

Predict Customer Personality to boost marketing campaign by using Machine Learning

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



Created by:

Devia Febyanti

devia.febyanti99@gmail.com

<https://www.linkedin.com/in/devia-febyanti/>

Experienced data operator with proven work history in the field of education. Currently motivated to become a Data Scientist to help driving financial inclusion through the use of big data and machine learning.



Bachelor of Geophysics
2018-2022

“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 ”

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
20	NumWebVisitsMonth	2240 non-null	int64
21	AcceptedCmp3	2240 non-null	int64
22	AcceptedCmp4	2240 non-null	int64
23	AcceptedCmp5	2240 non-null	int64
24	AcceptedCmp1	2240 non-null	int64
25	AcceptedCmp2	2240 non-null	int64
26	Complain	2240 non-null	int64
27	Z_CostContact	2240 non-null	int64
28	Z_Revenue	2240 non-null	int64
29	Response	2240 non-null	int64

dtypes: float64(1), int64(26), object(3)

Shape

- 2240 data rows
- 30 features

Data Type

- Float64 : 1 feature
- Int64 : 26 features
- Object : 3 features

Missing Value

Feature: **Income**

Duplicate

None

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

FEATURE EXTRACTION / ENGINEERING

Menghitung total purchase

Menghitung jumlah anak

Menghitung total accepted
campaign

Menghitung total hari
setelah bergabung

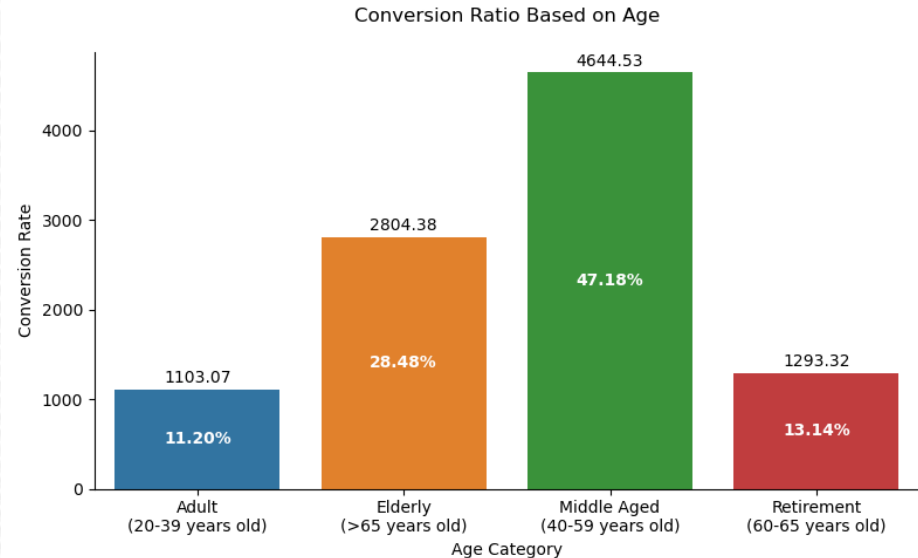
Menghitung total spent

Menghitung umur dan
kategori usia

Total_Purchases	Total_AcceptedCmp	Total_Spent	Total_Children	Datetime	Total_Days_Customer	conversion_rate	Age	Age_Category
9	0	39000	1	2024-07-11	3913	1.500000	53	Middle Aged
10	0	198000	1	2024-07-11	3757	5.000000	70	Elderly
22	0	716000	1	2024-07-11	3958	3.666667	57	Middle Aged
16	0	858000	0	2024-07-11	4015	8.000000	57	Middle Aged
29	0	1693000	0	2024-07-11	4352	9.666667	60	Retirement

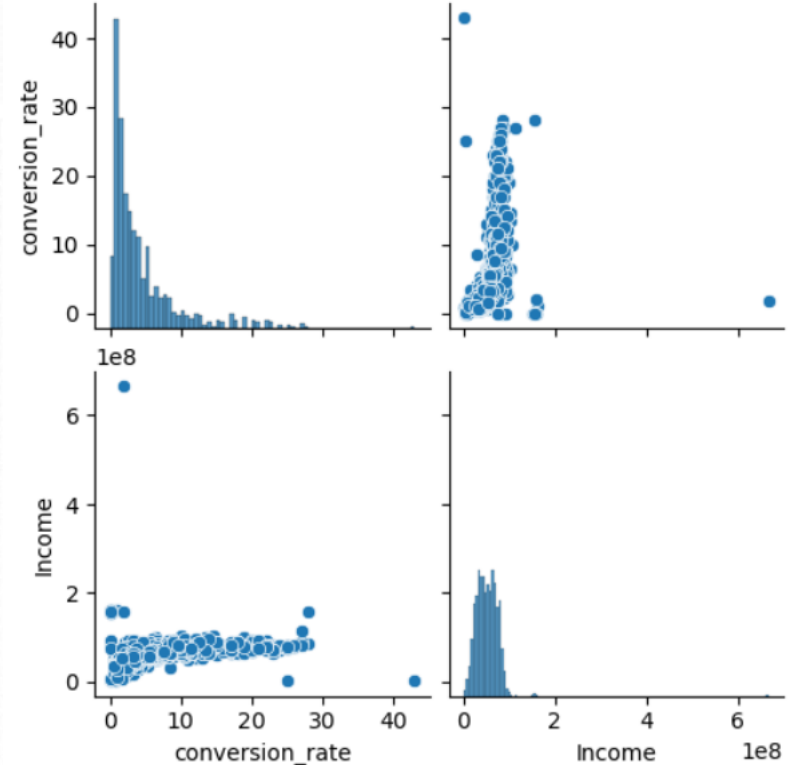
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

- Pelanggan dalam kategori usia Dewasa (Adult) memiliki tingkat konversi sebesar 1103.072264. Kelompok Dewasa menyumbang sekitar 11.20%. Ini menunjukkan bahwa meskipun kelompok ini cukup signifikan, mereka bukanlah kelompok usia dengan kontribusi tertinggi dalam hal tingkat konversi.
- Pelanggan dalam kategori usia Paruh Baya (Middle Aged) memiliki tingkat konversi sebesar 4644.532665. Mereka menyumbang 47.18% dari total tingkat konversi, menjadikan kelompok ini sebagai kontributor terbesar. Ini menunjukkan bahwa pelanggan paruh baya sangat responsif terhadap campaign pemasaran atau memiliki preferensi yang cocok dengan produk atau layanan yang ditawarkan.
- Pelanggan dalam kategori usia Pensiun (Retirement) memiliki tingkat konversi sebesar 1293.317251. Mereka menyumbang 13.14% dari total tingkat konversi. Meskipun kontribusi mereka lebih kecil dibandingkan kelompok Lansia dan Paruh Baya, mereka masih memberikan kontribusi yang signifikan terhadap total tingkat konversi.
- Pelanggan dalam kategori usia Lansia (Elderly) memiliki tingkat konversi sebesar 2804.375794. Mereka menyumbang 28.48% dari total tingkat konversi. Ini menunjukkan bahwa kelompok Lansia memiliki dampak yang signifikan terhadap tingkat konversi, lebih dari seperempat dari keseluruhan kontribusi.



