

# Predict Customer Personality to boost marketing campaign by using Machine Learning

Supported by:  
**Rakamin Academy**  
Career Acceleration School  
[www.rakamin.com](http://www.rakamin.com)



**Created by:**

**RAMADHANI nURUL FAZRI**

[rnurul08@gmail.com](mailto:rnurul08@gmail.com)

[www.linkedin.com/in/ramadhani-nurul-fazri](https://www.linkedin.com/in/ramadhani-nurul-fazri)

“Saya individu yang berkomitmen dan berdedikasi, memiliki latar belakang pendidikan di bidang Teknik Geologi. Kemampuan analitis yang kuat dan keterampilan komunikatif yang baik, saya mampu bekerja secara efektif dalam tim maupun secara mandiri. Saya memiliki minat yang besar dalam analisis data dan selalu berusaha untuk belajar dan berkembang, serta berkontribusi secara positif terhadap tujuan organisasi. Selain itu, saya memiliki ketertarikan dalam penerapan teknologi dan inovasi untuk meningkatkan efisiensi serta hasil kerja.”

“Sebuah perusahaan dapat berkembang dengan pesat saat mengetahui perilaku customer personality nya, sehingga dapat memberikan layanan serta manfaat lebih baik kepada customers yang berpotensi menjadi loyal customers. Dengan mengolah data historical marketing campaign guna menaikkan performa dan menyasar customers yang tepat agar dapat bertransaksi di platform perusahaan, dari insight data tersebut fokus kita adalah membuat sebuah model prediksi kluster sehingga memudahkan perusahaan dalam membuat keputusan ”

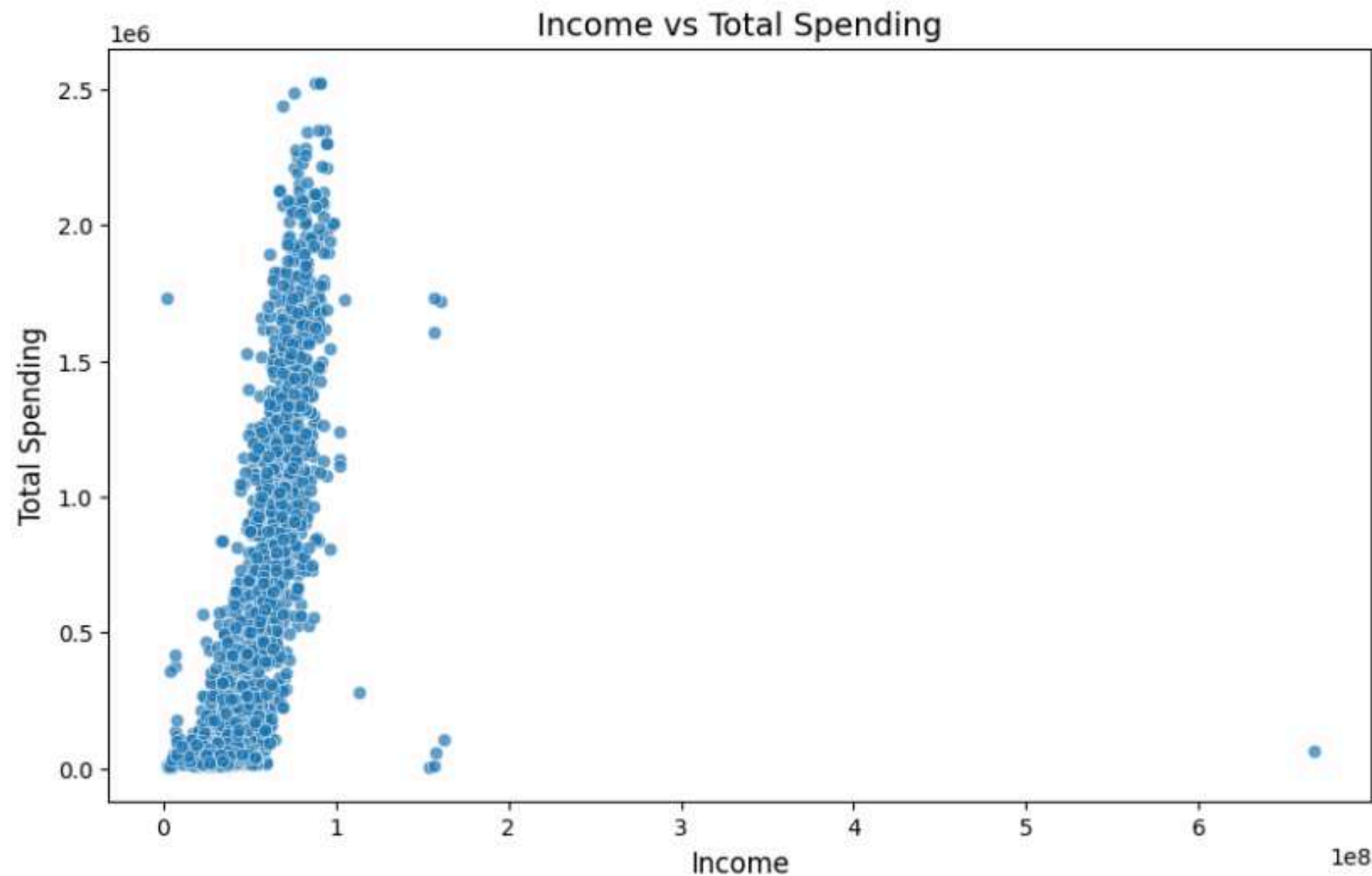


## 2440 Data Observasi 30 Variabel

- Tipe masing-masing kolom sudah sesuai
- Ditemukan baris kosong pada kolom Income sebanyak 24
- Tidak ditemukan data duplikat

