Presentasi Final Task

Kalbe Nutritionals Data Scientist Project Based Internship Program

Michael Fernandito Sanfia

SQL Query

Query 1

```
select case
   when "Marital Status"='Married' then 'Menikah'
   when "Marital Status"='Single' then 'Belum Menikah'
   else 'Tidak diketahui'
End as "Status Pernikahan", avg(age) as "Rata Rata usia"
from customer
group by "Marital Status"
```

Hasil:

	Status Pernikahan	Rata Rata usia
1	Tidak diketahui	31.3333333333
2	Menikah	43.0382352941
3	Belum Menikah	29.3846153846

Query 2

```
select case
   WHEN gender = 0 THEN 'Wanita'
   WHEN gender = 1 THEN 'Pria'
end as "Jenis Kelamin" ,avg(age) as "Rata Rata usia"
from customer
group by gender
```

Hasil:

	Jenis Kelamin	Rata Rata usia
1	Wanita	40.326446281
2	Pria	39.1414634146

Query 3

```
select store.storename as "Nama Toko",
count(transaction.storeid ) as "Jumlah Quantity"
from store
inner join transaction
on store.storeid=transaction.storeid
group by "Nama Toko"
order by "Jumlah Quantity" desc
```

Hasil:

	Nama Toko	Jumlah Quantity
1	Lingga	738
2	Sinar Harapan	698
3	Buana	368
4	Prima Kota	367
5	Prima Kelapa Dua	364
6	Prestasi Utama	363
7	Bonafid	362
8	Harapan Baru	355
9	Buana Indah	355
10	Prima Tendean	354
11	Gita Ginara	350
12	Priangan	346

Query 4

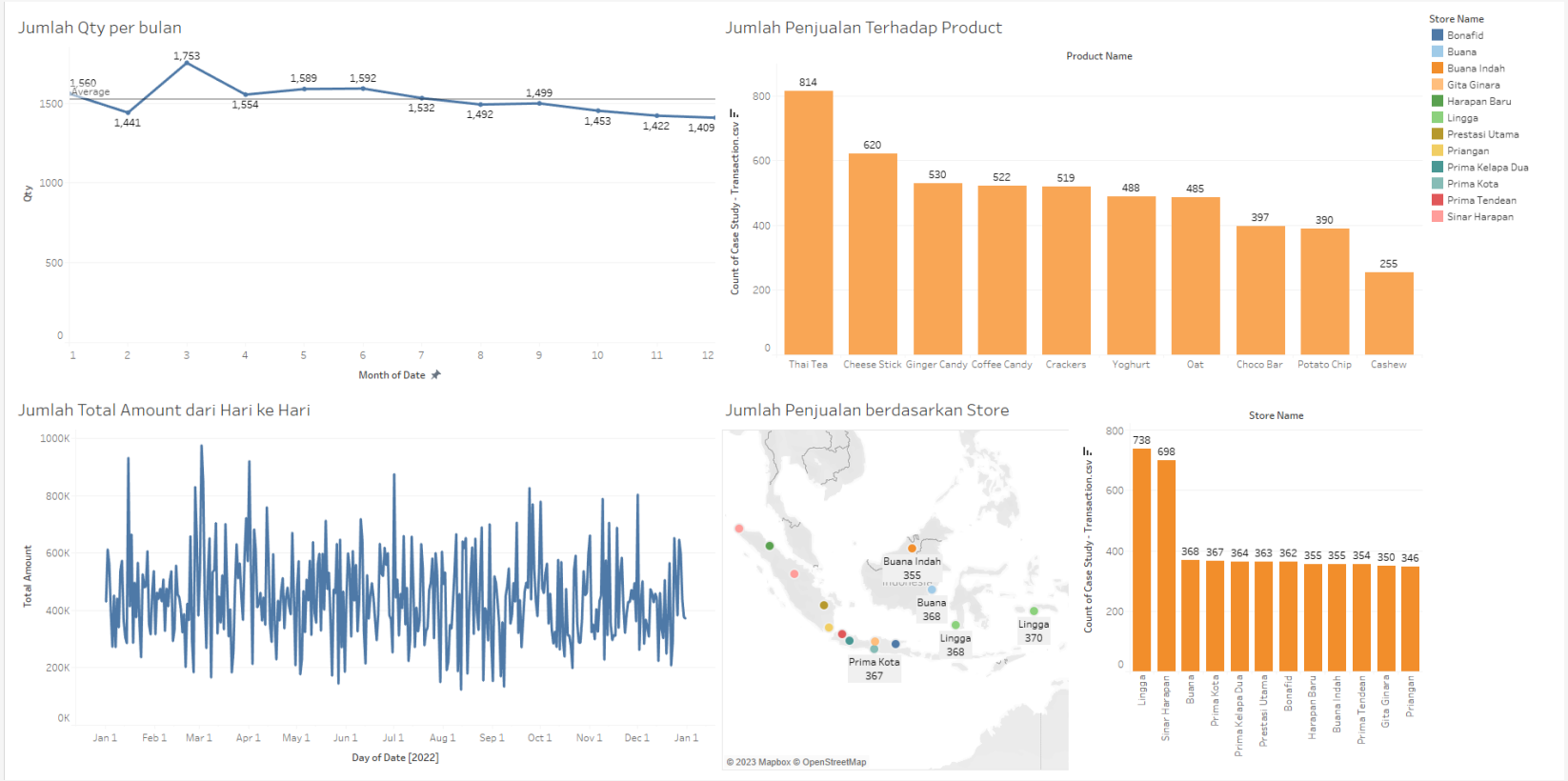
```
select product."Product Name" as "Nama Produk" ,
count(transaction.storeid ) as "Total Amount"
from product
inner join transaction
on product.productid=transaction.productid
group by "Nama Produk"
order by "Total Amount" desc
```