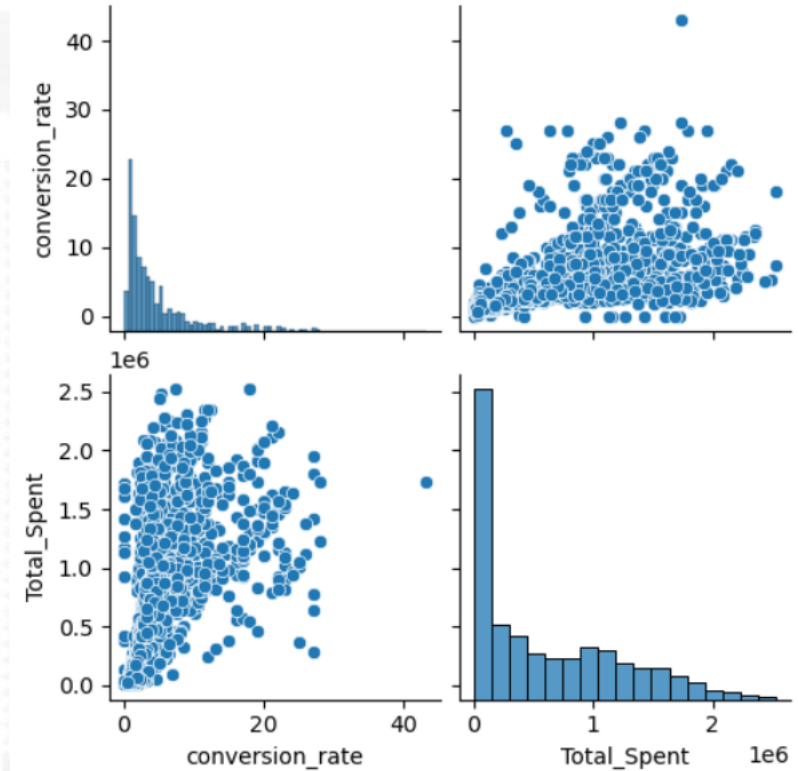
Conversion Rate Analysis Based on **Income**, Spending and Age

Pelanggan dengan pendapatan yang lebih tinggi cenderung memiliki tingkat konversi yang lebih tinggi. Tingkat konversi (conversion rate) mengukur seberapa efektif campaign atau interaksi dengan pelanggan menghasilkan tindakan yang diinginkan (misalnya, pembelian, pendaftaran, dll). Pelanggan dengan pendapatan lebih tinggi mungkin memiliki lebih banyak daya beli dan lebih mungkin merespons positif terhadap kampanye pemasaran yang menawarkan produk atau layanan yang sesuai dengan kemampuan finansial mereka.



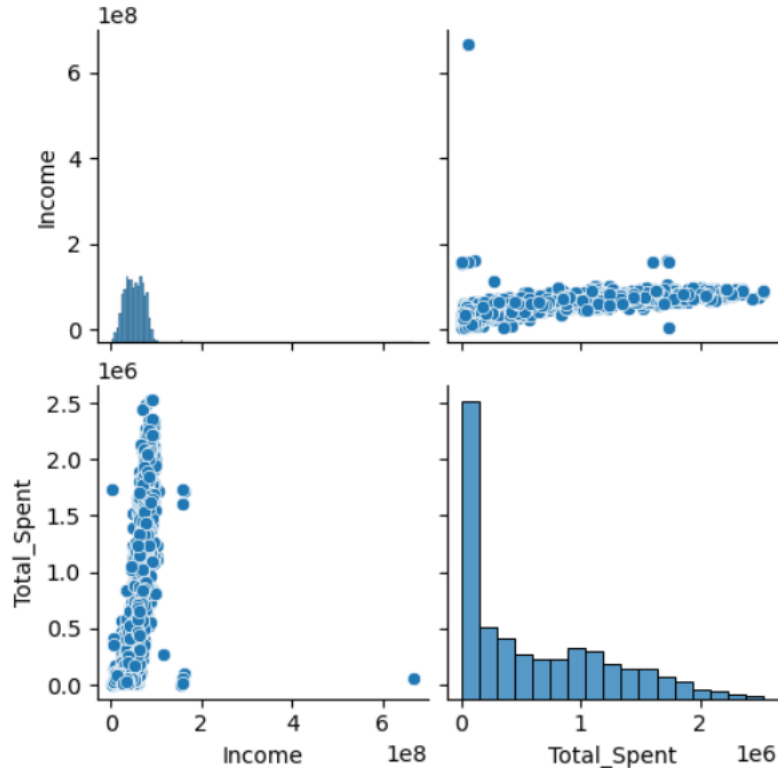
Conversion Rate Analysis Based on Income, **Spending** and Age

Pelanggan yang menghabiskan lebih banyak uang (total spent) juga cenderung memiliki tingkat konversi yang lebih tinggi. Pengeluaran yang lebih besar sering kali mencerminkan kepuasan dan keterlibatan yang lebih tinggi dengan produk atau layanan. Pelanggan yang puas dan terlibat lebih mungkin untuk merespons positif terhadap campaign pemasaran dan melakukan tindakan yang diinginkan.



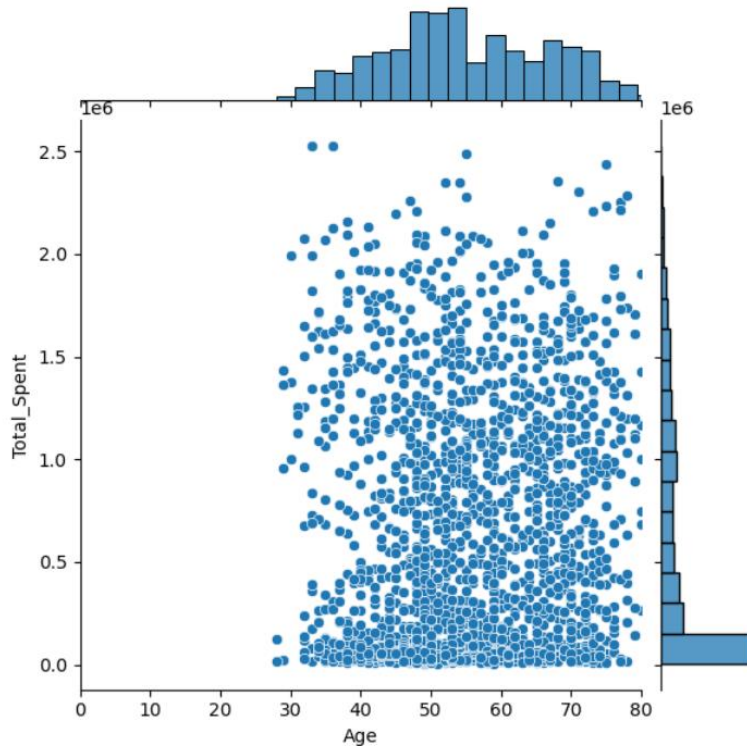
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Relation Income and Total Spent



Pelanggan dengan pendapatan yang lebih tinggi cenderung menghabiskan lebih banyak uang untuk produk atau layanan yang ditawarkan. Dengan daya beli yang lebih tinggi, pelanggan ini tidak hanya memiliki kemampuan untuk membeli lebih banyak, tetapi juga mungkin cenderung membeli produk yang lebih mahal atau dalam jumlah yang lebih besar.

