

Machine Learning Project

Kalbe Nutritional Data Scientist
Project Based Internship Program

Presented by
Nabila

Experiences

- Experience 1
Membuat aplikasi berbasis Website dan Mobile
- Experience 2
Desain UI/UX dan mahir menggunakan Photoshop
- Experience 3
Finalis KMIPN (Kompetisi Mahasiswa Informatika Politeknik Nasional) 2023 di Surabaya



Nabila

About You

Saya adalah seorang mahasiswi D3 di Politeknik Negeri Pontianak yang mengambil Jurusan Teknik Elektro dengan Program Studi Teknik Informatika.

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 Nutritionals:

- 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

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

```
SELECT
  CASE
    WHEN "Marital Status" = '' THEN 'Tidak Diketahui'
    ELSE "Marital Status"
  END AS MaritalStatus,
  AVG(Age) AS rata_rata_umur
FROM
  Customer
GROUP BY
  MaritalStatus;
```

	asc maritalstatus	123 rata_rata_umur
1	Tidak Diketahui	31.3333333333
2	Married	43.0382352941
3	Single	29.3846153846

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

```
SELECT
  CASE
    WHEN Gender = 0 THEN 'Wanita'
    WHEN Gender = 1 THEN 'Pria'
    ELSE 'Tidak Diketahui'
  END AS JenisKelamin,
  AVG(Age) AS rata_rata_umur
FROM
  Customer
GROUP BY
  Gender;
```

	asc jeniskelamin	123 rata_rata_umur
1	Wanita	40.326446281
2	Pria	39.1414634146

Exploratory Data Analysis

- query 3 : Tentukan nama store dengan total quantity terbanyak!

```
SELECT StoreName, SUM(Qty) AS TotalQuantity
FROM Transaction
JOIN Store ON Transaction.StoreID = Store.StoreID
GROUP BY StoreName
ORDER BY TotalQuantity desc
LIMIT 1;
```

