

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



**Created by:**

**Putri Kirey Eki Yogaswari**

Kirey.ey11@gmail.com

<https://www.linkedin.com/in/putri-eki-yogaswari-477952235>

I am a fresh graduate from Public Health with a focus on Health Statistics. During college, I have knowledge of situation analysis, designing health programs, leadership skills, systems thinking and data analysis. I have a career interest in data management and proficient in using SPSS and Ms. Excel. Has managed data up to 2000+.

“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 ”

## Feature Engineering



Tidak ada  
data duplikat

Terdapat 1 missing  
value pada feature  
Income

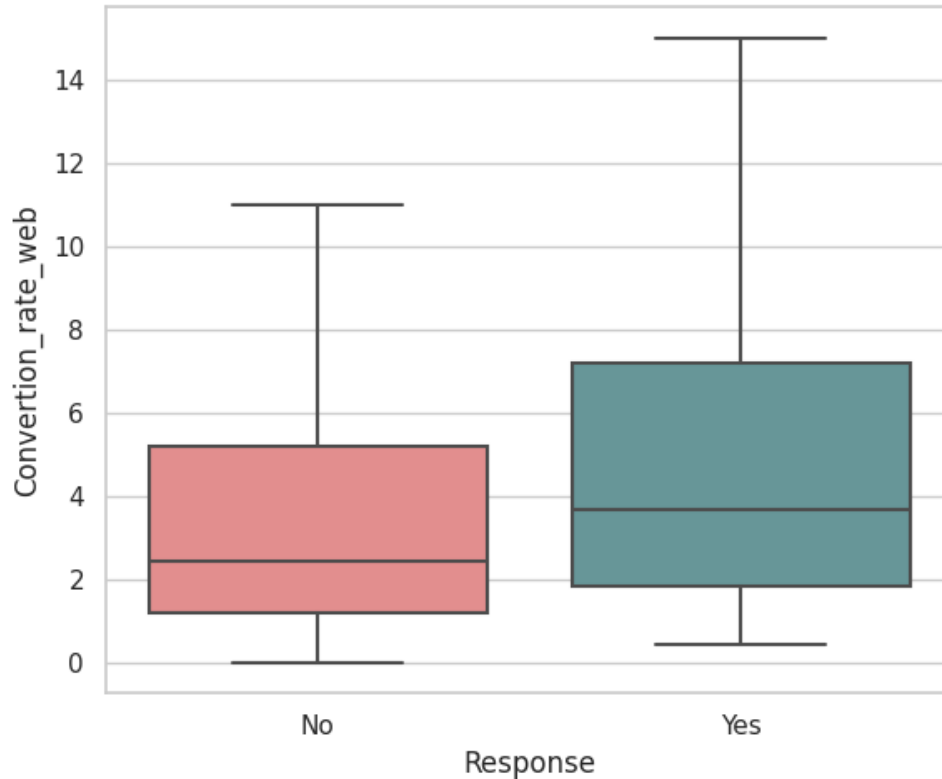
2240 baris  
30 kolom

Terdapat 723  
outlier all data

### **Feature Extraction:**

1. Spending
2. Age
3. Age Group
4. Has Child
5. Dependents
6. Lifetime
7. Primer\_purchase
8. Tersier\_purchase
9. Total Purchases
10. Total Cmp
11. Ever Accept
12. Total Revenue
13. Income Segment
14. Conversion Rate Web
15. Month & Year Joined

# Conversion Rate vs Response

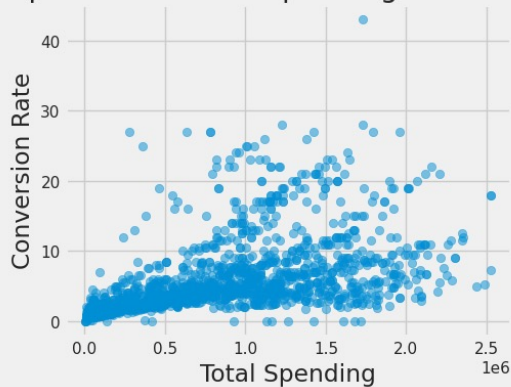


## Insight :

Semakin besar nilai conversion rate web semakin besar pelanggan merespon campaign.

