

E-COMMERCE CUSTOMER SEGMENTATION ANALYSIS



Presented By : M. Alwan Nur Ajie

Table Of Content

1

Background
Project

2

Data
Understanding

3

Data Cleaning

3

EDA

4

Preprocessing

5

Modelling - K
Means

6

Rekomendasi

Background

Background dari perusahaan merupakan e-commerce yang berbasis di Inggris yang telah menjual hadiah dan barang-barang rumah tangga untuk dewasa dan anak-anak melalui situs web. Dataset yang digunakan mencakup transaksi penjualan selama satu tahun (Desember 2018 hingga Desember 2019). Perusahaan ini perlu mengidentifikasi segmen pelanggan potensial untuk menargetkan pasar mereka dengan lebih baik dan lebih efisien. Informasi ini akan memastikan bahwa semua kampanye dan upaya pemasaran di masa depan diarahkan kepada pelanggan yang kemungkinan besar akan melakukan pembelian.

Objektif

- Mengidentifikasi segmen pelanggan potensial untuk membantu menargetkan pasar dengan lebih baik
- Dapat memberikan rekomendasi bisnis yang sesuai dengan karakteristik segmen pelanggan

Data Understanding

Dataset memiliki 536350 baris dan 8 kolom, dengan deskripsi kolom

- TransactionNo (categorical): nomor unik yang mendefinisikan setiap transaksi.
- Date (numeric): tanggal saat setiap transaksi dihasilkan.
- ProductNo (categorical): karakter unik yang digunakan untuk mengidentifikasi produk tertentu.
- Product (categorical): nama produk/barang
- Price (numeric): harga setiap produk per unit dalam pound sterling (£)
- Quantity (numeric): jumlah setiap produk per transaksi.
- CustomerNo (categorical): nomor unik yang mendefinisikan setiap pelanggan.
- Country (categorical): nama negara tempat tinggal pelanggan.

MISSING VALUE & DUPLICATE VALUE

Missing Value

Terdapat Missing Value di CustomerNo sebanyak 55 row (0,01%)

Handling Missing Value

Drop kolom yang terdapat missing Value

Duplicated Value

Terdapat 5200 duplicated Value

Handling Duplicated Value

Drop Kolom Duplicated Value

TREATMENT LANJUTAN

Kolom Quantity dan TransactionNo

Terdapat nilai dengan quantity minus dan TransactionNo yang nilainya berawalan 'C' yang berarti transaksi tersebut Cancel, Nilai ini invalid dalam transaksi maka akan kita drop
Jumlahnya ada 8494 kolom --> Drop

Kolom Date

Mengubah tipe data kolom date menjadi datetime

Membuat Fitur Revenue

Didapatkan dari kolom Price * Quantity



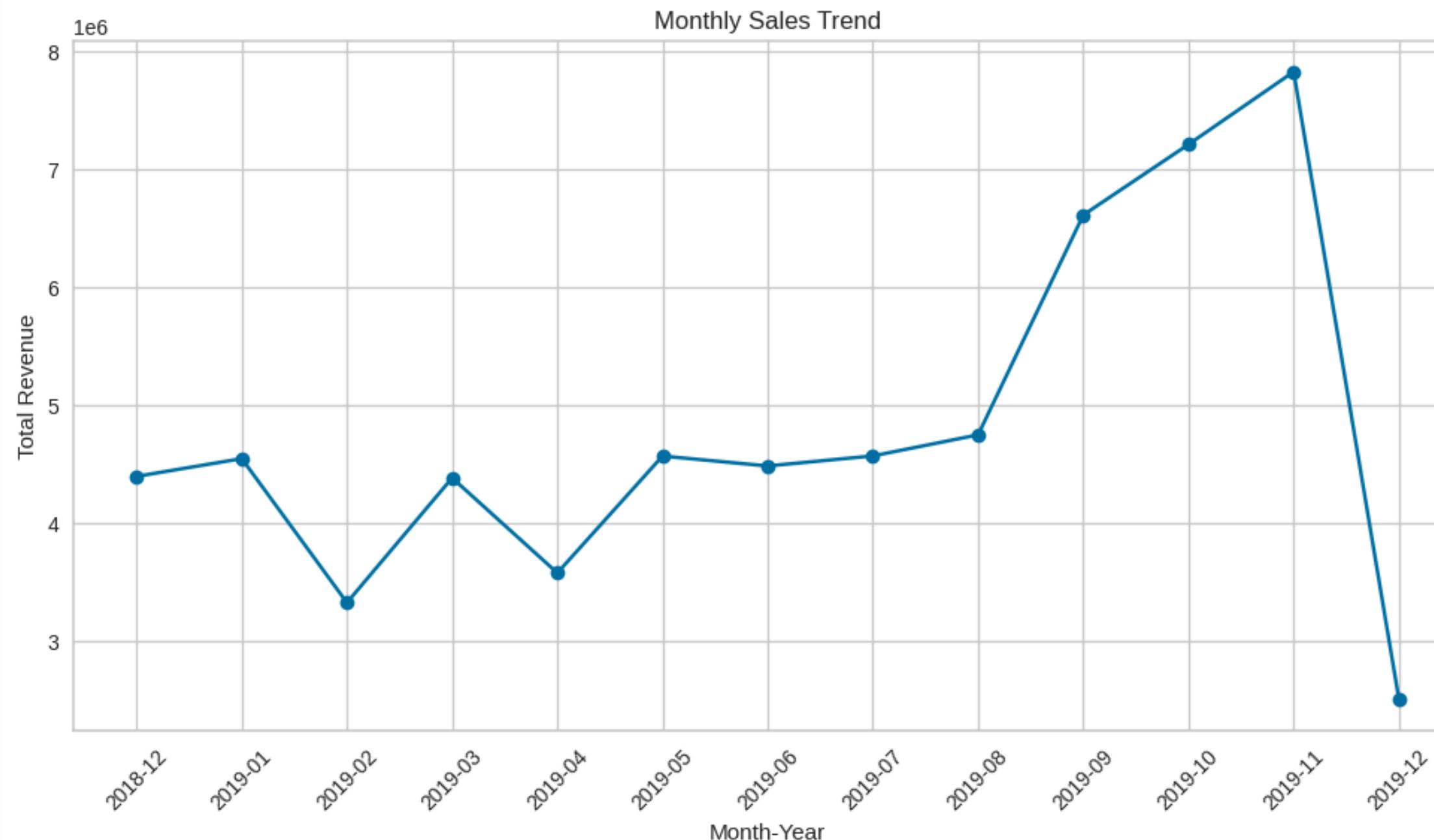
EDA

For Better Visualization:

<https://finalprojectcustsegment.streamlit.app>



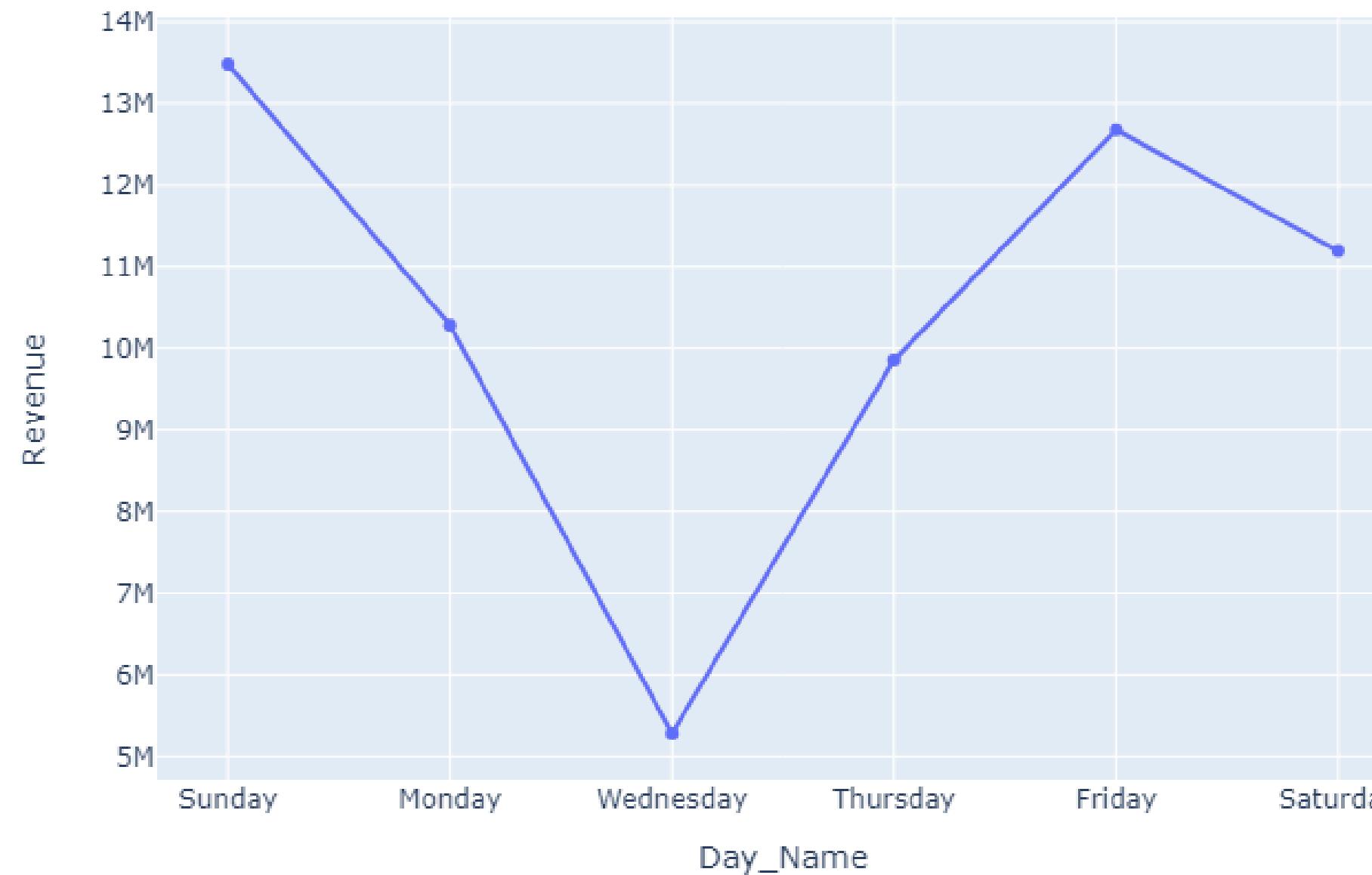
BAGAIMANA TREND PENJUALAN BERDASARKAN BULAN?