Hasil:

	Nama Produk	Total Amount
1	Thai Tea	814
2	Cheese Stick	620
3	Ginger Candy	530
4	Coffee Candy	522
5	Crackers	519
6	Yoghurt	488
7	Oat	485
8	Choco Bar	397
9	Potato Chip	390
10	Cashew	255

Tableau Dasboard

Tableau Dasboard

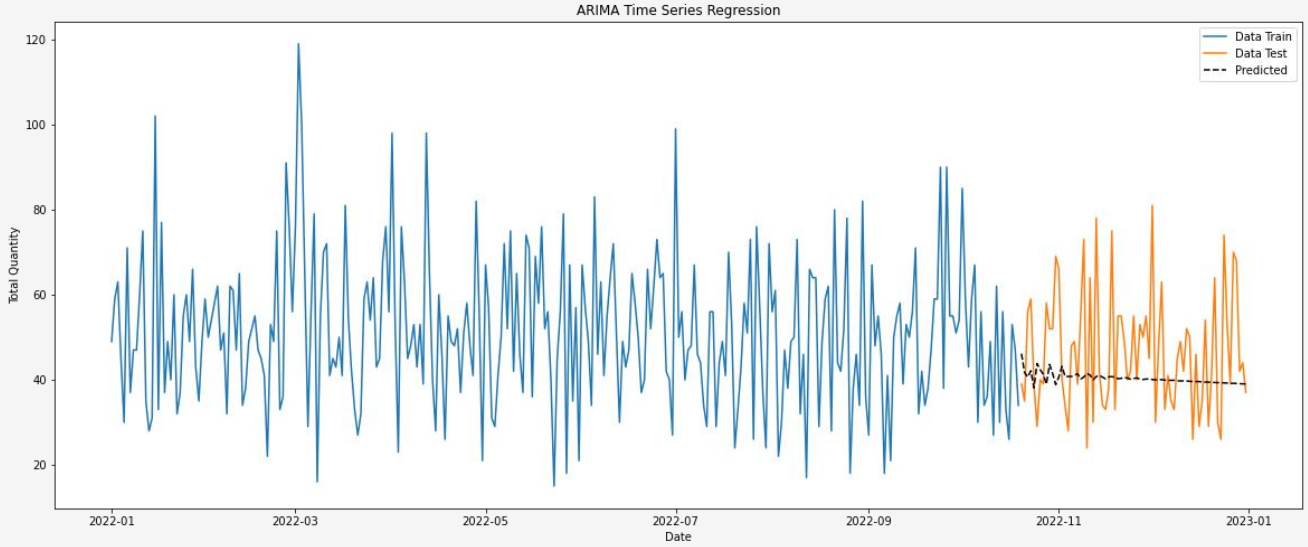


Regresion Model

Data Processing

- Melakukan Import data dari file csv
- Melakukan Pengecekan nilai null dan penghapusan
- Menyesuaikan format tipe data yang sesuai
- Melakukan Feature Selection
- Menggabungkan keseluruhan menjadi 1 dataframe

