

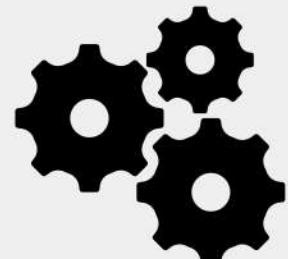
# **PROYEK DATA SCIENCE KALBE NUTRIRIONAL**

**MUHAMMAD RUSYDI HANAN**

# ABOUT PROJECT



Projek ini merupakan sebuah projek yang dilakukan untuk memperoleh insight atau informasi dari data yang telah tersedia



Projek ini dilakukan dengan bantuan teknologi machine learning untuk memudahkan pemerolehan informasi yang berkualitas



# PROJECT GOALS

## Objective n° 1

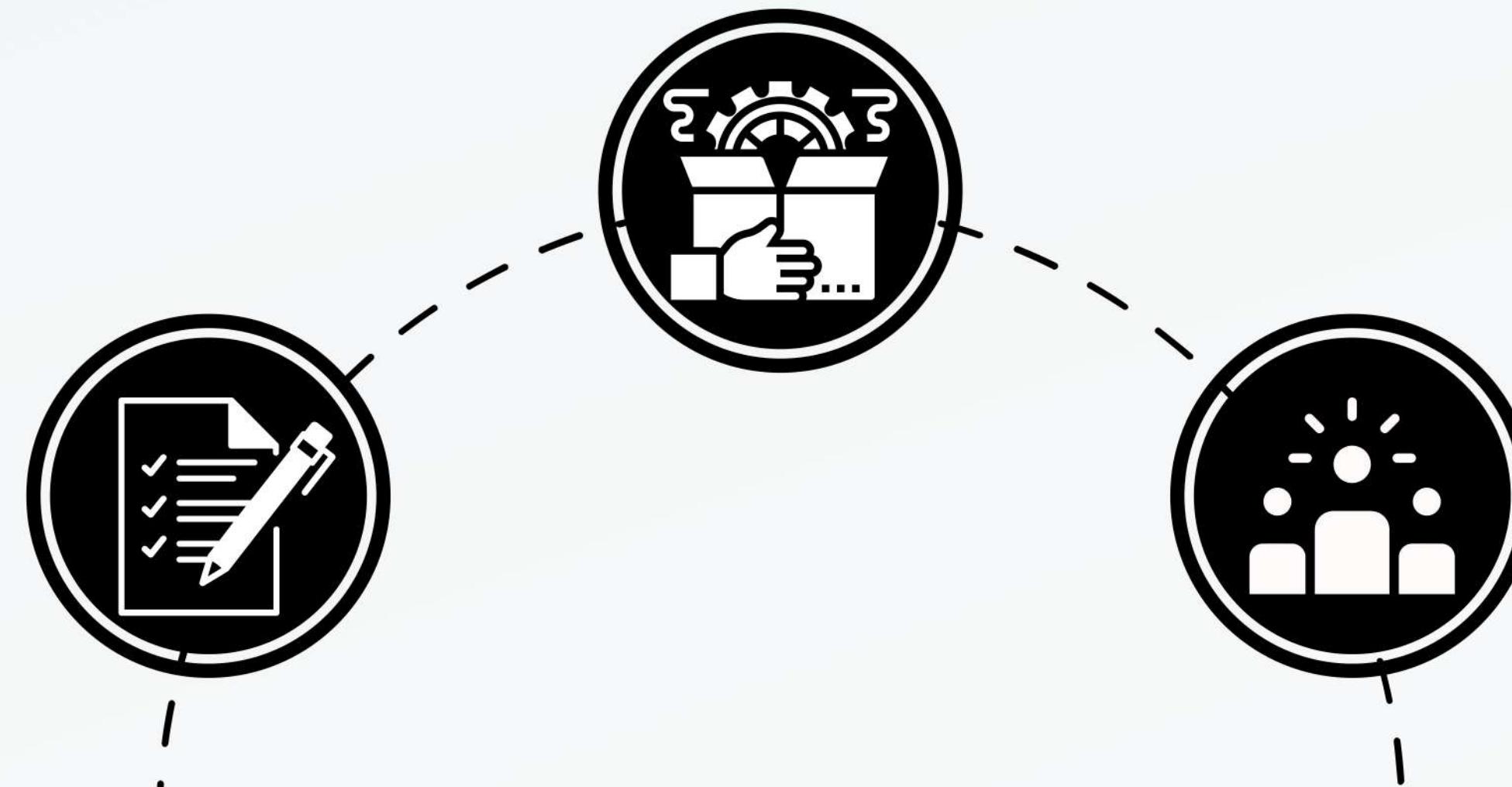
Mengetahui informasi penjualan perusahaan

## Objective n° 2

Memperikarakan quantitas produk yang dijual untuk membantu menentukan jumlah stock harian

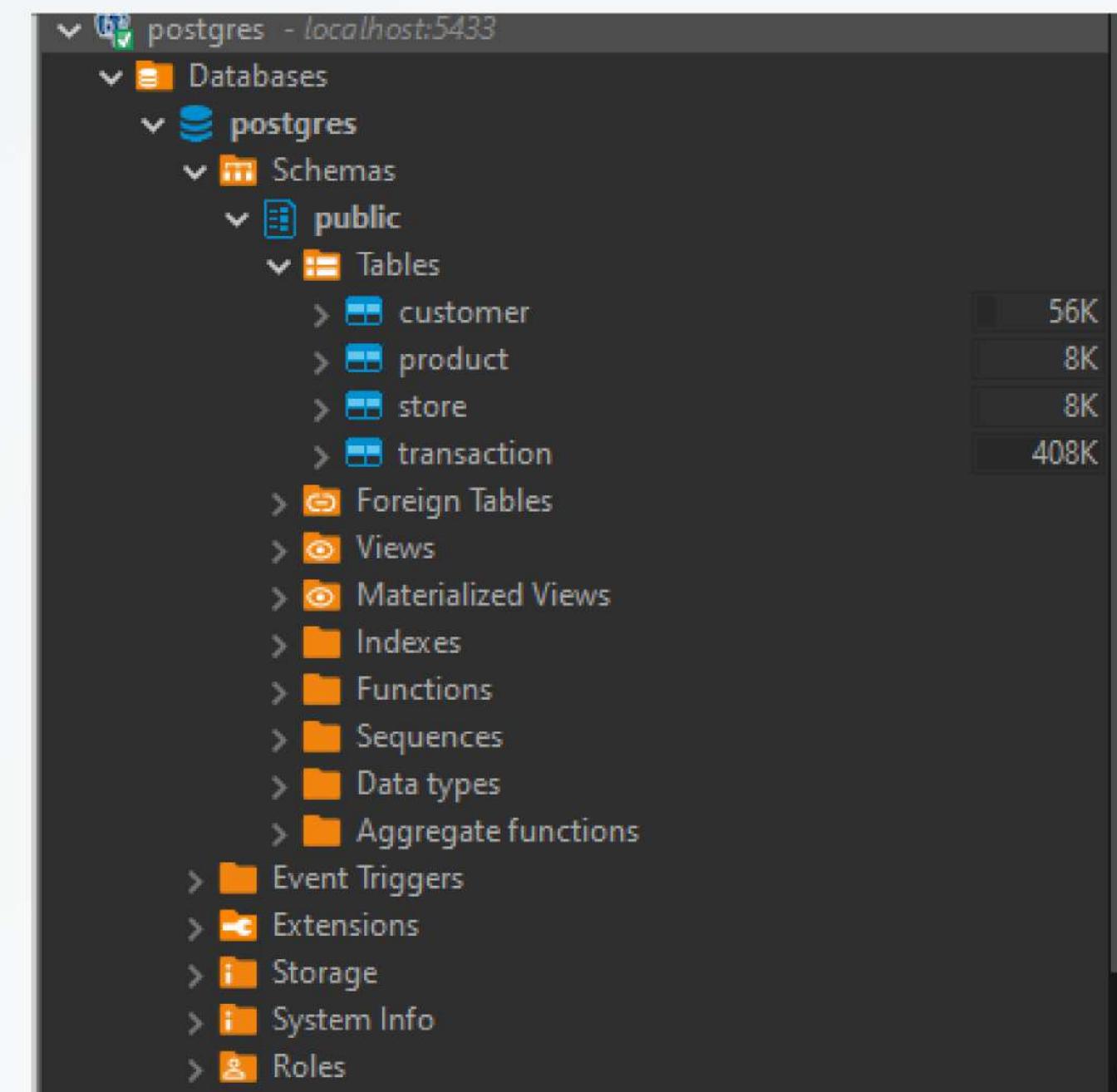
## Objective n° 3

Membuat segmentasi customer untuk membantu proses promosi

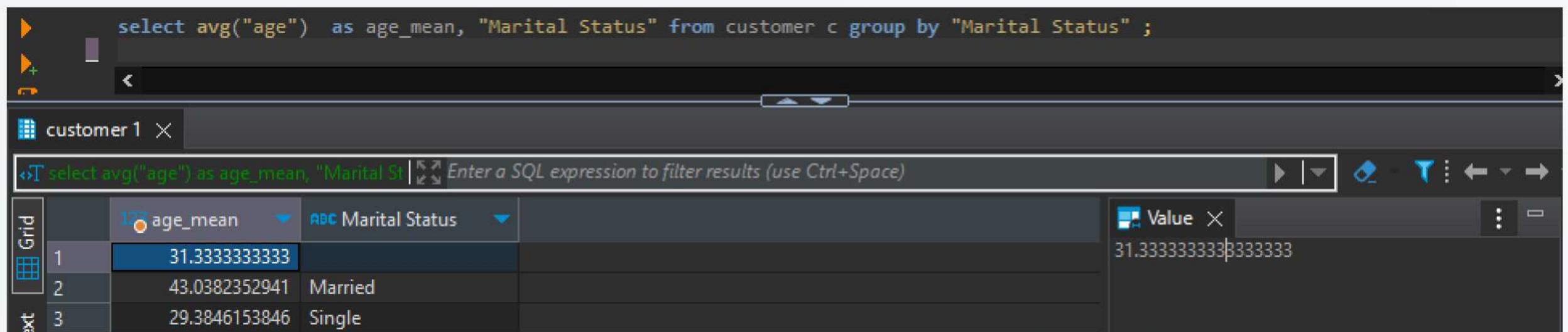


# DATA INGESTION

- Data Ingestion dilakukan dengan memasukkan data kedalam database
- Database yang digunakan adalah postgresql dengan DBeaver sebagai DBMS-nya
- Perintah-perintah query juga dapat dilakukan untuk pemerolehan informasi



# QUERY TEST 1



A screenshot of a SQL query results window. The query is:

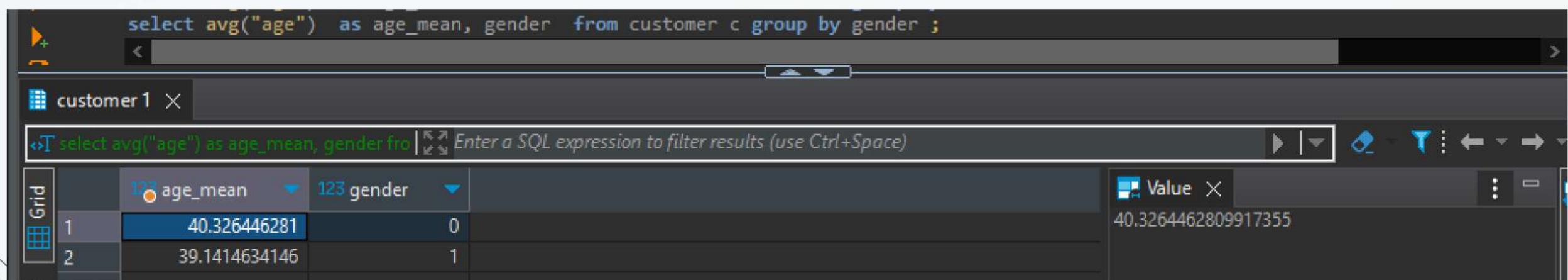
```
select avg("age") as age_mean, "Marital Status" from customer c group by "Marital Status";
```

The results are displayed in a grid:

	age_mean	Marital Status
1	31.3333333333	
2	43.0382352941	Married
3	29.3846153846	Single

A value panel on the right shows the first row's age\_mean value: 31.3333333333.

Melihat rata-rata umur cusomer yang sudah menikah dan belum



A screenshot of a SQL query results window. The query is:

```
select avg("age") as age_mean, gender from customer c group by gender;
```

The results are displayed in a grid:

	age_mean	gender
1	40.326446281	0
2	39.1414634146	1

A value panel on the right shows the first row's age\_mean value: 40.3264462809917355.