	asc storename	123 totalquantity
1	Lingga	2,777

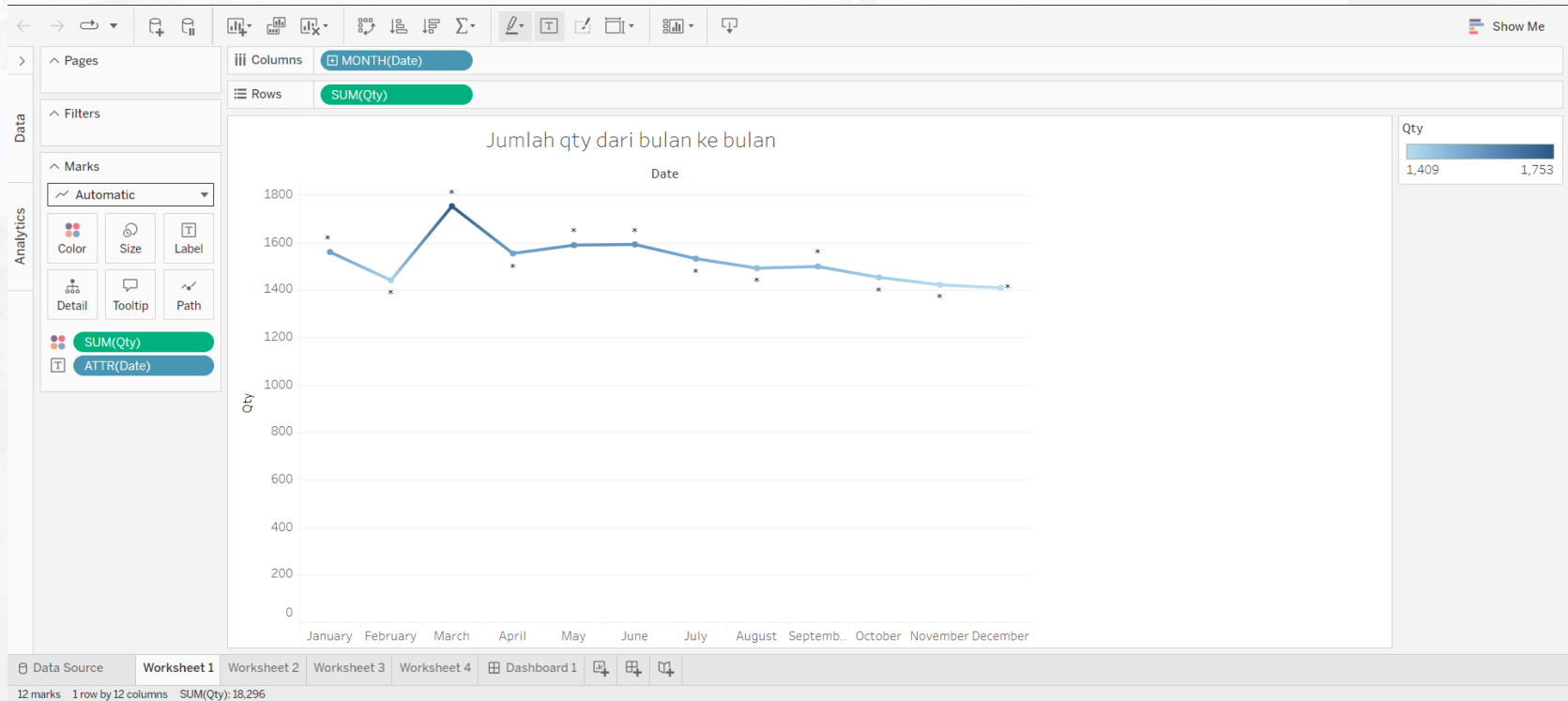
- query 4 : Tentukan nama produk terlaris dengan total amount terbanyak!

```
SELECT "Product Name" , SUM(TotalAmount) AS TotalSalesAmount
FROM Transaction
JOIN Product ON Transaction.ProductID = Product.ProductID
GROUP BY "Product Name"
ORDER BY TotalSalesAmount desc
LIMIT 1;
```

