Rakamin Academy X KALBE Nutritionals

Judul Project

Data Scientist Intern

Farid Wujdi Mubarok

Farid Wujdi Mubarok

About Me;

Saya merupakan seorang Fresh Graduate jurusan Ilmu Komputer. Saya memiliki ketertarikan dalam bidang data terutama data analis dan data scientis. Saat ini saya sedang fokus untuk mengembangkan kemampuan saya terutama kemampuan analisis dan pemrograman.

KALBE Nutritionals

Rakamin Academy

Case Study

Sebagai seroang Data Scientist di Kalbe Nutritionals mendapatkan tugas dari tim inventory dan tim marketing.

Dari tim inventory, 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, diminta untuk membuat cluster/segment customer berdasarkan beberapa kriteria.

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

KALBE Nutritionals

Rakamin Academy

EDA dengan PostgreSQL

Exploratory Data Analysis dengan Postgres SQL menggunakan dBeaver

KALBE Nutritionals

Rakamin Academy

Query 1:
Rata-rata umur customer berdasarkan status pernikahan.

```
select
    "Marital Status",
    concat(round(avg(age)), ' Tahun') "AVG Age"
from
    customers c
where
    "Marital Status" not in ('')
group by 1
```

	Marital Status	AVG Age
1	Married	43 Tahun
2	Single	29 Tahun

Query 2:
Rata-rata umur customer berdasarkan jenis kelamin mereka.

```
select
    case
        when gender = 0 then 'Wanita'
        when gender = 1 then 'Pria'
    end gender,
    concat(round(avg(age)), ' Tahun') "AVG Age"
from
    customers c
group by
    1
```

	gender	AVG Age
1	Wanita	40 Tahun
2	Pria	39 Tahun

Exploratory Data Analysis dengan Postgres SQL menggunakan dBeaver

KALBE Nutritionals

Rakamin Academy

Query 3:
Toko dengan jumlah penjualan tertinggi.

```
select
    s.storename,
    sum(t.qty) "Sum QTY"
from
    stores s
join transactions t on
    s.storeid = t.storeid
group by 1
order by 2 desc
```

	storename	Sum QTY
1	Lingga	2,777
2	Sinar Harapan	2,588
3	Prestasi Utama	1,395
4	Prima Kota	1,358

Query 4:
Produk dengan jumlah pendapatan tertinggi.

```
select
    p."Product Name",
    sum(t.totalamount) as "total amount"
from
    products p
join transactions t on
    p.productid = t.productid
group by 1
order by 2 desc
```

	Product Name	total amount
1	Cheese Stick	27,615,000
2	Choco Bar	21,190,400
3	Coffee Candy	19,711,800

KALBE Nutritionals Rakamin Academy

Data Visualization dengan Tableau

Kalbe Sales Dashboard

KALBE Nutritionals

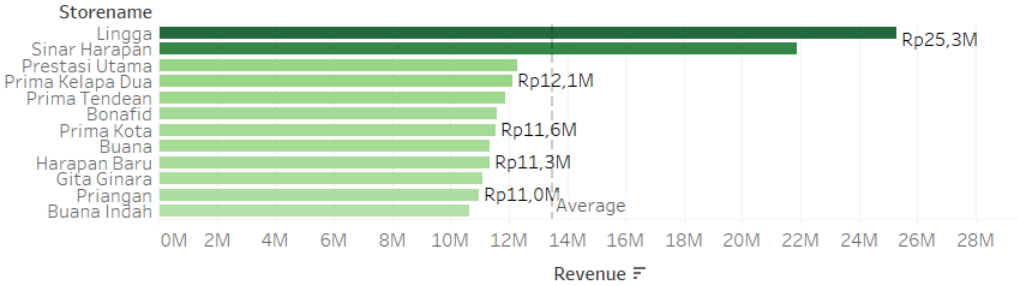
Rakamin Academy

Date
All values

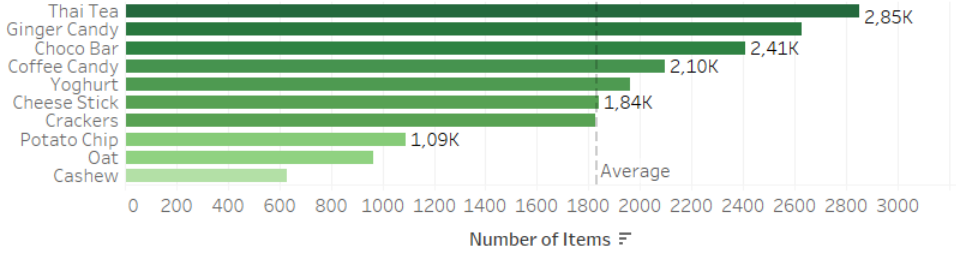
Number of Customers	Total Transactions
447	5,020

