

Predict Customer Personality to boost marketing campaign by using Machine Learning

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com



Created by:

Candraditya Dwaya Putra

Email: candradityaputra1@gmail.com

linkedin :

<https://www.linkedin.com/in/candraditya-dwaya-putra-77719a66>

“Have more than 11 years professional career in GIS and Management geodatabase. Good knowledge WebGIS Development, with strong experience Geospatial data in Forestry, environmental and regional planning. Skilled at data collection and analysis that elicits accurate and valuable information utilizing technical principles and theories. Technical proficiencies include SQL, Python, Machine Learning, MS Office, ESRI GIS, QGIS, HTML, CSS/Bootstrap, Javascript, geomorphology and Geospatial Software.”

Overview

“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 ”

Dataset

```
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 30 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   Unnamed: 0        2240 non-null    int64  
 1   ID               2240 non-null    int64  
 2   Year_Birth       2240 non-null    int64  
 3   Education        2240 non-null    object  
 4   Marital_Status   2240 non-null    object  
 5   Income            2216 non-null    float64 
 6   Kidhome          2240 non-null    int64  
 7   Teenhome         2240 non-null    int64  
 8   Dt_Customer      2240 non-null    object  
 9   Recency           2240 non-null    int64  
 10  MntCoke          2240 non-null    int64  
 11  MntFruits        2240 non-null    int64  
 12  MntMeatProducts  2240 non-null    int64  
 13  MntFishProducts  2240 non-null    int64  
 14  MntSweetProducts 2240 non-null    int64  
 15  MntGoldProds     2240 non-null    int64  
 16  NumDealsPurchases 2240 non-null    int64  
 17  NumWebPurchases  2240 non-null    int64  
 18  NumCatalogPurchases 2240 non-null    int64  
 19  NumStorePurchases 2240 non-null    int64  
 ...
 28  Z_Revenue         2240 non-null    int64  
 29  Response          2240 non-null    int64  
dtypes: float64(1), int64(26), object(3)
```

Description

Dataset mengandung kebiasaan customer melakukan transaksi dan interaksi di platform.

Shape

2240 Row, 30 Columns

Dtype

Float64 (1 features), int64(26 features), Object (3 features)

Missing Value

'income' merupakan feature yang memiliki missing value

Duplicate Data

0 Duplicate data

Untuk selengkapnya, dapat melihat jupyter notebook disini

<https://drive.google.com/file/d/1NYJ33LZXT7Ad1Hu2Cr2jW3broE8bknvl/view?usp=sharing>

Exploratory Data Analysis (EDA)

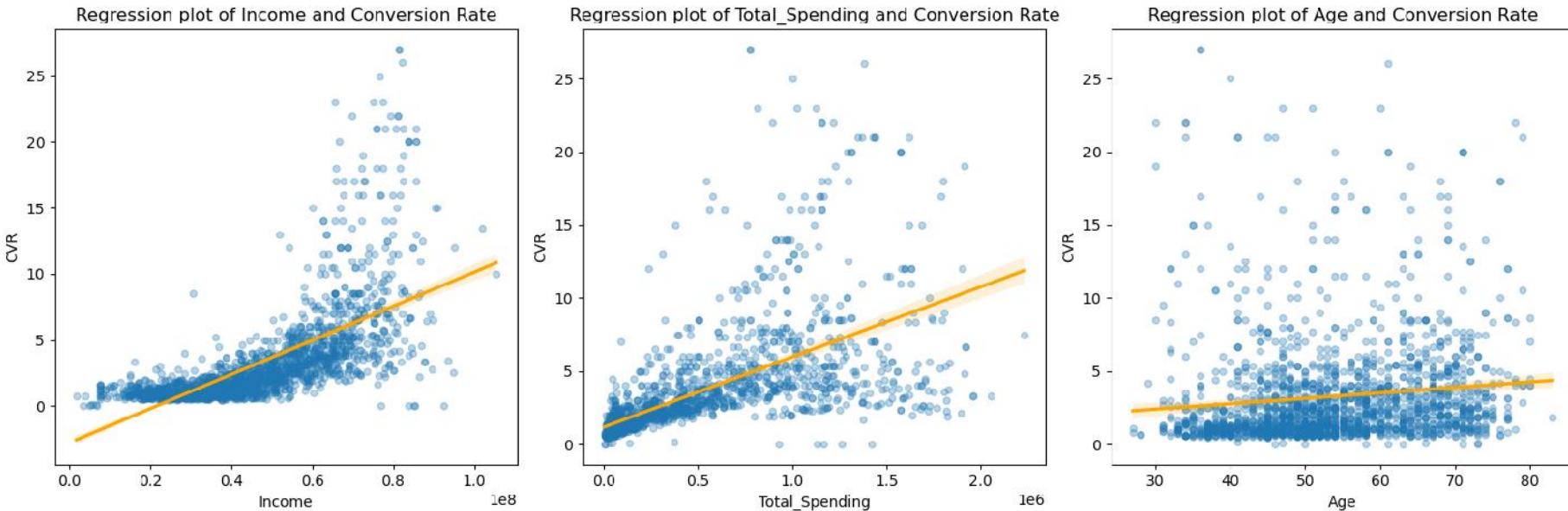
Features Extraction

- Membership Duration
- Customer Age
 - Age
 - Age Categories
- Total Children
- Total Transaction
- Total Spending
- Total Accepted Campaign
- Conversion Rate (CVR)

