

# Machine Learning Project

**Kalbe Nutritionals Data Scientist  
Virtual Internship Program**

Presented by  
Celine Paulsaeman

## Experiences

### Software Engineer

Berani Digital ID · Internship  
May 2023 - Present · 6 mos

### Mentor and Co-Founder of DACODE Bootcamp

Self-employed  
Feb 2023 - Present · 9 mos

### Laboratory Assistant

Bina Nusantara University · Part-time  
Sep 2022 - Feb 2023 · 6 mos



## Celine Paulsaeman

## About You

I am a diligent and detail-oriented Computer Science student entering my fifth semester at Bina Nusantara University, specializing in Database Technology. With a stellar GPA of 3.95, I am ardently committed to expanding my horizons and enriching my proficiency in both soft and hard skills.

# Case Study

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.

# Case Study

Data file dari Kalbe:

- Customer: berisi detail informasi customer: customer ID, age, gender, marital status, dan income
- Product: berisi detail informasi produk: produk ID, name, dan price
- Store: berisi detail informasi store: Store ID, name, group, type, latitude, dan longitude
- Transaction: berisi detail informasi transaction: transaction ID, Customer ID, date, Product ID, Price, Quantity, Total Amount, dan Store ID



# Exploratory Data Analysis

rata-rata umur customer jika dilihat dari marital statusnya

```
select
  case "Marital Status"
    when '' then 'Unknown'
    else "Marital Status"
  end "Marital Status",
  round(avg(age), 3) as "average age"
from customer
group by "Marital Status"
order by "average age" asc;
```

	ABC Marital Status	123 average age
1	Single	29.385
2	Unknown	31.333
3	Married	43.038

# Exploratory Data Analysis

rata-rata umur customer jika dilihat dari gender

```
select
  case gender
    when null then 'Unknown'
    when 0 then 'Female'
    when 1 then 'Male'
  end gender,
  round(avg(age), 3) as "average age"
from customer
group by gender
order by "average age" asc;
```

	gender	average age
1	Male	39.141
2	Female	40.326

# Exploratory Data Analysis

nama store dengan total quantity terbanyak

```
select
  store.storename as "Store Name",
  sum(transaction.qty) as "total quantity"
from store join "transaction" on transaction.storeid = store.storeid
group by "Store Name"
order by "total quantity" desc
limit 1;
```

	ABC Store Name	123 total quantity
1	Lingga	2,777

# Exploratory Data Analysis

nama produk terlaris dengan total amount terbanyak

```
select
  product."Product Name" as "Product Name",
  sum(transaction.totalamount) as "total amount"
from product join "transaction" on transaction.productid = product.productid
group by "Product Name"
order by "total amount" DESC
limit 1;
```