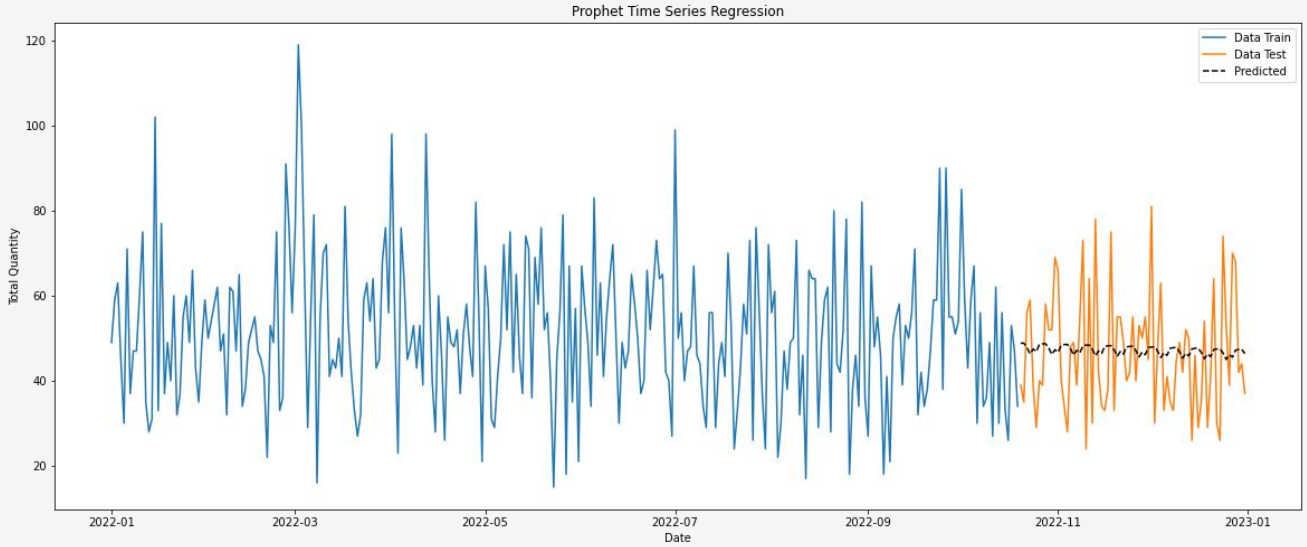
ARIMA Model Result



Hasil Performansi

- MAE: 11.78
- RMSE: 15.28
- MAPE: 23.96%

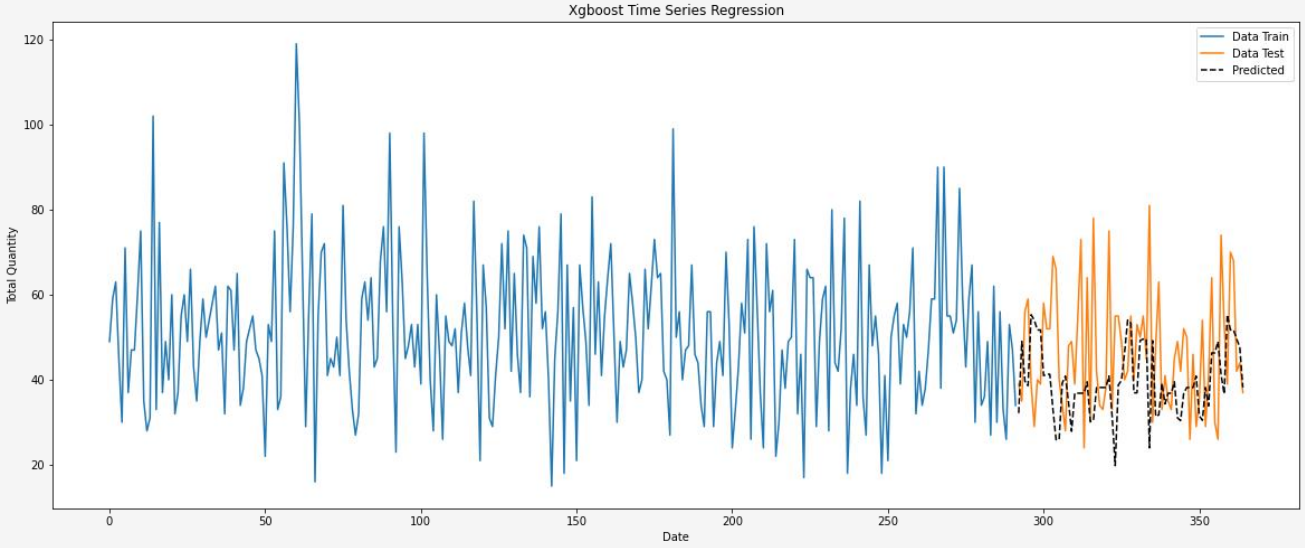
Prophet Model Result



Hasil Performansi

- MAE: 11.79
- RMSE: 14.06
- MAPE: 27.85%

