

# PROJECT CAPSTONE DICODING

DASHBOARD ANALISIS PENJUALAN UMKM FASHION  
UNTUK MENGIDENTIFIKASI TREND KONSUMEN DAN  
STRATEGI OPTIMASI PENDAPATAN

Disusun Oleh Kelompok DB4-PS002

# ANGGOTA KELompok

- B244035E - Adilah Widiasti - Aktif
- B244010E - I Gusti Bagus Ramadha Saverian  
Ranuh - Aktif
- B244012E - Muhammad Adib - Aktif
- B244003E - Zefanya Danovanta Tarigan - Aktif

# LATAR BELAKANG

- UMKM fashion berperan penting dalam perekonomian Indonesia, namun banyak menghadapi tantangan dalam pengelolaan penjualan dan pemasaran. Perkembangan teknologi membuka peluang bagi UMKM untuk beralih ke platform online, tetapi banyak pelaku usaha masih kesulitan dalam menganalisis data penjualan untuk strategi bisnis yang efektif.
- Industri fashion yang dinamis membutuhkan solusi berbasis data untuk memahami tren pasar, perilaku konsumen, serta mengoptimalkan penjualan dan stok. Tanpa sistem yang tepat, pengambilan keputusan menjadi kurang akurat, menyebabkan pemborosan dan hilangnya peluang bisnis.

# TUJUAN

- Membangun dashboard interaktif untuk menampilkan data penjualan fashion secara real-time dan memberikan wawasan tentang kinerja produk, tren pasar, serta perilaku konsumen.
- Menyediakan analisis data guna mendukung keputusan cepat dan tepat dalam perencanaan stok, strategi harga, serta kampanye pemasaran, sehingga meningkatkan efisiensi operasional.
- Memanfaatkan data-driven insights untuk optimasi harga, stok, dan pengalaman pelanggan, yang berkontribusi pada peningkatan pendapatan perusahaan.

## ALASAN MENGAPA MEMILIH PROJECT TERSEBUT

- Tim kami memilih proyek "Dashboard Penjualan Fashion untuk Meningkatkan Pendapatan" karena industri fashion yang dinamis menghadapi tantangan dalam pemantauan penjualan dan pengambilan keputusan. Banyak perusahaan kesulitan menganalisis data secara real-time, yang berdampak pada efisiensi strategi bisnis.
- Dashboard ini akan membantu perusahaan mengidentifikasi tren, memahami perilaku konsumen, dan merencanakan strategi pemasaran yang lebih efektif.

## PROJECT SERUPA SEBELUMNYA

- Perancangan Dashboard Penjualan Pakaian di Toko Cleo dengan Power BI: Penelitian ini bertujuan merancang dashboard interaktif yang menyajikan data penjualan untuk mendukung keputusan berbasis data di Toko Cleo. Metode yang digunakan mencakup pendekatan sembilan langkah Kimball dalam perancangan dashboard, serta proses ETL (Extract, Transform, Load) untuk pengolahan data penjualan dari 2021 hingga 2023, dan visualisasi data menggunakan Power BI. Hasilnya, dashboard ini berhasil menyediakan analisis mendalam berdasarkan kategori produk dan periode waktu, meningkatkan akurasi, serta mempercepat proses pengambilan keputusan di Toko Cleo, sehingga memaksimalkan efektivitas operasional bisnis. ([ojs.stmik-banjarbaru.ac.id](http://ojs.stmik-banjarbaru.ac.id))

## PROJECT SERUPA SEBELUMNYA

- Analisis Kanal Penjualan Produk Fashion pada Tiga E-Marketplace di Indonesia [Studi Kasus: Ziahijab.com]: Penelitian ini bertujuan untuk menganalisis kanal penjualan terbaik dengan menggunakan studi kasus tiga public e-marketplace di Indonesia. Data yang digunakan adalah data salah satu UKM di bidang fashion. Penelitian ini bertujuan untuk memberikan rekomendasi e-marketplace yang tepat untuk penjualan produk fashion di Indonesia. ([repository.its.ac.id](http://repository.its.ac.id))

## IMPLEMENTASI/PENINGKATAN & PENALARAN

Proyek ini akan mengembangkan dashboard interaktif yang memungkinkan pemilik UMKM fashion untuk menganalisis data penjualan secara real-time. Dashboard ini akan mengintegrasikan berbagai sumber data (seperti penjualan, inventaris, dan perilaku konsumen) untuk membantu mengidentifikasi tren pasar, preferensi konsumen, dan faktor-faktor yang memengaruhi pendapatan. Fitur utama yang akan dikembangkan meliputi:

1. Visualisasi Penjualan
2. Segmentasi Konsumen
3. Prediksi Tren Penjualan
4. Optimalisasi Inventaris
5. Rekomendasi Strategi Harga

# TAMPAKAN DATASET KAGGLE

The screenshot shows the Kaggle interface for the 'Fashion Retail Sales' dataset. The top navigation bar includes 'kaggle', 'Create', 'Home', 'Competitions', 'Datasets' (which is selected), 'Models', 'Code', 'Discussions', 'Learn', and 'More'. To the right is a search bar, 'Sign In', and 'Register' buttons. The main title 'Fashion Retail Sales' is displayed with the subtitle 'Unlocking Customer Insights and Sales Trends in the Fashion Industry'. Below the title is the 'family Fashion Clothing' logo. A navigation bar at the bottom of the page includes 'Data Card', 'Code (4)', 'Discussion (0)', and 'Suggestions (0)'. The 'About Dataset' section contains information about the 'Fashion Retail Sales Dataset', its introduction, and usage details like 'Usability 7.06', 'License GPL 2', and 'Expected update frequency Not specified'. At the bottom, a table provides sample data from the dataset.

Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method
0	Handbag	4619.0	2023-02-05	NaN	Credit Card
1	Tunic	2456.0	2023-07-11	2.0	Credit Card
2	Tank Top	2102.0	2023-03-23	4.1	Cash
3	Leggings	3126.0	2023-03-15	3.2	Cash
4	Wallet	3003.0	2022-11-27	4.7	Cash

# DATA PREPROCESSING

## MISSING VALUE & DUPLICATE VALUE

2 kolom memiliki missing value:

- Kolom purchase null value akan di delete
- Kolom review null value akan di fill dengan mean

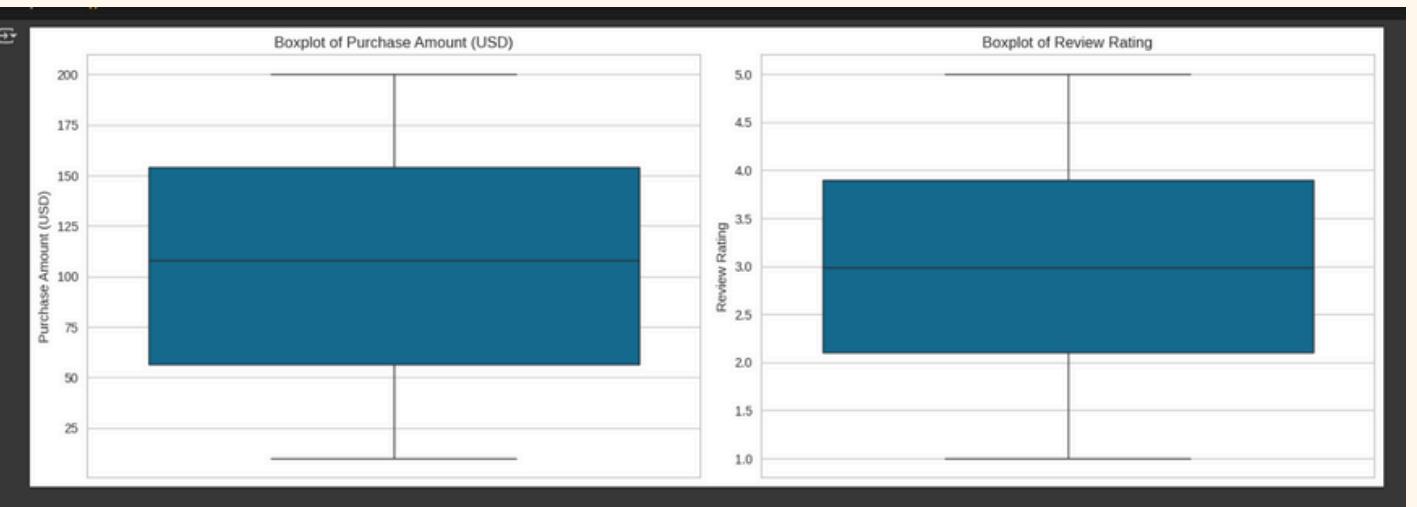
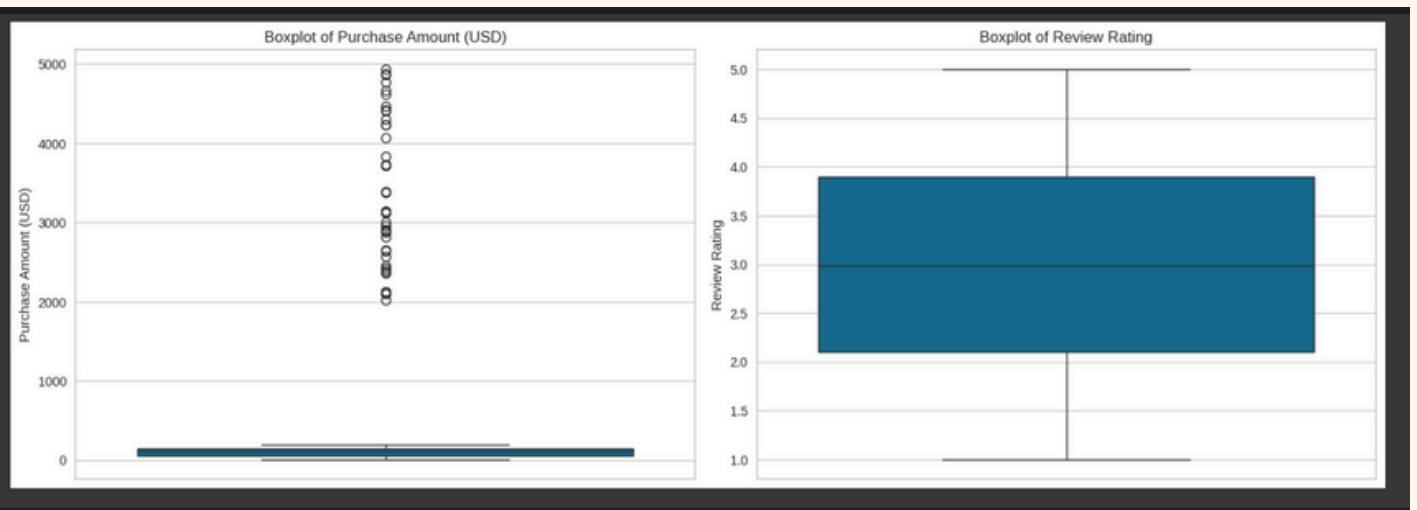
Tidak ada data duplikat.