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



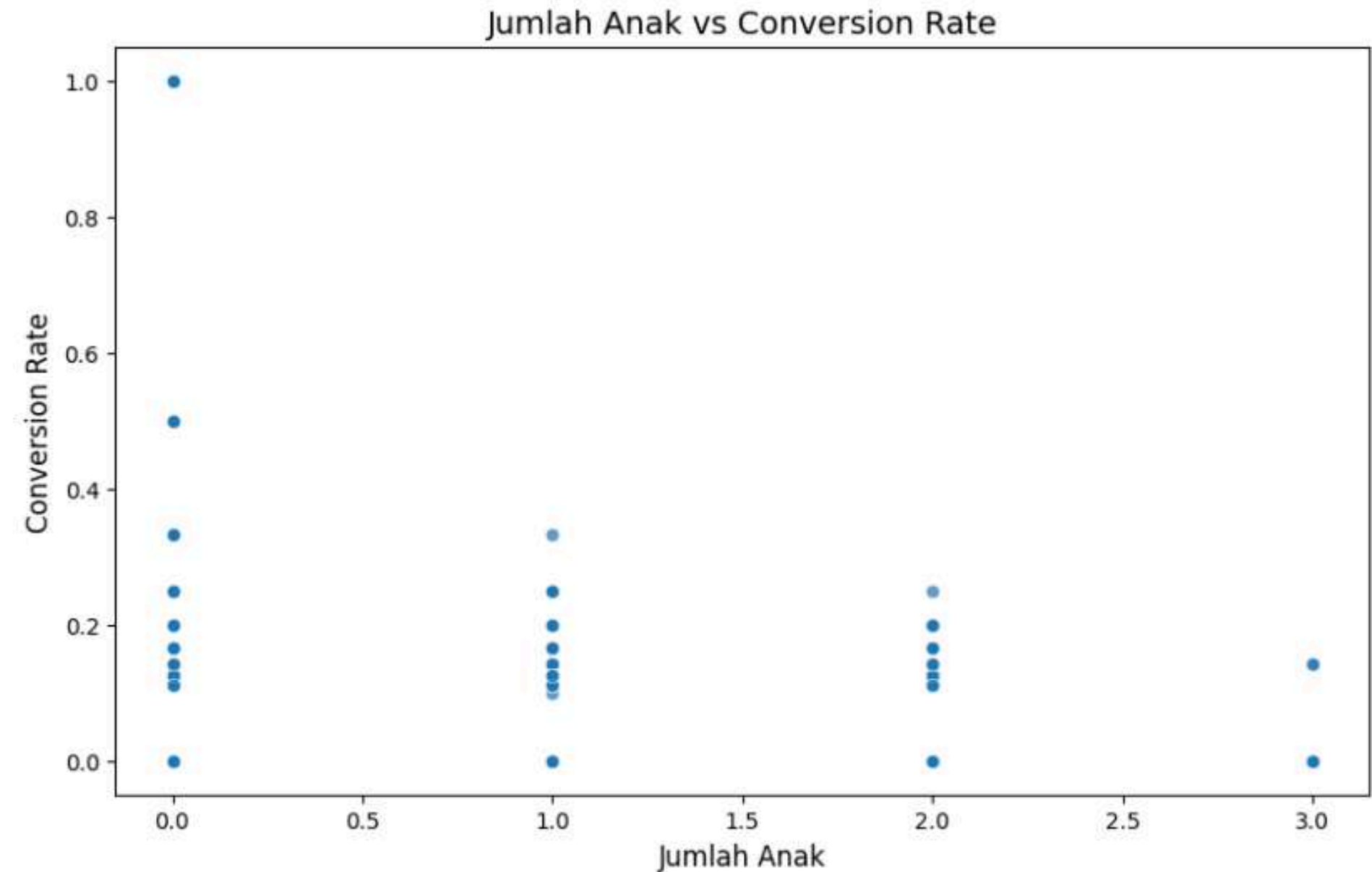


Income dengan Total Spending memiliki **korelasi linear** dimana semakin besar nilai income maka semakin besar total spending

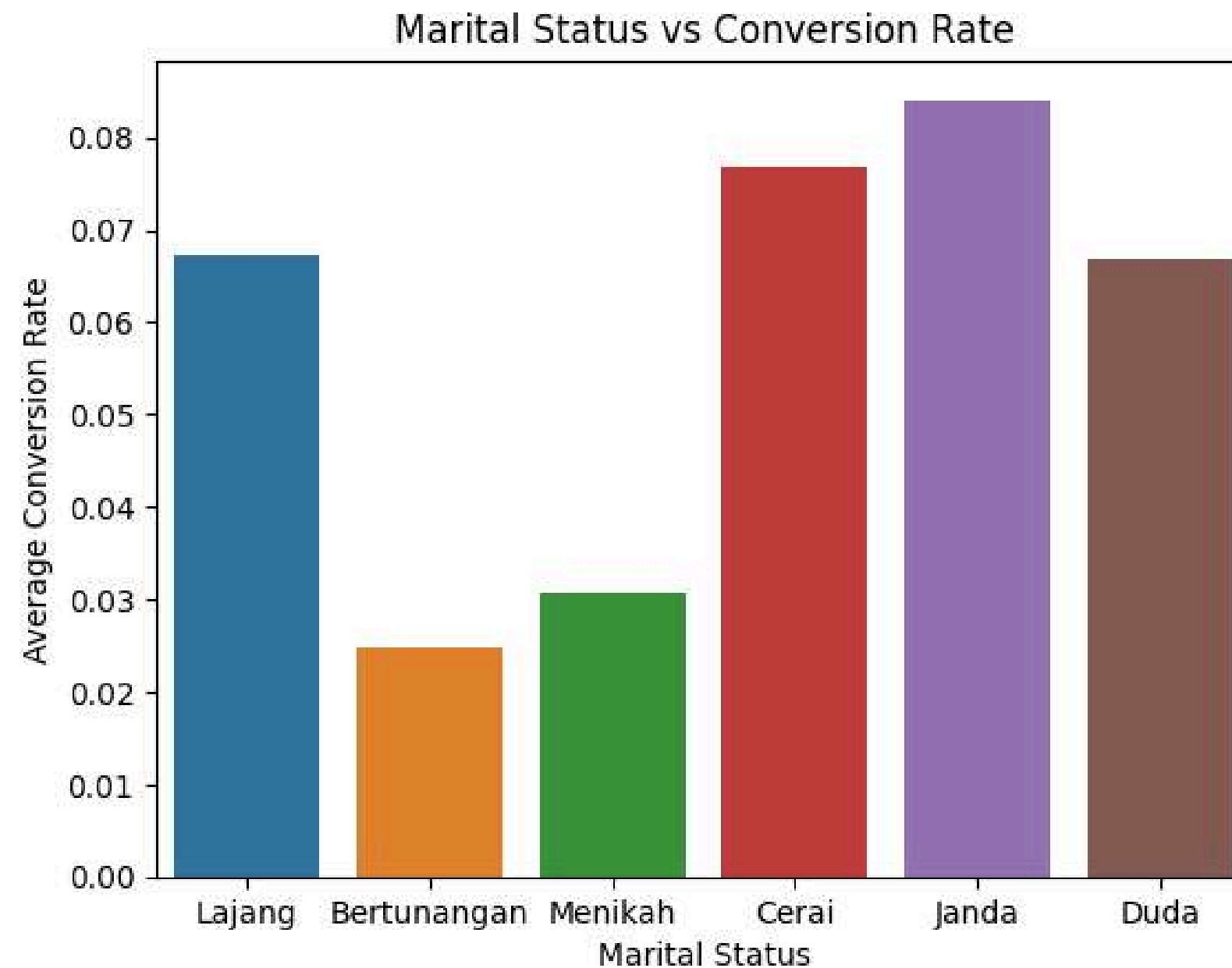
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Jumlah anak terhadap Conversion Rate **tidak memiliki korelasi**

