

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



**Created by:**

**Citta Mudita**

[citmud10@gmail.com](mailto:citmud10@gmail.com)

<https://www.linkedin.com/in/citta-mudita-40b91b22b/>

I have a strong interest in data analysis and data science. My experience in digital marketing has made me an expert in conducting exploratory data analysis and creating data visualizations to address business challenges. I have received awards as the Top 2 Student in the Data Science Bootcamp and as the Most Outstanding Student for the Final Project. Additionally, my team was recognized as the Best Final Project Team at Data Science Bootcamp.

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

“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 ”

# Conversion Rate Analysis Based on Income, Spending and Age

## Feature Engineering

### Total children

Menjumlahkan kidhome and teenhome

### Marital Status

Mengelompokan berdasarkan apakah ia orang tua atau bukan

### Total accepted campaign

Menjumlahkan total accepted 1 hingga 5

### Total purchases

Menjumlahkan total purchases deal, web, catalog, and store

### Total spent

Menjumlahkan total coke, fish, fruit, meat, sweat, and goldprods

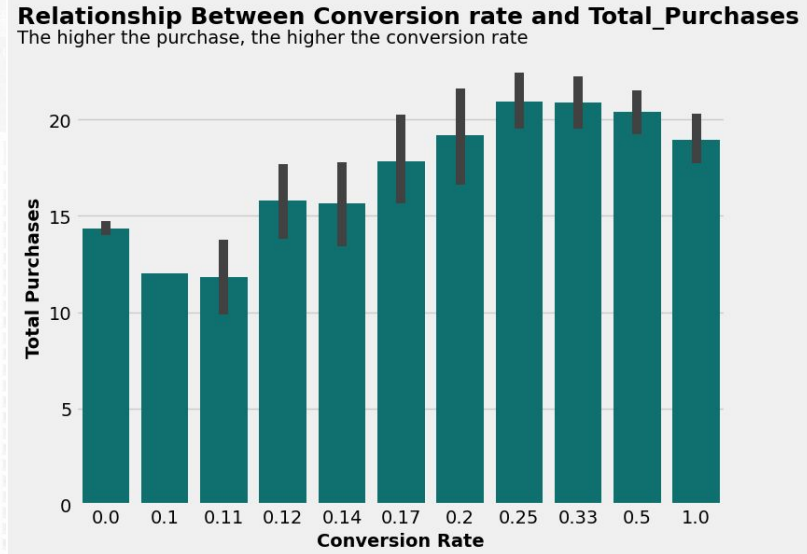
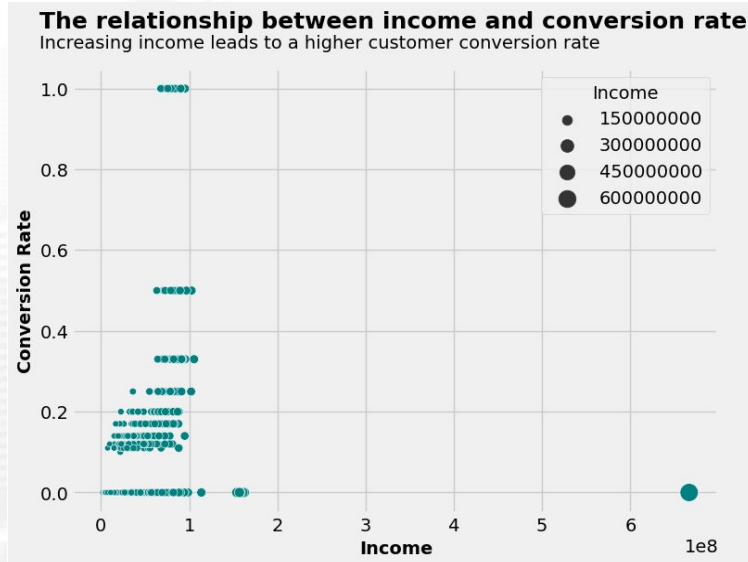
### Conversion rate

