

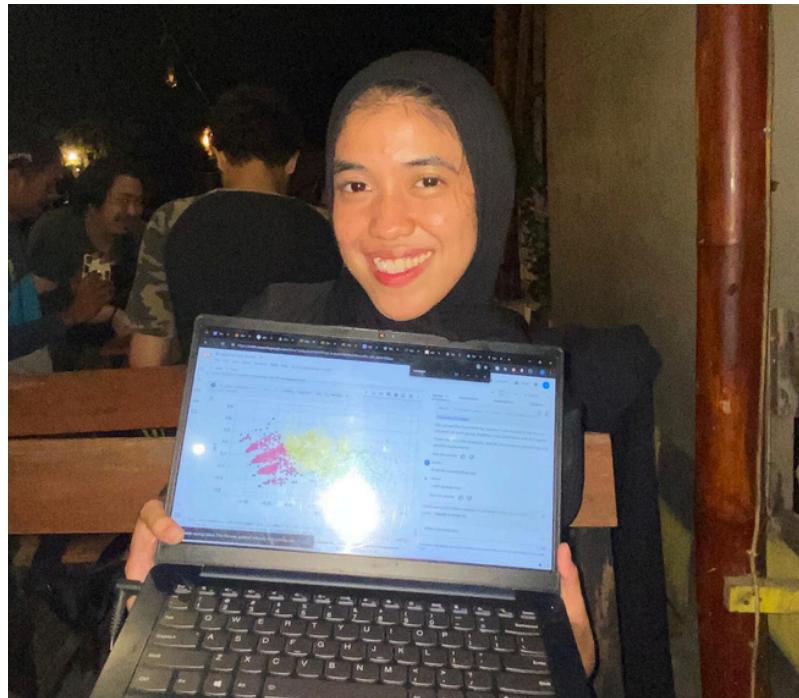


CUSTOMER PERSONALITY ANALYSIS

ANALISIS SEGMENTASI PELANGGAN
BERDASARKAN CUSTOMER PERSONALITY:
PENDEKATAN K-MEANS CLUSTERING DAN PCA

FINAL PROJECT

About Us



5026221043
Shof Watun Niswah



5026221091
Edward Yosafat Sirait



5026221110
M.Fazle M. Sidiki



5026221133
Baringga Aurico De E.



Daftar Isi

- 01** Overview Dataset
- 02** Latar Belakang
- 03** Tujuan dan Manfaat
- 04** Metodologi
- 05** Analisis
- 06** Kesimpulan

Overview Dataset

Sumber



Dataset

Customer Personality Analysis

[Akash Patel](#) · Updated 3 years ago

Usability 9.7 · 1 File (CSV) · 63 kB



2240 BARIS

29 KOLOM

- 25 Integer • 1 Float
- 2 String • 1 Date

Demografi Pelanggan			
1	ID	Nomor identifikasi unik pelanggan	Integer
2	Year_Birth	Tahun kelahiran pelanggan	Integer
3	Education	Jenjang pendidikan terakhir pelanggan	String (2n Cycle, Basic, Graduation, Master, PHD)
4	Marital_Status	Status pernikahan pelanggan	String (Absurd, Alone, Divorced, Married, Single, Together, Widow, YOLO)
5	Income	Pendapatan rumah tangga pelanggan per tahun	Integer
6	Kidhome	Jumlah anak-anak di rumah pelanggan	Integer
7	Teenhome	Jumlah remaja di rumah pelanggan	Integer
8	Dt_Customer	Tanggal pelanggan terdaftar di perusahaan	Date
9	Recency	Jumlah hari sejak transaksi terakhir pelanggan	Integer
10	Complain	Status pelanggan dalam melakukan komplain selama 2 tahun terakhir (pernah/tidak)	Boolean (1 = pernah melakukan complain, 0 = tidak pernah melakukan complain)

Overview Dataset

Produk			
11	MntWines	Jumlah produk wine yang dibeli dalam 2 tahun terakhir	Integer
12	MntFruits	Jumlah produk buah-buahan yang dibeli dalam 2 tahun terakhir	Integer
13	MntMeatProducts	Jumlah produk daging yang dibeli dalam 2 tahun terakhir	Integer
14	MntFishProducts	Jumlah produk ikan yang dibeli dalam 2 tahun terakhir	Integer
15	MntSweetProducts	Jumlah produk manisan yang dibeli dalam 2 tahun terakhir	Integer
16	MntGoldProducts	Jumlah produk emas yang dibeli dalam 2 tahun terakhir	Integer

Z_CostContact - 3
Z_Revenue - 11



Promosi			
17	NumDealsPurchases	Jumlah pembelian yang dilakukan menggunakan diskon	Integer
18	AcceptedCmp1	Penerimaan campaign pertama	Boolean
19	AcceptedCmp2	Penerimaan campaign kedua	Boolean
20	AcceptedCmp3	Penerimaan campaign ketiga	Boolean
21	AcceptedCmp4	Penerimaan campaign keempat	Boolean
22	AcceptedCmp5	Penerimaan campaign kelima	Boolean
23	Response	Penerimaan campaign terakhir	Boolean
Tempat transaksi			
24	NumWebPurchases	Jumlah pembelian yang dilakukan melalui website perusahaan	Integer
25	NumCatalogPurchases	Jumlah pembelian yang dilakukan melalui katalog	Integer
26	NumStorePurchases	Jumlah pembelian dilakukan secara langsung di toko	Integer
27	NumWebVisitsMonth	Jumlah kunjungan ke website selama sebulan	Integer



Latar Belakang

Customer Personality Analysis adalah proses dalam **memahami dan mengkarakterisasi** kepribadian, kebiasaan, psikologi, dan demografi **pelanggan** untuk mendapatkan insight dari preferensi, kebiasaan berbelanja, dan pembuatan keputusan.