Kelompok dengan lebih banyak anak (2 atau lebih) cenderung memiliki conversion rate yang lebih rendah, sementara kelompok tanpa anak atau dengan 1 anak menunjukkan conversion rate yang lebih tinggi

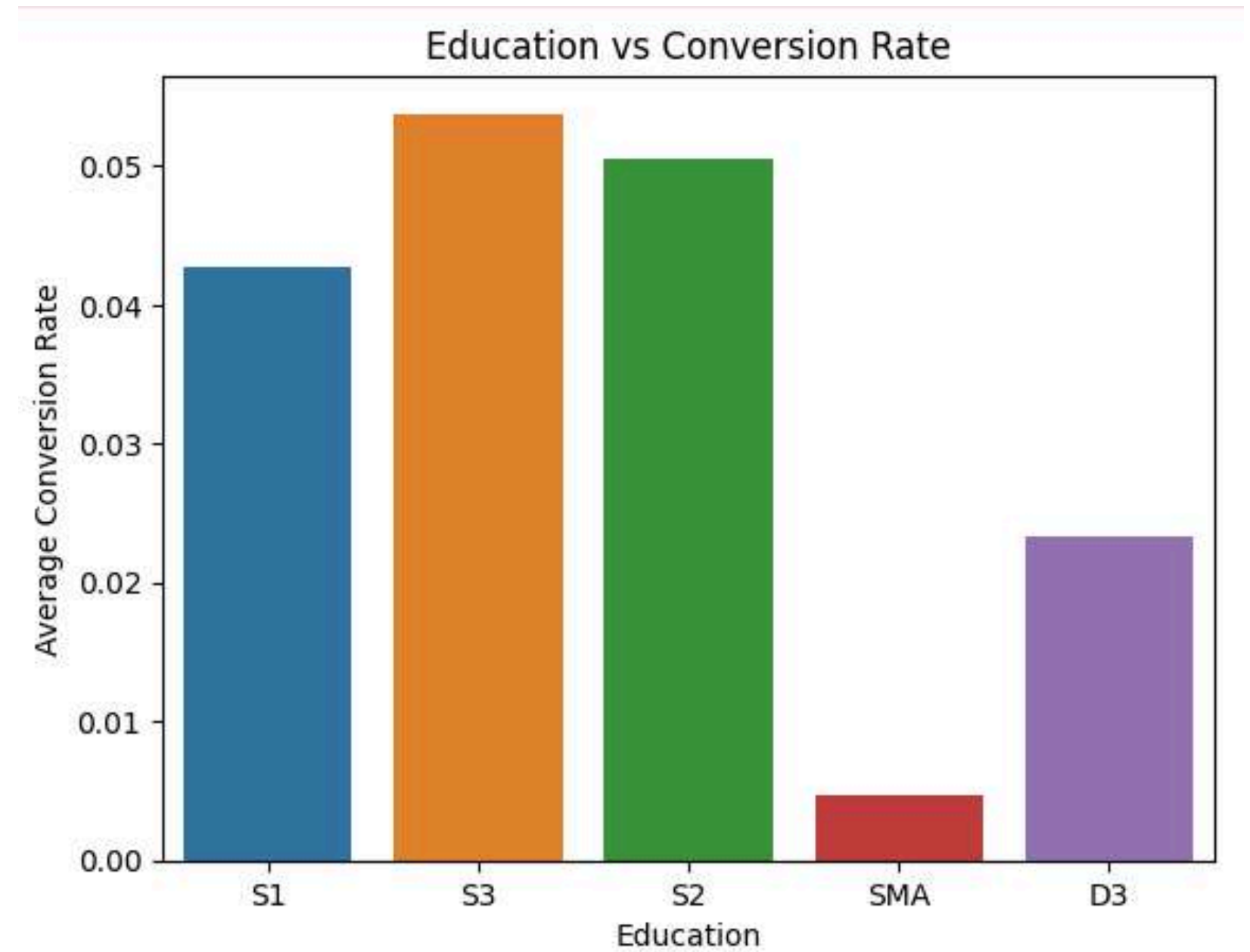


Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



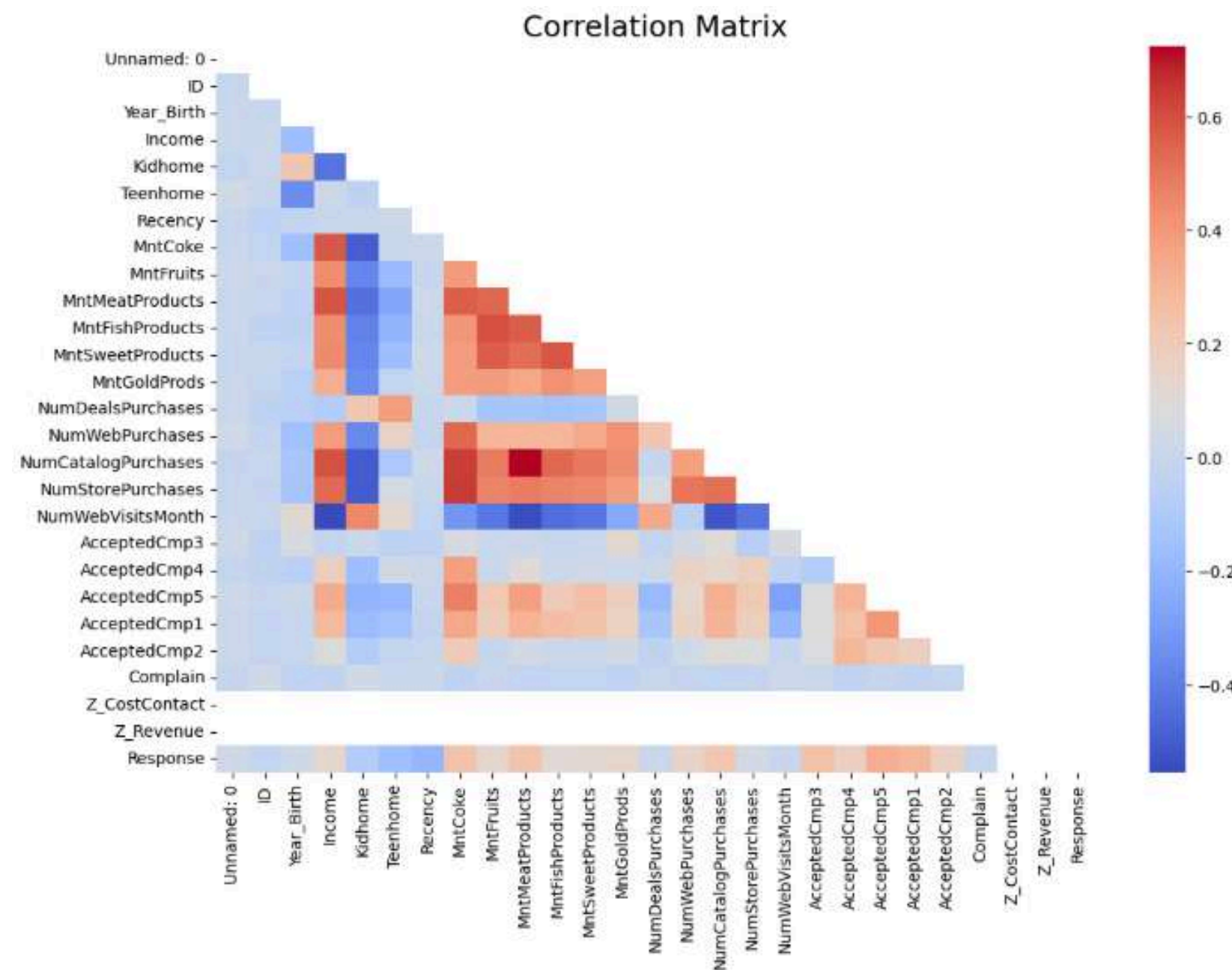
Kategori **janda** mendominasi dalam Conversion Rate berdasarkan Marital Status

Kategori **S3** mendominasi kolom Education dalam Conversion Rate



Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



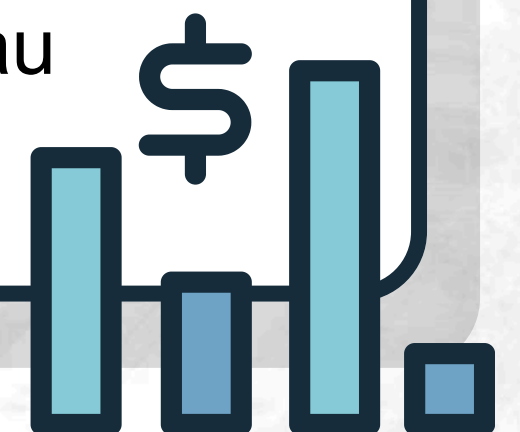


## 1. Korelasi Positif

- MntMeatProducts dengan Total Spending (MntCoke, MntFruits, dst.)
- NumCatalogPurchases dengan NumStorePurchases