Number of Items Sold	Revenue
18,30K	Rp162,04M

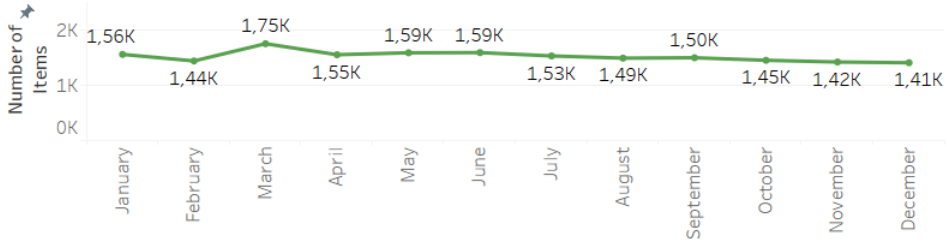
Total Sales Revenue by Stores



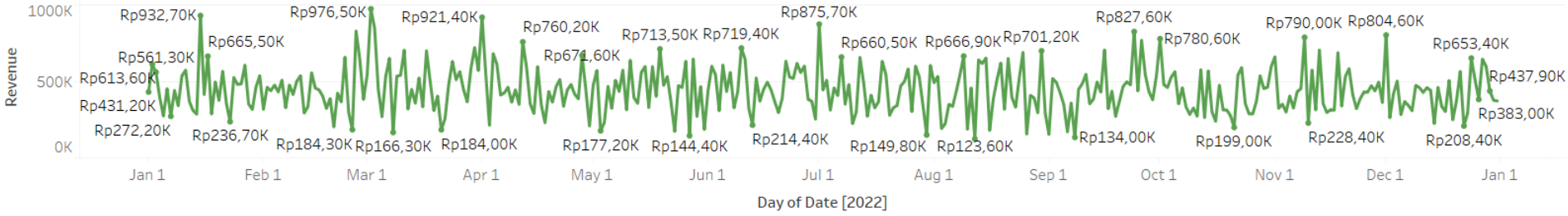
Number of Items Sold by Products



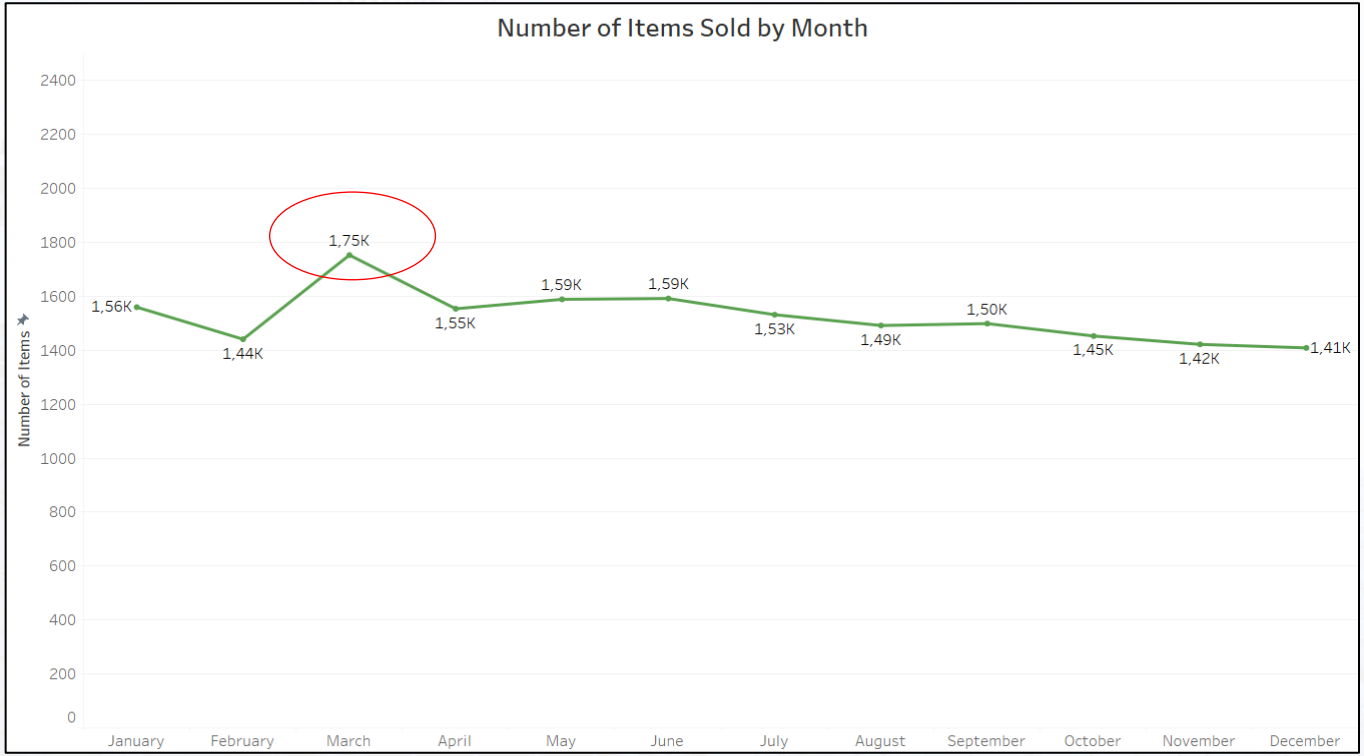
Number of Items Sold by Month



Total Sales Revenue by Day

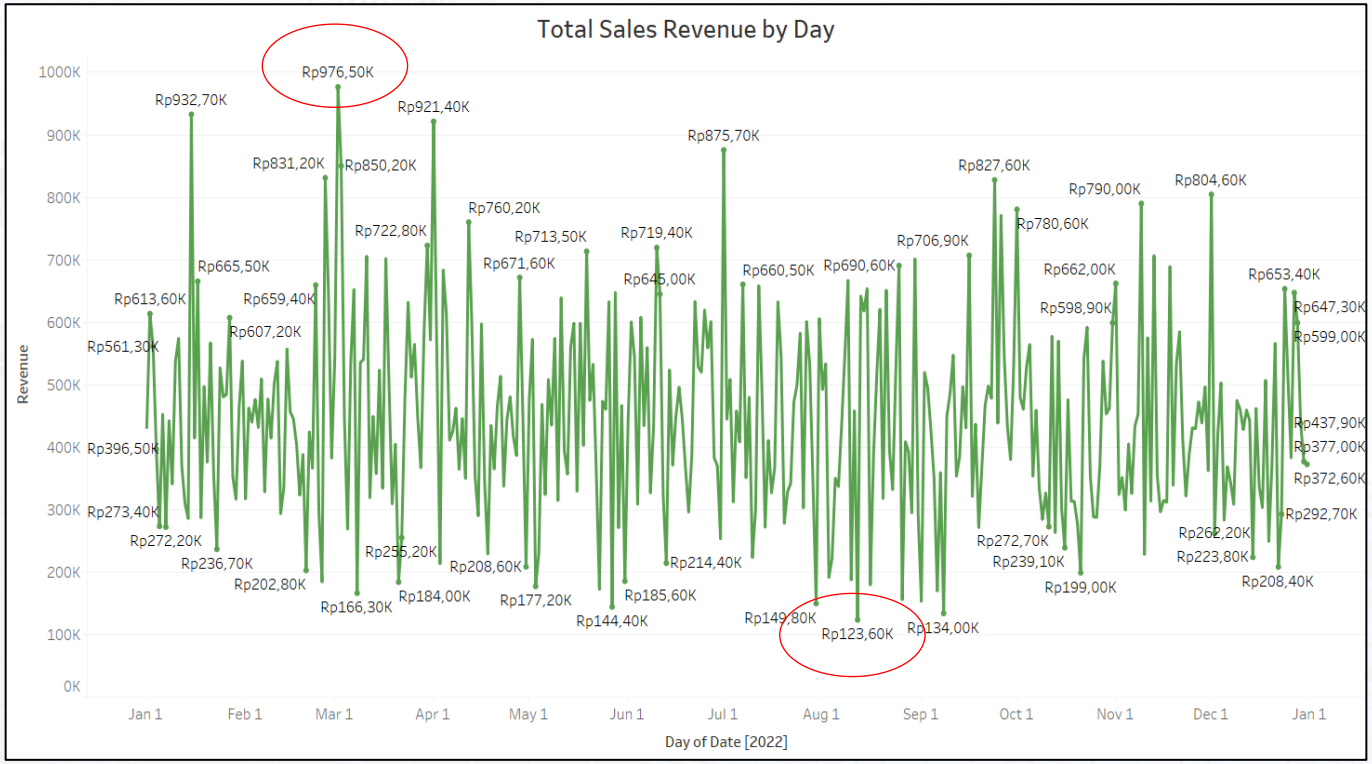


Data Visualization dengan Tableau



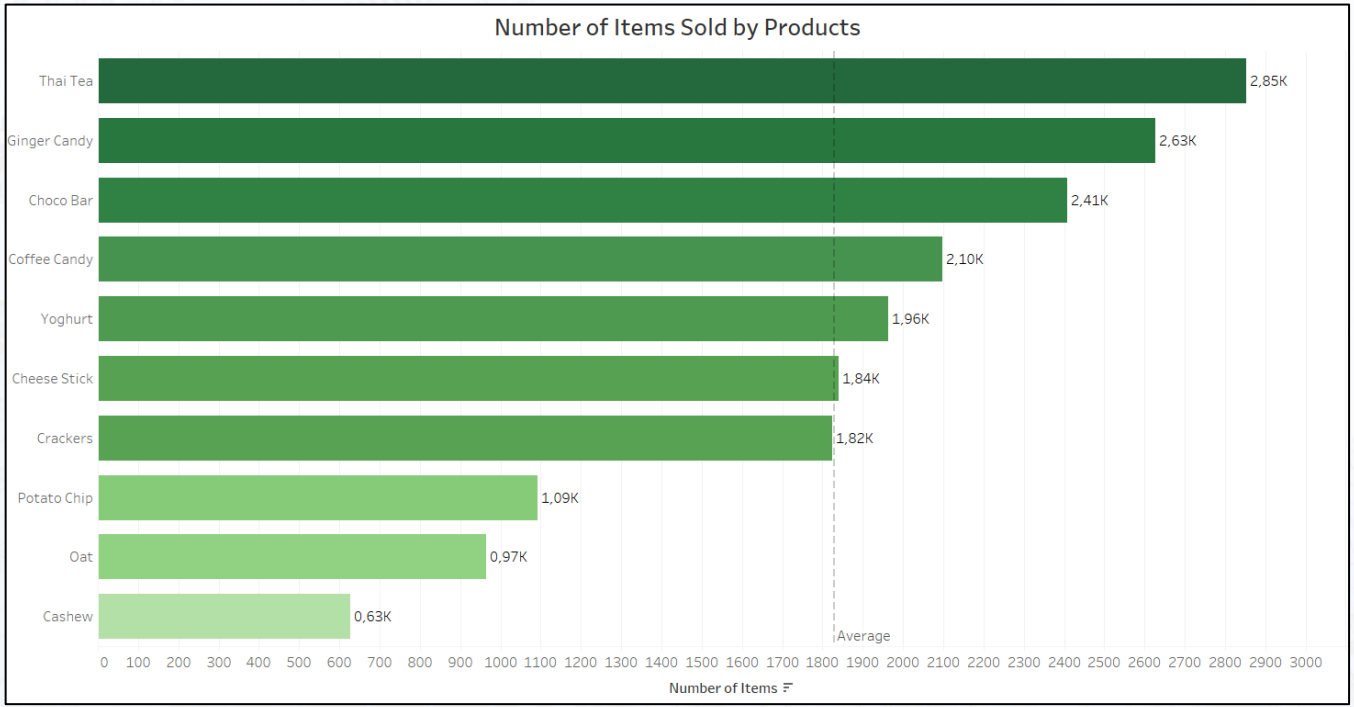
Dalam analisis grafik penjualan yang terlampir, dapat diidentifikasi bahwa puncak penjualan terjadi pada bulan Maret 2022, sementara titik terendah tercatat pada bulan Desember 2022. Selain itu, ditemukan tren penurunan secara berkelanjutan dalam penjualan mulai dari bulan Juni.