$\text{Conversion rate} = \text{response} / \text{visit}$

### Grup age

Mengelompokan umur berdasarkan kategori young adult, adult, and senior adult

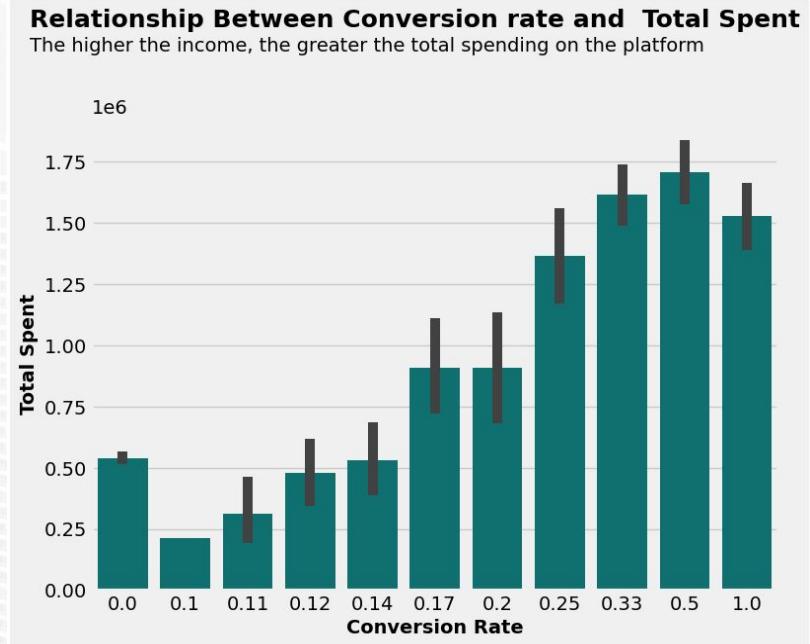
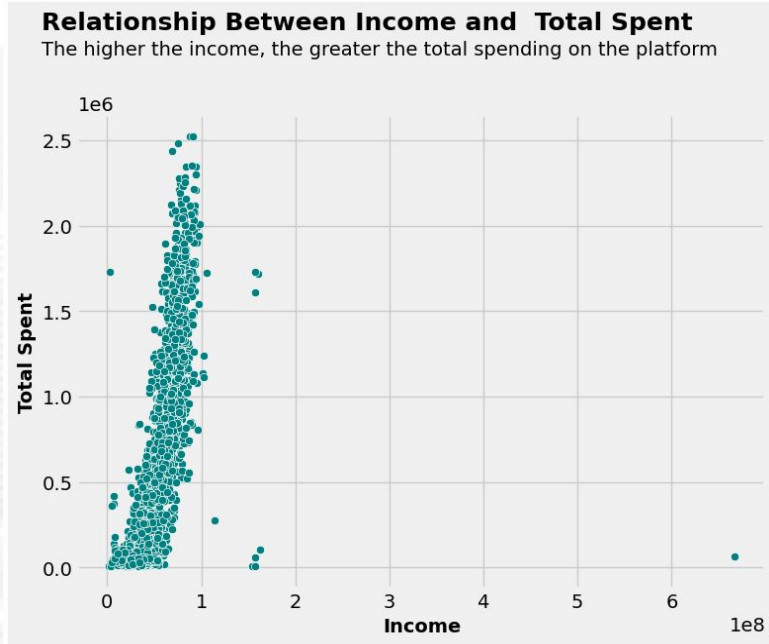
# Conversion Rate Analysis Based on Income, Spending and Age



Pendapatan yang lebih tinggi, menunjukkan tingkat konversi yang lebih tinggi. Dengan tingkat konversi tinggi maka total purchases juga tinggi. Dengan demikian kita bisa menargetkan kampanye kepada pelanggan dengan pendapatan yang lebih tinggi.

Untuk pelanggan dengan pendapatan lebih rendah, kita dapat menjelajah lebih jauh dengan menggunakan teknik pengelompokan untuk mengelompokkan mereka ke dalam cluster dan memberikan rekomendasi yang disesuaikan dengan masing-masing cluster, yang berpotensi meningkatkan jumlah transaksi di platform kita.