	asc Product Name	123 totalsalesamount
1	Cheese Stick	27,615,000

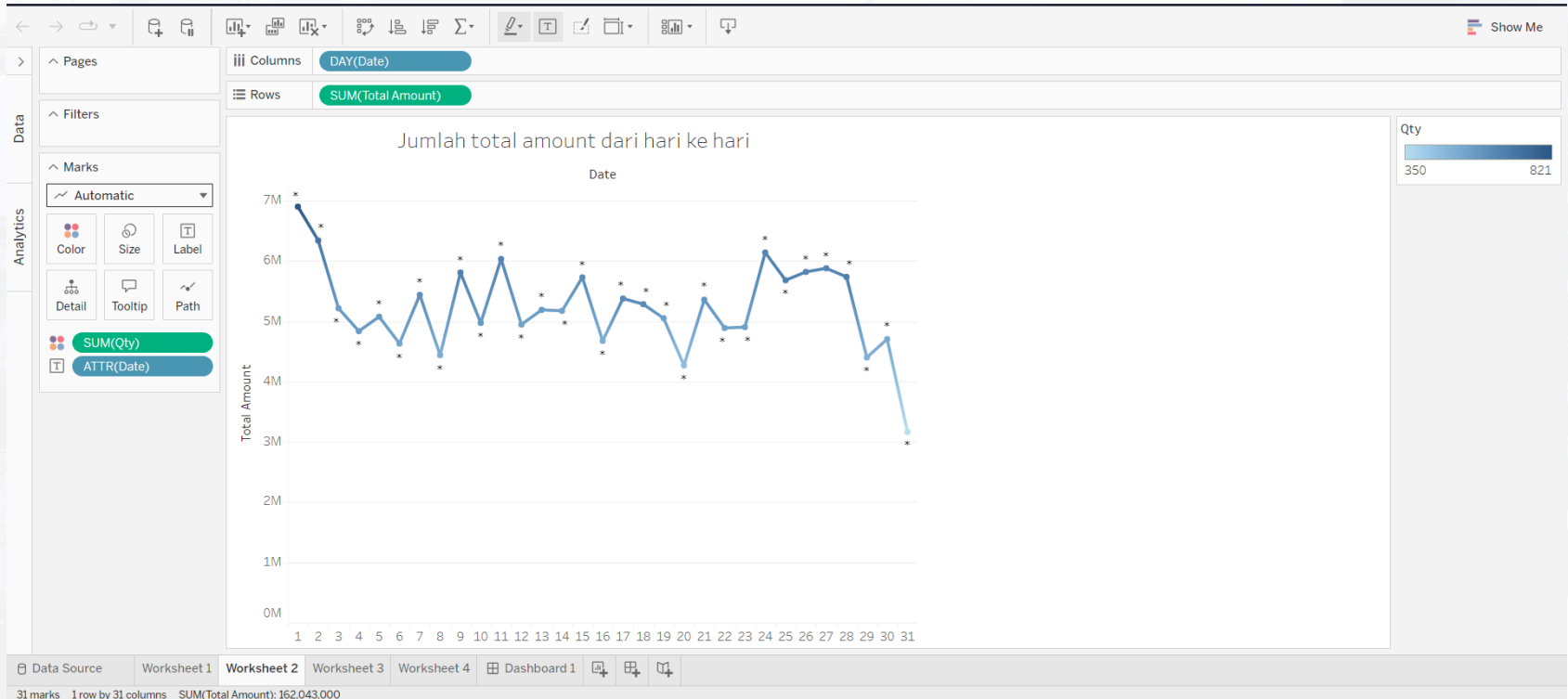
Worksheet

- Worksheet 1 Jumlah qty dari bulan ke bulan



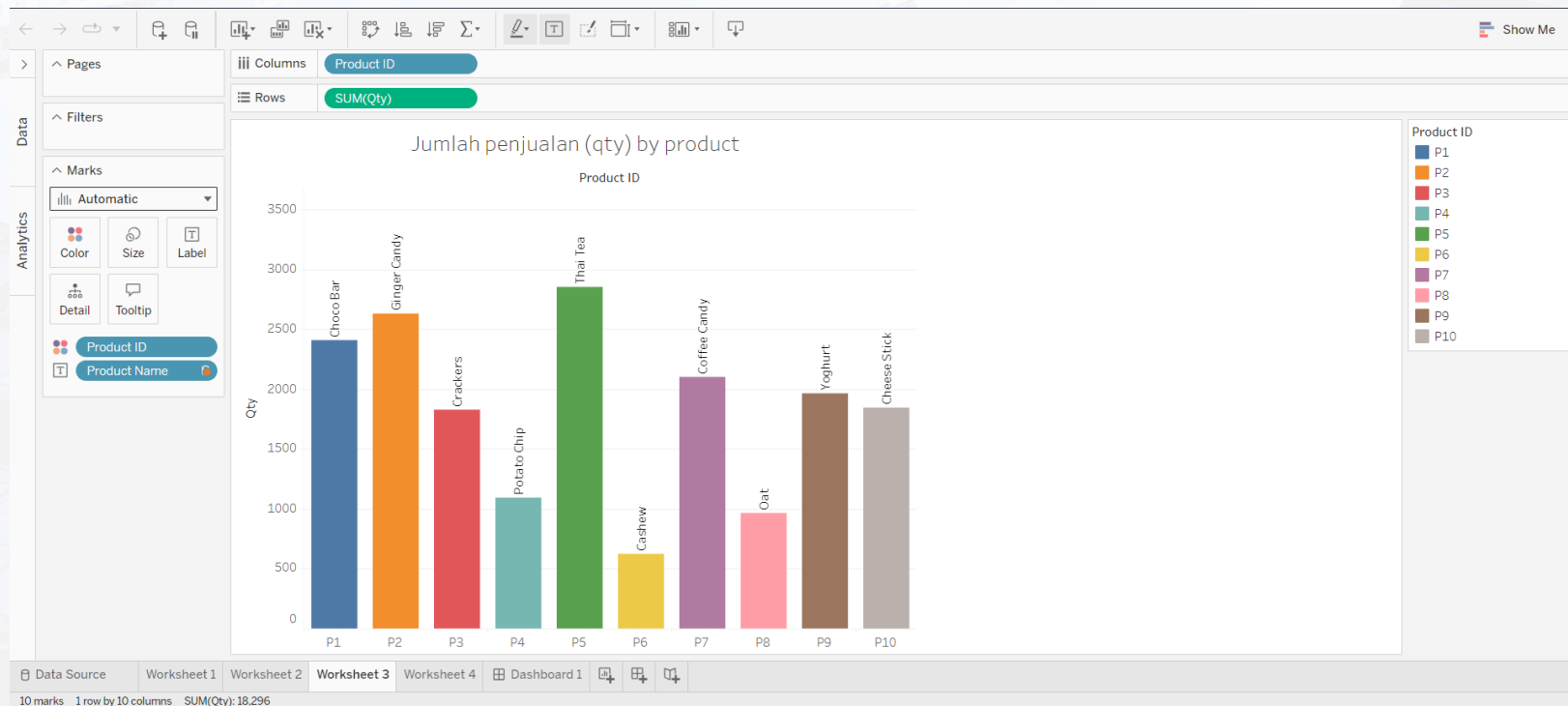
Worksheet

- Worksheet 2 Jumlah total amount dari hari ke hari



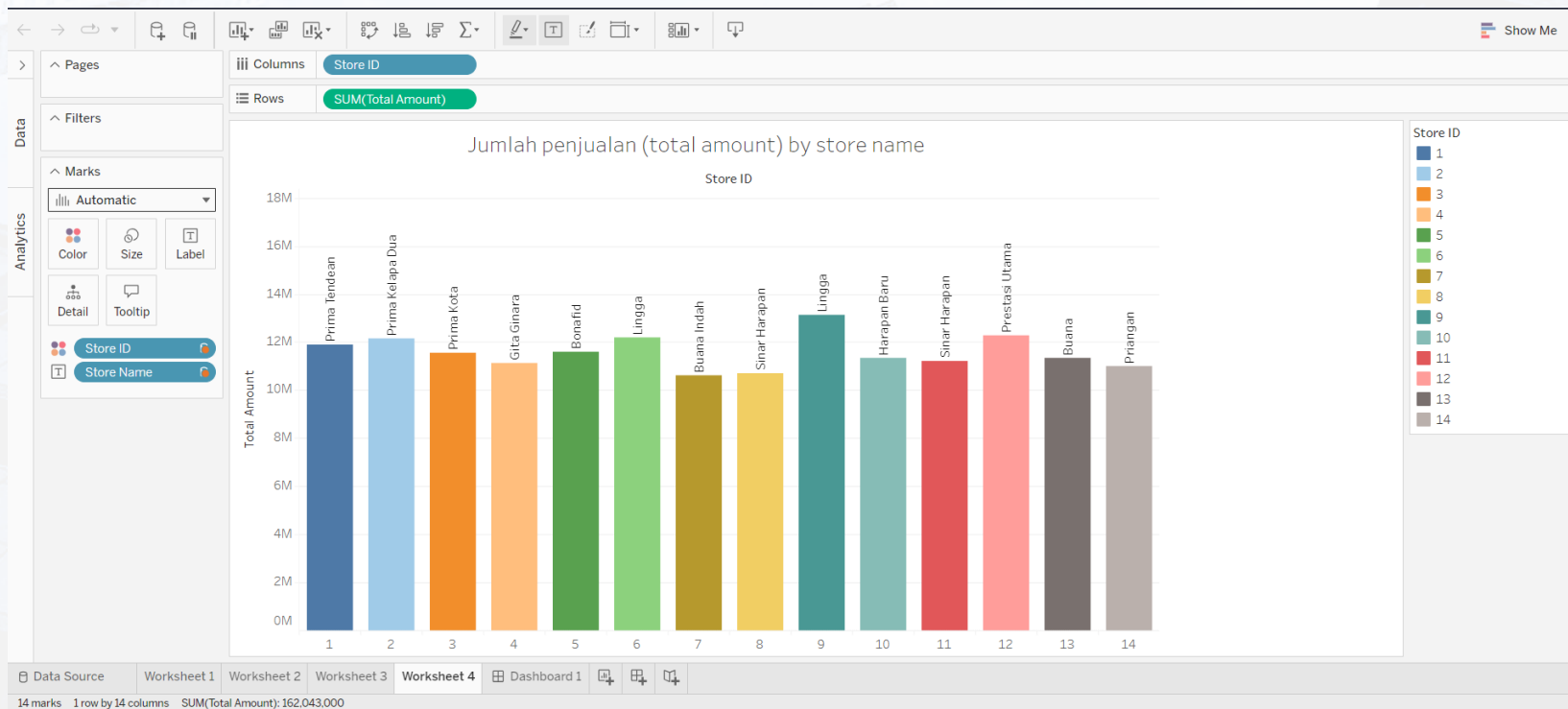
Worksheet

- Worksheet 3 Jumlah penjualan (qty) by product



Worksheet

- Worksheet 4 Jumlah penjualan (total amount) by store name

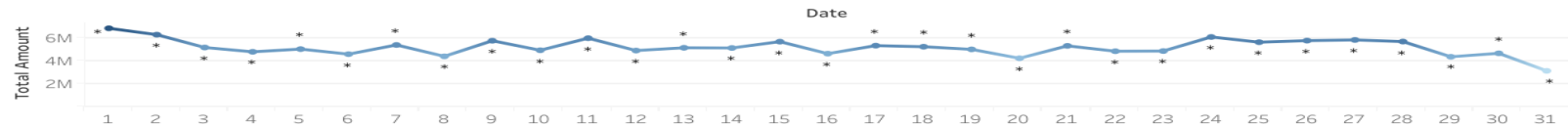


Dashboard

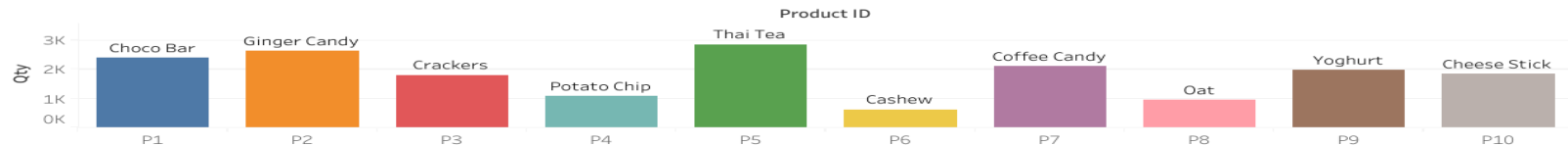
