

Penerapan Algoritma Clustering untuk Segmentasi Pelanggan E-Commerce pada Proses Pengambilan Keputusan

Decision Making



Presented By
Kelompok 7

i



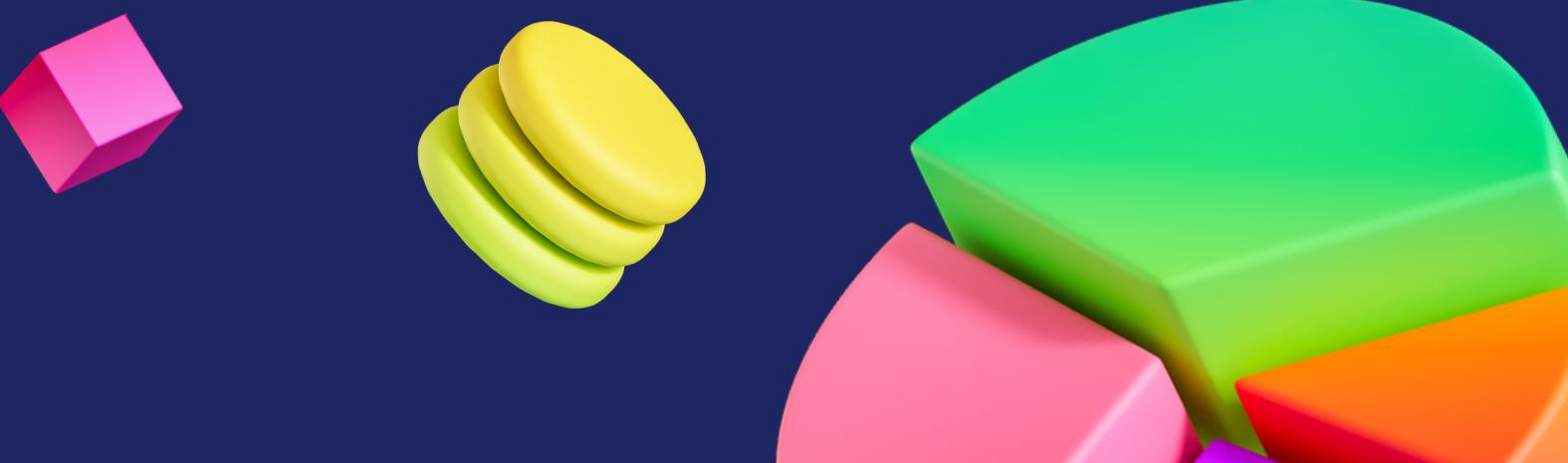


Decision Making



KELOMPOK 7

- Shahnaz Salsabila Ishak
- Kathrine Brigita Soraya
- Angga Pramana Putra Wibowo
- Mario Desendi Manurung
- M. Zaky Mahdavikia Zein
- Lisnurani



AGENDA



- 1** Latar Belakang
- 2** Sub-topik
- 3** Dataset
- 4** Data Cleaning and Exploration
- 5** Feature Engineering

LATAR BELAKANG

Dalam era yang semakin terdigitalisasi ini, pengambilan keputusan dalam bidang teknologi memiliki peran yang sangat penting dalam memastikan kesuksesan suatu organisasi. Teknologi informasi menjadi salah satu pilar utama yang membentuk efisiensi, daya saing, dan keunggulan suatu perusahaan. Untuk itu, tiga sub-topik kunci yang kami pilih, yaitu Evaluasi Teknologi Baru, Perencanaan Implementasi, Uji Coba, dan Evaluasi.



Sub Topik

**Evaluasi
Teknologi
Baru**

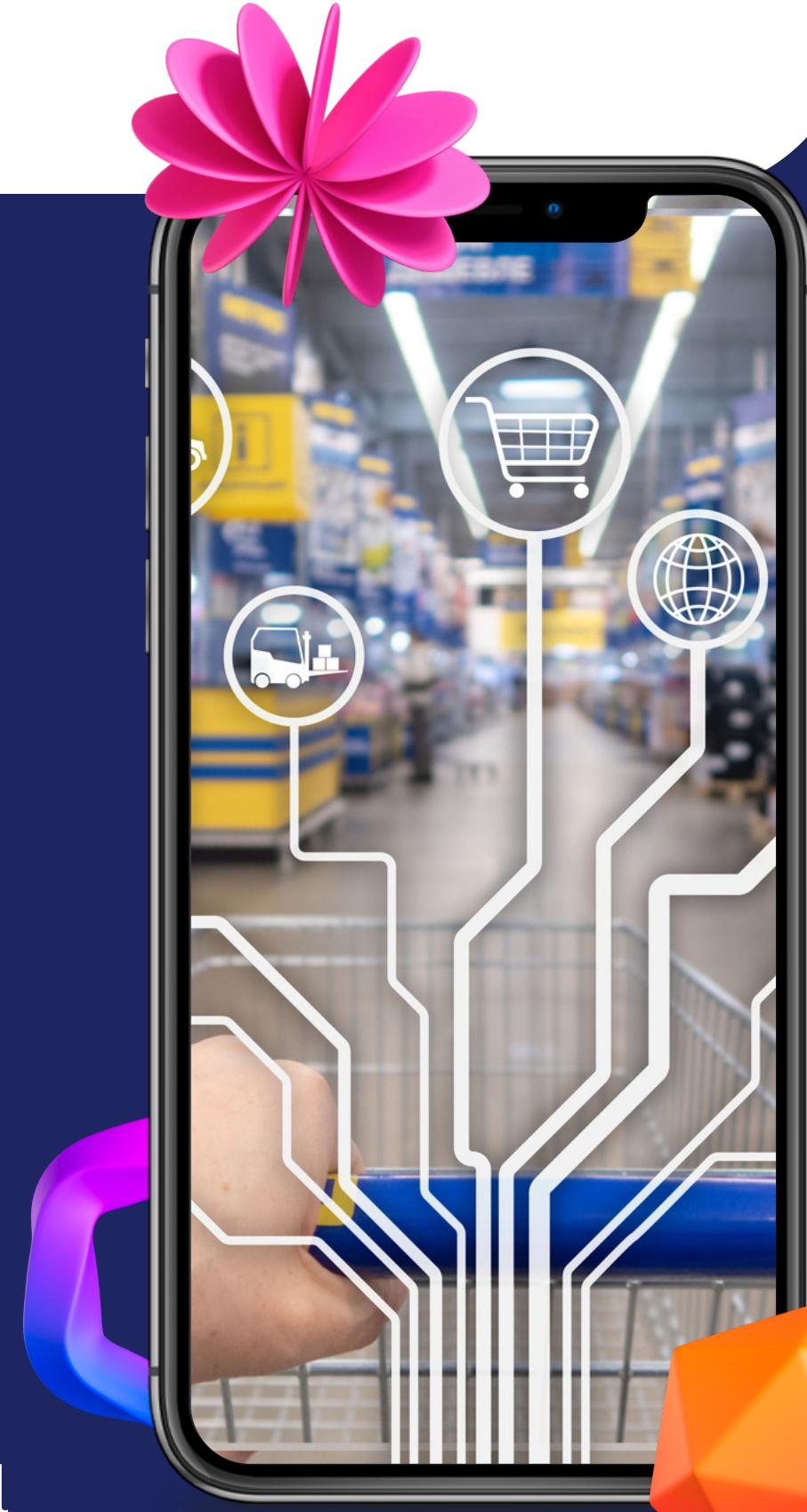
Pembaruan teknologi diperlukan untuk menghasilkan segmentasi yang lebih akurat dan relevan untuk memberikan pelayanan yang lebih baik

**Perencanaan
Implementasi**

Implementasi dari teknologi baru dalam melakukan segmentasi

**Uji Coba dan
Evaluasi**

Melakukan Uji Coba dan evaluasi dari hasil segmentasi untuk pengambilan keputusan



DATASET

Dataset Online Retail II berisi semua transaksi yang terjadi untuk perusahaan ritel online yang berbasis di Inggris dan terdaftar, tanpa toko fisik, antara tanggal 01/12/2009 dan 09/12/2011. Perusahaan ini utamanya menjual barang-barang hadiah serba guna yang unik.

Source Dataset : UCI Machine Learning



FITUR DATASET



Dataset terdiri dari 525461 baris dan 8 fitur

- Invoice: Nomor Faktur. Nominal. Nomor integral 6 digit yang diberikan secara unik untuk setiap transaksi. Jika kode ini dimulai dengan huruf 'C', itu menandakan pembatalan.
- StockCode: Kode Produk (barang). Nominal. Nomor integral 5 digit yang diberikan secara unik untuk setiap produk yang berbeda.
- Description: Nama Produk (barang). Nominal.
- Quantity: Jumlah dari setiap produk (barang) per transaksi. Numerik.
- InvoiceDate: Tanggal dan waktu Faktur. Numerik. Hari dan waktu ketika transaksi dihasilkan.
- UnitPrice: Harga per unit. Numerik. Harga produk per unit dalam mata uang pound sterling (£).
- Customer ID: Nomor Pelanggan. Nominal. Nomor integral 5 digit yang diberikan secara unik untuk setiap pelanggan.
- Country: Nama Negara. Nominal. Nama negara tempat pelanggan tinggal.

Dataset Preview

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

Statistics Summary

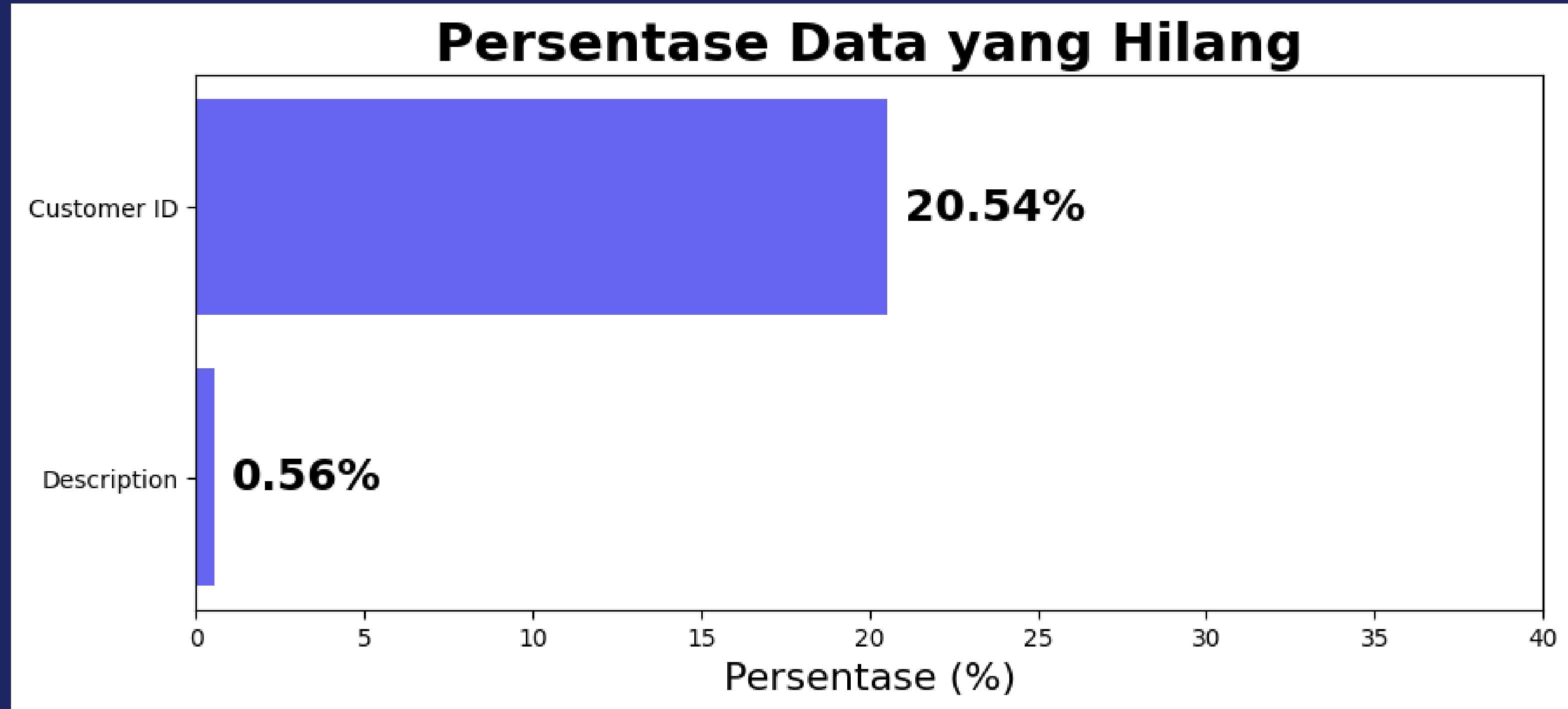
	Quantity	Price	Customer ID
count	525461.000000	525461.000000	417534.000000
mean	10.337667	4.688834	15360.645478
std	107.424110	146.126914	1680.811316
min	-9600.000000	-53594.360000	12346.000000
25%	1.000000	1.250000	13983.000000
50%	3.000000	2.100000	15311.000000
75%	10.000000	4.210000	16799.000000
max	19152.000000	25111.090000	18287.000000

- Dataset ini berisi beragam produk, tanpa satu produk atau nomor faktur yang dominan, dan sebagian besar pembelian berasal dari United Kingdom.
- Terdapat ketidakkonsistenan pada StockCode dan Description.

Rentang yang luas dan deviasi standar tinggi pada kedua kolom Kuantitas dan Harga menunjukkan variasi yang signifikan dalam fitur ini.