Tren fluktuatif di Desember 2018 – Mei 2019
Mei ke Agustus penjualan stagnan
setelahnya terjadi kenaikan hingga mencapai
puncaknya di November

BAGAIMANA TREND PENJUALAN BERDASARKAN HARI?

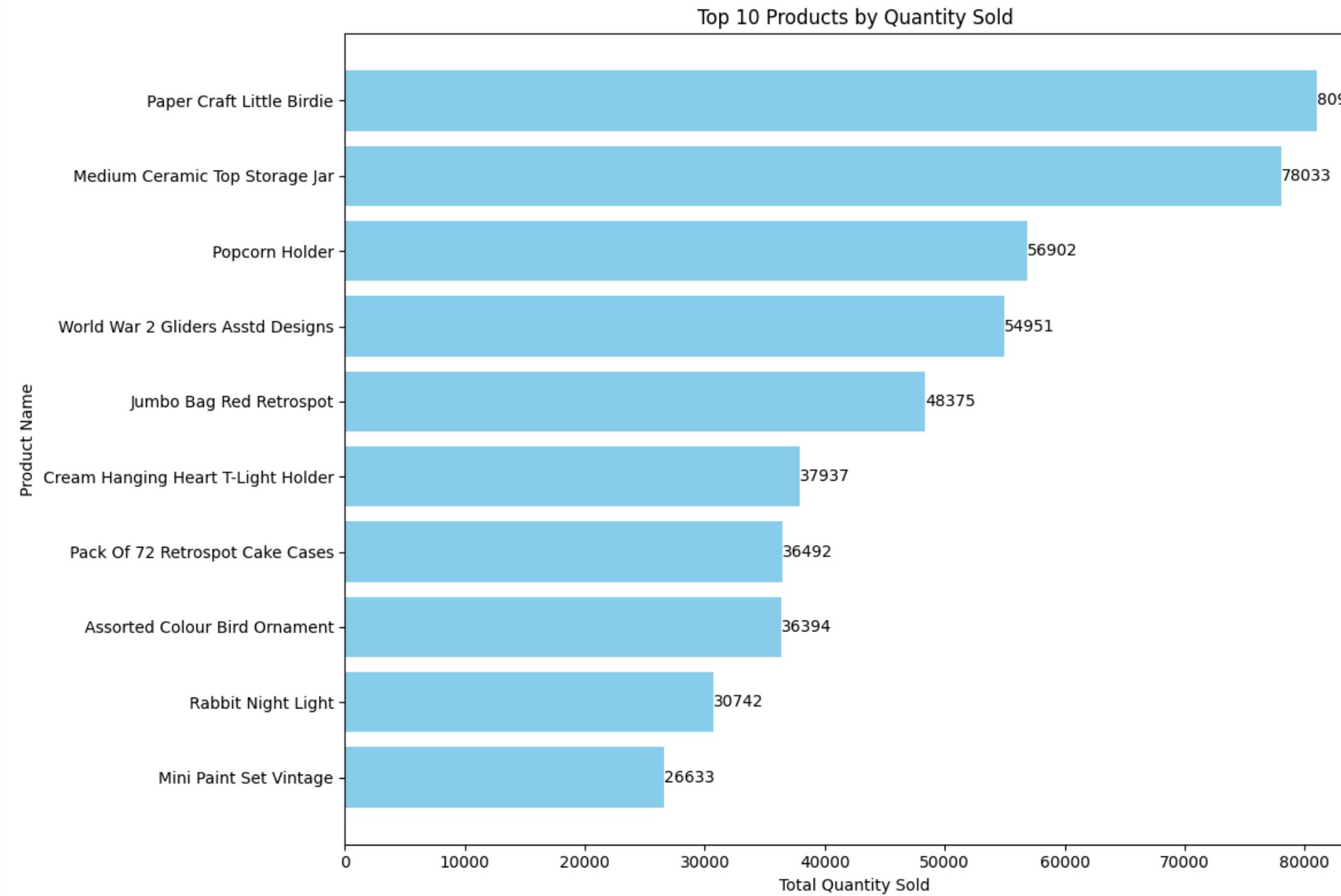
Sales Trend Over the Day



Pelanggan condong melakukan pembelian di weekend

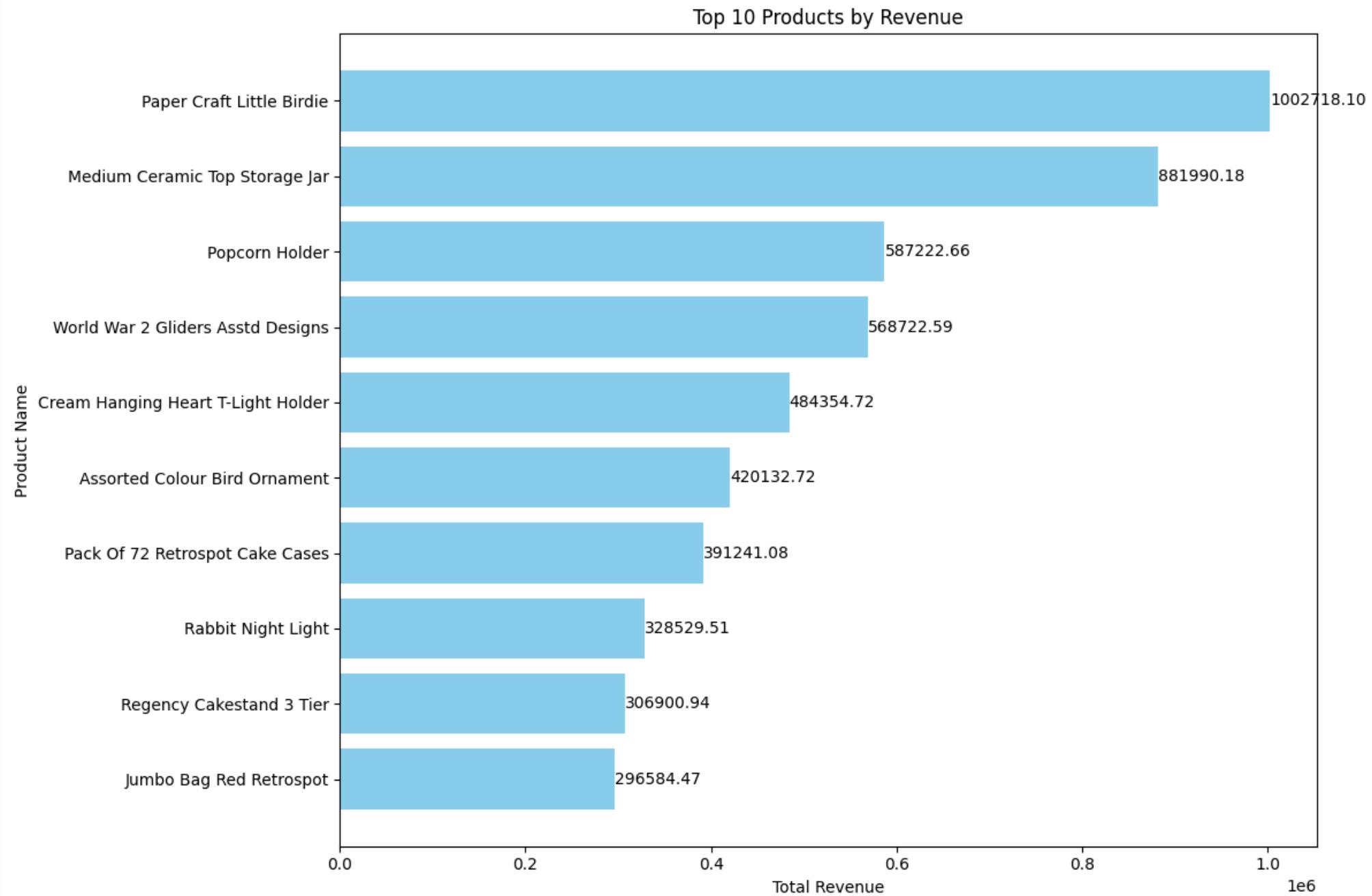
Hari Selasa storenya tutup

APA SAJA TOP 10 PRODUK YANG SERING DIBELI PELANGGAN?



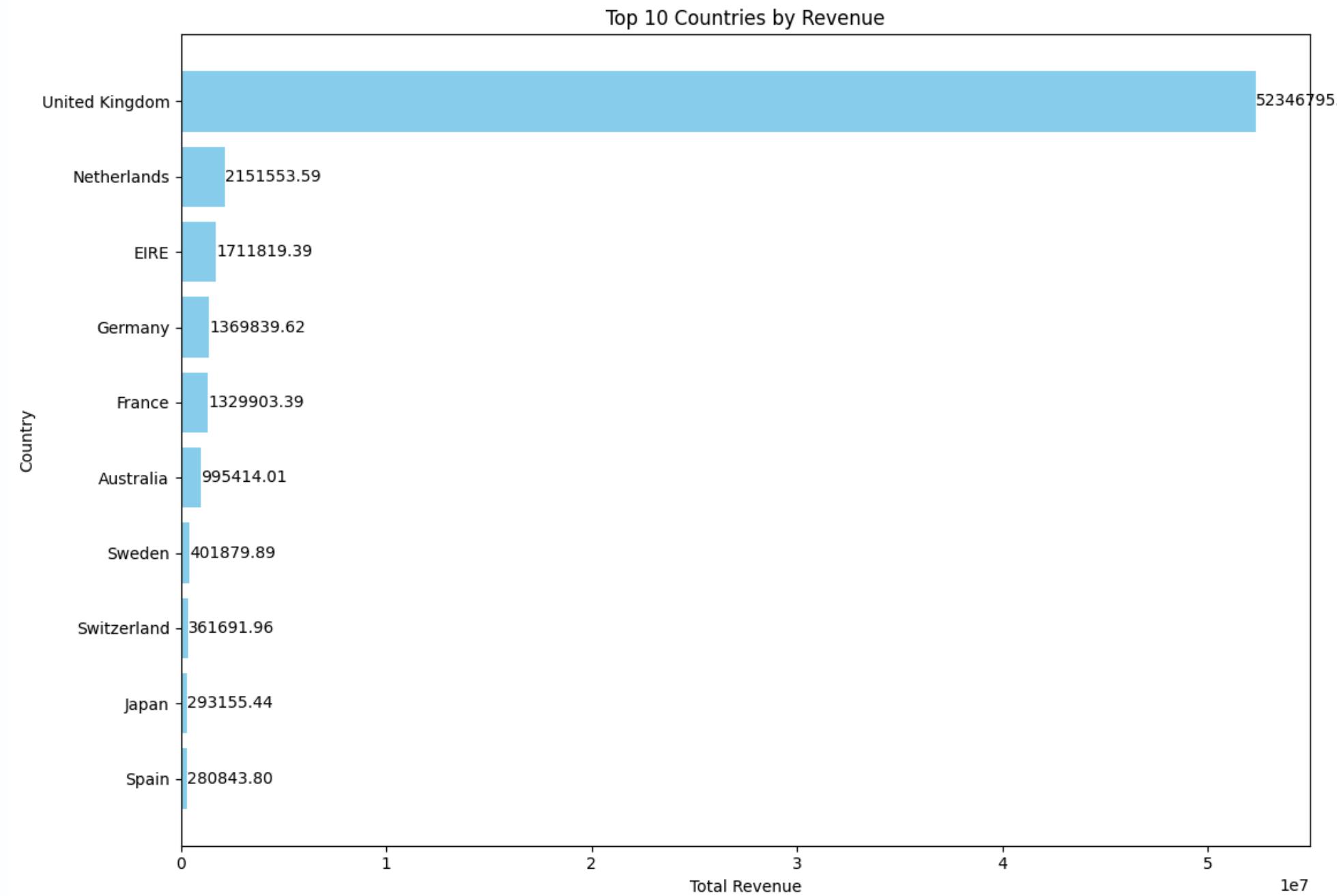
Paper Craft Little Birdie menjadi produk dengan penjualan tertinggi dengan total 80,995 terjual. Produk yang sering dibeli mencakup berbagai kategori seperti kerajinan kertas, penyimpanan keramik, dekorasi rumah, dan barang-barang untuk anak seperti mainan.