	Data Type	Non-Null Count	Missing Values
Customer Reference ID	int64	3400	0
Item Purchased	object	3400	0
Purchase Amount (USD)	float64	2750	650
Date Purchase	object	3400	0
Review Rating	float64	3076	324
Payment Method	object	3400	0

```
[ ] df.duplicated().sum()  
→ 0
```

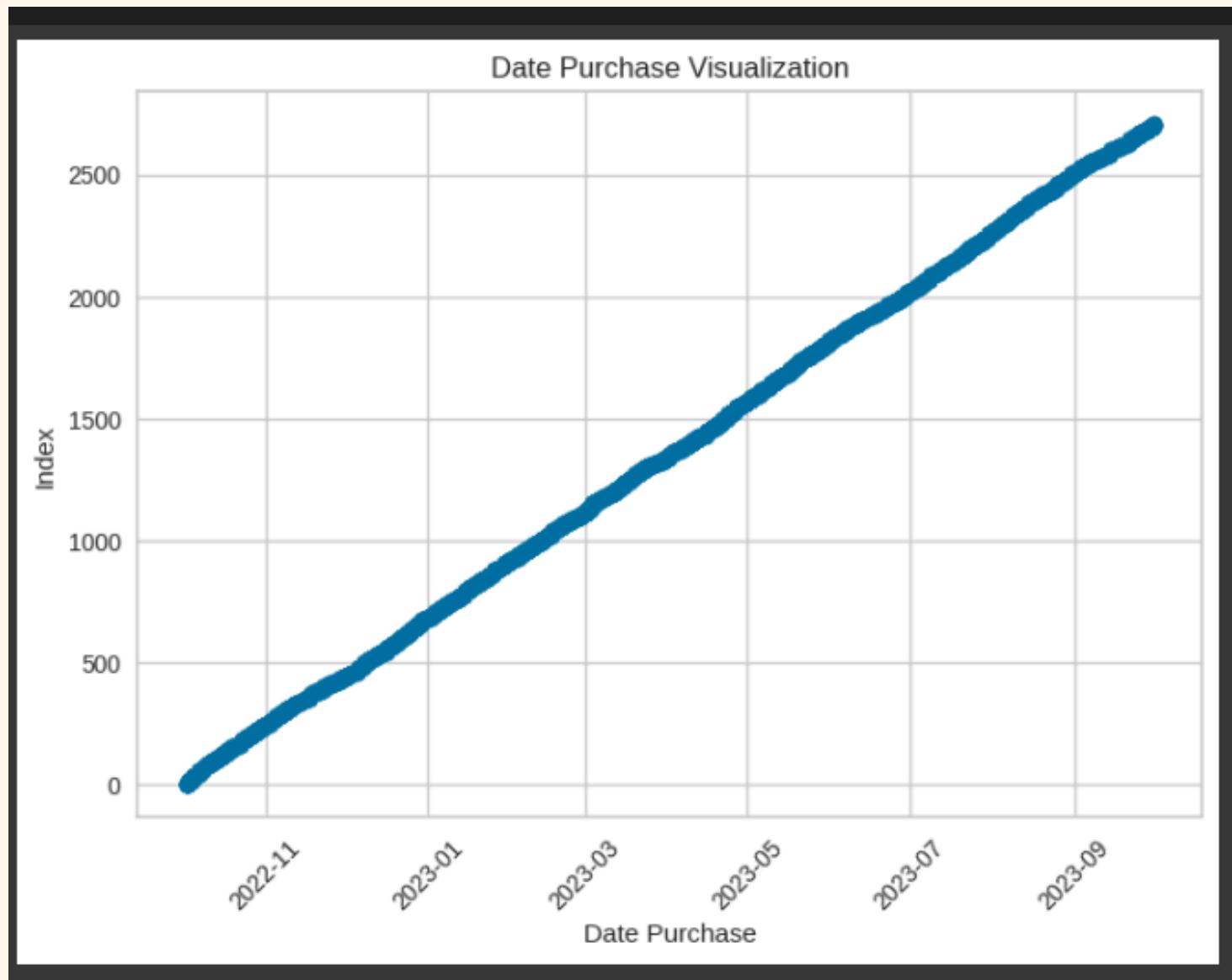
# DATA PREPROCESSING

## PEMANGKASAN OUTLIER



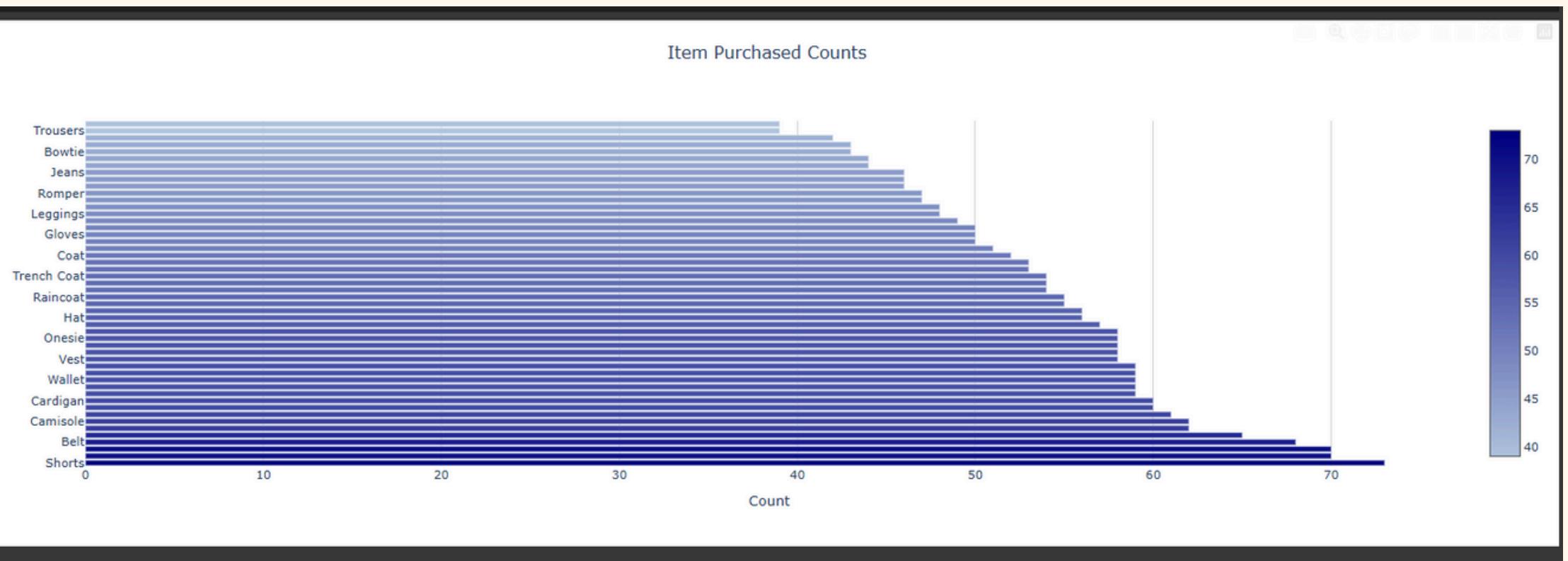
# EXPLORATORY DATA ANALYSIS

## PERTUMBUHAN PEMBELIAN



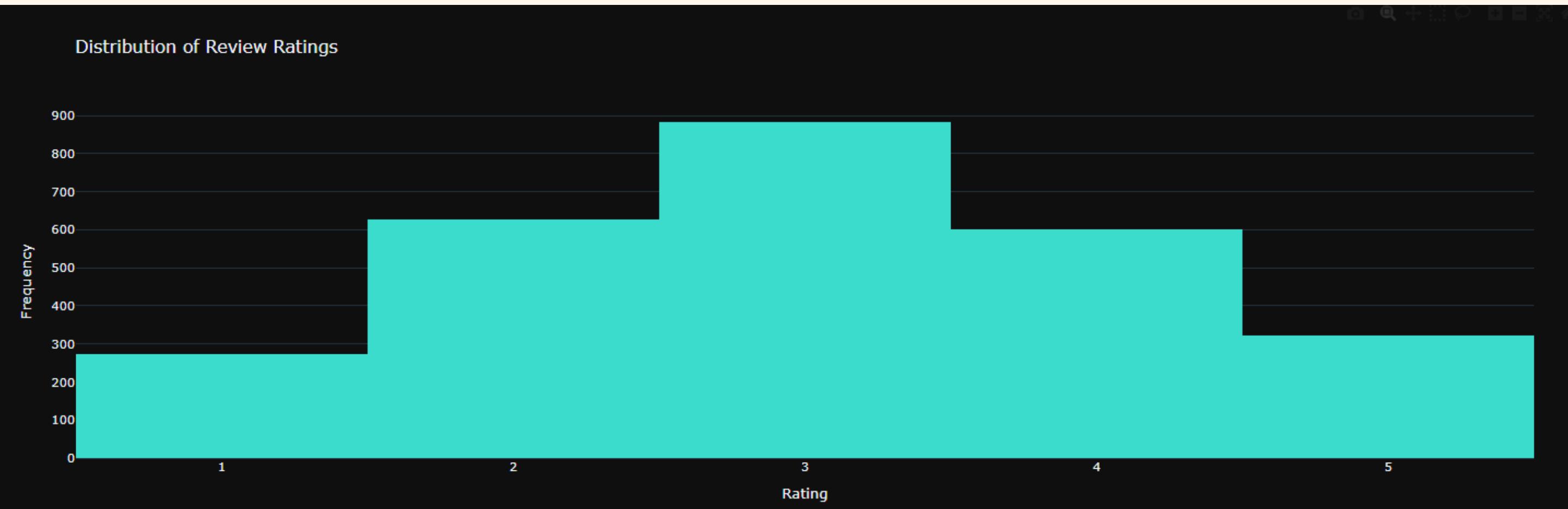
# EXPLORATORY DATA ANALYSIS

## PEMBELIAN ITEM TERBANYAK



# EXPLORATORY DATA ANALYSIS

## DISTRIBUSI REVIEW RATINGS (NORMAL DISTRIBUTION)



# CLUSTERING

## LABEL ENCODER

Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method
1392	4025	Blouse	126.0	2022-10-02	1.5
2286	4108	Handbag	50.0	2022-10-02	2.1
1250	4008	Belt	130.0	2022-10-02	3.4
3302	4093	Handbag	24.0	2022-10-02	2.8
1704	4014	Tank Top	99.0	2022-10-02	1.2



All features are now numerical

Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method
1392	4025	3	126.0	2022-10-02	1.5
2286	4108	13	50.0	2022-10-02	2.1
1250	4008	1	130.0	2022-10-02	3.4
3302	4093	13	24.0	2022-10-02	2.8
1704	4014	42	99.0	2022-10-02	1.2