## 2. Korelasi Negatif

- Recency dengan variabel seperti Response atau Purchases

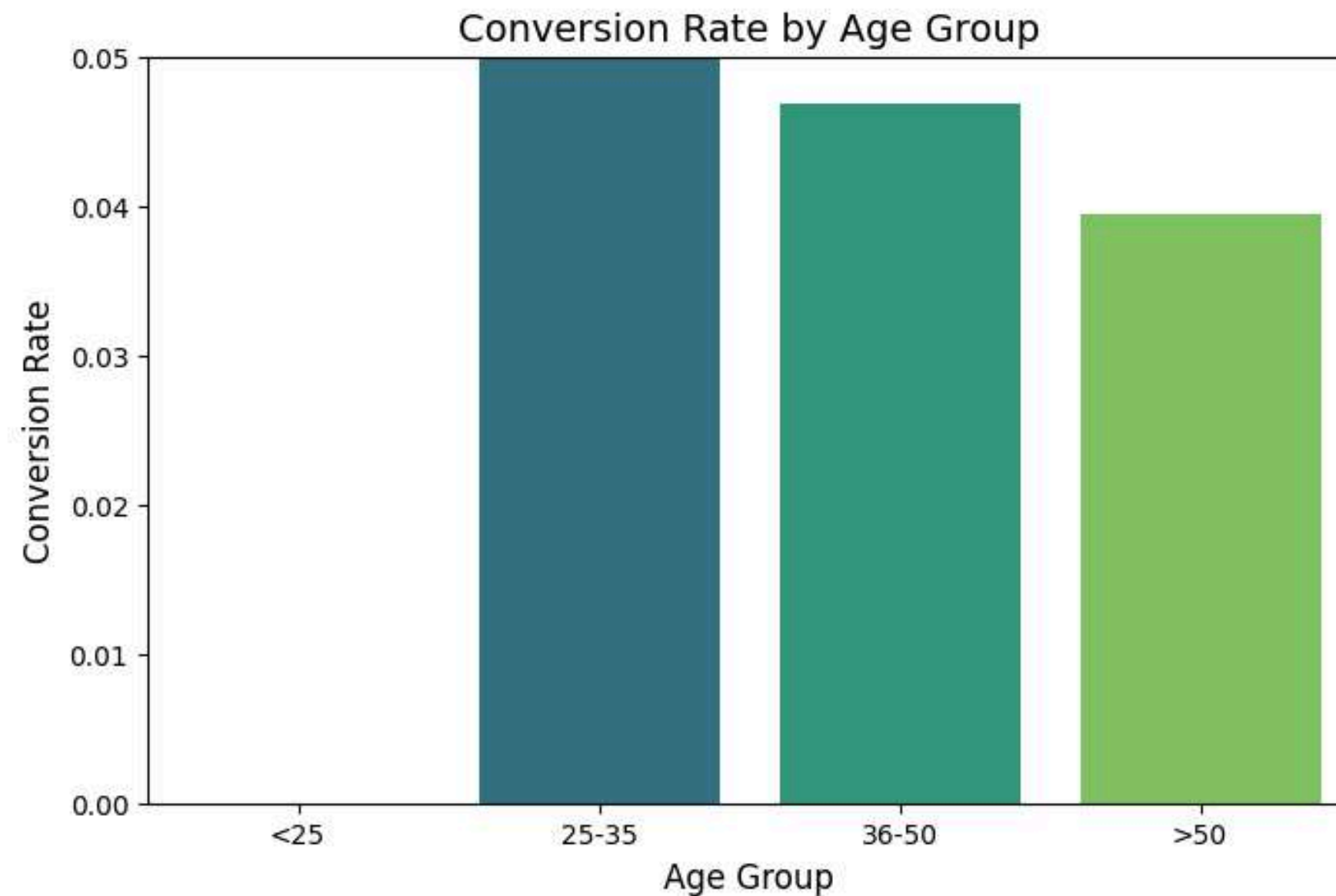


Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



1. Kelompok <25: Tidak ada data untuk kelompok umur ini
2. Kelompok 25-35: Memiliki conversion rate tertinggi sebesar 10.88%, yang menunjukkan bahwa mereka adalah kelompok yang paling responsif terhadap campaign.
3. Kelompok 36-50: Memiliki conversion rate 4.69%, menunjukkan penurunan responsivitas.
4. Kelompok >50: Memiliki conversion rate terendah, yaitu 3.95%, kelompok ini paling tidak responsif terhadap campaign

Age Group	Conversion Rate
<25	NaN
25-35	0.1088
36-50	0.0469
>50	0.0395



Terdapat tren yang jelas bahwa semakin tua umur customer, semakin rendah conversion rate mereka

Kelompok umur 25-35 lebih responsif terhadap campaign dibandingkan kelompok lainnya

# Rekomendasi Bisnis

- ① Fokus campaign pada kelompok 25-35, karena mereka memiliki responsivitas yang lebih baik
- ② Analisis lebih lanjut untuk memahami apa yang memotivasi kelompok 36-50 dan >50 agar lebih responsif. Misalnya, segmentasi lebih dalam berdasarkan preferensi, produk, atau waktu kunjungan.



## Handling Missing Value

Kolom **Income** memiliki baris  
kosong sebanyak **24**

Baris kosong akan di hapus karena  
jumlahnya lebih kurang dari 1%

Unnamed: 0	0	Unnamed: 0	0
ID	0	ID	0
Year_Birth	0	Year_Birth	0
Education	0	Education	0
Marital_Status	0	Marital_Status	0
Income	24	Income	0
Kidhome	0	Kidhome	0
Teenhome	0	Teenhome	0
Dt_Customer	0	Dt_Customer	0
Recency	0	Recency	0
MntCoke	0	MntCoke	0
MntFruits	0	MntFruits	0
MntMeatProducts	0	MntMeatProducts	0
MntFishProducts	0	MntFishProducts	0
MntSweetProducts	0	MntSweetProducts	0
MntGoldProds	0	MntGoldProds	0
NumDealsPurchases	0	NumDealsPurchases	0
NumWebPurchases	0	NumWebPurchases	0
NumCatalogPurchases	0	NumCatalogPurchases	0
NumStorePurchases	0	NumStorePurchases	0
NumWebVisitsMonth	0	NumWebVisitsMonth	0
AcceptedCmp3	0	AcceptedCmp3	0
AcceptedCmp4	0	AcceptedCmp4	0
AcceptedCmp5	0	AcceptedCmp5	0
AcceptedCmp1	0	AcceptedCmp1	0
AcceptedCmp2	0	AcceptedCmp2	0
Complain	0	Complain	0
Z_CostContact	0	Z_CostContact	0
Z_Revenue	0	Z_Revenue	0
Response	0	Response	0
.	.	.	.

## Handling Duplicate Value

```
data.duplicated().sum()
```

0