# Hubungan antara Total Spending, Income dengan Conversion Rate

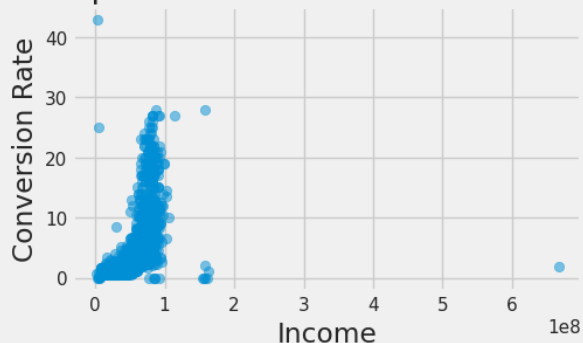
Relationship between Total Spending and Conversion Rate



## Insight:

Pada grafik scatter plot menunjukkan korelasi yang positif dimana semakin tinggi total spending maka semakin tinggi nilai conversion rate, oleh karena itu pelanggan dengan total pengeluaran yang lebih tinggi cenderung merespons kampanye

Relationship between Income and Conversion Rate

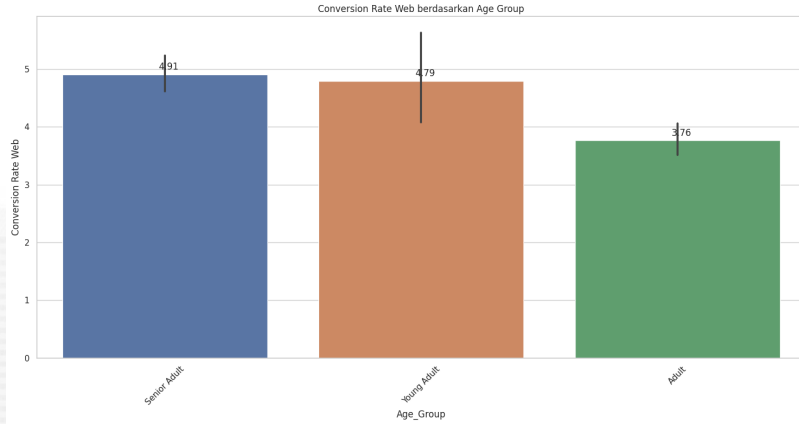


## Insight:

Pada grafik scatter plot menunjukkan korelasi positif, dimana semakin tinggi pendapatan maka semakin tinggi nilai conversion rate, oleh karena itu pelanggan dengan pendapatan yang tinggi cenderung lebih merespons kampanye.

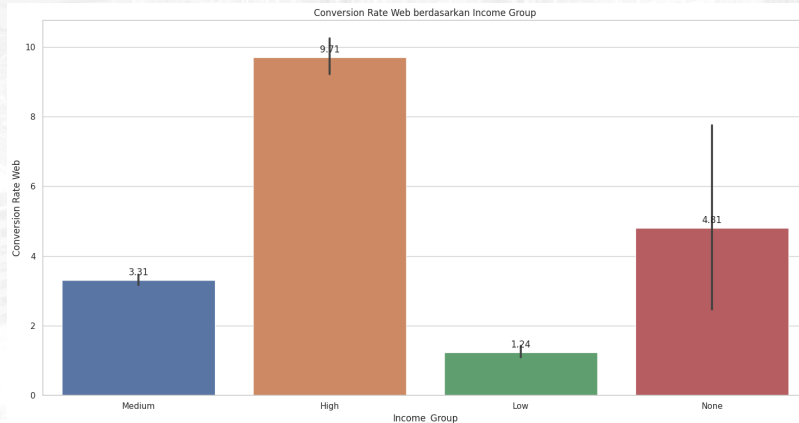


# Conversion Rate Analysis Berdasarkan Age dan Income Group



## Insight :

Pada diagram bar kelompok usia 'Senior Adults (> 45 tahun)' menunjukkan nilai conversion rate web yang paling tinggi, berbeda sedikit dengan kelompok 'Young Adults (17-30 tahun)' dan yang paling rendah adalah kelompok 'Adults (31-45 tahun)'. Maka kelompok 'senior adults (> 45 tahun)' dan 'young adults (17-30 tahun)' lebih banyak merespon campaign dibandingkan kelompok 'adults (31-45 tahun)'



## Insight :

Pada diagram bar menunjukkan bahwa kelompok customer dengan pendapatan tinggi (High) memiliki nilai conversion rate yang paling tinggi, dan nilai conversion rate yang paling rendah terdapat pada kelompok pelanggan dengan pendapatan yang rendah (Low).