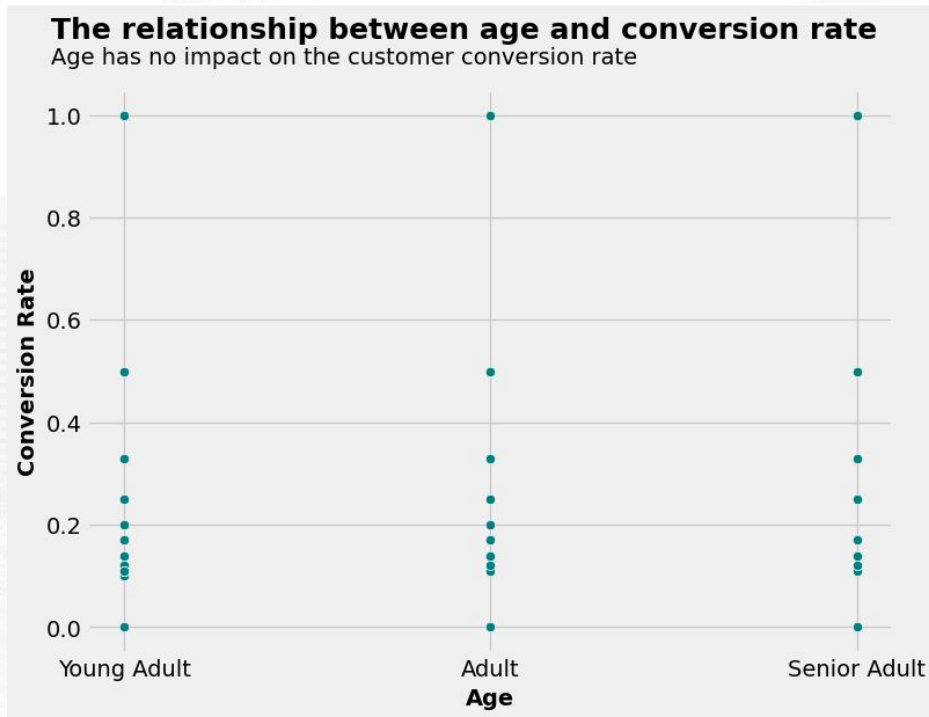
# Conversion Rate Analysis Based on Income, Spending and Age



Semakin tinggi pendapatan seorang pelanggan, semakin tinggi pula tingkat konversinya. Temuan ini didukung insight lain, di mana peningkatan pendapatan berhubungan dengan total pengeluaran yang lebih tinggi di platform, dan terdapat hubungan linier antara total pengeluaran dan tingkat konversi.



# Conversion Rate Analysis Based on Income, Spending and Age



Namun pada feature umur, baik untuk grup umur young adult, adult, dan senior adult tidak memiliki hubungan yang spesifik dengan conversion. Tidak ada umur spesifik yang memiliki conversion rate yang tinggi.

## Missing Value



Income

Mengisi nilai null dengan nilai median

## Duplicated Data



no duplicate data

## Feature Selection



- 'Unnamed: 0'
- 'AcceptedCmp1'
- 'AcceptedCmp2'
- 'AcceptedCmp3'
- 'AcceptedCmp4'
- 'AcceptedCmp5'
- 'NumDealsPurchases'
- 'NumWebPurchases'
- 'NumCatalogPurchases'
- 'Z\_Revenue'
- 'NumStorePurchases'
- 'MntCoke'
- 'MntFishProducts'
- 'MntFruits'
- 'MntMeatProducts'
- 'MntSweetProducts'
- 'MntGoldProds'
- 'Kidhome'
- 'Teenhome'
- 'age\_order'
- 'Year\_Birth'
- 'ID'
- 'Z\_CostContact'

## Feature Engineering



**Label Encoding :**  
Education

**One Hot Encoding :**  
Marital\_Status,  
age\_grup, Has\_Parents

## Standarisasi



Handling  
standarisasi using  
MinMaxScaler

## Handling Outlier



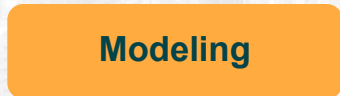
Handling outlier  
using Persentil  
1% - 99%