Relation Age and Total Spent



1. Kontributor Utama:

Kelompok usia Paruh Baya (Middle Aged) memiliki total pengeluaran tertinggi, menunjukkan bahwa mereka memiliki daya beli yang sangat tinggi dan merupakan target pasar utama untuk perusahaan.

2. Kontributor Signifikan Lainnya:

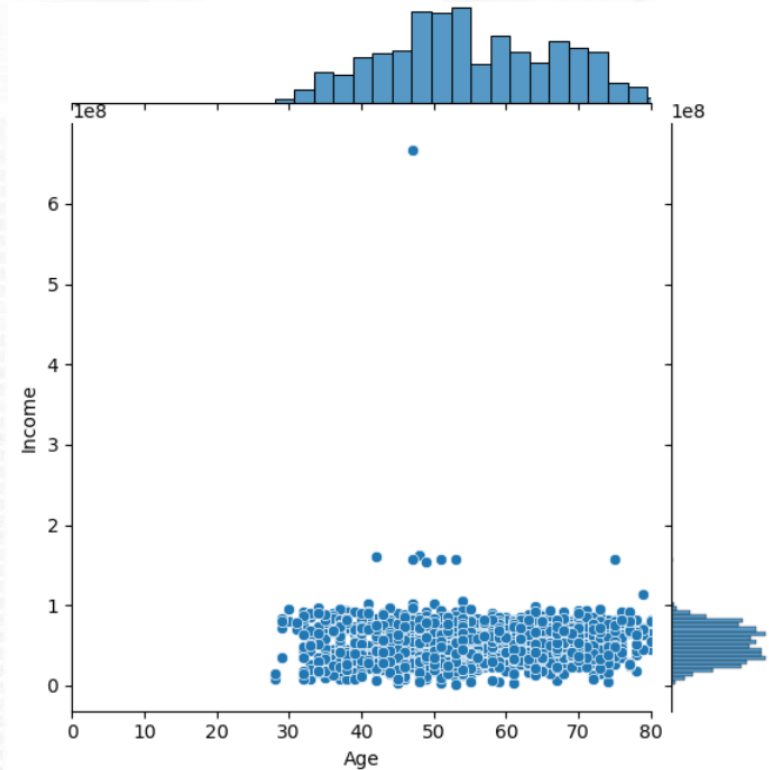
Kelompok usia Pensiun (Retirement) juga memiliki pengeluaran yang signifikan, menjadikannya kelompok penting untuk difokuskan dalam strategi pemasaran.

3. Kelompok Dewasa dan Lansia:

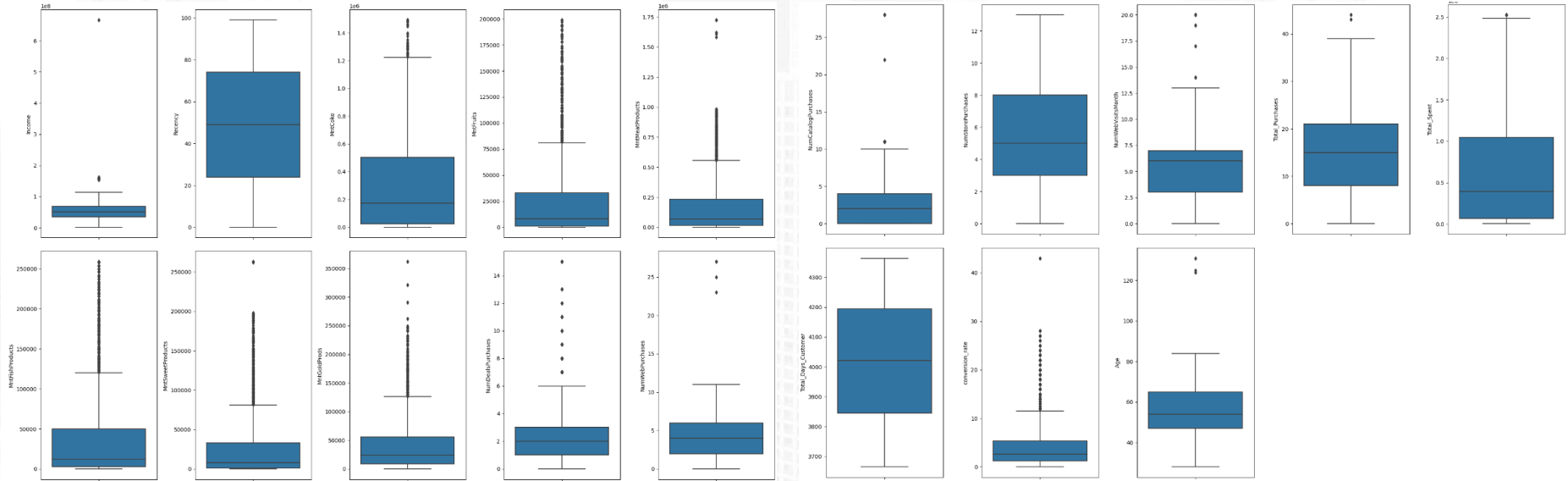
- Meskipun memiliki total pengeluaran yang lebih rendah, kelompok usia Dewasa (Adult) masih penting dan dapat ditingkatkan dengan strategi pemasaran yang tepat.
- Pengeluaran lebih rendah dari kelompok Lansia (Elderly) mungkin mencerminkan pendapatan tetap atau prioritas pengeluaran yang berbeda. Campaign yang lebih disesuaikan mungkin diperlukan untuk meningkatkan pengeluaran dalam kelompok ini.

Relation Age and Income

Tidak menunjukkan adanya keterhubungan satu sama lain, dikarenakan pendapatan tidak tergantung pada usia.