Correlation

- Income
- Total Spending
- CVR

Conversion Rate Analysis Based on Income, Spending and Age



- Secara keseluruhan pendapatan dan total spending memiliki korelasi positif terhadap konversi pembelian, Hal ini berarti semakin tinggi pendapatan dan total spending, kemungkinan konversi atau pembelian dari calon konsumen juga meningkat.
- Korelasi positif ini dapat disebabkan oleh beberapa faktor, seperti kemampuan finansial yang lebih baik untuk memenuhi kebutuhan atau keinginan pelanggan, serta persepsi nilai yang tinggi terhadap produk atau layanan yang ditawarkan. Dengan demikian, perusahaan dapat memperhatikan dan mengoptimalkan strategi pemasaran mereka dengan mempertimbangkan target audiens dengan pendapatan dan total spending yang lebih tinggi, sehingga meningkatkan peluang konversi dan keberhasilan kampanye pemasaran secara keseluruhan.
- Sementara itu, fitur usia cenderung tidak memiliki korelasi yang signifikan terhadap conversion rate. Ini berarti usia tidak menjadi faktor dominan yang mempengaruhi keputusan konsumen untuk melakukan konversi atau pembelian.

Untuk selengkapnya, dapat melihat jupyter notebook disini

<https://drive.google.com/file/d/1NYJ33LZXT7Ad1Hu2CrjW3broE8bknl/view?usp=sharing>

Data Cleaning & Preprocessing

	feature	data_type	null_values	null_percentage	unique_values	unique_sample
0	Unnamed: 0	int64	0	0.00	2240	[0, 1, 2, 3, 4]
1	ID	int64	0	0.00	2240	[5524, 2174, 4141, 6182, 5324]
2	Year_Birth	int64	0	0.00	59	[1957, 1954, 1965, 1984, 1981]
3	Education	object	0	0.00	5	[S1, S3, S2, SMA, D3]
4	Marital_Status	object	0	0.00	6	[Lajang, Bertunangan, Menikah, Cerai, Janda]
5	Income	float64	24	1.07	1974	[58138000.0, 46344000.0, 71613000.0, 26646000...]
6	Kidhome	int64	0	0.00	3	[0, 1, 2]
7	Teenhome	int64	0	0.00	3	[0, 1, 2]
8	Dt_Customer	object	0	0.00	663	[04-09-2012, 08-03-2014, 21-08-2013, 10-02-201...]
9	Recency	int64	0	0.00	100	[58, 38, 26, 94, 16]
10	MntCoke	int64	0	0.00	776	[635000, 11000, 426000, 173000, 520000]
11	MntFruits	int64	0	0.00	158	[88000, 1000, 49000, 4000, 43000]
12	MntMeatProducts	int64	0	0.00	558	[546000, 6000, 127000, 20000, 118000]
13	MntFishProducts	int64	0	0.00	182	[172000, 2000, 111000, 10000, 46000]
14	MntSweetProducts	int64	0	0.00	177	[88000, 1000, 21000, 3000, 27000]
15	MntGoldProds	int64	0	0.00	213	[88000, 6000, 42000, 5000, 15000]
16	NumDealsPurchases	int64	0	0.00	15	[3, 2, 1, 5, 4]
17	NumWebPurchases	int64	0	0.00	15	[8, 1, 2, 5, 6]
18	NumCatalogPurchases	int64	0	0.00	14	[10, 1, 2, 0, 3]

Sebelum dilakukan Handling

Fitur income memiliki missing value sebesar 1,07 % dan telah dilakukan handling dengan median (robust)

	feature	data_type	null_values	null_percentage	unique_values	unique_sample
0	Unnamed: 0	int64	0	0.000	1840	[1, 2, 3, 4, 5]
1	ID	int64	0	0.000	1840	[2174, 4141, 6182, 5324, 7446]
2	Year_Birth	int64	0	0.000	55	[1954, 1965, 1984, 1981, 1967]
3	Education	object	0	0.000	5	[S1, S3, S2, SMA, D3]
4	Marital_Status	object	0	0.000	6	[Lajang, Bertunangan, Menikah, Cerai, Janda]
5	Income	float64	0	0.000	1631	[46344000.0, 71613000.0, 26646000.0, 58293000...]
6	Kidhome	int64	0	0.000	3	[1, 0, 2]
7	Teenhome	int64	0	0.000	3	[1, 0, 2]
8	Dt_Customer	datetime64[ns]	0	0.000	644	[2014-08-03T00:00:00.000000000, 2013-08-21T00:...]
9	Recency	int64	0	0.000	100	[38, 26, 94, 16, 34]
10	MntCoke	int64	0	0.000	647	[11000, 426000, 173000, 520000, 235000]
11	MntFruits	int64	0	0.000	130	[1000, 49000, 4000, 43000, 42000]
12	MntMeatProducts	int64	0	0.000	367	[6000, 127000, 20000, 118000, 98000]
13	MntFishProducts	int64	0	0.000	144	[2000, 111000, 10000, 46000, 0]
14	MntSweetProducts	int64	0	0.000	79	[1000, 21000, 3000, 27000, 42000]
15	MntGoldProds	int64	0	0.000	184	[6000, 42000, 5000, 15000, 14000]
16	NumDealsPurchases	int64	0	0.000	15	[2, 1, 5, 4]
17	NumWebPurchases	int64	0	0.000	12	[1, 8, 2, 5, 6]
18	NumCatalogPurchases	int64	0	0.000	11	[1, 2, 0, 3]