## Metode RMF

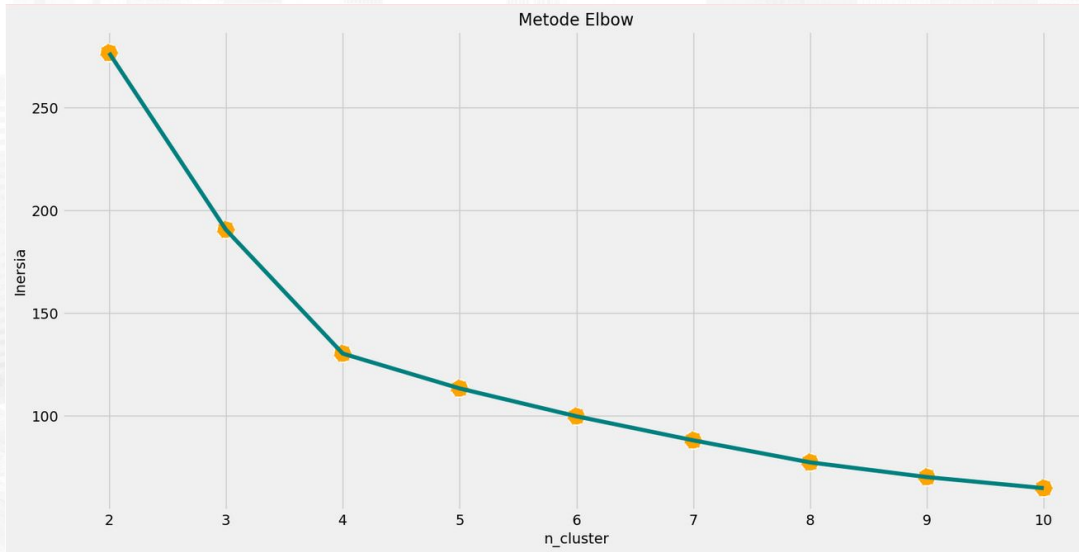


Recency  
Total\_Purchases  
Total\_Spent

## Modeling



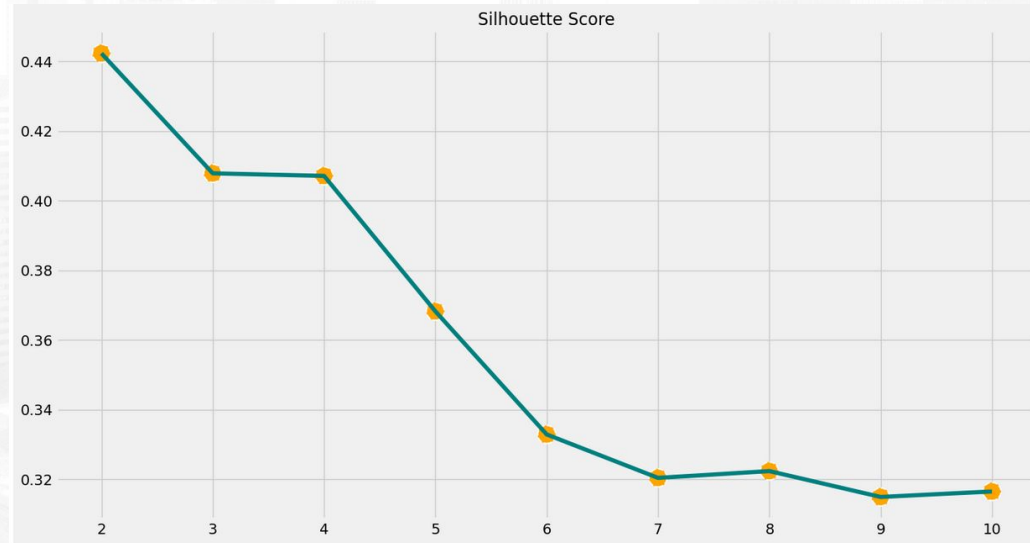
## ***Elbow Method*** menggunakan ***K-Means Clustering***



Dalam gambar di atas, kita dapat melihat bahwa ketika  $n\_cluster = 4$ , skor inersia tidak mengalami perubahan yang signifikan, sehingga kita akan menggunakan  $n\_cluster = 4$ .  $n=4$  dimana perbedaan inerti dengan menambah cluster baru berkurang drastis