	Invoice	StockCode	Description	Country
count	525461	525461	522533	525461
unique	28816	4632	4681	40
top	537434	85123A	WHITE HANGING HEART T-LIGHT HOLDER	United Kingdom
freq	675	3516	3549	485852

Handle Missing Value



Handle Missing Value

- Missing data pada Customer ID dan Description akan dihapus
- Missing data pada CustomerID dihapus karena CustomerID merupakan fitur yang penting dan tidak bisa digantikan
- Missing data pada description dihapus karena tidak terlalu banyak missing value dan tidak bisa dilakukan penyesuaian dengan StockCode karena terdapat ketidakkonsistensi berdasarkan deskripsi statistik sebelumnya

```
Invoice          0  
StockCode       0  
Description     0  
Quantity        0  
InvoiceDate    0  
Price           0  
Customer ID    0  
Country         0  
dtype: int64
```

Remove Duplicate

```
[ ] # Jumlah Baris Duplikat  
df.duplicated().sum()
```

6771

```
▶ # Hapus Data Duplikat  
df.drop_duplicates(inplace=True)
```

```
[ ] # Dimensi data sekarang  
df.shape
```

(410763, 8)



Remove Anomaly

Kode barang anomali:

POST

D

M

C2

BANK CHARGES

TEST001

TEST002

PADS

ADJUST

ADJUST2

SP1002

- Terdapat anomali pada StockCode dengan total 11 kode menyimpang dari aturan penulisan StockCode yang seharusnya terdiri dari 5 karakter.
- Persentase kode barang yang anomali dalam dataset adalah: 0.44%
- Kode barang anomaly akan dihapus

Correct Anomaly

Deskripsi yang mengandung karakter huruf kecil :

BAG 500g SWIRLY MARBLES
POLYESTER FILLER PAD 40x40cm
POLYESTER FILLER PAD 60x40cm
POLYESTER FILLER PAD 65CMx65CM
POLYESTER FILLER PAD 45x45cm
BAG 125g SWIRLY MARBLES
ESSENTIAL BALM 3.5g TIN IN ENVELOPE
FOLK ART GREETING CARD, pack/12
BAG 250g SWIRLY MARBLES
POLYESTER FILLER PAD 30CMx30CM
POLYESTER FILLER PAD 45x30cm
SET OF 6 SOLDIER SKITTLES
THE KING GIFT BAG 25x24x12cm
FRENCH BLUE METAL DOOR SIGN, No
NUMBER TILE COTTAGE GARDEN, No
NUMBER TILE VINTAGE FONT, No
NUMBER TILE VINTAGE FONT No
FRENCH BLUE METAL DOOR SIGN No
NUMBER TILE COTTAGE GARDEN No
3 TRADITIONAL BISCUIT CUTTERS SET

- terdapat ketidakkonsistensi karena tanda baca seperti koma.
- Hapus tanda baca dan buat semua deskripsi menjadi huruf besar

Remove Anomaly

Transaksi dengan harga 0

	Quantity	Price	Customer ID
count	28.000000	28.0	28.000000
mean	30.678571	0.0	14078.142857
std	121.605024	0.0	1736.849931
min	1.000000	0.0	12417.000000
25%	1.000000	0.0	12647.000000
50%	4.000000	0.0	13321.500000
75%	10.500000	0.0	14752.000000
max	648.000000	0.0	18071.000000

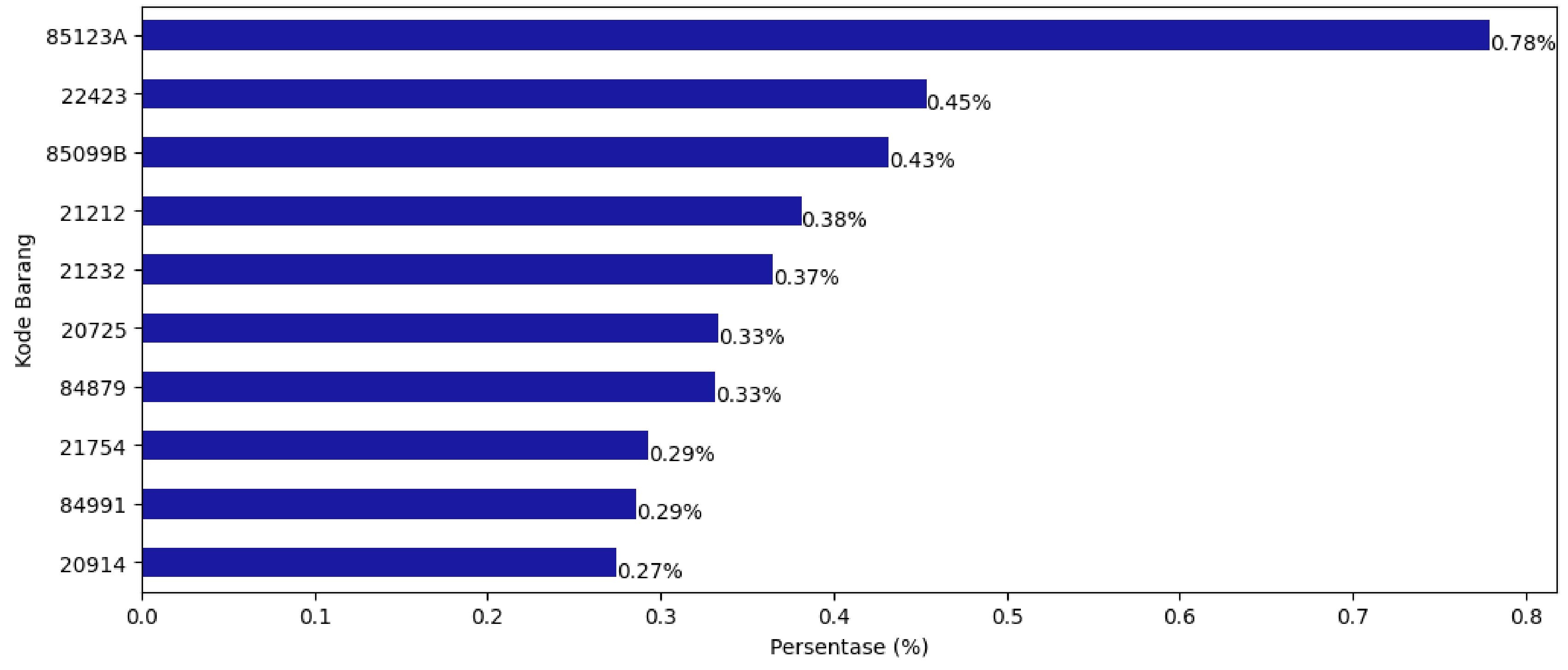
- terdapat 28 baris yang memiliki harga 0.
- Jika dilihat dari kuantitasnya, maksimal kuantitas yang memiliki harga 0 mencapai 648 barang.
- Karena jumlahnya sedikit, baris yang harganya 0 dapat dihapus.

Hasil Cleaning Data

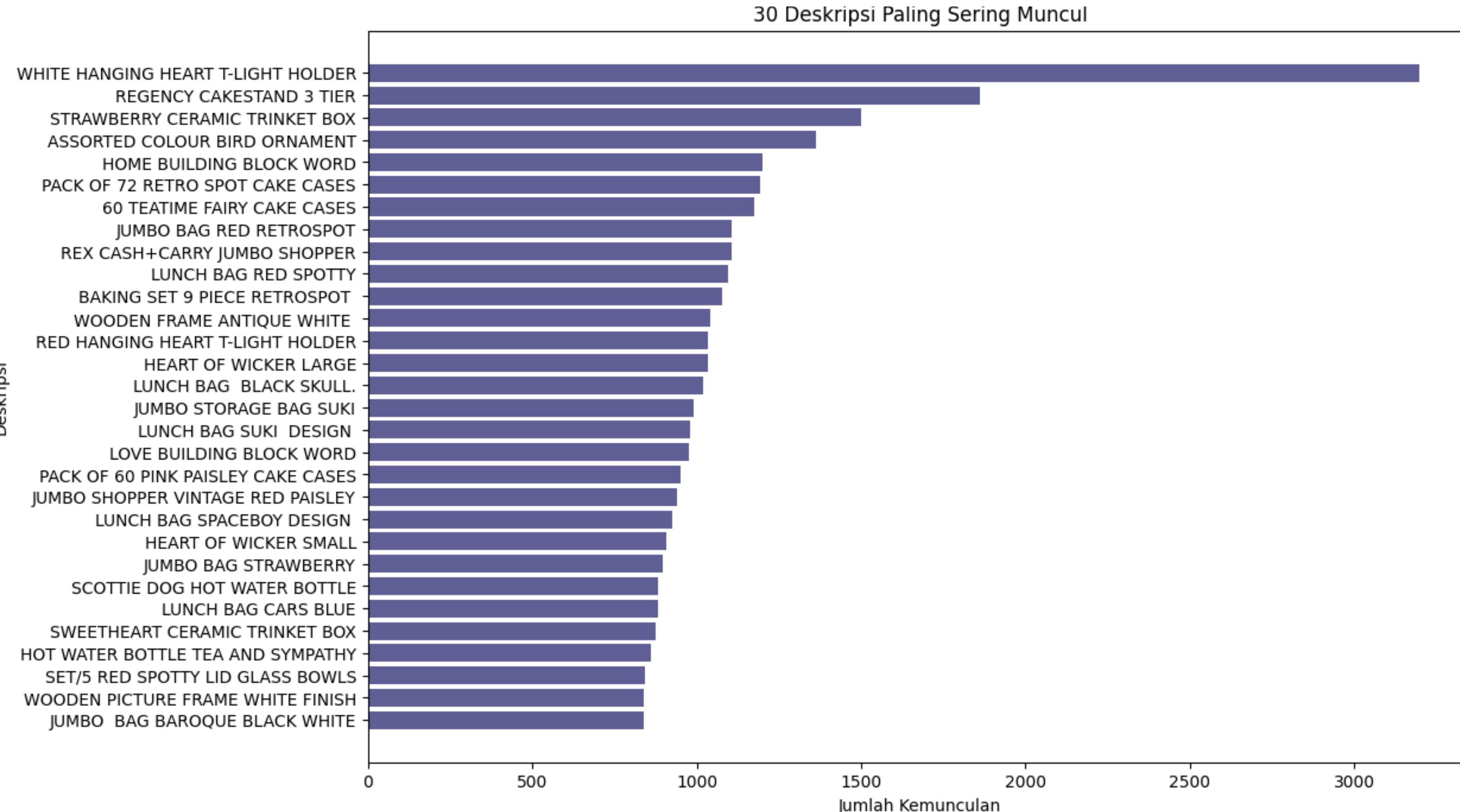
Hasil cleaning mendapatkan data akhir dengan 408924 baris dan 8 kolom

EXPLORE DATA

Top 10 Kode Barang Paling Sering Muncul

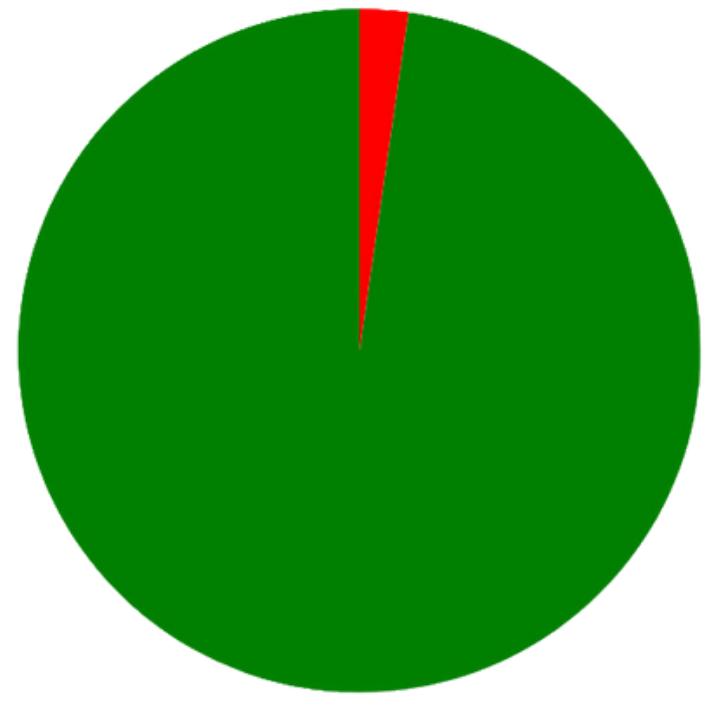


EXPLORE DATA (DESKRIPSI)