Melihat rata-rata umur cusomer wanita dan pria

# QUERY TEST 2

```
select s.storename, sum(qty) as tot_qty from "transaction" t join store s on t.storeid = s.storeid group by s.storename order by tot_qty desc limit 1;
```

The screenshot shows a database interface with a query editor at the top containing the SQL code. Below it is a results grid with columns for storename and tot\_qty. A single row is selected, showing 'Lingga' with a value of 2,777. A tooltip 'Enter a SQL expression to filter results (use Ctrl+Space)' is visible above the grid.

storename	tot_qty
Lingga	2,777

Melihat toko dengan jumlah penjualan barang terbanyak

```
select p."Product Name" , totalamount from "transaction" t join product p on t.productid = p.productid limit 1;
```

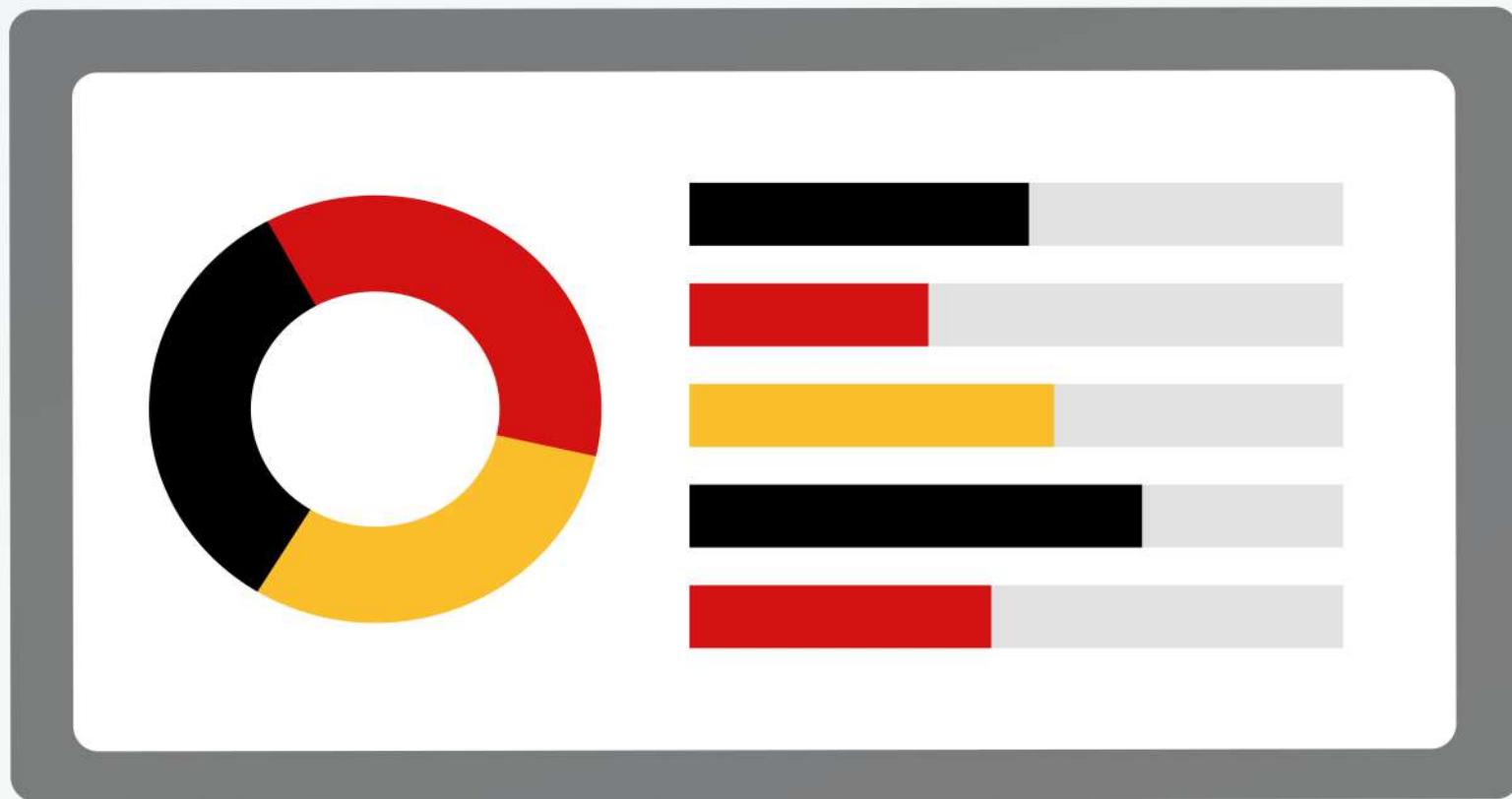
The screenshot shows a database interface with a query editor at the top containing the SQL code. Below it is a results grid with columns for Product Name and totalamount. A single row is selected, showing 'Crackers' with a value of 30,000. A tooltip 'Enter a SQL expression to filter results (use Ctrl+Space)' is visible above the grid.

Product Name	totalamount
Crackers	30,000

Melihat produk dengan jumlah profit terbesar

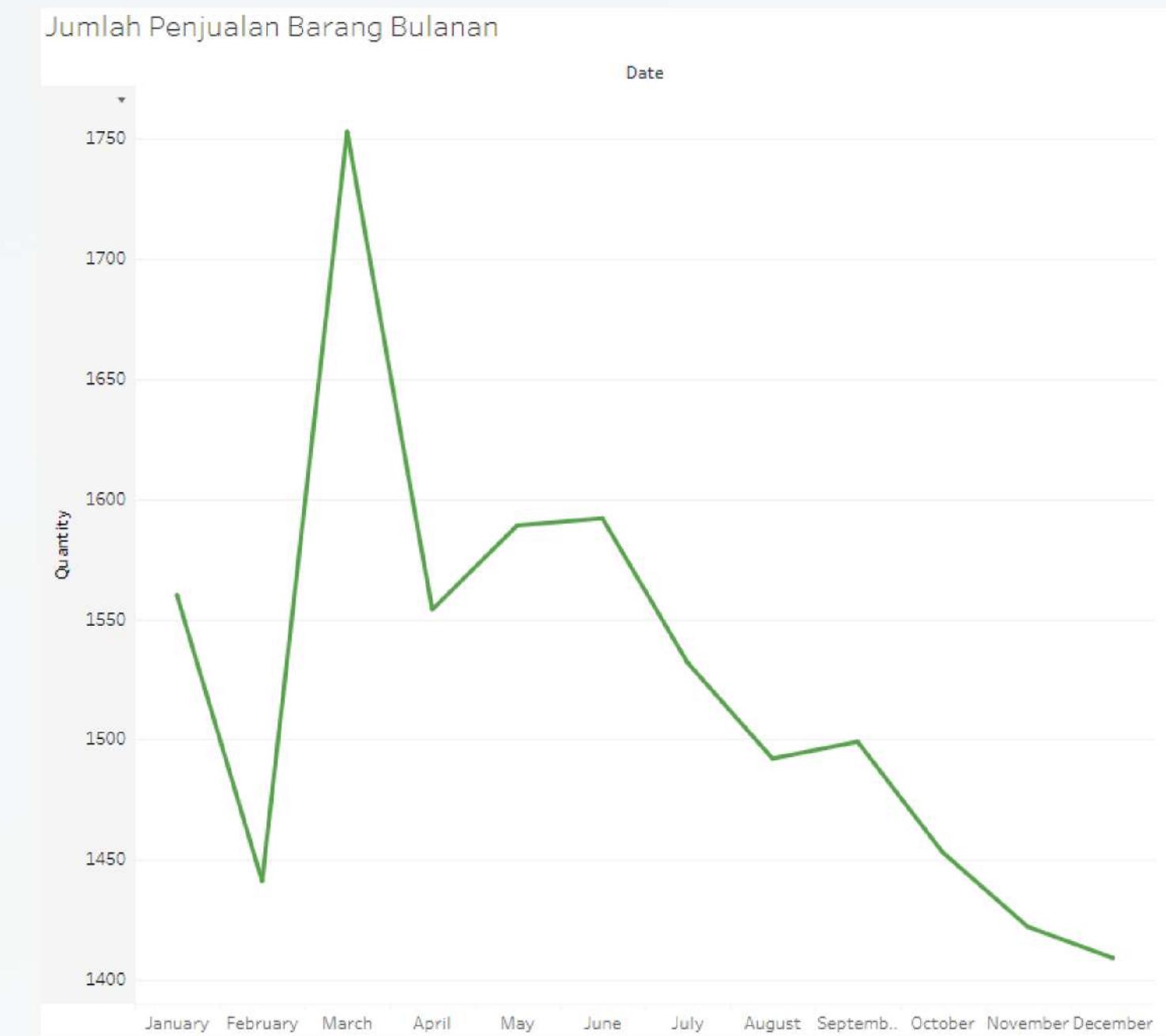
# DASHBOARDING

- Databoarding adalah proses pembuatan dashboard sebagai bentuk visualisasi data untuk memudahkan penyampaian data pada stakeholder
- Dashboard disini dilakukan dengan menggunakan tableau dengan menyatukan beberapa worksheet



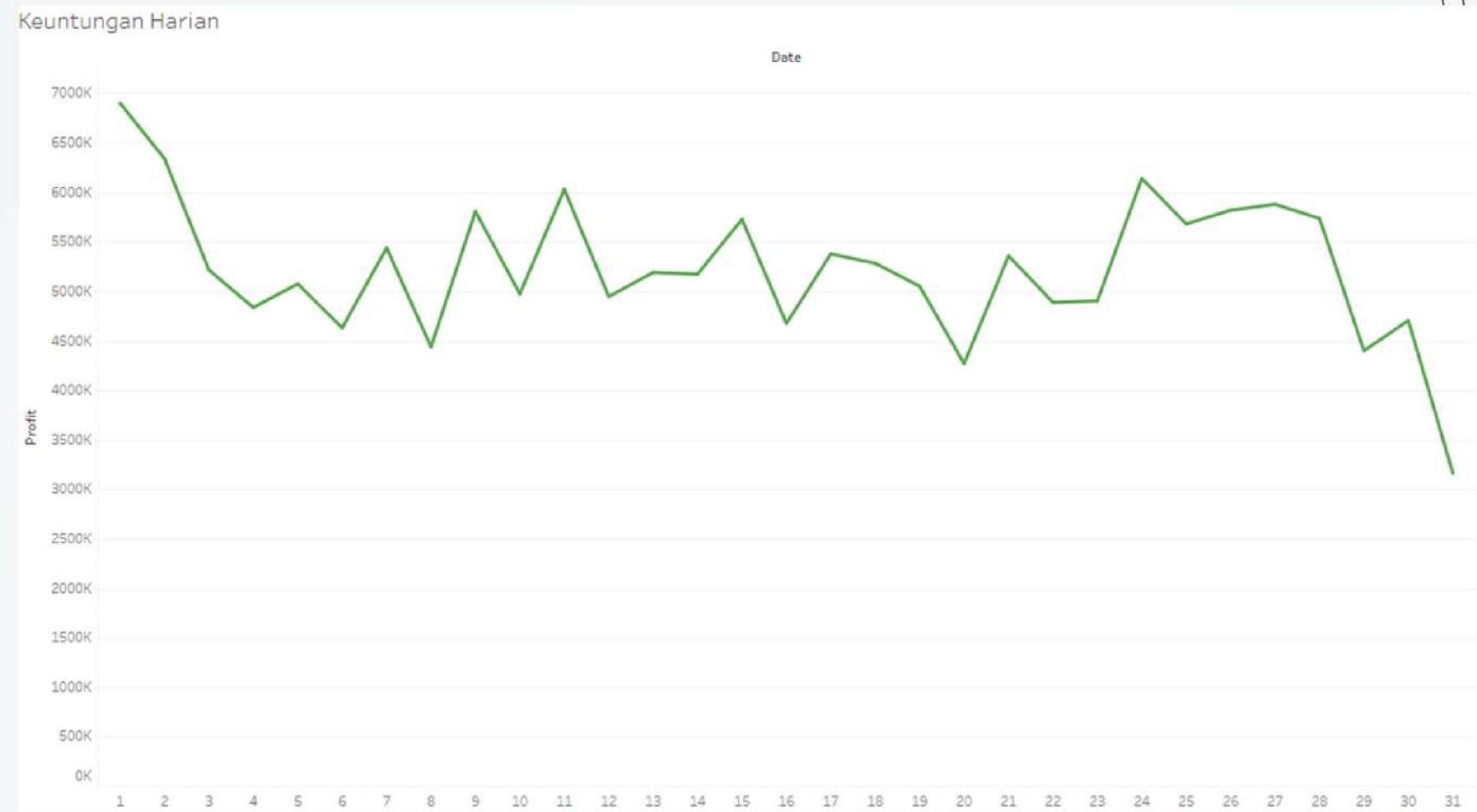
# WORKSHEET 1

- worksheet 1 menunjukkan jumlah penjualan barang perbulannya dalam satu tahun
- Dapat kita lihat penjualan tertinggi berada pada bulan maret yang kemudian berangsur turun hingga akhir tahun



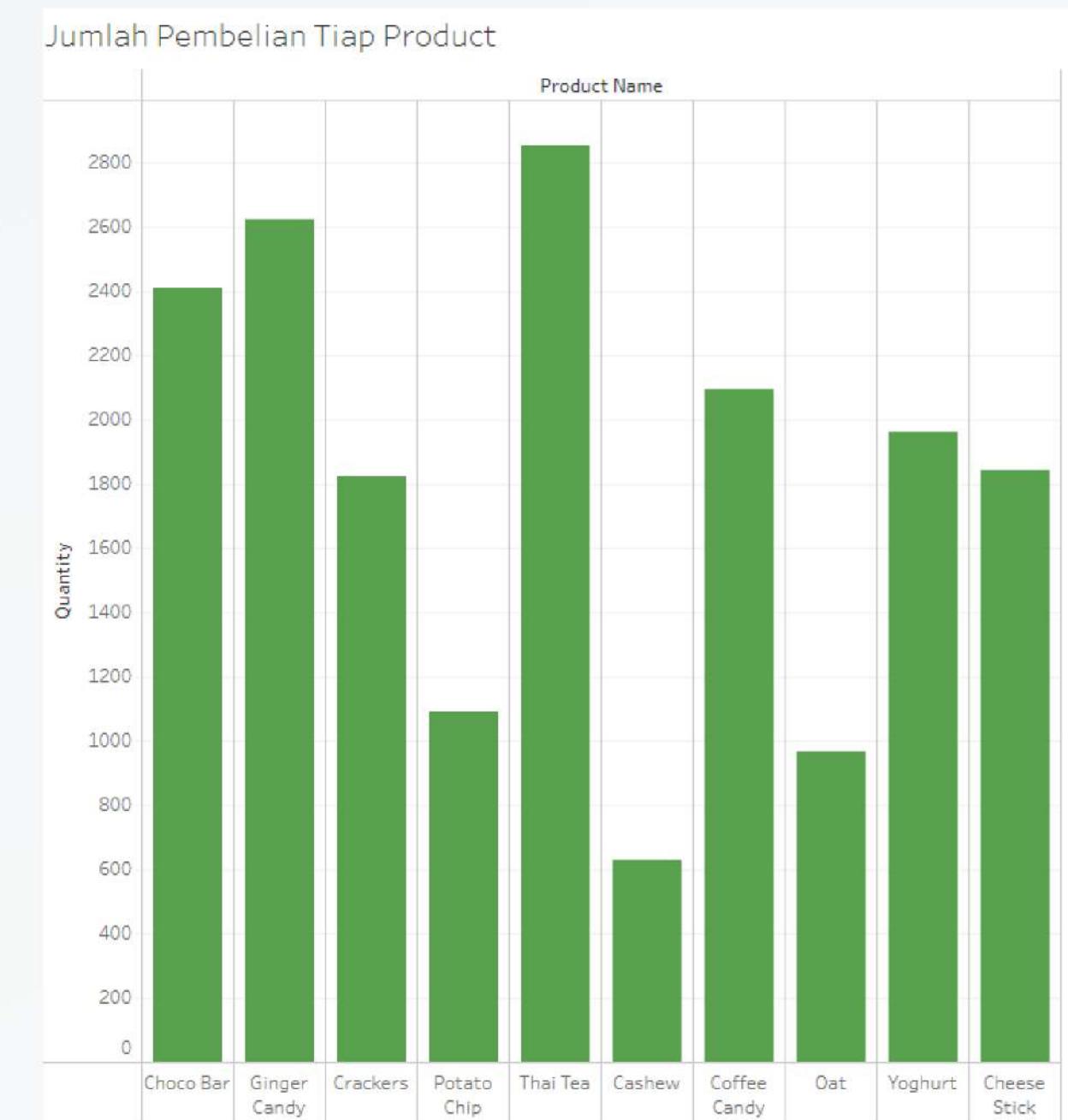
# WORKSHEET 2

- worksheet 2 menunjukkan jumlah penjualan barang perbulannya dalam satu tahun
- Dapat kita lihat keuntungan harian yang didapatkan oleh perusahaan yang terus naik turun



