ONLINE RETAIL CUSTOMER SEGMENTATION AND PRODUCT RECOMMENDATION

Yodi Ramadhani Alfariz

github.com/yodialfa

Bussines Problem

- Segmentasi customer belum diterapkan agar strategi marketing tepat sasaran
- Segmentasi adalah langkah awal untuk Customer Relationship Management
- Banyaknya product memungkinkan customer sulit untuk mencari product yang mungkin sangat customer cari.

Objectives

Clustering

Mengcluster Customer Berdasarkan RFM Analisis sehingga kelas tiap customer terfokus ke feature tertentu

Recommendations

Product recommendation untuk memudahkan customers mencari product yang mungkin sangat mereka cari.

Customer Segmentation

Data Understanding

Dataset consist of : 541.909 rows and 8 columns

No.	Columns	Details
1	InvoiceNo	transaction number code
2	StockCode	item code
3	Description	item description
4	Quantity	number of items purchased
5	InvoiceDate	purchase date
6	UnitPrice	item price
7	CustomerID	customers code
8	Country	country of customers

Data Cleansing

```
InvoiceNo            0
StockCode            0
Description       1454
Quantity             0
InvoiceDate          0
UnitPrice            0
CustomerID      135080
Country              0
dtype: int64
```

Filling Description with same stockcode

For CustomerID we have to focus fro customer segmentation, so we drop it

RFM Analysis

Recency

Variabel untuk mengukur nilai pelanggan berdasarkan rentang waktu (tanggal, bulan, tahun) transaksi terakhir pelanggan sampai saat ini

Frequency

Mengacu pada berapa kali pelanggan berinteraksi dengan produk atau melakukan transaksi dalam periode waktu tertentu.

Monetary

Menunjukkan jumlah total yang dihabiskan oleh pelanggan untuk membeli produk Anda dalam periode waktu tertentu.

Data Preprocessing with RFM

Recency

	CustomerID	LastPurchaseDate
0	12346.0	2011-01-18
1	12347.0	2011-12-07
2	12348.0	2011-09-25
3	12349.0	2011-11-21
4	12350.0	2011-02-02

	CustomerID	recency
0	12346.0	325
1	12347.0	2
2	12348.0	75
3	12349.0	18
4	12350.0	310