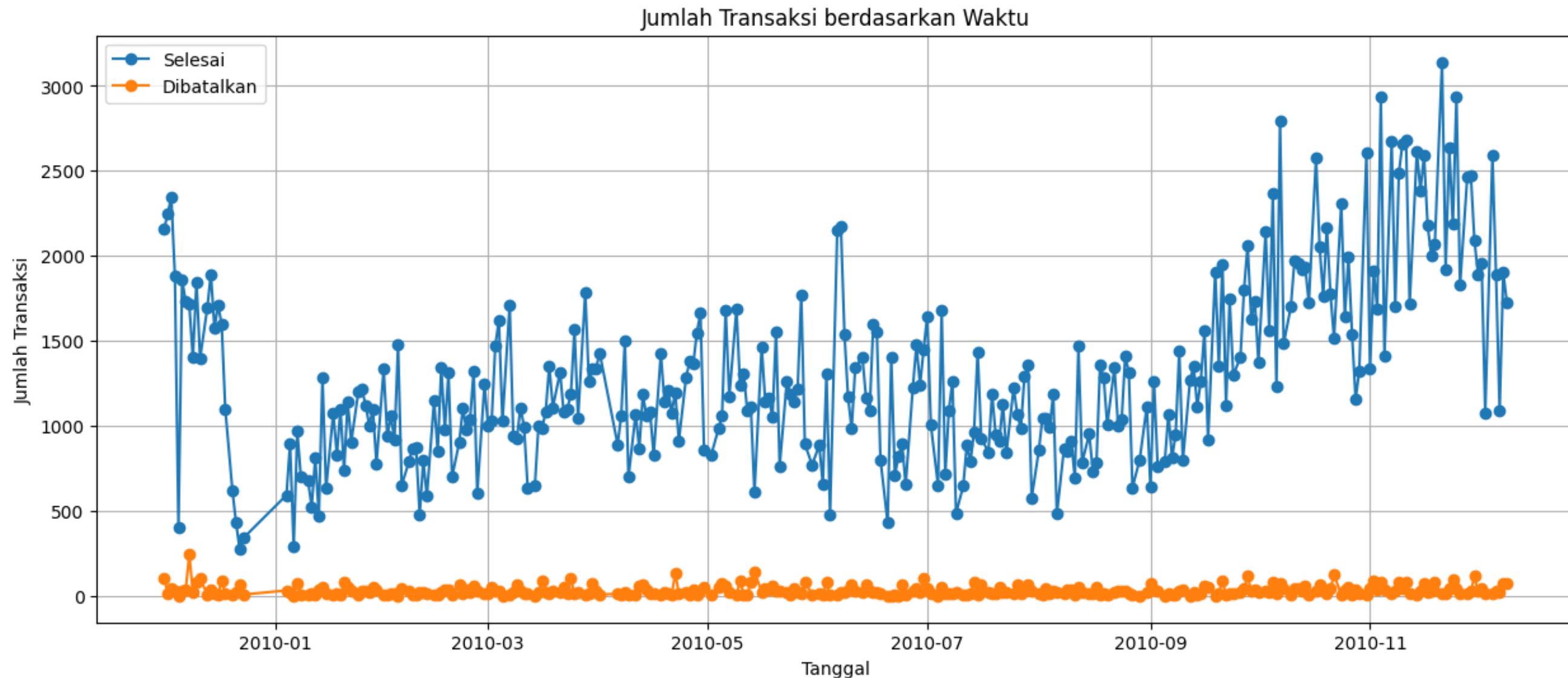
Data Exploration

EDA

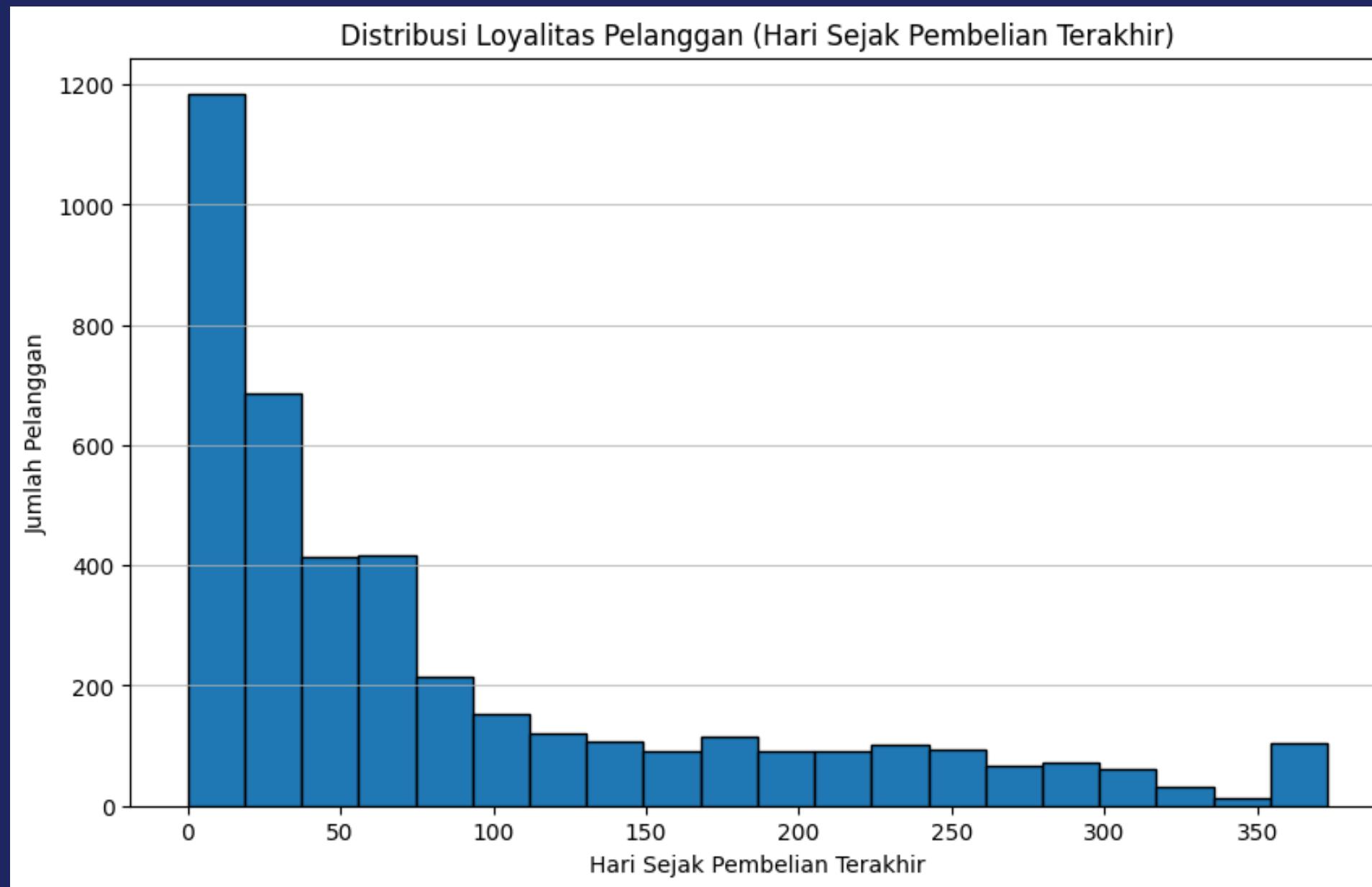


terdapat 9372 (2.3%) transaksi yang dibatalkan dan transaksi selesai sebesar 399552 (97.7%) transaksi berstatus selesai.

Dapat dilihat terjadi kenaikan jumlah transaksi dimulai dari tahun 2010 bulan 9 hingga bulan 12 yang ditandai adanya tren naik.

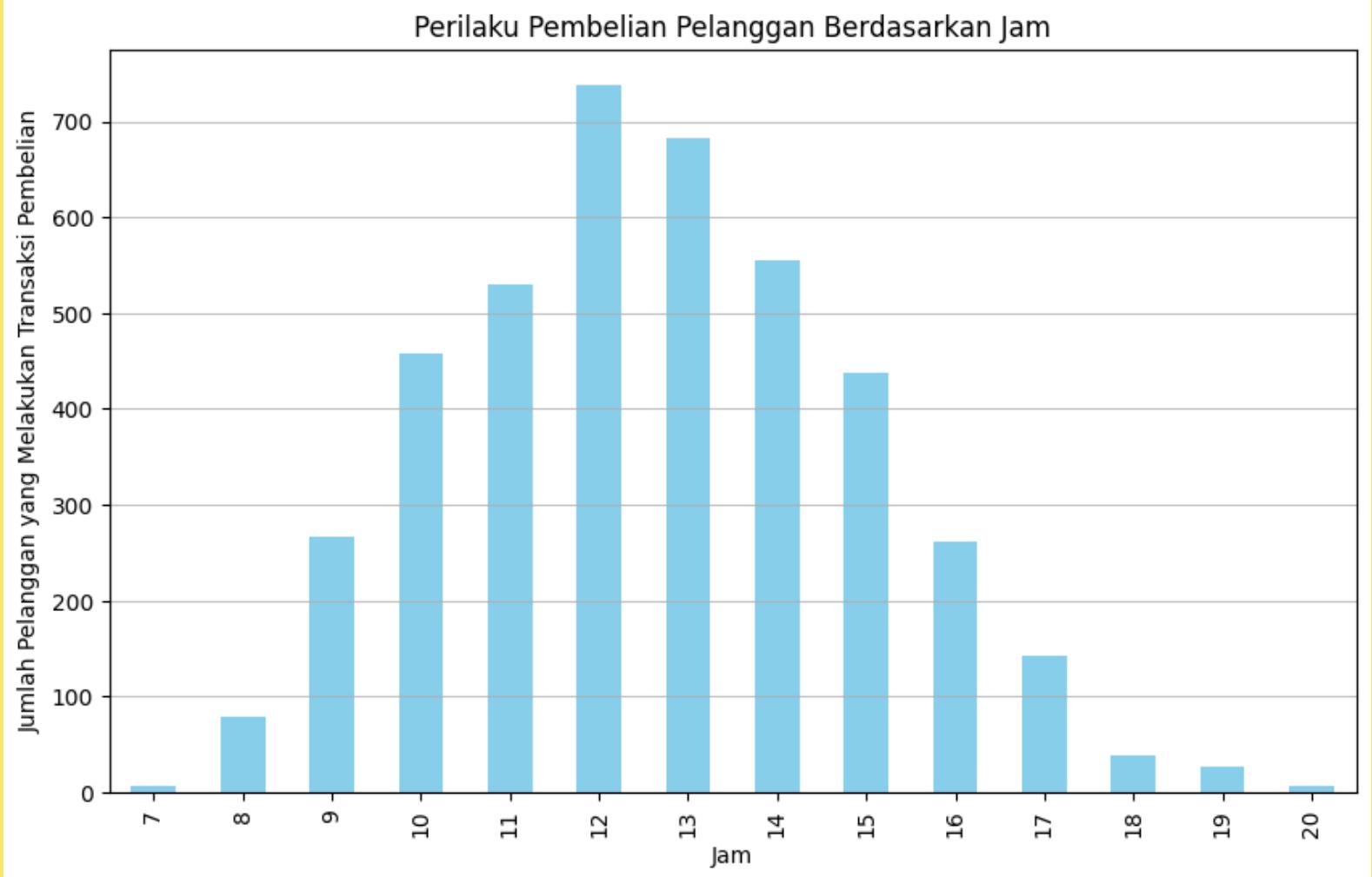


EDA



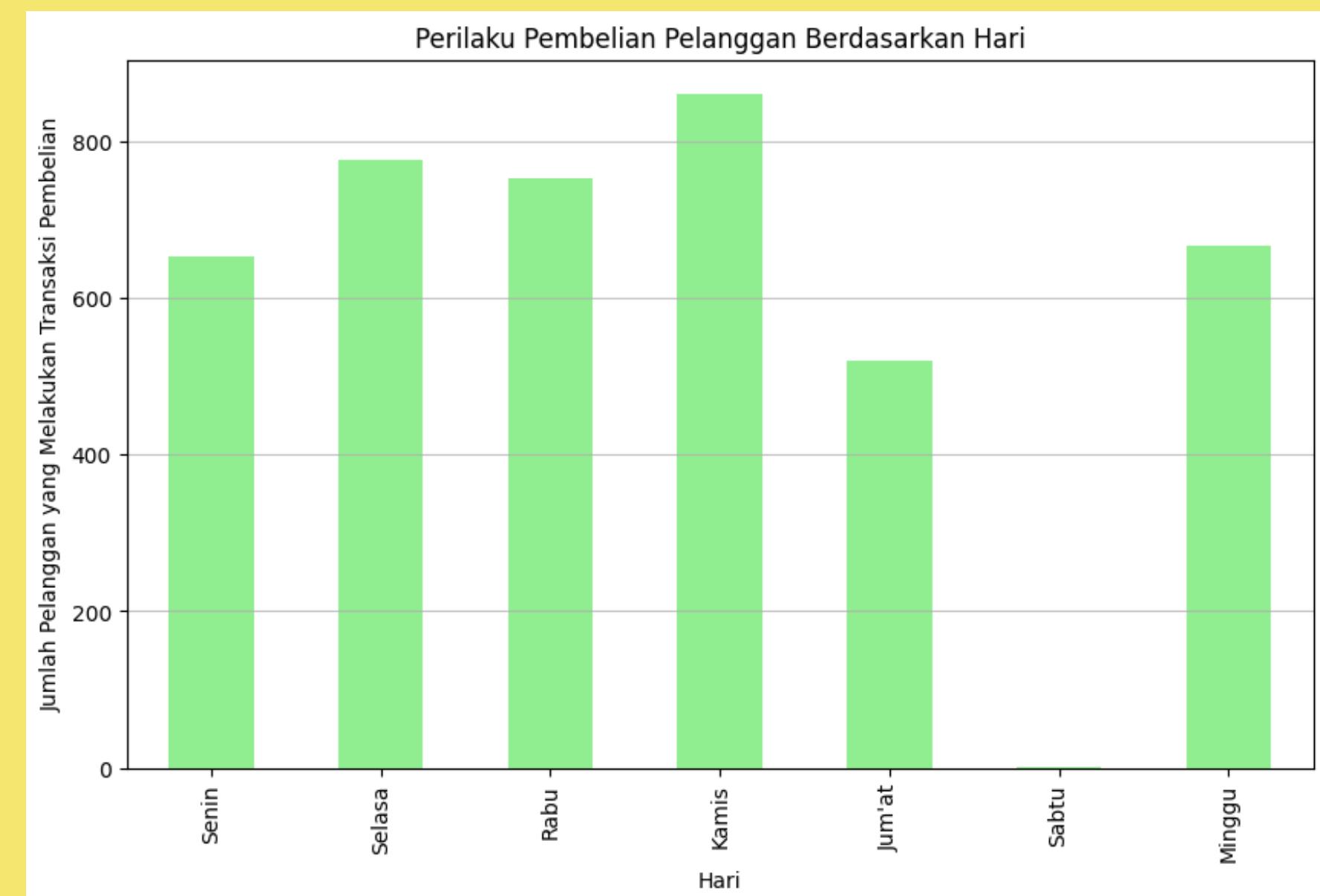
Skewness positif menunjukkan skenario yang menjelaskan perilaku pembelian pelanggan terkonsentrasi pada transaksi yang lebih baru, namun terdapat outlier yang tidak melakukan pembelian dalam jangka waktu yang cukup lama.

EDA



Mayoritas pelanggan lebih suka melakukan pembelian di pagi hingga sore hari, mulai dari jam 9 pagi sampai jam 4 sore dengan transaksi terbanyak terjadi di jam 12 siang yang menjadikan jam 12 siang menjadi *golden hour*.

Lebih banyak pelanggan yang cenderung melakukan pembelian di hari kamis. Sedangkan pada hari Sabtu, sangat sedikit jumlah pelanggan yang melakukan pembelian pada hari tersebut. Sehingga Sabtu menjadi hari yang paling tidak disukai bagi pelanggan untuk berbelanja atau melakukan transaksi.