Dengan memahami lebih dalam tentang preferensi produk dan perilaku pembelian pelanggan, **perusahaan dapat meningkatkan strategi pemasaran, personalisasi layanan, dan pengembangan produk** yang lebih sesuai dengan kebutuhan pasar.

Tujuan

Memahami profil demografis dan perilaku pelanggan serta mengelompokkan pelanggan ke dalam segmen-semen yang dapat digunakan untuk menyesuaikan strategi pemasaran dan penawaran produk.



Manfaat

1. Memungkinkan perusahaan untuk **mengidentifikasi segmen pelanggan** yang berbeda-beda berdasarkan demografi dan kebiasaan pelanggan
2. Dengan **memahami preferensi dan kebutuhan masing-masing segmen**, perusahaan dapat menyesuaikan komunikasi, produk, dan layanan mereka untuk meningkatkan kepuasan pelanggan.
3. Mengetahui segmen pelanggan yang lebih cenderung untuk merespons kampanye pemasaran tertentu, memungkinkan perusahaan untuk **mengalokasikan sumber daya secara lebih efisien**.
4. Menggunakan wawasan dari analisis untuk **mengembangkan produk baru atau meningkatkan produk** yang ada yang lebih sesuai dengan preferensi pelanggan.

Metodologi



Pengumpulan
Data



Penyesuaian
Data

- Dataset format CSV -> XLS
- Drop Missing Values
- Menambah Features Baru
- Filtering