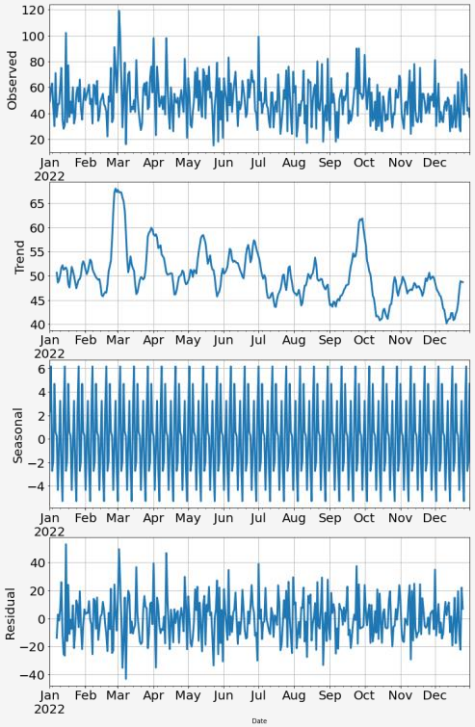
Xgboost Model Result



Hasil Performansi

- MAE: 14.44
- RMSE: 18.43
- MAPE: 29.92%

Peyebab



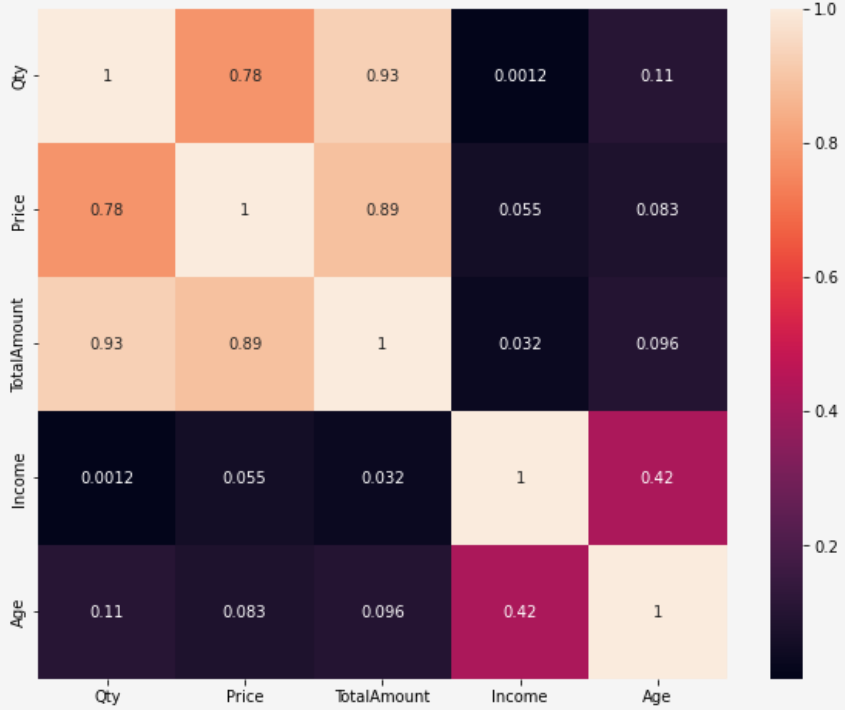
- Tidak Ada Trend yang jelas
- Residualnya sangat tinggi