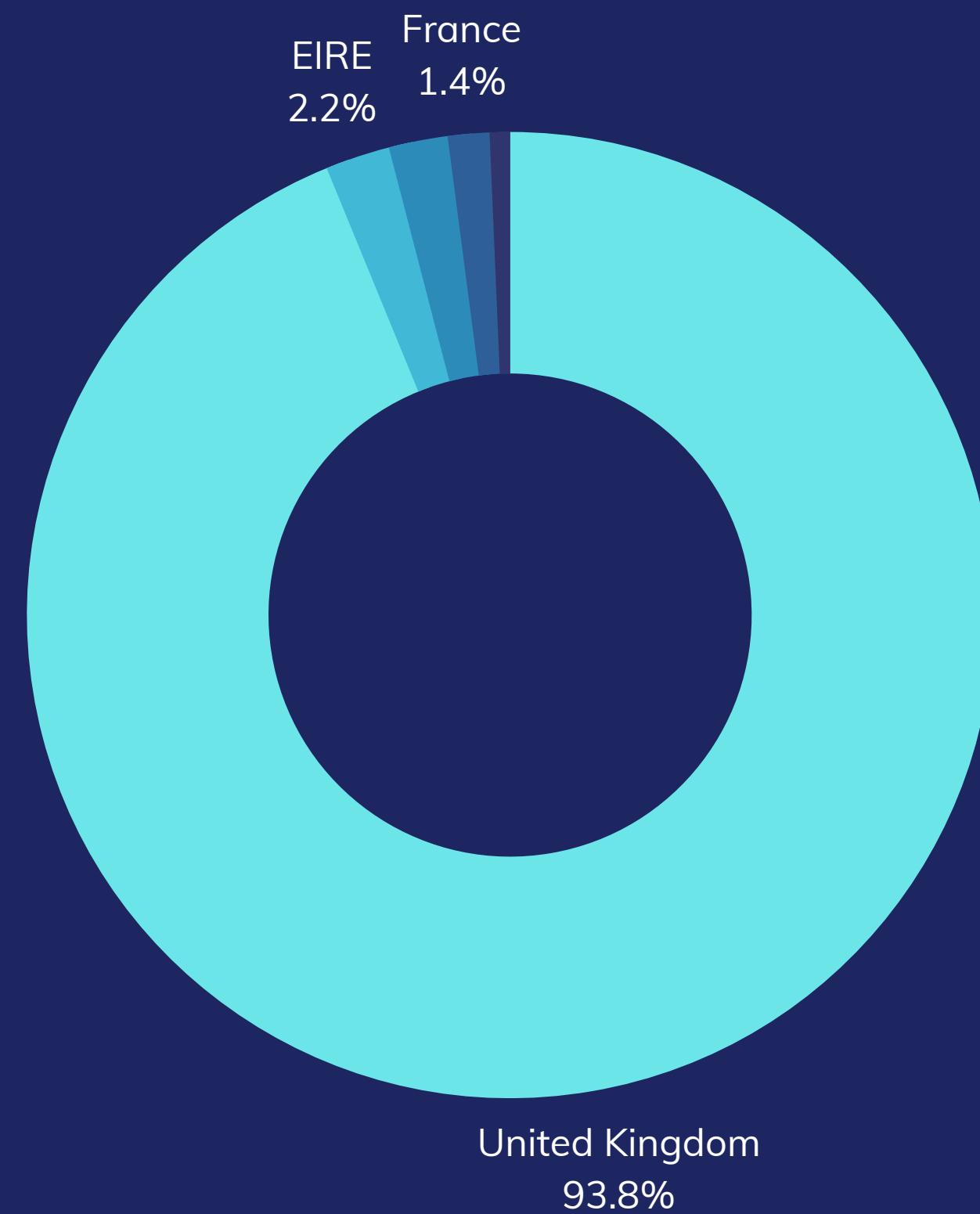
TRANSAKSI DIBATALKAN

Harga barang yang dibatalkan cukup bervariasi, tidak ada kecondongan terhadap suatu barang.

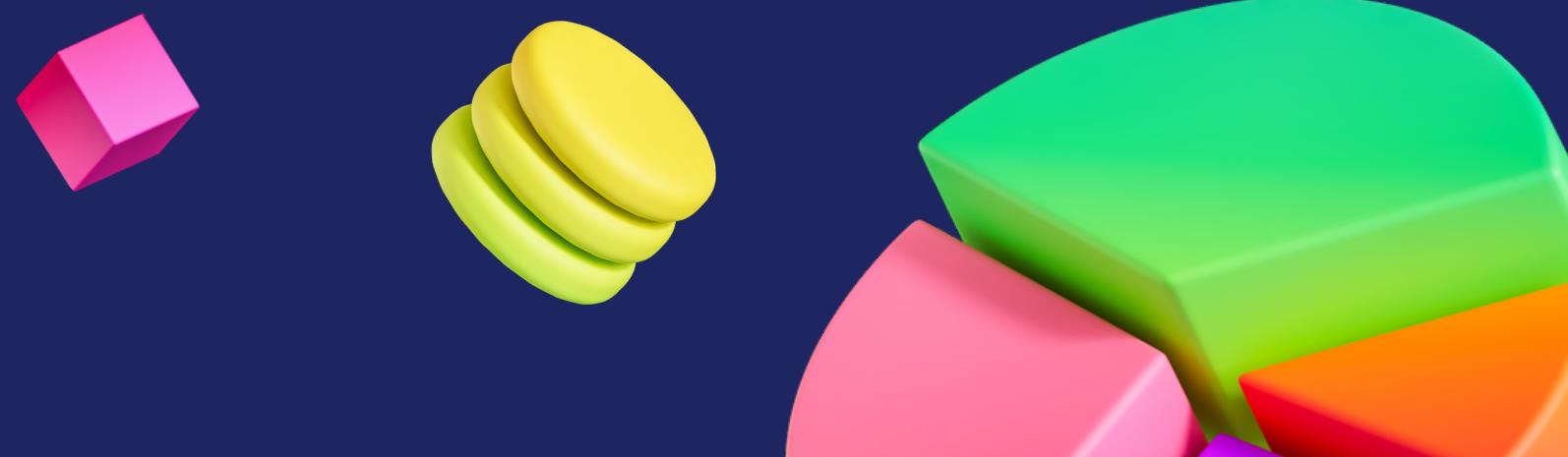
Persentase transaksi yang dibatalkan dalam dataset adalah: 2.39%

	Quantity	Price
count	9816.000000	9816.000000
mean	-21.582722	28.529997
std	212.324392	405.514214
min	-9360.000000	0.010000
25%	-6.000000	1.565000
50%	-2.000000	2.950000
75%	-1.000000	6.750000
max	-1.000000	25111.090000

ASAL NEGARA



Customer sebagian
berasal besar dari
United Kingdom



FEATURE ENGINEERING

Pada tahap Feature Enggining didapatkan informasi diantaranya:

- Data Pelanggan
- Total Transaksi
- Total Produk yang dibeli tiap Customer
- Merge total transaksi dan total produk dalam data pelanggan
- Total Belanja
- Rata-rata Belanja
- Dll,

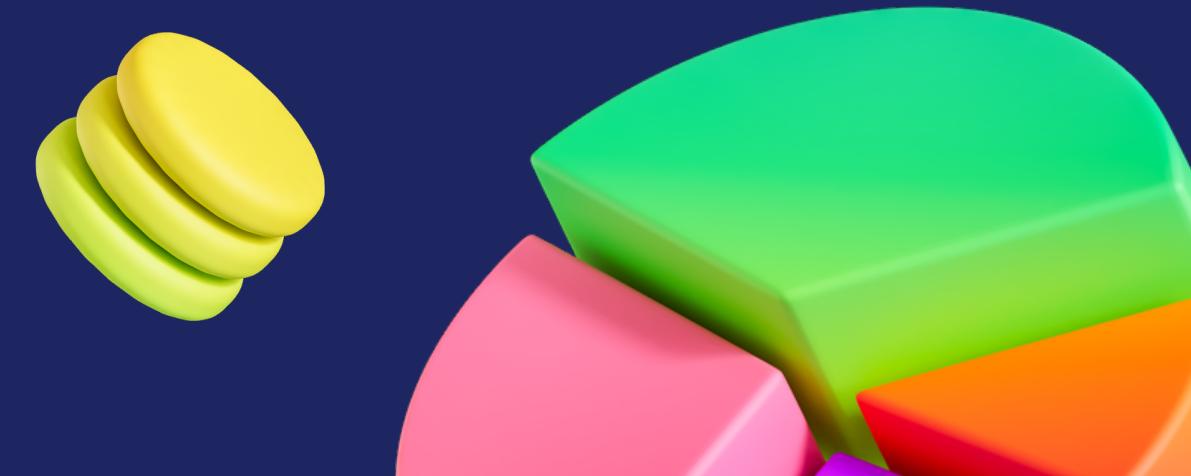
Data Pelanggan

Customer ID	Hari_Sejak_Pembelian_Terakhir
0	12346.0
1	12347.0
2	12348.0
3	12349.0
4	12351.0

HASIL FEATURE ENGINEERING

	Customer_ID	Hari_Sejak_Pembelian_Terakhir	Total_Transaksi	Total_Produk_Dibeli	Total_Belanja	Rata_rata_Belanja	Total_Jenis_Produk_Dibeli
0	12346.0	164	2	24	169.36	84.68	24
1	12347.0	2	2	828	1323.32	661.66	70
2	12348.0	73	1	372	221.16	221.16	19
3	12349.0	42	3	986	2196.99	732.33	91
4	12351.0	10	1	261	300.93	300.93	21

Rata_Hari_Antara_Pembelian	Hari	Jam	Is_UK	Frekuensi_Pembatalan	Tingkat_Pembatalan	Rata_Bulan	Std_Bulan
5.130435	0	13	1	0	0.0	84.68	81.501128
0.528571	6	14	0	0	0.0	661.66	70.894526
0.0	0	14	0	0	0.0	221.16	0.0
3.144231	3	8	0	1	0.333333	732.33	656.479014
0.0	0	15	0	0	0.0	300.93	0.0

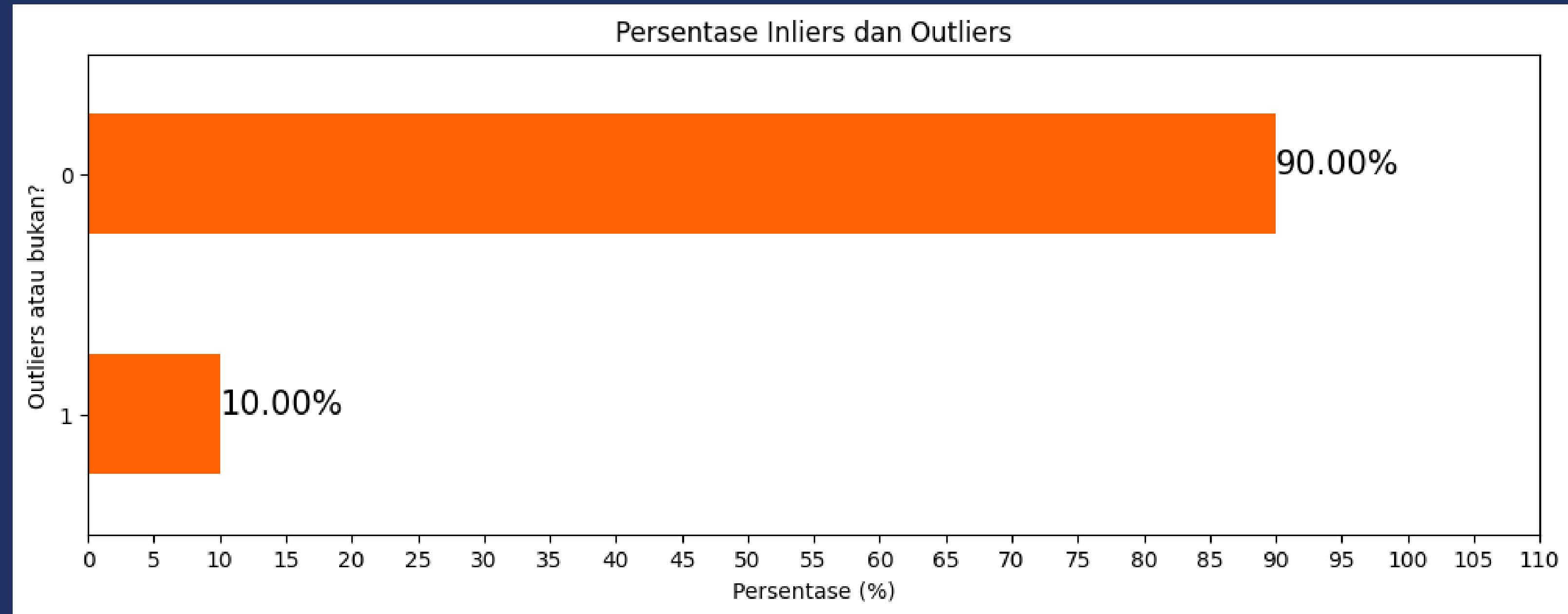


HASIL FEATURE ENGINEERING

15 fitur didapatkan dari feature engineering

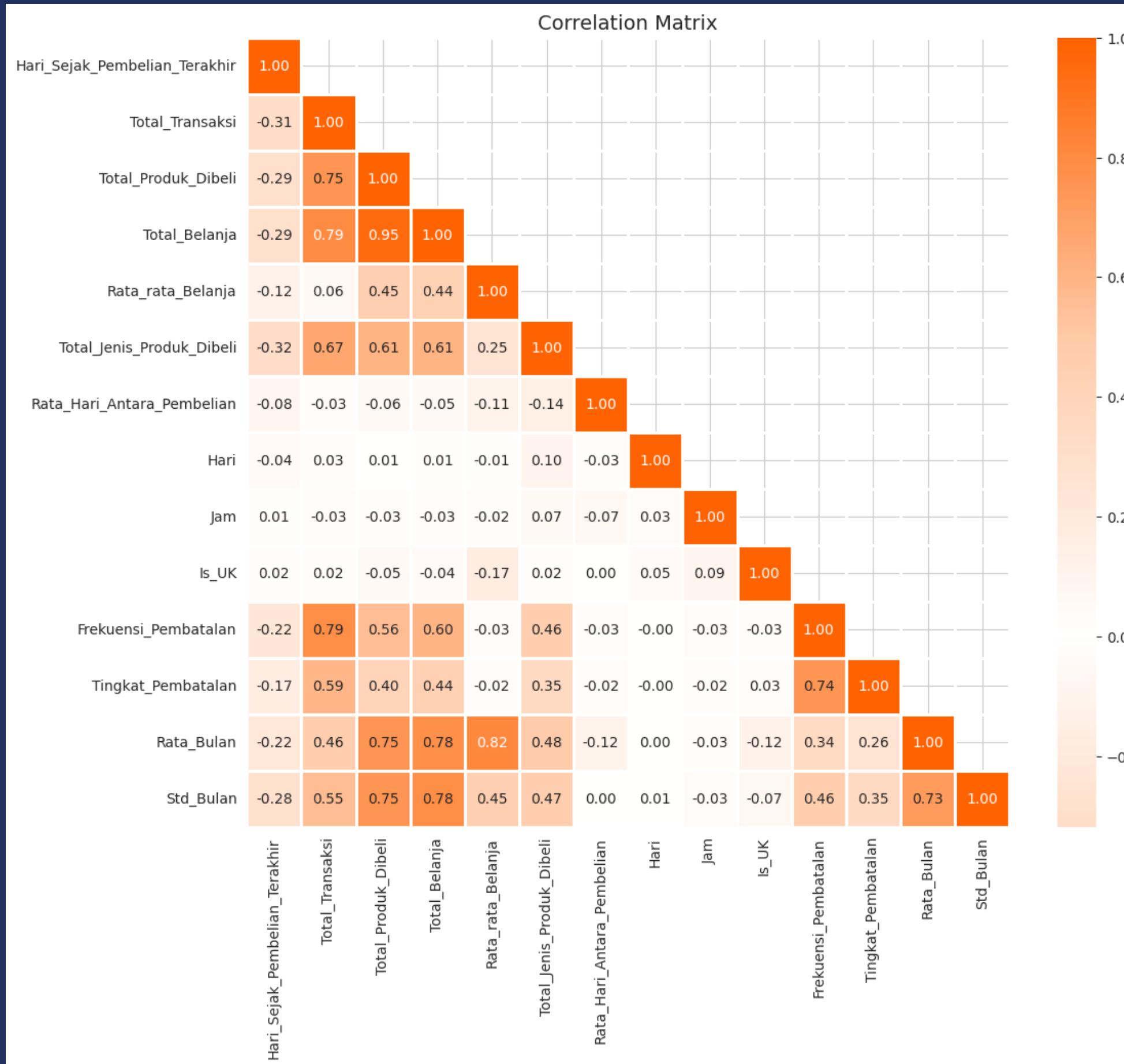
- Customer ID: Identifier unik untuk membedakan setiap pelanggan.
- Hari_Sejak_Pembelian_Terakhir: Jumlah hari sejak pembelian terakhir pelanggan.
- Total_Transaksi: Total jumlah transaksi yang dilakukan pelanggan.
- Total_Produk_Dibeli: Jumlah total produk yang dibeli pelanggan.
- Total_Belanja: Total uang yang dihabiskan pelanggan.
- Rata_rata_Belanja: Nilai rata-rata setiap transaksi pelanggan.
- Total_Jenis_Produk_Dibeli: Jumlah produk berbeda yang telah dibeli pelanggan.
- Rata_Hari_Antara_Pembelian: Rata-rata hari antara pembelian berturut-turut.
- Hari: Hari pelanggan cenderung berbelanja.
- Jam: Jam pelanggan cenderung berbelanja.
- Is_UK: Variabel biner menunjukkan apakah pelanggan berbasis di Inggris.
- Frekuensi_Pembatalan: Jumlah total transaksi yang dibatalkan oleh pelanggan.
- Tingkat_Pembatalan: Proporsi transaksi yang dibatalkan dari total transaksi.
- Rata_Bulan: Rata-rata pengeluaran bulanan pelanggan.
- Std_Bulan: Deviasi standar pengeluaran bulanan, menunjukkan variabilitas dalam pola pengeluaran pelanggan.