Setelah dilakukan Handling

Untuk selengkapnya, dapat melihat jupyter notebook disini
<https://drive.google.com/file/d/1NYJ33LZXT7Ad1Hu2Cr2jW3broE8bknvl/view?usp=sharing>

Feature Encoding

- Melakukan encoding untuk fitur-fitur kategorikal
- Education
 - SMA : 0
 - D3 : 1
 - S1 : 2
 - S2 : 3
 - S3 : 4

```
# education
mapping_education = {'SMA' : 0,
                     'D3' : 1,
                     'S1' : 2,
                     'S2' : 3,
                     'S3' : 4}

df_prep['education_encl'] = df_prep['Education'].map(mapping_education)
```

Feature Encoding

Untuk selengkapnya, dapat melihat jupyter notebook disini
<https://drive.google.com/file/d/1NYJ33LZXT7Ad1Hu2Cr2jW3broE8bknvl/view?usp=sharing>

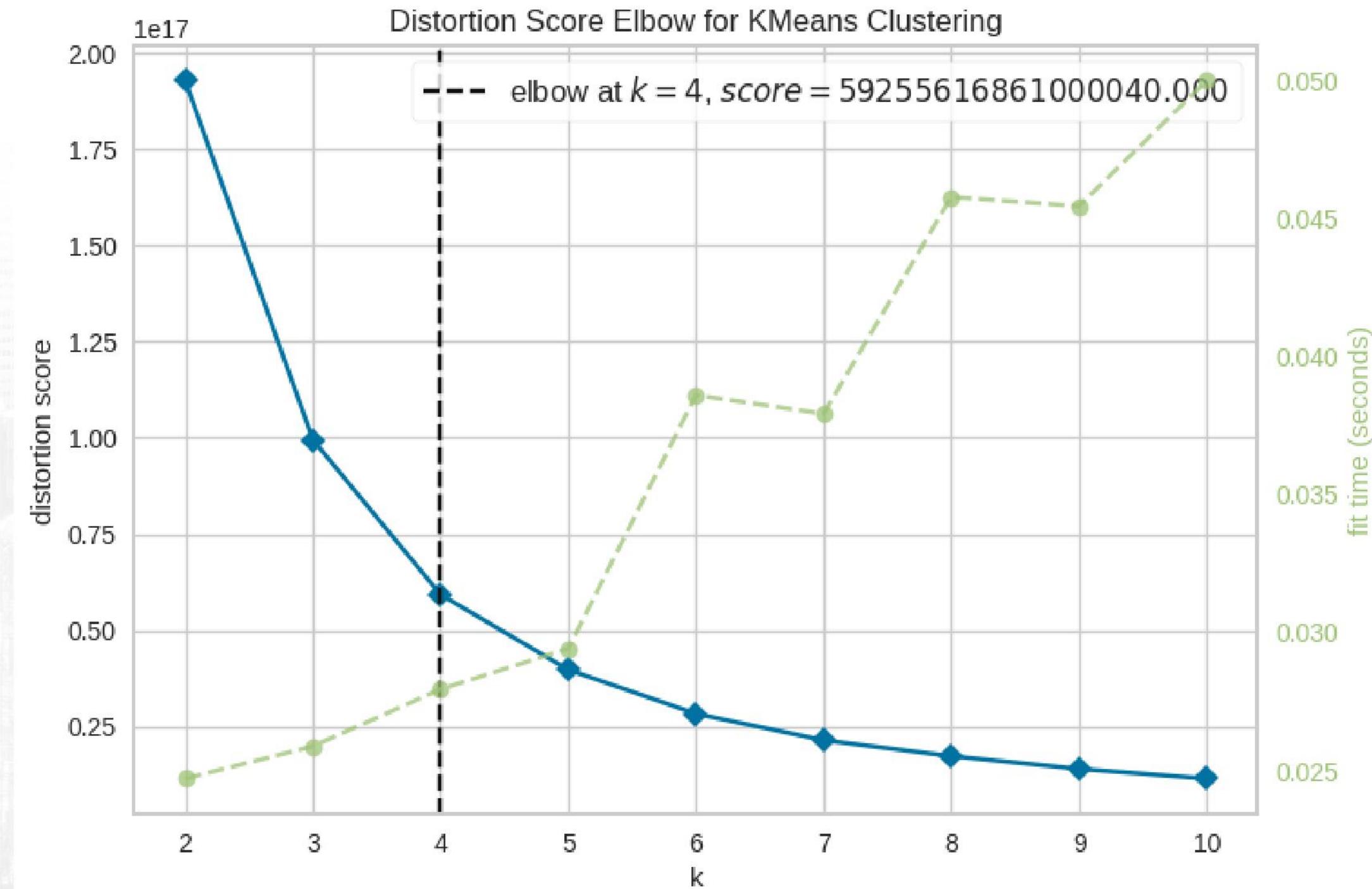
Data Cleaning & Preprocessing

	Income	Kidhome	Teenhome	Recency	MntCoke	MntFruits	MntMeatProducts	MntFishProd	MntSweetProducts	MntGoldProds
0	-0.032256	0.876749	0.809803	-0.365148	-0.760019	-0.550891		-0.733803		-0.000100
1	1.321897	-0.950225	-1.022962	-0.780333	0.551282	1.078317		0.235689		0.000100
2	-1.087862	0.876749	-1.022962	-0.780333	-0.760019	-0.449065		-0.621630		-0.000100

3 rows × 32 columns

Fitur income memiliki missing value sebesar 1,07 % dan telah dilakukan handling dengan median (robust)