Analisis

- EDA
- Correlation Matrix
- Minmax Scalling
- K-Means Clustering
- PCA



Penyesuaian Data

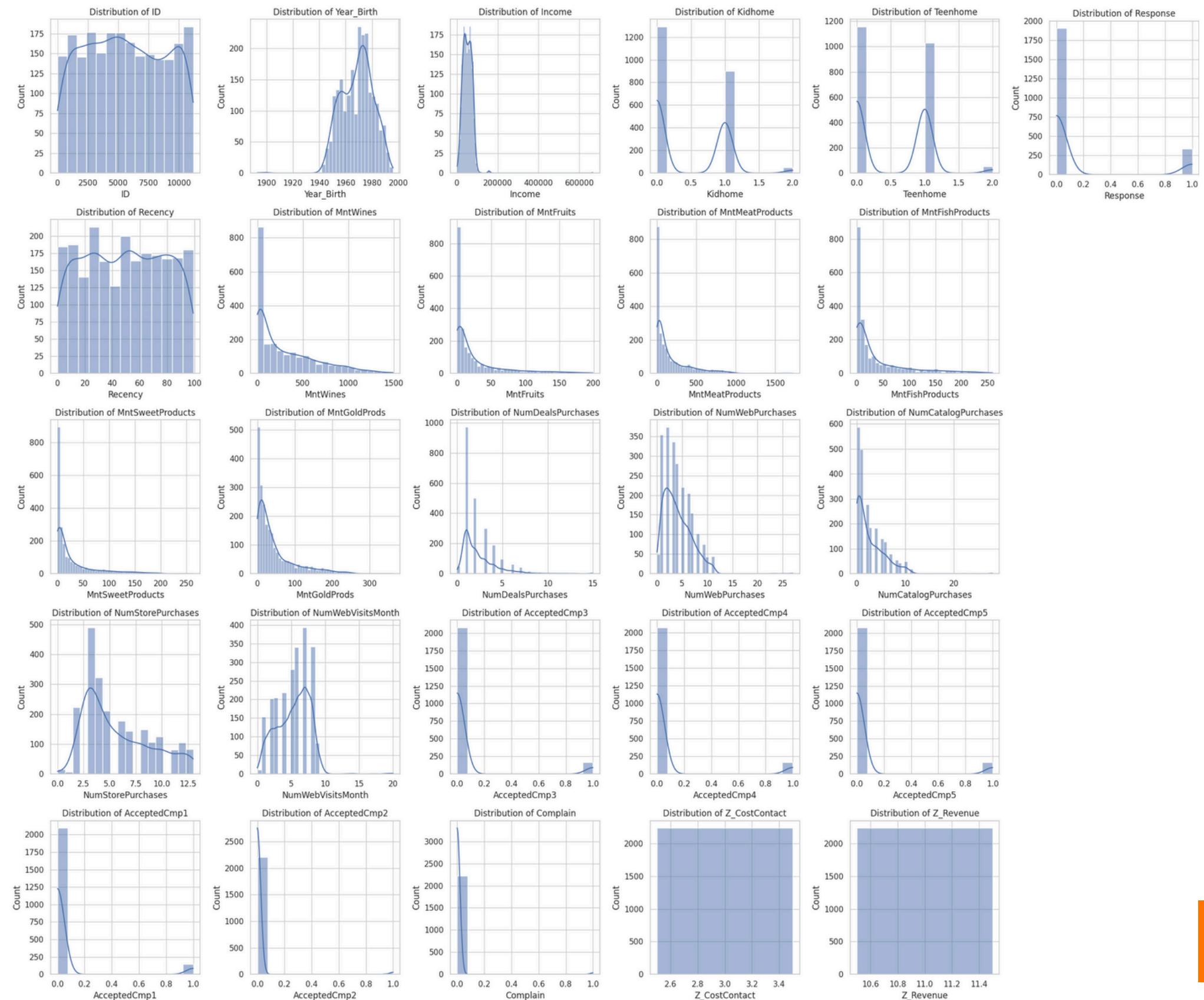
Data Awal

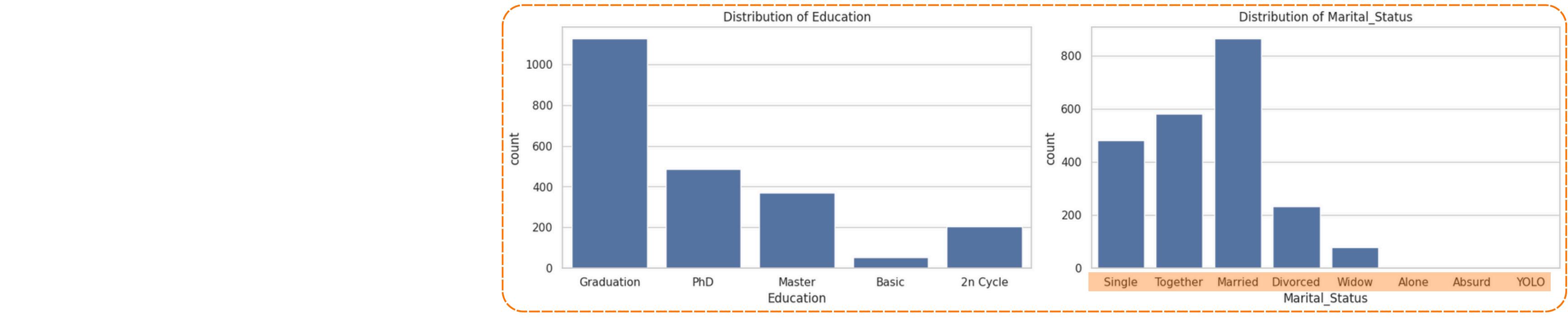
2216 BARIS

Drop Missing Value
24 Baris

29 KOLOM

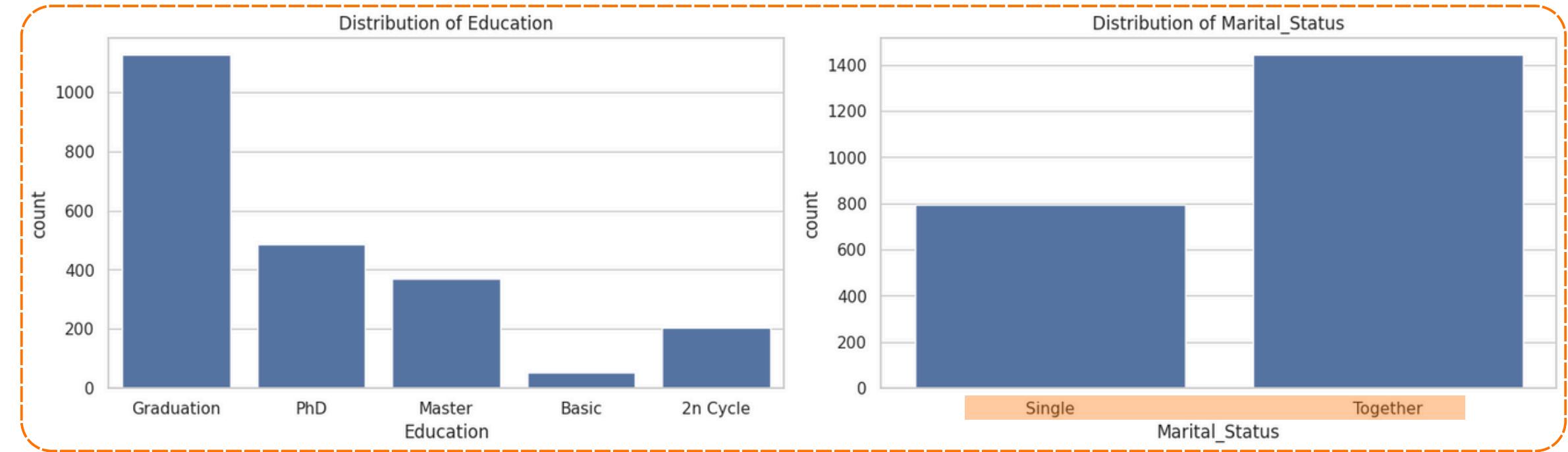
- 25 Integer
- 1 Float
- 2 Object
- 1 Date





Penyesuaian Data Kategorikal

Mengubah Data Kolom *Marital Status*



Divorced, Widow, Absurd, YOLO -> **Single** && Married -> **Together**

Penyesuaian Data

Mengubah data kategorikal menjadi numerik

Data Setelah

Menambah *Features* Baru

2216 BARIS

PeopleInHome	<ul style="list-style-type: none">Single = 1 + KidHome + TeenHomeTogether = 2 + KidHome + TeenHome
TotalPurchases	NumWebPurchases + NumCatalogPurchases + NumStorePurchases
TotalCampaign Accepted	AcceptedCmp1 + AcceptedCmp2 + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + Response