Outliers



- Data outliers akan dihapus dari data yang dipakai untuk segmentasi
- Data outliers akan dianalisis secara terpisah

Correlation Matrix



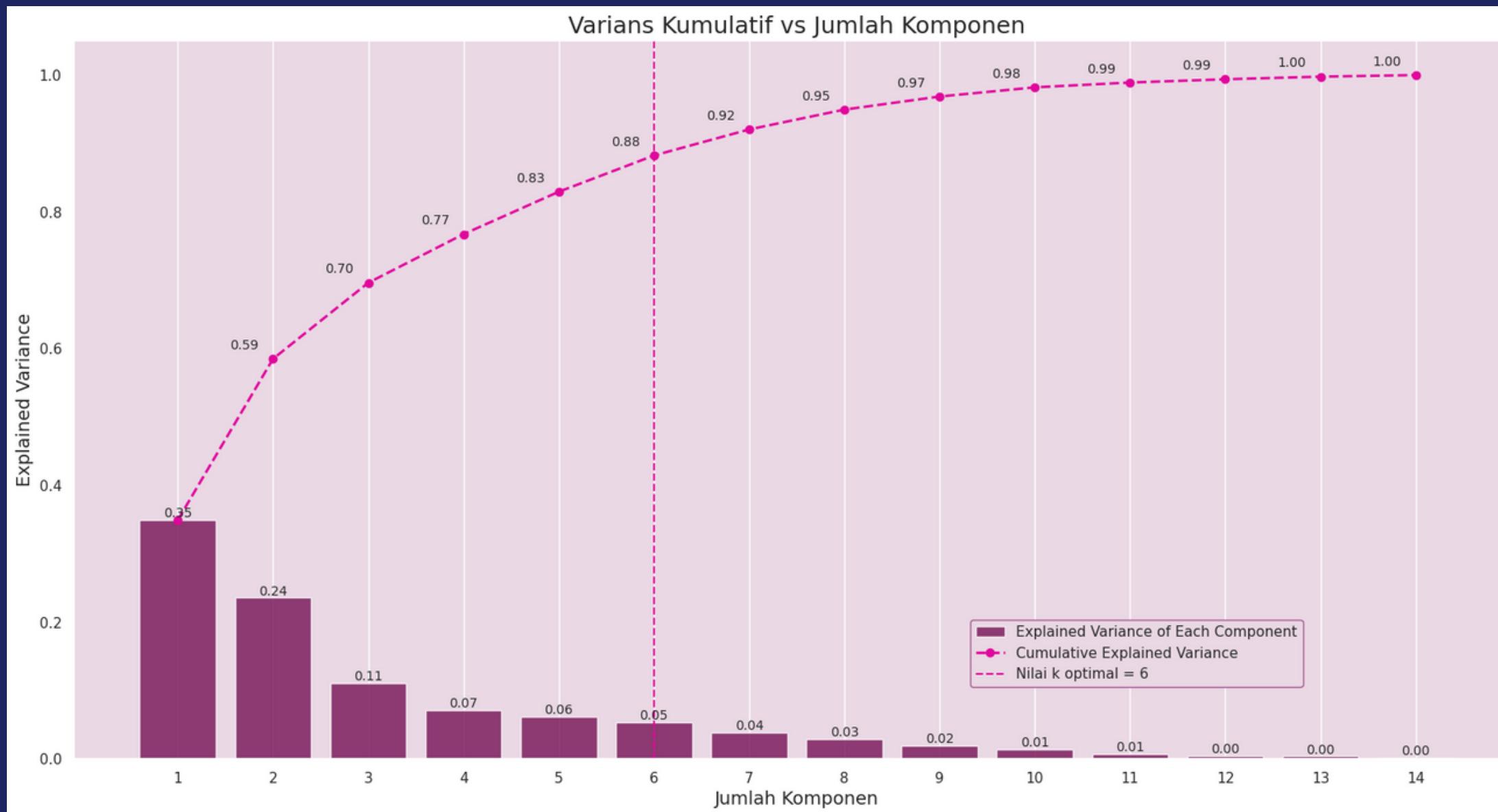
- Banyak fitur yang memiliki korelasi positif dengan nilai lebih dari 0.5 sehingga mengindikasikan adanya multikolinearitas
- data dapat direduksi menjadi dimensi yang lebih kecil berdasarkan variansnya menggunakan PCA.

FEATURE SCALING

- penggunaan **StandardScaler** dari pustaka **Scikit-learn (sklearn)** dalam proses pra-pemrosesan data untuk penskalaan fitur (feature scaling)
- Melakukan penskalaan pada kolom-kolom yang diperlukan menggunakan **fit_transform** dari objek **StandardScaler**, sehingga nilai-nilai dalam kolom-kolom tersebut diubah agar memiliki rata-rata nol dan varians satu (distribusi normal standar).

	Customer_ID	Hari_Sejak_Pembelian_Terakhir	Total_Transaksi	Total_Produk_Dibeli	Total_Belanja	Rata_rata_Belanja	Total_Jenis_Produk_Dibeli	Rata_Hari_Antara_Pembelian	Hari	Jam	Is_UK	Frekuensi_Pembatalan	Tingkat_Pembatalan	Rata_Bulan	Std_Bulan
0	12346.0	0.835966	-0.438042	-0.600730	-0.537908	-1.047724	-0.520785	0.228689	0	0.150301	1	-0.482834	-0.358638	-1.029984	-0.287261
1	12347.0	-0.882890	-0.438042	0.026633	-0.037174	2.290348	0.085704	-0.303518	6	0.590676	0	-0.482834	-0.358638	1.218095	-0.337137
2	12348.0	-0.129564	-0.582075	-0.329185	-0.515430	-0.258130	-0.586685	-0.364645	0	0.590676	0	-0.482834	-0.358638	-0.498219	-0.670507
3	12349.0	-0.458481	-0.294008	0.149921	0.341935	2.699203	0.362569	-0.001015	3	-2.051572	0	0.065069	-0.066059	1.493446	2.416480
4	12351.0	-0.798008	-0.582075	-0.415799	-0.480816	0.203373	-0.560317	-0.364645	0	1.031051	0	-0.482834	-0.358638	-0.187412	-0.670507

PCA (PRINCIPAL COMPONENT ANALYSIS)



- * Komponen pertama bersama-sama dapat menjelaskan 35% varians
- * komponen kedua bersama-sama dapat menjelaskan 59% varians dst

Untuk memilih jumlah komponen optimal, umumnya kita mencari titik di mana penambahan komponen lain tidak secara signifikan meningkatkan varians yang dijelaskan kumulatif, sering disebut sebagai titik "siku" dalam kurva.

PCA

nilai kumulatif varians dijelaskan mengindikasikan seberapa banyak varians total dalam dataset yang ditangkap oleh setiap komponen utama, serta varians kumulatif yang dijelaskan oleh n komponen pertama.

	PC1	PC2	PC3	PC4	PC5	PC6
Hari_Sejak_Pembelian_Terakhir	-0.165658	0.009561	-0.056886	-0.296579	-0.537142	0.683990
Total_Transaksi	0.358912	0.014222	0.297433	-0.026231	-0.004874	-0.017989
Total_Prodук_Dibeli	0.391547	0.031617	-0.081565	0.021342	-0.011157	0.067027
Total_Belanja	0.401607	0.029792	-0.058130	0.022494	-0.022521	0.088477
Rata_rata_Belanja	0.197833	0.028286	-0.595688	0.022618	-0.033726	0.052333
Total_Jenis_Prodук_Dibeli	0.308565	-0.033063	0.055015	-0.167444	0.153408	-0.245568
Rata_Hari_Antara_Pembelian	-0.035655	0.017724	0.114373	0.727093	0.360377	0.509308
Hari	0.059357	-0.995710	-0.012755	0.033919	-0.034657	0.032888
Jam	-0.010401	-0.027626	-0.002630	-0.576666	0.713636	0.372646
Is_UK	-0.005692	-0.007387	0.025893	-0.017663	0.018444	0.003687
Frekuensi_Pembatalan	0.304646	0.027652	0.418784	-0.048514	-0.127118	0.093569
Tingkat_Pembatalan	0.246876	0.024682	0.418405	-0.068162	-0.160072	0.147314
Rata_Bulan	0.340870	0.032010	-0.391045	0.003895	-0.058136	0.098135
Std_Bulan	0.346276	0.030951	-0.152037	0.110254	0.017357	0.130495

Clustering Algorithm

1

K-means with
Random
Initialization

2

K-Means++

3

DenMune : A density-
peak clustering
algorithm

4

K-Medoids with
BanditPAM

K-Means

mengelompokkan data menjadi k kelompok
berdasarkan kesamaan atribut

1

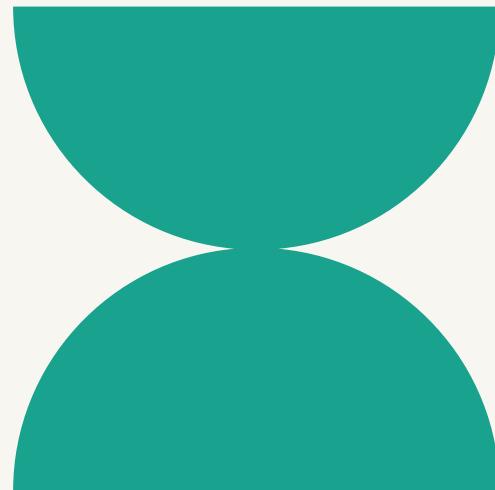
K-Means

Centroid dipilih secara acak dari antara titik data.

2

K-Means++

Memilih centroid yang jauh dari satu sama lain, mengurangi kemungkinan hasil konvergensi ke solusi suboptimal.





3

Menemukan kelompok dengan ukuran, bentuk, dan kerapatan yang sembarang dalam dua dimensi. Dimensi yang lebih tinggi pertama-tama direduksi menjadi 2-D menggunakan t-sne. Algoritma ini bergantung pada satu parameter K (jumlah tetangga terdekat).