Jumlah qty dari bulan ke bulan



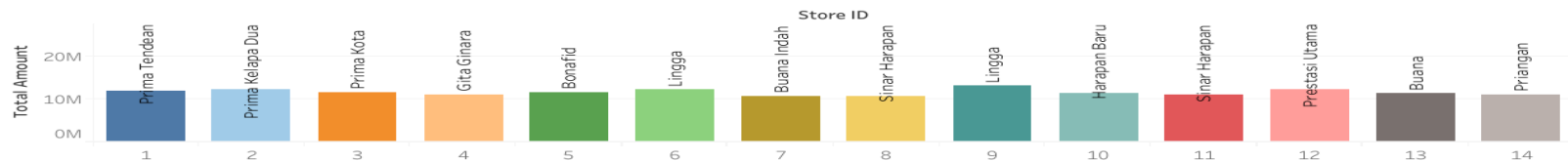
Jumlah total amount dari hari ke hari



Jumlah penjualan (qty) by product

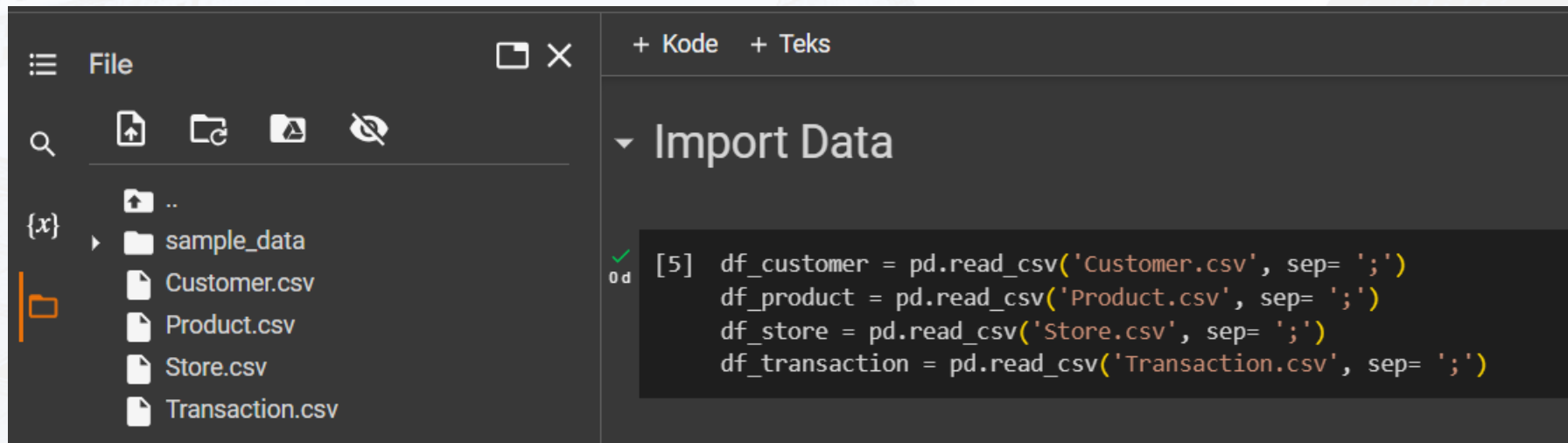


Jumlah penjualan (total amount) by store name



Machine Learning

- Membaca data CSV



The screenshot displays a Jupyter Notebook interface. On the left, the 'File' sidebar shows a file explorer with a folder named 'sample_data' containing five CSV files: 'Customer.csv', 'Product.csv', 'Store.csv', and 'Transaction.csv'. The main area on the right is titled '+ Kode + Teks' and features a section 'Import Data'. Below this, a code cell (labeled '[5]') contains the following Python code to read CSV files using pandas:

```
df_customer = pd.read_csv('Customer.csv', sep= ';')  
df_product = pd.read_csv('Product.csv', sep= ';')  
df_store = pd.read_csv('Store.csv', sep= ';')  
df_transaction = pd.read_csv('Transaction.csv', sep= ';')
```

The code cell shows a green checkmark and the output '0 d'.

Machine Learning

- Melakukan data cleansing

```
✓ [53] #drop missing values karena tidak signifikan jumlahnya  
0d df_merged.dropna(inplace = True)  
df_merged.isna().sum()
```

```
TransactionID    0  
CustomerID      0  
Date            0  
ProductID       0  
Price_x         0  
Qty            0  
TotalAmount     0  
StoreID         0  
Age            0  
Gender          0  
Marital Status  0  
Income         0  
StoreName       0  
GroupStore      0  
Type           0  
Latitude        0  
Longitude       0  
Product Name    0  
dtype: int64
```