# CLUSTERING

## FEATURE EXTRACTION & NORMALIZATION

All features are now numerical

Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method
1392	4025	3	126.0	2022-10-02	1.5
2286	4108	13	50.0	2022-10-02	2.1
1250	4008	1	130.0	2022-10-02	3.4
3302	4093	13	24.0	2022-10-02	2.8
1704	4014	42	99.0	2022-10-02	1.2



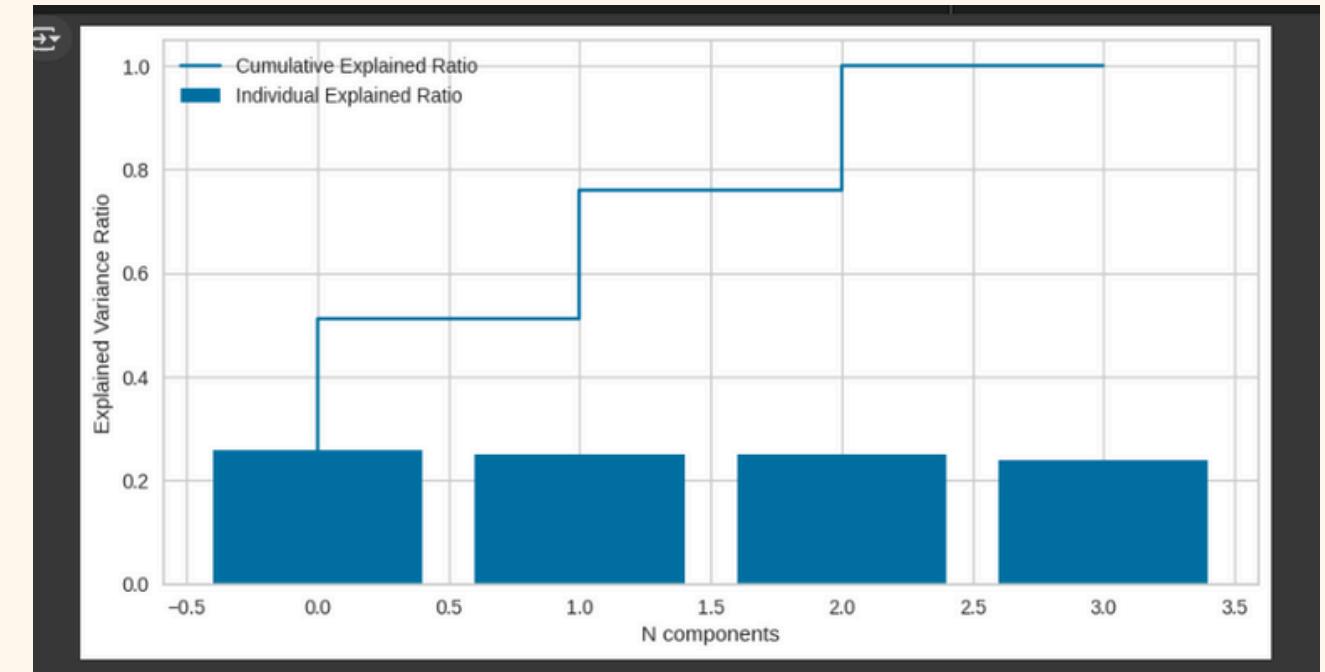
All features are now scaled

	Item Purchased	Purchase Amount (USD)	Review Rating	Payment Method
0	-1.495009	0.356177	-1.354772	-1.045744
1	-0.801461	-1.019092	-0.805916	0.956257
2	-1.633718	0.428560	0.383271	0.956257
3	-0.801461	-1.489579	-0.165585	0.956257
4	1.209827	-0.132405	-1.629200	-1.045744

# CLUSTERING

## EXPLAINED RATIO

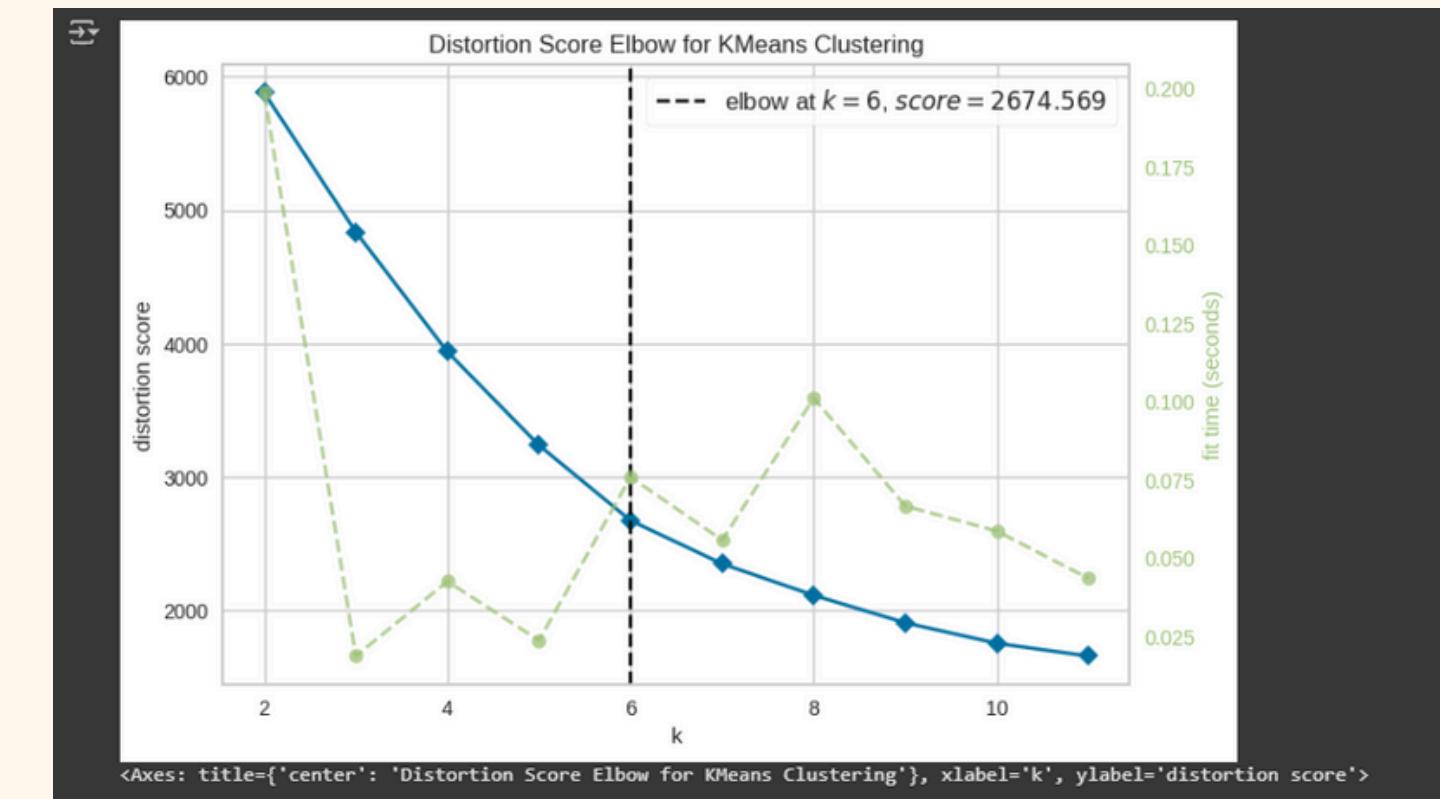
- membantu kita menentukan jumlah komponen utama yang optimal untuk digunakan dalam analisis selanjutnya.
- Kita ingin memilih jumlah komponen utama yang cukup untuk menjelaskan sebagian besar varians data dengan nilai 0.9 - 0.95 dan menggunakan jumlah komponen utama yang sesuai.



# CLUSTERING

## DISTORTION SCORE

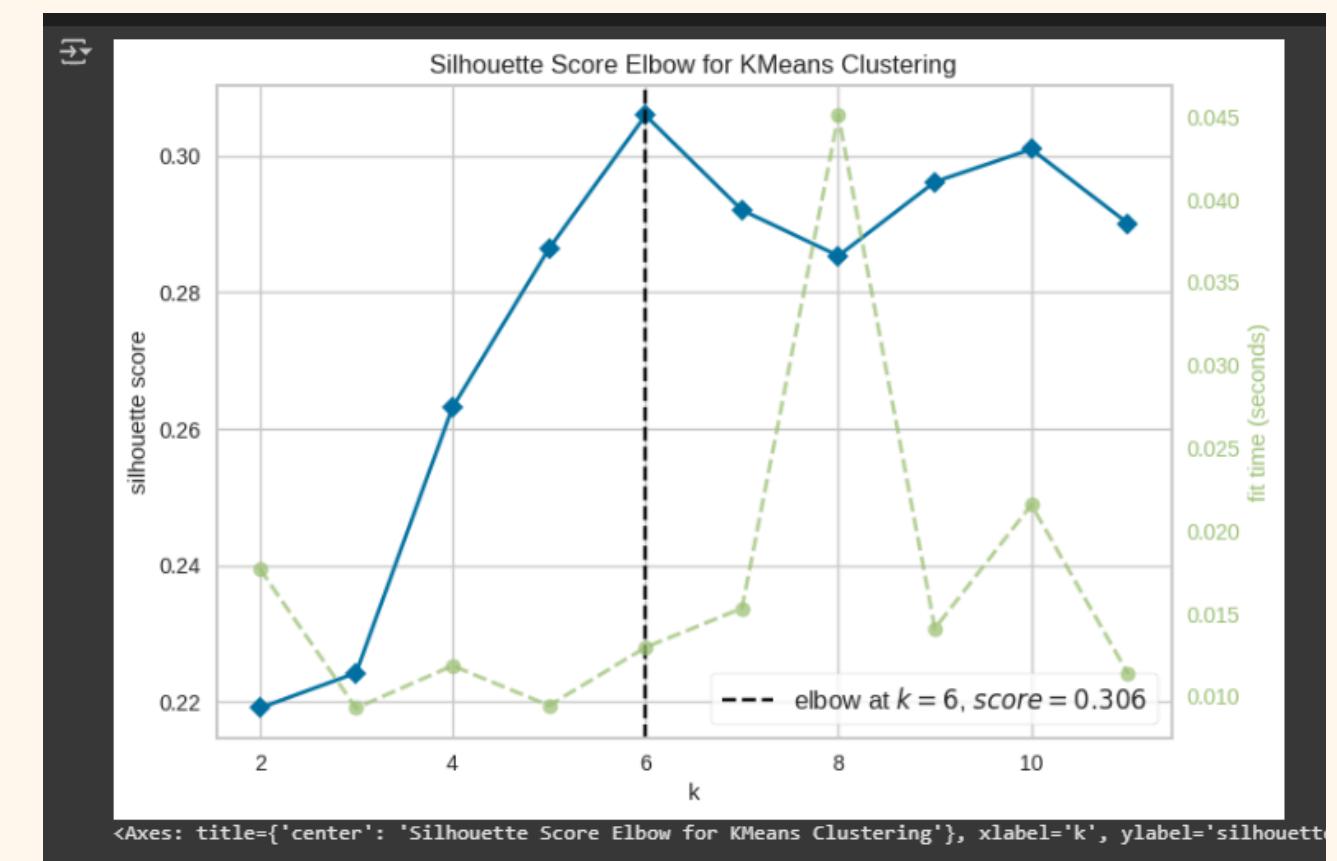
- Distortion score cenderung menurun seiring dengan bertambahnya jumlah cluster ( $k$ ). Ini karena semakin banyak cluster, semakin dekat titik-titik data dalam satu cluster ke pusat clusternya.
- Elbow Point: Pada gambar ini, "elbow" atau titik tekuk terlihat jelas pada  $k = 6$ .