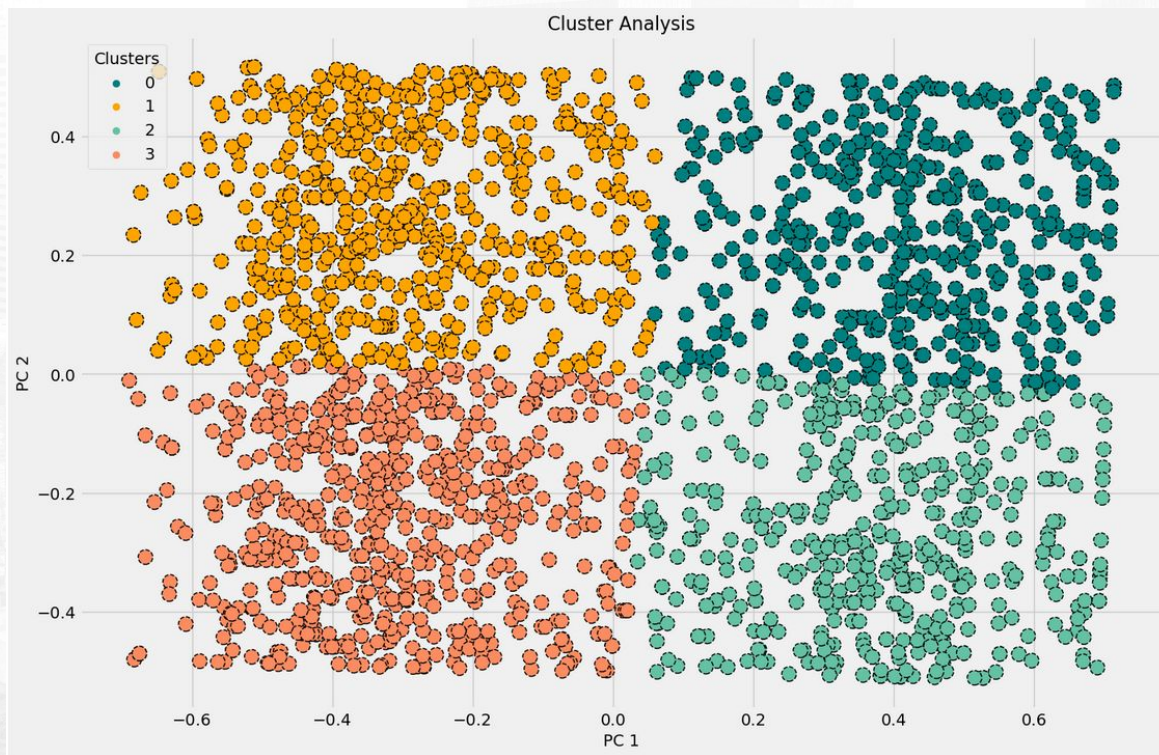


## Evaluasi menggunakan *Silhouette Score*

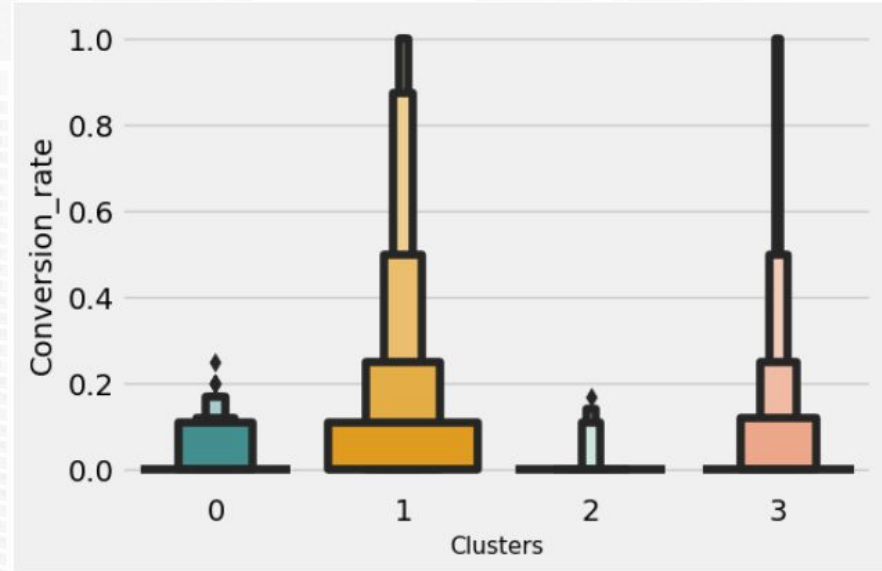
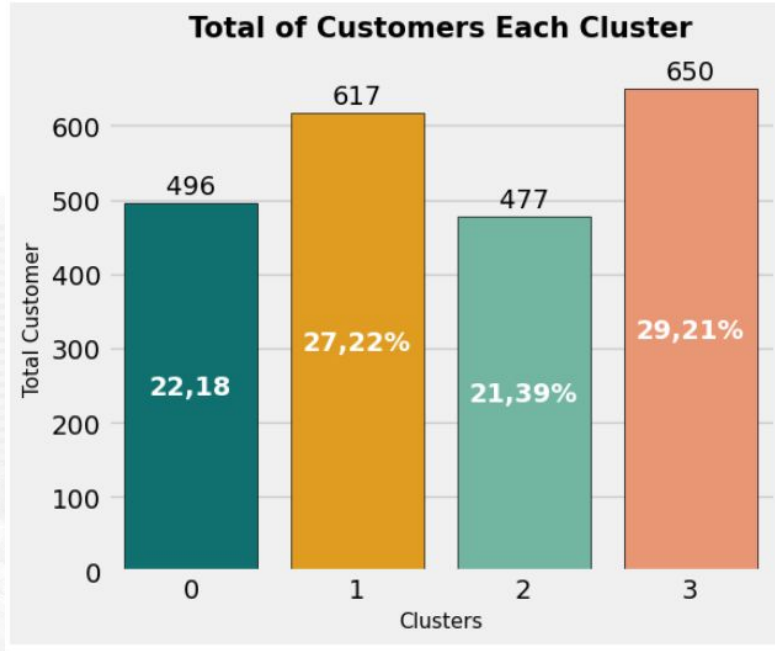