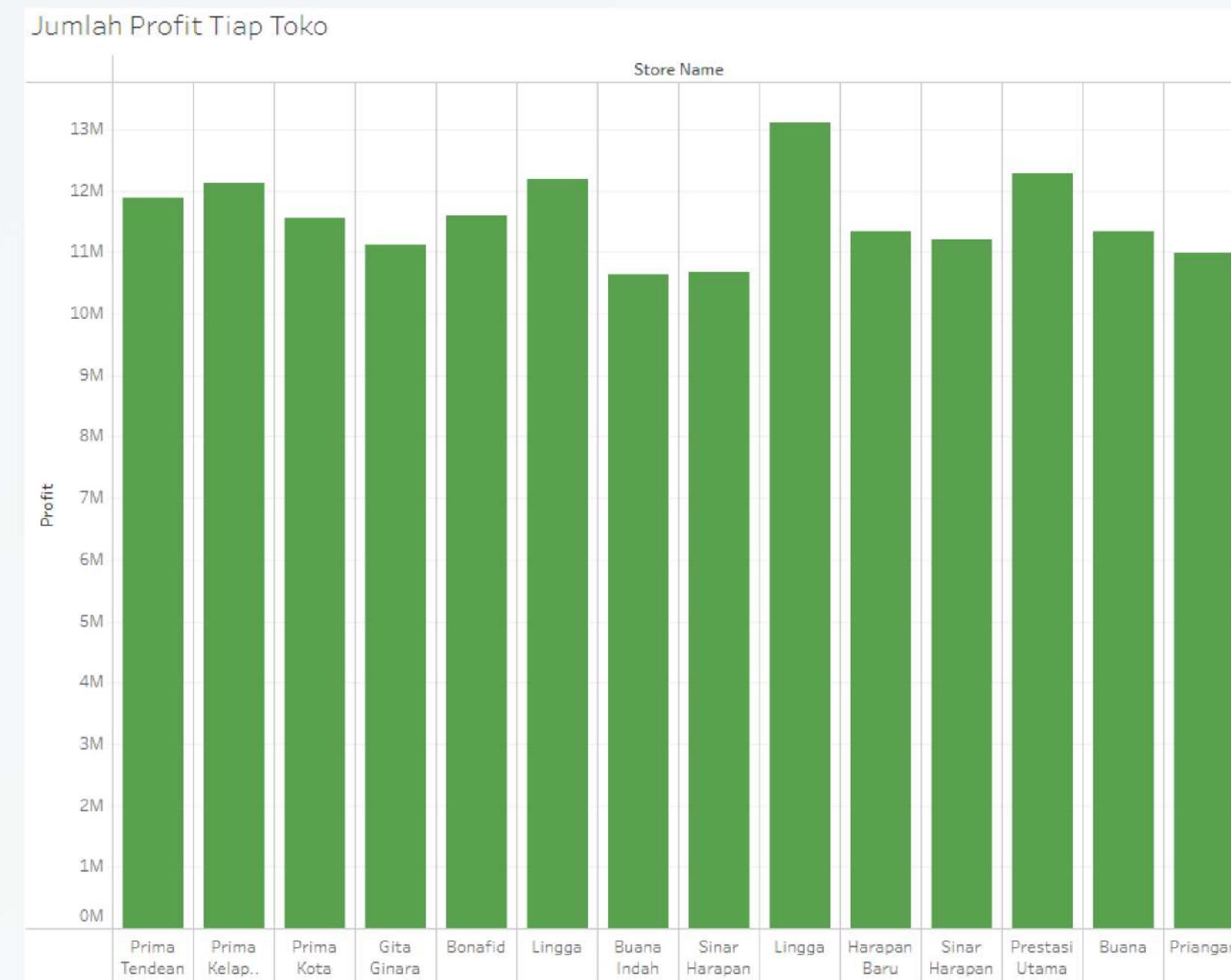
# WORKSHEET 3

- worksheet 3 menunjukkan jumlah total kuantitas penjualan barang dalam 1 tahun
- Dapat kita lihat produk thai tea merupakan produk yang paling sering dibeli



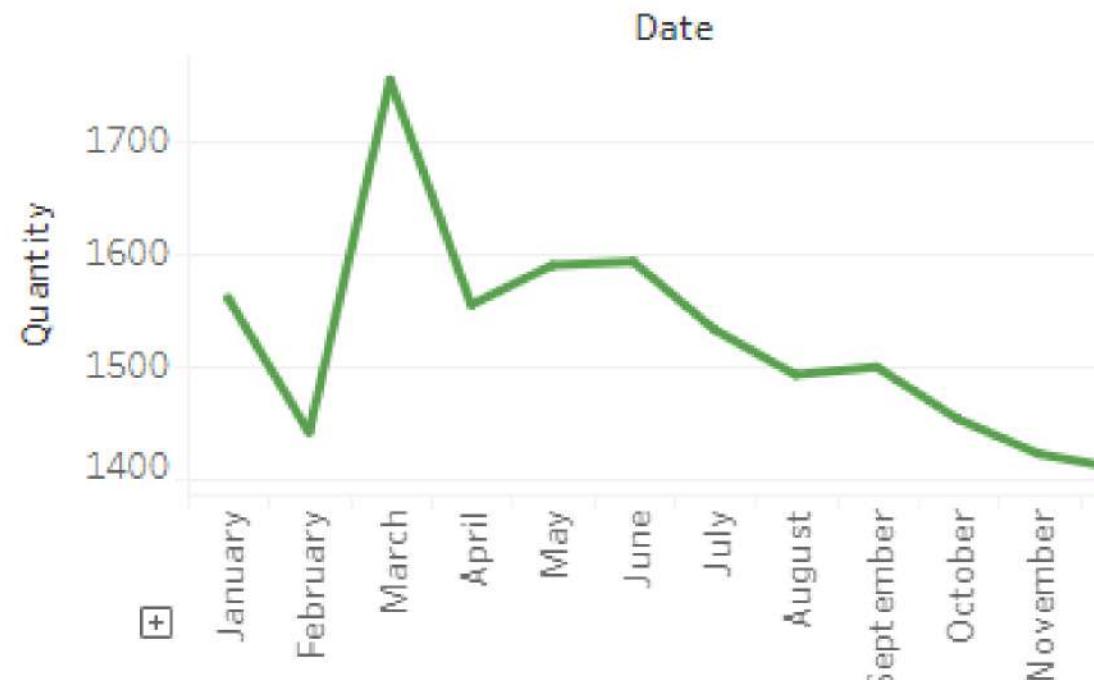
# WORKSHEET 4

- worksheet 4 menunjukkan jumlah profit yang dihasilkan oleh tiap toko
- Dapat kita lihat Lingga merupakan toko yang paling banyak menghasilkan keuntungan