## Rekomendasi Bisnis:

- **Penargetan Lebih Spesifik:** Fokuskan upaya pemasaran dan kampanye pada kelompok usia 'Senior Adults (> 45 tahun)' dan 'Young Adults (17-30 tahun)', karena mereka menunjukkan respon yang lebih baik terhadap campaign. Perusahaan dapat mengalokasikan lebih banyak anggaran dan sumber daya pada pemasaran yang ditujukan khusus untuk kelompok usia ini.
- **Penyelarasan Konten:** Pastikan konten pemasaran dan kampanye yang dibuat sesuai dengan preferensi dan kebutuhan kelompok usia 'Senior Adults (> 45 tahun)'. Perhatikan apa yang menarik bagi kelompok usia ini dan ciptakan konten yang relevan dan menarik bagi mereka.
- **Penargetan Berdasarkan Pendapatan:** Fokuskan upaya pemasaran dan kampanye pada kelompok pelanggan dengan pendapatan tinggi. Identifikasi dan targetkan segmen pasar yang memiliki potensi tinggi untuk merespon kampanye perusahaan dengan baik. Pelanggan dengan pendapatan tinggi cenderung lebih mampu melakukan pembelian dan memiliki nilai conversion rate yang lebih tinggi.
- **Peningkatan Retensi Pelanggan:** Fokus pada retensi pelanggan dengan pendapatan tinggi. Pastikan untuk memberikan layanan yang memuaskan dan mempertahankan hubungan baik dengan pelanggan setia. Pelanggan dengan pendapatan tinggi cenderung menjadi pelanggan berulang jika mereka merasa dihargai dan mendapatkan nilai yang baik dari bisnis perusahaan.
- **Targeting Pelanggan Potensial:** Identifikasi pelanggan potensial yang dapat memberikan nilai conversion rate yang tinggi. Gunakan data dan analisis untuk mengetahui profil pelanggan yang cenderung merespon pemasaran dengan baik, dan fokuskan pemasaran pada kelompok pelanggan ini.
- **Strategi Penawaran:** Pertimbangkan untuk menyusun strategi penawaran yang menarik untuk mendorong pelanggan untuk mengambil tindakan pembelian. Diskon khusus, penawaran eksklusif, atau paket bundle dapat meningkatkan nilai conversion rate.

## Missing Value

- Dapat diketahui bahwa terdapat data kosong pada kolom income sebanyak 24 baris dengan persentase sebesar 1,0796 dari keseluruhan data.
- Pada proses handling missing values kita menggunakan Imputation using MICE with LightGBM

## Duplicate Rows

- Tidak ditemui baris data yang memiliki duplikat
- Tidak ditemukan adanya ID Customer yang sama

## Feature Encoding

Label encoding pada feature education, age group, income group, education level, living status, dan marital status menggunakan metode mapping

## Invalid Value

- kolom Dt\_Customer sebelumnya masih berbentuk string/object sehingga diubah menjadi Datetime untuk diolah pada tahap Feature Engineering
- Menghapus column1 karena tidak kita gunakan datanya

## Outliers

Handling outlier dengan Manually Trimmed :