Machine Learning

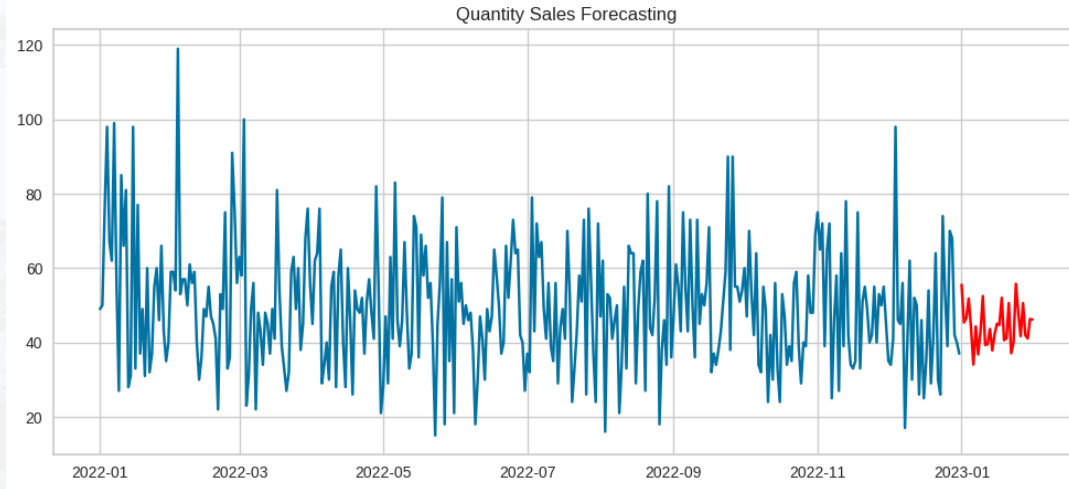
- Menggabungkan semua data menjadi 1 data

```
[10] df1 = pd.merge(df_transaction, df_customer, on='CustomerID', how='inner')
df2 = pd.merge(df1, df_store, on = 'StoreID', how = 'inner')
df_merged = pd.merge(df2, df_product, on = 'ProductID', how = 'inner')
df_merged.head()
```

	TransactionID	CustomerID	Date	ProductID	Price_x	Qty	TotalAmount	StoreID	Age	Gender	Marital Status	Income	StoreName	GroupStore	Type	Latitude	Longitude	Product Name	Price_
0	TR11369	328	01/01/2022	P3	7500	4	30000	12	36	0	Married	10,53	Prestasi Utama	Prestasi	General Trade	-2,990934	104,756554	Crackers	7500
1	TR89318	183	17/07/2022	P3	7500	1	7500	12	27	1	Single	0,18	Prestasi Utama	Prestasi	General Trade	-2,990934	104,756554	Crackers	7500
2	TR9106	123	26/09/2022	P3	7500	4	30000	12	34	0	Married	4,36	Prestasi Utama	Prestasi	General Trade	-2,990934	104,756554	Crackers	7500
3	TR4331	335	08/01/2022	P3	7500	3	22500	12	29	1	Single	4,74	Prestasi Utama	Prestasi	General Trade	-2,990934	104,756554	Crackers	7500
4	TR6445	181	10/01/2022	P3	7500	4	30000	12	33	1	Married	9,94	Prestasi Utama	Prestasi	General Trade	-2,990934	104,756554	Crackers	7500

Machine Learning

- Membuat model machine learning regression (time series)