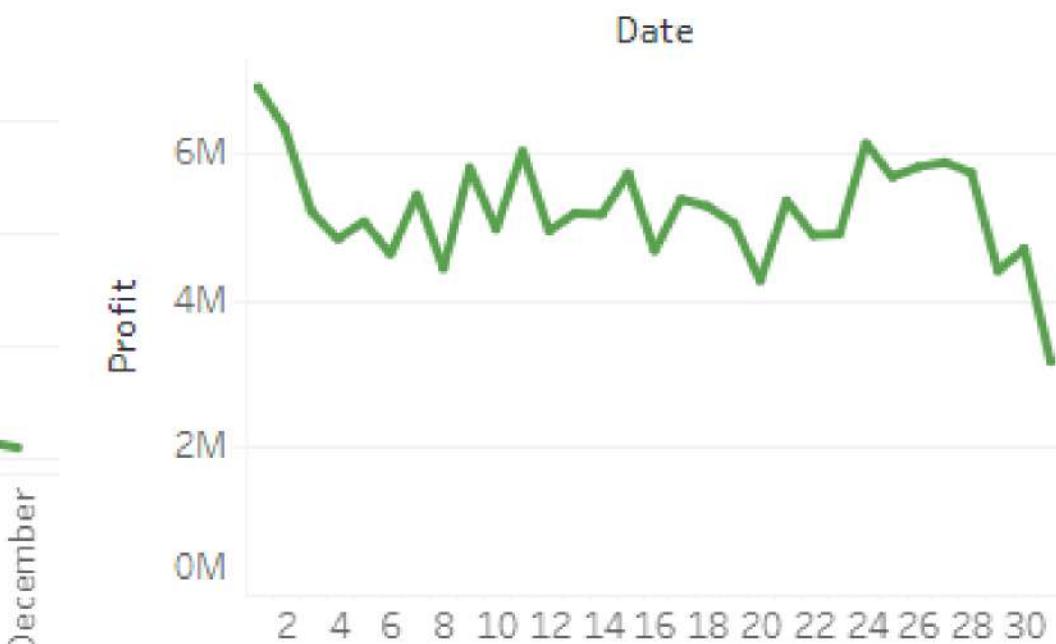


# DASH BOA RD

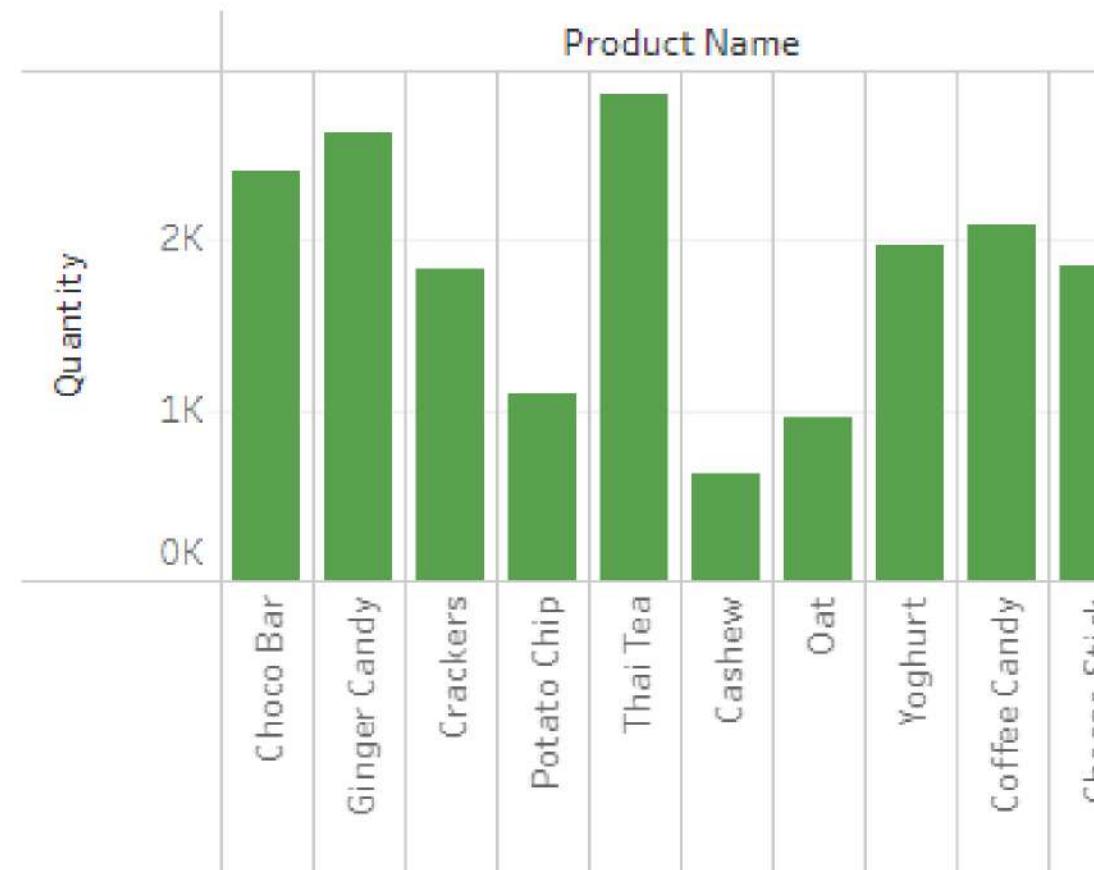
Jumlah Penjualan Barang Bulanan



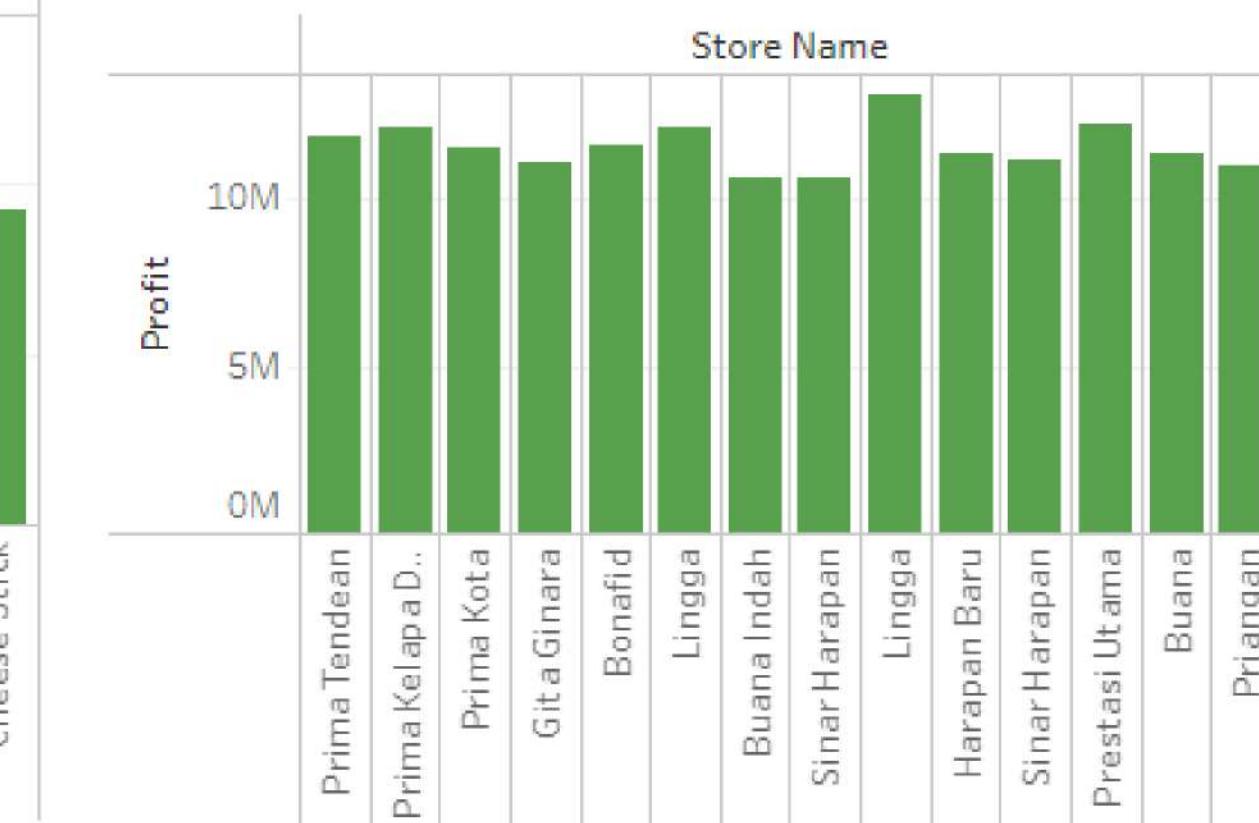
Keuntungan Harian



Jumlah Pembelian Tiap Product



Jumlah Profit Tiap Toko



# PREDIKSI DAN CUSTOMER SEGMENTATION



# DATA

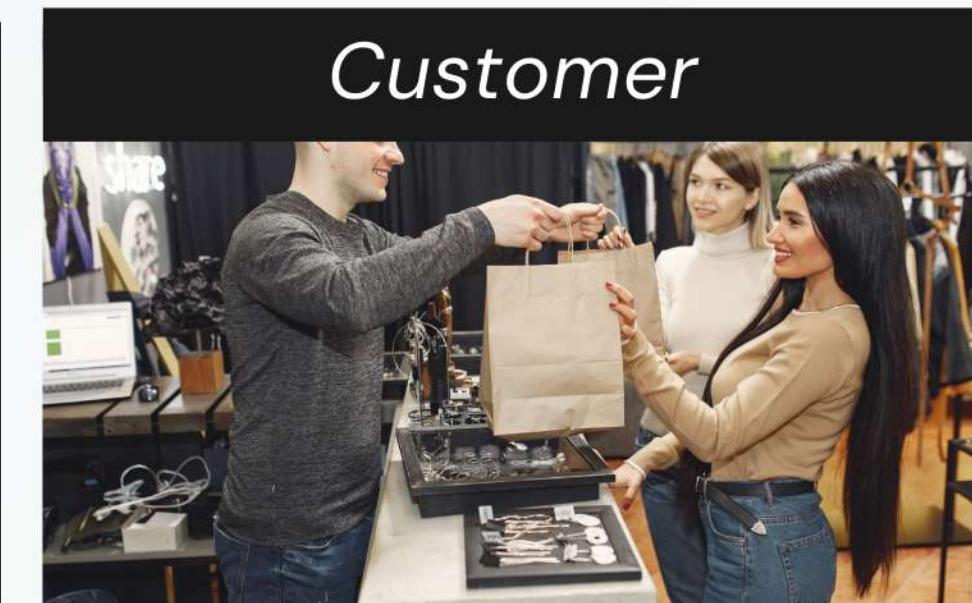
Data yang digunakan merupakan data yang terdiri dari 4 file csv. Data tersebut yaitu data transaksi, data customer, data produk, dan data toko

# DATA



- Data Transaksi merupakan catatan transaksi produk yang telah dilakukan
- Data ini mencatat kode unik transaksi, tanggal transaksi, jumlah item yang dibeli, dan jumlah nominal yang telah dibayarkan

- Data Customer merupakan data dari customer perusahaan yang telah tercatat
- Data ini mencatat id customer, umur customer, gender customer, status pernikahan, dan pendapatan customer



# DATA



- Data Store merupakan catatan terkait toko-toko dilakukan transaksi produk yang diproduksi perusahaan
- Data ini mencatat kode unik toko, nama toko, nama grup toko, tipe toko, kode latitud, dan kode longitude

- Data Product merupakan catatan terkait produk-produk yang telah dipasarkan oleh perusahaan
- Data ini mencatat kode unik produk, nama produk, dan harga produk.



# EDA

EDA atau Exploratory Data Analysis dilakukan untuk mendapatkan pengetahuan dari data yang dapat berupa statistik deskriptif atau pola pada data

# STATISTIK DESKRIPTIF

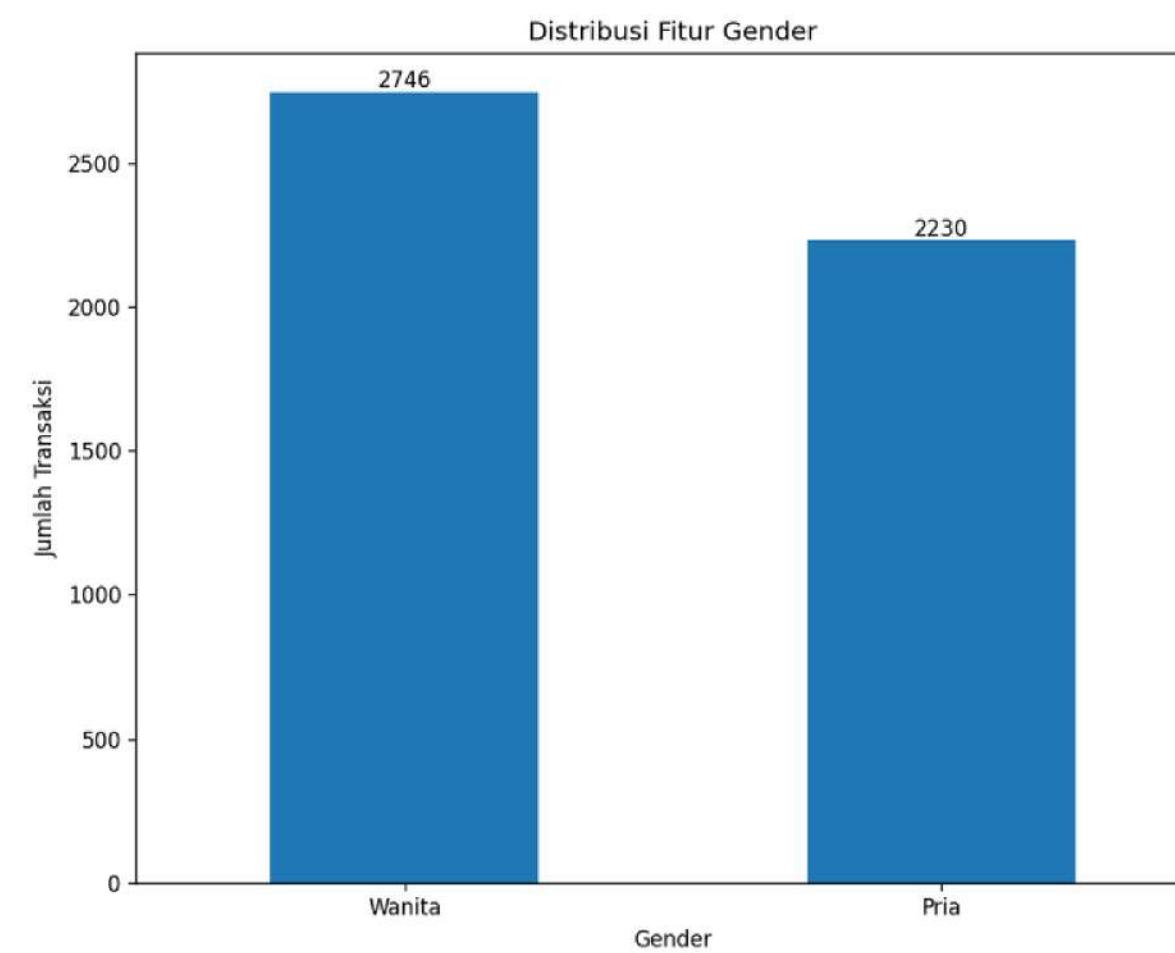
	CustomerID	Price	Qty	TotalAmount	StoreID	Age	Gender	Income	Latitude	Longitude
count	4976.000000	4976.000000	4976.000000	4976.000000	4976.000000	4976.000000	4976.000000	4976.000000	4976.000000	4976.000000
mean	220.597267	9690.293408	3.642484	32268.629421	7.492162	40.079582	0.448151	8.649224	-2.944228	109.602110
std	128.735807	4601.844969	1.855968	19661.512914	4.027835	12.862586	0.497354	6.536447	4.321477	8.368621
min	1.000000	3200.000000	1.000000	7500.000000	1.000000	0.000000	0.000000	0.000000	-7.797068	95.323753
25%	108.750000	4200.000000	2.000000	16000.000000	4.000000	30.000000	0.000000	4.250000	-6.914864	104.756554
50%	220.000000	9400.000000	3.000000	28200.000000	7.000000	39.000000	0.000000	7.720000	-5.135399	110.370529
75%	330.000000	15000.000000	5.000000	47000.000000	11.000000	51.000000	1.000000	10.840000	0.533505	114.590111
max	447.000000	18000.000000	10.000000	88000.000000	14.000000	72.000000	1.000000	71.300000	5.548290	128.190643

# TIPE DATA DAN JUMLAH DATA

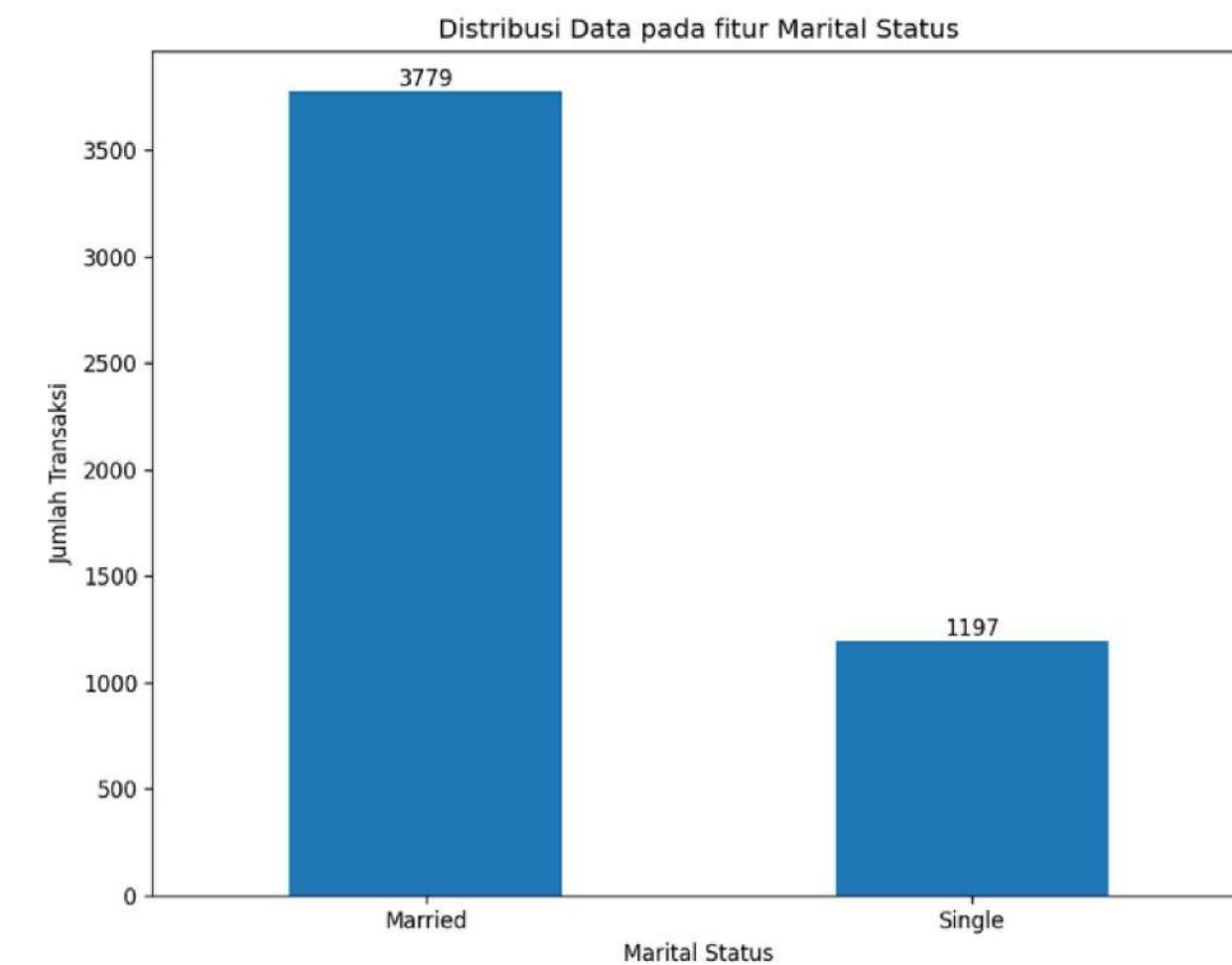
#	Column	Non-Null Count	Dtype
0	TransactionID	4976 non-null	object
1	CustomerID	4976 non-null	int64
2	Date	4976 non-null	datetime64[ns]
3	ProductID	4976 non-null	object
4	Price	4976 non-null	int64
5	Qty	4976 non-null	int64
6	TotalAmount	4976 non-null	int64
7	StoreID	4976 non-null	int64
8	Age	4976 non-null	int64
9	Gender	4976 non-null	int64
10	Marital Status	4976 non-null	object
11	Income	4976 non-null	float64
12	Product Name	4976 non-null	object
13	StoreName	4976 non-null	object
14	GroupStore	4976 non-null	object
15	Type	4976 non-null	object
16	Latitude	4976 non-null	float64
17	Longitude	4976 non-null	float64

dtypes: datetime64[ns](1), float64(3), int64(7), object(7)  
memory usage: 738.6+ KB