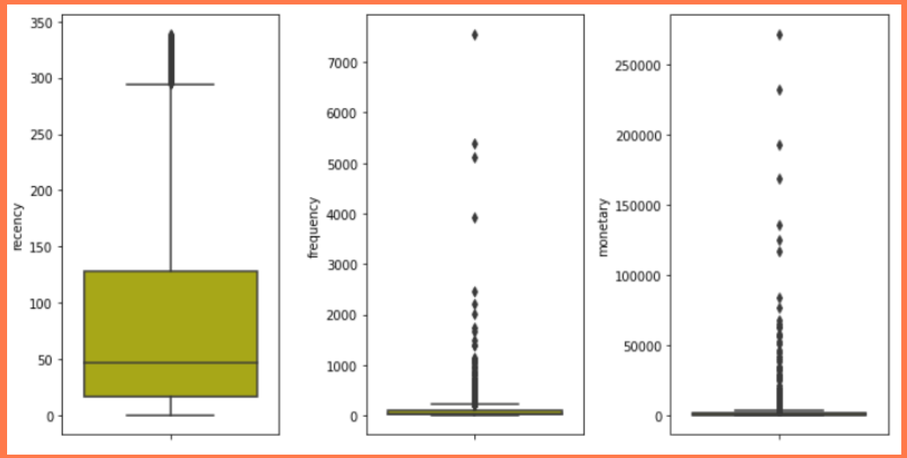
Frequency

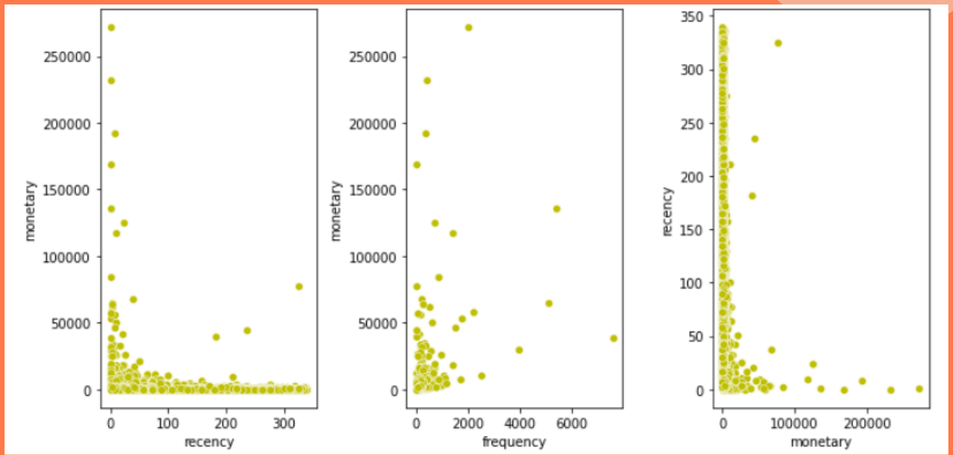
	CustomerID	frequency
0	12346.0	1
1	12347.0	151
2	12348.0	14
3	12349.0	73
4	12350.0	17

Monetary

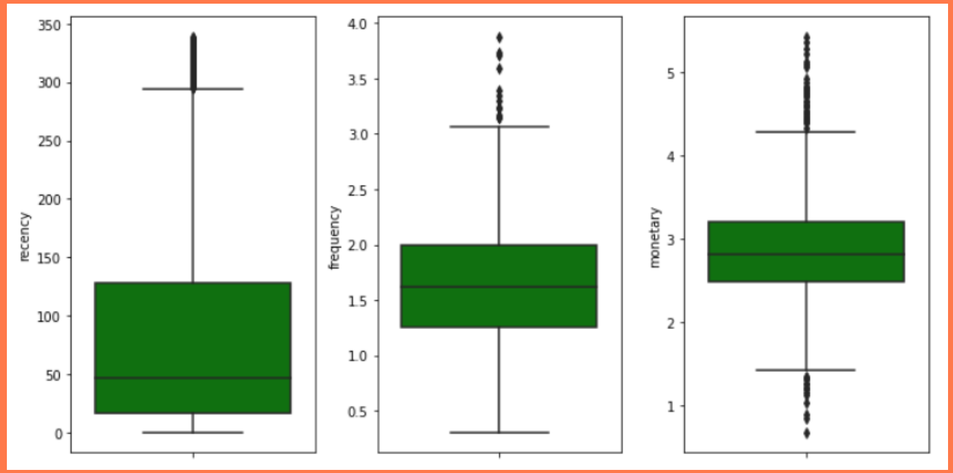
	CustomerID	monetary
0	12346.0	77183.60
1	12347.0	3598.21
2	12348.0	904.44
3	12349.0	1757.55
4	12350.0	334.40