Tidak di temukan data yang duplikat

## Penambahan Kolom Baru

- Age
- Age\_Group
- Total\_Spending
- Jumlah\_Anak
- Conversion\_Rate
- Total\_Respons
- avg\_web\_visits
- online\_purchase\_ratio

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

## Value Range & Statistik Kolom

	Unnamed: 0	ID	Year_Birth	Income	Kidhome	Teenhome	Recency	MntCoke	MntFruits	MntMeatProducts	...	NumWebVisitsMonth
count	2216.000000	2216.000000	2216.000000	2.216000e+03	2216.000000	2216.000000	2216.000000	2.216000e+03	2216.000000	2.216000e+03	...	2216.000000
mean	1121.867329	5588.353339	1968.820397	5.224725e+07	0.441787	0.505415	49.012635	3.050916e+05	26356.046931	1.669959e+05	...	5.319043
std	643.164161	3249.376275	11.985554	2.517308e+07	0.536896	0.544181	28.948352	3.373279e+05	39793.916518	2.242833e+05	...	2.425359
min	0.000000	0.000000	1893.000000	1.730000e+06	0.000000	0.000000	0.000000	0.000000e+00	0.000000	0.000000e+00	...	0.000000
25%	566.750000	2814.750000	1959.000000	3.530300e+07	0.000000	0.000000	24.000000	2.400000e+04	2000.000000	1.600000e+04	...	3.000000
50%	1120.500000	5458.500000	1970.000000	5.138150e+07	0.000000	0.000000	49.000000	1.745000e+05	8000.000000	6.800000e+04	...	6.000000
75%	1678.250000	8421.750000	1977.000000	6.852200e+07	1.000000	1.000000	74.000000	5.050000e+05	33000.000000	2.322500e+05	...	7.000000
max	2239.000000	11191.000000	1996.000000	6.666660e+08	2.000000	2.000000	99.000000	1.493000e+06	199000.000000	1.725000e+06	...	20.000000

- Rata-rata Income adalah **5,224,725**
- Rata-rata **NumWebVisitMonth** adalah **5x** dan paling banyak **20x**

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



## Value Range & Statistik Kolom

	Education	Marital_Status	Dt_Customer
count	2216	2216	2216
unique	5	6	662
top	S1	Menikah	31-08-2012
freq	1116	857	12

- Jumlah customer pada perusahaan ini berjumlah **2240**
- Kolom Education berisikan 5 kategori yang di dominasi oleh **S1** dengan jumlah **1127**
- Kolom **Marital\_Status** didominasi dengan status **Menikah** sebanyak **864**

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

## Featured Transformation

Transformasi fitur adalah metode dalam machine learning dan statistik yang digunakan untuk mengubah atribut data agar lebih mudah dianalisis atau dimanfaatkan oleh model prediktif.