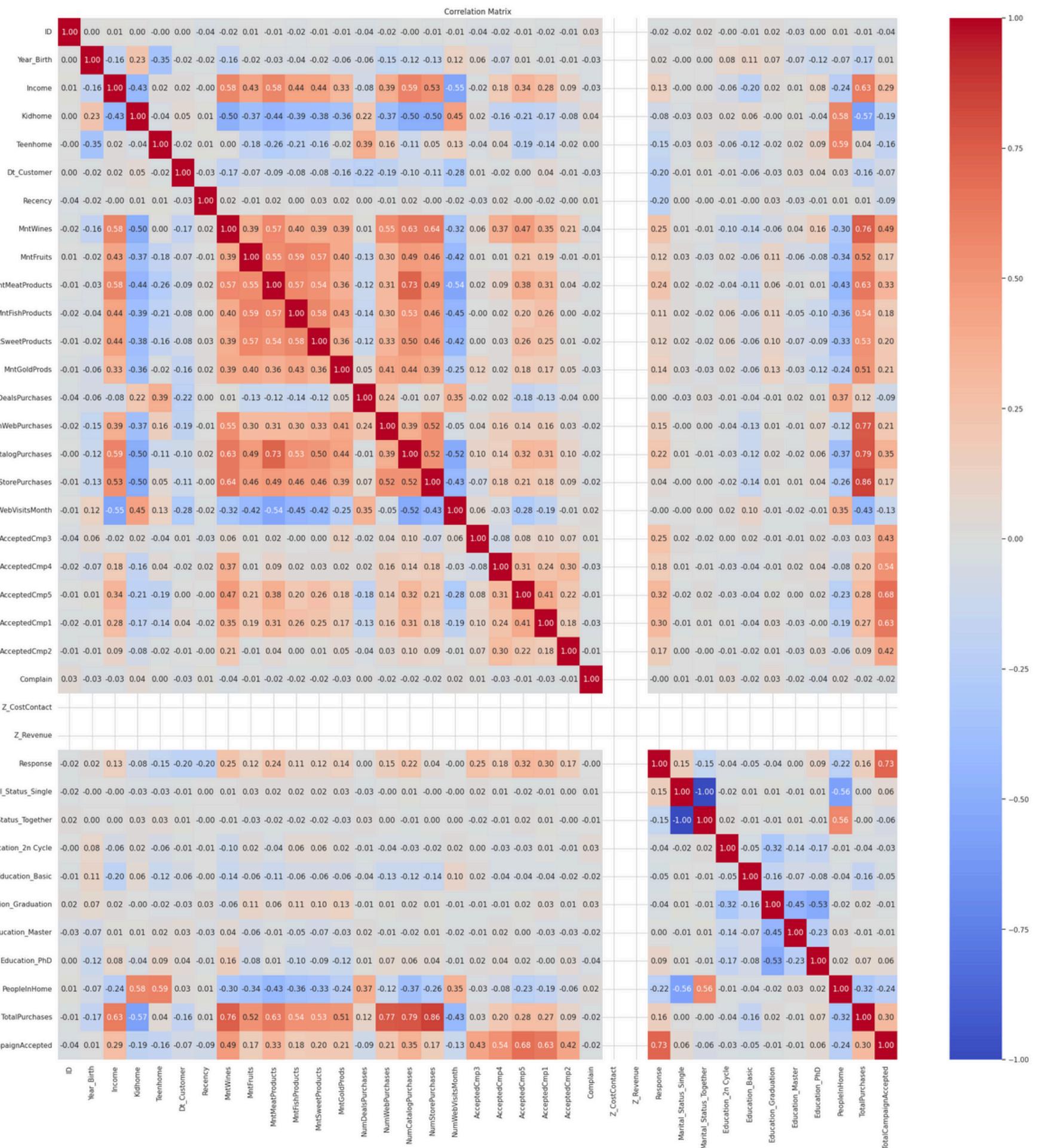
37 KOLOM

- 35 Integer
- 1 Float
- 1 Date



Correlation Matrix

Data Awal



Filtering

Menentukan **features** yang ingin digunakan, berdasarkan hasil pembahasan dan penyesuaian kebutuhan data terhadap tujuan clustering.