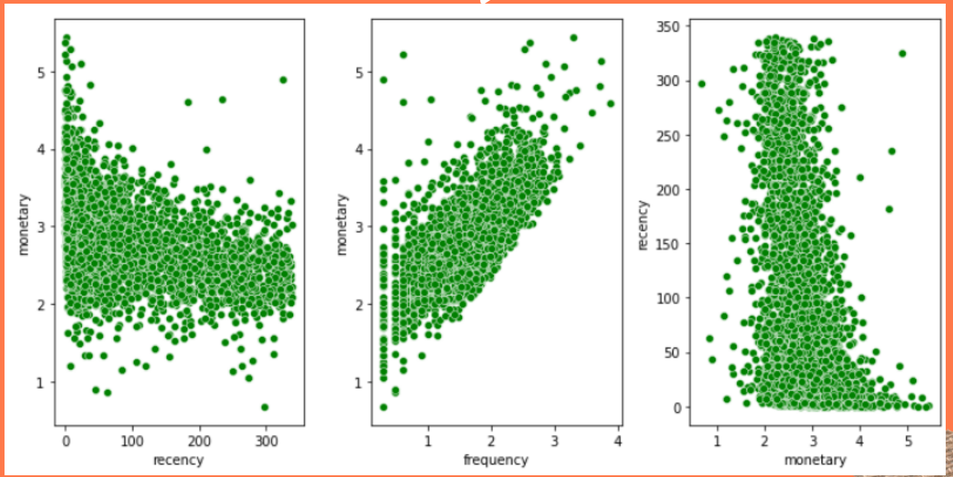
Data Distribution





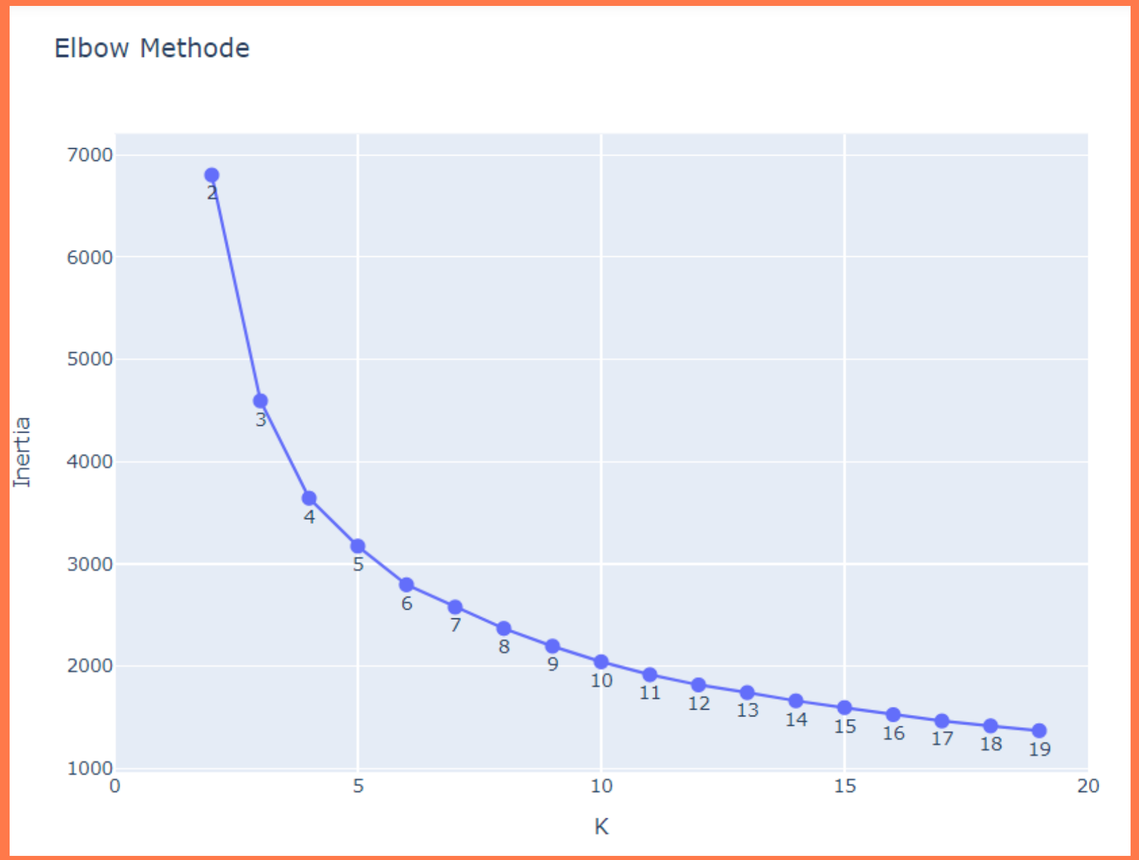
Transform data to reduce outlier



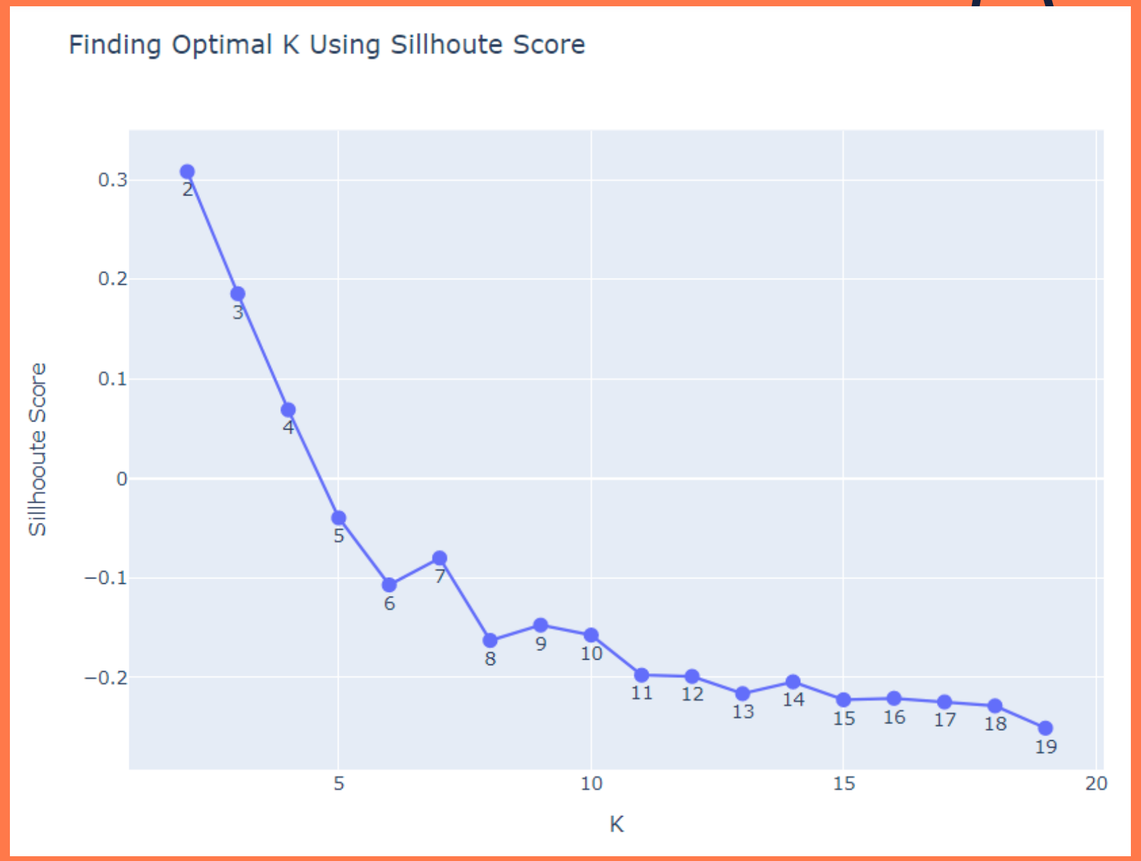


Clustering With K-Means