# CLUSTERING

## SILHOUETTE SCORE

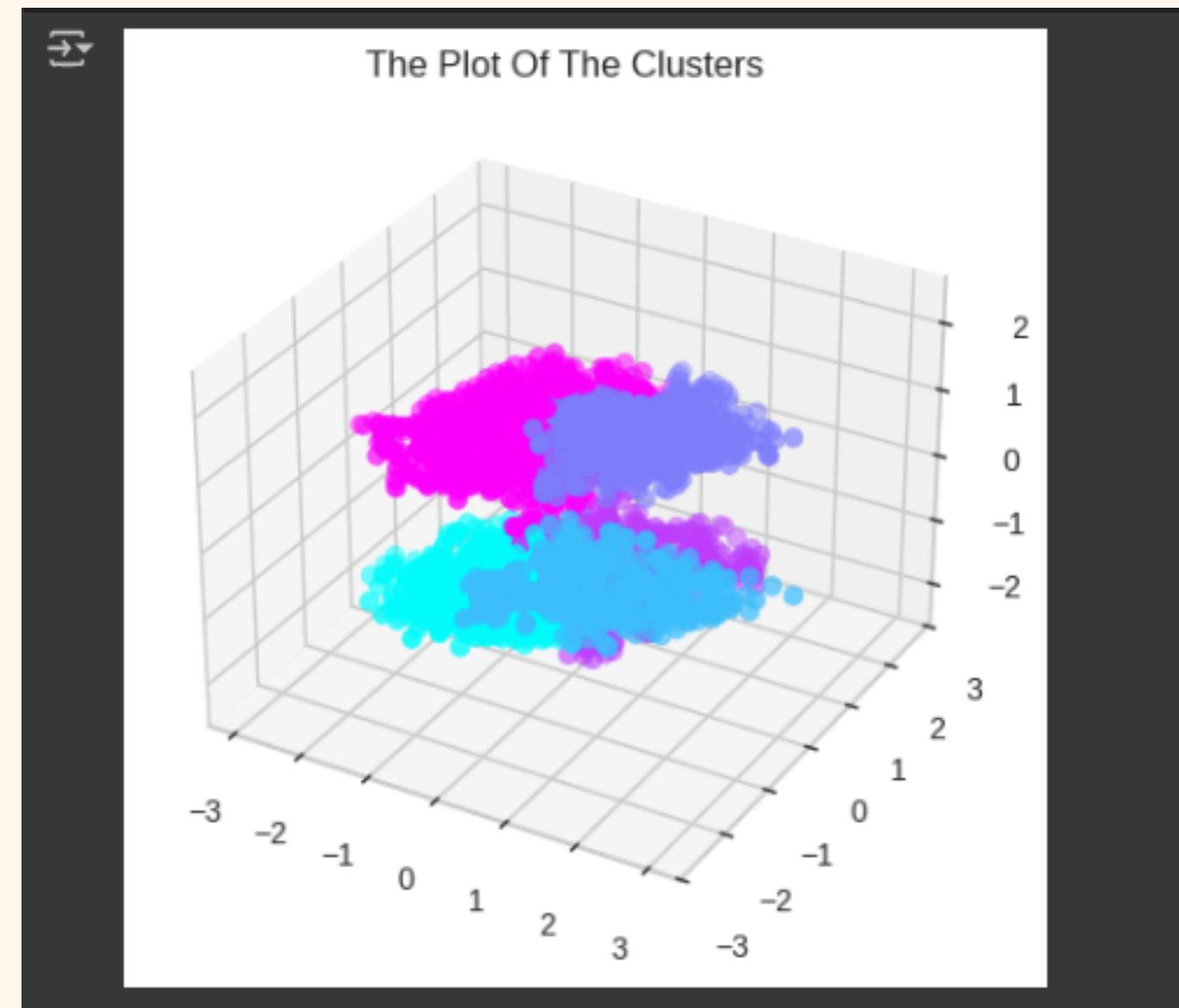
- Silhouette score umumnya meningkat seiring dengan bertambahnya jumlah cluster ( $k$ ), mencapai puncaknya pada titik tertentu, kemudian menurun lagi.
- Elbow Point: Pada gambar ini, "elbow" atau titik tertinggi silhouette score terlihat jelas pada  $k = 6$ .



# CLUSTERING

## PCA VISUALIZATION

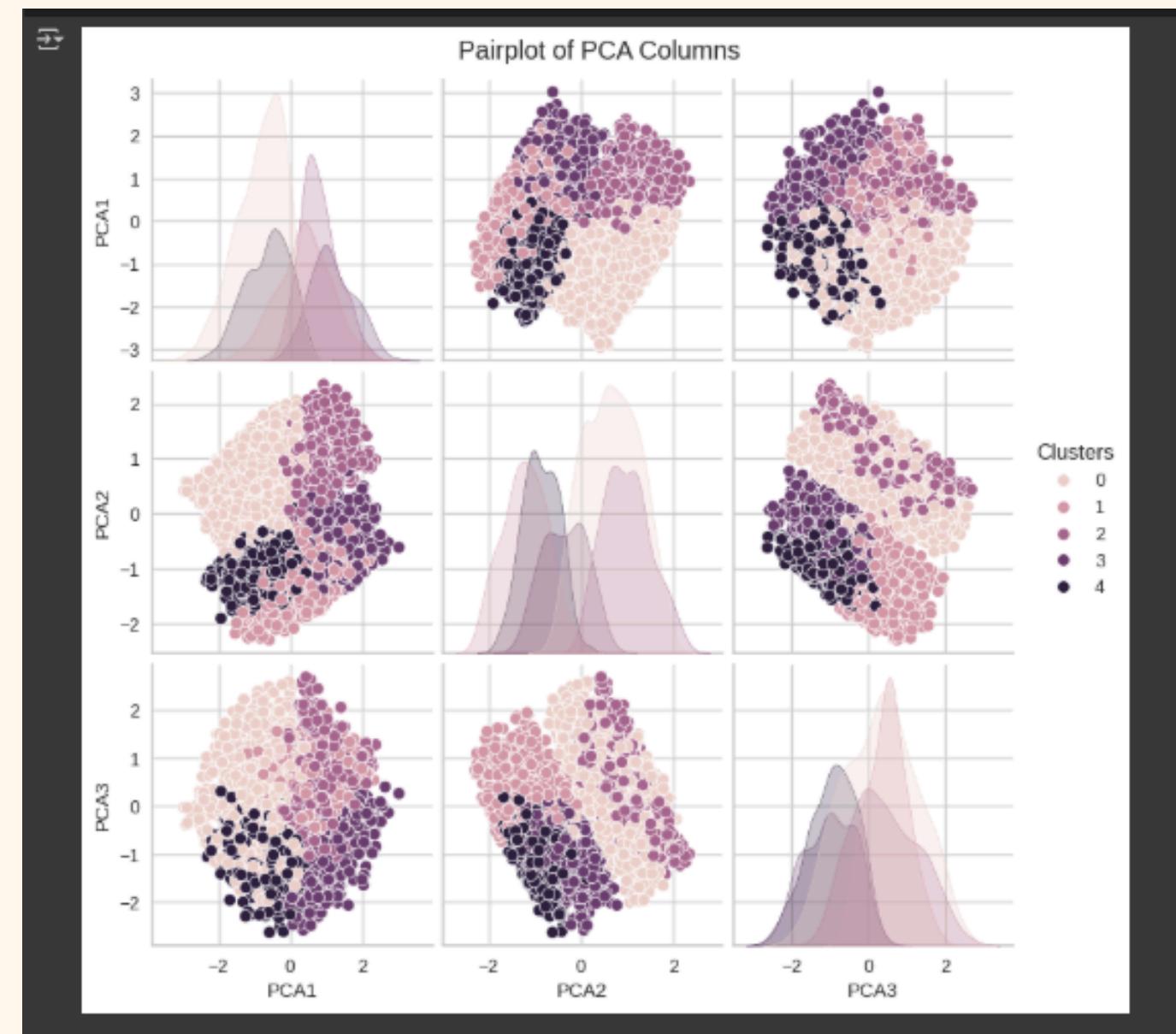
- Gambar 3D scatter plot ini memberikan wawasan tentang bagaimana data terdistribusi dan mengelompok dalam tiga dimensi. Ini membantu kita memahami hubungan antara tiga fitur yang berbeda dan mengidentifikasi potensi kelompok atau cluster dalam data.



# CLUSTERING

## PAIRPLOT PCA VISUALIZATION

- Gambar "Pairplot of PCA Columns" ini memberikan wawasan tentang bagaimana data terdistribusi dan mengelompok setelah mengalami reduksi dimensi menggunakan PCA.