Univariate Analysis

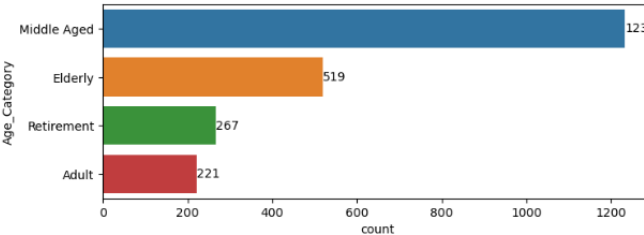
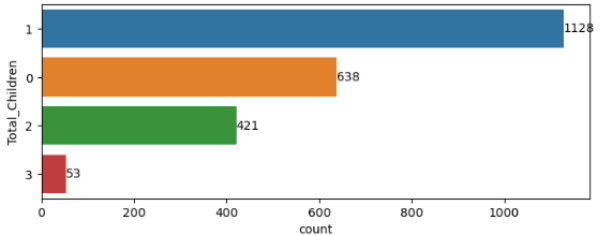
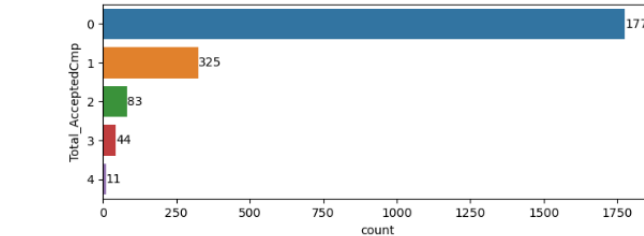
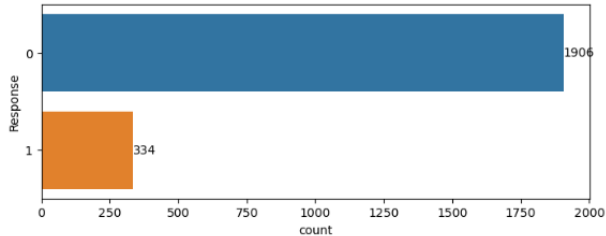
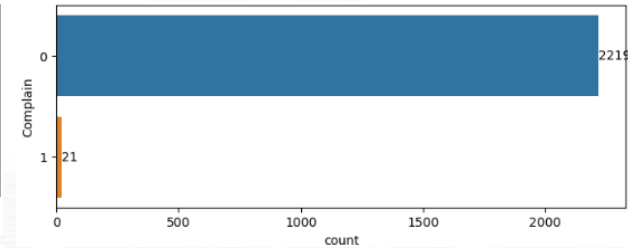
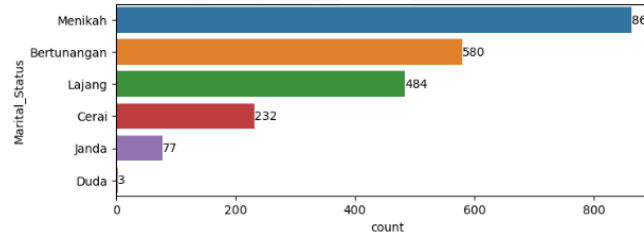
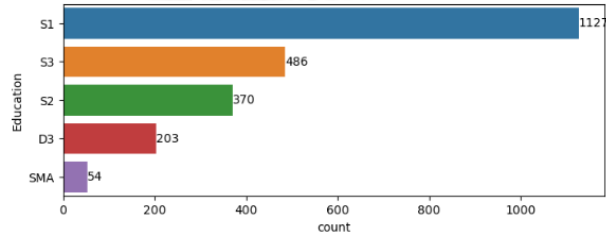


Terlihat banyak terdapat outlier, akan di handling saat data preprocessing

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

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

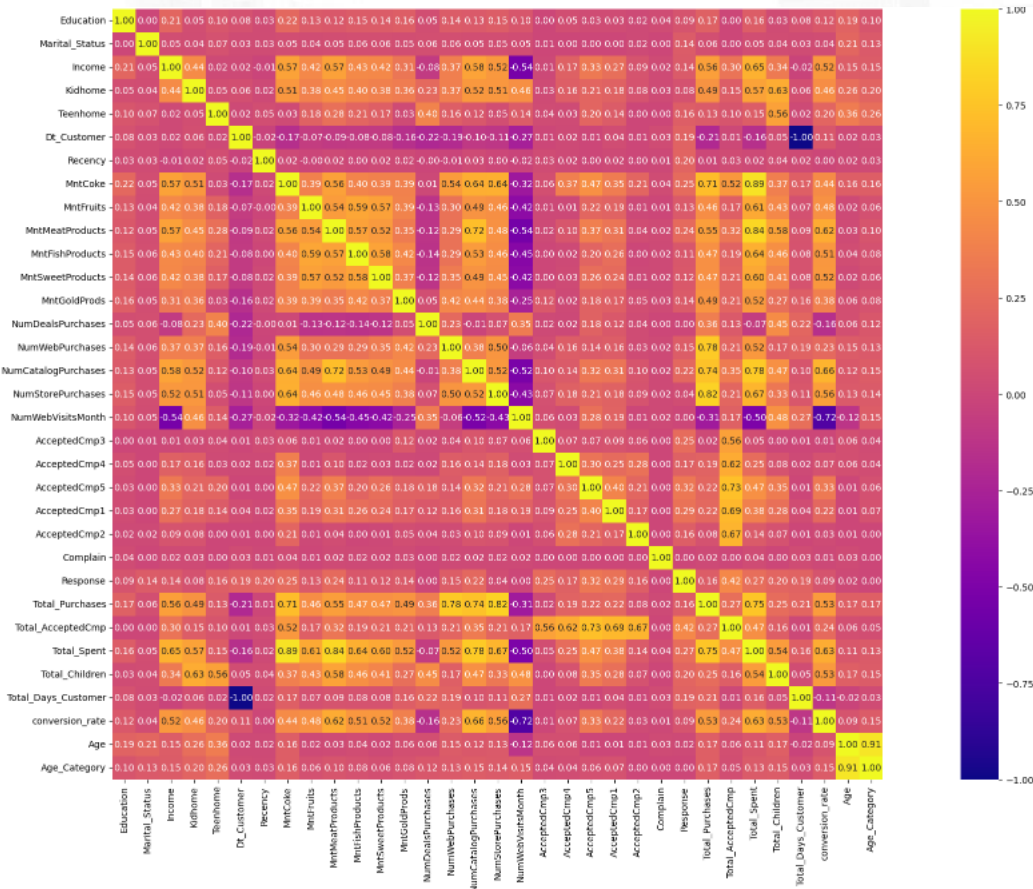
Univariate Analysis



- Berdasarkan pendidikannya, paling banyak customer yang berpendidikan S1
- Berdasarkan statusnya, paling banyak customer sudah menikah
- Berdasarkan tingkat complain, paling banyak customer tidak mengajukan complain
- Berdasarkan tingkat jumlah menerima campaign, paling banyak customer yang tidak menerima campaign
- Berdasarkan jumlah anak, paling banyak customer memiliki 1 anak
- Berdasarkan kategori usia, paling banyak customer yang termasuk Middle Aged

Exploratory Data Analysis (EDA)

Multivariate Analysis



- Masih terlalu banyak fitur
- Belum dapat digunakan, akan di filter beberapa fitur yang akan digunakan dalam pemodelan ML

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

Education 0
Marital_Status 0
Income 24
Kidhome 0
Teenhome 0
Dt_Customer 0
Recency 0
MntCoke 0
MntFruits 0
MntMeatProducts 0
MntFishProducts 0
MntSweetProducts 0
MntGoldProds 0
NumDealsPurchases 0
NumWebPurchases 0
NumCatalogPurchases 0
NumStorePurchases 0
NumWebVisitsMonth 0
AcceptedCmp3 0
AcceptedCmp4 0
AcceptedCmp5 0
AcceptedCmp1 0
AcceptedCmp2 0
Complain 0
Response 0
Total_Purchases 0
Total_AcceptedCmp 0
Total_Spent 0
Total_Children 0
Total_Days_Customer 0
conversion_rate 0
Age 0
Age_Category 0
dtype: int64

HANDLING MISSING VALUE

- 24 row data has missing value on Income
- Fill it with median

HANDLING DUPLICATED DATA

- 0 duplicated data

DROP FITUR YANG TIDAK DIPERLUKAN, MENJADI:

#	Column	Non-Null Count	Dtype
0	Education	2240 non-null	object
1	Marital_Status	2240 non-null	object
2	Income	2240 non-null	float64
3	Dt_Customer	2240 non-null	datetime64[ns]
4	Recency	2240 non-null	int64
5	NumWebVisitsMonth	2240 non-null	int64
6	Complain	2240 non-null	object
7	Total_Purchases	2240 non-null	int64
8	Total_AcceptedCmp	2240 non-null	object
9	Total_Spent	2240 non-null	int64
10	Total_Children	2240 non-null	object
11	Total_Days_Customer	2240 non-null	int64
12	conversion_rate	2240 non-null	float64
13	Age	2240 non-null	int64
14	Age_Category	2240 non-null	object

dtypes: datetime64[ns](1), float64(2), int64(6), object(6)

FEATURE ENCODING

Feature: Education

FEATURE SELECTION

Kolom yang akan dipilih berdasarkan RFM dengan metode reduce dimensionality:

- R** (Recency): Recency (Jumlah hari sejak pelanggan terakhir berinteraksi dengan perusahaan)
- F** (Frequency) : Total_Purchases (Jumlah total pembelian yang dilakukan oleh pelanggan)
- M** (Monetary) : Total_Spent (Total pengeluaran pelanggan)
- L** (Loyalty) : Total_Days_Customer (jumlah hari sejak pelanggan pertama kali mendaftar atau menjadi pelanggan)
- C** : Age (Usia pelanggan)

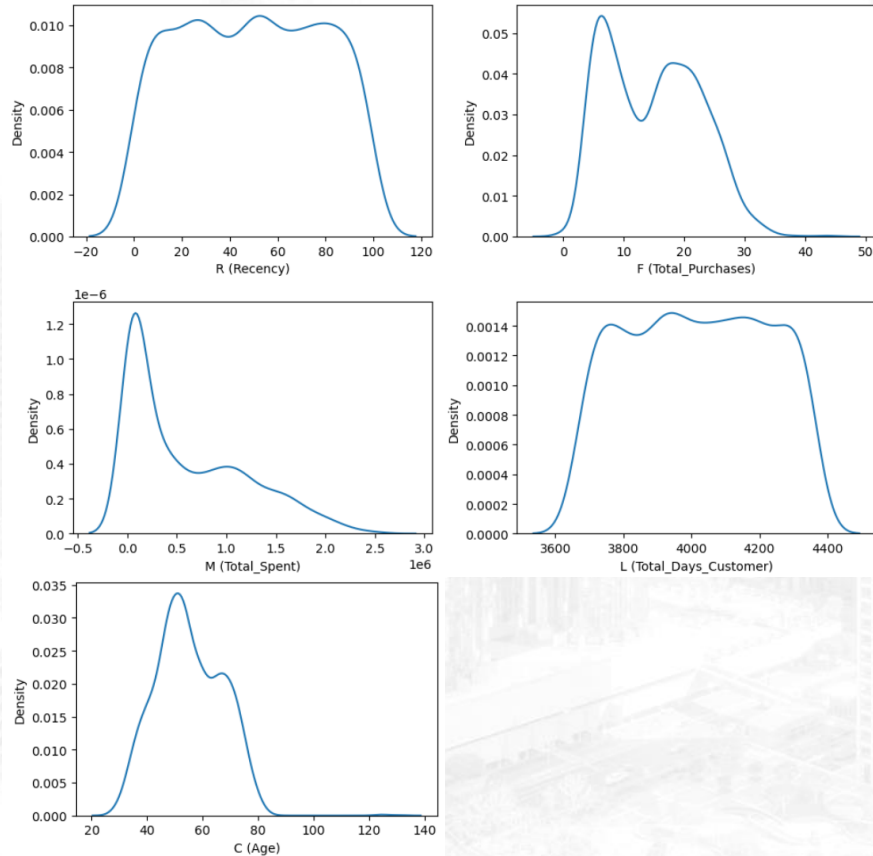
HANDLING OUTLIER

Menggunakan IQR method (Q1=1%; Q3=99%)

FEATURE TRANSFORMATION

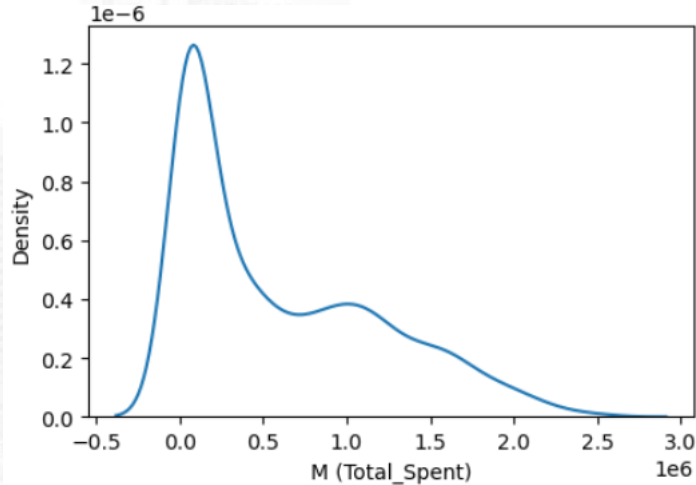
Standardization pada 5 features yang digunakan untuk modeling menggunakan MinMaxScaler

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



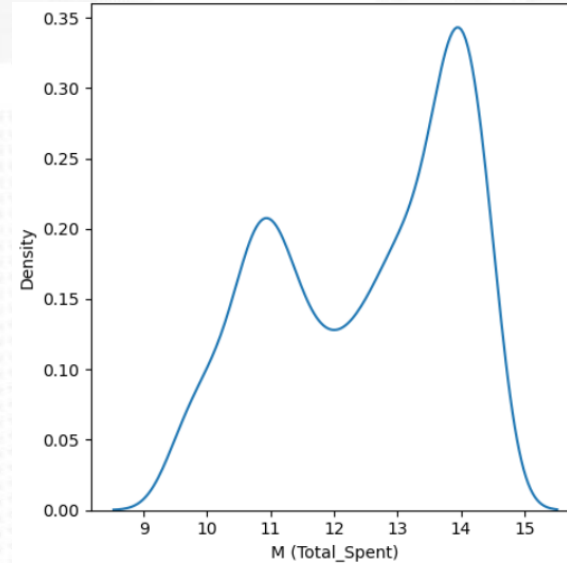
Fitur M (Total_Spent) memiliki distribusi long-right skewed sehingga dilakukan log transformation