Standardisasi adalah proses transformasi fitur numerik sehingga memiliki mean = 0 dan standard deviation = 1, yang berguna untuk model machine learning yang sensitif terhadap skala data

	Income	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth	Age	Total_Spending	Jumlah_Anak	Conversion_Rate	Total_Respons	avg_web_visits	online_purchase_ratio
0	0.234063	0.978226	1.549429	1.690227	2.454568	1.484827	0.850031	1.428553	2.504712	0.554143	0.693232	0.986443	1.675488	1.581389	0.697514	0.439265	0.693232	0.046998
1	0.234559	0.872024	0.637328	0.717986	0.651038	0.633880	0.732867	1.125881	0.571082	1.169518	0.131574	1.236801	0.962358	0.632356	0.307283	0.439265	0.131574	0.037209
2	0.769478	0.358511	0.569159	0.178368	1.340208	0.146821	0.037937	1.428553	0.229327	1.291982	0.543978	0.318822	0.280250	1.581389	0.307283	0.439265	0.543978	0.096224
3	1.017239	0.872024	0.561922	0.655551	0.904892	0.585174	0.752171	0.760962	0.912837	0.554143	0.280829	1.256777	0.919224	0.632356	0.307283	0.439265	0.280829	0.020469
4	0.240221	0.391671	0.416348	0.218505	0.152766	0.000703	0.559135	0.333796	0.112428	0.061232	0.131574	1.016420	0.307044	0.632356	0.307283	0.439265	0.131574	0.123779

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

## One-hot Encoding

Encoding adalah proses mengubah variabel kategorikal menjadi representasi numerik sehingga bisa digunakan dalam model analisis dan pemodelan.

**One-Hot Encoding:** Menciptakan kolom baru untuk setiap kategori dan memberi nilai 1 atau 0.

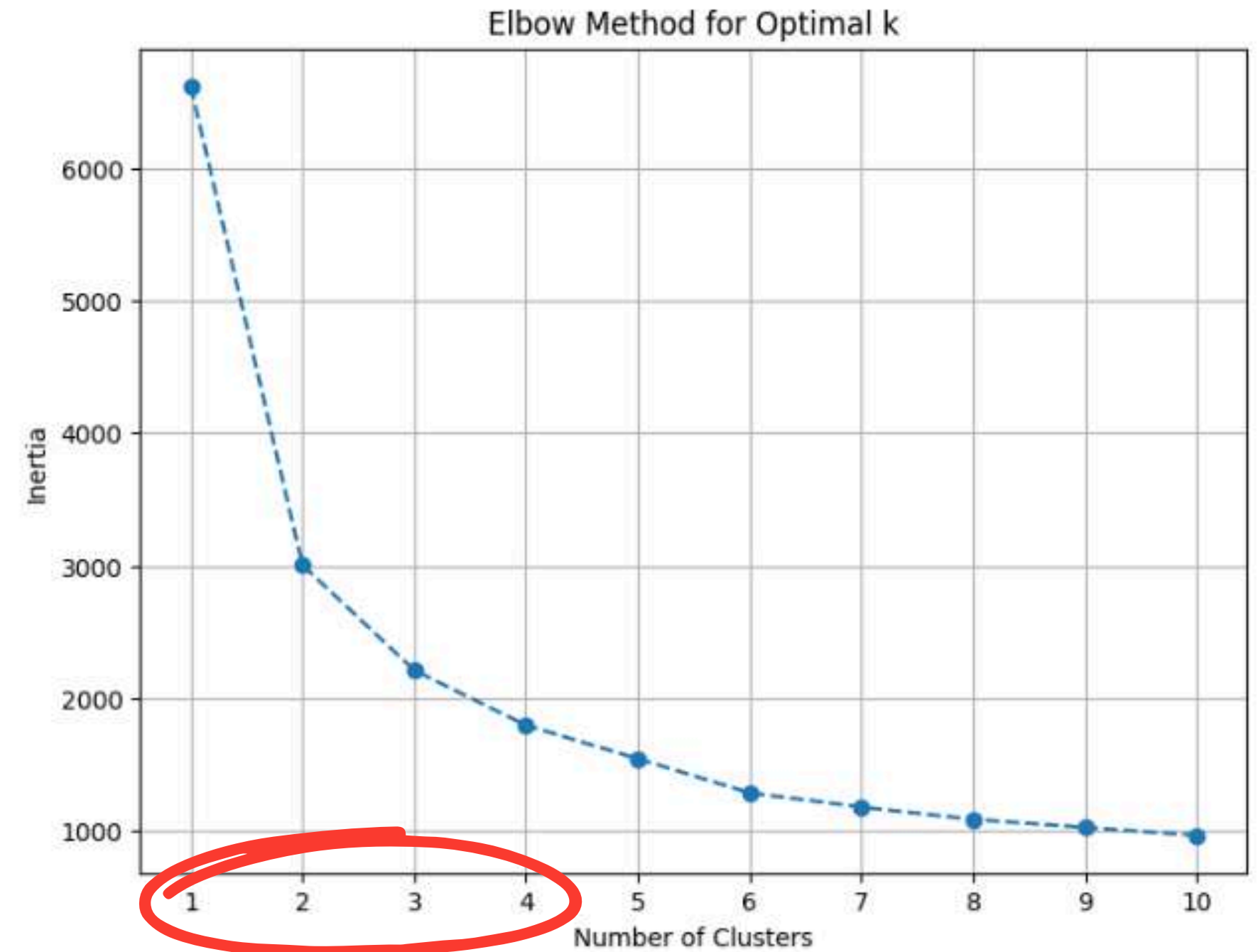
Pada analisis kali ini encoding di lakukan pada kolom  
**Age\_Group**