# CLUSTERING

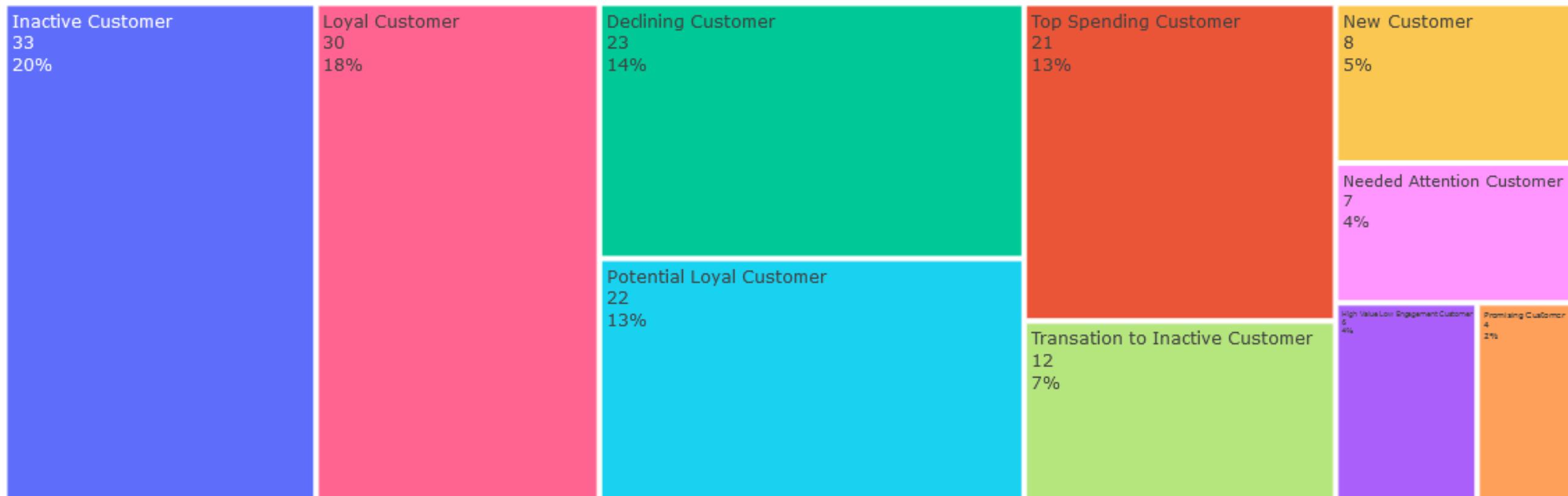
## RFM ANALYSIS

mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. Pertama, kode menghitung Recency (berapa minggu sejak pembelian terakhir), Frequency (jumlah pembelian), dan Monetary (total pengeluaran dalam USD) untuk setiap pelanggan.

Customer Reference ID	Recency	Frequency	Monetary
3957	12.142857	11	164
3958	2.428571	17	190
3959	6.857143	19	292
3960	2.428571	15	204
3961	4.714286	17	258

# CLUSTERING

## Distribution of the RFM Segments



# CLASSIFICATION MODEL

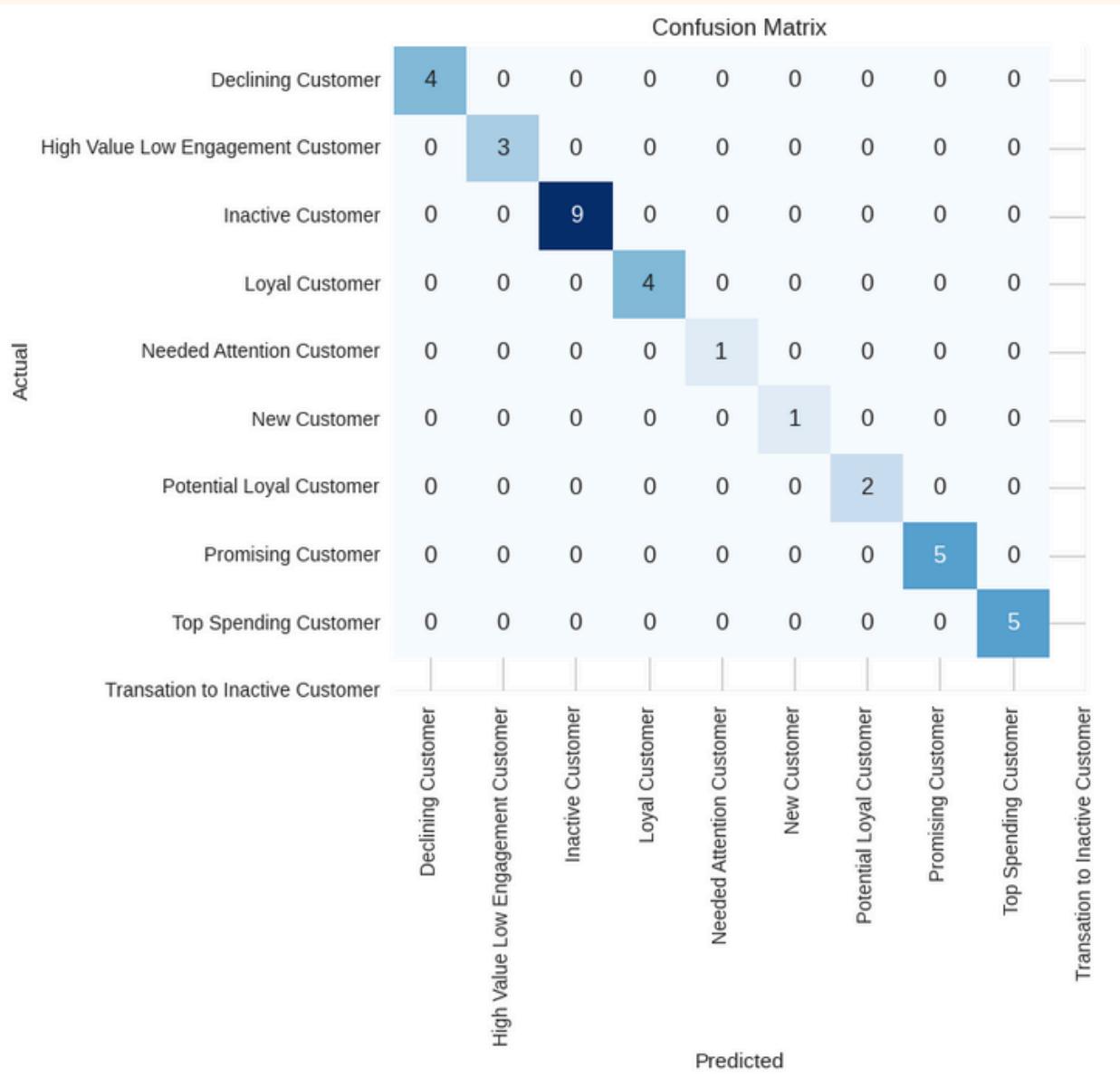
## DATASET BARU

Menggunakan dataset baru setelah  
melakukan clustering dan RFM analysis

Customer Reference ID	Recency	Frequency	Monetary	R	F	M	RFM_SCORE	segment
3957	12.142857	11	164	1	1	1	11	Inactive Customer
3958	2.428571	17	190	3	3	1	33	Needed Attention Customer
3959	6.857143	19	292	1	4	4	14	Declining Customer
3960	2.428571	15	204	3	2	2	32	Transation to Inactive Customer
3961	4.714286	17	258	2	3	3	23	Declining Customer

# CLASSIFICATION MODEL

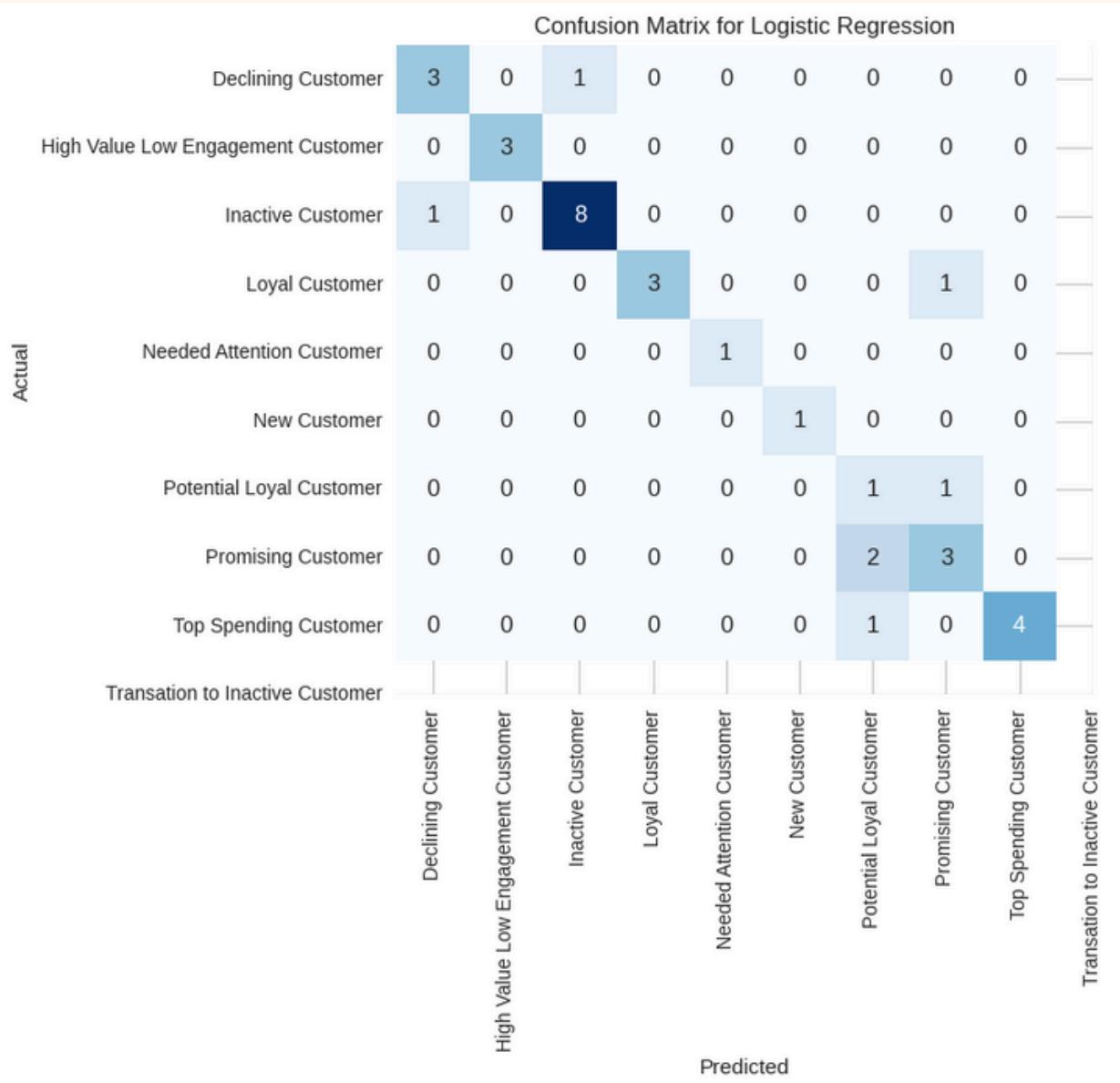
## SUPPORT VECTOR MACHINE



	precision	recall	f1-score	support
0	1.00	1.00	1.00	4
1	1.00	1.00	1.00	3
2	1.00	1.00	1.00	9
3	1.00	1.00	1.00	4
4	1.00	1.00	1.00	1
5	1.00	1.00	1.00	1
6	1.00	1.00	1.00	2
8	1.00	1.00	1.00	5
9	1.00	1.00	1.00	5
accuracy			1.00	34
macro avg	1.00	1.00	1.00	34
weighted avg	1.00	1.00	1.00	34
Accuracy: 1.0				