Before Log Transformation

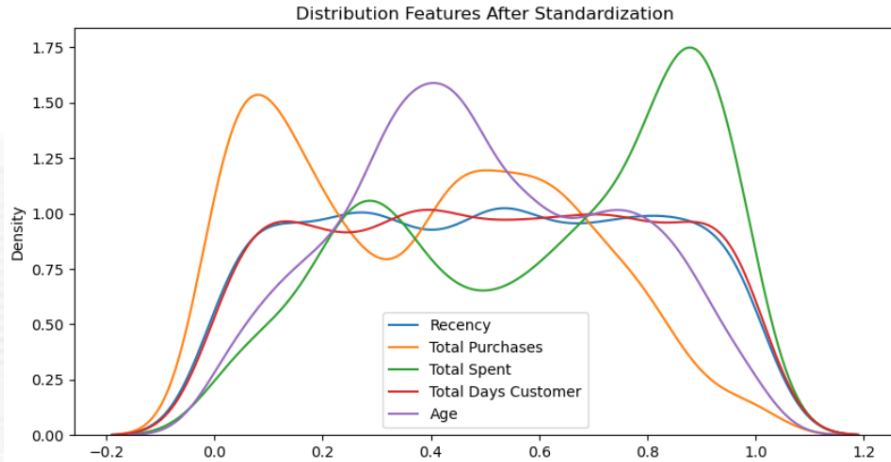


Distribusi right-skewed

After Log Transformation



Distribusi menjadi mendekati normal

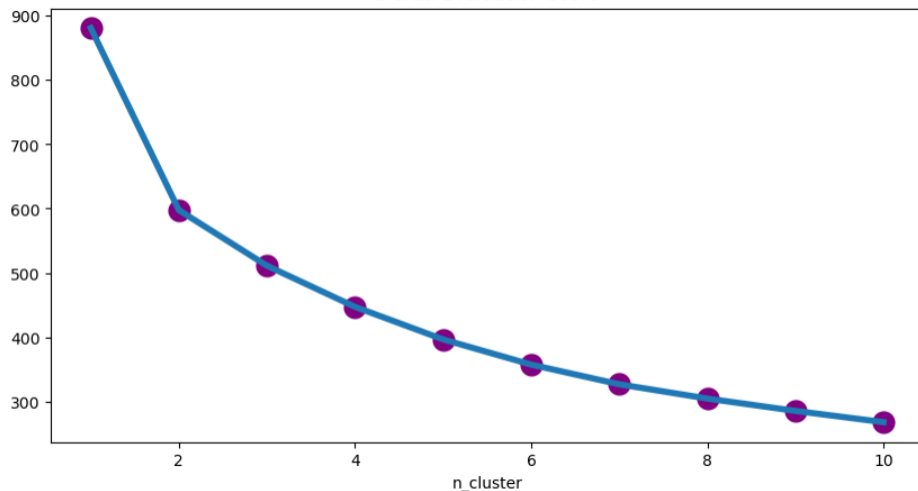


	R (Recency)	F (Total_Purchases)	M (Total_Spent)	L (Total_Days_Customer)	C (Age)
count	2240.000000	2240.000000	2240.000000	2240.000000	2240.000000
mean	0.501039	0.387675	0.597922	0.506694	0.492173
std	0.295403	0.271121	0.289446	0.295306	0.248215
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.244898	0.142857	0.326759	0.254020	0.319149
50%	0.500000	0.392857	0.670283	0.509503	0.468085
75%	0.755102	0.607143	0.860753	0.763158	0.702128
max	1.000000	1.000000	1.000000	1.000000	1.000000

Fitur yang akan digunakan untuk pemodelan Machine Learning sudah distandardisasi menggunakan MinMaxScaler

Elbow Method using Inertia

Inertia Evaluation Score



Inertia score:

```
0    32.153658
1    14.467672
2    12.510600
3    11.319539
4     9.892008
5     8.564164
6     6.792799
7     6.303135
8     6.127182
9         NaN
dtype: float64
```

Titik k yang optimal di angka 2

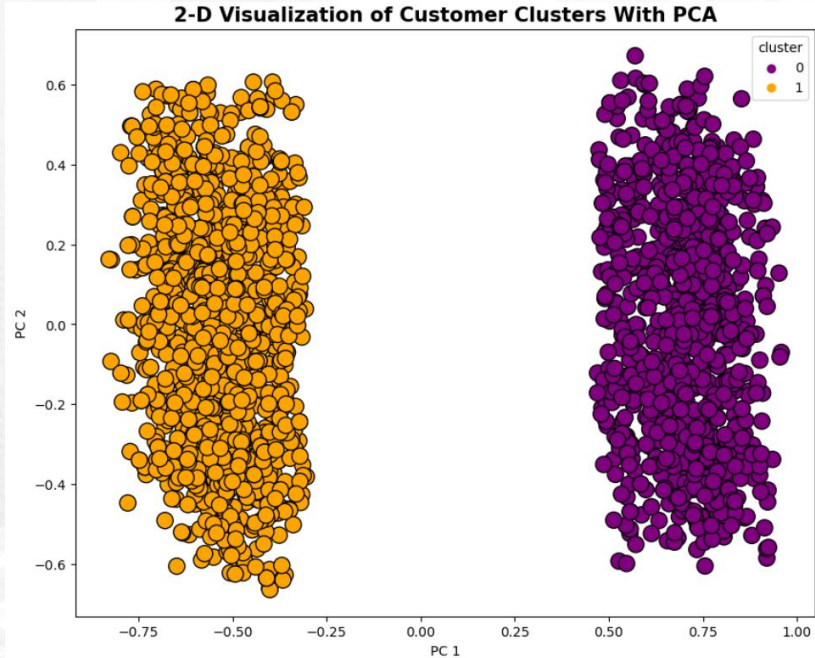
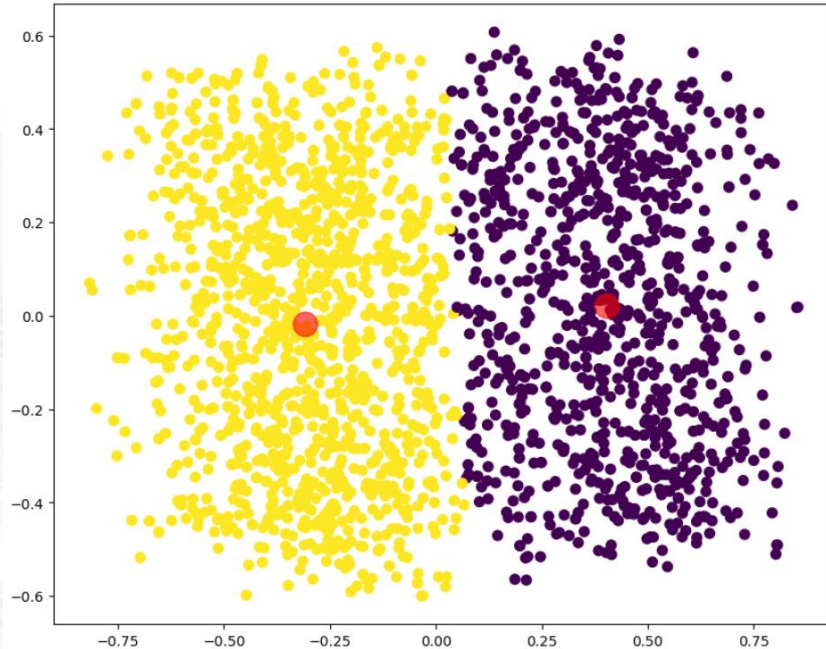
Silhouette Score

Untuk `n_clusters = 2`, rata-rata `silhouette_score` adalah : 0.16603883507156295
Untuk `n_clusters = 3`, rata-rata `silhouette_score` adalah : 0.16064647080746355
Untuk `n_clusters = 4`, rata-rata `silhouette_score` adalah : 0.15062150868807853
Untuk `n_clusters = 5`, rata-rata `silhouette_score` adalah : 0.1518335888265661
Untuk `n_clusters = 6`, rata-rata `silhouette_score` adalah : 0.1509015990003672
Untuk `n_clusters = 7`, rata-rata `silhouette_score` adalah : 0.15650434033497063
Untuk `n_clusters = 8`, rata-rata `silhouette_score` adalah : 0.15530079074300795
Untuk `n_clusters = 9`, rata-rata `silhouette_score` adalah : 0.14653941383523314
Untuk `n_clusters = 10`, rata-rata `silhouette_score` adalah : 0.155075819179872