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

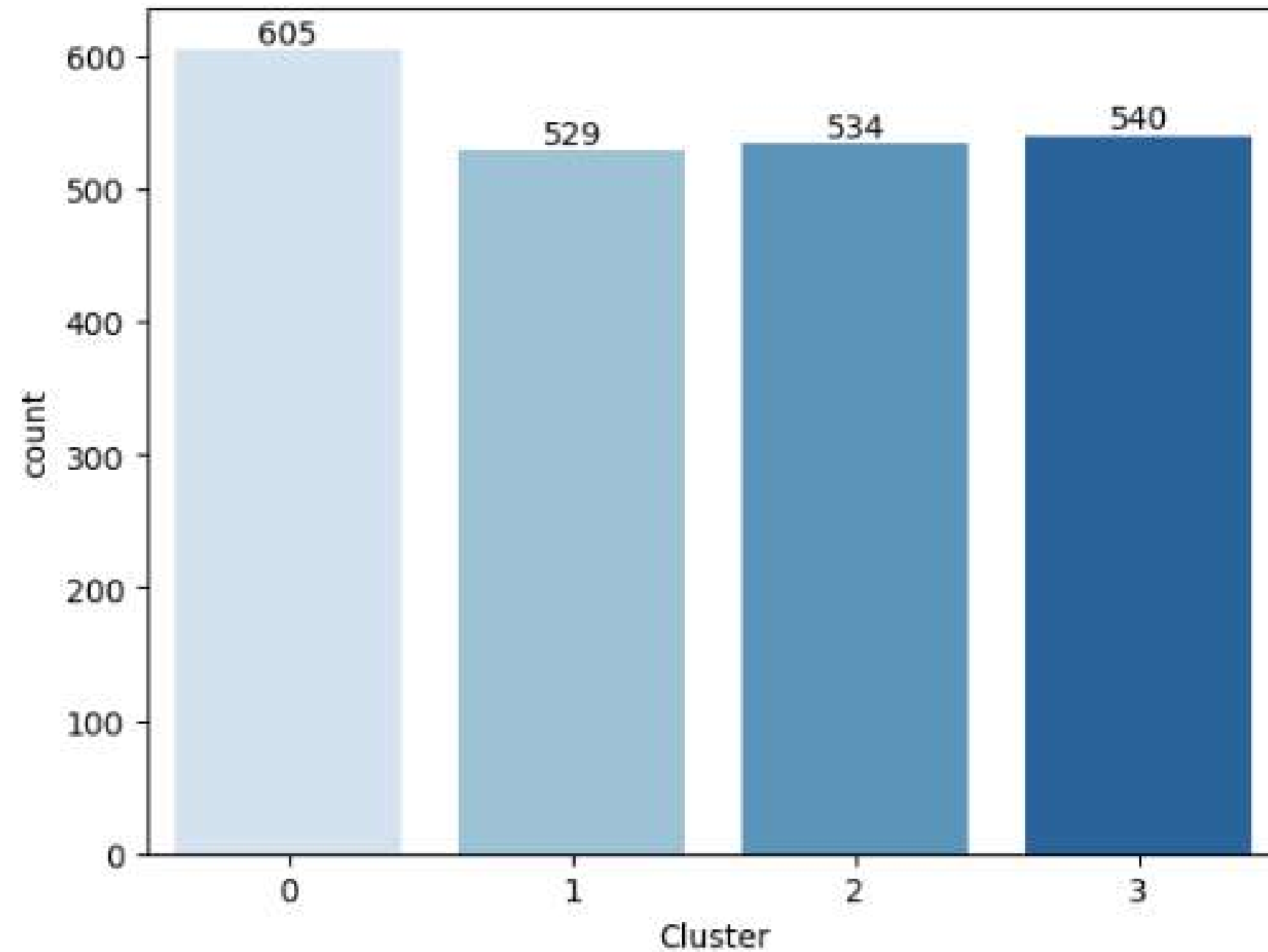


## Elbow Method

Berdasarkan grafik ini, **4 cluster** dapat dianggap sebagai jumlah cluster yang optimal untuk digunakan dalam analisis K-Means



Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



Jumlah pelanggan tiap Cluster:

1. Pada **Cluster 1** pelanggan berjumlah **605**
2. Pada **Cluster 2** pelanggan berjumlah **529**
3. Pada **Cluster 3** pelanggan berjumlah **534**
4. Pada **Cluster 4** pelanggan berjumlah **540**

## Cluster

	Income	age	total_spending	jumlah_anak	conversion_rate	total_respons	avg_web_visits	total_purchases
Cluster								
0	32313000.0	51.0	68000.0	1.0	0.0	0.0	0.666667	8.0
1	75922000.0	56.0	1305000.0	0.0	0.0	0.0	0.166667	20.0
2	60614000.0	57.0	841500.0	1.0	0.0	0.0	0.500000	22.0
3	41495500.0	54.0	78000.0	1.0	0.0	0.0	0.416667	8.0

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)



## Cluster

①

- Cluster ini terdiri dari individu dengan pendapatan yang relatif rendah dibandingkan dengan cluster lain, tetapi usia mereka berada di rentang menengah.
- Total pengeluaran juga rendah, dan conversion rate serta total respons menunjukkan bahwa cluster ini tidak aktif dalam melakukan pembelian.
- Rata-rata kunjungan ke situs web pun rendah, yang dapat berarti kurangnya minat atau keterlibatan dalam produk yang ditawarkan.

②

- Cluster ini memiliki pendapatan tertinggi di antara semua cluster, dan pengeluaran total juga signifikan (1,305,000) yang menunjukkan daya beli yang kuat.
- Usia lebih tua (56 tahun) dan tidak memiliki anak mungkin mengindikasikan bahwa mereka adalah individu yang lebih mandiri secara finansial.
- Meski menghabiskan banyak, angka conversion rate dan total respons menunjukkan bahwa mereka juga kurang terlibat dalam pembelian atau mungkin lebih selektif dalam pembelian.