APA SAJA TOP 10 PRODUK YANG MEMBERIKAN REVENUE PALING TINGGI?



Top 4 Produk yang memberikan revenue tertinggi merupakan top 4 barang yang paling sering di beli pelanggan.

APA SAJA TOP 10 NEGARA MEMBERIKAN REVENUE PALING TINGGI?



- Inggris menjadi negara dengan pendapatan tertinggi dengan total pendapatan sebesar \$52,346,795.60, make sense karena store based nya sendiri di Inggris

Rekomendasi EDA

- Lakukan promosi pasca liburan untuk mengurangi penurunan penjualan, seperti diskon akhir tahun atau clearance sale.
- Mengadakan promosi atau diskon khusus pada hari Rabu dapat membantu meningkatkan penjualan pada hari dengan aktivitas belanja yang lebih rendah.
- Fokuskan kampanye pemasaran dan promosi pada produk terlaris ini untuk meningkatkan penjualan lebih lanjut. Misalnya, bisa diberikan diskon khusus dan juga memastikan produk selalu tersedia stoknya
- Untuk negara diluar dari UK seperti Belanda, Irlandia, dan Jerman menunjukkan pendapatan yang cukup besar. Memperluas operasi dan kampanye pemasaran di negara-negara ini dapat meningkatkan pendapatan tambahan.



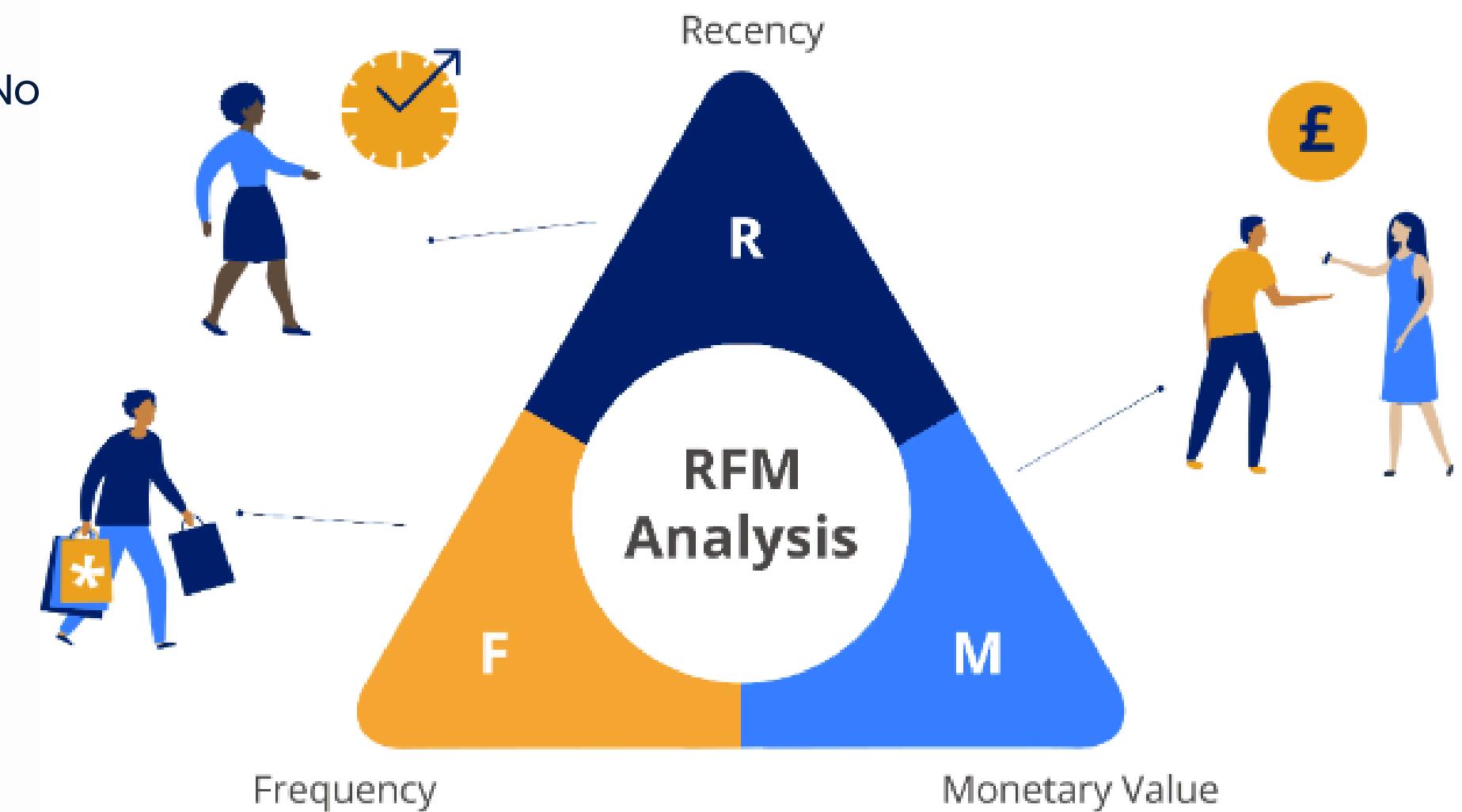
PREPROCESSING



RFM

Dilakukan preprocessing untuk memfasilitasi analisis RFM, yang menggunakan tiga metrik utama: Recency, Frequency , dan Monetary value . Dataset saat ini adalah data long, di mana satu pelanggan (terutama CustomerNo) dapat muncul lebih dari sekali di tiap baris.

Saya akan mengelompokkan data berdasarkan CustomerNo masing-masing pelanggan sebelumnya.



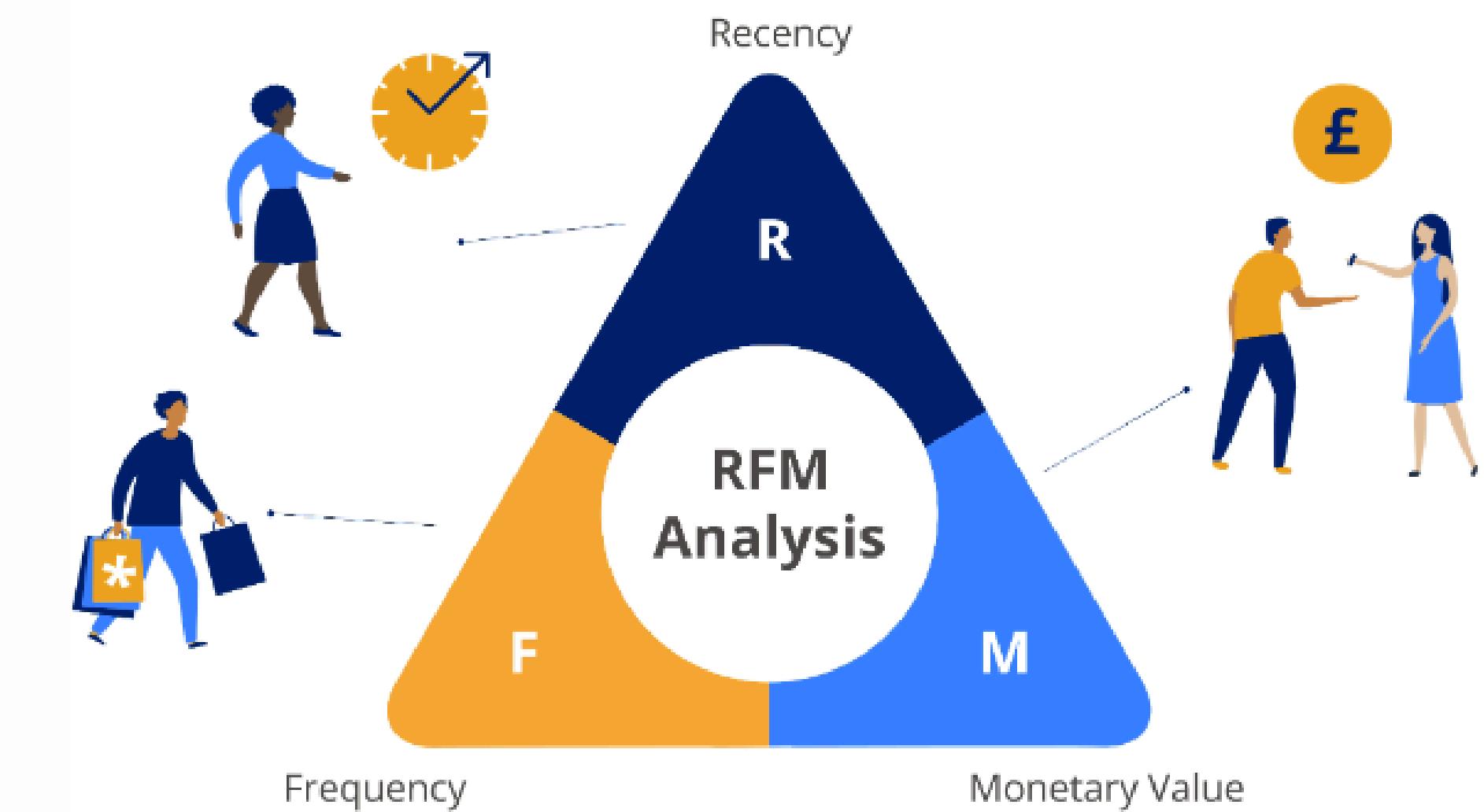
Recency

Waktu terkini pesanan baru pelanggan

recent_date: 2019-12-10 00:00:00

	CustomerNo	LastPurchaseDate	Recency
0	12004	2019-04-26	228
1	12006	2019-05-05	219
2	12008	2019-03-08	277
3	12013	2018-12-15	360
4	12024	2019-06-16	177
...
4713	18280	2019-03-07	278
4714	18281	2019-06-12	181
4715	18282	2019-12-02	8
4716	18283	2019-12-06	4
4717	18287	2019-10-28	43