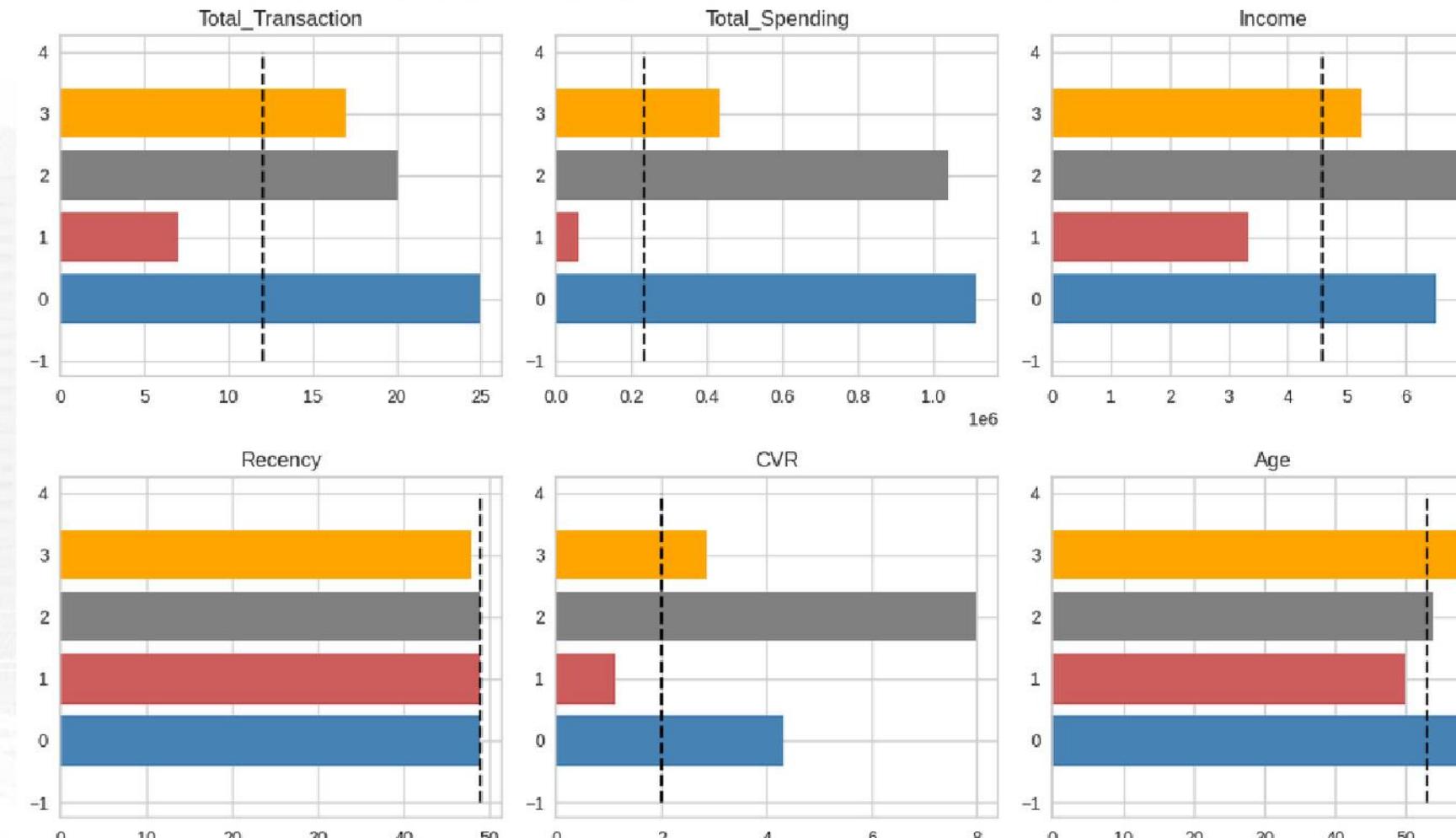
Data Modeling



Berdasarkan Distortion Score dan Elbow method didapatkan jumlah cluster terbaik adalah 4

Untuk selengkapnya, dapat melihat jupyter notebook disini
<https://drive.google.com/file/d/1xkORHL6pIWcoxo28WeLA9zpAw3hUeXx/view?usp=sharing>

Customer Personality Analysis for Marketing Retargeting



Penjelasan:

- Berdasarkan hasil clustering dapat diketahui karakteristik dari setiap segment cluster

Cluster 0

- Angka transaksi dan spending tertinggi yaitu mayoritas 25 transaksi dan Rp.1.116.000 perbulan
- Pendapatan cukup tinggi, mayoritas Rp.65.215.000/tahun
- Conversion rate sedang, yaitu 4%
- Mayoritas usia 60 tahun

Cluster 1

- Angka transaksi dan spending terendah yaitu mayoritas hanya 7 transaksi dan Rp.58.000 perbulan
- Pendapatan terendah, mayoritas Rp.33.297.500/tahun
- Conversion terendah, yaitu 1%
- Mayoritas usia 50 tahun

Cluster 2

- Angka transaksi dan spending cukup tinggi yaitu mayoritas 20 transaksi dan Rp.1.040.000 perbulan
- Pendapatan tertinggi, mayoritas Rp.71.488.000/tahun
- Conversion rate tertinggi, yaitu 8%
- Mayoritas usia 54 tahun

Cluster 3

- Angka transaksi dan spending sedang yaitu mayoritas 17 transaksi dan Rp.434.000 perbulan
- Pendapatan cukup sedang, mayoritas Rp.52.597.000/tahun
- Conversion rate cukup sedang, yaitu 3%
- Mayoritas usia 58 tahun