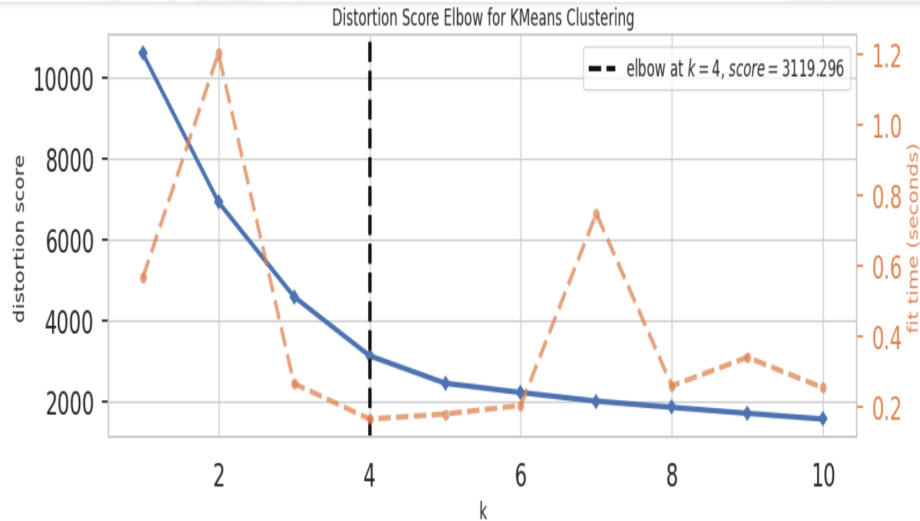
- Kolom Year\_Birth, menghapus nilai yang sangat jauh di tahun 1893-1900
- Kolom Income menghapus nilai yang sangat tinggi sebesar 666M

## Feature Transformation

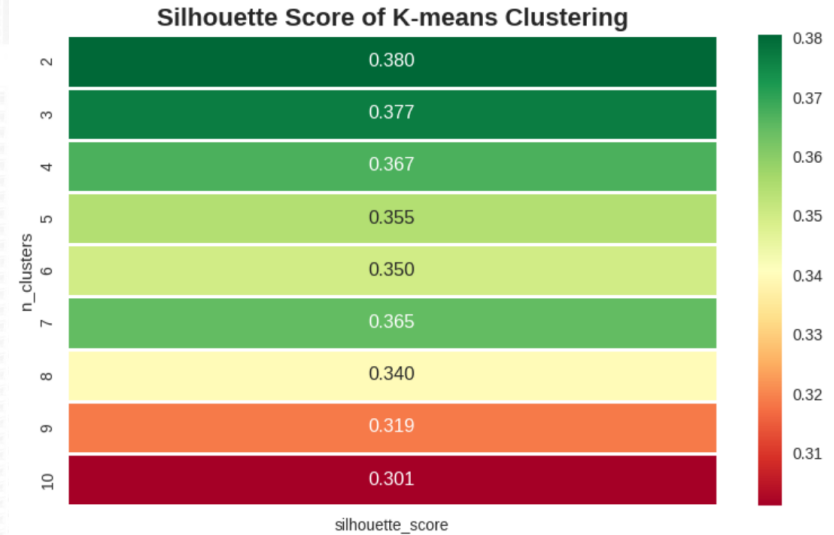
- Log Transformation dengan Yeo Johnson Transformation
- Melakukan Normalization



## Elbow Method K-Means Clustering



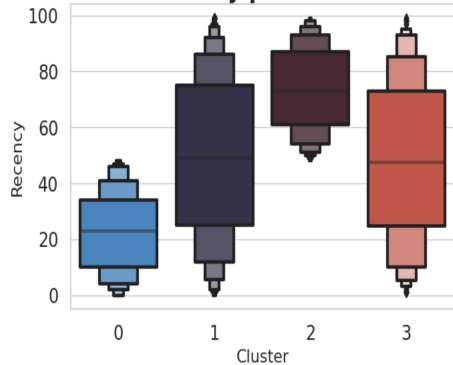
## Silhouette Score



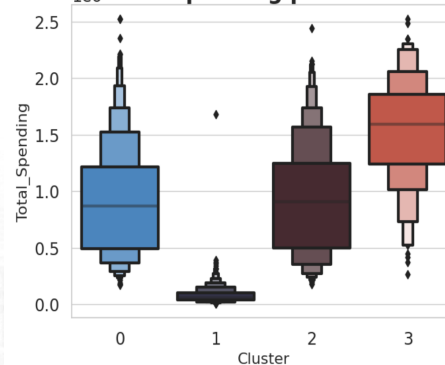
Berdasarkan grafik di atas, dengan parameter distorsi nilai optimal adalah  $k = 4$

# Customer Personality Analysis for Marketing Retargeting

Recency per Cluster



1e6 Total Spending per Cluster



## Recency per cluster :

- Recency terendah ada pada klaster 0 (waktu sejak pembelian terakhir baru-baru ini).
- Recency tertinggi ada pada klaster 2 (pelanggan dalam klaster ini belum melakukan pembelian dalam waktu yang lama).