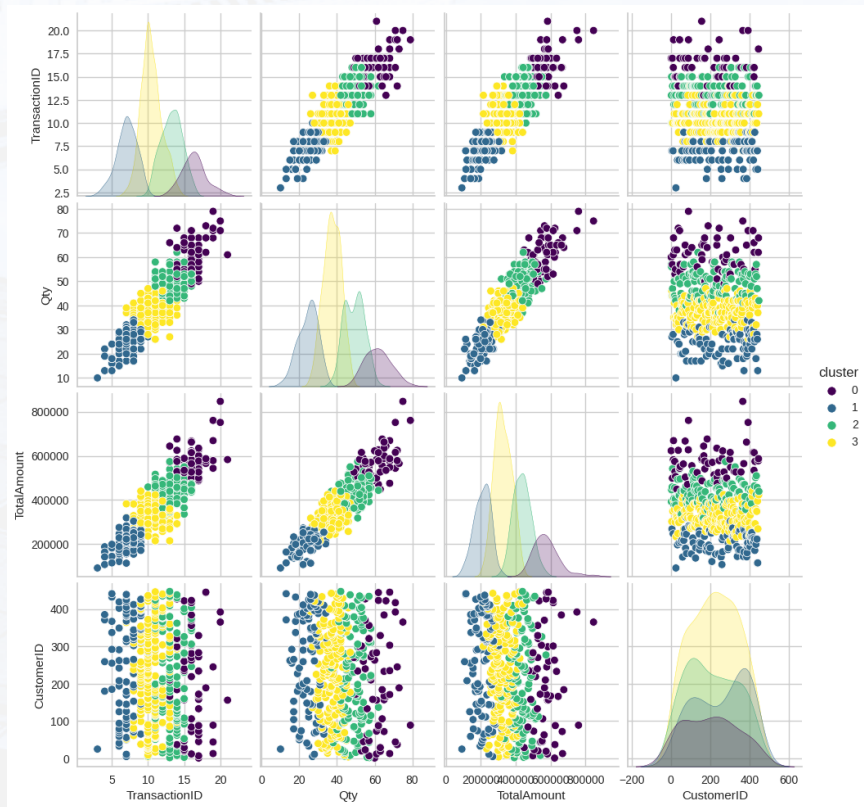


```
✓ [60] forecast.describe()
0 d
count    31.000000
mean     44.489775
std       5.541639
min      34.053507
25%      40.847933
50%      43.811827
75%      47.331878
max      55.767671
Name: predicted_mean, dtype: float64
```

- Dari prakiraan tersebut, rata-rata kuantitas penjualan pada bulan Januari 2023 adalah 44.489775 atau dibulatkan menjadi sekitar 44 pcs/hari.

Machine Learning

- Membuat model machine learning clustering



	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

Kesimpulan

- **Cluster 3**

Cluster dengan jumlah pelanggan paling banyak.

Karakteristik pelanggan menempati posisi ketiga dari setiap metrik (transaction, quantity, total amount).

Rekomendasi:

- Membangun hubungan baik dengan pelanggan.
- Memberikan survey untuk mengembangkan minat pelanggan terbanyak.

- **Cluster 2**

Karakteristik pelanggan yang menempati posisi ke dua tertinggi pada setiap metrik.

Rekomendasi:

- Memberikan promo secara rutin untuk meningkatkan transaksi.
- Melakukan upselling produk-produk dengan harga tinggi.

Kesimpulan

- **Cluster 1**

Karakteristik pelanggan dengan nilai terendah pada setiap metriknya.

Rekomendasi:

- Memberikan discount price yang cukup besar untuk meningkatkan Transaksi pelanggan.
- Memberikan promo pada transaksi dengan Quantity lebih tinggi.
- Memberikan survey untuk mengetahui potensi pengembangan produk.

- **Cluster 0**

Cluster dengan jumlah pelanggan paling sedikit

Karakteristik pelanggan dengan nilai tertinggi pada setiap metriknya.

Rekomendasi:

- Memberikan promo loyalty untuk mempertahankan transaksi
- Memberikan survey kepuasan pelanggan.
- Melakukan upselling produk dengan harga lebih tinggi

Link

- Github: https://github.com/Nabila13n/FinalTask_Kalbe-Nutritionals-Data-Scientist

Video Presentation

- Google Drive:
https://drive.google.com/drive/folders/12eL5YBp_k-P72ojDXIJDXdADGsEx7fne?usp=sharing

Thank You