Data Visualization dengan Tableau



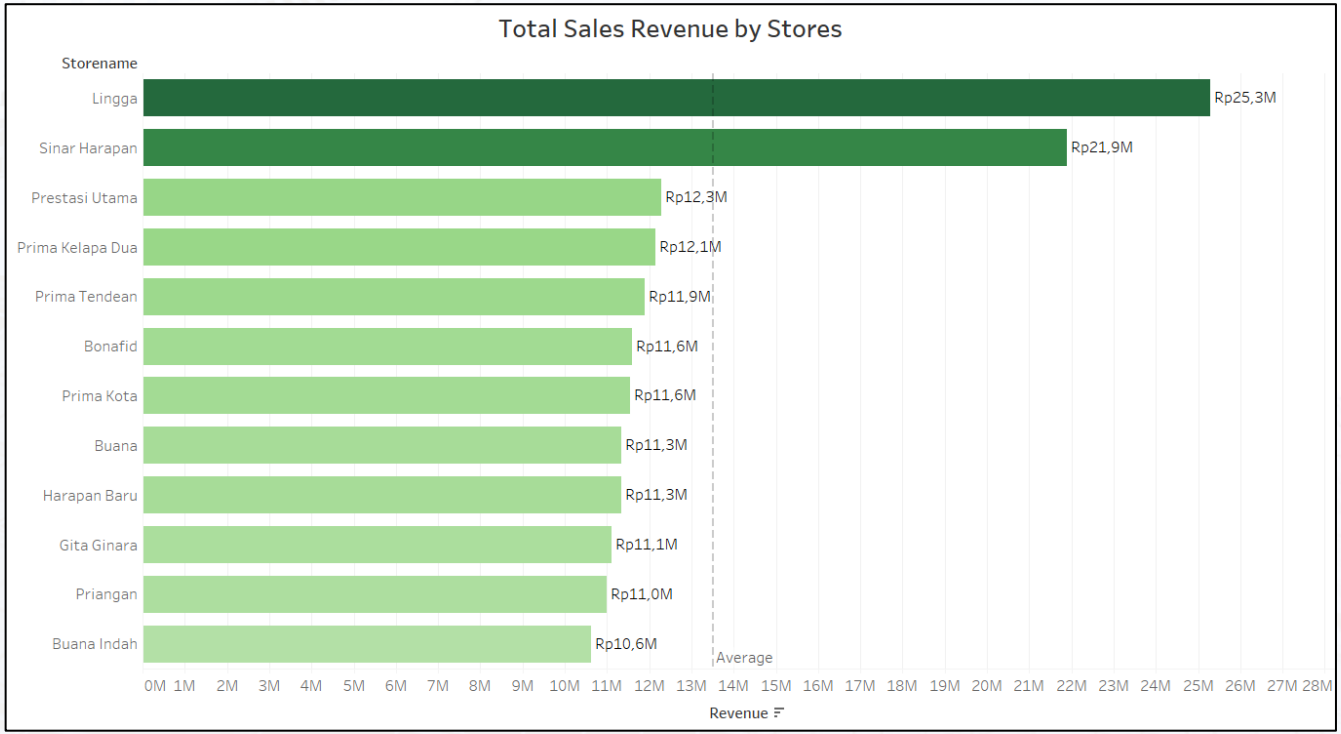
Dari gambar disamping, pendapatan tertinggi ada pada tanggal 2 Maret 2022 dengan total pendapatan Rp976,50K dan pendapatan terendah ada pada tanggal 12 Agustus 2022 dengan Rp123,60K

Data Visualization dengan Tableau



Dapat kita lihat dari bagan disamping, produk Thai Tea merupakan produk yang paling diminati oleh customer. Hal ini dapat diketahui dengan total penjualan Thai Tea selama tahun 2022 sebanyak 2.853. Sementara produk Potaoi Chip, Oat dan Cashew merupakan produk yang kurang diminati oleh customer.

