

# Customer Segmentation and Profiling by Modified LRFM Analysis using CLARA Algorithm to Optimize Marketing Strategy

## DATAVANGERS

MENTOR



Kak Bachtiyar

FASIL



Kak Renita



Andri Maulana



Nadia  
Adyutarahma P.



Nabila Amina R.



Syifa Mufidah



Rahmad  
Kholid



Satria  
Imawan



# SUPERAPP

Persaingan di industri yang semakin ketat menuntut perusahaan untuk terus **berinovasi** agar dapat **menarik** dan **mempertahankan pelanggan**. Strategi yang digunakan perusahaan adalah menawarkan berbagai bentuk promosi, salah satunya dengan pemberian **DISKON**.

Untuk **mengoptimalkan** kinerja perusahaan dalam **menyusun strategi promosi**, perlu dilakukan **pemahaman perilaku pelanggan** lebih mendalam. Dengan begitu, diharapkan strategi promosi yang dilakukan tepat sasaran sehingga akan memberikan dampak positif bagi perusahaan ke depannya.

>>





# Problem Analysis

## Discount

| Month of 1st Purchase | Month After 1st Purchase |         |        |        |        |        |        |        |        |        |        |        |
|-----------------------|--------------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|                       | 0                        | 1       | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 10     | 11     |
| Jan                   | 100.00%                  | 90.88%  | 90.91% | 91.30% | 90.91% | 91.15% | 91.00% | 90.77% | 90.87% | 91.00% | 90.88% | 90.95% |
| Feb                   | 100.00%                  | 89.92%  | 88.36% | 89.27% | 89.92% | 89.44% | 89.87% | 89.23% | 89.10% | 89.61% | 89.74% |        |
| Mar                   | 100.00%                  | 86.11%  | 83.73% | 85.32% | 86.11% | 84.52% | 88.49% | 84.13% | 85.32% | 83.33% |        |        |
| Apr                   | 100.00%                  | 73.91%  | 82.61% | 76.09% | 69.57% | 78.26% | 86.96% | 76.09% | 80.43% |        |        |        |
| May                   | 100.00%                  | 65.00%  | 65.00% | 85.00% | 85.00% | 70.00% | 65.00% | 75.00% |        |        |        |        |
| Jun                   | 100.00%                  | 60.00%  | 53.33% | 80.00% | 60.00% | 66.67% | 60.00% |        |        |        |        |        |
| Jul                   | 100.00%                  | 70.00%  | 70.00% | 70.00% | 70.00% | 64.29% |        |        |        |        |        |        |
| Aug                   | 100.00%                  | 87.50%  | 75.00% | 75.00% | 87.50% |        |        |        |        |        |        |        |
| Sep                   | 100.00%                  | 100.00% | 57.14% | 85.71% |        |        |        |        |        |        |        |        |
| Oct                   | 100.00%                  | 66.67%  | 77.78% |        |        |        |        |        |        |        |        |        |
| Nov                   | 100.00%                  | 75.00%  |        |        |        |        |        |        |        |        |        |        |
| Dec                   | 100.00%                  |         |        |        |        |        |        |        |        |        |        |        |

## Non-Discount

| Month of 1st Purchase | Month After 1st Purchase |        |        |        |        |        |        |        |        |        |        |        |
|-----------------------|--------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|                       | 0                        | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | 10     | 11     |
| Jan                   | 100.00%                  | 91.04% | 90.81% | 90.89% | 90.95% | 90.81% | 90.96% | 90.84% | 91.02% | 90.92% | 90.98% | 90.95% |
| Feb                   | 100.00%                  | 89.06% | 88.63% | 88.46% | 87.64% | 88.81% | 88.59% | 89.16% | 88.95% | 88.38% | 89.67% |        |
| Mar                   | 100.00%                  | 83.83% | 84.98% | 84.98% | 84.82% | 85.97% | 84.32% | 86.63% | 83.99% | 84.32% |        |        |
| Apr                   | 100.00%                  | 84.40% | 84.40% | 79.43% | 82.98% | 77.30% | 78.72% | 84.40% | 79.43% |        |        |        |
| May                   | 100.00%                  | 70.45% | 75.00% | 68.18% | 72.73% | 70.45% | 68.18% | 65.91% |        |        |        |        |
| Jun                   | 100.00%                  | 70.00% | 65.00% | 72.50% | 75.00% | 70.00% | 72.50% |        |        |        |        |        |
| Jul                   | 100.00%                  | 80.00% | 80.00% | 80.00% | 80.00% | 80.00% | 80.00% |        |        |        |        |        |
| Aug                   | 100.00%                  | 74.29% | 82.86% | 80.00% | 82.86% |        |        |        |        |        |        |        |
| Sep                   | 100.00%                  | 68.18% | 77.27% | 81.82% |        |        |        |        |        |        |        |        |
| Oct                   | 100.00%                  | 89.29% | 75.00% |        |        |        |        |        |        |        |        |        |
| Nov                   | 100.00%                  | 74.07% |        |        |        |        |        |        |        |        |        |        |
| Dec                   | 100.00%                  |        |        |        |        |        |        |        |        |        |        |        |