## Total Purchases per cluster :

- Total pembelian terendah ada pada klaster 1.
- Total pembelian di klaster 0, 2, dan 3 berada dalam kategori tinggi.

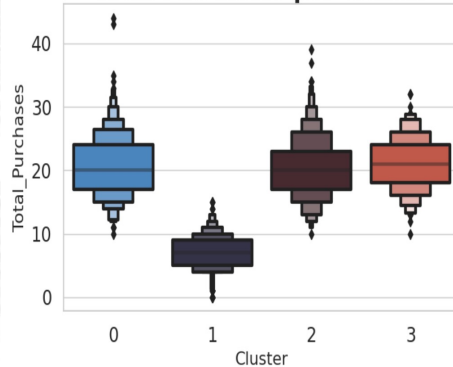
## Total Spending per cluster :

- Total pengeluaran terendah ada pada klaster 1.
- Total pengeluaran tertinggi ada pada klaster 3, diikuti oleh klaster 1, dan kemudian klaster 2.

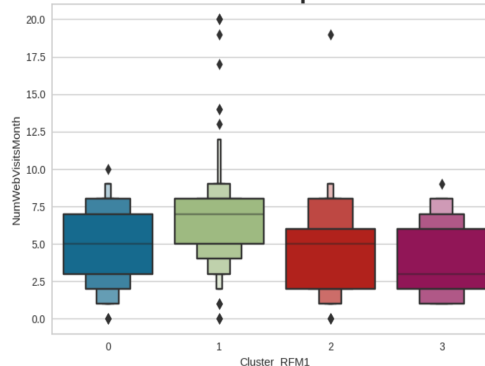
## Total Visit Website per cluster :

- Total visit website terendah ada pada klaster 3.
- Total visit website tertinggi ada pada klaster 1.
- Total visit website klaster 0 dan 2 adalah sama.