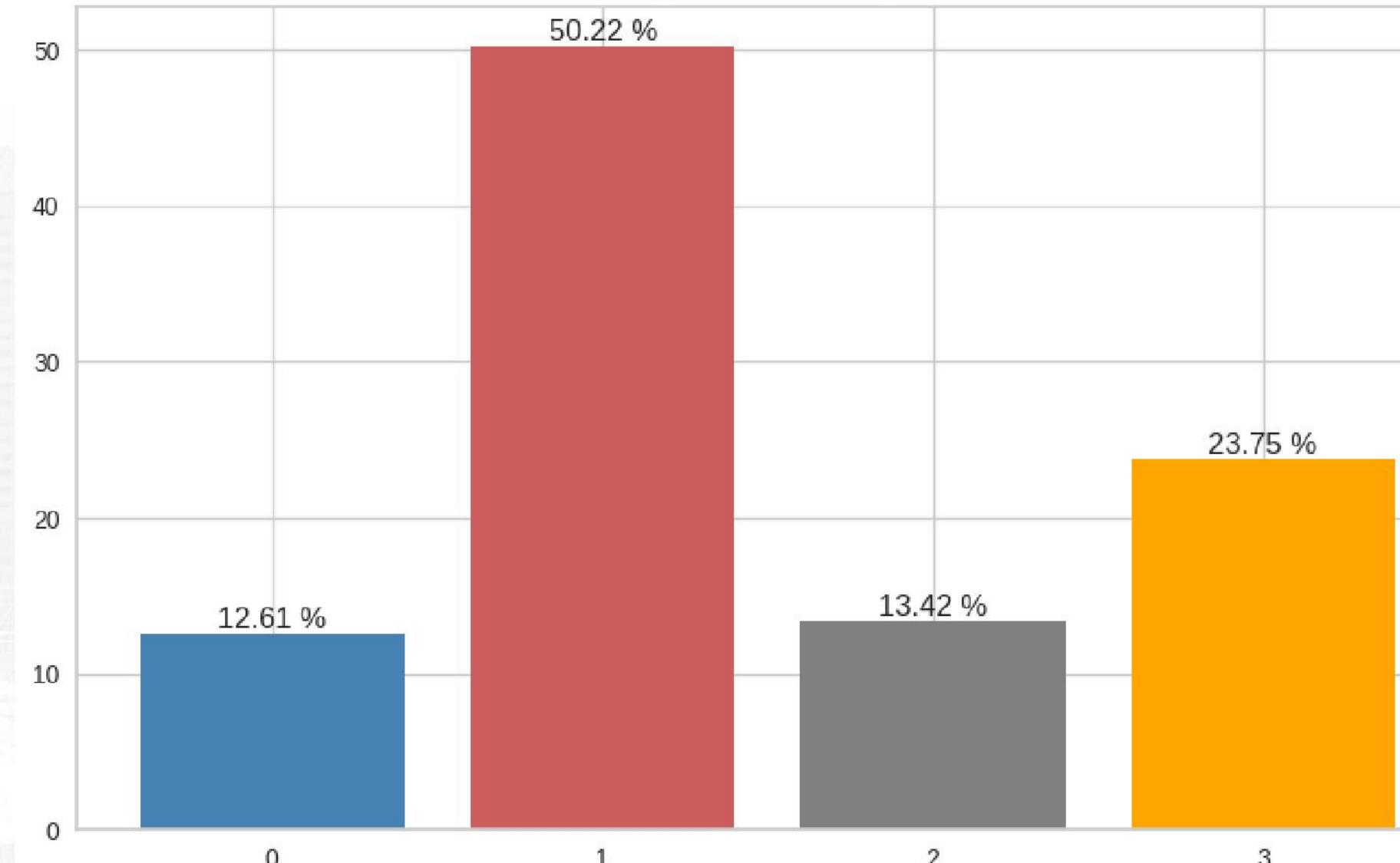
Dari karakteristik tersebut, dapat setiap cluster dapat dikelompok berdasarkan kategori, yaitu:

- Cluster 0: "High-Transaction High-Spending Group" - High Customer A
- Cluster 1: "Low-Transaction Low-Spending Group" - Low Customer
- Cluster 2: "High-Income High-Conversion Group" - High Customer B
- Cluster 3: "Moderate-Transaction Moderate-Spending Group" - Moderate Customer