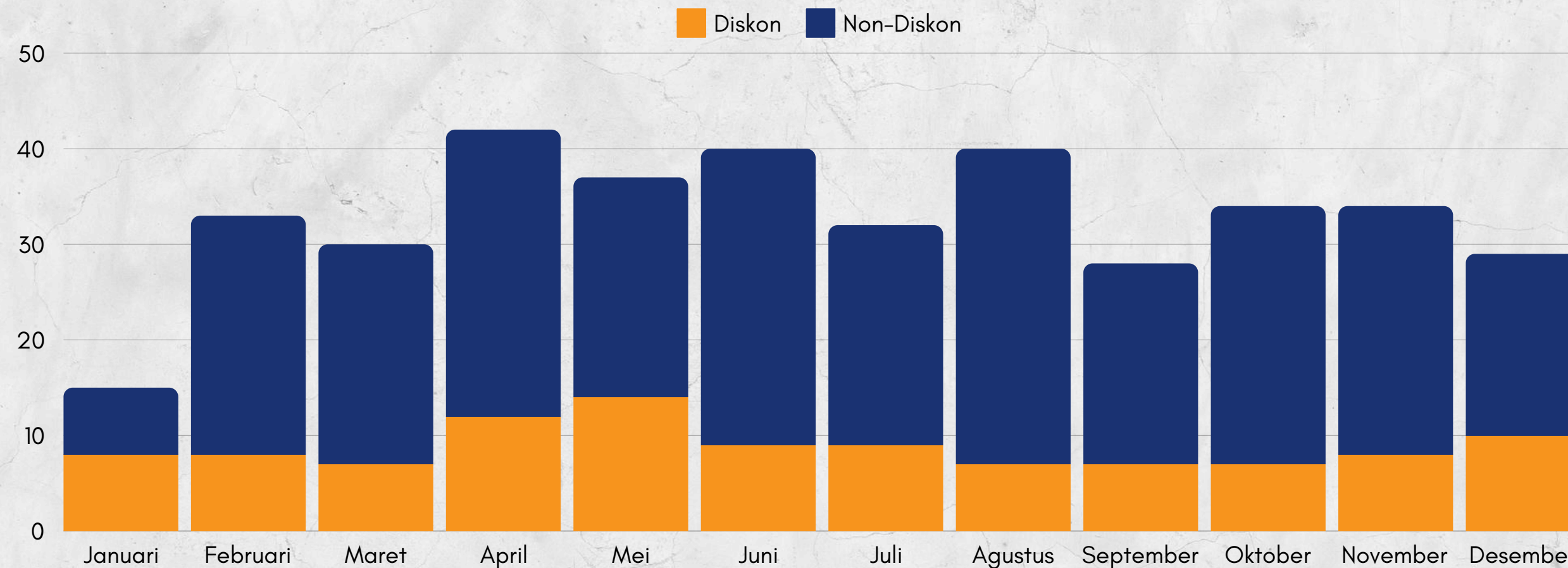
Pengguna dengan transaksi pertama **menggunakan DISKON** memiliki **retensi yang seimbang/ terkadang lebih buruk** dibandingkan dengan pengguna dengan transaksi pertama **tanpa DISKON**.

Promosi **DISKON** belum **optimal** dalam **mempertahankan retensi pelanggan** dalam **jangka panjang**.



# Problem Analysis

## Proporsi Pengguna Baru 2024 Berdasarkan Tipe Diskon pada Transaksi Pertama



Pengguna dengan transaksi pertama **menggunakan DISKON** < pengguna dengan transaksi pertama **tanpa DISKON**

Promosi **DISKON** belum optimal dalam **menarik pelanggan baru**.