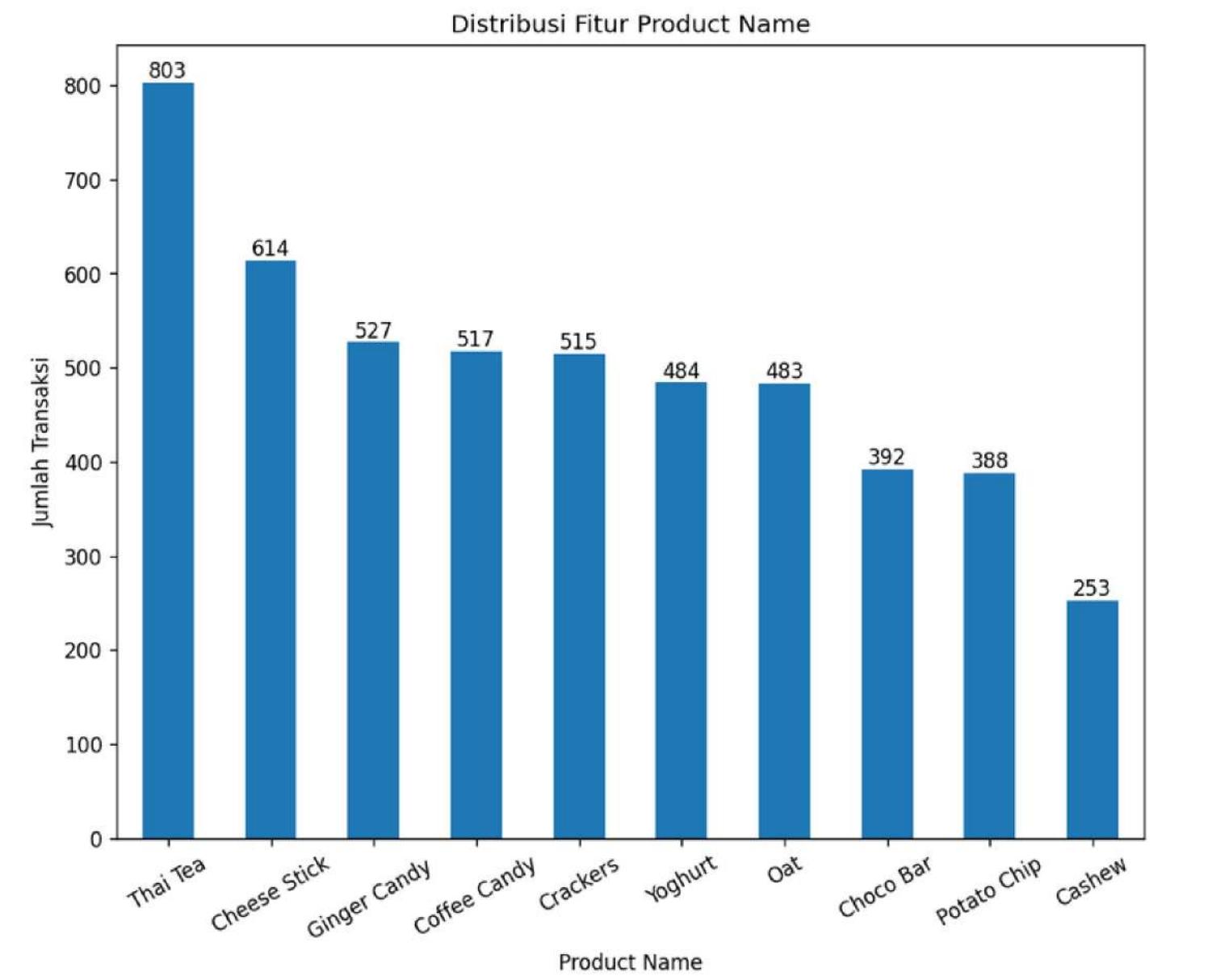
TransactionID	4866
CustomerID	444
Date	365
ProductID	10
Price	10
Qty	10
TotalAmount	44
StoreID	14
Age	54
Gender	2
Marital Status	2
Income	366
Product Name	10
StoreName	12
GroupStore	7
Type	2
Latitude	14
Longitude	14



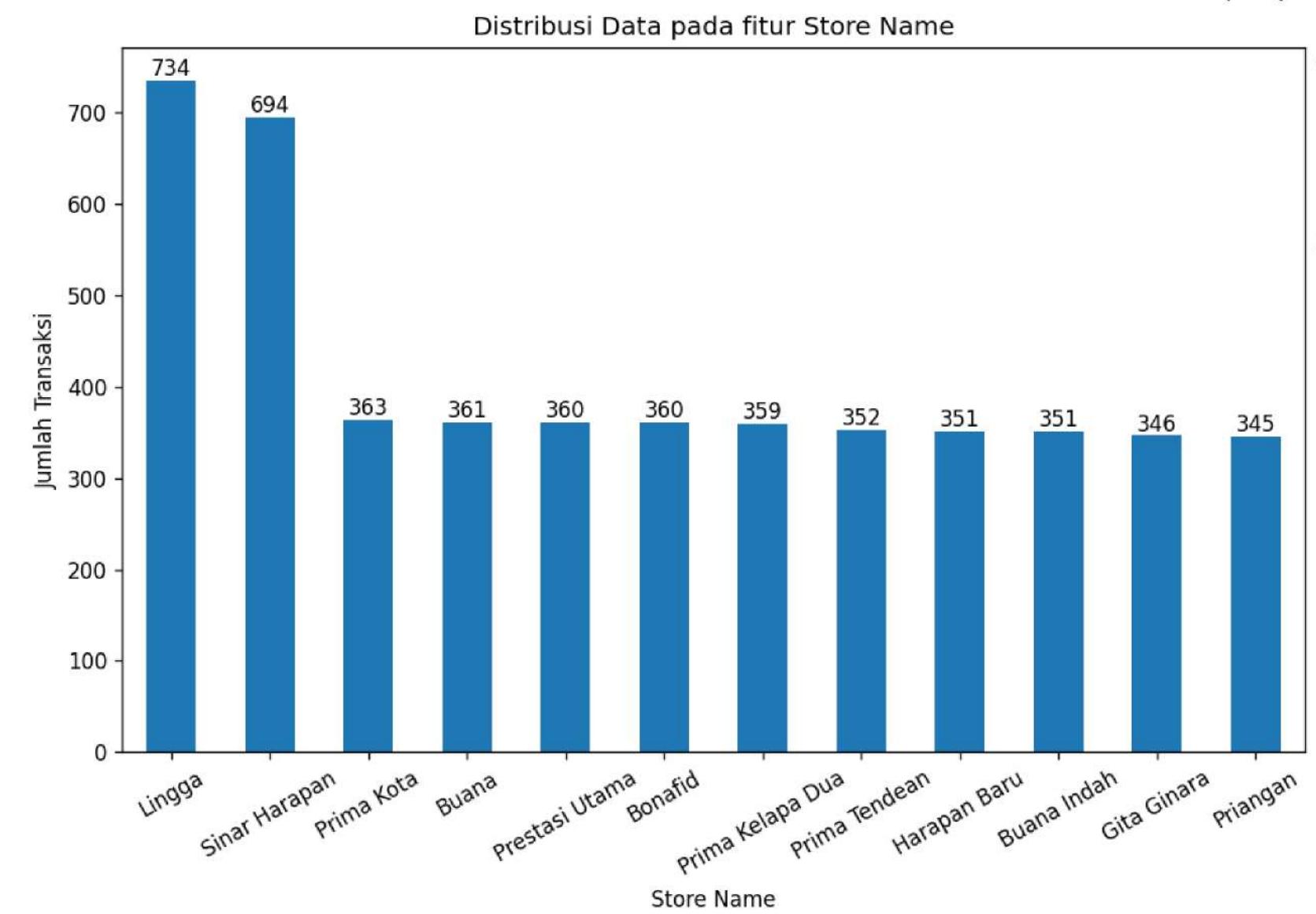
## GENDER



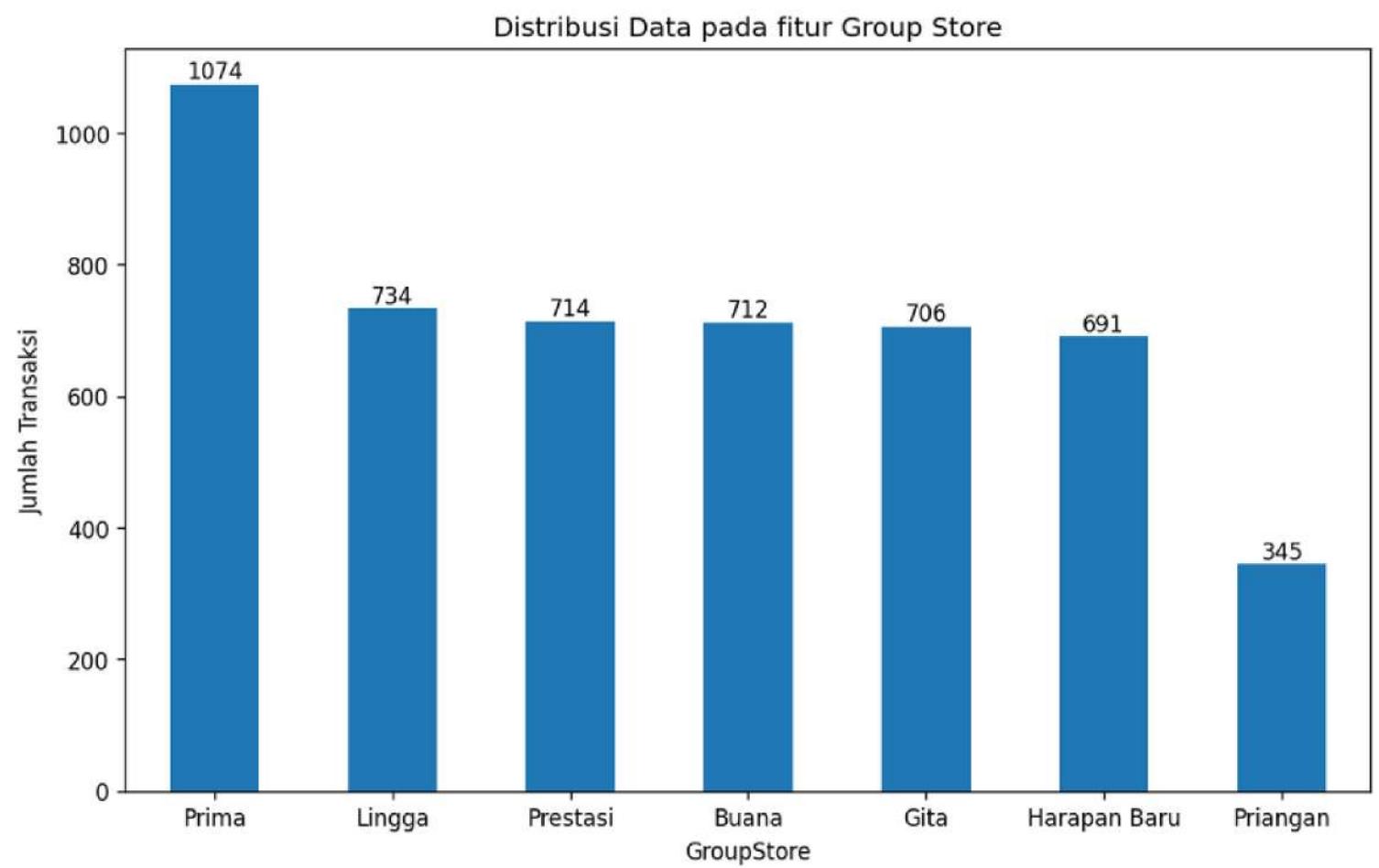
## STATUS PERINKAHAN



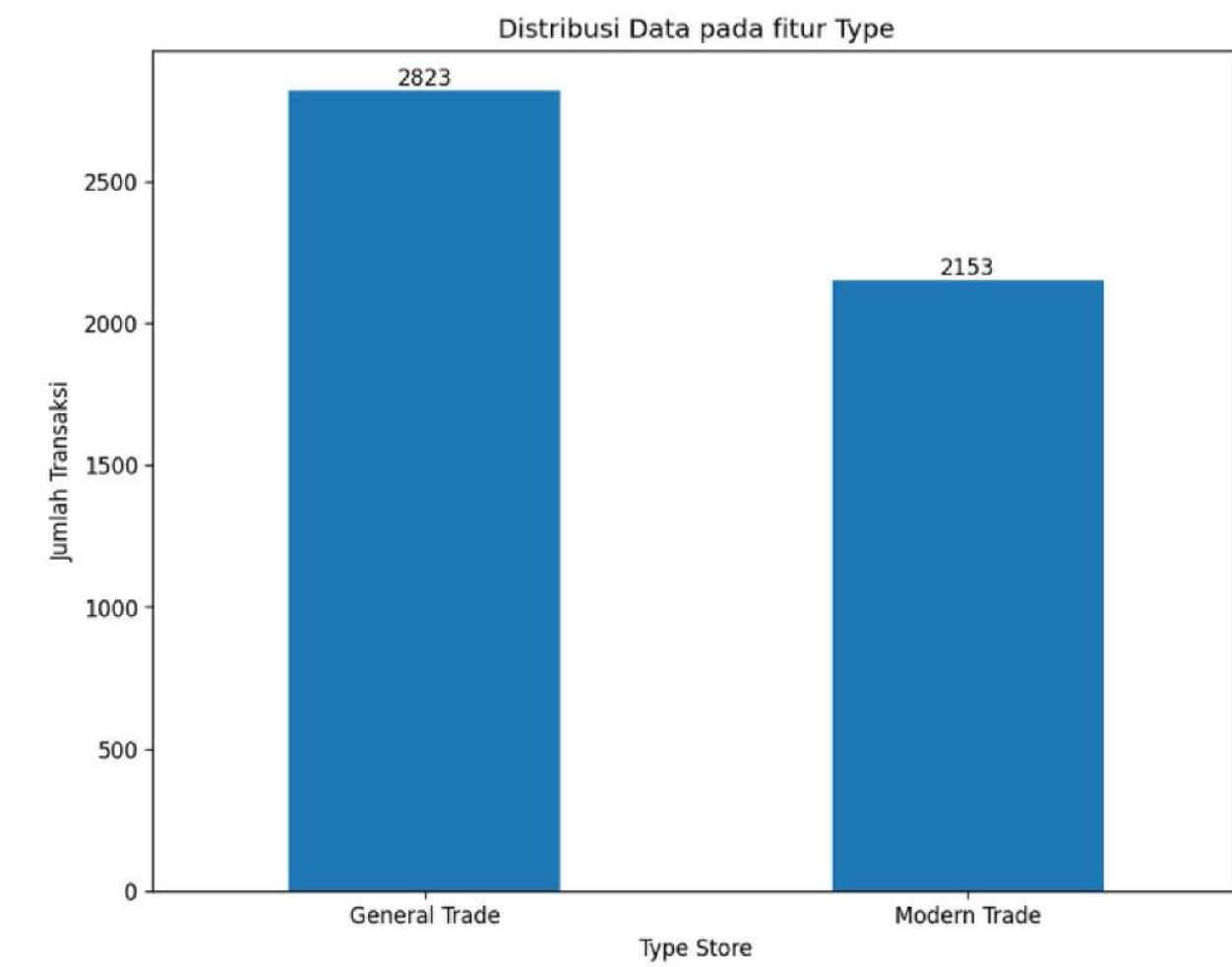
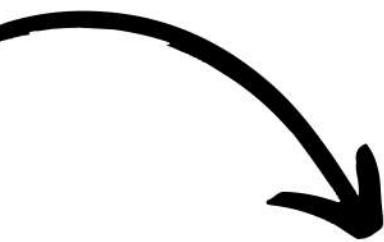
PRODUK