Data Visualization dengan Tableau



Dari grafik dapat dilihat, toko Lingga memiliki jumlah pendapatan tertinggi selama tahun 2022, yaitu sebesar Rp25,3M diikuti oleh toko Sinar Harapan sebesar Rp21,9M. Sementara itu, toko-toko lain menghasilkan pendapatan dibawah rata-rata total pada tahun 2022.

KALBE Nutritionals

Rakamin Academy

Forecast Regression dan Segmentasi Pelanggan

Load Datasets

```
df_customers = pd.read_csv("Case Study Data Scientist\Case Study - Customer.csv", sep=';')
df_products = pd.read_csv("Case Study Data Scientist\Case Study - Product.csv", sep=';')
df_stores = pd.read_csv("Case Study Data Scientist\Case Study - Store.csv", sep=';')
df_transactions = pd.read_csv("Case Study Data Scientist/Case Study - Transaction.csv", sep=';')
```

```
df_customers.head()
```

	CustomerID	Age	Gender	Marital Status	Income
0	1	55	1	Married	5,12
1	2	60	1	Married	6,23
2	3	32	1	Married	9,17
3	4	31	1	Married	4,87
4	5	58	1	Married	3,57

```
df_stores.head()
```

	StoreID	StoreName	GroupStore	Type	Latitude	Longitude
0	1	Prima Tendean	Prima	Modern Trade	-6,2	106,816666
1	2	Prima Kelapa Dua	Prima	Modern Trade	-6,914864	107,608238
2	3	Prima Kota	Prima	Modern Trade	-7,797068	110,370529
3	4	Gita Ginara	Gita	General Trade	-6,966667	110,416664
4	5	Bonafid	Gita	General Trade	-7,250445	112,768845

```
df_products.head()
```

	ProductID	Product Name	Price
0	P1	Choco Bar	8800
1	P2	Ginger Candy	3200
2	P3	Crackers	7500
3	P4	Potato Chip	12000
4	P5	Thai Tea	4200

```
df_transactions.head()
```

	TransactionID	CustomerID	Date	ProductID	Price	Qty	TotalAmount	StoreID
0	TR11369	328	01/01/2022	P3	7500	4	30000	12
1	TR16356	165	01/01/2022	P9	10000	7	70000	1
2	TR1984	183	01/01/2022	P1	8800	4	35200	4
3	TR35256	160	01/01/2022	P1	8800	7	61600	4
4	TR41231	386	01/01/2022	P9	10000	1	10000	4

Cleaning Datasets

Proses pembersihan data dilakukan dengan adanya temuan-temuan berikut:

- Pada dataset ‘Customer’, ditemukan nilai null pada kolom ‘Marital Status'. Di dataset yang sama, kolom ‘Income' seharusnya berjenis data float karena merujuk pada pendapatan setiap pelanggan.
- Di dataset ‘Store’, kolom 'Latitude' dan 'Longitude' seharusnya memiliki tipe data float, bukan objek.
- Sementara itu, di dataset ‘Transactions’, kolom ‘Date' sebaiknya berjenis data datetime karena mencatat waktu transaksi. Selain itu, terdapat baris duplikat dalam dataset ini, yang dapat disimpulkan dari duplikasi pada kolom ‘TransactionID’. Seharusnya kolom ini bersifat unik pada setiap record/baris-nya.

Cleaning Datasets

KALBE Nutritionals Rakamin Academy

Customers Datasets

```
# Ubah tipe data pada kolom Income
df_customers['Income'] = df_customers['Income'].replace('[,]', '.', regex=True).astype('float')
```

```
# Isi data Marital Status yang kosong menjadi 'Unknown'
df_customers['Marital Status'] = df_customers['Marital Status'].fillna('Unknown')
```

Stores Datasets

```
# ubah tipe data kolom latitude dan longitude
df_stores['Latitude'] = df_stores['Latitude'].replace('[,]', '.', regex=True).astype('float')
df_stores['Longitude'] = df_stores['Longitude'].replace('[,]', '.', regex=True).astype('float')
```

Transactions Datasets

```
df_transactions['Date'] = pd.to_datetime(df_transactions['Date'], format='%d/%m/%Y')
```

```
df_transactions['TransactionID'].value_counts()
```

...

```
df_transactions[df_transactions['TransactionID'] == 'TR71313']
```

...

```
df_transactions = df_transactions.drop_duplicates(subset='TransactionID', keep='last')
```

Merge Datasets

```
# Buat merge df_transactions dan df_customers
df_merge = pd.merge(df_transactions, df_customers, how='inner', on='CustomerID')

# merge hasil merge sebelumnya/df_merge dengan df_products
df_merge = pd.merge(df_merge, df_products[['ProductID', 'Product Name']], how='inner', on='ProductID')

# merge hasil merge sebelumnya/df_merge dengan df_stores
df_merge = pd.merge(df_merge, df_stores, how='inner', on='StoreID')
```

```
df_all = df_merge.copy()
```