4718 rows × 3 columns

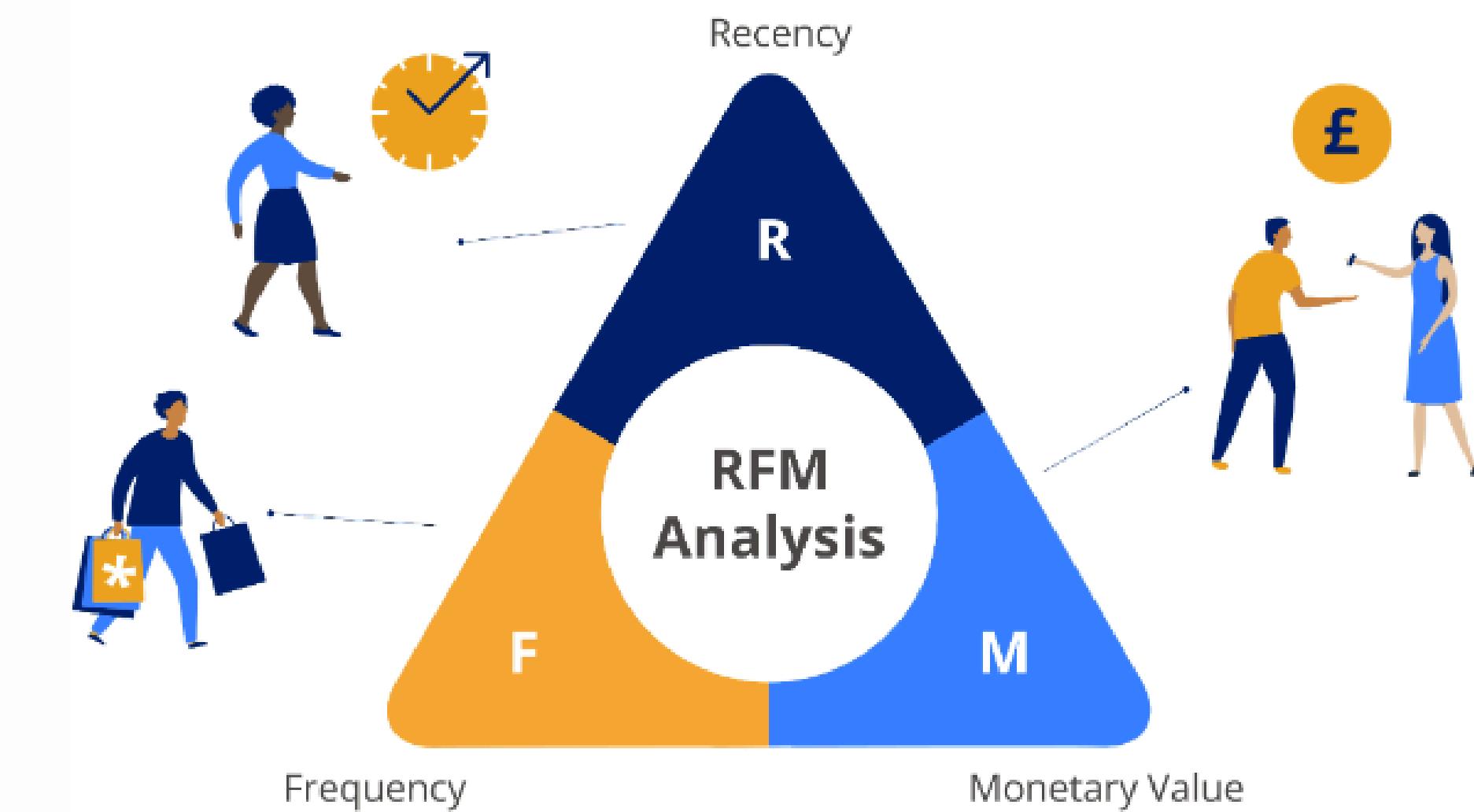


Frequency

Seberapa sering pelanggan kembali untuk membeli

CustomerNo	Frequency
408	12748
2085	14911
4384	17841
667	13089
2406	15311
...	...
2339	15226
2345	15234
2347	15236
2357	15247
2359	15250

4718 rows × 2 columns

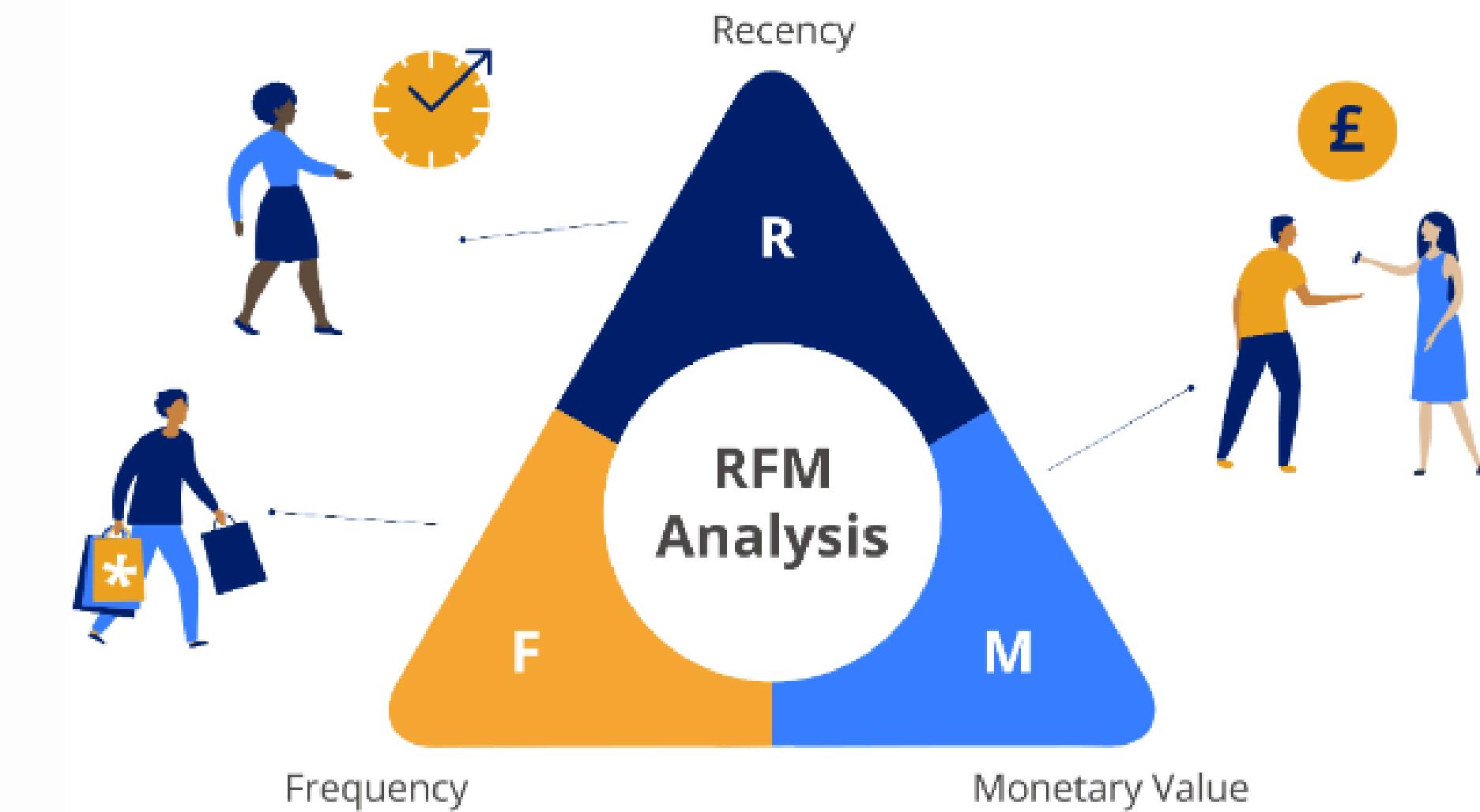


Monetary

Total uang yang dibelanjakan oleh pelanggan

CustomerNo	Monetary
1880	14646 2112282.03
3302	16446 1002741.57
2085	14911 914204.19
126	12415 900545.54
4581	18102 897137.36
...	...
61	12309 12.86
1716	14435 12.38
453	12810 11.98
1208	13775 11.53
3693	16937 5.97