TOKO



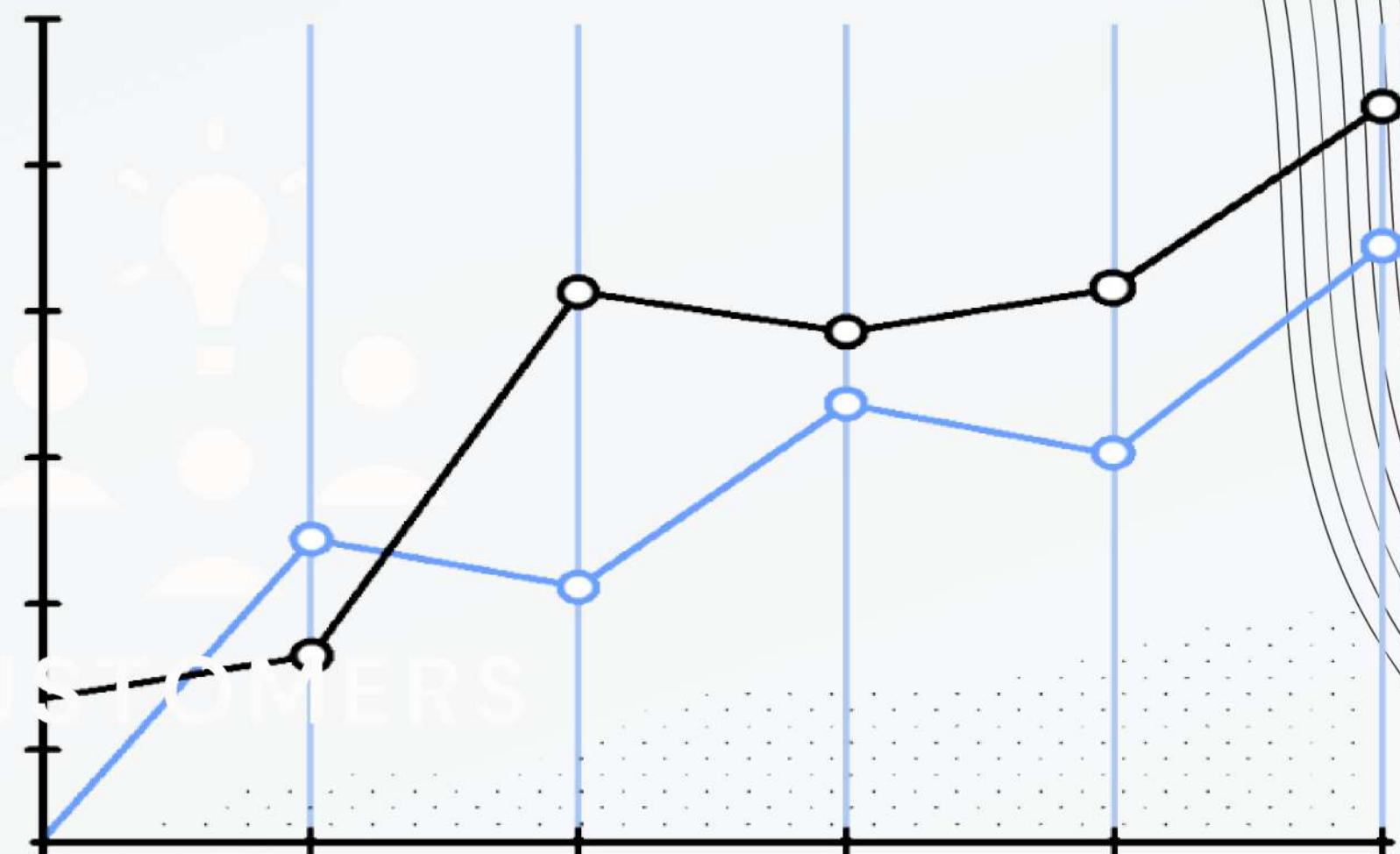
GRUP TOKO



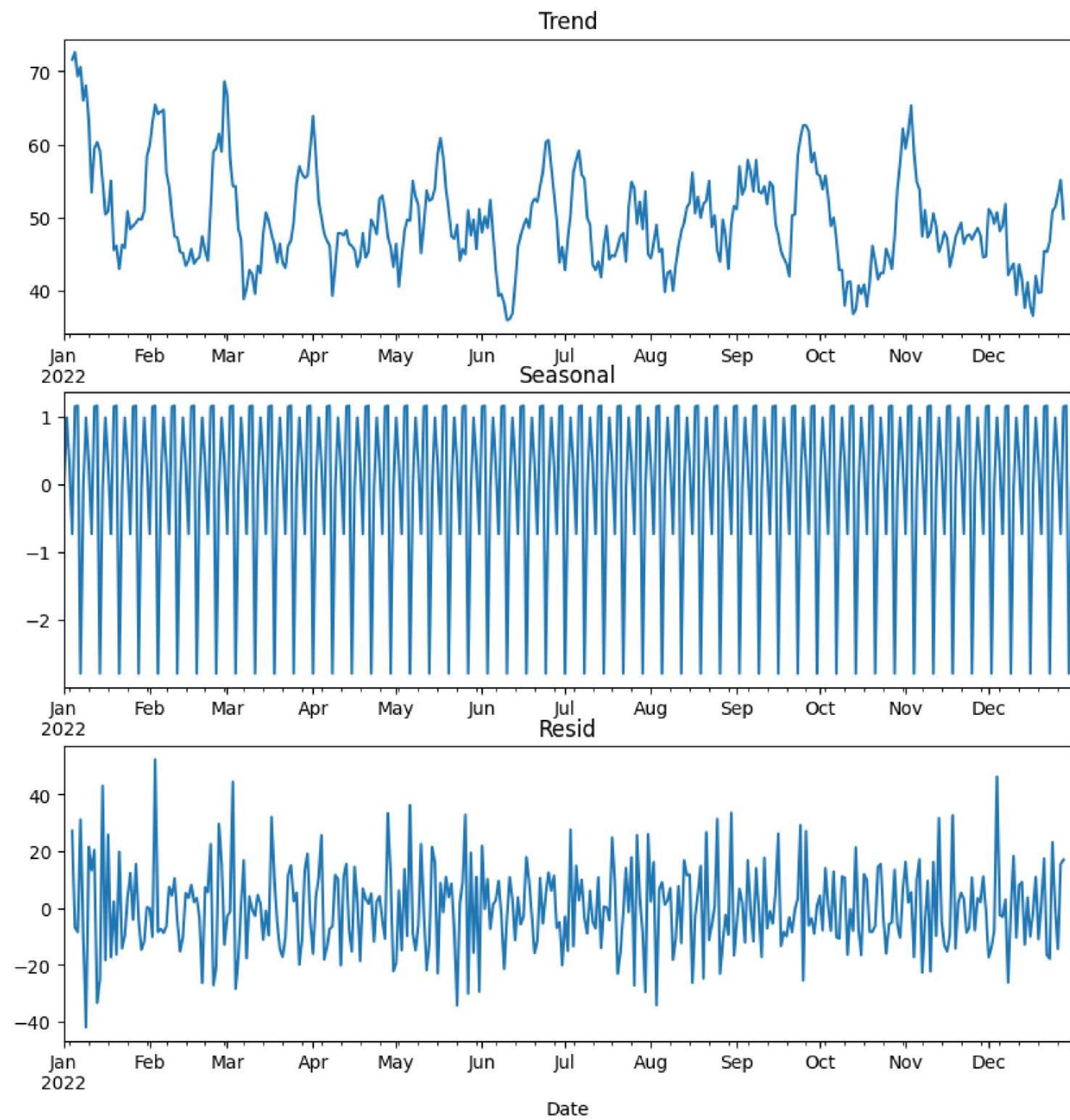
TIPE TOKO

# PREDICTION

- Prediksi digunakan untuk memprediksi stok harian
- Prediksi dilakukan dengan cara membentuk model ARIMA yang kemudian akan melakukan prediksi