KALBE Nutritionals

Rakamin Academy

Forecast Qty dengan ARIMA

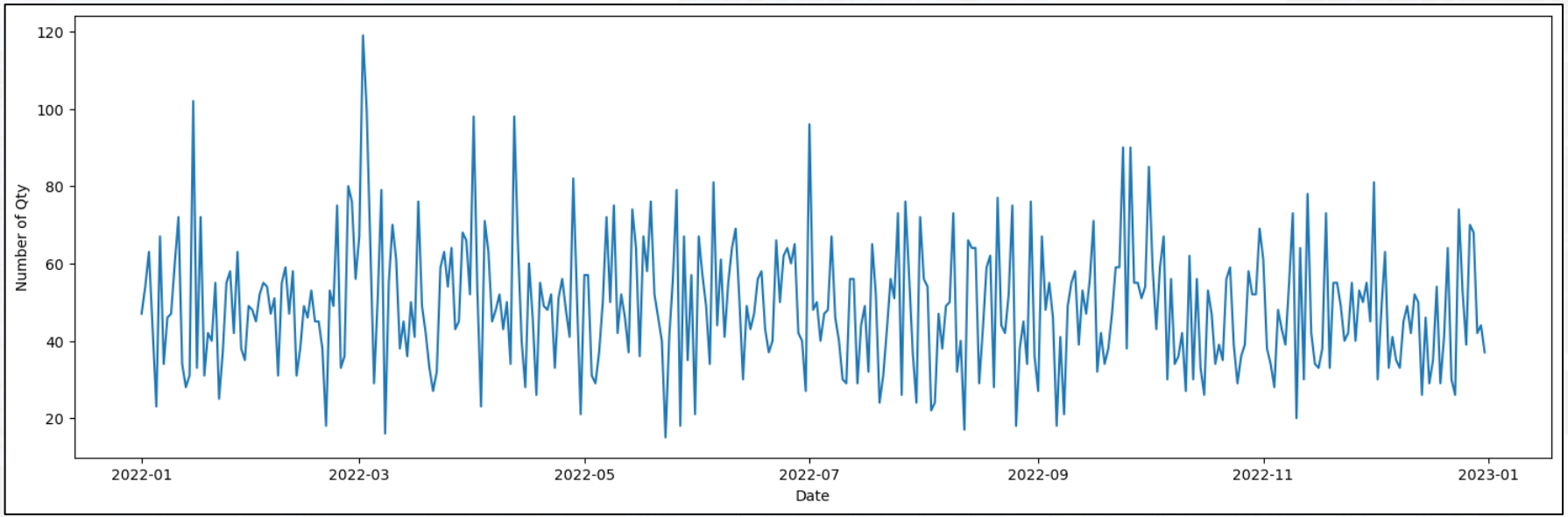
Pemilihan Data

```
1 df_tsa = df_all[['Date', 'Qty']]
```

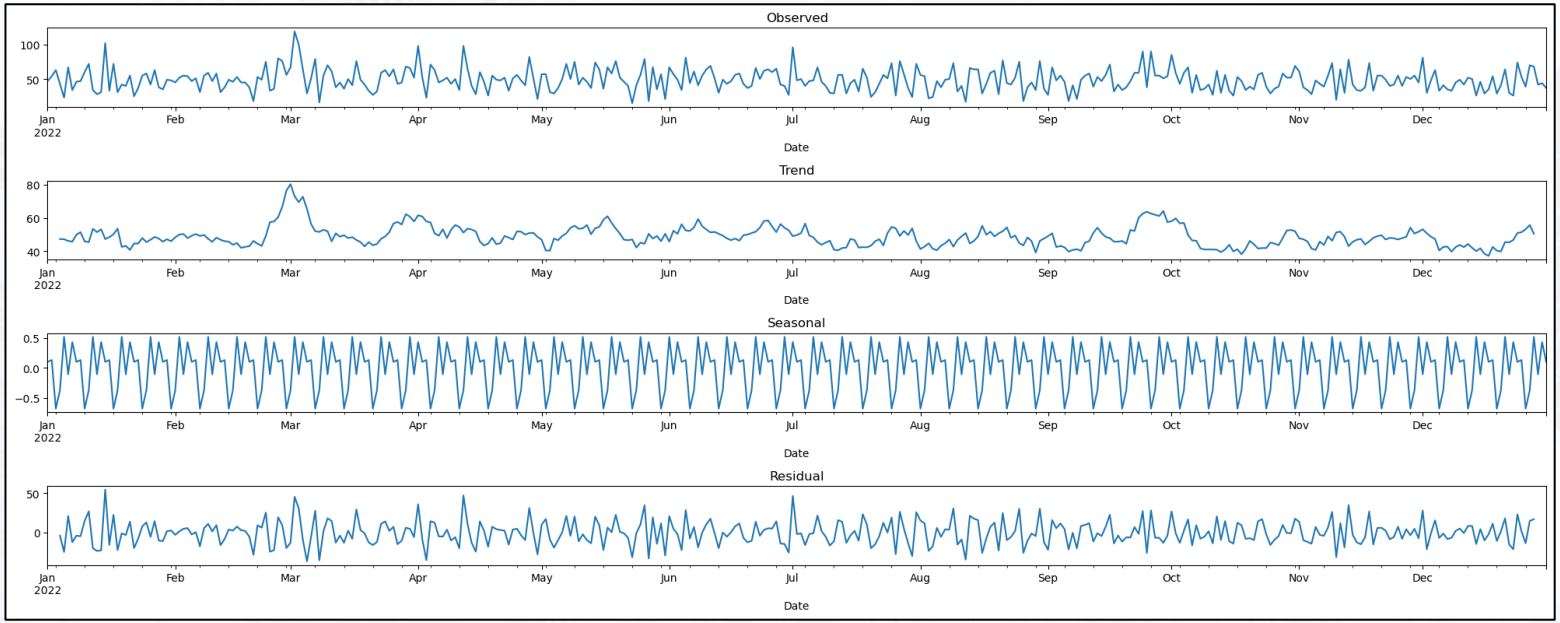
```
1 df_tsa = df_tsa.groupby('Date').sum()
```

Group by dilakukan untuk mendapatkan jumlah penjualan perhari

Date	Qty
2022-01-01	47
2022-01-02	54
2022-01-03	63
2022-01-04	43
2022-01-05	23



Analisis Trend dan Season



Uji Stationer

```
adf_result = adfuller(df_tsa)
```

```
print('ADF Statistik: {:.4f}'.format(adf_result[0]))
print('p-value: {:.4f}'.format(adf_result[1]))
print('Critical Values:')
for key, value in adf_result[4].items():
    print('\t{}: {:.3f}'.format(key, value))
if adf_result[1] > 0.05:
    print("Terima H0: Data Non-Stationary")
    print("Cari d optimal")
else:
    print("Tolak H0: Data Stationary")
    print("d = 0")
```

```
ADF Statistik: -19.4260
p-value: 0.0000
Critical Values:
        1%: -3.448
        5%: -2.870
        10%: -2.571
Tolak H0: Data Stationary
```

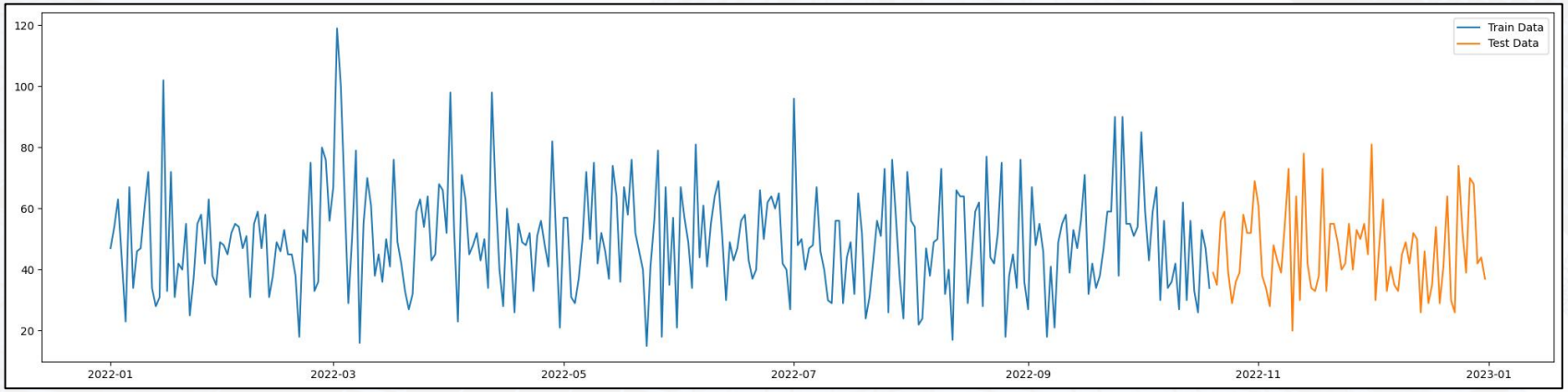
H0 : Data Tidak Stasioner
H1 : Data Stasioner
Jika nilai p-value < 0.05 maka H0 ditolak.