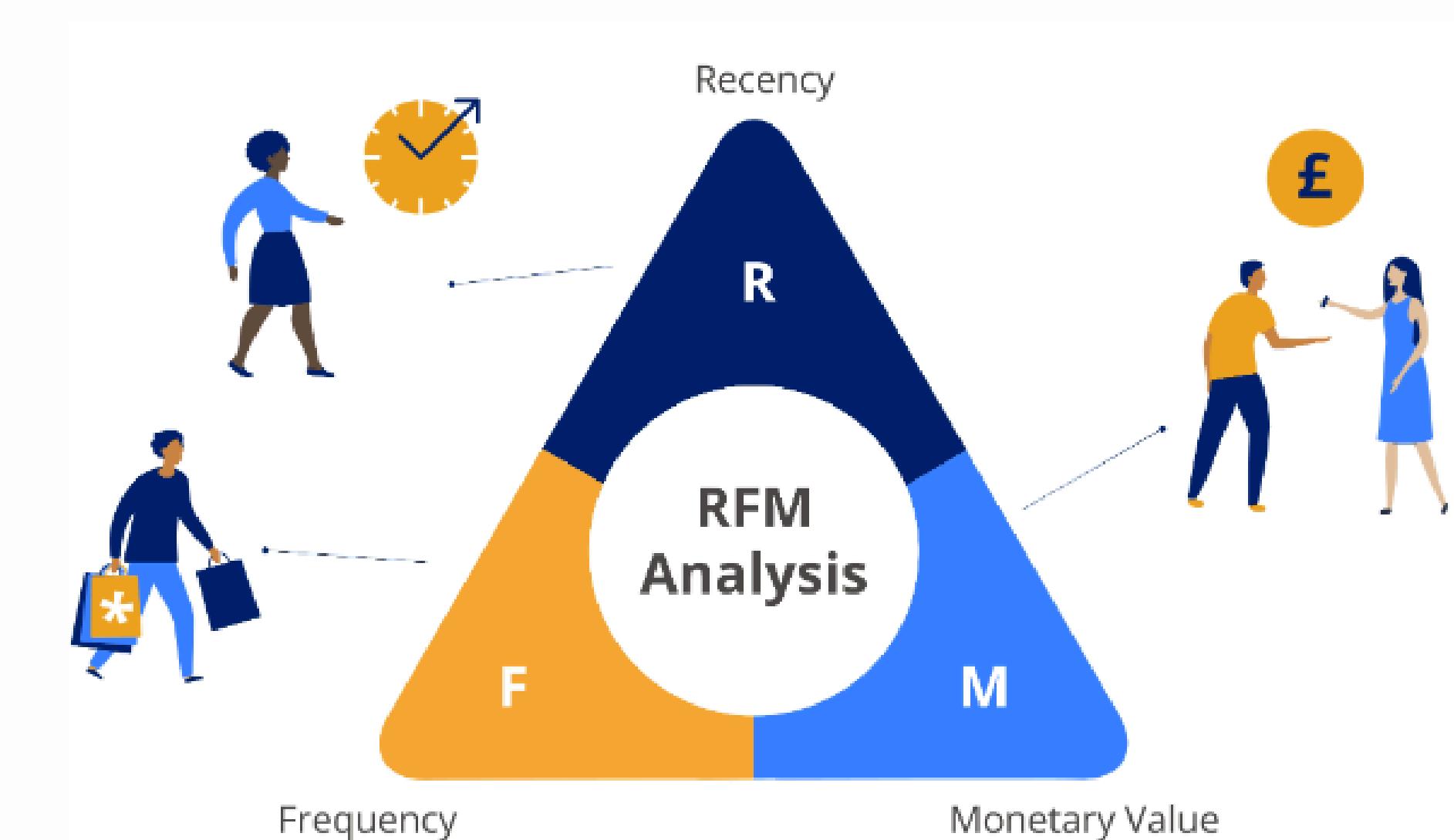
4718 rows × 2 columns



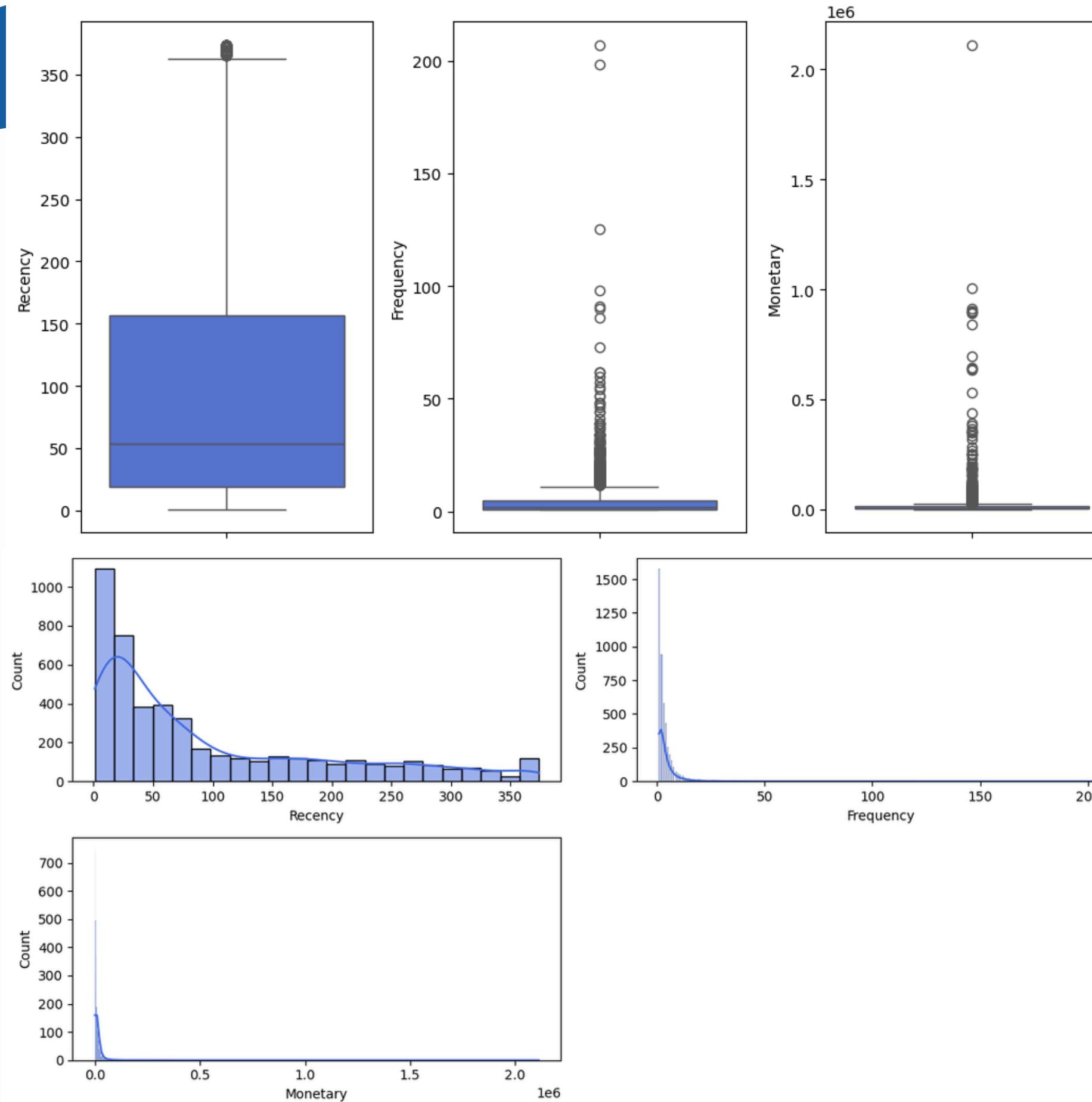
Menggabungkan RFM Menjadi Data Frame

	CustomerNo	Recency	Frequency	Monetary
0	12004	228	1	1509.60
1	12006	219	1	24.76
2	12008	277	1	5689.57
3	12013	360	1	69.96
4	12024	177	1	149.52
...
4713	18280	278	1	623.26
4714	18281	181	1	576.58
4715	18282	8	2	1044.86
4716	18283	4	16	11773.90
4717	18287	43	3	18139.56

4718 rows × 4 columns



Brief Singkat Distribusi RFM



- rata-rata pelanggan melakukan pembelian terakhir mereka dalam 50 hari terakhir, sementara beberapa pelanggan sudah lama tidak melakukan pembelian.
- beberapa pelanggan melakukan pembelian dalam jumlah besar
- Sebagian besar pendapatan berasal dari sejumlah kecil pelanggan dengan nilai pembelian yang tinggi.

SCALING

Dilakukan untuk mencegah variabel dengan rentang yang terlalu luas mempengaruhi perhitungan jarak. Misalnya pada Monetary Scaling dilakukan dengan menggunakan standard scaler

	Recency	Frequency	Monetary
0	1.287572	-0.431944	-0.217001
1	1.199118	-0.431944	-0.244313
2	1.769156	-0.431944	-0.140113
3	2.584901	-0.431944	-0.243482
4	0.786331	-0.431944	-0.242018