## Cluster

③

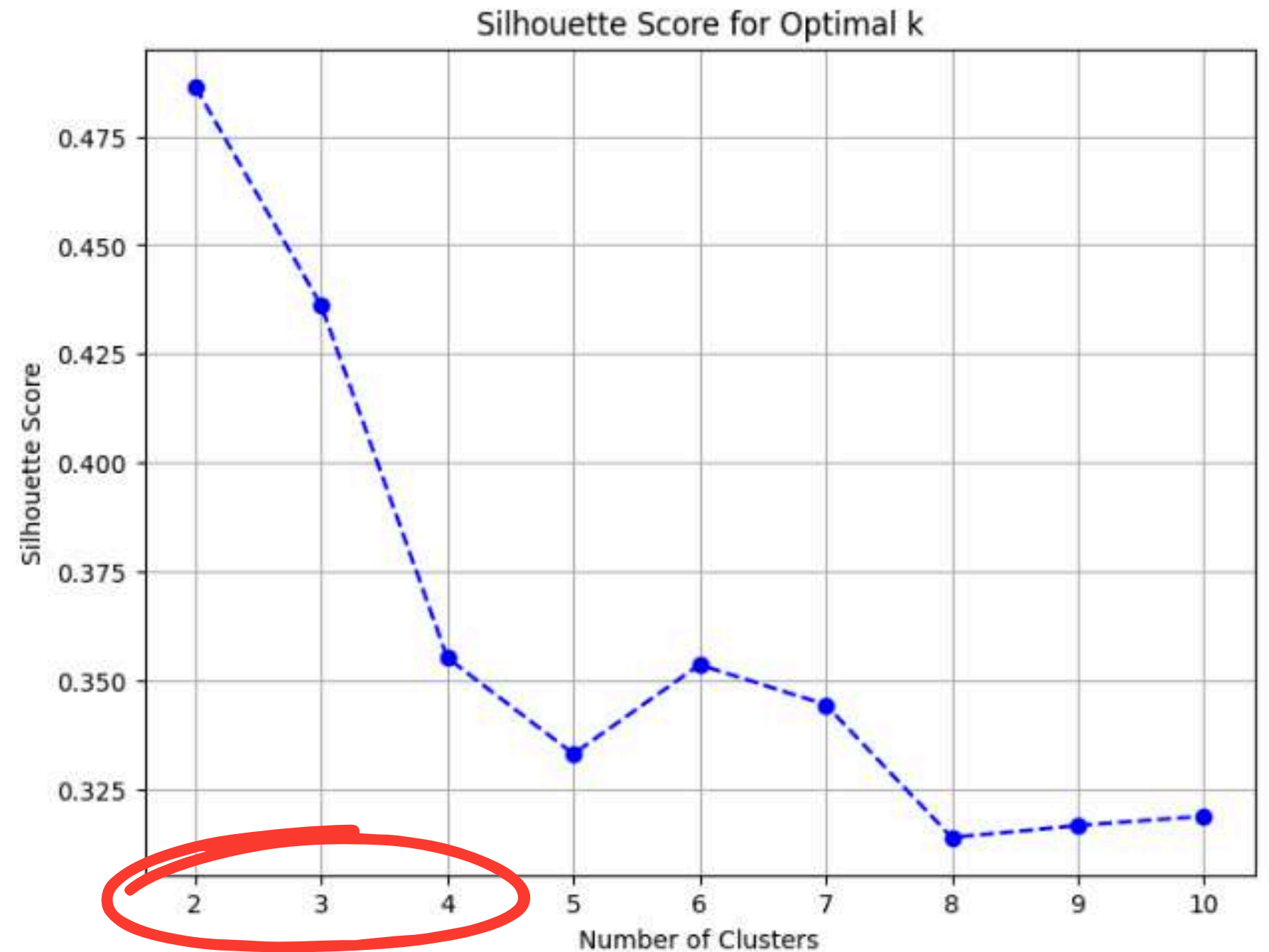
- Pendapatan dan usia mirip dengan Cluster 1, tetapi pengeluaran totalnya lebih rendah.
- Meskipun cluster ini memiliki sedikit lebih banyak total pembelian, conversion rate menunjukkan bahwa mereka juga mengalami kurangnya keterlibatan dalam membeli produk.
- Kehadiran satu anak dapat menunjukkan bahwa mereka adalah keluarga dengan pengeluaran yang dapat dipertimbangkan.

④

- Cluster ini menunjukkan pendapatan yang lebih rendah dibandingkan dengan cluster lainnya, dengan total pengeluaran dan total pembelian yang minimal.
- Usia menunjukkan bahwa mereka berada dalam rentang yang relatif lebih tua, tetapi mirip dengan Cluster 0 dalam hal keterlibatan dan pengeluaran.
- Seperti cluster lainnya, conversion rate 0% menunjukkan bahwa meskipun mereka melakukan beberapa kunjungan web, dukungan untuk pembelian tetap rendah.

## Silhouette Score

**4 cluster** dapat meningkatkan pemisahan antara segmen



Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

## Hasil Evaluasi Elbow Method & Silhouette Score

Silhouette Score: 0.3738156670879627
Inertia : 1784.1800300830537

- Silhouette Score bernilai 0,3738 menunjukkan bahwa model clustering memiliki performa yang cukup baik, tetapi ada ruang untuk perbaikan. Nilai tersebut menunjukkan bahwa beberapa titik data mungkin berada di batas antara cluster yang berbeda, yang berarti pemisahan antara cluster tidak sangat jelas
- Nilai Inertia sebesar 1784.1800 menunjukkan total jarak kuadrat antara setiap titik data dalam cluster dan centroid cluster tersebut.



## Segmentasi dan Marketing Retargeting

Cluster	Segment	Marketing_Strategy
0	Risk of Churn	Kirim email khusus dengan diskon besar untuk menghidupkan kembali minat.
1	High Spenders	Berikan penghargaan eksklusif, seperti poin loyalitas atau diskon VIP.
2	Mid Spenders	Dorong pembelian lebih besar dengan bundling produk atau diskon terbatas.
3	Low Spenders	Tawarkan diskon kecil untuk pembelian berikutnya atau promosi edukatif.

## Potential GMV Impact

GMV Impact (Gross Merchandise Value Impact) mengacu pada perkiraan peningkatan pendapatan yang dihasilkan dari serangkaian perilaku pelanggan, biasanya berdasarkan perubahan tingkat konversi.

**Total Potential GMV Impact: Rp 122,296,600.00**

Mengindikasikan adanya peluang yang besar untuk pertumbuhan pendapatan di seluruh segmen pelanggan yang dianalisis.

## Potential GMV Impact

Cluster	Expected New Conversions	Potential GMV Impact
0	60,5	4.114.000
1	52,9	69.034.500
2	53,4	44.936.100
3	54,0	4.212.000

Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

## Rekomendasi Bisnis

### 1. Kampanye Pemasaran Bertarget:

- Fokus pada **Cluster 1**: Kembangkan kampanye bernilai tinggi yang disesuaikan dengan preferensi mereka. Pemasaran email yang dipersonalisasi, program loyalitas, atau penawaran eksklusif dapat secara efektif meningkatkan pengeluaran.
- **Cluster 2**: Terapkan promosi spesifik yang sesuai dengan kebiasaan belanja Cluster 2 untuk mendorong tingkat konversi yang lebih tinggi.

### 2. Tingkatkan Keterlibatan Pelanggan:

- Untuk **Cluster 0 dan 3**, pertimbangkan strategi untuk meningkatkan keterlibatan:
  - Perkenalkan program loyalitas yang memberikan penghargaan untuk pembelian berulang.
  - Tawarkan diskon atau paket yang bertujuan untuk meningkatkan nilai pesanan rata-rata dan mengubah pelanggan yang acuh tak acuh menjadi pembeli.





**Thank You**  
**See You Soon**