2216 BARIS

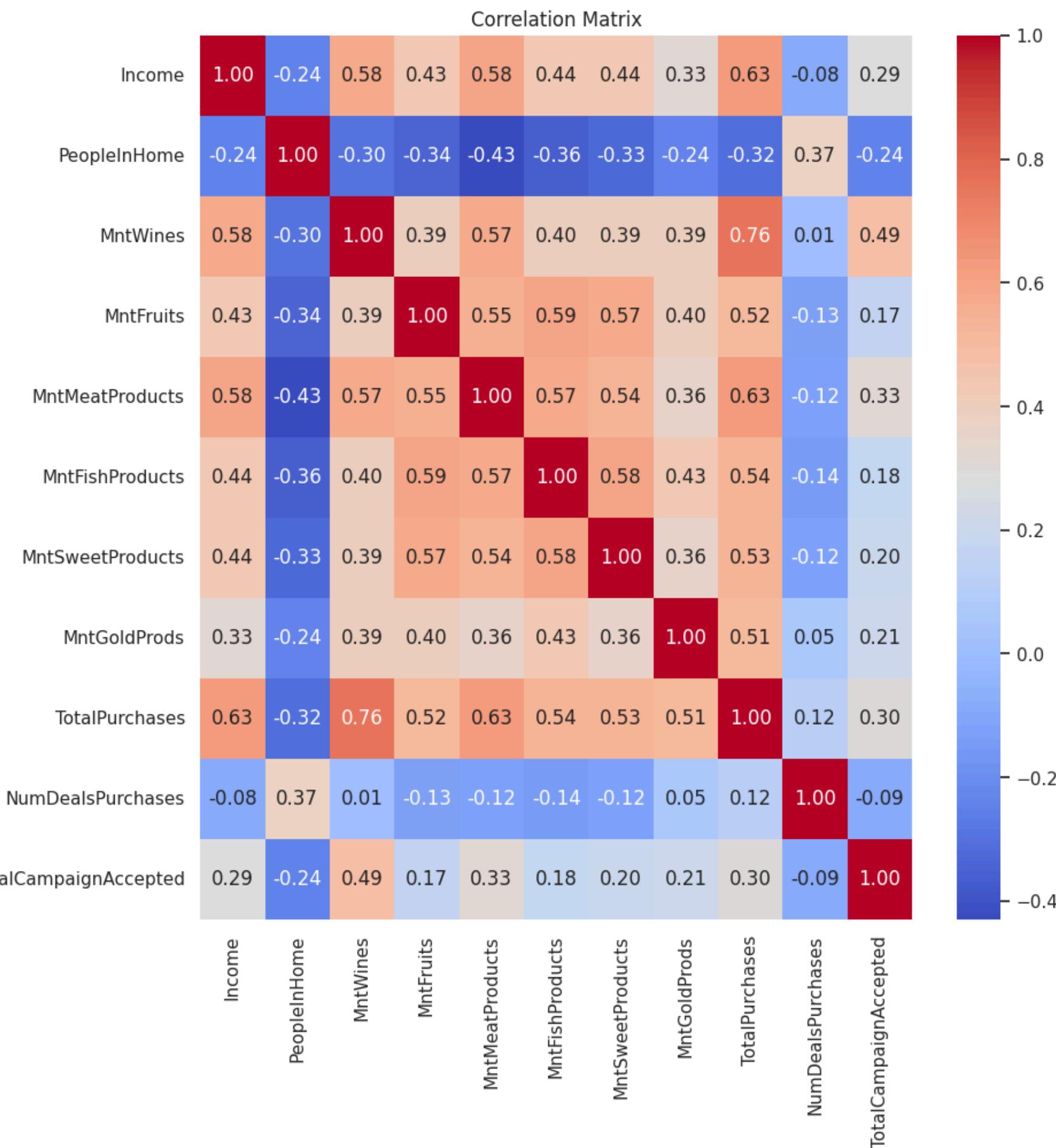
11 KOLOM

Income	Pendapatan pelanggan
PeopleInHome	Jumlah orang yang tinggal di rumah pelanggan.
MntWines	Jumlah pembelian produk wine
MntFruits	Jumlah pembelian produk buah-buahan
MntMeatProducts	Jumlah pembelian produk daging
MntFishProducts	Jumlah pembelian produk ikan
MntSweetProducts	Jumlah pembelian produk gula-gulaan
MntGoldProds	Jumlah pembelian produk emas
TotalPurchases	Total pembelian yang dilakukan pelanggan.
NumDealsPurchases	Total pembelian menggunakan diskon
TotalCampaignAccepted	Frekuensi pelanggan menerima kampanye promosi dari perusahaan.



Correlation Matrix

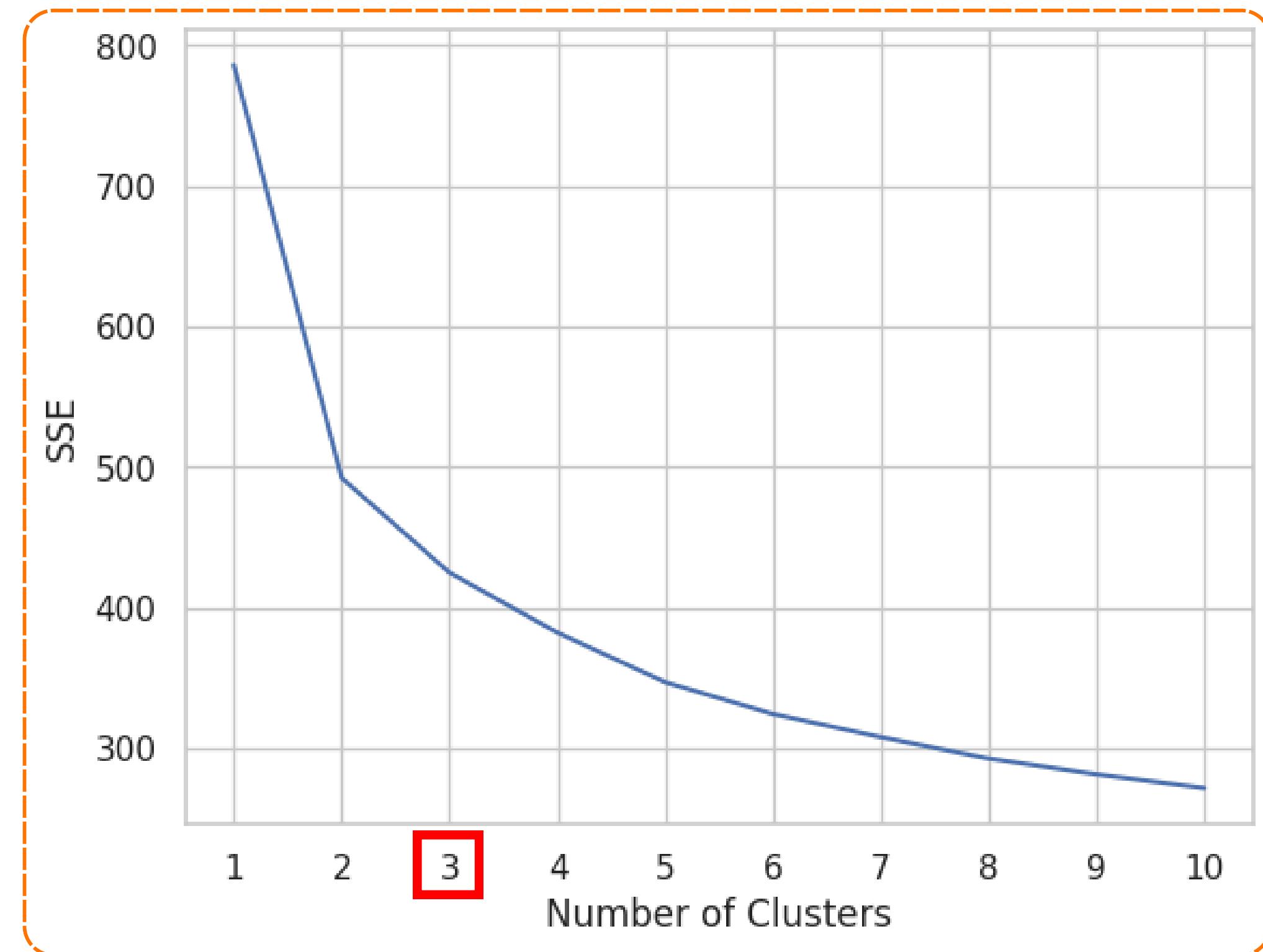
Setelah di *Minmax Scaling* dan *Filtering*