Untuk selengkapnya, dapat melihat jupyter notebook disini

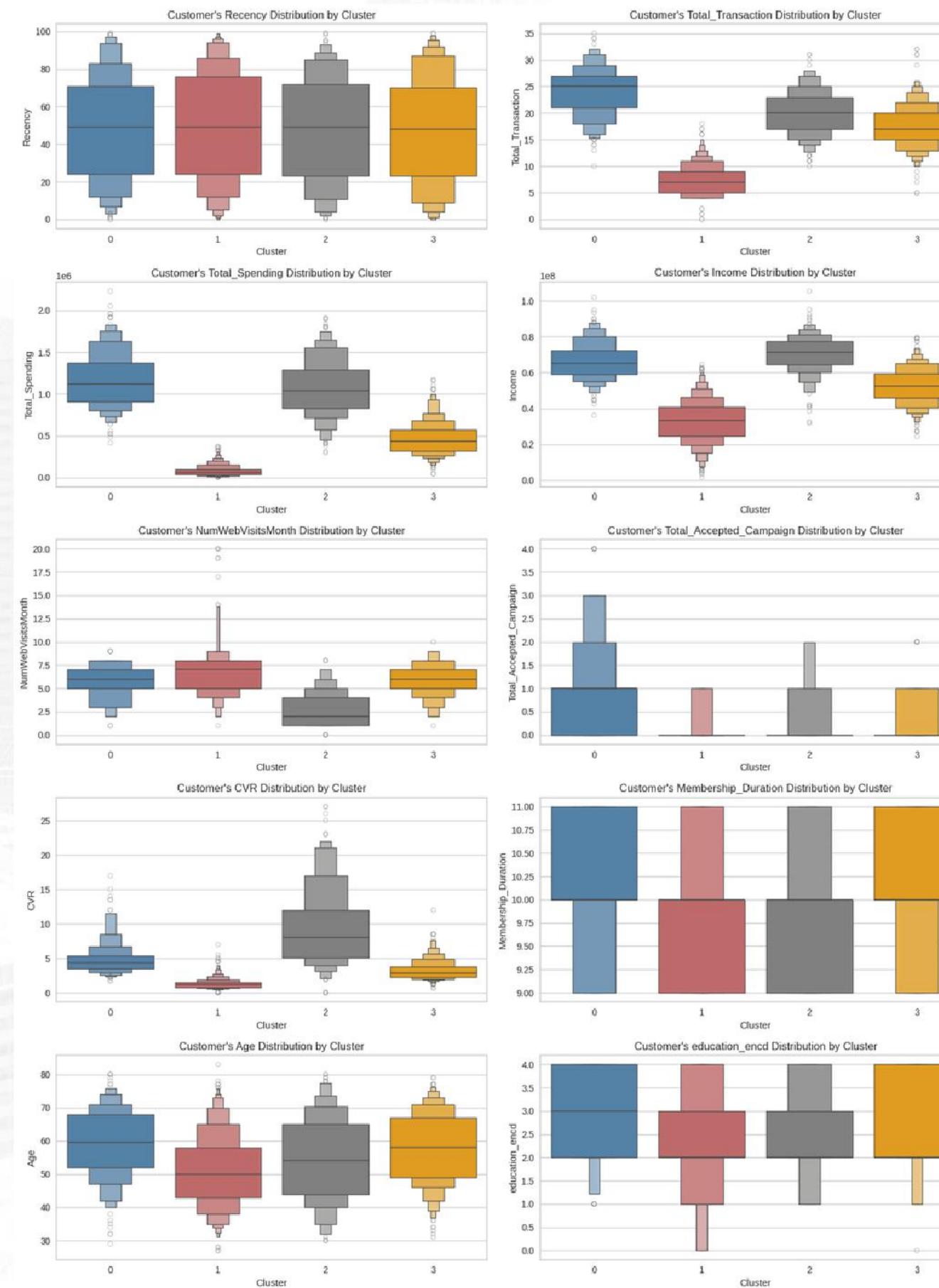
Customer Personality Analysis for Marketing Retargeting

Percentage of Customer by Cluster



- Berdasarkan hasil analisis, ditemukan bahwa sekitar 50% dari populasi termasuk dalam kelompok Low Customer. Meskipun kelompok ini memiliki angka transaksi dan pengeluaran yang rendah, namun karena populasi mereka yang besar, perusahaan dapat fokus untuk menarik perhatian mereka.
- Sedangkan populasi High Customer A dan B cenderung rendah, namun memiliki potensi transaksi, spending, konversi, dan income yang tinggi. Perusahaan dapat mempertimbangkan strategi pemasaran yang lebih personal dan eksklusif untuk menarik minat mereka.