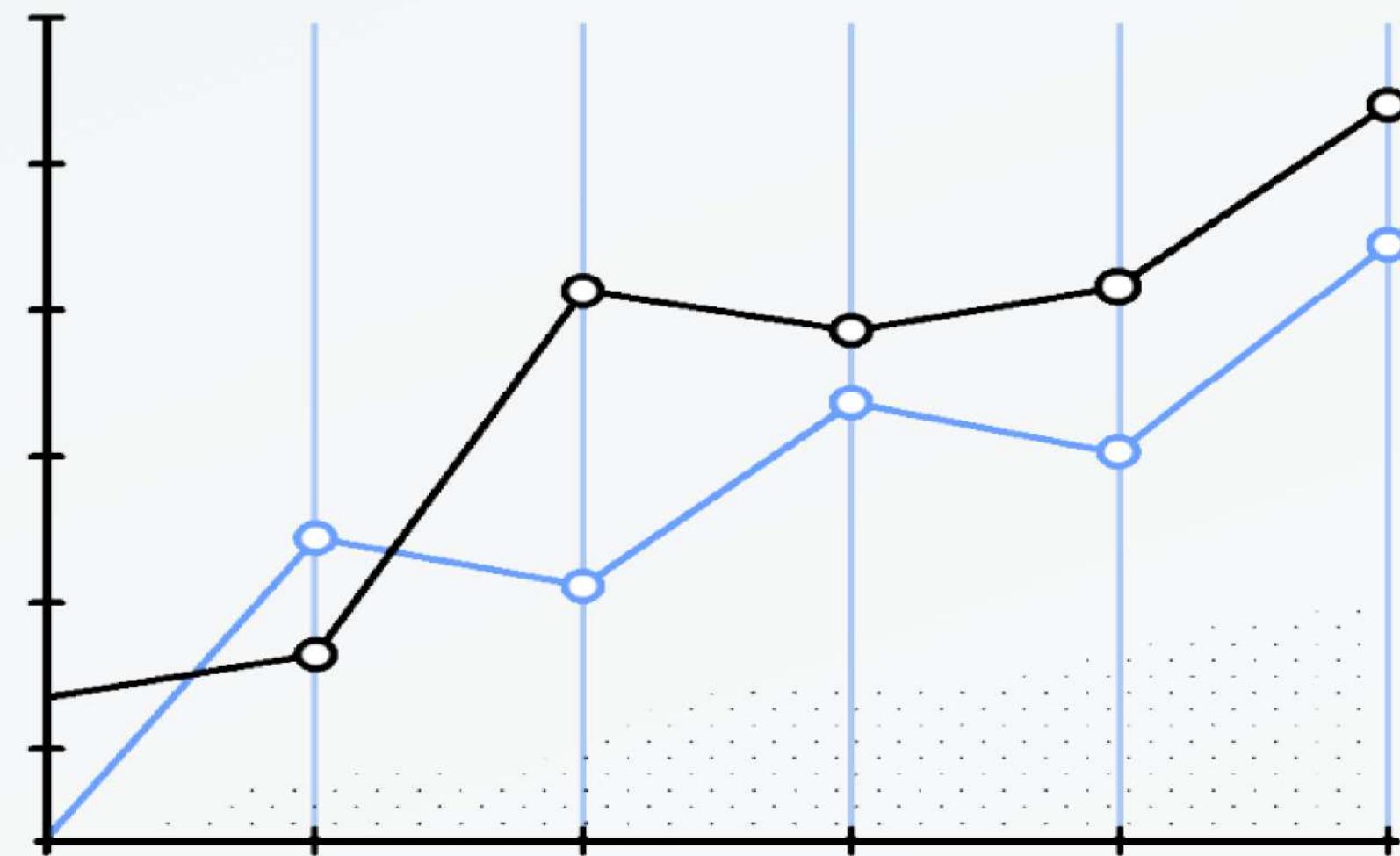


# SEASONAL DECOMPOSE



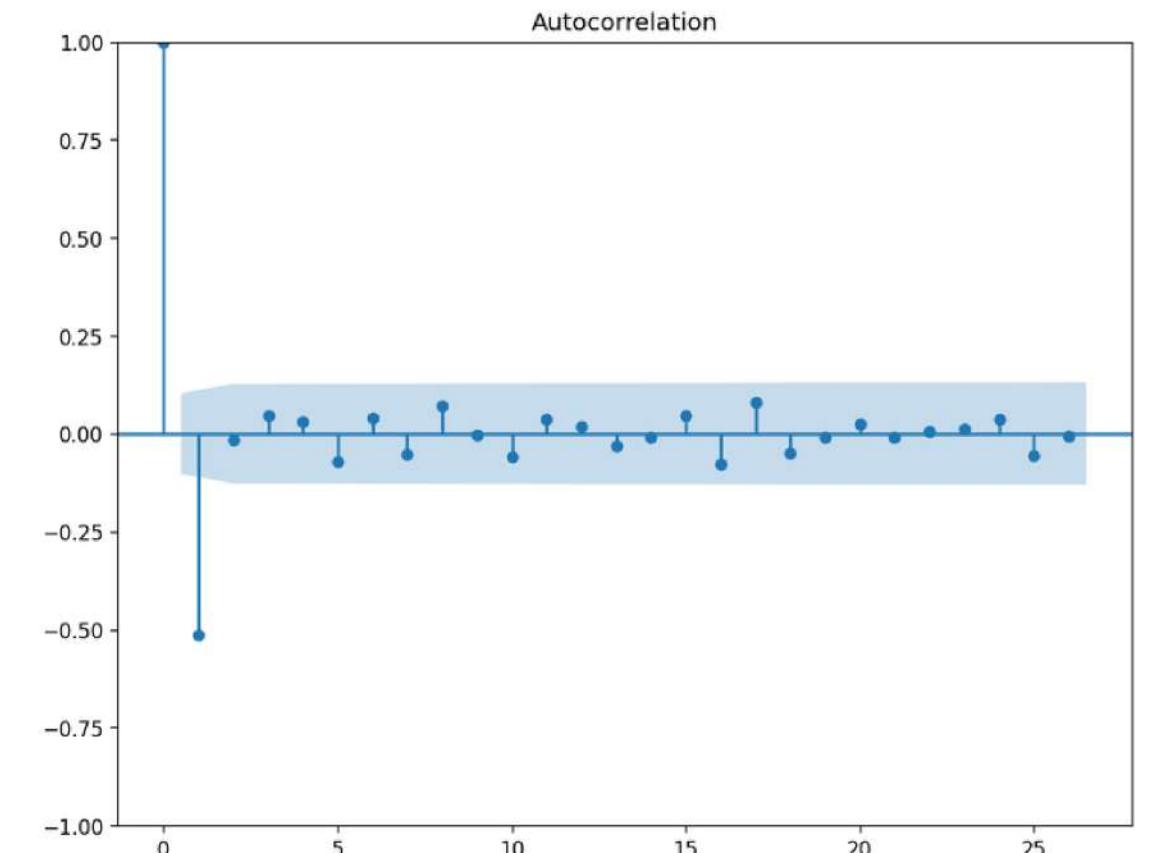
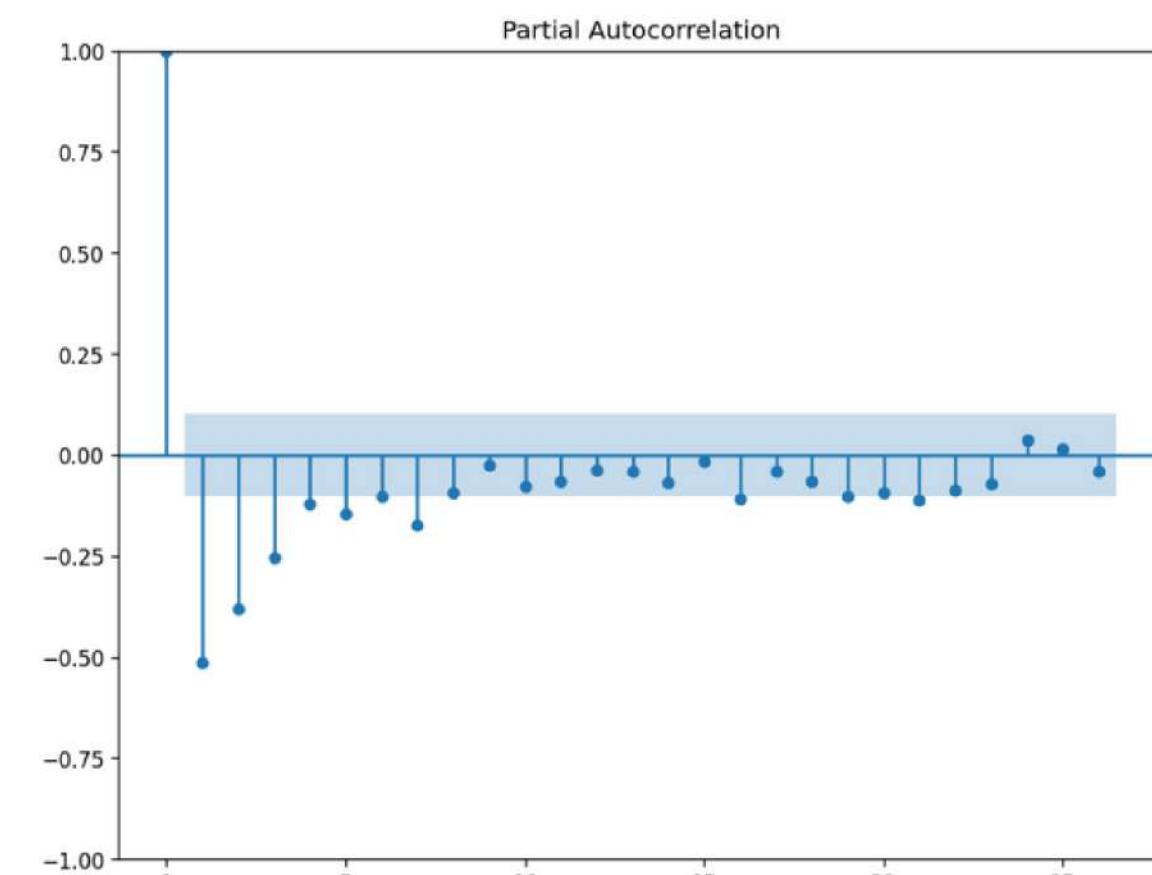
# MODEL ARIMA

- Pencarian parameter p,d,dan q perlu dilakukan untuk membentuk model ARIMA yang baik
- Nilai d otomatis 0 karena merupakan data stasioner



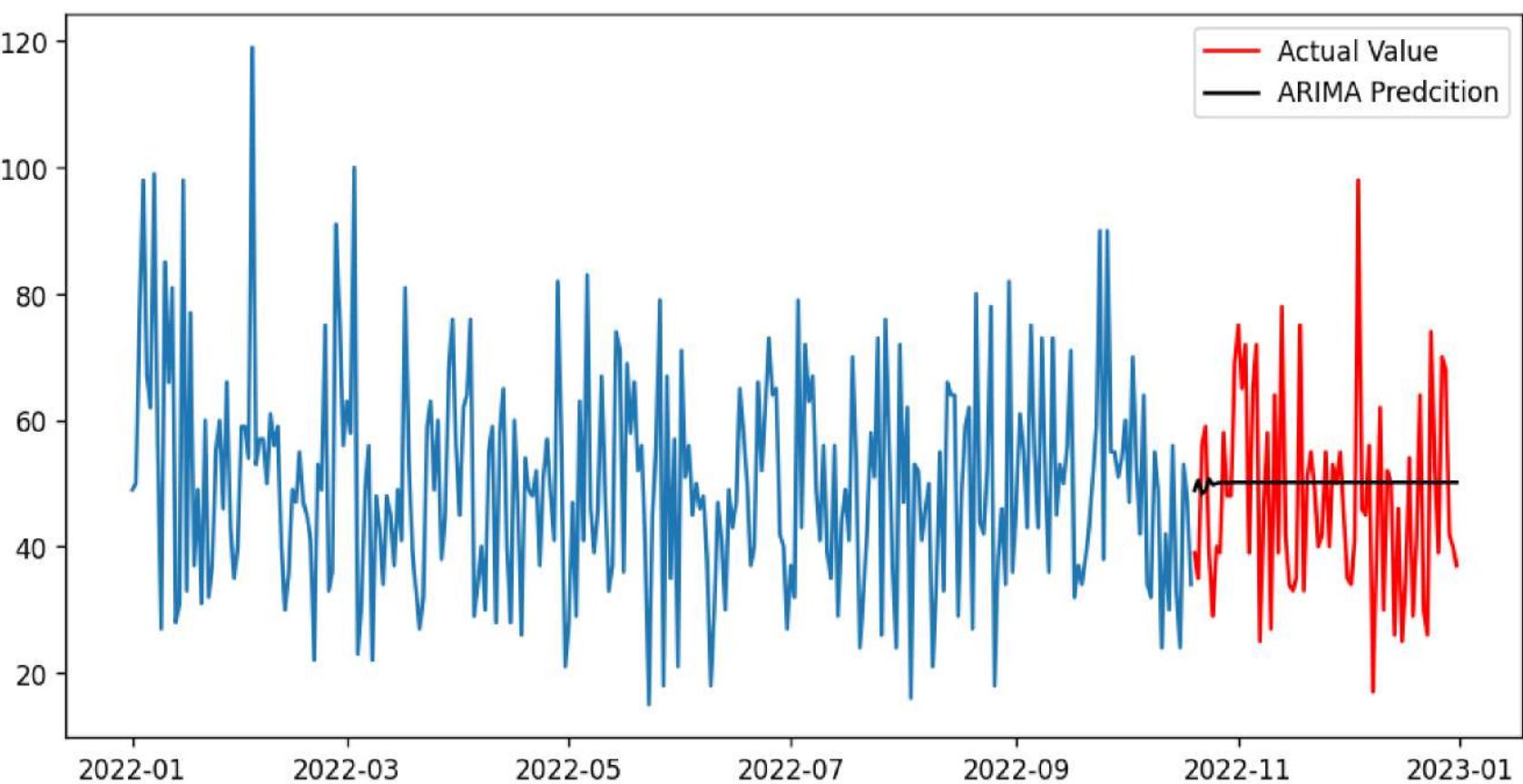
# PENCARIAN P DAN Q

- Dilihat dari grafik partial correlation dan autocorrelation disamping dapat disimpulkan bahwa nilai p yang baik adalah 4 dan nilai q yang baik adalah 2



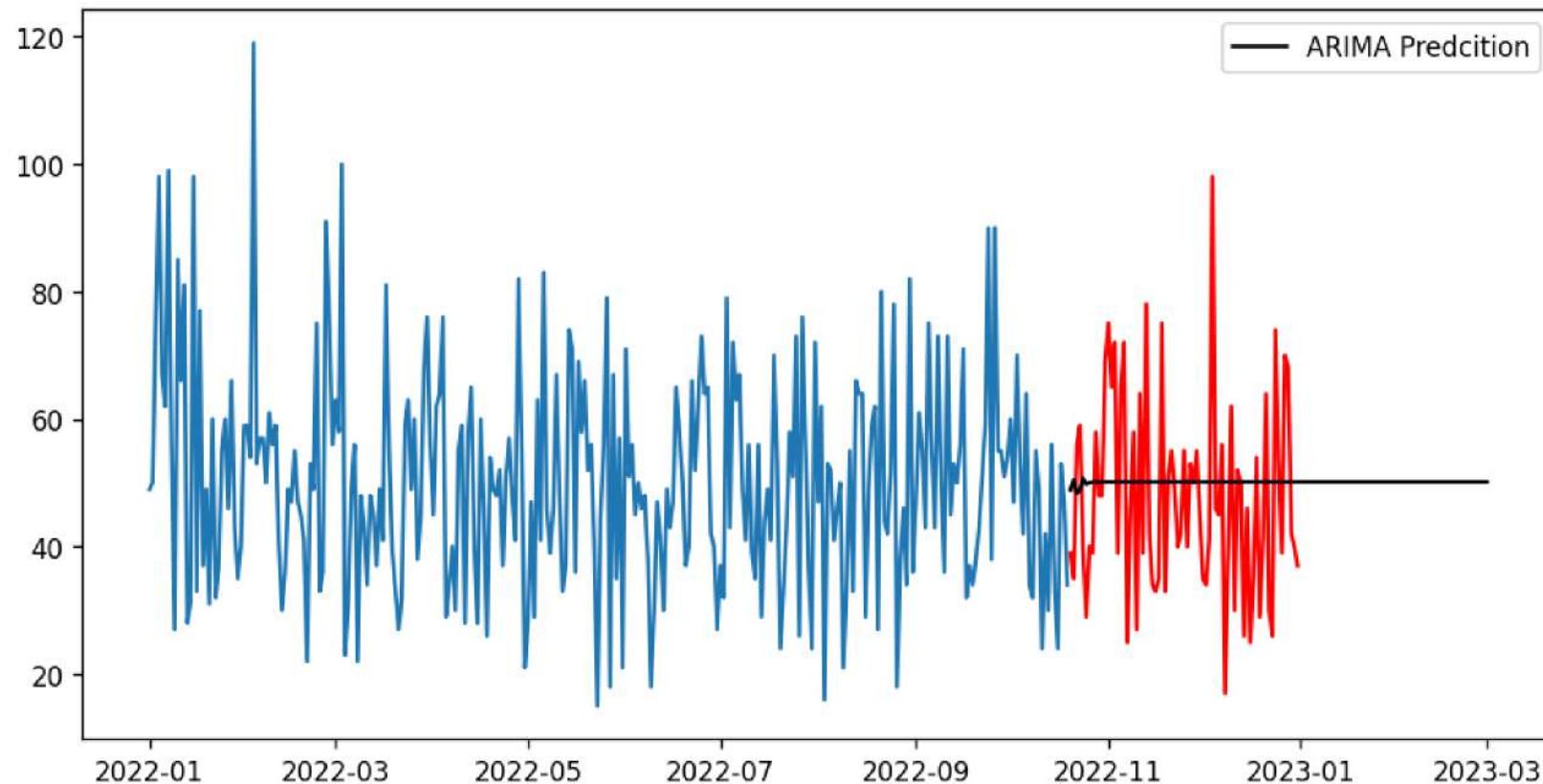
# PENGUJIAN MODEL

- hasil prediksi dari model ARIMA yang telah dibuat terhadap data uji bisa dibilang masih belum cukup bagus karena nilai RMSE yang bernilai 15 lebih dan MAE yang bernilai 12 lebih.



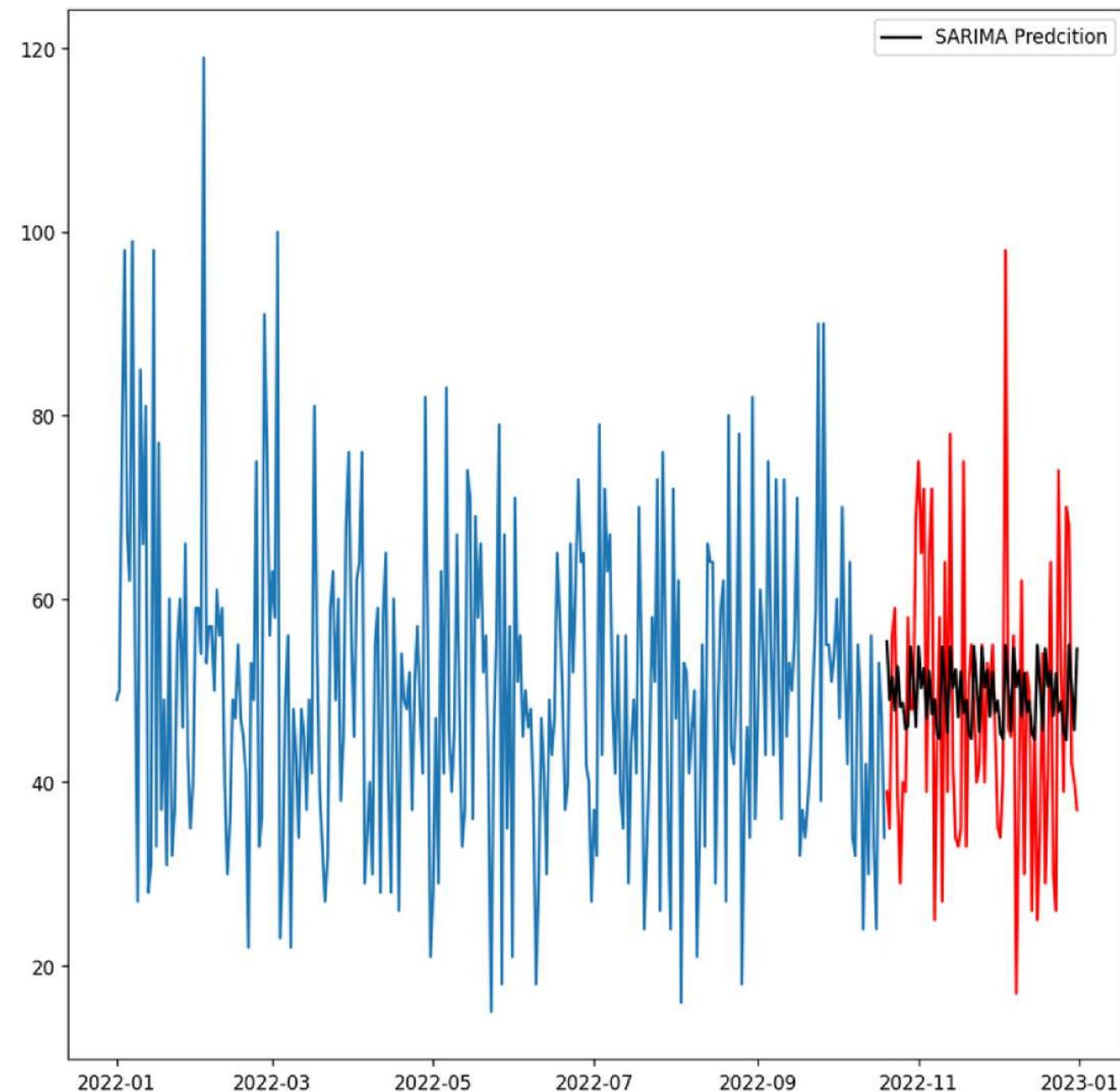
# PREDIKSI

- Dan prediksi untuk jumlah quantity untuk dua bulan kedepan berdasarkan model ARIMA yang telah dibuat yaitu sekitar 50. Tentunya prediksi tersebut bisa dibilang belum cukup baik dan meyakinkan mengingat hasil pengujian yang belum memuaskan.