# CLASSIFICATION MODEL

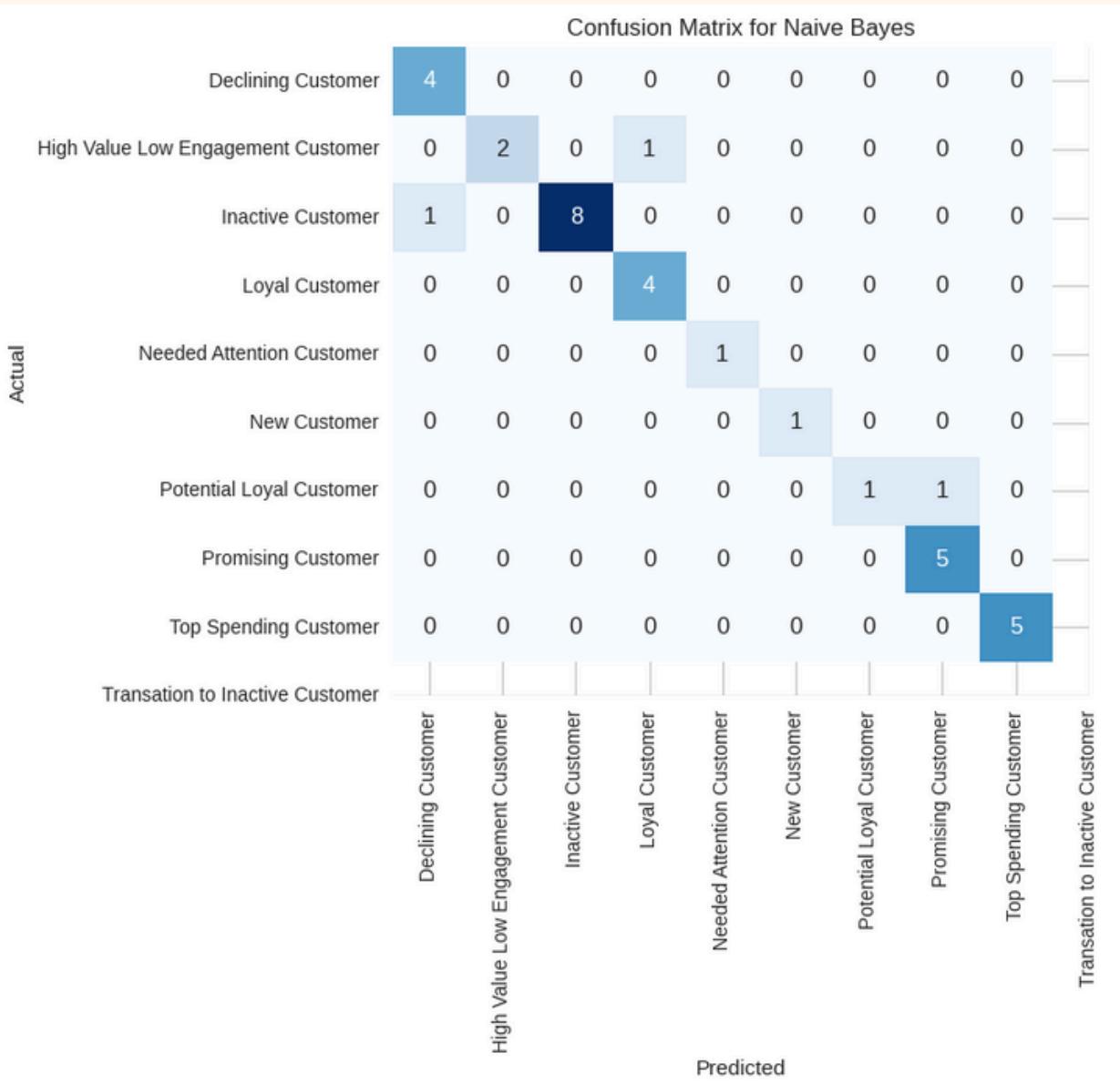
## LOGISTIC REGRESSION



	precision	recall	f1-score	support
0	0.75	0.75	0.75	4
1	1.00	1.00	1.00	3
2	0.89	0.89	0.89	9
3	1.00	0.75	0.86	4
4	1.00	1.00	1.00	1
5	1.00	1.00	1.00	1
6	0.25	0.50	0.33	2
8	0.60	0.60	0.60	5
9	1.00	0.80	0.89	5
accuracy			0.79	34
macro avg	0.83	0.81	0.81	34
weighted avg	0.84	0.79	0.81	34
Accuracy:	0.7941176470588235			

# CLASSIFICATION MODEL

## NAIVE BAYES

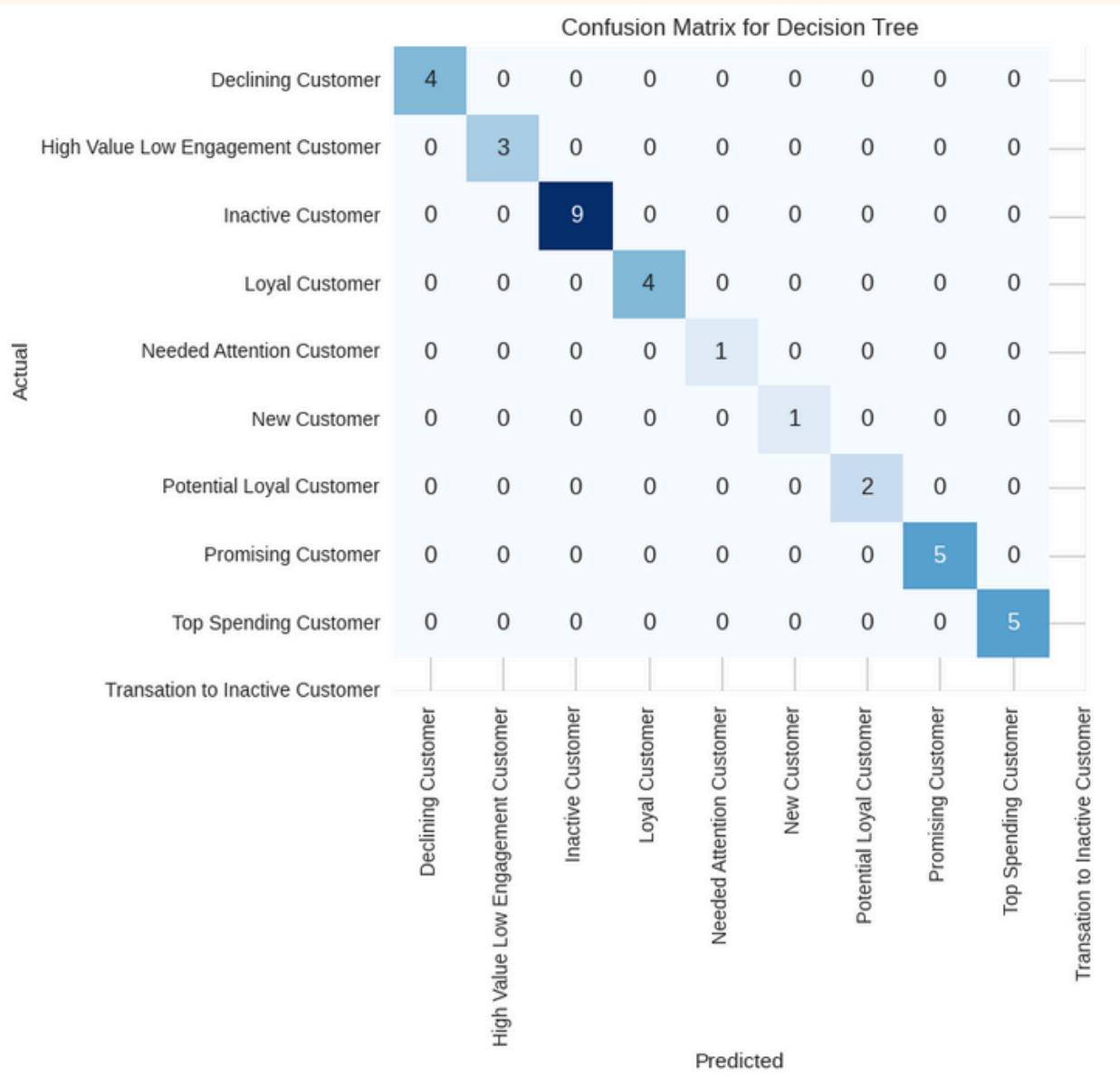


	precision	recall	f1-score	support
0	0.80	1.00	0.89	4
1	1.00	0.67	0.80	3
2	1.00	0.89	0.94	9
3	0.80	1.00	0.89	4
4	1.00	1.00	1.00	1
5	1.00	1.00	1.00	1
6	1.00	0.50	0.67	2
8	0.83	1.00	0.91	5
9	1.00	1.00	1.00	5
accuracy			0.91	34
macro avg	0.94	0.90	0.90	34
weighted avg	0.93	0.91	0.91	34

Accuracy: 0.9117647058823529

# CLASSIFICATION MODEL

## DECISION TREE



	precision	recall	f1-score	support
0	1.00	1.00	1.00	4
1	1.00	1.00	1.00	3
2	1.00	1.00	1.00	9
3	1.00	1.00	1.00	4
4	1.00	1.00	1.00	1
5	1.00	1.00	1.00	1
6	1.00	1.00	1.00	2
8	1.00	1.00	1.00	5
9	1.00	1.00	1.00	5
accuracy			1.00	34
macro avg	1.00	1.00	1.00	34
weighted avg	1.00	1.00	1.00	34
Accuracy: 1.0				