Berdasarkan tes Augmented Dicky-Fuller, karena nilai p-value = 0 maka H0 ditolak dan data stasioner.

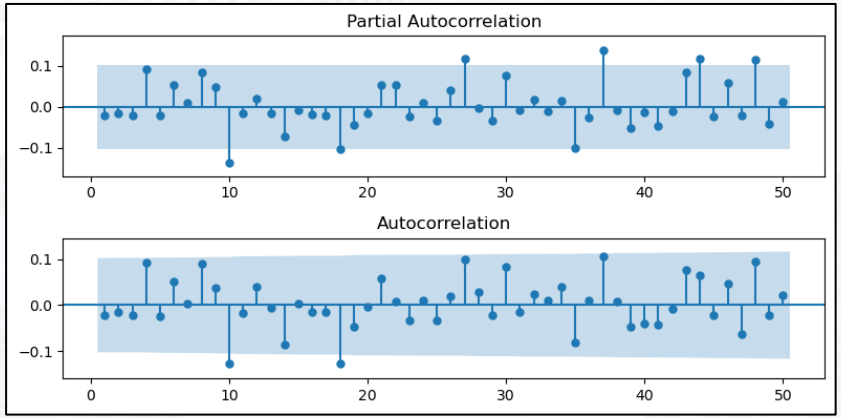
Uji Stationer

```
# split data
split = int(np.round(df_tsa.shape[0] * 0.8))
df_tsa_train, df_tsa_test = df_tsa.iloc[:split], df_tsa.iloc[split:]
```

Split data dengan 80% data latih dan 20% data tes



Model ARIMA




```
y = df_tsa_train['Qty']
ARIMA_model = ARIMA(y, order=(10, 0, 10))
ARIMA_model = ARIMA_model.fit()
```

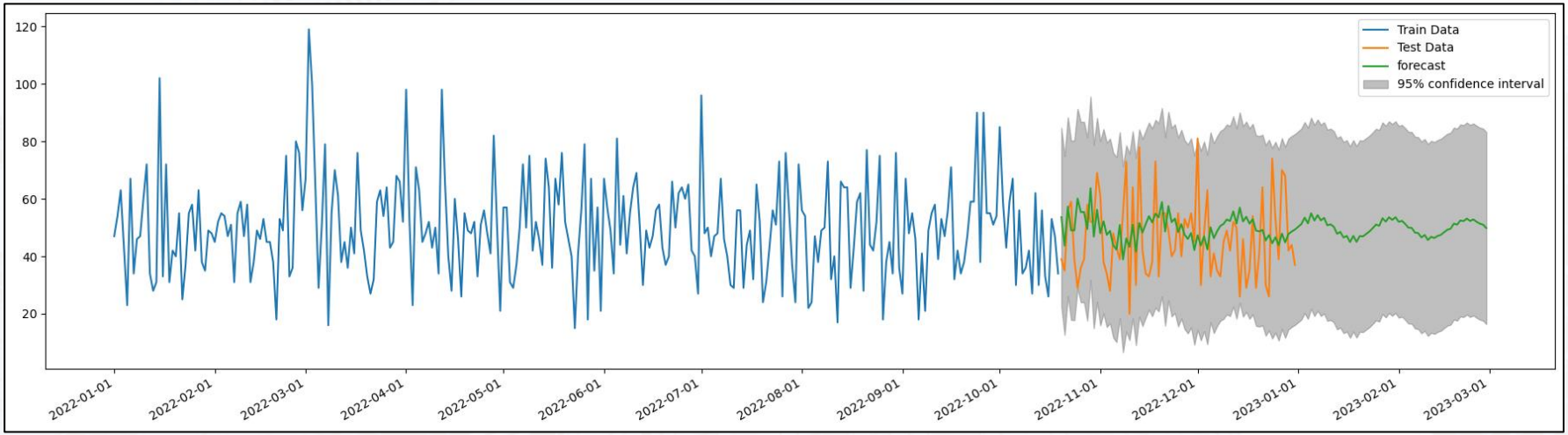
```
ARIMA_model.summary()
```

SARIMAX Results

Dep. Variable:	Qty	**No. Observations:**	292
Model:	ARIMA(10, 0, 10)	**Log Likelihood**	-1221.424
Date:	Sun, 01 Oct 2023	**AIC**	2486.849
Time:	10:02:51	**BIC**	2567.737
Sample:	01-01-2022	**HQIC**	2519.249
	- 10-19-2022		
Covariance Type:	opg		

Nilai p dan q didaptkan dengan Partial Auto Correlation Function (PACF) dan Auto Correlation Function (ACF). Dapat kita lihat visualisasi PACF dan ACF disamping, jumlah lag yang keluar dari limit adalah lag ke-10. Maka nilai p dan q yang digunakan adalah 10. Sementara nilai d yang digunakan adalah 0 karena data merupakan data stasioner.