- Walau terdapat sessional yang cukup jelas tetapi kedua variable diatas sudah menyebabkan

Solusi Peningkatan Perfoma

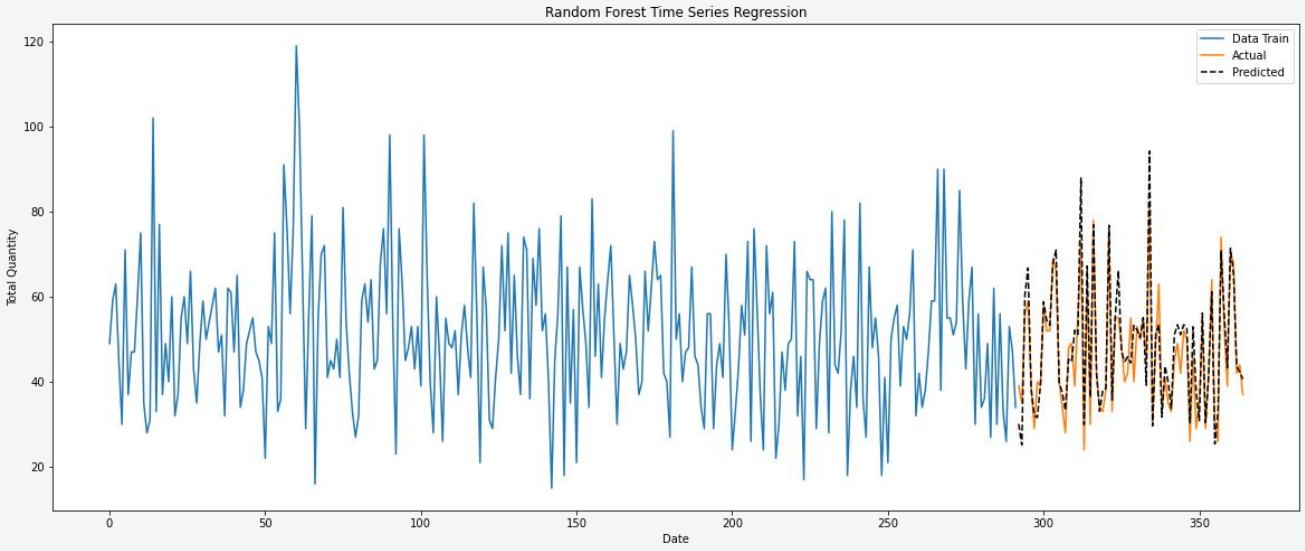
Melakukan Regresi mengunakan fitur fitur yang tersedia berupa agregasi pada fitur berikut:

- Qty dijumlahkan berdasarkan hari
- Harga dijumlahkan berdasarkan hari
- Total Amount dijumlahkan berdasarkan hari
- Income pelanggan direratakan berdasarkan hari
- Umur pelanggan direratakan berdasarkan hari

Hasil Korelasi Fitur-Fitur



Random Forest Model Result

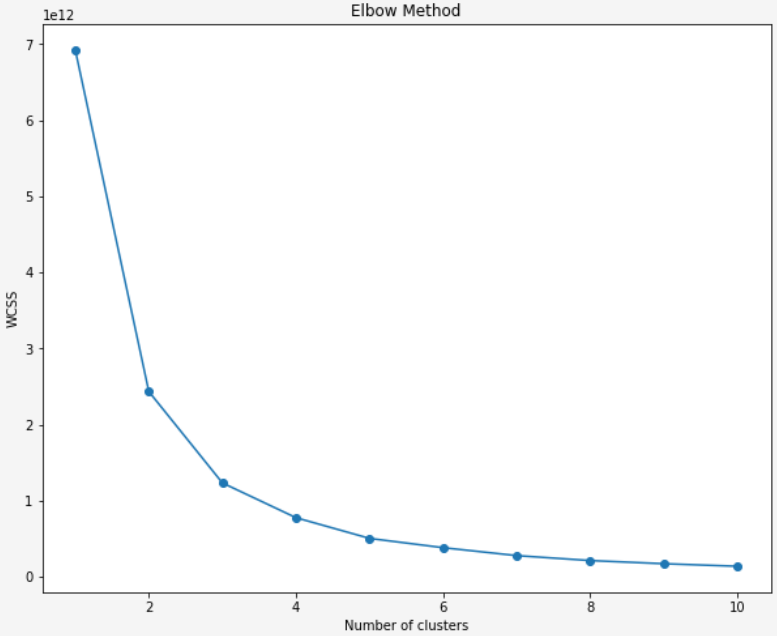