# Focus Target



## PROBLEM

Promosi **diskon** yang sudah ada **belum optimal** dalam menarik pelanggan **baru** dan mempertahankan **retensi** pelanggan dalam jangka panjang.



## SOLUTION

Analisis strategi promosi yang lebih terpersonalisasi dengan **mengelompokkan pelanggan** berdasarkan **pola perilaku pembelian** menggunakan metrik **LRFM** (Length, Recency, Frequency, Monetary) yang **dimodifikasi**.

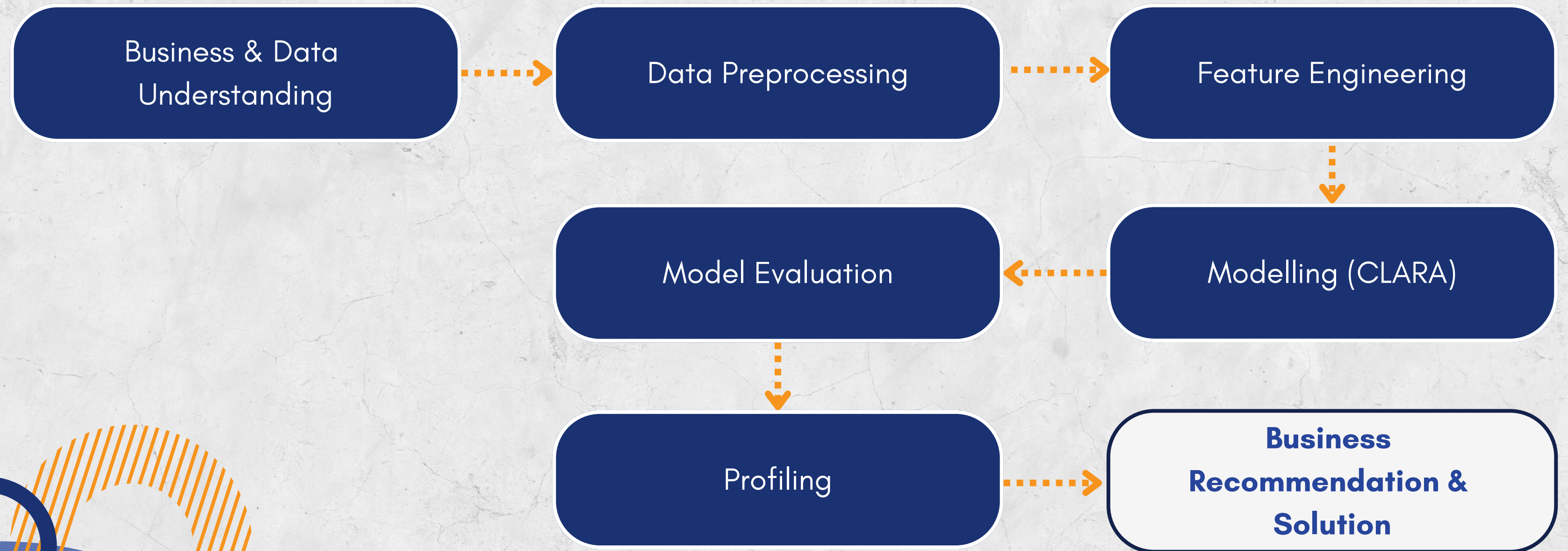


## GOALS

Memahami pola perilaku pembelian pelanggan untuk mengoptimalkan **strategi promosi** agar **tepat sasaran** sehingga meningkatkan **retensi** dan mempertahankan **loyalitas** pelanggan yang ada.