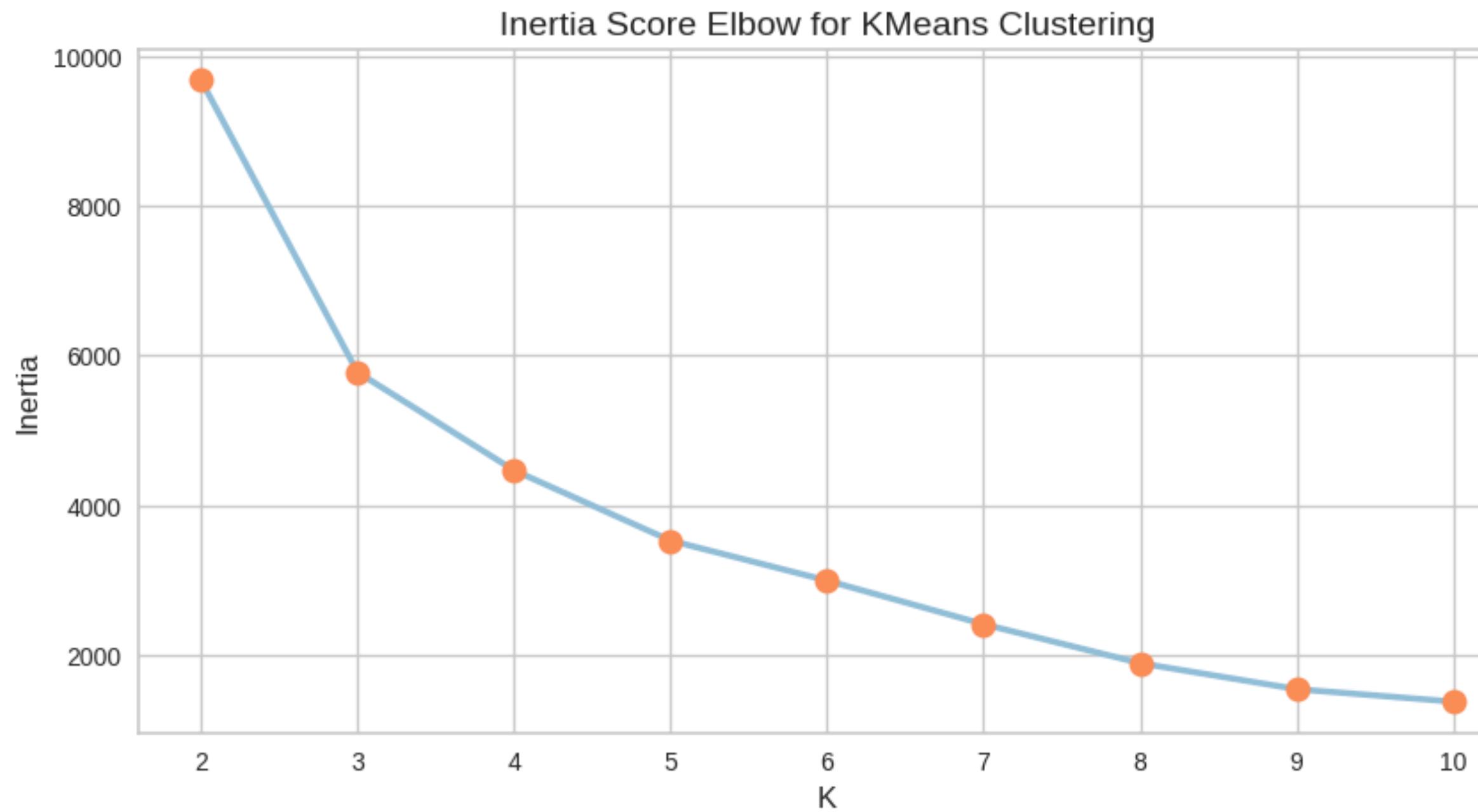


MODELLING



MENCARI NILAI K OPTIMAL

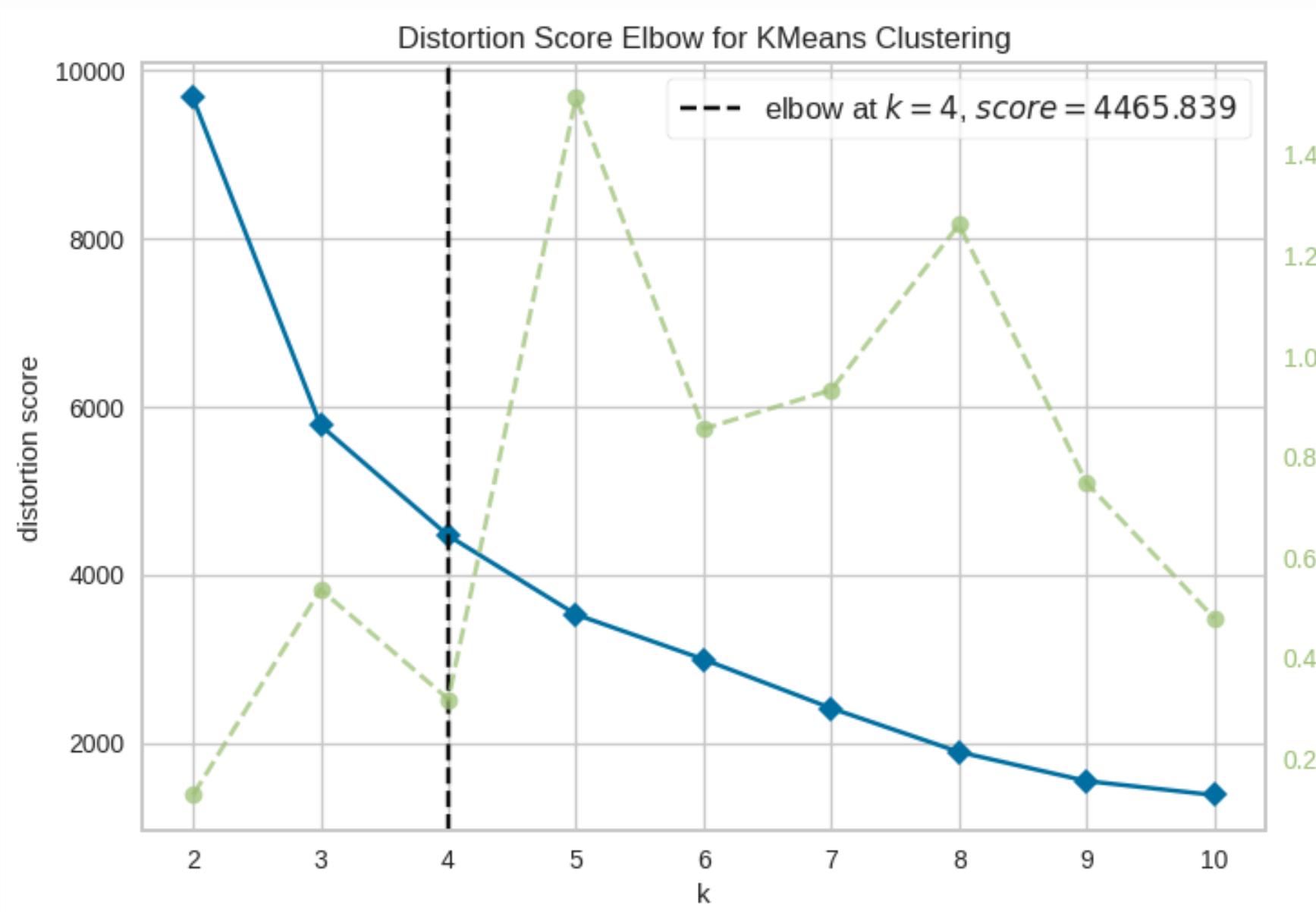
ELBOW METHOD



Berdasarkan grafik yang dihasilkan tidak terlalu terlihat penurunan nilai dan patahan yang tajam, sehingga nilai k yang optimal tidak terlalu diketahui dengan pasti

MENCARI NILAI K OPTIMAL

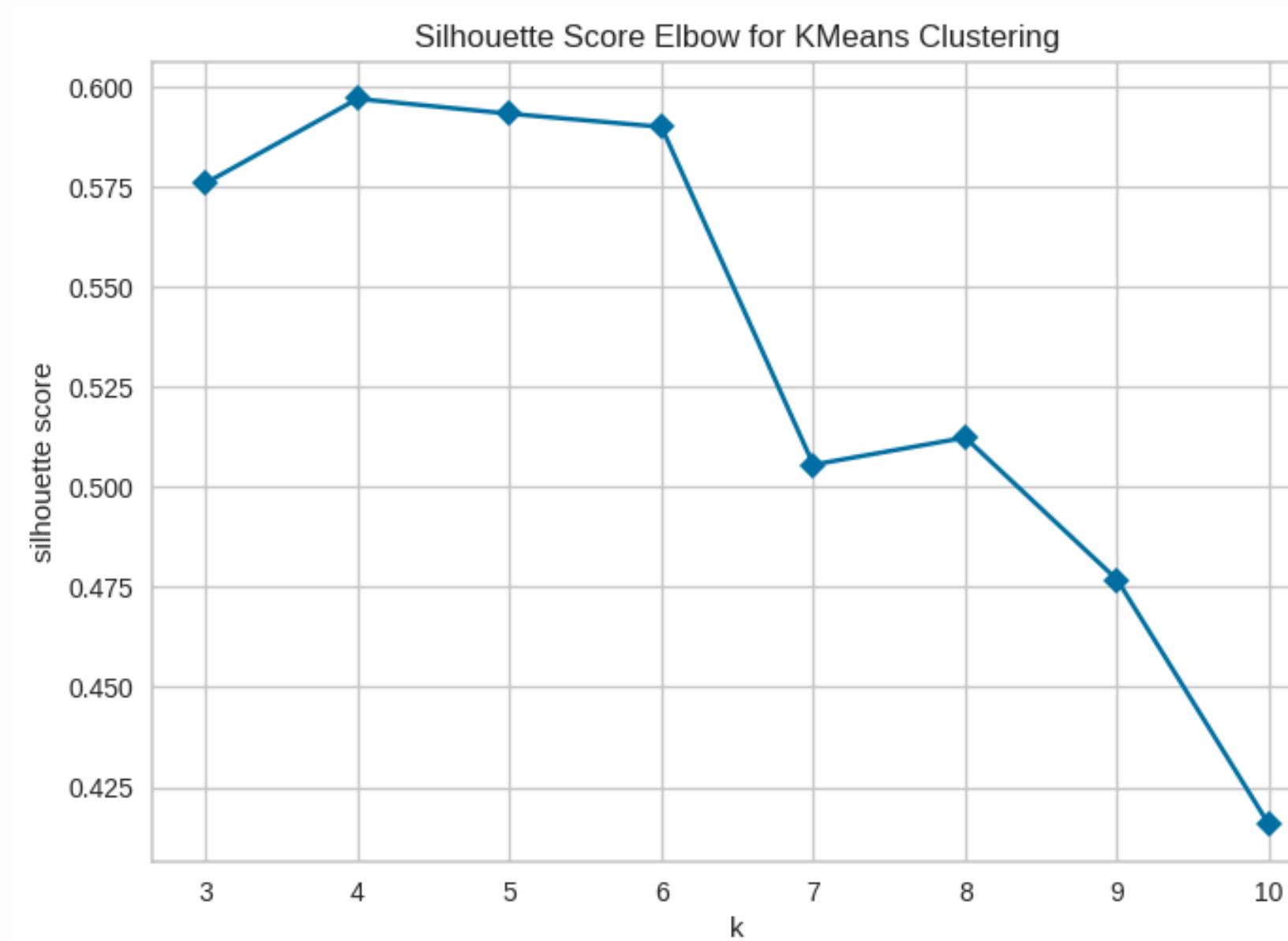
ELBOW METHOD



Berdasarkan grafik diatas, dengan parameter distortion nilai optimal k = 4

MENCARI NILAI K OPTIMAL

Silhouette Score



Berdasarkan grafik silhouette nilai optimal k = 4

K-MEAN CLUSTERING

Setelah menemukan kluster optimal nya yaitu 4 maka dilakukan modelling untuk menentukan Cluster.