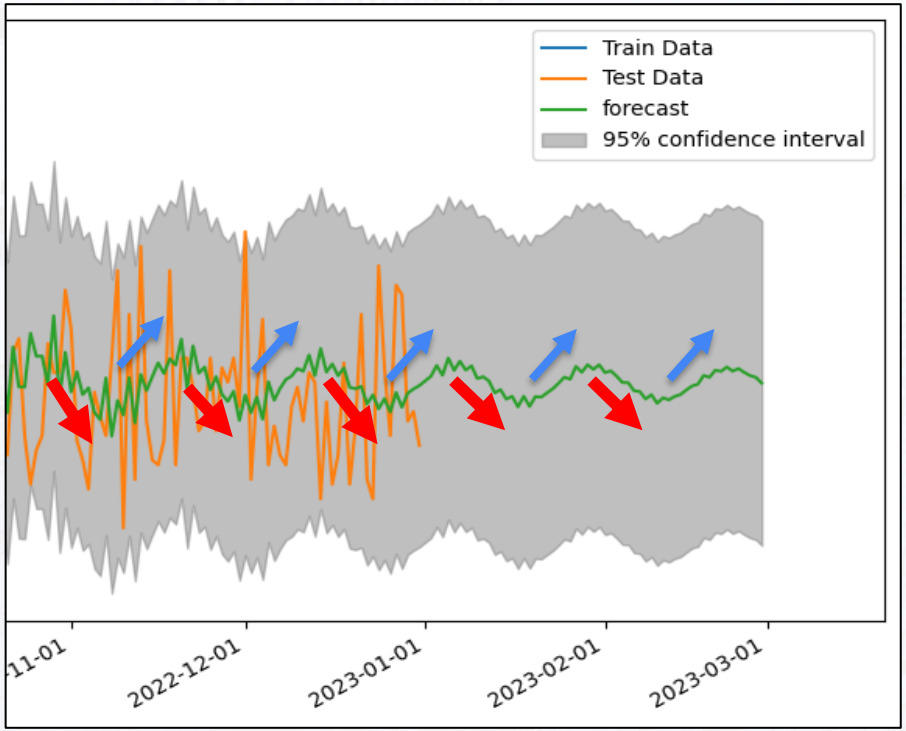
ARIMA Forecast




```
1 evaluation(df_tsa_test['Qty'], y_pred_ARIMA_df['predictions'])
```

```
Mean Absolute Error (MAE): 12.82
Root Mean Squared Error (RMSE): 15.12
Mean Absolute Percentage Error (MAPE): 31.78%
```

ARIMA Forecast



Dari hasil forecast dan prediksi, ditemukan pola penjualan akan menurun di minggu awal tiap bulan kemudian akan kembali naik pada minggu terakhir tiap bulan

KALBE Nutritionals

Rakamin Academy

Customer Segmentation/Segmentasi Pelanggan

Pemilihan Data

```python
df_cust_seg = df_all.groupby('CustomerID').agg(
    Total_Transactions = ('TransactionID', 'count'),
    Total_Qty = ('Qty', 'sum'),
    Total_Amount = ('TotalAmount', 'sum')
)
```

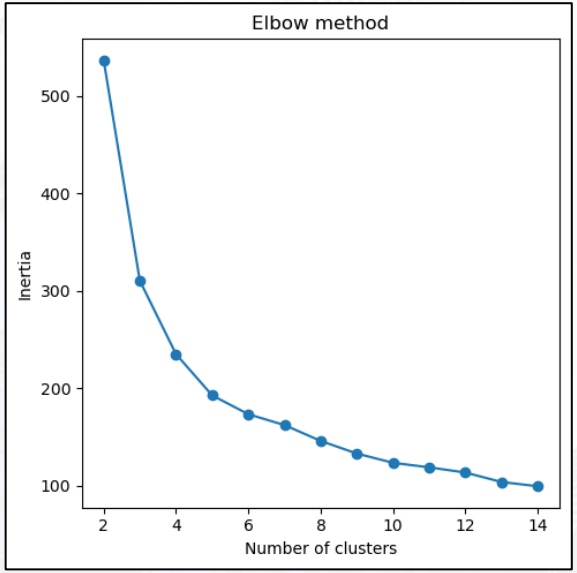
```python
df_cust_seg
```

CustomerID	Total_Transactions	Total_Qty	Total_Amount
1	17	60	623300
2	12	56	382300
3	15	56	446200
4	10	46	302500
5	7	27	268600
...	...	...	...
443	16	59	485100
444	18	62	577700
445	17	62	530800
446	11	42	423300
447	12	37	401800

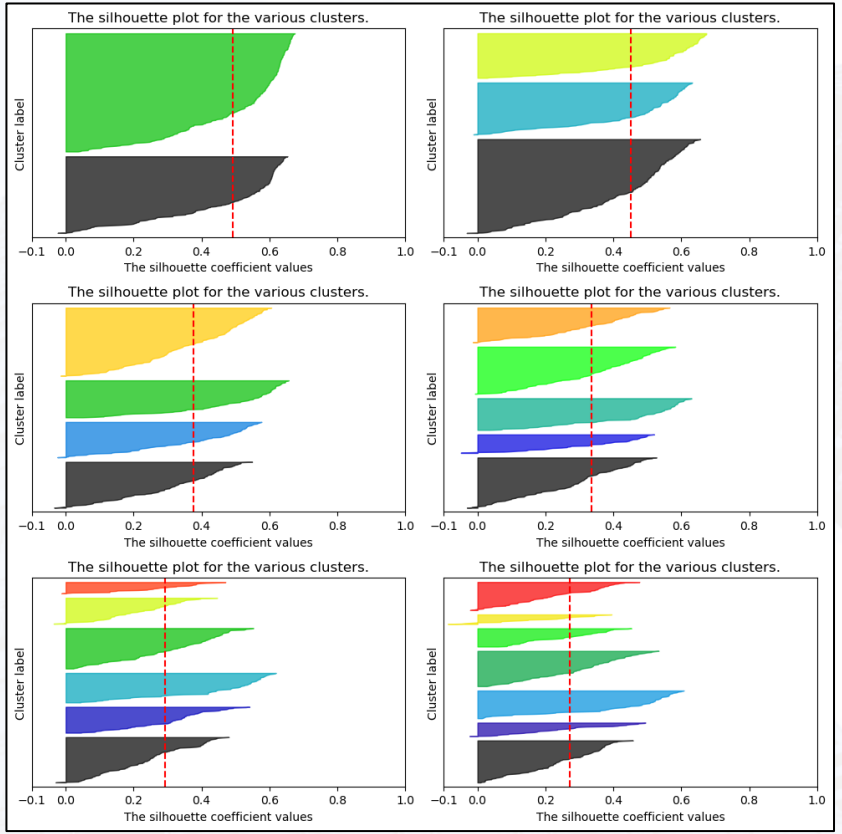
447 rows × 3 columns

Segmentasi pelanggan akan dilakukan berdasarkan jumlah transaksi, total pembelian dan total pembayaran yang dilakukan masing-masiong pelanggan selama tahun 2022