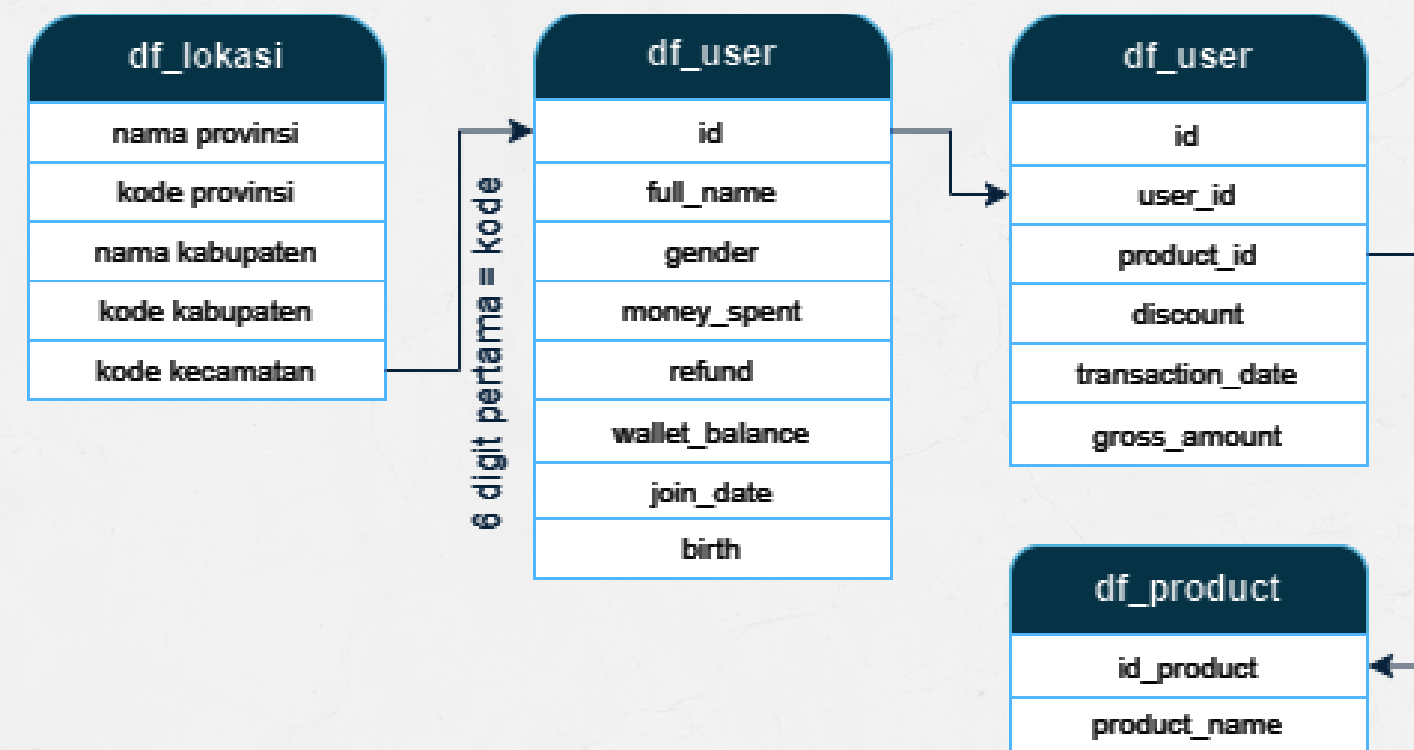
# Workflow





# Data Preprocessing

## Data Merges



## Data Formatting

### Standarisasi Nama Kolom

### Konversi Tipe Data

- menyesuaikan tipe data kolom sesuai deskripsi

## Data Cleaning

### Missing Value

- gross\_amount = dihapus (**0,57%**)
- discount = diganti dengan nilai 0 (**69,83%**)

### Duplicated Data

### Invalid Data

- pelanggan yg bertransaksi sebelum mendaftar = dihapus (**4.991 baris**)
- baris bernilai negatif pada kolom gross\_amount, discount, refund, dan wallet\_balance = dihapus (**0,80%**)
- pelanggan yg tidak pernah bertransaksi = dihapus

### Outlier

- dibiarkan -> mempertahankan informasi transaksi yg berharga (gross\_amount = **9.91%**; discounts = **15.95%**)



# Feature Engineering

## Why Metrik LRFM?

untuk menargetkan strategi promosi agar tepat sasaran, mengidentifikasi pelanggan yg beresiko churn dan pelanggan yg paling berharga

## Enhance Cluster Result

### Heatmap Analysis

Melihat korelasi antar feature clustering

### Modifikasi Feature

- Tiga Variable Clustering: efisiensi algoritma clustering dan kemudahan visualisasi hasil model
- Recency Score: menurunkan tingkat korelasi antar variable (heatmap analysis)

### Feature Scalling (MinMax)

Menangani perbedaan skala yang signifikan

## Feature Importance

### Length

Selisih tanggal terakhir transaksi dengan tanggal pertama transaksi (day)