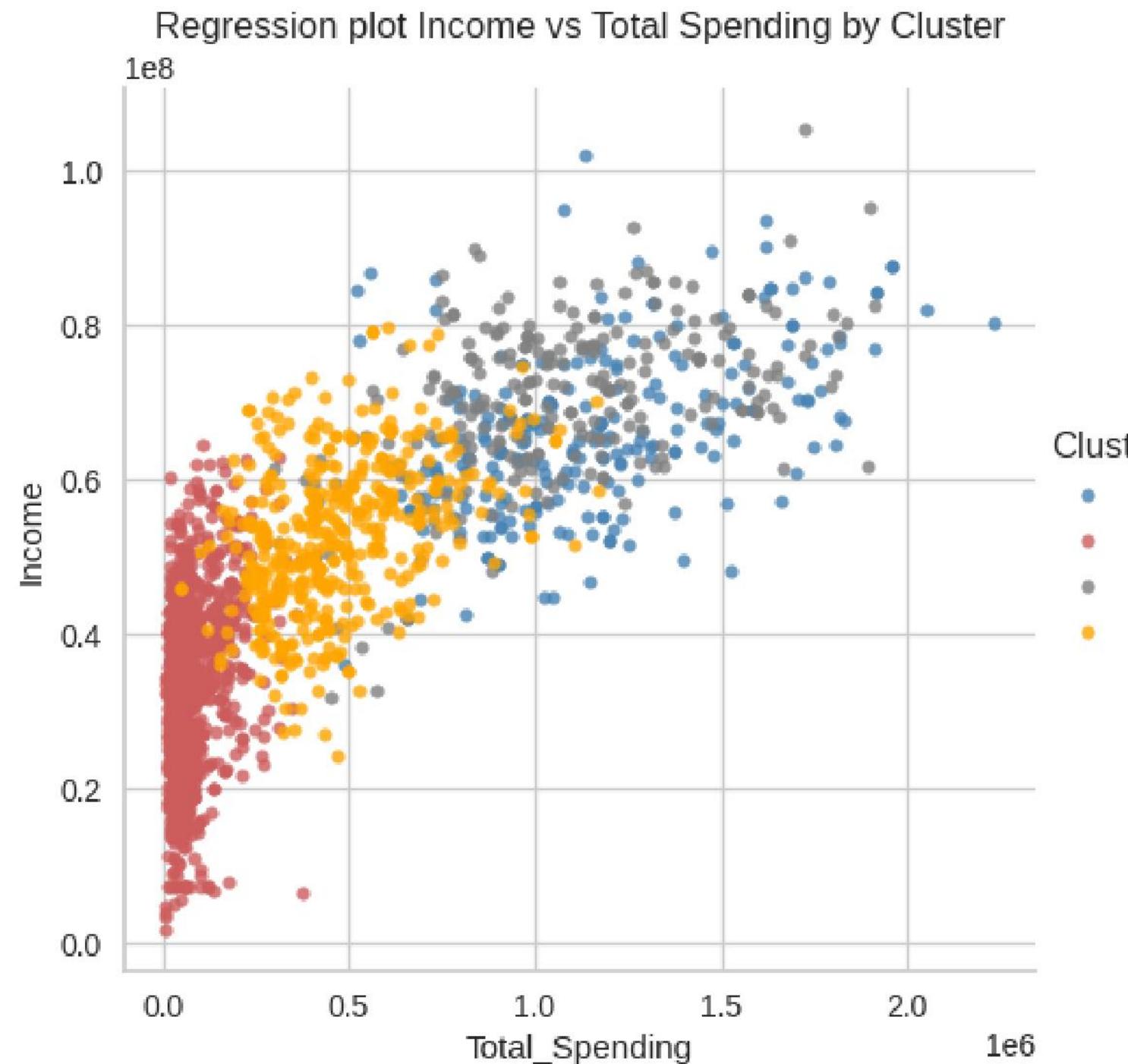
Customer Personality Analysis for Marketing Retargeting



- Hal yang menarik adalah adanya kategori Low Customer atau Cluster 1 yang memiliki distribusi jumlah kunjungan website yang tinggi, namun memiliki total acceptance campaign yang rendah. Ini menunjukkan bahwa kelompok ini sangat sering mengunjungi website perusahaan, tetapi tidak sepenuhnya menyadari atau tidak responsif terhadap campaign yang ditawarkan. Mengingat kelompok ini memiliki populasi yang paling banyak, perusahaan perlu mengembangkan strategi yang tepat untuk menarik perhatian dan meningkatkan keterlibatan mereka. Mungkin perlu dilakukan analisis lebih lanjut untuk memahami alasan di balik rendahnya acceptance campaign dalam kelompok ini.
- Cluster yang paling banyak merespon campaign adalah High Customer A atau Cluster 0 dengan tingkat konversi yang sedang. Ini menunjukkan bahwa mayoritas pelanggan dalam kelompok ini sangat responsif terhadap campaign yang ditawarkan oleh perusahaan. Hal ini dapat menjadi kesempatan yang baik untuk meningkatkan interaksi dan pembelian dari kelompok ini dengan meluncurkan campaign yang lebih menarik dan relevan sesuai dengan preferensi mereka.
- Sementara itu, High Spender B atau Cluster 2 menarik perhatian karena mayoritas pelanggannya tidak terlalu sering mengunjungi website perusahaan, namun memiliki distribusi konversi rate yang lebih tinggi dengan respon campaign yang tidak terlalu tinggi. Fenomena ini menunjukkan bahwa kelompok ini memiliki kecenderungan pengeluaran yang tinggi dan cenderung merespons positif terhadap kampanye yang ditawarkan, meskipun mereka tidak begitu aktif dalam kunjungan ke website. Perusahaan dapat memanfaatkan informasi ini dengan mengoptimalkan saluran komunikasi lain seperti email, media sosial, atau platform online lainnya untuk efektif menjangkau kelompok ini.

Untuk selengkapnya, dapat melihat jupyter notebook disini
<https://drive.google.com/file/d/1xkORHL6plWcoxco28WeLA9zpAw3hUeXx/view?usp=sharing>

Customer Personality Analysis for Marketing Retargeting



- Berdasarkan plot korelasi antara pendapatan (Income) dan total pengeluaran (Total Spending), terlihat bahwa terbentuknya cluster atau kelompok yang dapat dibedakan. Terlihat bahwa High cluster 0 dan 3 cenderung berada dalam satu kelompok, yaitu dalam kategori high customer. Hal ini menunjukkan adanya hubungan positif antara tingkat pendapatan dan pengeluaran pelanggan dalam kelompok tersebut. Dengan mengetahui pola korelasi ini, perusahaan dapat mengoptimalkan strategi pemasaran dan penawaran produk mereka.

Recommendation

Berdasarkan analisis yang telah dilakukan, kita dapat mengidentifikasi personalitas atau karakteristik pelanggan berdasarkan kluster yang terbentuk. Mengetahui personalitas ini sangat berharga dalam merancang strategi pemasaran yang lebih efektif. Dengan memahami preferensi, kebutuhan, dan perilaku konsumen dalam setiap kluster, perusahaan dapat menghasilkan kampanye yang lebih relevan dan menarik bagi setiap kelompok pelanggan.