Score tertinggi dari Silhouette Score terletak pada `n_clusters` ke 2

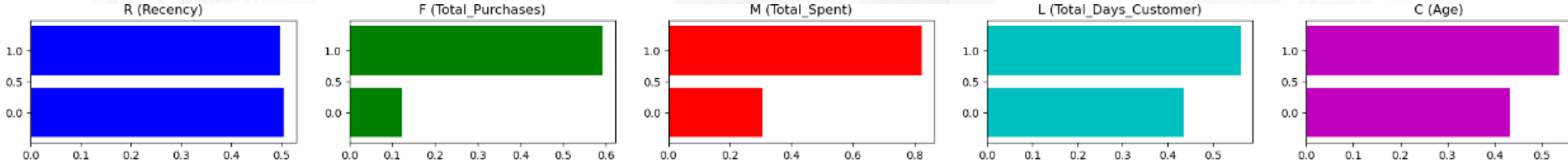
Untuk selengkapnya, dapat melihat jupyter notebook [disini](#)

Evaluasi Cluster menggunakan PCA (Principal Component Analysis)



Customer terbagi menjadi 2 kelompok

Evaluasi Cluster menggunakan PCA (Principal Component Analysis)



Cluster 0: General Customer

1. Jumlah hari sejak pelanggan terakhir berinteraksi dengan perusahaan (Recency) yang relatif lebih panjang.
2. Jumlah pembelian yang dilakukan oleh pelanggan (Total Purchase) yang sangat rendah.
3. Total pengeluaran (Total Spent) yang cukup rendah.
4. Durasi member pelanggan (Total Days Customer) cukup rendah.
5. Usia pelanggan (Age) yang relatif lebih muda.

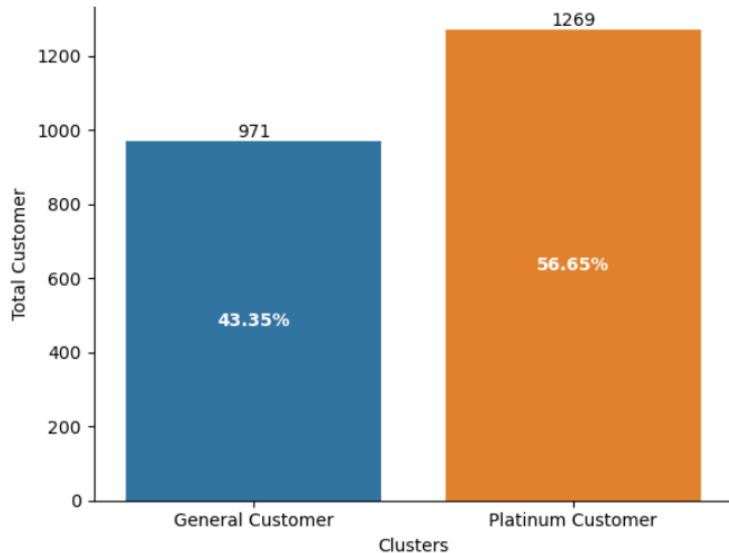
Penjelasan: Pelanggan dalam cluster ini cenderung terdiri dari pengunjung biasa yang tidak terlalu sering melakukan konversi/pembelian.

Cluster 1: Platinum Customer

1. Jumlah hari sejak pelanggan terakhir berinteraksi dengan perusahaan (Recency) yang relatif lebih rendah.
2. Jumlah pembelian yang dilakukan oleh pelanggan (Total Purchase) yang sangat tinggi.
3. Total pengeluaran (Total Spent) yang sangat tinggi.
4. Durasi member pelanggan (Total Days Customer) sangat tinggi.
5. Usia pelanggan (Age) yang relatif lebih berusia lanjut.

Penjelasan: Pelanggan dalam cluster ini merupakan pengunjung rata-rata yang memiliki tingkat keaktifan dan kesetiaan yang cukup tinggi, dengan sering melakukan konversi/pembelian dan menerima marketing campaign.

Total of Customers Each Cluster



Rekomendasi Bisnis

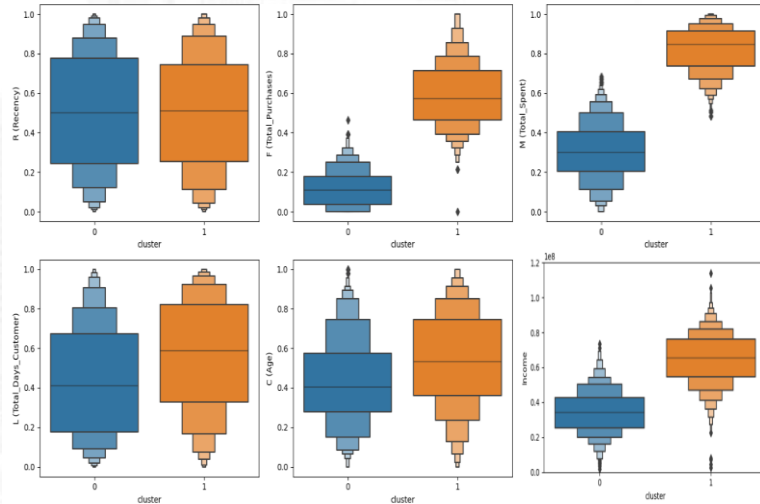
General Customer

- Peningkatan Engagement: Memberikan insentif untuk meningkatkan frekuensi pembelian, seperti program loyalty atau poin reward untuk setiap pembelian.
- Promosi Khusus: Mengirimkan penawaran diskon atau promosi eksklusif untuk mendorong pelanggan berbelanja lebih banyak.
- Produk Bundle: Menawarkan bundling produk dengan harga lebih rendah untuk meningkatkan total pengeluaran.

Platinum Customer

- Personalized Marketing: Mengirimkan email atau pesan yang dipersonalisasi berdasarkan preferensi pembelian mereka.
- Early Access: Memberikan akses awal ke produk baru atau penjualan eksklusif untuk meningkatkan rasa eksklusivitas.
- Feedback Loops: Mengumpulkan umpan balik untuk terus meningkatkan pengalaman pelanggan dan menyesuaikan penawaran produk.

CLUSTER FOR MARKETING RETARGETING



Cluster 1 (Platinum Customer) memiliki nilai Total_Purchases, Total_Spent, dan Income yang lebih tinggi dibandingkan dengan Cluster 0 (General Customer).

Cluster 1 (Platinum Customer) menjadi target marketing campaign

STRATEGI MARKETING CAMPAIGN

Personalized Marketing

- Kirimkan penawaran dan promosi yang dipersonalisasi berdasarkan preferensi pembelian dan riwayat transaksi pelanggan.
- Gunakan analisis data untuk mengidentifikasi produk atau layanan yang paling menarik bagi pelanggan di Cluster 1.