Jumlah Klaster berdasarkan Elbow Method

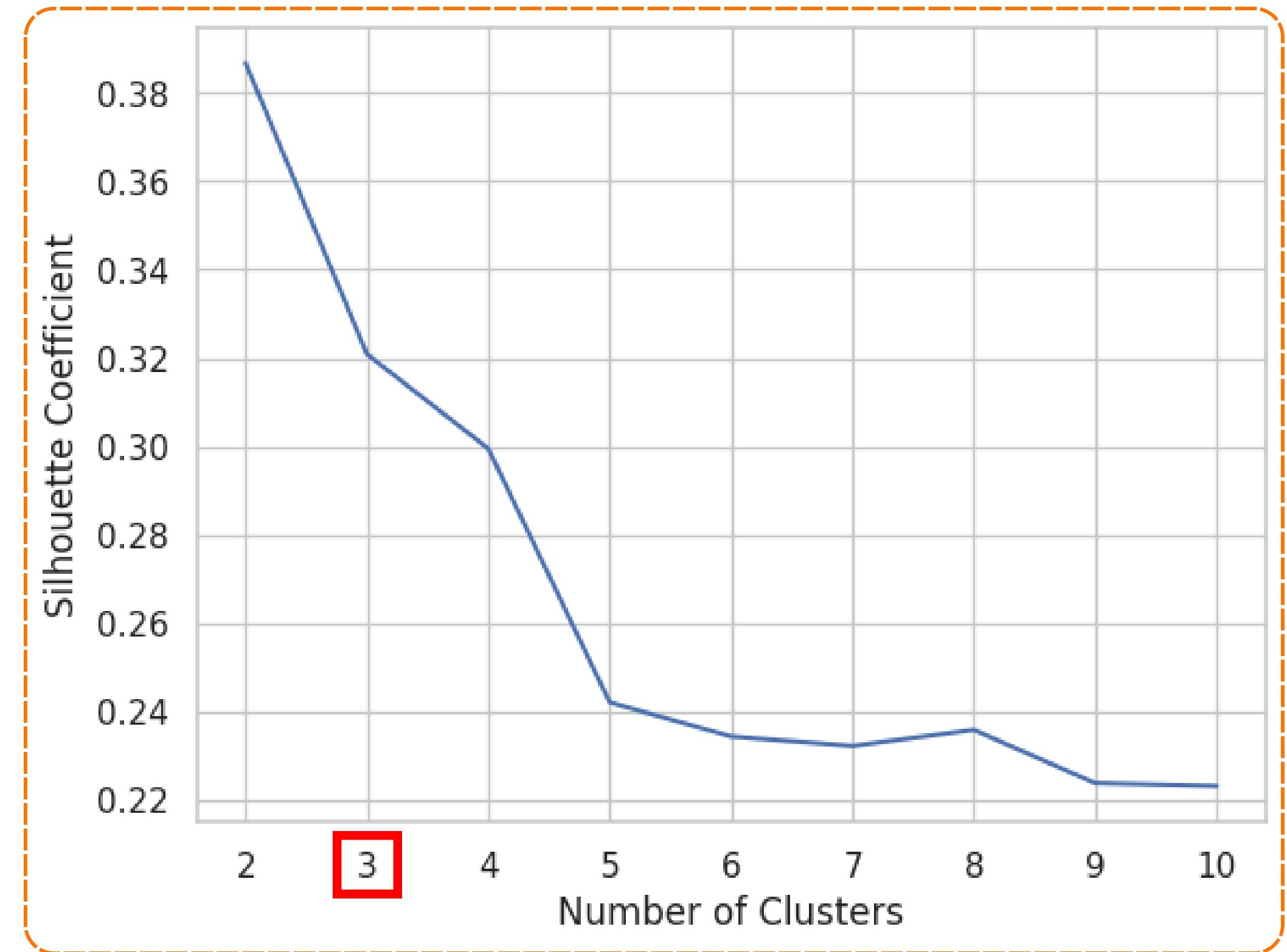
3

```
[28] kl = KneeLocator(range(1, 11), sse, curve="convex", direction="decreasing")  
kl.elbow  
3
```