DenMune : A density-peak clustering algorithm



4

K-Medoids with BanditPAM

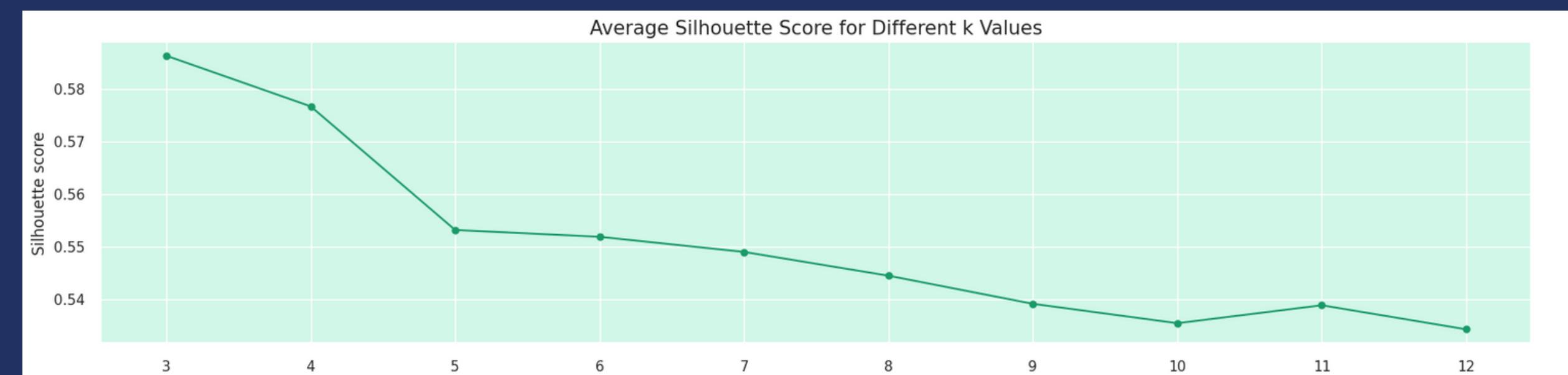
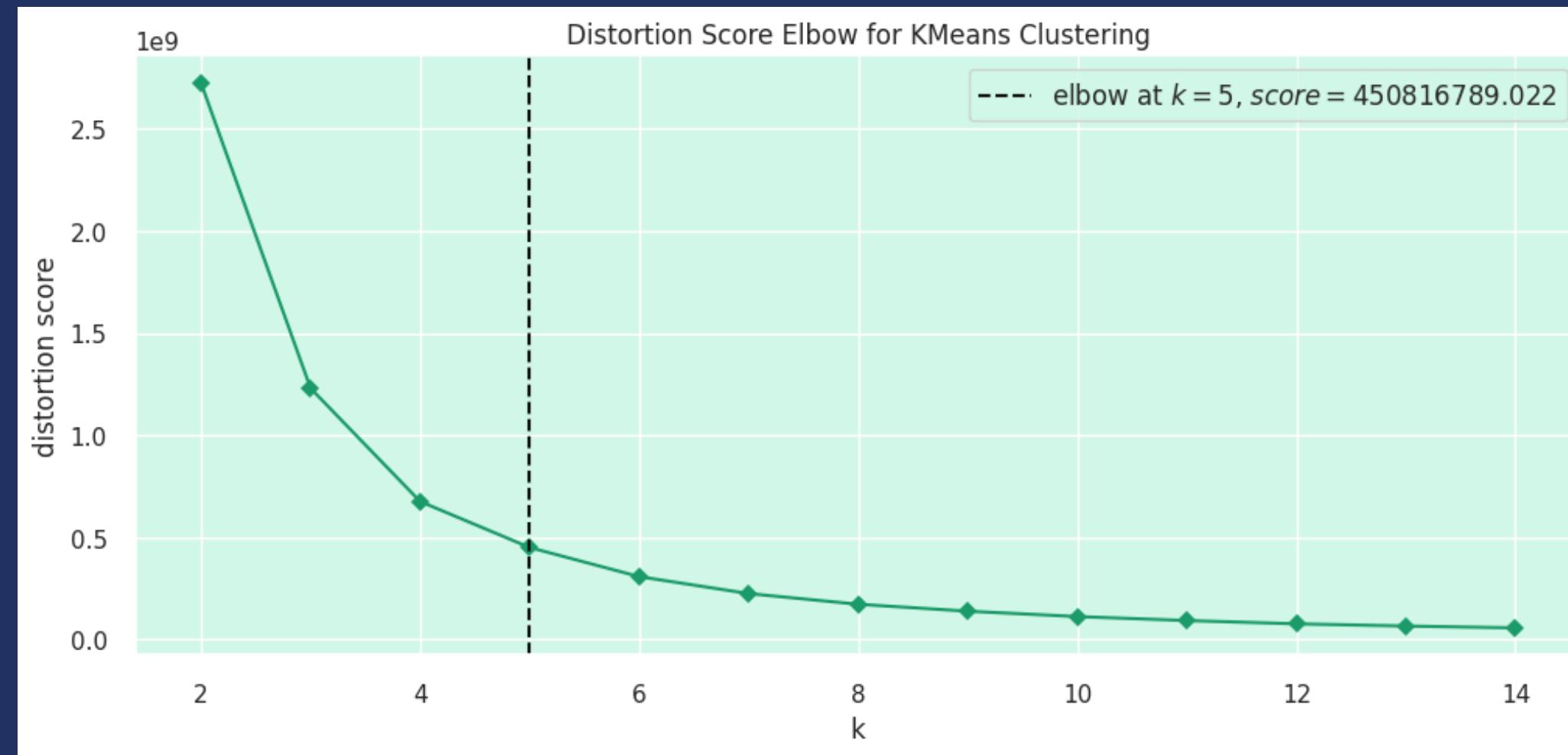
Menggunakan medoid (titik data yang mewakili kelompok secara optimal) sebagai pusat kelompoknya. K-Medoid memiliki harga komputasi yang mahal. BanditPAM mengurangi kompleksitas pada ukuran sampel n dari $O(n^2)$ menjadi $O(n \log n)$ sehingga dapat mengurangi beban komputasi dari K-Medoid.