### Recency Score (1/R)

Recency = selisih tanggal terakhir transaksi dengan tanggal pertama transaksi

### Monetary per Frequency (M/F)

Rata-rata uang yang dibelanjakan pelanggan tiap transaksi

Monetary = total uang yg dibelanjakan pelanggan

Frequency = total transaksi pelanggan



# Modelling (CLARA)

| Algoritma | Robust Outlier | Dataset Besar | Time Complexity | Low Dimension |
|-----------|----------------|---------------|-----------------|---------------|
| K-Means   | ✗              | ✓             | Low             | ✓             |
| K-Medoid  | ✓              | ✗             | High            | ✓             |
| CLARA     | ✓              | ✓             | Middle          | ✓             |

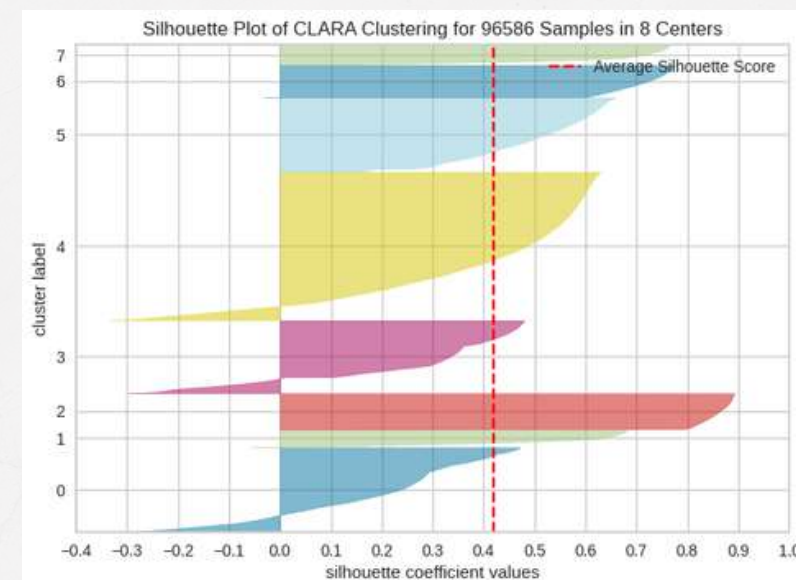
Zhang, Huang, Niu, et al. (2023)

## Model Algorithm CLARA

- Modifikasi dari algoritma PAM (Partitioning Around Medoids)
- **Cara kerja:** memilih medoid sebagai pusat cluster, mengelompokkan data disekitarnya, menggunakan metode sampling (efisiensi untuk data besar)
- Parameter default: n\_clusters = 8, init = 'build', n\_sampling = None

\*) medoid: titik aktual dalam dataset

## Hasil Clustering Parameter Default





# Model Evaluation

## Kombinasi Parameter Algorithm CLARA

'n\_clusters': [2, 3, 4, 5, 6, 7, 8]  
'n\_sampling': [None, 100]  
'init': ['random', 'heuristic', 'k-medoids++', 'build']

'n\_clusters': [3, 4],  
'n\_sampling': [300],  
'init': ['random', 'k-medoids++']

'n\_clusters': [3, 4, 5],  
'n\_sampling': [300, 500, 1000],  
'init': ['k-medoids++']