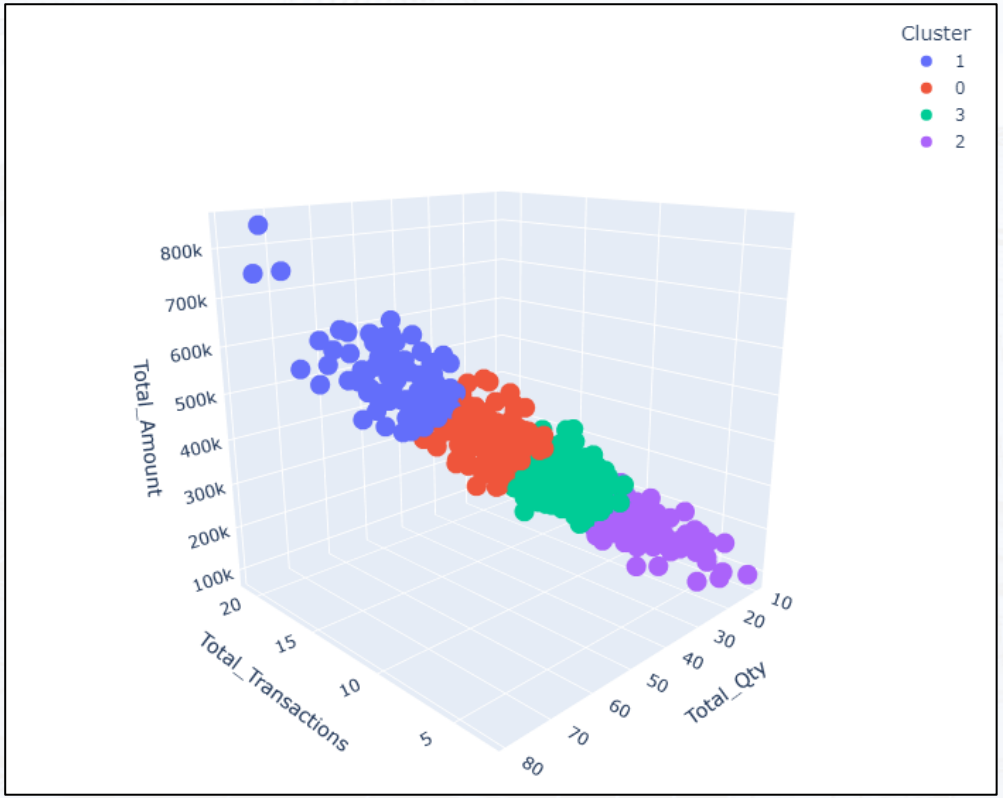
Pemilihan Jumlah Klaster



Untuk pemilihan jumlah klaster menggunakan metode elbow dan nilai skor silhoutte. Dari kedua metode tersebut, maka diambil jumlah klaster sebanyak 4.



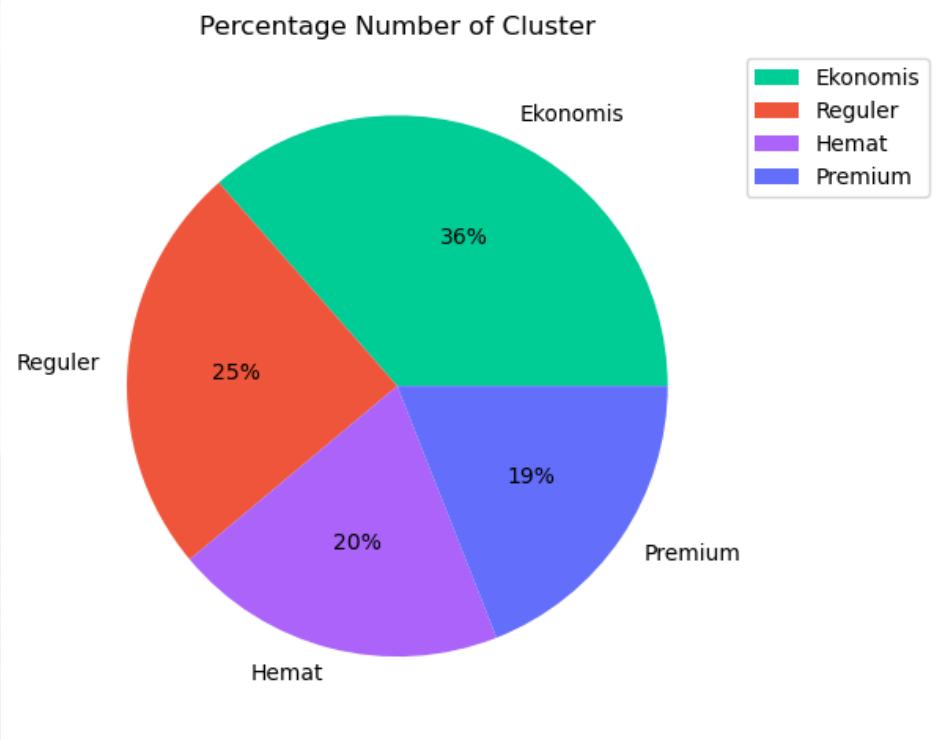
Hasil Klasterisasi



Dari hasil pengklasteran didapat hasil:

1. Klaster 0 / Reguler : Merupakan Pelanggan dengan jumlah transaksi, jumlah pembelian produk dan pengeluaran uang yang sedang
2. Klaster 1 / Premium : Merupakan pelanggan dengan tingkat transaksi, pembelian produk dan jumlah pengeluaran uang yang tinggi.
3. Klaster 2 / Hemat : Merupakan pelanggan dengan transaksi, pembelian dan pengeluaran uang yang rendah.
4. Klaster 3 / Ekonomis : Merupakan pelanggan dengan transaksi dan pembelian yang tinggi tetapi tingkat pengeluaran uang yang rendah.

Persentase jumlah pelanggan



Strategi bisnis yang ditawarkan

KALBE Nutritionals Rakamin Academy

Pelanggan Premium

- Produk Eksklusif: berikan penawaran produk eksklusif yang hanya tersedia untuk pelanggan premium, menciptakan rasa eksklusivitas dan keinginan untuk memiliki.
- Pengalaman Belanja Premium: Tingkatkan layanan pelanggan, termasuk pengiriman cepat, pelayanan pelanggan pribadi, dan paket hadiah eksklusif.

Pelanggan Reguler

- Penawaran Spesial: Berikan penawaran eksklusif atau produk terbatas kepada pelanggan reguler sebagai tanda terima kasih atas dukungan mereka.
- Program Loyalitas: Buat program loyalitas yang memberikan poin atau diskon khusus untuk pelanggan yang sering berbelanja produk

Strategi bisnis yang ditawarkan

Pelanggan Ekonomis

- Promosi Bundel: Tawarkan paket produk yang dipasangkan bersama dengan harga yang lebih murah daripada membeli produk secara terpisah.
- Diskon Khusus untuk Pembelian Besar: Berikan diskon atau hadiah untuk pembelian dalam jumlah besar untuk menarik pelanggan ekonomis yang cenderung berbelanja dalam jumlah besar.

Pelanggan Hemat

- Paket Hemat: Tawarkan paket produk dalam jumlah besar dengan harga diskon yang menarik untuk menarik pembeli yang mencari nilai terbaik.
- Promosi Diskon Rutin: Selenggarakan promosi diskon rutin pada produk-produk yang sering dibeli oleh pelanggan hemat.

Link Source Code

- Github
- Tableau

Thank You