Menentukan Nilai K pada K-Means

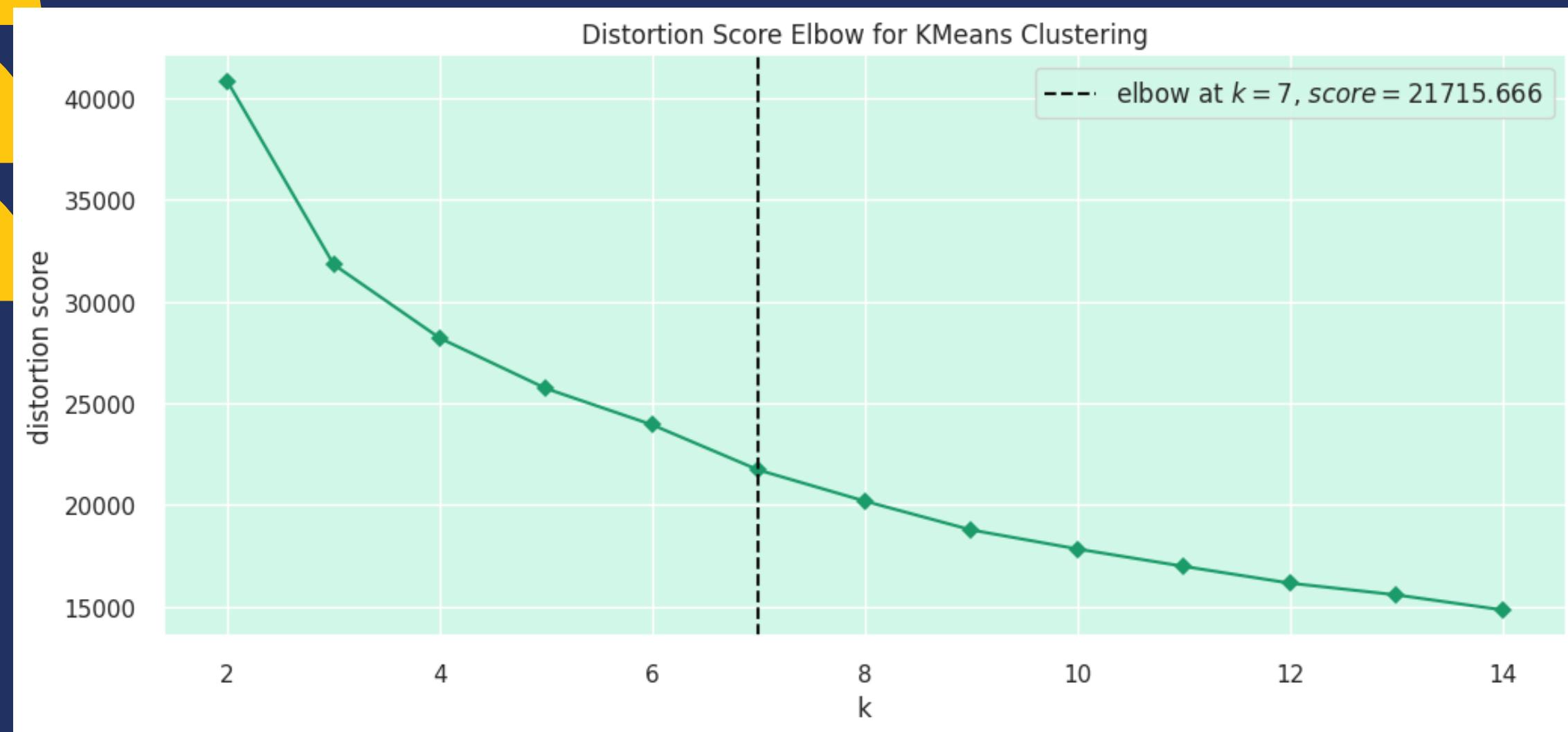


Silhouette
Score

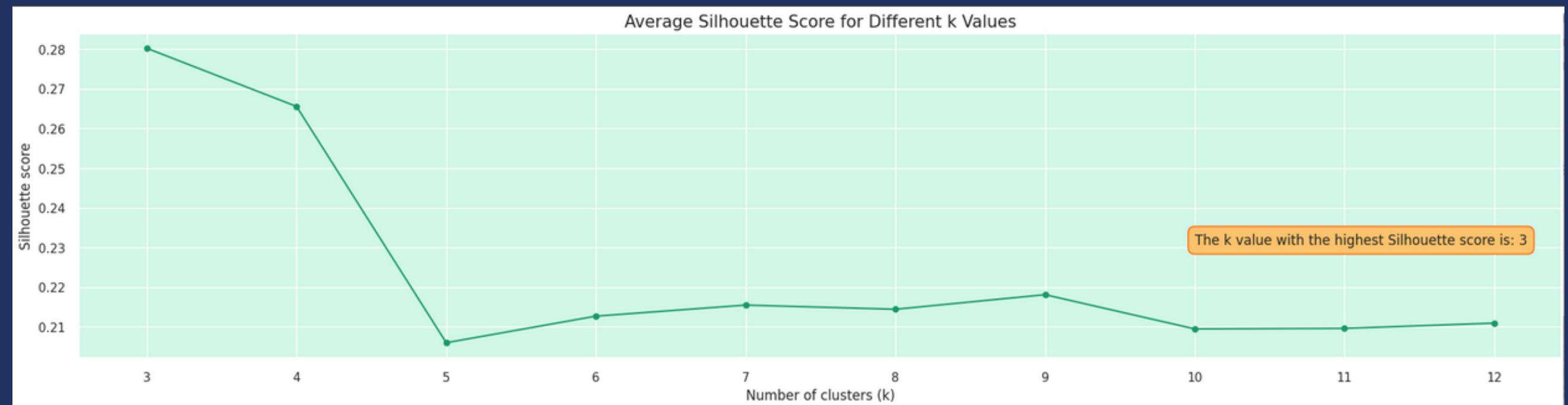
Elbow
Method



Menentukan Nilai K pada K-Means++



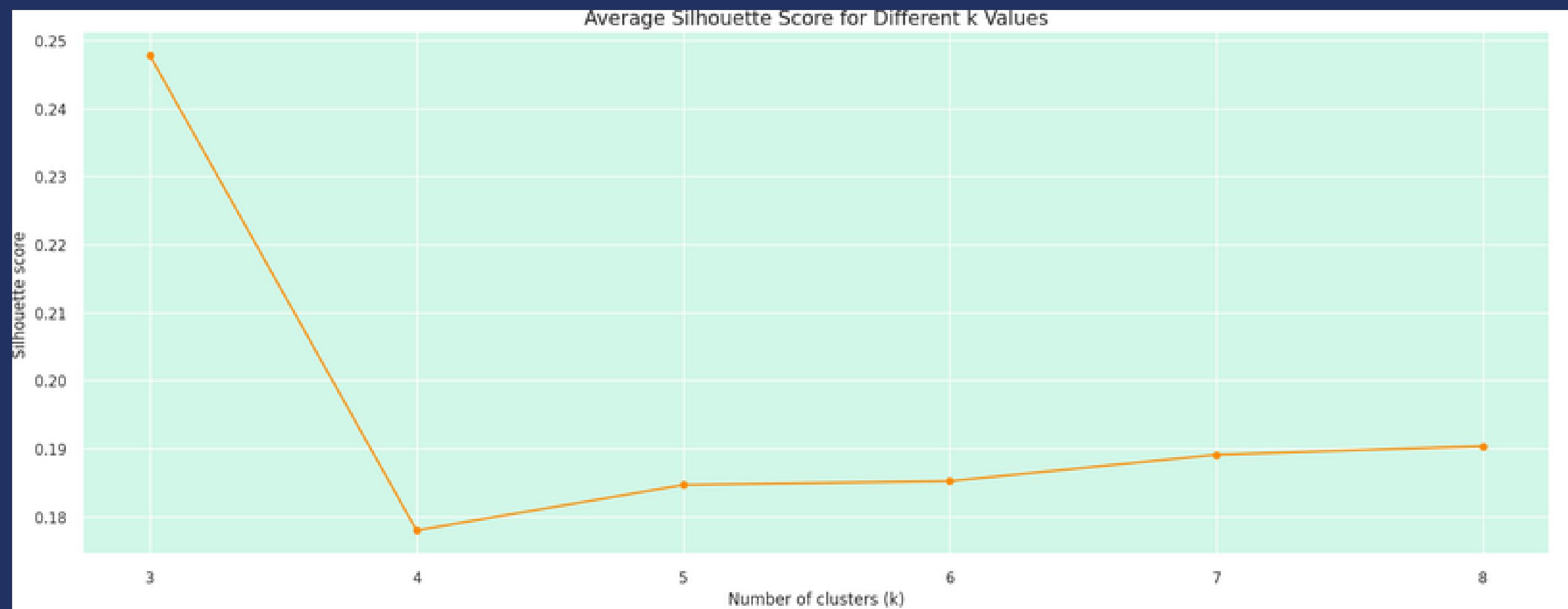
Silhouette Score



Elbow Method

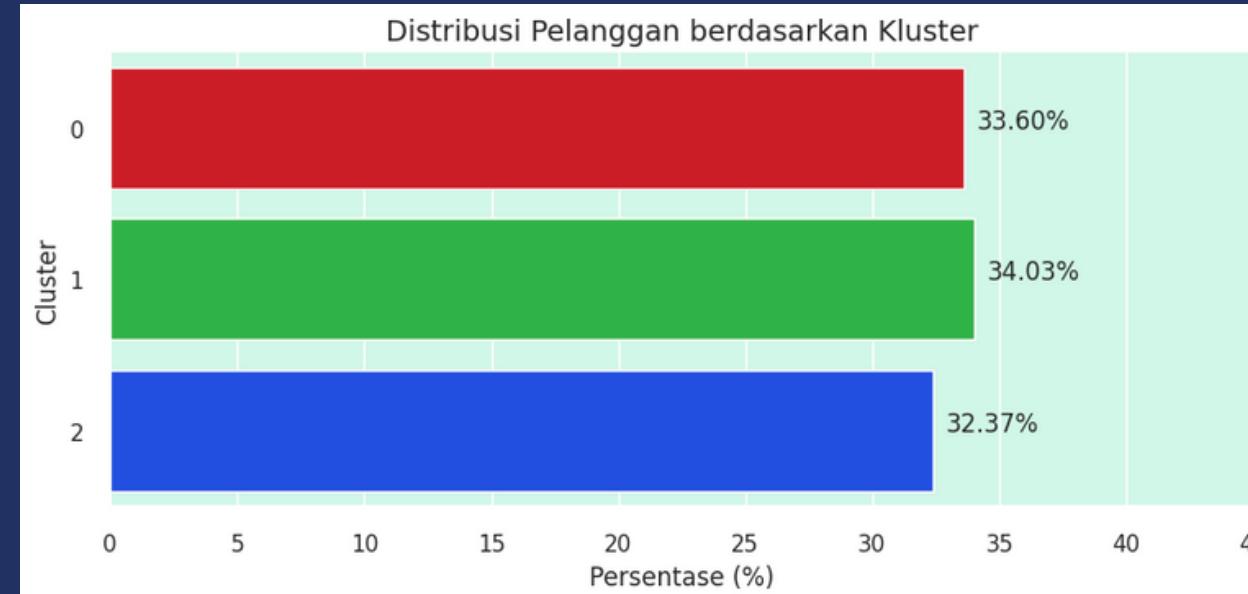
Menentukan Nilai K pada K-Medoids

Silhouette
Score

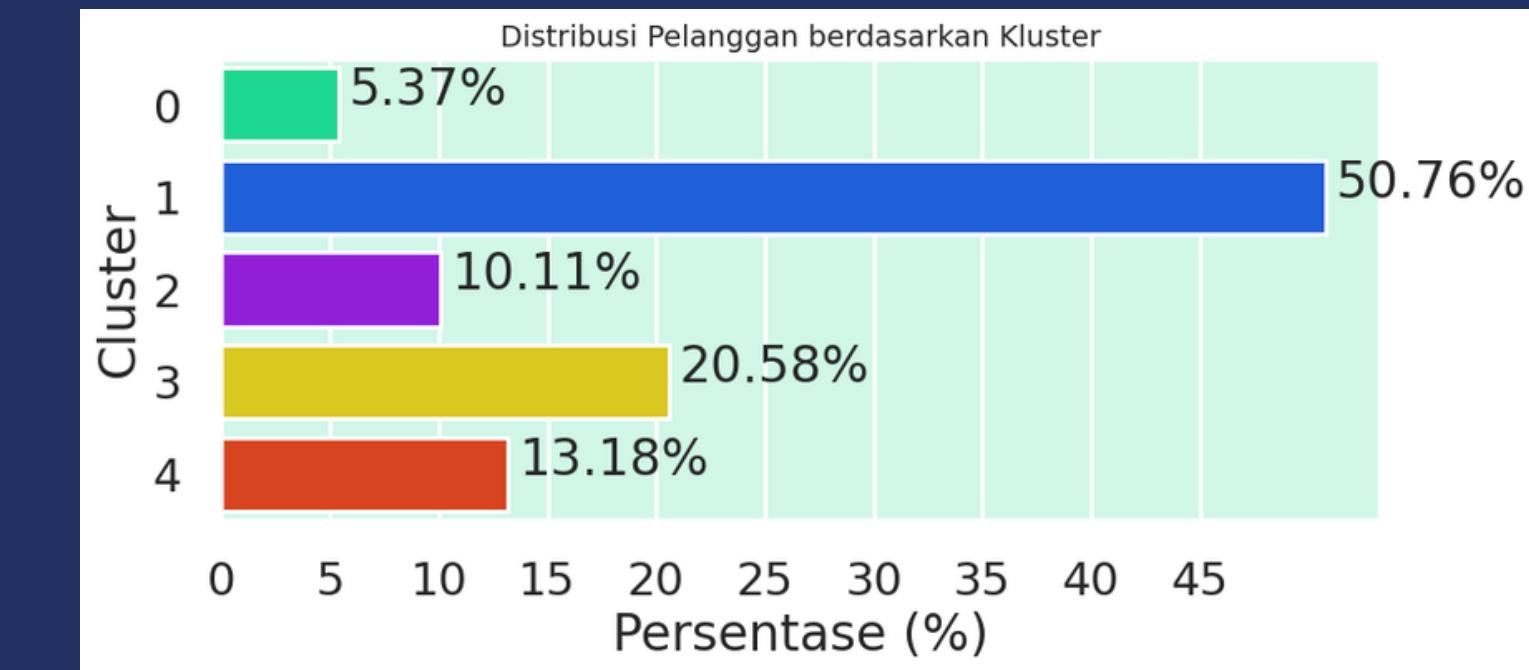


Hasil Distrbusi Cluster pada Masing-masing Algoritma

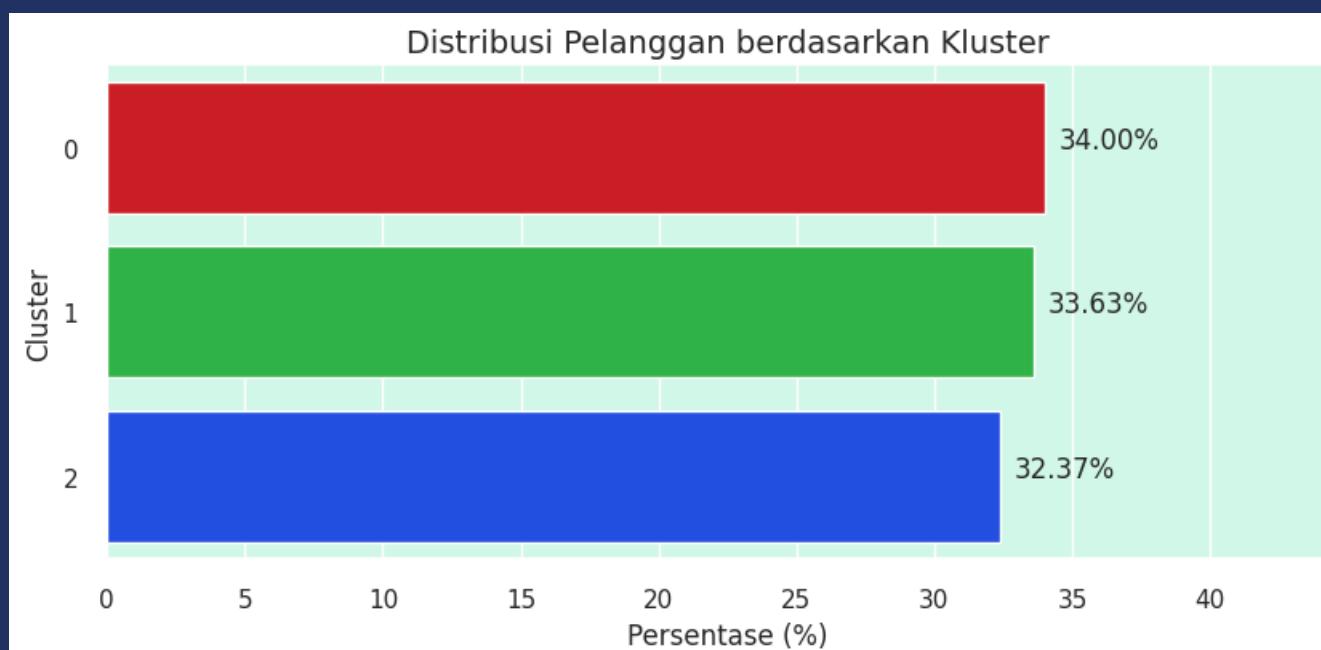
K-means with Random Initialization



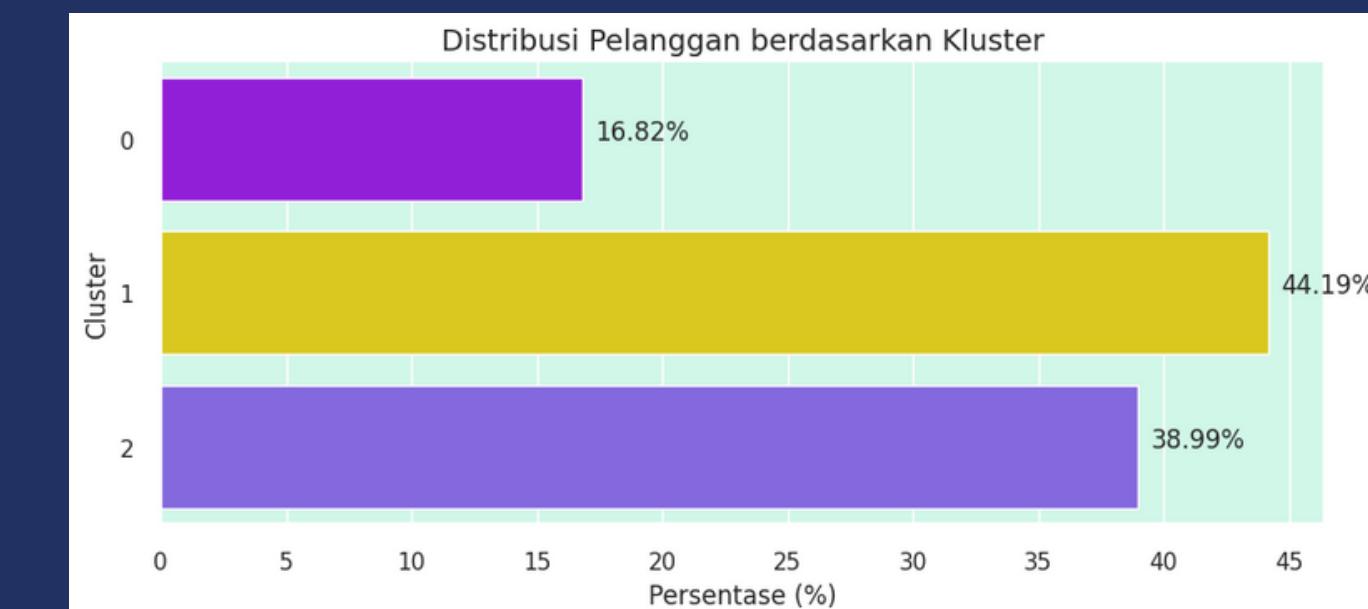
DenMune



K-Means++

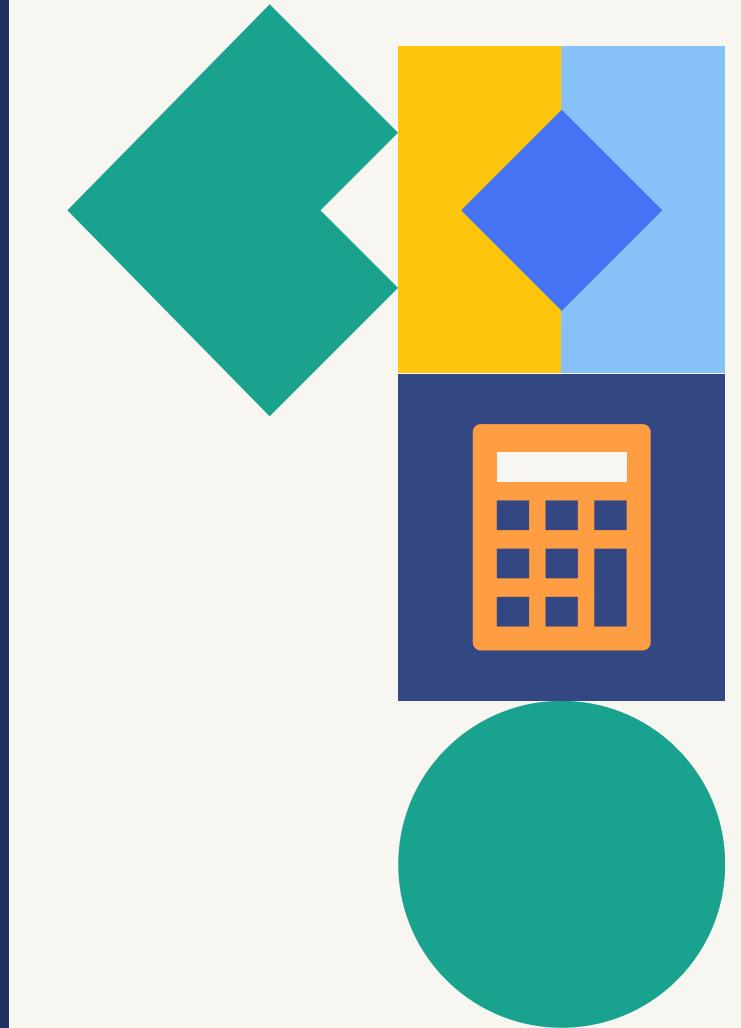


K-Medoids with BanditPAM



Comparison

	K-Means	K-Means++	DenMune	K-Medoid BanditPAM
Silhouette Score	0.5863	0.5863	0.3424	0.5861
Calinski Harabasz Score	14902.534	14902.429	5628.0305	14900.36
Davies Bouldin Score	0.50458	0.50459	0.5033	0.50470
Waktu Eksekusi	0.6698 s	1.2425 s	69.982 s	22.35073 s



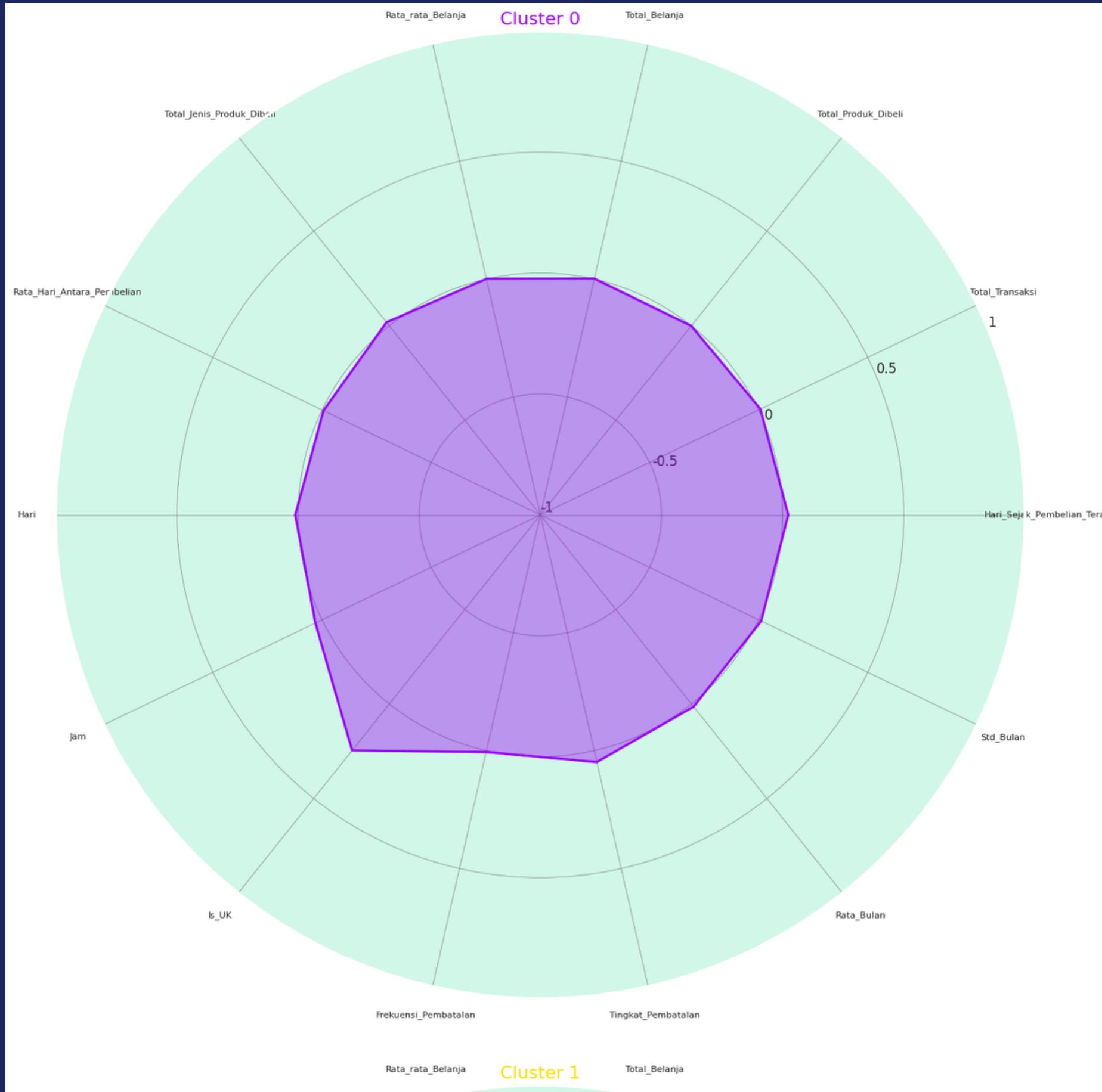
- Skor Silhouette: Nilai yang lebih tinggi menunjukkan pemisahan kelompok yang lebih baik. Rentang nilainya dari -1 hingga 1.
- Skor Calinski Harabasz: Skor yang lebih tinggi menunjukkan kelompok yang lebih terdefinisi dengan baik.
- Skor Davies Bouldin: Nilai yang lebih rendah menunjukkan pemisahan kelompok yang lebih baik.