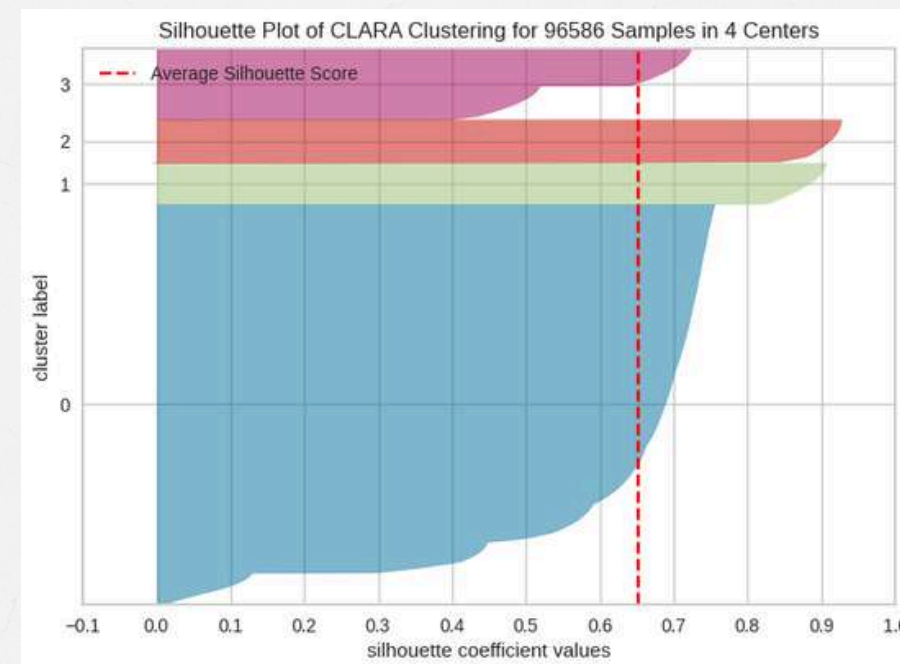
## Kriteria Keputusann

- Nilai silhouette tinggi (mendekati 1)
- Semua cluster harus memiliki nilai silhouette yg lebih besar dari rata-rata nilai dataset
- Jumlah nilai negatif sedikit
- Persebaran pembagian cluster merata (tdk ada fluktuasi besar)

## Silhouette Score

Nilai 0 = terdapat cluster overlapping  
Nilai -1 = objek berada di kelompok yg salah  
Nilai 1 = obje berada di cluster yg sesuai

## Kombinasi Parameter Terbaik



n\_cluster : 4  
n\_sampling : 300  
init : k-medoids++  
random state : 42

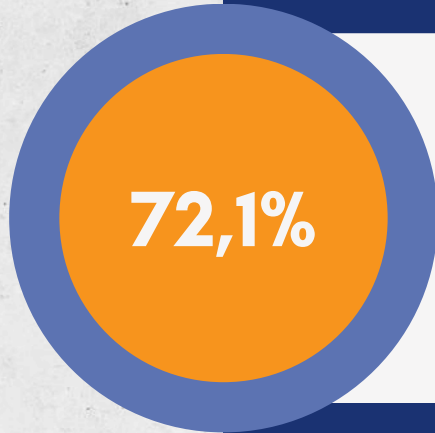
**Nilai Silhouette = 0,652**



# Profiling

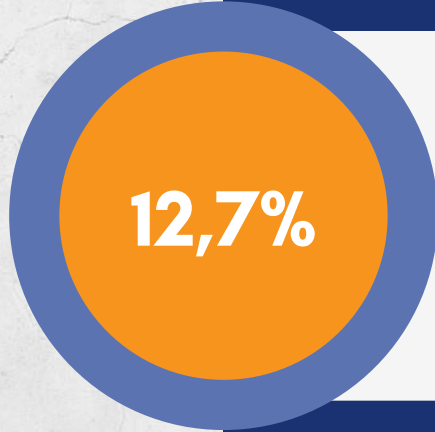
|                    | L | 1/R | M/F |
|--------------------|---|-----|-----|
| At Risk            | 4 | 1   | 3   |
| Loyal              | 4 | 5   | 3   |
| Potential Loyalist | 4 | 5   | 2   |
| Lost               | 1 | 1   | 1   |

L = Length  
1/R = Recency Score  
M/F = Monetary per Frequency



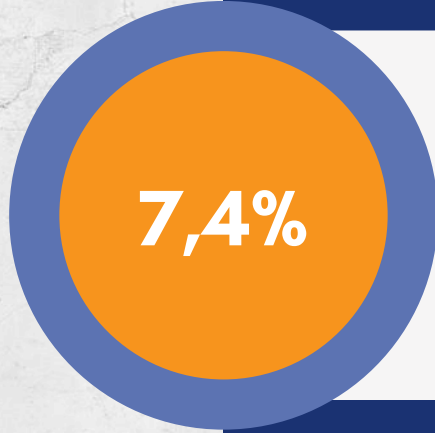
## At Risk (413)

Long-lasting customers who spent medium amounts of money per transaction a long time ago



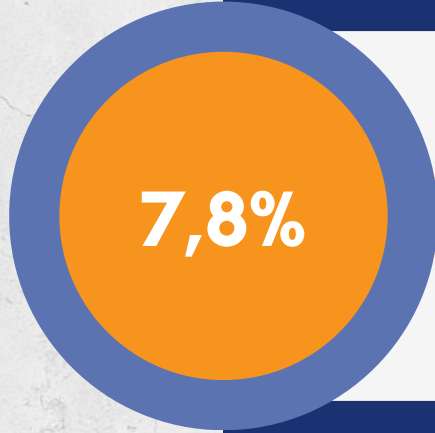
## Loyal (453)