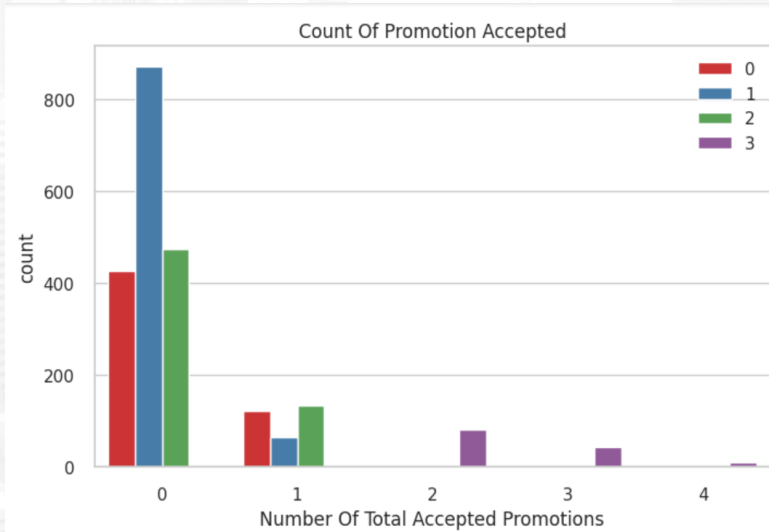
Total Purchases per Cluster



Total Visit Web per Cluster

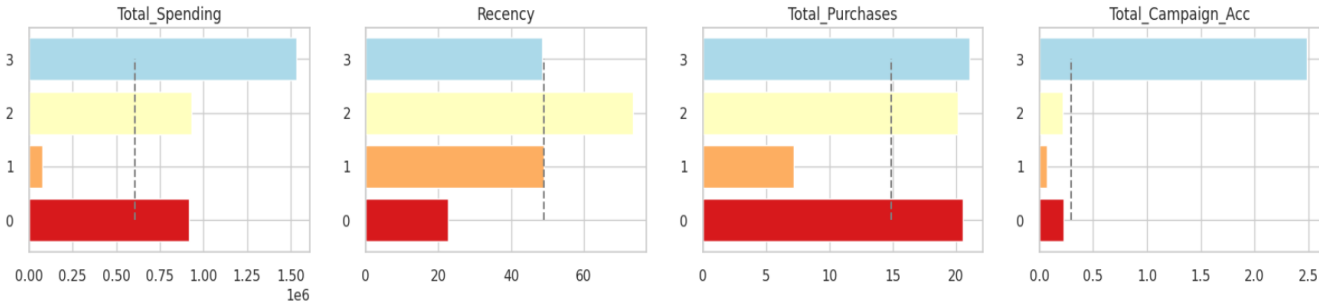


# Customer Personality Analysis for Marketing Retargeting



Belum ada respons yang terlihat jelas pada kampanye ini. Jumlah pelanggan yang terlibat sangat sedikit secara keseluruhan. Selain itu, tidak ada bagian yang mendapatkan semua dalam lima kali promosi. Mungkin diperlukan kampanye yang lebih ditargetkan dan direncanakan dengan baik untuk meningkatkan penjualan.

# Customer Personality Analysis for Marketing Retargeting



## 1. Cluster 0 (Loyal Customer):

- Kelompok ini adalah kelompok dengan jumlah pelanggan ketiga terbesar sebanyak 550 orang
- Pelanggan di kelompok ini mempunyai rata-rata waktu sejak pembelian terakhir (22 hari), dan rata-rata total pembelian (21 items) itu berarti mereka berbelanja dalam waktu yang dekat, sering berbelanja dan mereka menghabiskan banyak uang di platform kami (sekitar Rp 919k/tahun)
- Pelanggan dalam kelompok ini kurang dalam merespon kampanye tetapi kelompok ini mengunjungi website cukup sering dengan median sebanyak 5 kali dalam sebulan

## 2. Cluster 1 (Need Attention):

- Kelompok ini adalah kelompok dengan jumlah pelanggan terbesar sebanyak 941 orang
- pelanggan di kelompok ini mempunyai rata-rata waktu sejak pembelian terakhir (48 hari), dan rata-rata total pembelian (7 items) itu berarti mereka sudah cukup lama tidak berbelanja, bukan pembeli rutin dan mereka menghabiskan sedikit uang di platform kami (sekitar Rp 79k/tahun)
- Pelanggan dalam kelompok ini paling sedikit dalam merespon kampanye tetapi kelompok ini mengunjungi website paling sering dengan median sebanyak 7 kali dalam sebulan



### 3. Cluster 2 (At Risk):

- Kelompok ini adalah kelompok dengan jumlah pelanggan kedua terbesar sebanyak 609 orang
- Pelanggan di kelompok ini mempunyai rata-rata waktu sejak pembelian terakhir (73 hari) yang berarti paling lama tidak berbelanja di platform kami dibanding kelompok lainnya, dan rata-rata total pembelian (20 items) itu berarti mereka sering berbelanja dan menghabiskan banyak uang di platform kami (sekitar Rp 930k/tahun)
- Pelanggan dalam kelompok ini kurang dalam merespon kampanye tetapi kelompok ini mengunjungi website cukup sering dengan median sebanyak 5 kali dalam sebulan

### 4. Cluster 3 (Potential Loyalist):

- Kelompok ini adalah kelompok dengan jumlah pelanggan terkecil sebanyak 136 orang
- Pelanggan di kelompok ini mempunyai rata-rata waktu sejak pembelian terakhir (48 hari), dan rata-rata total pembelian (21 items) itu berarti mereka sering berbelanja dan menghabiskan paling banyak uang di platform kami (sekitar Rp 1.5jt/tahun) dibanding kelompok lain
- Pelanggan dalam kelompok ini yang paling banyak merespon kampanye yaitu sebanyak 2 kali tetapi kelompok ini mengunjungi website hanya 3 kali dalam sebulan

## Business Recommendation:

### 1. Cluster 0 (Loyal Customer):

- **Pertahankan Keterlibatan:** Lanjutkan komunikasi secara teratur melalui kampanye pemasaran yang relevan dan menarik.
- **Penawaran Khusus:** Berikan penawaran eksklusif atau diskon kepada pelanggan sebagai hadiah atas keterlibatan mereka.
- **Program Hadiah:** Pertimbangkan untuk memperkenalkan program hadiah yang menawarkan insentif untuk partisipasi aktif dan pembelian.

