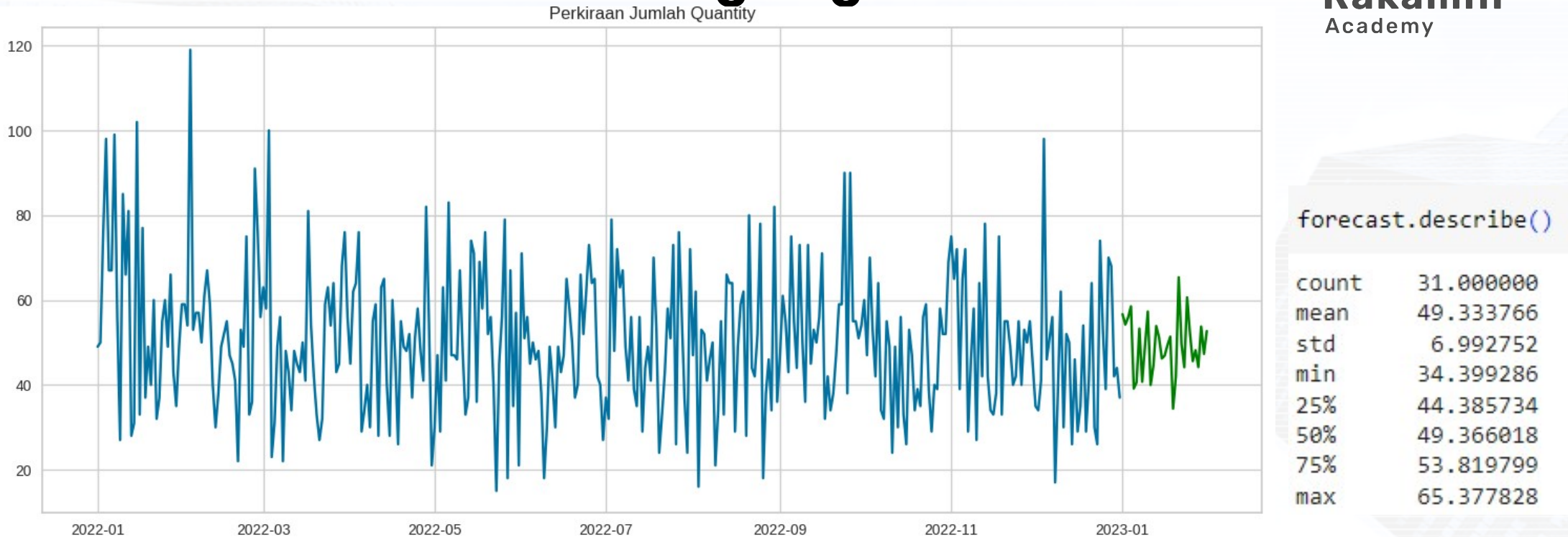
1. Jumlah penjualan Quantity terbanyak berada pada bulan Maret dan terendah pada bulan Desember;
2. Jumlah total amount mengalami naik turun setiap harinya, dan tertinggi pada hari pertama, serta terendah pada hari terakhir;
3. Jumlah penjualan produk tertinggi adalah produk Thai Tea;
4. Toko dengan penjualan terbanyak adalah Lingga dan Sinar Harapan. Untuk penjualan toko lain hampir sama.



# Link Task 3 dan Task 4

<https://github.com/Putri4501/Kalbe-Nutritionals.git>

# Task 3 - Machine Learning Regression



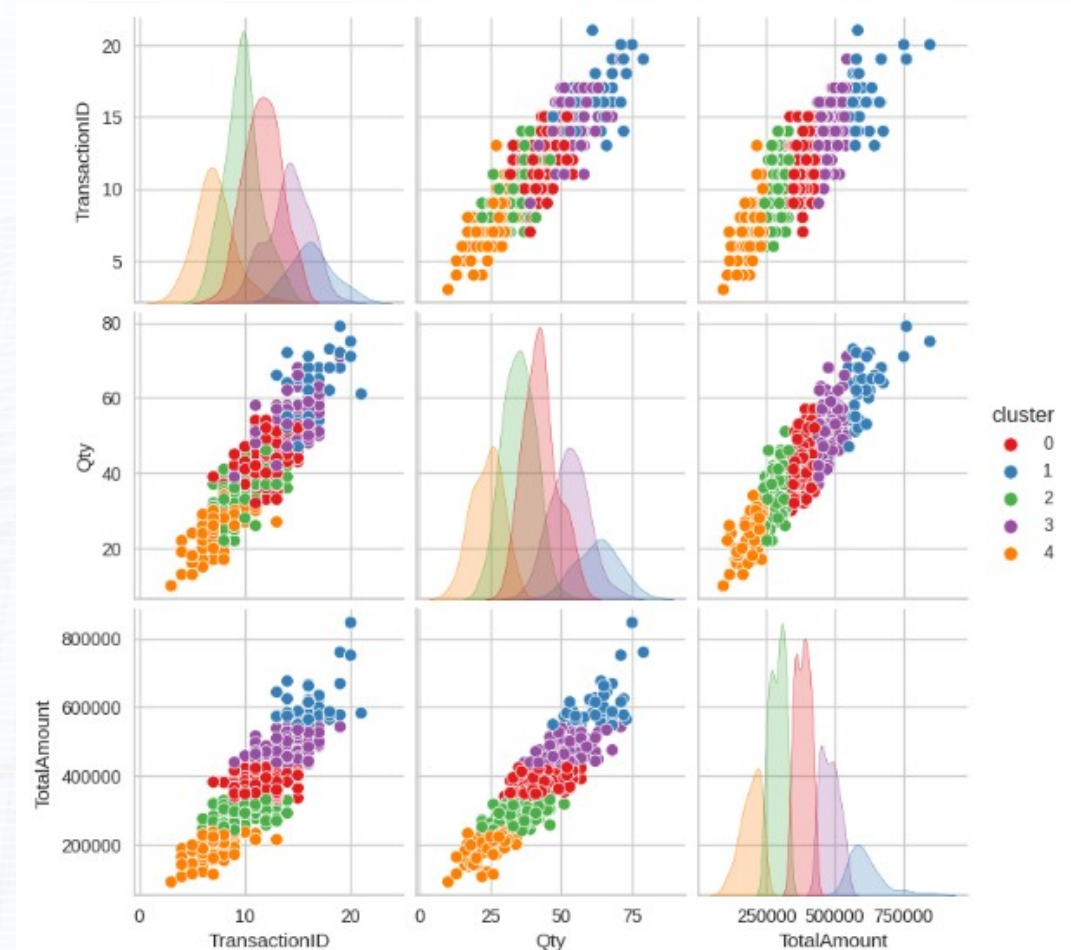
## Interpretasi :