# TIME SERIES FORECASTING

## PRICE VS DATE ANALYSIS

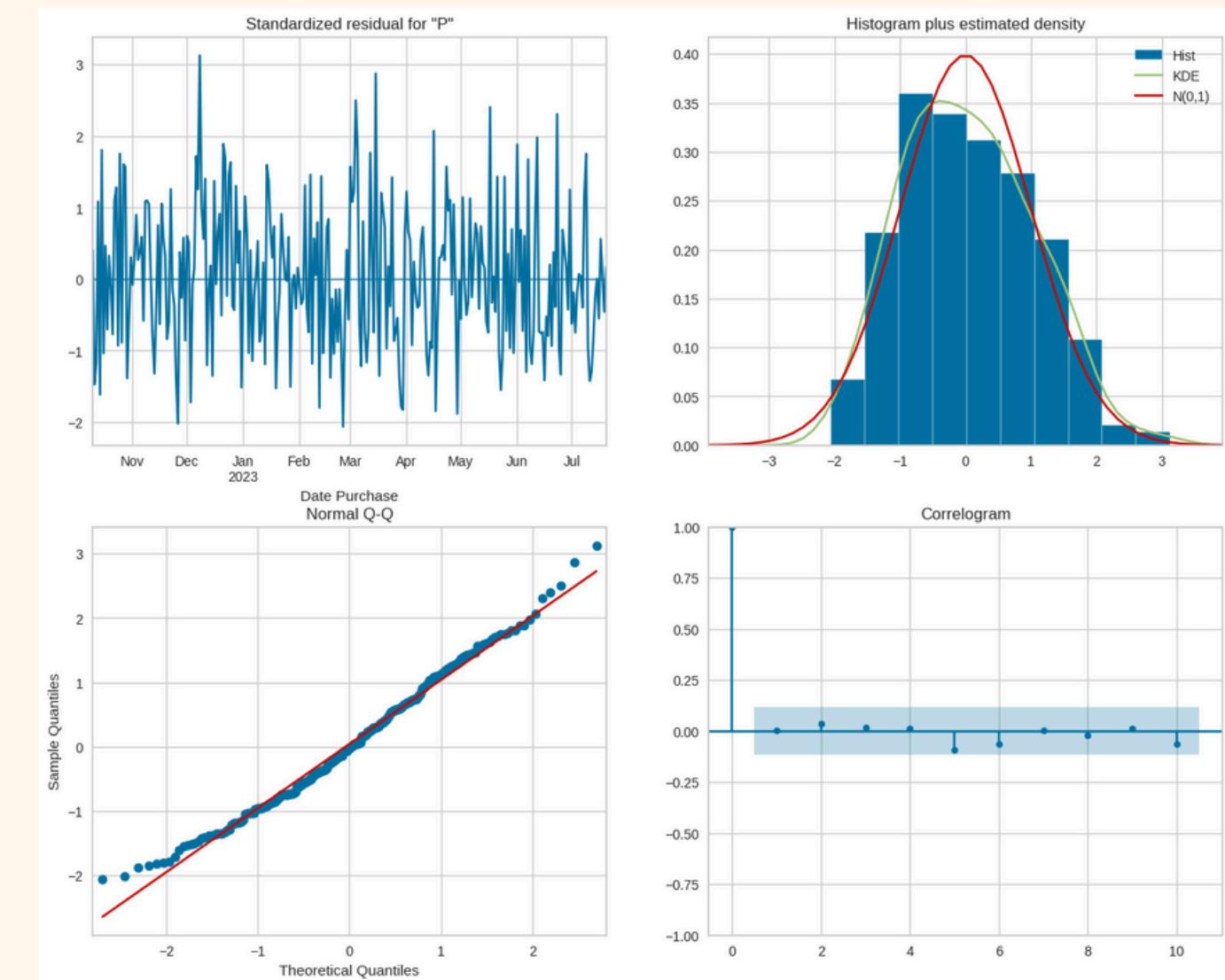
Grafik "Price vs. Date" ini memberikan gambaran tentang dinamika harga aset dari waktu ke waktu. Fluktuasi harga yang tinggi menunjukkan bahwa aset ini memiliki risiko yang tinggi juga. Investor atau trader yang tertarik dengan aset ini perlu dengan hati-hati mempertimbangkan risiko ini.



# TIME SERIES FORECASTING

## PLOT DIAGNOSTICS

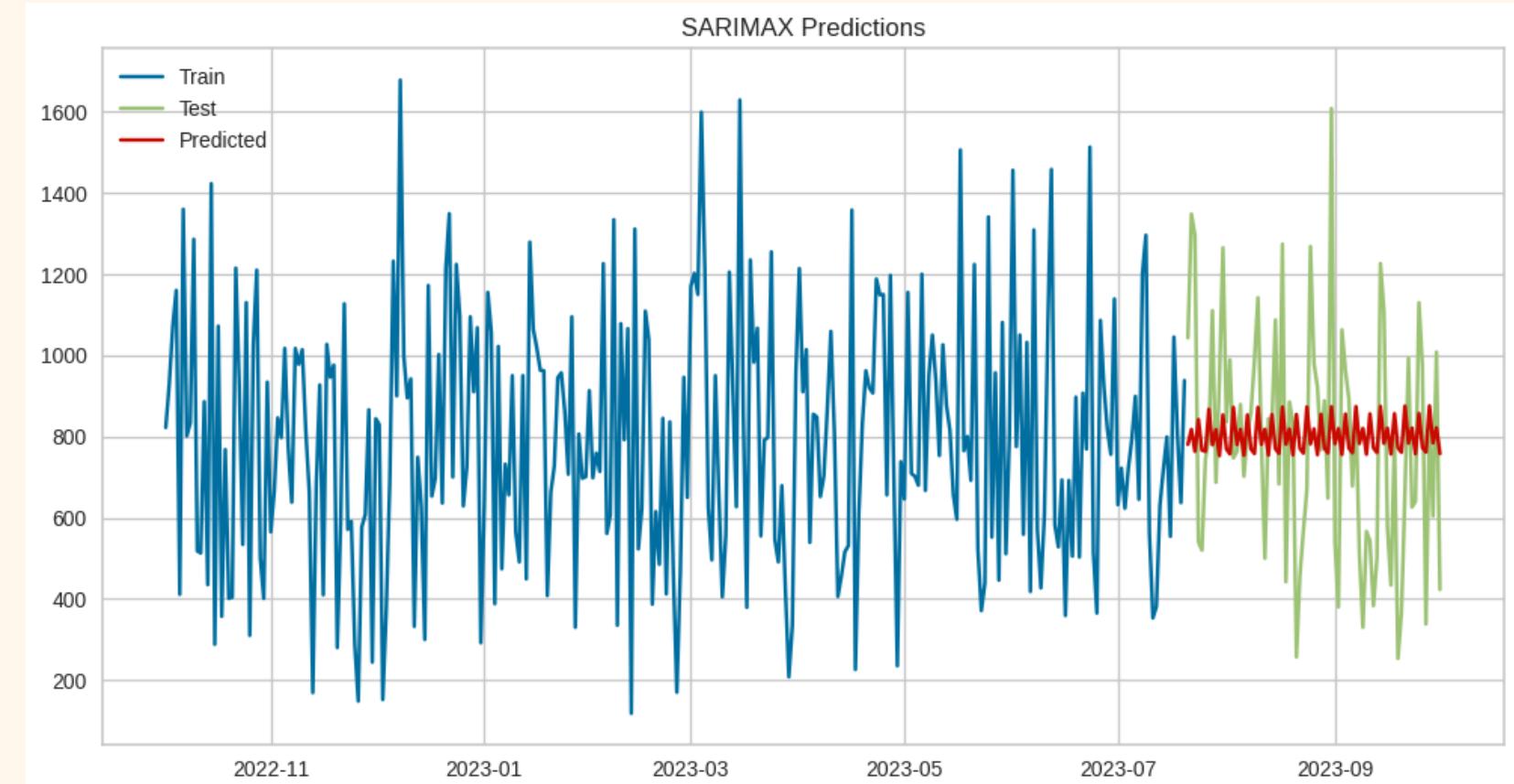
Plot diagnostik ini menunjukkan bahwa model regresi yang dianalisis cukup baik dan memenuhi asumsi-asumsi regresi linear. Namun, ada beberapa penyimpangan kecil dari normalitas pada Normal Q-Q plot, yang mungkin perlu dipertimbangkan jika normalitas adalah asumsi yang sangat penting dalam konteks ini.



# TIME SERIES FORECASTING

## USE TRAIN DATA TO PREDICT TEST DATA

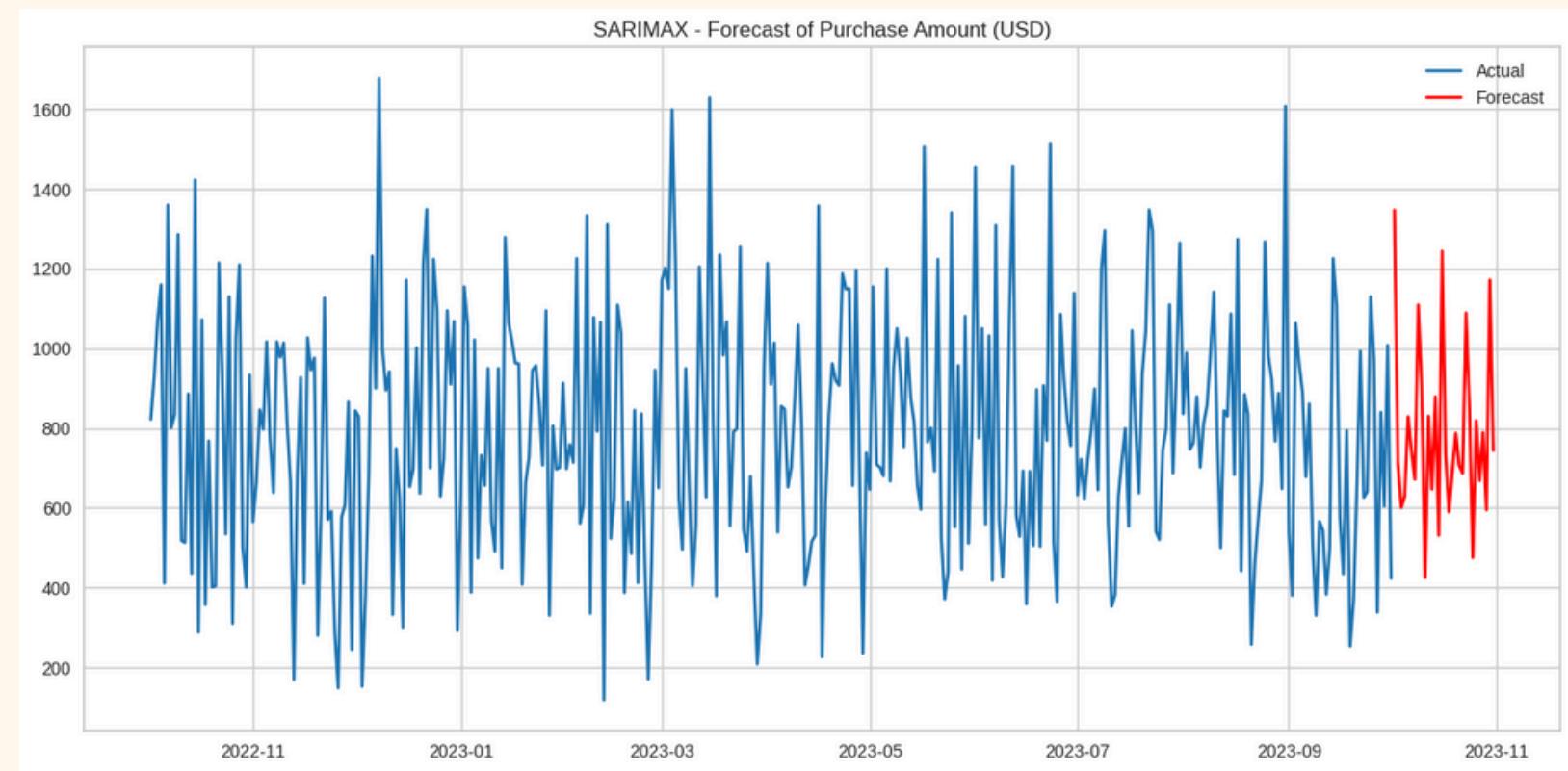
Dari hasil prediksi menggunakan model SARIMAX, terlihat bahwa model tersebut mampu memprediksi nilai Purchase Amount (USD) dengan tingkat kesalahan yang diukur melalui RMSE (Root Mean Squared Error) sebesar 279.91. Nilai ini dapat dianggap cukup baik, tergantung pada skala data dan konteks bisnisnya.