Hasil Performansi

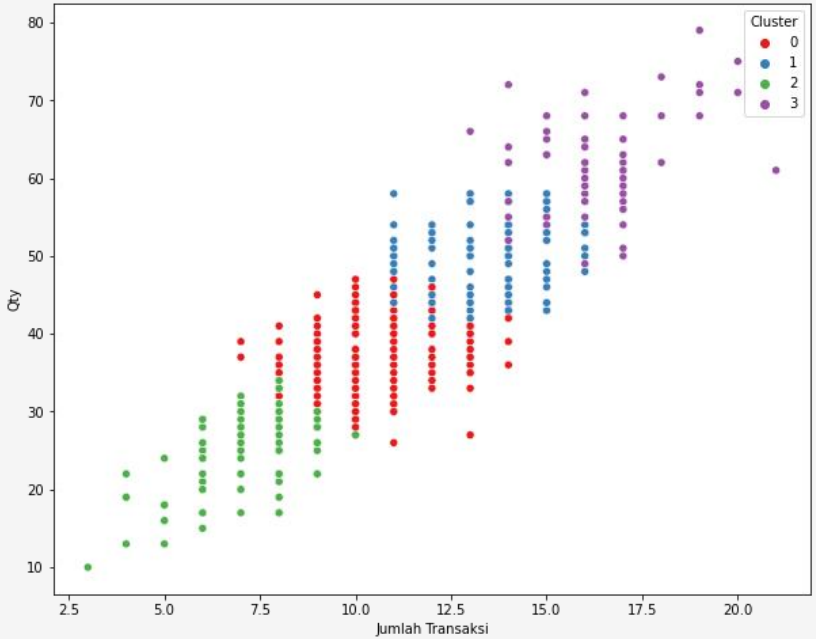
- MAE: 4.19
- RMSE: 5.52
- MAPE: 9.82%

Clustering Model

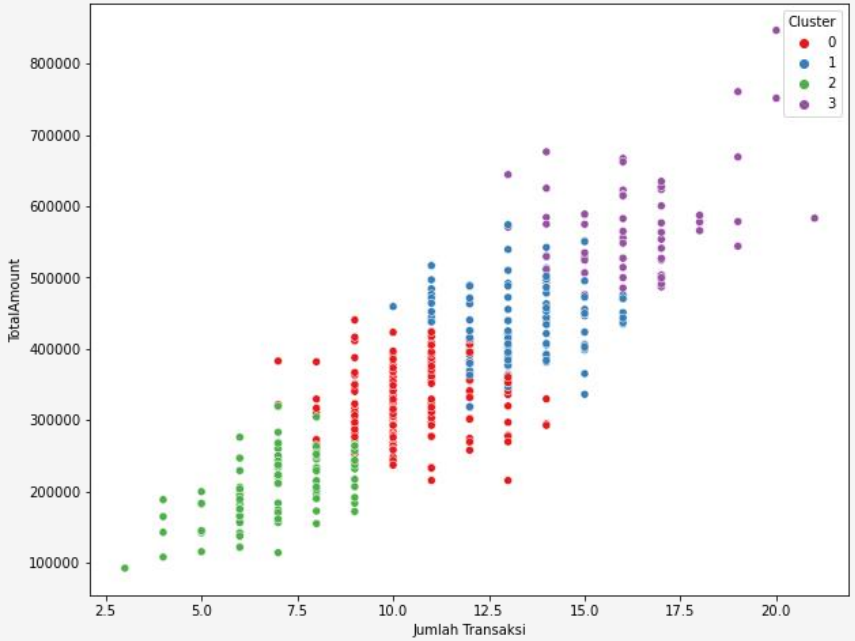
Elbow Method



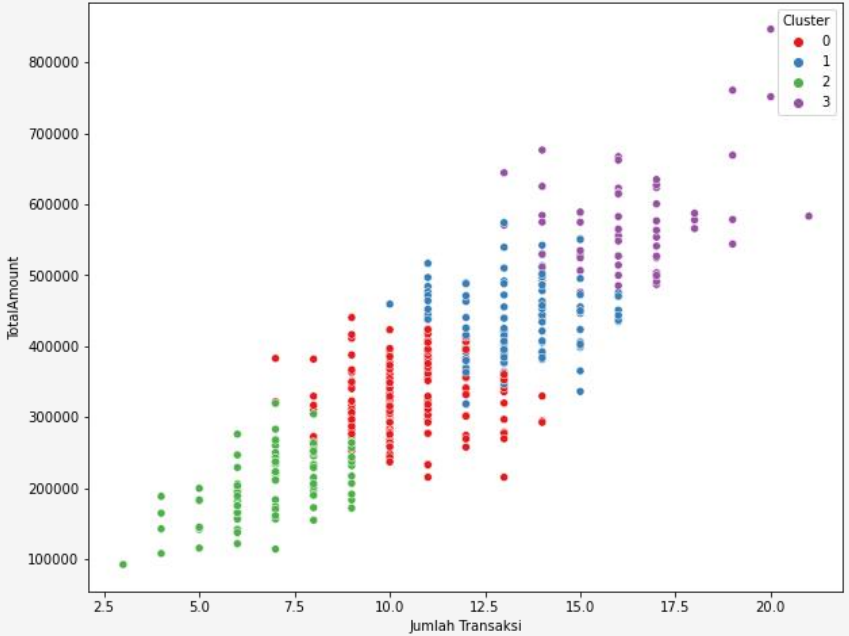
Result Knn-Clustering



Result Knn-Clustering



Result Knn-Clustering



Terima kasih Atas Perhatianya