Diperoleh perkiraan jumlah quantity untuk bulan selanjutnya yaitu 1 Januari-31 Januari 2023 seperti pada grafik di atas, dengan rata-rata perkiraan jumlah quantity yang akan terjual sebesar 50 per-harinya. Untuk nilai perkiraan lebih lanjut dapat dilihat pada link yang telah disediakan sebelumnya.

# Task 4 - Machine Learning Clustering

|     | CustomerID | TransactionID | Qty | TotalAmount | cluster |
|-----|------------|---------------|-----|-------------|---------|
| 0   | 1          | 17            | 60  | 623300      | 1       |
| 1   | 2          | 13            | 57  | 392300      | 0       |
| 2   | 3          | 15            | 56  | 446200      | 3       |
| 3   | 4          | 10            | 46  | 302500      | 2       |
| 4   | 5          | 7             | 27  | 268600      | 2       |
| ... | ...        | ...           | ... | ...         | ...     |
| 442 | 443        | 16            | 59  | 485100      | 3       |
| 443 | 444        | 18            | 62  | 577700      | 1       |
| 444 | 445        | 18            | 68  | 587200      | 1       |
| 445 | 446        | 11            | 42  | 423300      | 0       |
| 446 | 447        | 13            | 42  | 439300      | 3       |

447 rows x 5 columns





# Task 4 - Machine Learning Clustering

|         | CustomerID | TransactionID | Qty       | TotalAmount   |
|---------|------------|---------------|-----------|---------------|
| cluster |            |               |           |               |
| 0       | 124        | 11.661290     | 42.508065 | 381419.354839 |
| 1       | 38         | 16.342105     | 62.815789 | 611818.421053 |
| 2       | 127        | 9.748031      | 34.842520 | 288907.086614 |
| 3       | 85         | 14.105882     | 52.717647 | 480628.235294 |
| 4       | 73         | 7.068493      | 23.726027 | 191141.095890 |

Interpretasi :

- **Klaster 0** : Jumlah customer sebanyak 124, dengan rata-rata jumlah item yang dibeli sebanyak 43 (nilai tersebut berada dalam kategori sedang), dan rata-rata jumlah total penghasilan juga dalam kategori sedang.

- **Klaster 1** : Jumlah customer sebanyak 38 (sangat sedikit) namun rata-rata jumlah produk yang dibeli sebanyak 67 (nilai tersebut terbanyak dibanding dengan jumlah item yang dibeli pada klaster lain) dan rata-rata jumlah total penghasilan juga tinggi

Interpretasi :

- **Klaster 2** : Jumlah customer sebanyak 127, dengan rata-rata jumlah item yang dibeli sebanyak 35 item (nilai tersebut berada dalam kategori sedang), dan rata-rata jumlah total penghasilan dalam kategori rendah.

- **Klaster 3** : Jumlah customer sebanyak 85 dengan rata-rata jumlah produk yang dibeli sebanyak 52 item (nilai tersebut berada dalam kategori tinggi/banyak) dan rata-rata jumlah total penghasilan dalam kategori rendah.

- **Klaster 4** : Jumlah customer sebanyak 73 dengan rata-rata jumlah produk yang dibeli sebanyak 23 item (nilai tersebut berada dalam kategori rendah) dan rata-rata jumlah total penghasilan juga dalam kategori rendah. Klaster ini menghasilkan Total Amount terendah.

# Thank You



**Rakamin**  
Academy



**KALBE**  
Nutritional