# TIME SERIES FORECASTING

## FORECAST 1 MONTH AHEAD

- Gambar di samping di atas menggunakan model SARIMAX untuk forecasting Purchase Amount (USD) selama 1 bulan ke depan menggunakan data 2 bulan sebelumnya.
- Model SARIMAX terlihat mampu mengikuti tren data historis, meskipun terdapat deviasi pada beberapa titik.



# COLLABORATIVE FILTERING

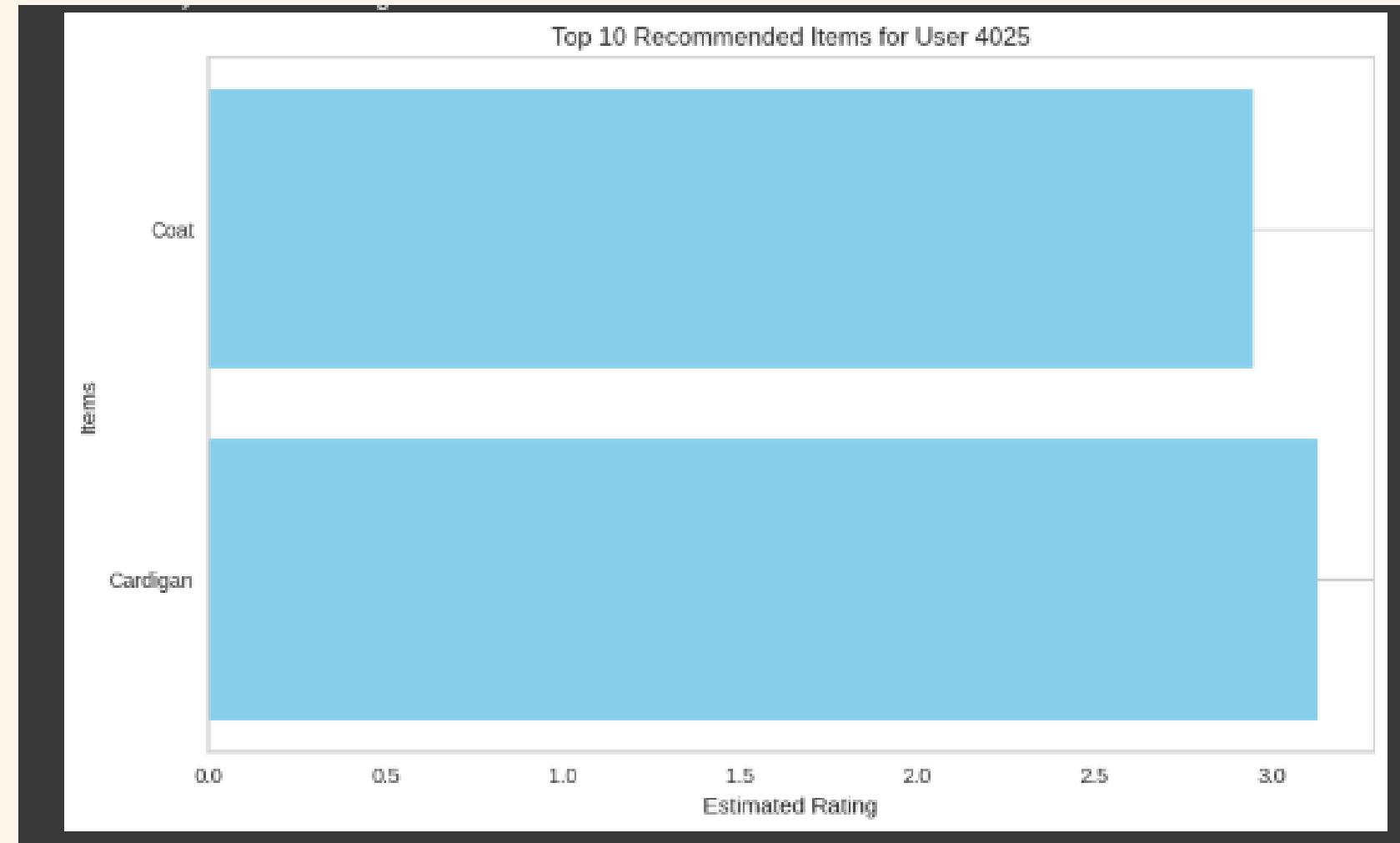
## DATASET

Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method	Clusters	
1392	4025	Blouse	126.0	2022-10-02	1.5	Cash	4
2286	4108	Handbag	50.0	2022-10-02	2.1	Credit Card	0
1250	4008	Belt	130.0	2022-10-02	3.4	Credit Card	0
3302	4093	Handbag	24.0	2022-10-02	2.8	Credit Card	0
1704	4014	Tank Top	99.0	2022-10-02	1.2	Cash	1

# COLLABORATIVE FILTERING

## PREDIKSI PREFERENSI USER

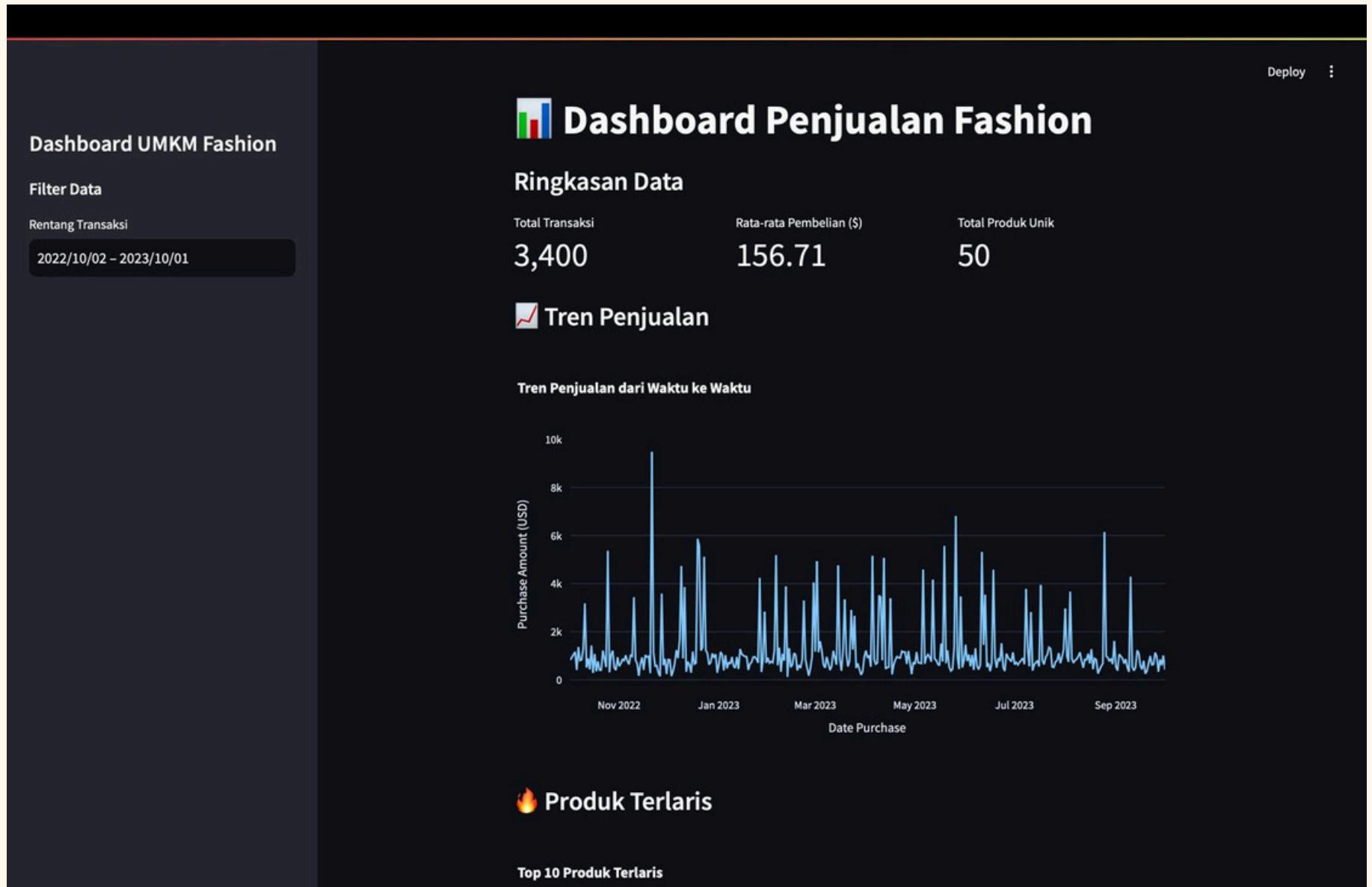
- Diagram batang horizontal ini menyajikan rekomendasi yang jelas dan mudah dipahami untuk pengguna dengan ID 4025.
- User dapat dengan cepat melihat item mana yang paling direkomendasikan dan mempertimbangkan untuk memilih item tersebut.



# TAMPILAN DASHBOARD METABASE



# TAMPAKAN DASHBOARD STREAMLIT



## KUMPULAN LINK PROJEK

- Link Dataset : <https://www.kaggle.com/datasets/fekihmea/fashion-retail-sales/data>
- Link Python : [https://colab.research.google.com/drive/15wr0rDD\\_0-FwTKoL\\_GMtp4NipW8YJrQn?usp=sharing#scrollTo=oLglutlb3L\\_Q](https://colab.research.google.com/drive/15wr0rDD_0-FwTKoL_GMtp4NipW8YJrQn?usp=sharing#scrollTo=oLglutlb3L_Q)
- Link Dashboard Streamlit : <https://supreme-orbit-76q9w6w4v4627w4-8501.app.github.dev/>
- Link Github : <https://github.com/RamadhaRanuh/Capstone-UMKM-Fashion>
- Link Presentasi :  
<https://www.canva.com/design/DAGdq0duGuk/efnlUfSQ6H43n2NjHZENqw/edit?ui=eyJEljp7IlAiOnsiQiI6ZmFsc2V9fx0>
- Link Video Presentasi :

# KESIMPULAN

- Proyek ini bertujuan untuk membangun sebuah dashboard prediksi penjualan Fashion yang dapat memberikan insight prediktif terkait penjualan fashion seperti penjualan pakaian, aksesoris dan sepatu. Berdasarkan data historis dan faktor-faktor yang mempengaruhi penjualan, seperti genre, harga, rating, dan lain-lain. Dashboard ini akan membantu pembeli, penjual pakaian, penjualan sepatu dan penjualan aksesoris untuk merencanakan strategi pemasaran, manajemen stok dan produk, perencanaan penjualan dan penentuan harga dengan lebih akurat dan berbasis data.

**TERIMAKASIH**