Long-lasting customers who spent medium amounts of money per transaction most recently



## Potential Loyalist (452)

Long-lasting customers who spent small amounts of money per transaction most recently



## Lost (111)

Extremely short-term customers who spent very small amounts of money per transaction a long time ago



# Business Recommendation



## At Risk

*Re-engage with personalized offers and nostalgic campaigns*

- **Kampanye nostalgia** => membangkitkan pengalaman positif & menciptakan hubungan emosional untuk mendorong mereka kembali.
- **Diskon khusus** => menghargai **kesetiaan** mereka.

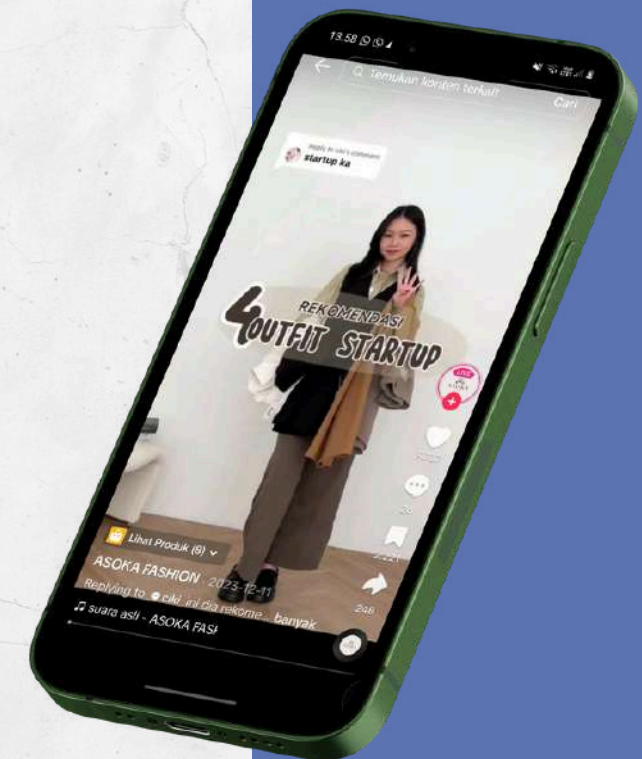


## Potential Loyalist

*Encourage repeat purchases through subscription models, free samples, and customer education*

- **Edukasi** pelanggan dengan **membuat konten** yang relevan sesuai **preferensi** mereka, contohnya konten tren fashion wanita, tips berpakaian
- Mengirim **pesan** kepada pelanggan yang berisi **informasi produk baru, ulasan pelanggan lain**, atau **manfaat produk**.

"Kami merindukan Anda! Nikmati diskon 20% untuk semua produk Food & Drink favorit Anda dan nikmati momen-momen berharga bersama kami lagi!."





# Business Recommendation



## Lost

*Reactivate with targeted product recommendations and limited-time offers*

- Penawaran **eksklusif** berupa **diskon besar** dengan harapan pelanggan yang hilang akan tertarik kembali dan melakukan pembelian lagi.
- **Penawaran hanya dalam waktu terbatas** agar pelanggan terdorong untuk mengambil tindakan kembali.



## Loyal

*Reward loyalty with community building, early access programs, and customer appreciation initiatives*

- Memberikan **perlakuan khusus** agar pelanggan merasa dihargai, contohnya memberikan **akses awal ke produk baru** atau penawaran **member eksklusif** (VIP).

"Kami ingin Anda kembali! Dapatkan diskon 55% untuk perjalanan pertama Anda dengan Ride Hailing. Ayo, coba lagi dan nikmati perjalanan dengan nyaman dan hemat! Ayo tunggu apa lagi? Hanya berlaku HARI INI!"


"Sebagai pelanggan setia, Anda mendapatkan akses pertama ke koleksi terbaru kami dan undangan untuk bergabung dengan program VIP kami dengan manfaat eksklusif."





# Reference

Zhang, C., Huang, W., Niu, T. et al. Review of Clustering Technology and Its Application in Coordinating Vehicle Subsystems. Automot. Innov. 6, 89-115 (2023).  
<https://doi.org/10.1007/s42154-022-00205-0>







**THANK YOU**

**DATAVENGERS**