Pembentukan Klaster berdasarkan Koefisien Silhouette

3



K-Means Clustering

Data Asli (Tanpa PCA)

SSE	424.9063152344339
Sentroid K1	[0.10934255, 0.20263158, 0.37859344, 0.40123777, 0.25573028, 0.45662264, 0.30736842, 0.25429743, 0.61171053, 0.09403509, 0.17431579]
Sentroid K2	[0.09336129, 0.39922145, 0.39017491, 0.11800351, 0.12697157, 0.11697239, 0.09936871, 0.20693874, 0.58515355, 0.2171857, 0.13114187]
Sentroid K3	[0.05370237, 0.47742906, 0.04082739, 0.02983533, 0.01691129, 0.03226246, 0.0216471, 0.054264, 0.20706148, 0.14881055, 0.03456578]
Iterasi	7

Dengan PCA (Semua Komponen)

SSE	424.9063152344339
Sentroid K1	[5.93648084e-01, -1.19238494e-01, -8.92616918e-02, 5.79851435e-02, -1.29549749e-02, -1.61963662e-02, 4.77459353e-03, 6.03773755e-03, -2.67620865e-03, 8.23218060e-03, 2.23362365e-04]
Sentroid K2	[1.83623606e-01, 9.11514648e-02, 6.69259865e-02, -9.60638107e-02, 8.54443685e-03, 1.22506837e-04, 5.64802051e-04, -2.09399311e-03, 1.12960030e-02, -9.29592387e-03, 1.57417038e-03]
Sentroid K3	[-3.33275146e-01, -4.60309999e-02, -2.19525657e-03, 2.40601371e-02, 1.04465025e-03, 1.26573292e-04, -2.13137616e-03, -1.42527715e-03, -4.52097124e-03, 1.25774567e-03, -8.73574895e-04]
Iterasi	7

K-Means Clustering

Dengan PCA (1 Komponen)

SSE	41.28055497004375
Sentroid K1	[0.59780979]
Sentroid K2	[-0.35220038]
Sentroid K3	[0.10658273]
Iterasi	15

Dengan PCA (2 Komponen)

SSE	127.88225002783344
Sentroid K1	[0.57603482 -0.0948964]
Sentroid K2	[-0.33907671 -0.05887377]
Sentroid K3	[0.11696315 0.21629235]
Iterasi	19

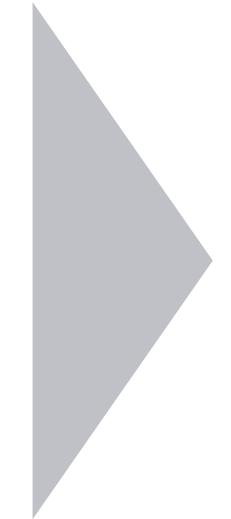
Dengan PCA (3 Komponen)

SSE	213.24247831908565
Sentroid K1	[0.57577455 -0.11893705 -0.10351418]
Sentroid K2	[-0.32759368 -0.04572504 0.0020551]
Sentroid K3	[0.19303336 0.2131341 0.09211869]
Iterasi	36