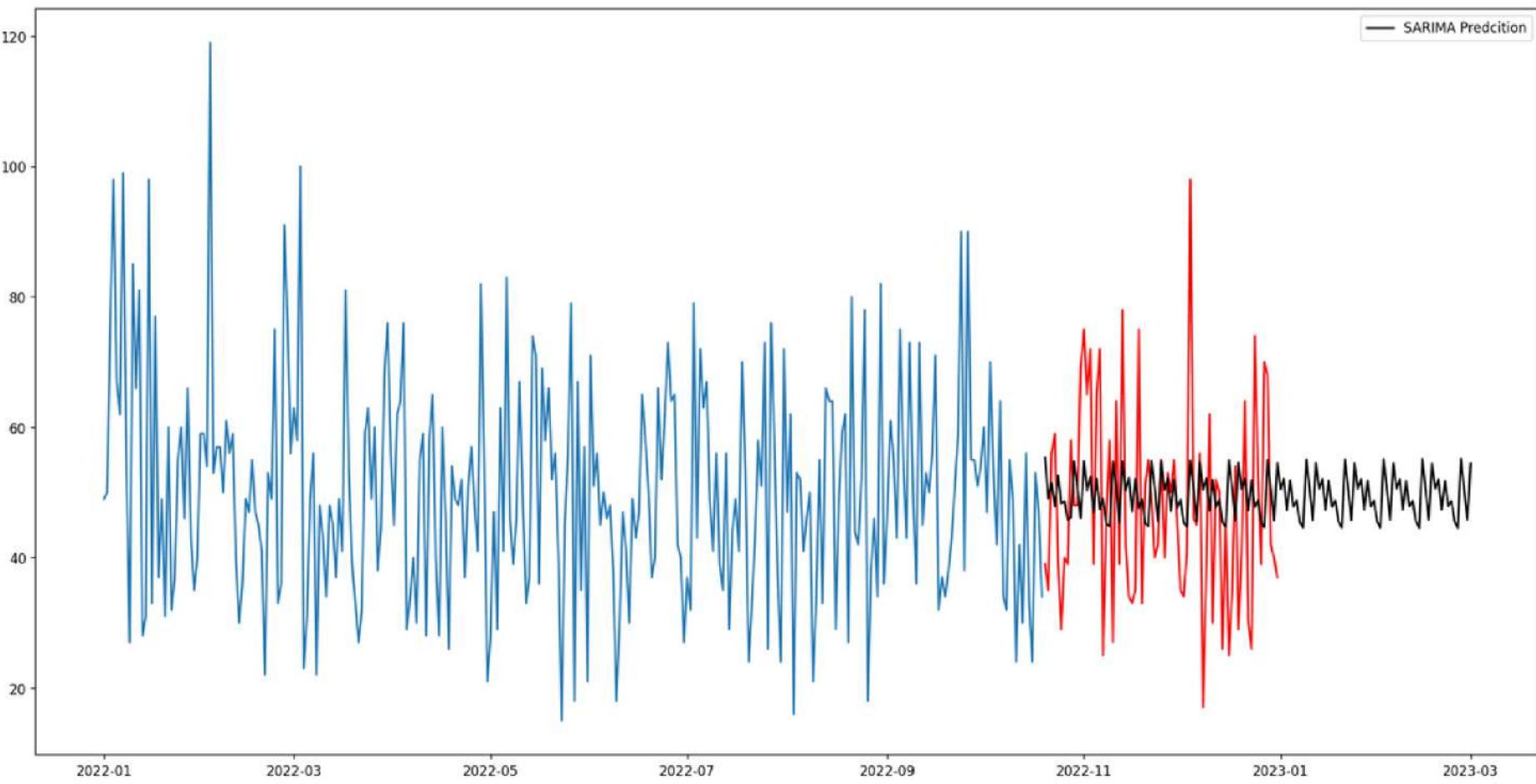
# MODEL SEASONAL ARIMA

- Penggunaan model seasonal ARIMA dapat dilakukan untuk mendapatkan hasil yang baik mengingat data merupakan data seasonal
- Penggunaan model seasonal ARIMA terlihat lebih baik dibandingkan model ARIMA meskipun nilai RMSE dan MAE tidak terlihat begitu jauh.



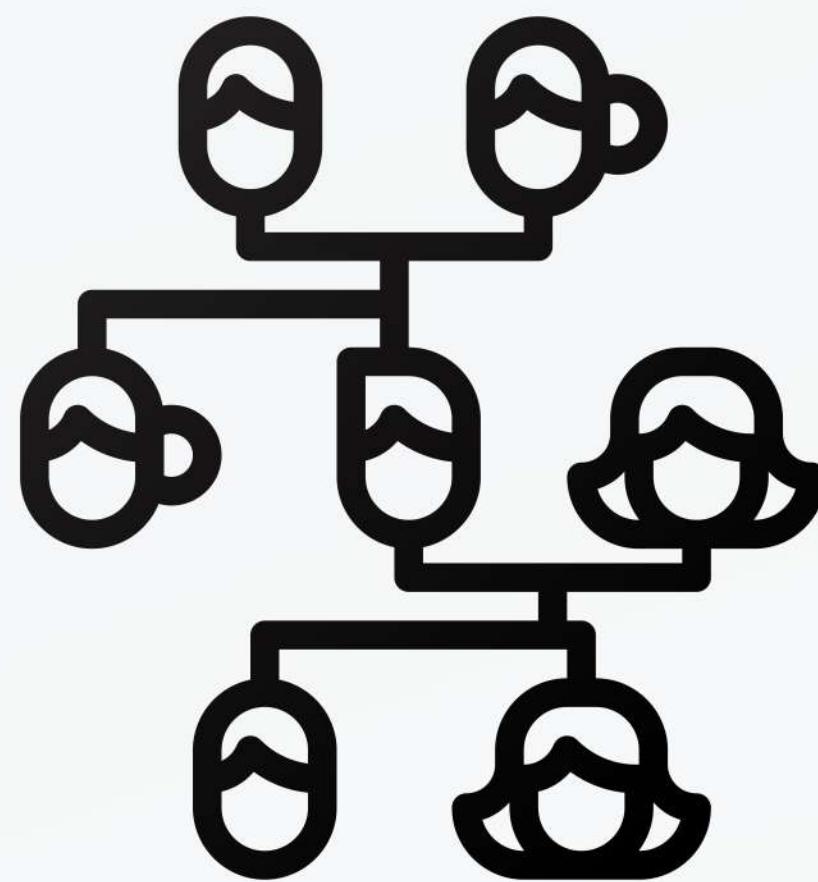
# PREDIKSI DENGAN SEASONAL ARIMA

- Hasil prediksi jumlah kauntitas atau jumlah barang untuk 2 bulan kedepan dengan menggunakan model seasonal ARIMA atau SARIMA adalah sekitar 50-55 barang. Tentunya prediksi yang dihasilkan dari model ini bisa dibilang lebih terpercaya dibanding model ARIMA sebelumnya dilihat dari hasil pengujian yang telah dilakukan walau peningkatannya tidak terlalu signifikan.



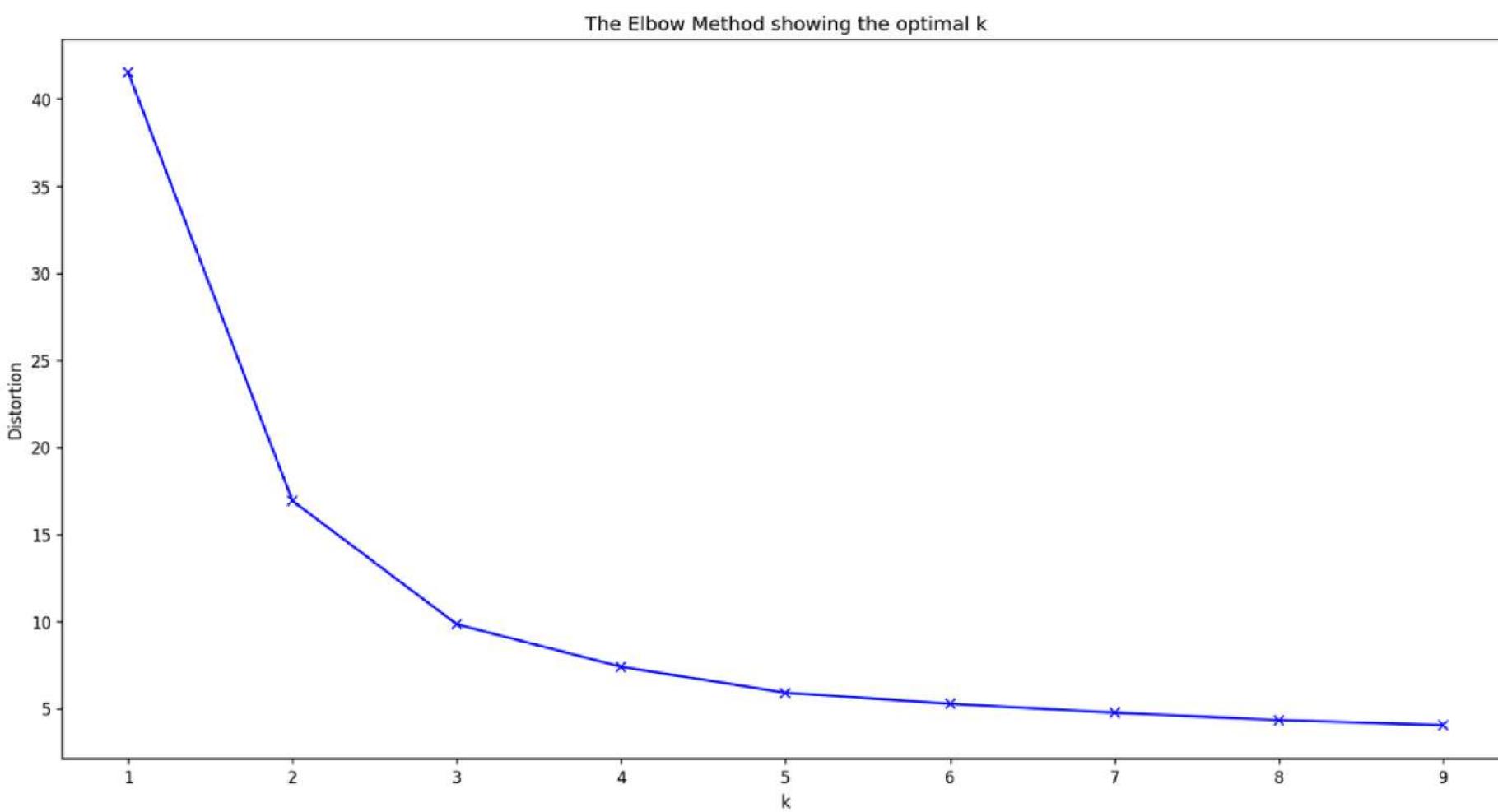
# CUSTOMER SEGMENTATION

- Customer segmentation dilakukan untuk membagi customer kedalam kelompok-kelompok berdasarkan kemiripan perilaku
- Hal ini dilakukan untuk membantu proses promosi produk yang akan dilakukan
- Customer segmentation dilakukan dengan membentuk cluster-cluster dengan bantuan algoritma KMeans



# MENCARI NILAI K

- Hal ini dilakukan untuk mendapatkan nilai k atau jumlah cluster yang optimal



# CLUSTERING

- clustering dilakukan dengan membentuk model KMeans dengan menggunakan nilai k hasil elbow method
- Evaluasi hasil clustering dilakukan dengan menghitung silhouette score.
- Hasil penghitungan silhouette score pada hasil clustering model yang telah dibuat menunjukkan hasil yang baik yaitu 0.835

Cluster	Count_Transaction	Total_Quantity	TotalAmount
0	0	207	92.753623 73.432547
1	1	108	74.463768 61.881262
2	2	129	31.115942 23.281871

# ANALISIS HASIL SEGMENTASI

- cluster pertama merupakan cluster yang sering melakukan transaksi dan menghasilkan keuntungan yang paling besar.
- cluster selanjutnya merupakan cluster yang sering melakukan transaksi namun tidak sesering cluster pertama dan keuntungan yang didapatkan juga paling sedikit dikarenakan barang yang dibeli juga tidak terlalu banyak dilihat dari total quantitas.
- cluster yang terakhir merupakan cluster dengan jumlah transaksi paling sedikit namun menghasilkan keuntungan yang lebih besar daripada cluster kedua dan jumlah barang yang dibeli juga lebih banyak dibandingkan cluster kedua.

# SARAN PROMOSI

- Mempromosikan produk yang lebih menghasilkan profit pada cluster pertama dengan memberikan promo-promo yang bergandengan seperti memberikan promo pada pembelian produk yang sering dibeli dengan produk yang menghasilkan profit paling banyak.
- Menawarkan produk-produk baru serta memberikan diskon yang terus berubah-ubah dalam jangka waktu tertentu sehingga mereka lebih sering membeli untuk mengincar diskon-diskon yang berbeada-beada pada cluster 3
- memberikan promo bonus seperti 2 gratis 1 pada cluster 2. Hal ini dilakukan mengingat cluster ini sering melakukan pembelian namun jumlah barang yang dibeli tidak terlalu banyak

**SEKIAN,  
TERIMA KASIH**