Exclusive Promotions

- Berikan akses eksklusif kepada pelanggan Cluster 1 untuk penjualan produk terbaru atau penawaran khusus.
- Tawarkan diskon khusus atau paket bundling yang menarik untuk meningkatkan loyalitas dan frekuensi pembelian.

Loyalty Programs

- Buat program loyalitas yang menawarkan poin atau hadiah untuk setiap pembelian, yang dapat ditukarkan dengan produk atau layanan eksklusif.
- Pertimbangkan untuk memberikan status VIP atau platinum dengan manfaat tambahan seperti pengiriman gratis atau layanan pelanggan prioritas.

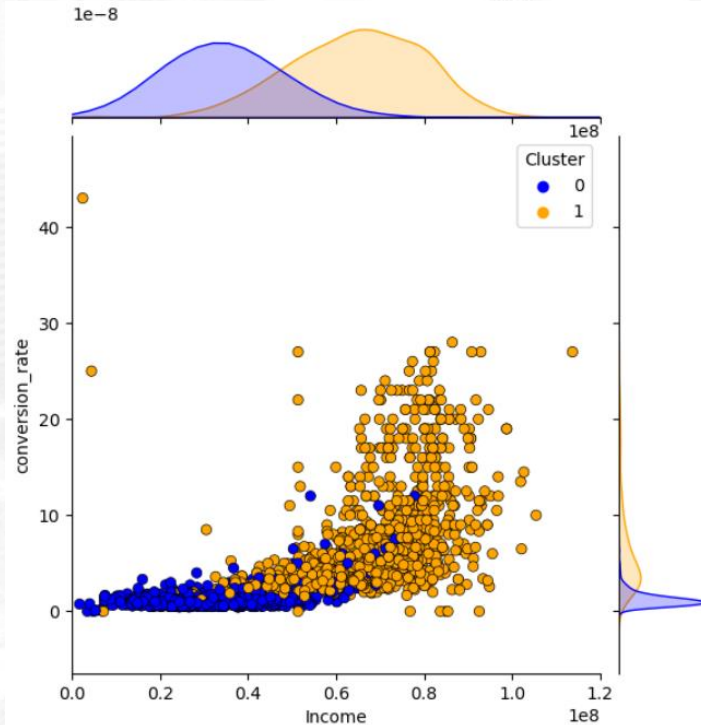
Customer Feedback and Engagement

- Minta umpan balik dari pelanggan di Cluster 1 untuk terus meningkatkan produk dan layanan yang ditawarkan.
- Gunakan feedback ini untuk menyesuaikan strategi marketing dan meningkatkan pengalaman pelanggan.

Data-Driven Insights

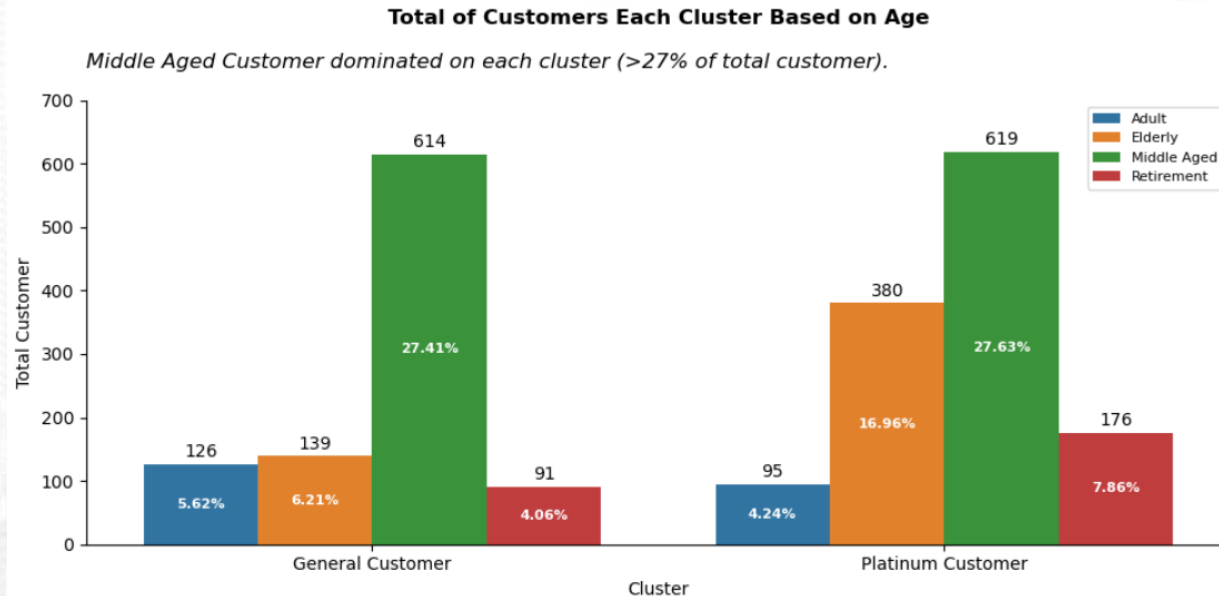
- Gunakan analisis data lanjutan untuk terus memantau perilaku pembelian dan menyesuaikan kampanye marketing sesuai kebutuhan.
- Identifikasi tren dan pola pembelian untuk mengoptimalkan strategi pemasaran.

Conversion Rate Based On Income Each Cluster



- Cluster 1 memiliki income yang lebih tinggi dan memiliki conversion rate yang lebih tinggi juga.
- Cluster 1 memang cocok dijadikan target marketing campaign berikutnya.

Customer Personality Analysis for Marketing Retargeting



- Pelanggan paruh baya (Middle Aged) mendominasi kedua cluster dengan lebih dari 27% dari total pelanggan.
- Platinum Customer memiliki lebih banyak pelanggan dalam kategori Lansia (16.96%) dan Pensiun (7.86%) dibandingkan dengan General Customer.
- Adult dan Lansia memiliki jumlah yang relatif lebih rendah di kedua cluster.

Rekomendasi:

Untuk marketing retargeting, dapat fokus pada pelanggan paruh baya (Middle Aged) karena mereka adalah mayoritas dalam kedua cluster. Selain itu, strategi yang berbeda mungkin diperlukan untuk mengakomodasi kebutuhan kelompok usia lainnya seperti lansia dan dewasa.

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

Potential Impact of Marketing Retargeting on Cluster 1



Peningkatan Total Pembelian: 2956.40 pembelian

Memilih cluster 1 sebagai target marketing campaign berpotensi meningkatkan jumlah pembelian sebanyak 2956.40 unit atau item. Ini menunjukkan bahwa cluster 1 memiliki potensi besar untuk merespons campaign pemasaran dengan meningkatkan aktivitas pembelian mereka secara signifikan.

Peningkatan Total Pengeluaran: \$2503.27

Campaign pemasaran yang ditargetkan pada cluster 1 juga berpotensi meningkatkan total pengeluaran konsumen dalam cluster tersebut sebesar \$2503.27. Artinya, dengan menargetkan kampanye pada cluster 1, ada potensi untuk meningkatkan volume penjualan dan dengan demikian meningkatkan pendapatan yang dihasilkan dari penjualan kepada konsumen dalam cluster ini.