### 2. Cluster 1 (Need Attention Customer):

- **Reaktivasi:** Kirim tawaran menarik atau diskon untuk mendorong pelanggan berbelanja kembali.
- **Penawaran Khusus:** Berikan penawaran khusus atau diskon terbatas untuk merangsang pembelian lebih lanjut.
- **Konten Edukatif:** Kirim konten yang memberikan informasi berharga tentang produk atau metode penggunaan untuk meningkatkan minat dan keterlibatan.

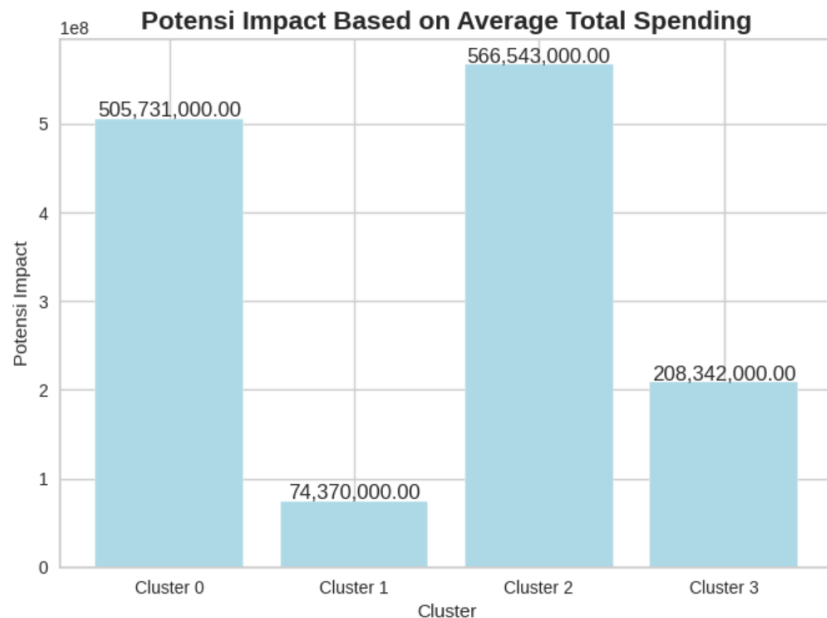
## Business Recommendation:

### 3. Cluster 2 (At Risk Customer):

- **Kampanye Personalisasi:** Upayakan untuk mengirimkan kampanye yang lebih relevan dan menarik kepada pelanggan dalam kelompok ini.
- **Penawaran Eksklusif:** Berikan penawaran eksklusif atau diskon khusus kepada pelanggan dalam kelompok ini. Ini dapat menjadi insentif tambahan bagi mereka untuk berbelanja kembali meskipun jarang menerima kampanye pemasaran.
- **Program Loyalitas:** Ajak mereka untuk bergabung dalam program loyalitas yang menawarkan manfaat eksklusif bagi pelanggan setia.
- **Eksperimen dengan Media dan Saluran Berbeda:** Cobalah menguji kampanye melalui media atau saluran yang berbeda. Beberapa pelanggan mungkin merespons lebih baik jika kampanye disampaikan melalui email, SMS, media sosial, atau periklanan online. Eksperimen ini dapat membantu kami menemukan saluran komunikasi yang paling efektif untuk kelompok ini.

### 4. Cluster 3 (Potential Loyalist Customer):

- **Kampanye Personalisasi:** Tingkatkan upaya untuk mengirim kampanye pemasaran yang lebih personal dan relevan kepada pelanggan dalam klaster ini.
- **Penawaran Diskon:** Berikan penawaran khusus atau diskon kepada pelanggan untuk mendorong pembelian lebih lanjut.
- **Program Loyalitas:** Ajak pelanggan untuk bergabung dengan program loyalitas yang menawarkan insentif tambahan untuk pembelian tinggi.
- **Pemberitahuan Produk Baru:** Informasikan mereka tentang produk-produk baru yang sejalan dengan minat dan preferensi mereka.



## Potential Impact:

Jika kita bisa memprioritaskan pada semua kelompok dan mereka tidak beralih (churn), kita masih memiliki potensi GMV sekitar Rp 1,3 miliar/tahun

- Loyal Customer = IDR 505jt/tahun
- Need Attention Customer = IDR 74jt/tahun
- At Risk Customer = IDR 566jt/tahun
- Potential Loyalist Customer= IDR 208jt/tahun