Perbandingan V-Score

Klasterisasi **Data Asli**
& **Semua PCA**

1,0



Klasterisasi **Data Asli & 1 PCA**

0.6518462467250382

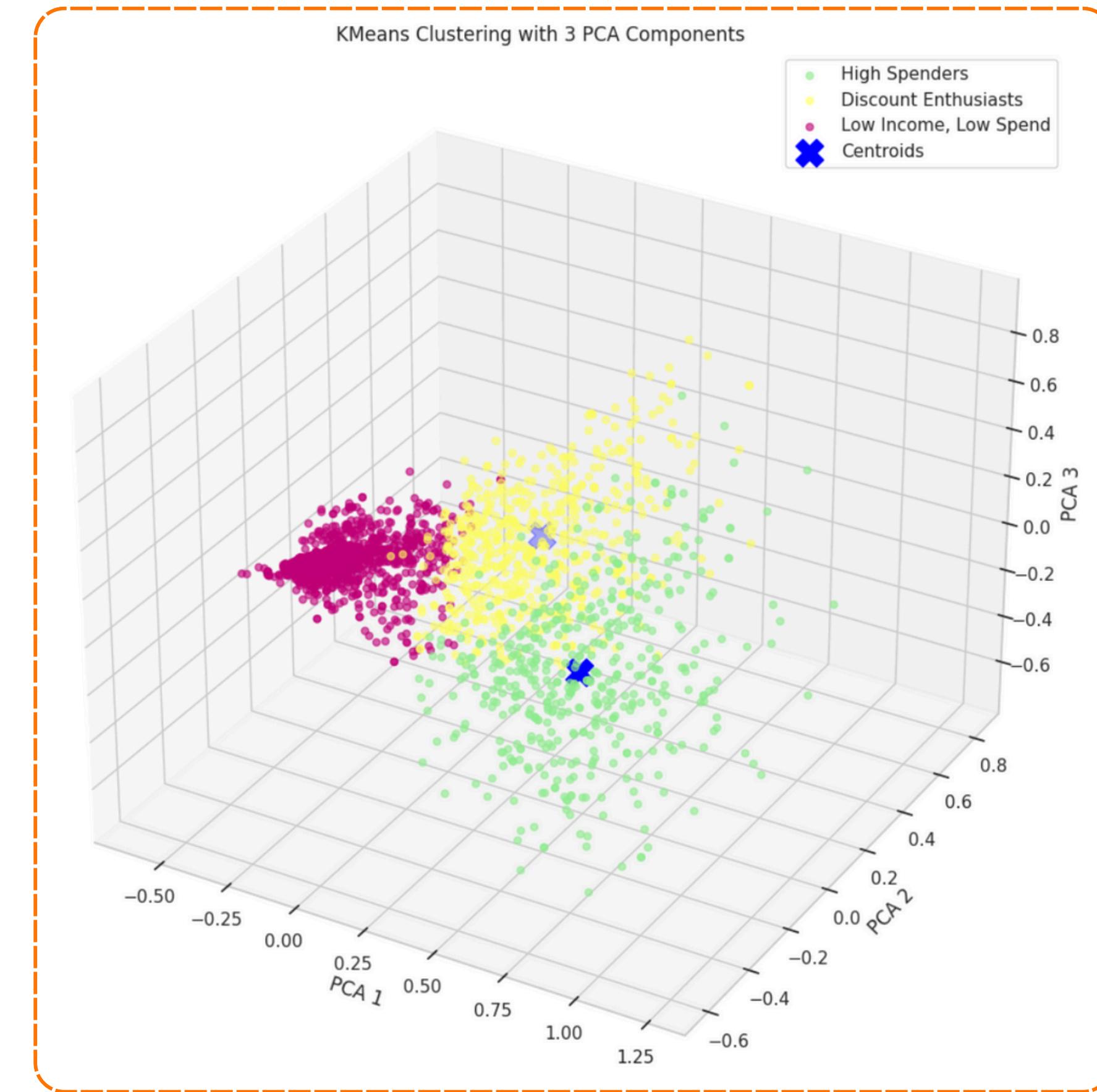
Klasterisasi **Data Asli & 2 PCA**

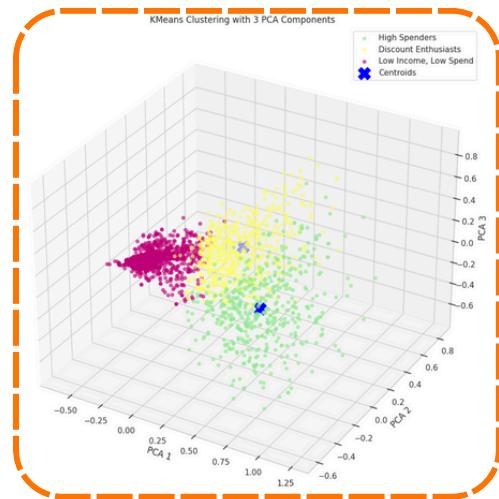
0.7794171966012491

Klasterisasi **Data Asli & 3 PCA**

0.8372755937747758

Scatter Plot 3 PCA





Hasil Analisis

	Income	PeopleInHome	MntWines	MntFruits	MntMeatProducts	\
Cluster	0.109343	0.202632	0.378593	0.401238	0.255730	
0	0.093361	0.399221	0.390175	0.118004	0.126972	
1	0.053702	0.477429	0.040827	0.029835	0.016911	
	MntFishProducts	MntSweetProducts	MntGoldProds	TotalPurchases	\	
Cluster	0.456623	0.307368	0.254297	0.611711		
0	0.116972	0.099369	0.206939	0.585154		
1	0.032262	0.021647	0.054264	0.207061		
	NumDealsPurchases	TotalCampaignAccepted				
Cluster	0.094035	0.174316				
0	0.217186	0.131142				
1	0.148811	0.034566				

Klaster 0: High Spenders

475/2216

Pelanggan dalam klaster ini memiliki pendapatan tertinggi, cenderung mengeluarkan lebih banyak uang untuk ikan dan buah, serta melakukan banyak pembelian. Mereka kurang tertarik pada diskon tetapi merespons tinggi terhadap kampanye promosi.

Klaster 1: Discount Enthusiast

578/2216

Pelanggan klaster ini berpendapatan rendah, tinggal di rumah tangga besar, dan mengeluarkan lebih banyak uang untuk wine tetapi kurang untuk makanan lainnya. Mereka sering membeli, sangat tertarik pada diskon, dan merespons kampanye promosi cukup tinggi, meskipun lebih rendah dibandingkan High Spenders.

Klaster 2: Low Income, Low Spend

1163/2216

Pelanggan klaster ini berpendapatan paling rendah, tinggal di rumah tangga besar, dan mengeluarkan sangat sedikit untuk semua produk dengan total pembelian rendah. Mereka menunjukkan ketertarikan yang biasa saja pada diskon dan sangat sedikit merespons kampanye promosi.

Kesimpulan

01

Analisis bertujuan untuk memahami profil demografis dan segmentasi pelanggan yang bisa digunakan untuk menyesuaikan strategi pemasaran dan penawaran produk.

EDA

MINMAX SCALING

ELBOW METHOD

SILHOUETTE COEFF.

K-MEANS

PCA

2240 BARIS

29 KOLOM

2216 BARIS

29 KOLOM

2216 BARIS

37 KOLOM

2216 BARIS

11 KOLOM

02

3 Segmentasi Pelanggan:

03

HIGH SPENDERS

DISCOUNT ENTHUSIAST

LOW INCOME, LOW SPEND

04

Segmentasi pelanggan ini memberikan berbagai manfaat strategis bagi perusahaan dalam **menyesuaikan komunikasi, produk, dan layanan mereka untuk meningkatkan kepuasan pelanggan**. Sehingga, dapat digunakan untuk mengembangkan atau meningkatkan produk sesuai **preferensi pelanggan**, serta meningkatkan loyalitas pelanggan.