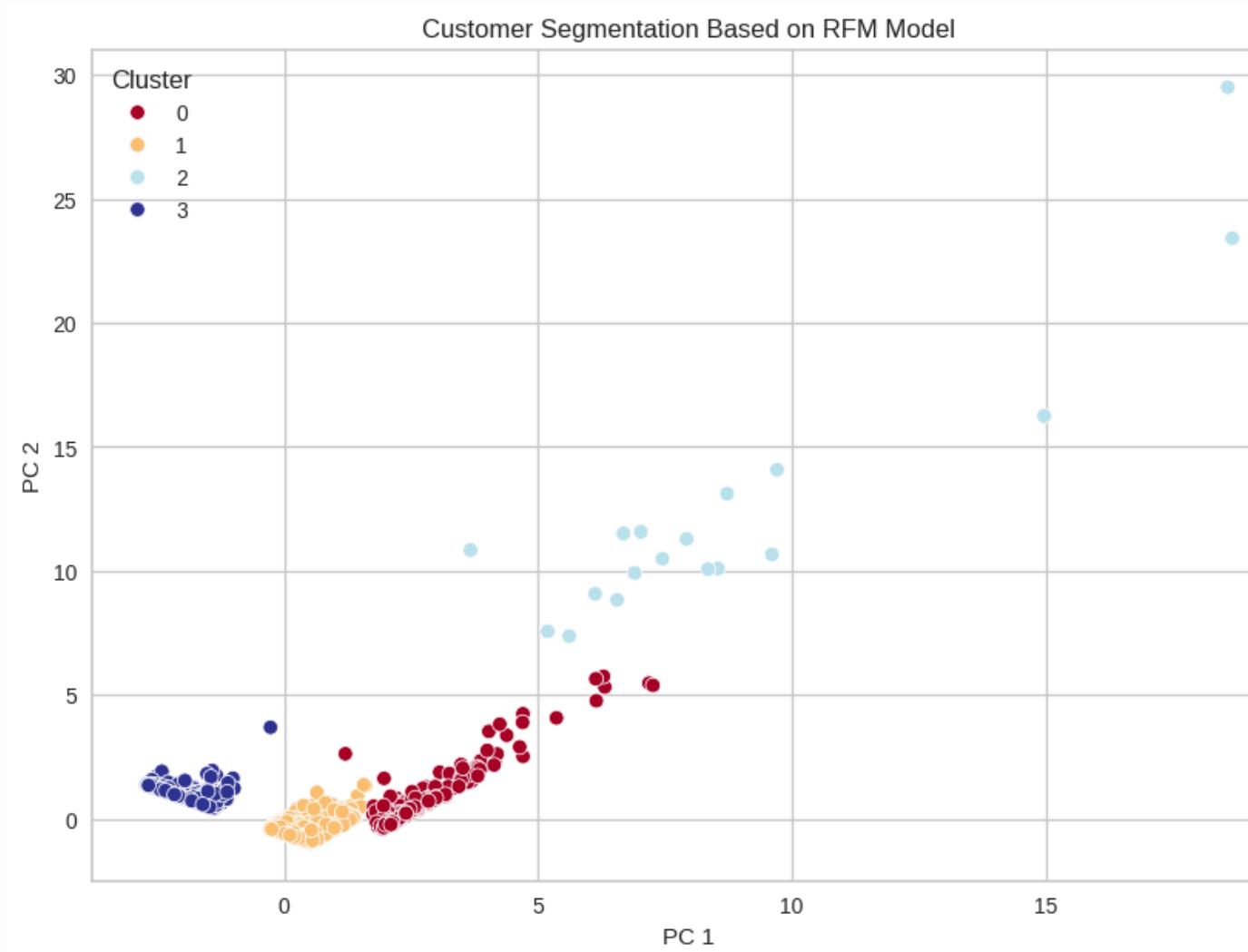
```
▶ from sklearn.cluster import KMeans  
  
# fit model  
kmeans = KMeans(n_clusters=4, random_state = 123)  
kmeans.fit(rfm_features_std.values)  
rfm_features_std['Cluster'] = kmeans.labels_
```

```
▶ rfm_features_std.head()
```

	Recency	Frequency	Monetary	Cluster	grid	info
0	1.287572	-0.431944	-0.217001	3		
1	1.199118	-0.431944	-0.244313	3		
2	1.769156	-0.431944	-0.140113	3		
3	2.584901	-0.431944	-0.243482	3		
4	0.786331	-0.431944	-0.242018	3		

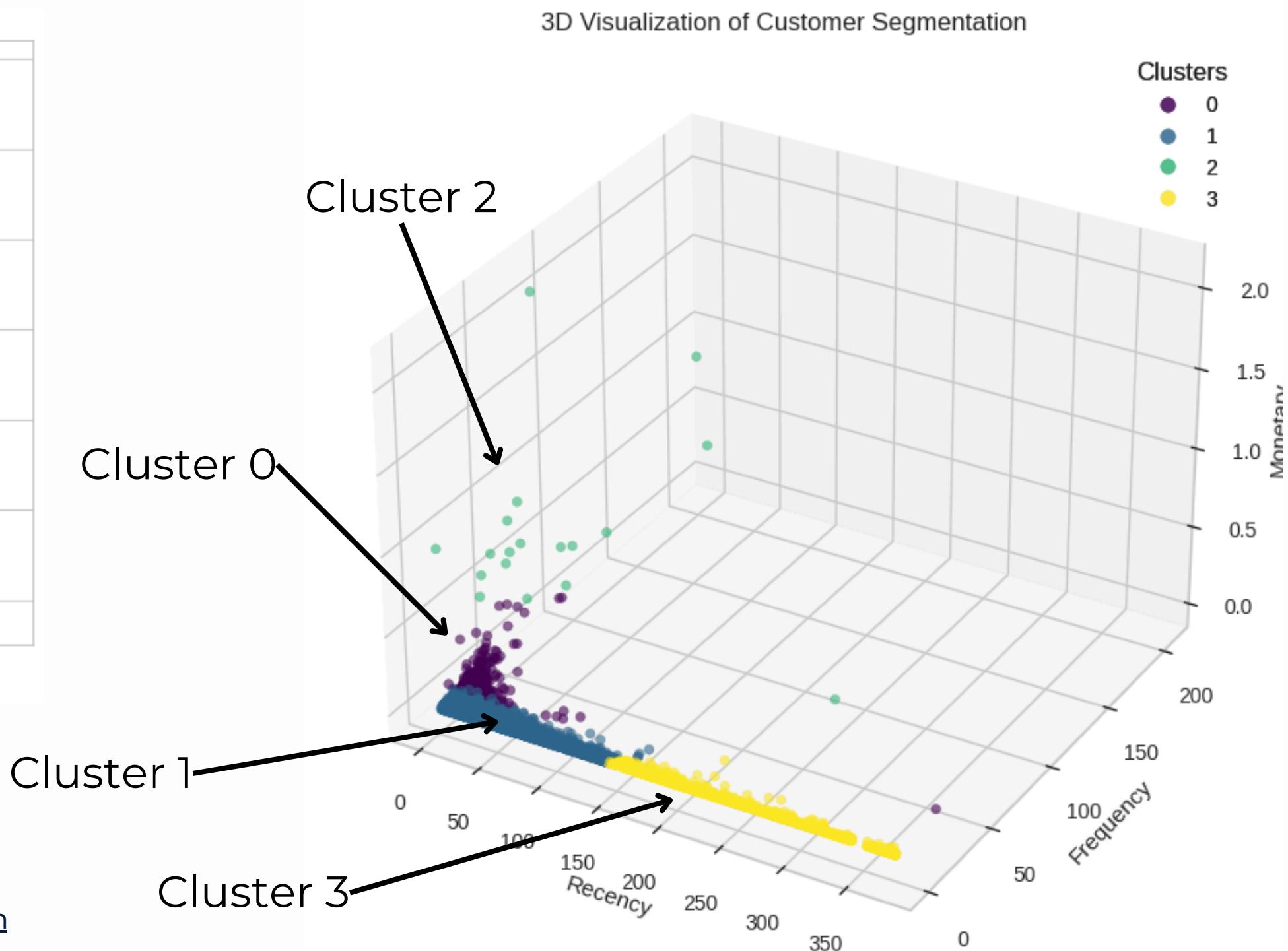
VISUALISASI CUSTOMER SEGMENTATION BERDASARKAN RFM

Setelah fitting model melakukan visualisasi untuk melakukan evaluasi dari segmentasi di tiap cluster



Visualisasi dengan PCA --> 2D

For Better Visualization:
https://finalprojectcustsegment.streamlit.app/Customer_Segmentation