Cluster Analysis Overall

Analisis

- * K-Means dan K-Means+++ menghasilkan 3 cluster dengan karakteristik yang berbeda dan distribusi yang tidak berbeda jauh antar cluster.
- * DenMune menghasilkan 5 Cluster tetapi selain cluster 0, semua cluster tidak memiliki karakteristik yang jauh berbeda. Skor metrik DenMune juga merupakan yang terendah dari semua algoritma, sehingga hasil cluster DenMune tidak akan dipakai untuk implementasi.
- * K-Medoids berhasil membagi 3 cluster customer dengan karakteristik yang berbeda tiap cluster dan distribusi tiap cluster juga tidak berbeda jauh antara satu sama lain.
- * K-Means, K-Means++, K-Medoids menghasilkan cluster dengan karakteristik yang sama namun hanya ada sedikit perbedaan pada K-Medoids karena hasil distribusinya berbeda
- * Kelompok kami akan memutuskan untuk menggunakan hasil cluster dari K-Medoids berhasil membuat segmentasi pelanggan yang diinginkan dan mudah diimplementasikan

Visualisasi cluster 0 K-Medoids



Cluster 0: Pelanggan Beraktivitas Tinggi dan Setia

Pelanggan dalam Cluster 0 memiliki karakteristik yang menunjukkan tingkat aktivitas yang tinggi dan kesetiaan terhadap bisnis.

- Cenderung melakukan pembelian dengan frekuensi yang lebih tinggi, ditandai dengan jumlah transaksi, produk yang dibeli, dan total pengeluaran yang sedikit di atas rata-rata.
- Cenderung berbelanja lebih sering, dengan jeda waktu antar pembelian yang sedikit lebih pendek.
- Sebagian besar dari mereka berbasis di Inggris.
- Meskipun terdapat tingkat pembatalan yang sedikit lebih tinggi, pola pengeluaran bulanan mereka tetap konsisten.

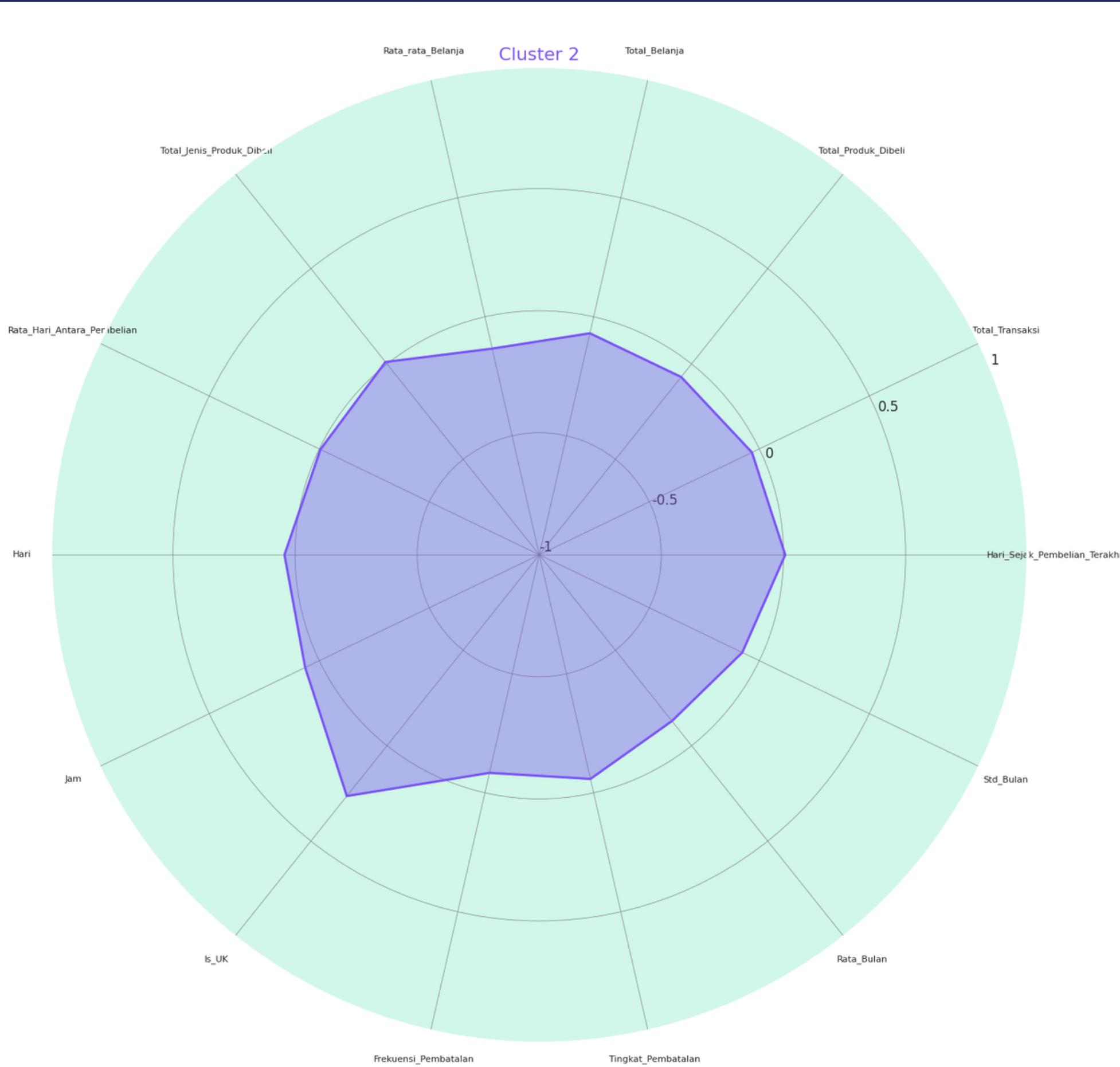
visualisasi cluster 1 K-Medoids



Cluster 1: Pelanggan Berbelanja Intensif tapi Jarang

- Mencakup pelanggan yang aktif berbelanja dengan karakteristik transaksi yang tinggi.
- Meskipun lebih jarang, setiap transaksi dilakukan dengan nilai yang lebih tinggi, tercemin dalam jumlah total transaksi, produk yang dibeli, dan total pengeluaran yang di atas rata-rata.
- Meskipun terdapat tingkat pembatalan yang lebih tinggi, pola pengeluaran bulanan mereka tetap tinggi.

visualisasi cluster 2 K-Medoids



Cluster 2: Pelanggan Tidak Aktif

- Pelanggan dalam cluster ini cenderung tidak aktif, dengan frekuensi pembelian yang sangat rendah.
- Mereka melakukan sedikit transaksi, membeli sedikit produk, dan menghabiskan jumlah uang yang terbatas.
- Kurang cenderung membatalkan transaksi, dan pola pengeluaran bulanan mereka sangat konsisten.
- Mereka juga cenderung tidak berbasis di Inggris.



Marketing Strategies Results and Decision Recommendations

Cluster 0 : Pelanggan beraktivitas tinggi dan setia

- Program loyalitas yang memberikan penghargaan kepada pelanggan untuk pembelian mereka.
- Penawaran khusus dan diskon untuk pelanggan setia untuk mendorong pembelian lebih lanjut.
- Melakukan survei kepuasan pelanggan untuk memahami faktor yang dapat meningkatkan tingkat kepuasan dan mengurangi tingkat pembatalan.
- Tingkatkan komunikasi personal dengan pelanggan,



Marketing Strategies Results and Decision Recommendations

Cluster 1 : Pelanggan Berbelanja Intensif tapi Jarang

- Menawarkan insentif atau diskon khusus untuk pembelian berulang.
- Mengembangkan program loyalitas yang memberikan penghargaan berdasarkan nilai transaksi atau frekuensi pembelian tinggi.
- Pesan pemasaran/notifikasi yang mengingatkan pelanggan untuk berbelanja kembali, seperti email yang dikirimkan dua minggu setelah pembelian terakhir mereka.
- Personalisasi tawaran produk berdasarkan preferensi pelanggan untuk meningkatkan nilai transaksi.
- Program yang menawarkan manfaat tambahan untuk pelanggan yang berbelanja secara teratur, seperti akses ke konten eksklusif atau layanan pelanggan yang lebih cepat.



Marketing Strategies

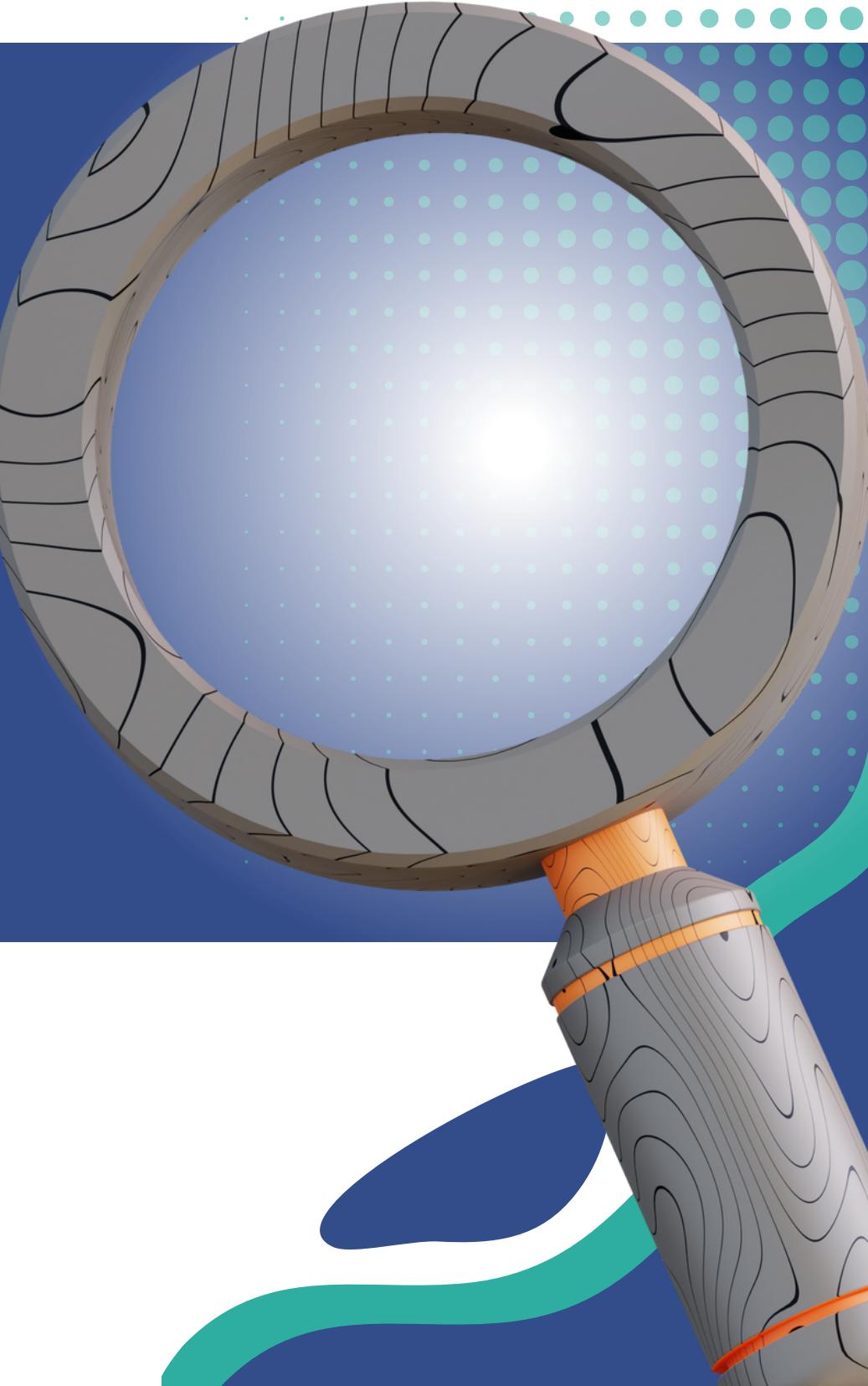
Results and Decision Recommendations

Cluster 2 : Pelanggan tidak aktif

- 
- Mengadakan program diskon atau penawaran khusus seperti diskon untuk pengguna yang sudah tidak pernah berbelanja minimal 1 bulan untuk memancing pembelian
 - Kirim notifikasi atau promosi yang menarik untuk merangsang minat pembelian seperti email yang memperkenalkan produk atau layanan baru.
 - Melakukan survei untuk memahami penyebab rendahnya aktivitas dan menyesuaikan strategi berdasarkan umpan balik pelanggan.

Kesimpulan

- Algoritma *clustering* untuk melakukan segmentasi pelanggan yang telah dievaluasi adalah K-Means, K-Means++, DenMune, dan K-Medoids with BanditPAM++
 - Berdasarkan hasil evaluasi algoritma yang berhasil melakukan clustering adalah K-Means, K-Means++, dan K-Medoids
 - Hasil *cluster* yang dipilih adalah K-Medoids
 - Hasil *cluster* diimplementasikan strategi bisnis/marketing untuk setiap karakteristik cluster demi menaikkan keuntungan perusahaan
-





THANK YOU