	ABC Product Name	123 total amount
1	Cheese Stick	27,615,000

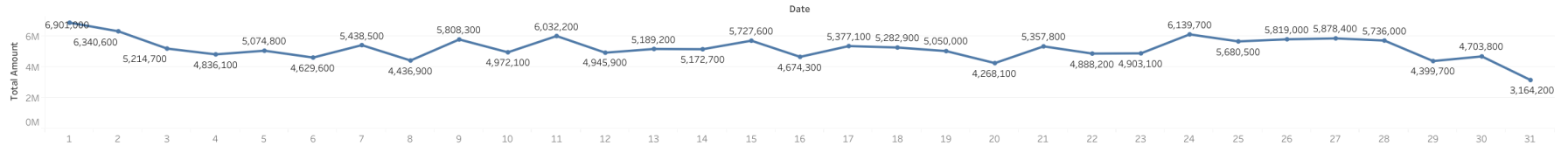


# Sales Dashboard

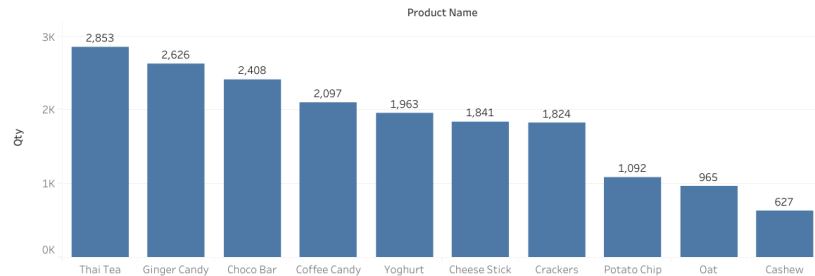
Total quantity per month



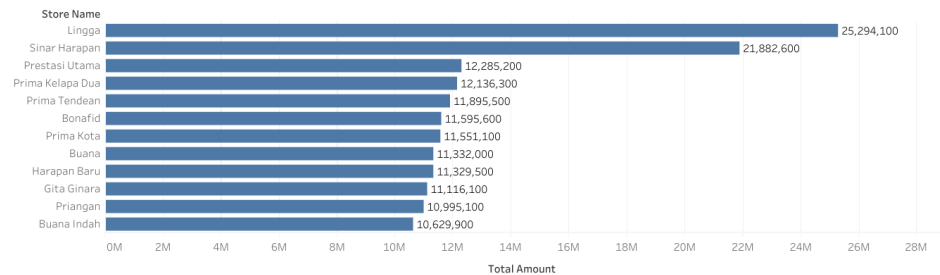
Total amount per day



Number of sales by product



Jumlah penjualan (total amount) by store name



# Forecast Daily Sales Quantity

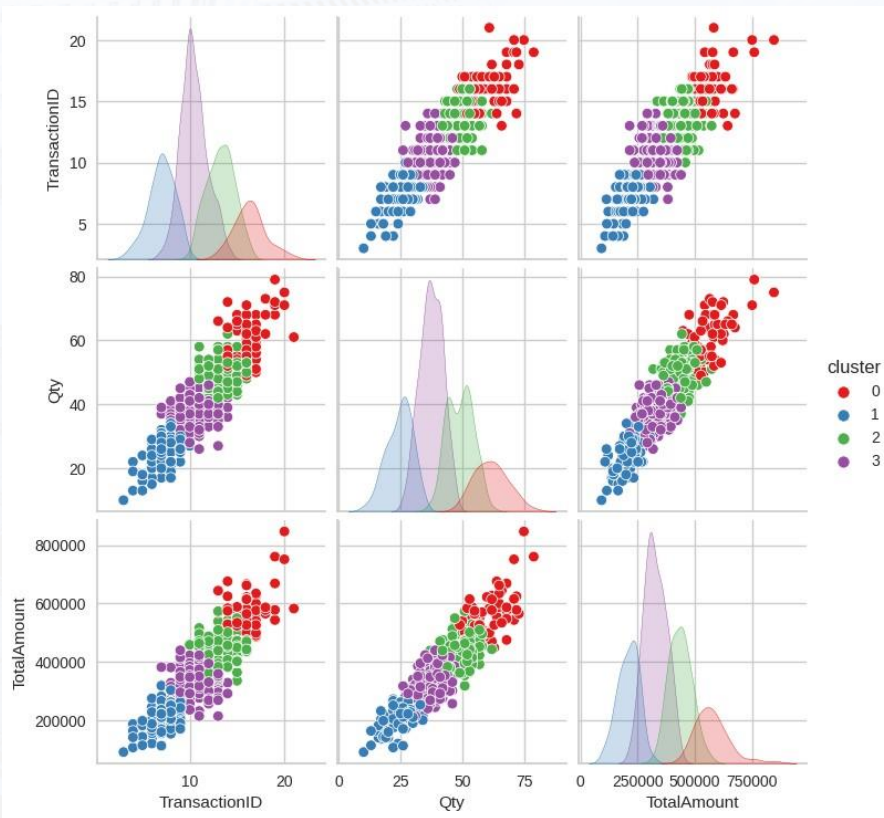


```
[261] forecast.describe()
```

```
count    31.000000
mean     44.339338
std       4.272234
min      35.687121
25%      41.282627
50%      43.228274
75%      47.339633
max      52.771970
Name: predicted_mean, dtype: float64
```

From the forecast, the average quantity sales in January 2023 is 44.339338 or up rounded to around 44 pcs/day.

# Clustering Analysis



	CustomerID	TransactionID	Qty	TotalAmount
cluster				
3	180	10.427778	37.350000	325663.333333
2	113	13.247788	49.070796	435833.628319
1	93	7.021505	24.505376	208283.870968
0	58	16.362069	61.689655	574650.000000

# Conclusion

- **Cluster 0**

Kluster ini memiliki jumlah pelanggan paling sedikit. Pelanggan dalam kelompok ini memiliki karakteristik yang menduduki peringkat tertinggi dalam setiap metrik.

**Rekomendasi:**

1. Menawarkan program loyalitas (poin) untuk mempertahankan tingkat transaksi.
2. Melakukan survei kepuasan pelanggan.
3. Membuat produk atau paket premium dengan harga yang lebih tinggi
4. Menawarkan program afiliasi yang tersedia untuk semua orang

- **Cluster 1**

Kluster ini terdiri dari pelanggan dengan peringkat terendah dalam setiap metrik.

**Rekomendasi:**

1. Menawarkan Bundling Produk atau Penawaran Diskon Volume
2. Melakukan survei untuk mengidentifikasi potensi pengembangan produk.
3. Menyiapkan tim dukungan pelanggan yang responsif dan berpengetahuan untuk membantu pelanggan dengan pertanyaan atau masalah yang mereka alami.
4. Membuat section FAQ secara online, artikel tutorial, atau forum komunitas tempat pelanggan dapat berinteraksi dan saling berbagi informasi.



- **Cluster 2**

Pelanggan dalam kelompok ini memiliki karakteristik yang menduduki peringkat kedua tertinggi dalam setiap metrik.

**Rekomendasi:**

1. Fokus melakukan up-selling
2. Mengadakan program loyalitas
3. Mengadakan umpan balik secara teratur dari pelanggan untuk memahami kebutuhan dan preferensi mereka.

- **Cluster 3**

Kluster ini memiliki jumlah pelanggan terbanyak. Para pelanggan dalam kelompok ini memiliki karakteristik yang menempati posisi ketiga tertinggi.

**Rekomendasi:**

1. Mengadakan program loyalitas atau memberikan penghargaan kepada pelanggan yang setia sebagai bentuk pengakuan dan apresiasi atas bisnis mereka.
2. Menyelenggarakan survei secara berkala untuk lebih memahami minat utama dari pelanggan.
3. Tingkatkan komunikasi dengan pelanggan melalui email, pesan teks, atau media sosial untuk tetap terhubung.

# Github

<https://github.com/CelinePaulsaeman/VIX-Kalbe-Nutritionals>

# Video Presentation

[https://youtu.be/S\\_Fl82M0SsQ](https://youtu.be/S_Fl82M0SsQ)

# Thank You