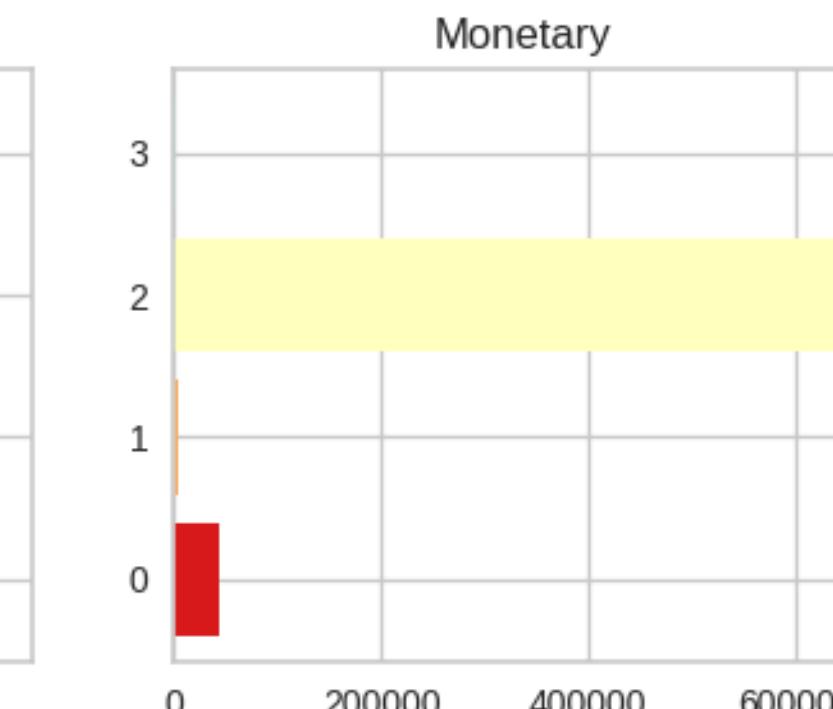
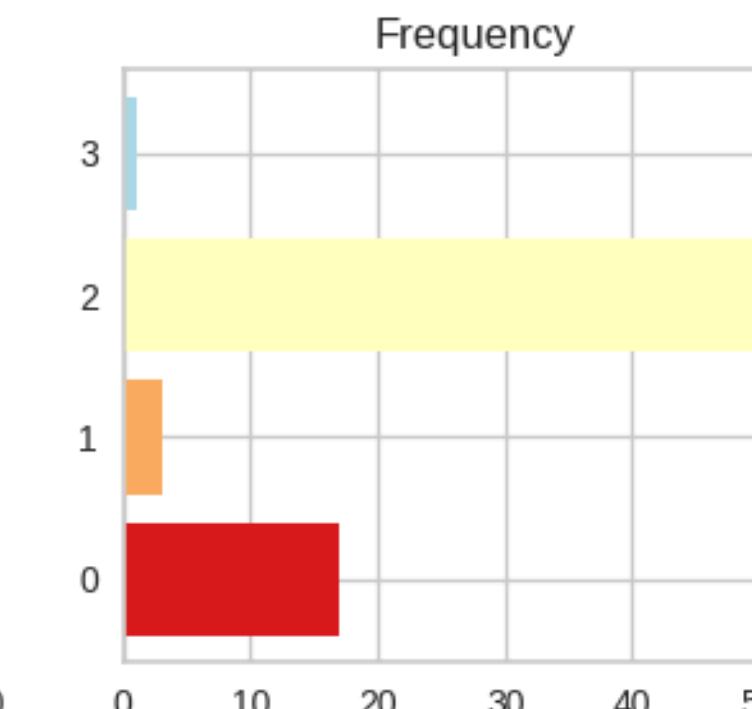
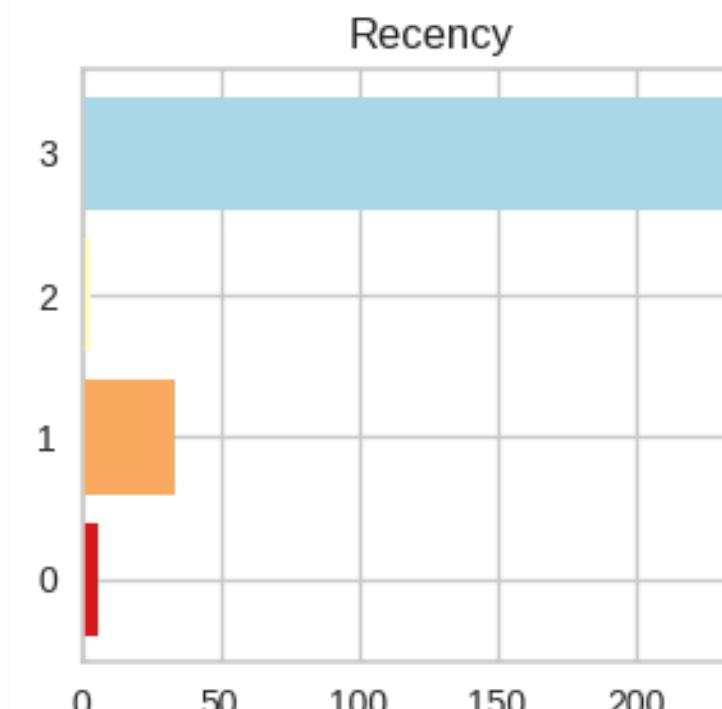
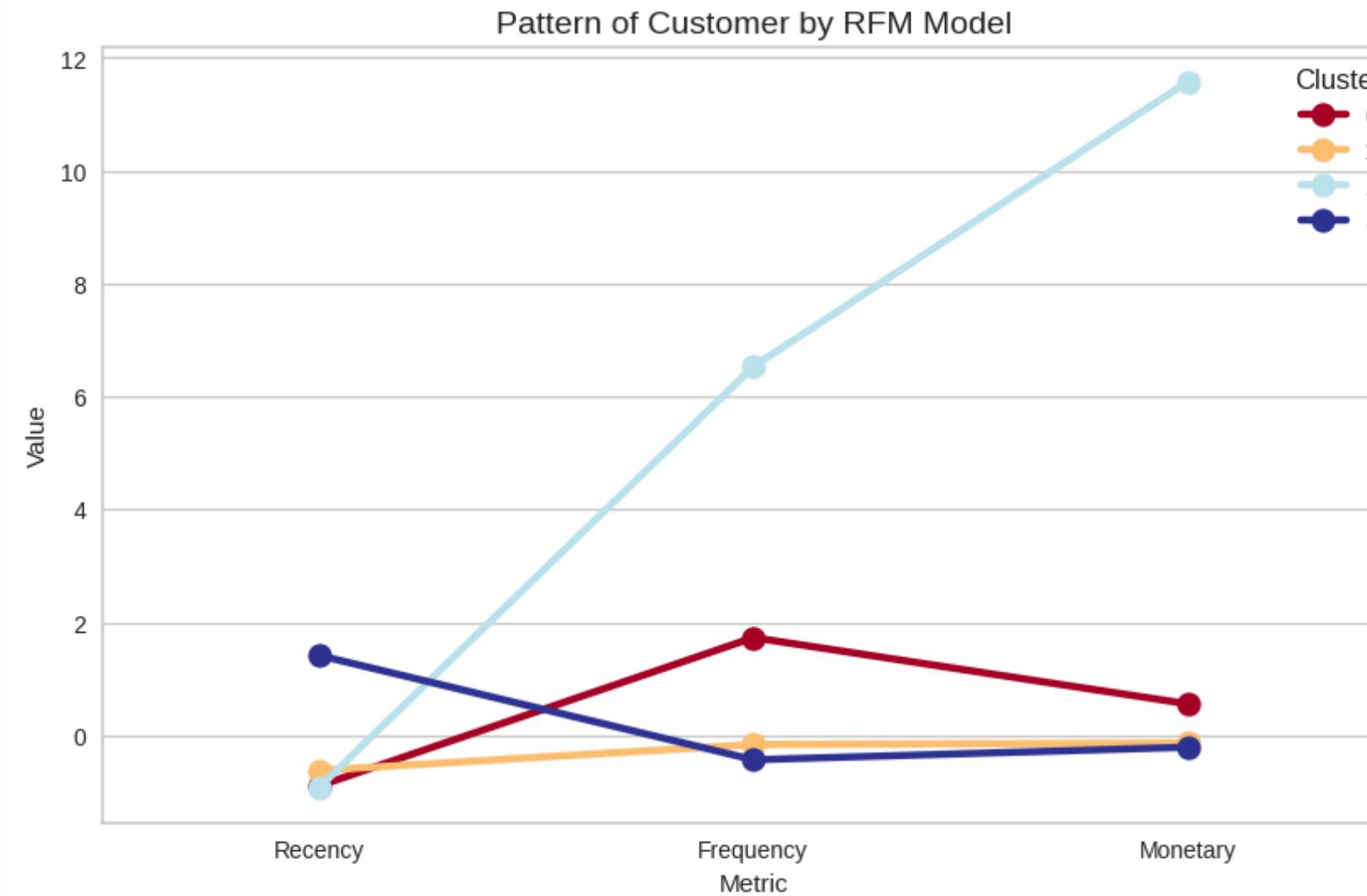
PERSENTASE CUSTOMER DI TIAP CLUSTER



Cluster	count	percentage (%)
0	0	257
1	1	3187
2	2	18
3	3	1256

Custer 1 memiliki persentase jumlah customer terbanyak
Cluster 2 memiliki persentase jumlah customer yang kecil

ANALISA SEGMENTASI



ANALISA SEGMENTASI

Cluster 0 : Frequent High Spender

- Pelanggan dalam cluster ini lumayan sering melakukan pembelian dan melakuakn pembelian dalam jumlah relatif besar
- Mereka baru saja melakukan pembelian (nilai Recency rendah), yang menunjukkan bahwa mereka sangat aktif.

Cluster 1: Regular Shopper

- Pelanggan dalam cluster ini memiliki nilai transaksi yang moderat dan melakukan pembelian dengan frekuensi yang juga moderat.
- Mereka memiliki recency yang lebih tinggi dibandingkan beberapa cluster lain, menunjukkan bahwa mereka mungkin belum melakukan pembelian baru-baru ini, tetapi masih cukup aktif.

Cluster 2 : Loyal High Spender

- Pelanggan dalam cluster ini adalah yang paling berharga dengan nilai transaksi yang sangat tinggi dan frekuensi pembelian yang sangat sering.
- Mereka baru saja melakukan pembelian, menunjukkan tingkat aktivitas yang sangat tinggi.

Cluster 3 : Inactive Low Spender

- Pelanggan dalam cluster ini memiliki nilai transaksi yang rendah dan frekuensi pembelian yang sangat rendah.
- Mereka tidak melakukan pembelian baru-baru ini, menunjukkan bahwa mereka mungkin sudah tidak aktif.

REKOMENDASI

Cluster 0 : Frequent High Spender

- Memberikan program loyalitas seperti poin reward tambahan atau gratis ongkos kirim
- Promosikan produk terkait atau produk lebih premium yang mungkin menarik bagi mereka dengan rekomendasi khusus di halaman produk
- Kirimkan penawaran yang dipersonalisasi berdasarkan riwayat pembelian mereka melalui email

Cluster 1 : Regular Shopper

- Mengirimkan penawaran khusus atau diskon pada momen spesial atau hari libur untuk mendorong mereka berbelanja lebih sering.
- Berikan diskon khusus untuk pembelian berikutnya yang berlaku untuk jangka waktu terbatas.
- Menawarkan insentif bagi pelanggan untuk merujuk teman atau keluarga, seperti diskon untuk setiap referral yang berhasil.

REKOMENDASI

Cluster 2 : Loyal High Spender

- Membuat program VIP yang memberikan manfaat eksklusif seperti hadiah khusus, prioritas layanan pelanggan yang eksklusif.
- Berikan penghargaan tambahan atas kesetiaan mereka, seperti poin reward ekstra atau voucher belanja..
- Kirimkan ucapan terima kasih pribadi atau hadiah kecil untuk memperkuat hubungan dan meningkatkan loyalitas.

Cluster 3 : Inactive Low Spender

- Melakukan kampanye reaktivasi dengan penawaran yang menarik, seperti diskon besar atau gratis ongkos kirim, untuk mendorong mereka kembali berbelanja.
- Minta umpan balik untuk memahami mengapa mereka tidak aktif dan apa yang bisa diperbaiki untuk menarik minat mereka kembali.
- Mengirimkan pesan yang dipersonalisasi untuk mengingatkan mereka tentang produk yang pernah mereka beli atau yang mungkin mereka minati



THANK YOU

*We look forward to working
with you!*

CONTACT ME ON



[LinkedIn](#)



[Github](#)



m.alwannur@gmail.com