Mencari Potensial K Dengan Elbow Method Pada RFM DataFrame



Mencari Potensial K Degan Sillhoute Score Menggunakan Euclidean



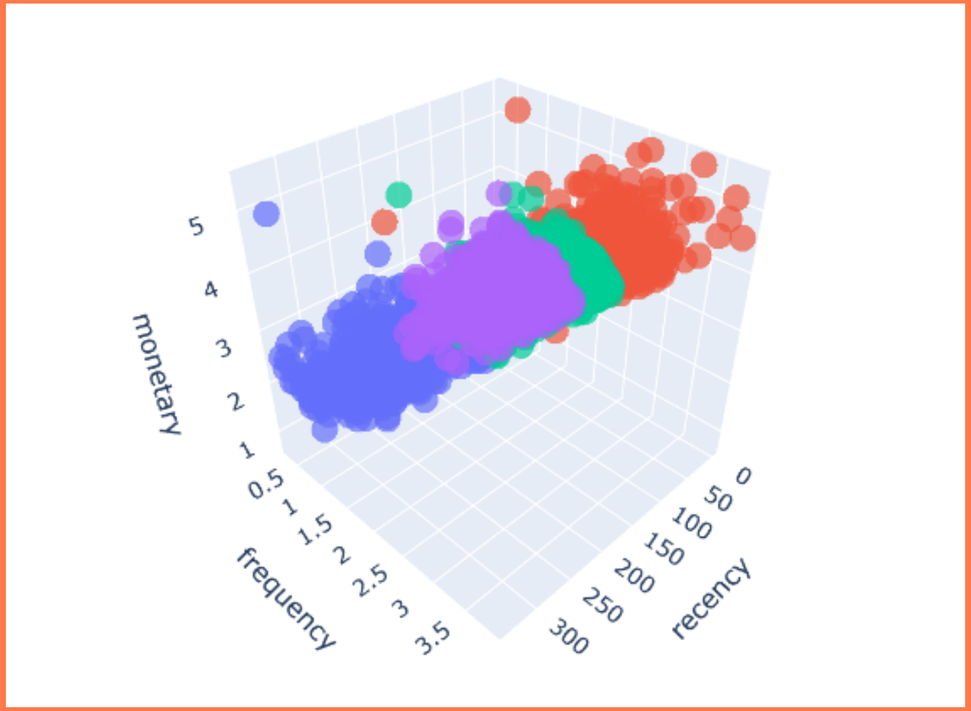
K = 4 diambil dikarenakan pada elbow method menjukan nilai yang tidak begitu signifikan dengan nilia k setelahnya, begitupun dengan sillhoute method, pada k yang tertinggi adalah 2,3,4 dan 5 , dikarenakan pada elbow method k potensial adalah 4 maka kita akan ambil 4 untuk nilai k.