Dari gambar di atas, kita dapat melihat bahwa ketika  $n\_cluster = 4$ , skor silhouette adalah 0.41.

Ini menunjukkan bahwa, secara umum, pengamatan dalam kelompok yang sama memiliki jarak yang cukup seragam satu sama lain dan juga cukup terpisah dari kelompok lain. Nilai 0.41 mengindikasikan bahwa pengelompokan ini mungkin memadai, tetapi masih memiliki ruang untuk perbaikan. Semakin tinggi nilai Silhouette Score, semakin baik pengelompokan data tersebut.

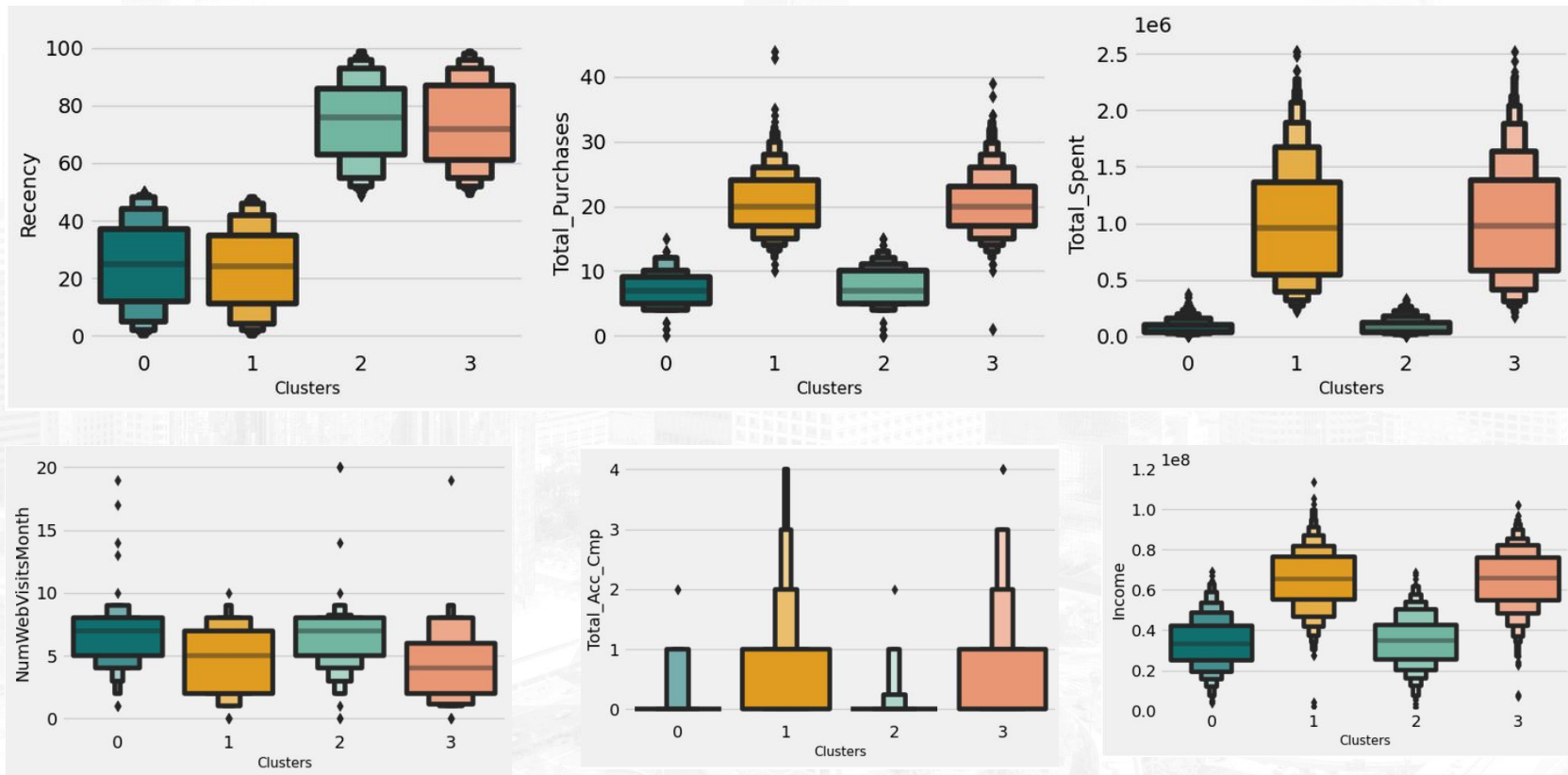


# Customer Personality Analysis for Marketing Retargeting



Untuk selengkapnya, dapat melihat jupyter notebook disini

# Customer Personality Analysis for Marketing Retargeting





# Customer Personality Analysis for Marketing Retargeting

## Cluster 0

### Risk of Churn

Kelompok ini memiliki pendapatan dan pengeluaran terendah. Mereka sering mengunjungi situs web kita, namun jarang melakukan transaksi atau menggunakan promo-promo di platform kami. Mereka lebih cenderung bertransaksi secara organik.

## Cluster 1

### Mid Spender

Kelompok ini memiliki pendapatan dan pengeluaran terbesar kedua setelah pengeluar besar. Meskipun mereka cukup jarang mengunjungi platform, mereka merespons kampanye dengan baik. Ciri khasnya adalah penggunaan promo yang lebih tinggi daripada pengeluar besar. Kami perlu memahami preferensi pembelian kelompok ini agar dapat mengelola promosi dengan lebih efisien dan mengurangi biaya kami.

## Cluster 2

### Low Spender

Kelompok ini memiliki pengeluaran lebih rendah daripada pengeluar besar dan pengeluar menengah. Mereka cukup sering mengunjungi situs web dan mencari promo, tetapi mereka tidak mengaplikasikan promo se sering kelompok pengeluar menengah.