High Customer A

Personaliti:

- Populasi 12.61%.
- Cluster 0 yang merupakan pelanggan yang memiliki total taransaksi dan total spending yang tinggi tertinggi.
- Paling responsif terhadap campaign, dengan konversi ke pembelian sedang.
- Tingkat kunjungan website sedang.

Rekomendasi:

- Mengingat kelompok High Customer A cenderung memiliki total transaksi dan total spending yang tinggi, perusahaan dapat memberikan penawaran khusus dan insentif tambahan untuk mendorong pelanggan melakukan pembelian secara terus-menerus. Perusahaan dapat menerapkan program diskon eksklusif, hadiah loyalitas, atau akses ke produk atau layanan khusus untuk kelompok ini.
- Perusahaan dapat meningkatkan kualitas pengalaman pengguna dalam berselancar di website, mengingat tingkat kunjungan website yang sedang. Perusahaan dapat memastikan tampilan yang menarik, customer journey yang efisien, dan lain sebagainya.
- Mengingat kelompok High Customer A sangat responsif terhadap campaign, memanfaatkan kepuasan mereka dengan memperkenalkan program referral dapat menjadi strategi yang efektif. Memberikan insentif kepada pelanggan untuk merekomendasikan produk atau layanan perusahaan kepada teman dan keluarga dapat membantu dalam memperluas jangkauan dan memperoleh pelanggan baru.
- Perusahaan dapat mengirimkan pesan yang dipersonalisasi seperti info promo atau diskon berdasarkan preferensi kelompok ini. Hal ini dilakukan untuk menjaga loyalitas pelanggan.

High Customer B

Personaliti :

- Populasi 13.42%.
- Sama seperti Cluster 0, namun kategori ini termasuk Cluster 2 yang merupakan pelanggan yang memiliki income paling tinggi.
- Tingkat konversi paling tinggi, respon terhadap campaign relatif sedang.
- Kurang mengunjungi website secara aktif.

Rekomendasi:

- Sama halnya dengan High Customer A, perusahaan dapat memberikan penawaran khusus seperti diskon, program loyalti, dan sebagainya agar pelanggan selalu tertarik untuk berbelanja terus menerus.
- Mengingat kelompok ini kurang aktif dalam kunjungan website, perusahaan dapat memanfaatkan saluran komunikasi alternatif seperti email, pesan teks, atau media sosial. Hal ini dapat membantu meningkatkan interaksi dan kesadaran pelanggan.
- Untuk meningkatkan respon pelanggan terhadap campaign, perusahaan dapat memberikan campaign-campaign yang tertarget sesuai dengan preferensi dan kebutuhan pelanggan.
- Mengingat pelanggan dalam kelompok High Customer B memiliki tingkat konversi yang tinggi, perusahaan dapat mempertimbangkan untuk meluncurkan program loyalitas yang memberikan insentif tambahan, penghargaan khusus, atau akses ke acara atau produk eksklusif dapat memperkuat loyalitas pelanggan.

Moderate Customer

Personaliti :

- Populasi 23.75%.
- Termasuk kedalam cluster 4 yang memiliki total transaksi, total spending, dan income sedang atau rata-rata.
- Tingkat konversi, kunjungan website dan respon terhadap campaign relatif sedang.

Rekomendasi:

- Perusahaan dapat memberikan penawaran khusus dan diskon untuk mendorong pembelian lebih lanjut. Hal ini dapat memberikan insentif tambahan kepada pelanggan dalam kelompok ini untuk memilih produk atau layanan perusahaan dibandingkan dengan pesaing.
- Perusahaan dapat mengirim pesan yang relevan dan menarik kepada pelanggan.
- Memastikan pengalaman pengguna yang baik saat mengunjungi website atau berinteraksi dengan produk atau layanan perusahaan.
- Membangun program hadiah atau loyalitas dapat membantu memperkuat keterikatan pelanggan. Seperti dengan memberikan poin, penghargaan, atau manfaat khusus kepada pelanggan setia, perusahaan dapat mendorong mereka untuk terus memilih dan membeli produk atau layanan perusahaan.

Low Customer

Personaliti :

- Populasi 50.22%, didominasi oleh customer kategori ini.
- Termasuk kedalam cluster 1 yang memiliki total transaksi, total spending, dan income rendah.
- Tingkat konversi paling rendah, cenderung tidak merespon campaign.
- Namun kategori ini paling sering mengunjungi website.

Rekomendasi:

- Mengingat kelompok Low Customer sering mengunjungi website, perusahaan dapat memanfaatkan informasi kunjungan website untuk menyajikan konten yang personalisasi dan penawaran khusus yang sesuai dengan minat dan preferensi mereka.
- Perusahaan dapat melakukan retargeting campaign dengan mengingatkan pelanggan dalam kelompok ini tentang produk atau layanan yang mereka telah kunjungi di website. Dengan menampilkan iklan yang disesuaikan di berbagai platform digital yang mereka gunakan, perusahaan dapat membangun kesadaran dan mendorong mereka untuk melanjutkan proses pembelian.
- Mengingat kelompok Low Customer memiliki tingkat konversi yang rendah dan cenderung tidak merespon campaign dengan baik, perusahaan dapat menggunakan strategi konten yang lebih fokus pada edukasi dan informasi. Memberikan konten yang memberikan nilai tambah, memberikan solusi untuk masalah atau kebutuhan pelanggan, dan membantu mereka membuat keputusan yang lebih informasi dapat meningkatkan keterlibatan dan kepercayaan pelanggan dalam kelompok ini.