## Cluster 3

### High Spender

Kelompok pelanggan ini jumlahnya lebih sedikit jika dibandingkan dengan total midspender, total low spender dan risk of churn, tetapi mereka adalah pengeluar besar dengan pendapatan dan pengeluaran yang tinggi di platform. Mereka tidak sering mengunjungi platform kita, tetapi ketika mereka melakukannya, mereka cenderung bertransaksi. Kelompok ini merupakan target yang sangat menarik karena memiliki tingkat konversi yang tinggi.



## Recommendation

### Risk of Churn

Melihat perilaku mereka saat berhadapan dengan produk, apakah produk yang ditawarkan kurang sesuai, atau apakah sensitivitas terhadap harga memengaruhi mereka untuk tidak melakukan transaksi di platform kita.

### Mid Spender

Lakukan analisis lebih mendalam untuk memahami cara meningkatkan frekuensi transaksi dalam kelompok ini. Berikan rekomendasi yang lebih personal dan melakukan analisis lebih detail tentang bagaimana mengoptimalkan promosi untuk kelompok ini. Meskipun kami menawarkan promosi lebih sedikit, kelompok ini tetap berbelanja di platform kita.

### Low Spender

Berikan insentif kepada pelanggan dalam kelompok ini untuk mengaktifkan kembali aktivitas berbelanja mereka, misalnya dengan menawarkan diskon khusus atau promosi untuk pembelian pertama setelah jangka waktu yang lama.

### High Spender

Memantau transaksi dan retensi dari kelompok ini. Fokus pada peningkatan pelayanan agar kelompok ini tetap loyal dan tidak melakukan churn.

## Potential Impact GMV

Risk of Churn

GMV Risk of Churn: 38259000

Mid Spender

GMV Mid Spender: 614752000

Low Spender

GMV Low Spender: 41176000

High Spender

GMV High Spender: 662801000

Jika kita fokus untuk memfokuskan pada high spender, kita bisa mendapatkan gmv\_high\_spender sebesar 662801000, sedangkan pada mid\_spender sebesar 614752000.