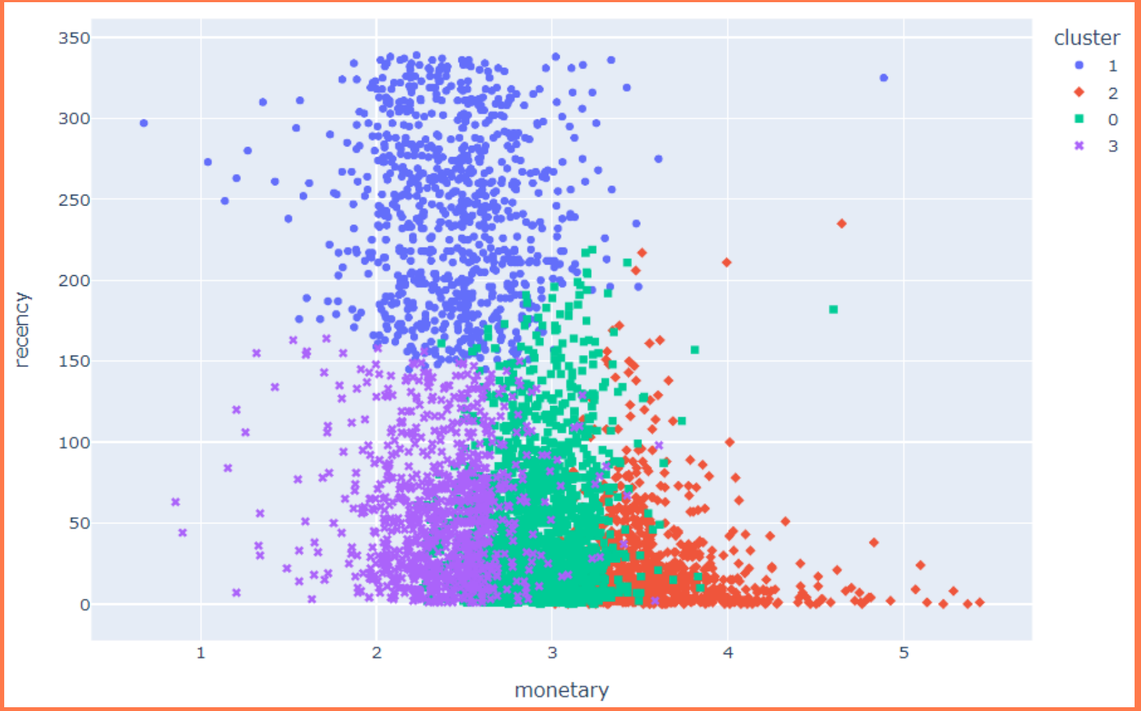
Clustering With K-Means

Sebaran Data Dengan Scatterplot
Setelah Pengklusteran

Clustering With K-Means

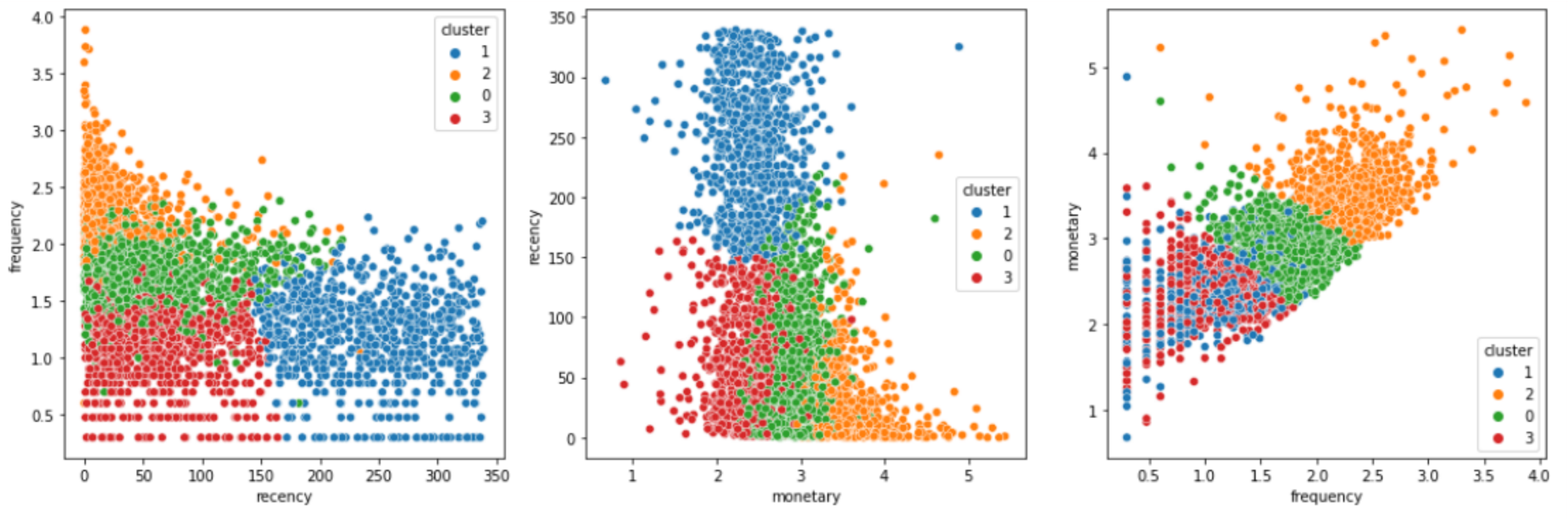
Sebaran Data dengan Scatterplot
Setelah Pengclusteran





Clustering With K-Means

Scatterplot
After Clustering



Clustering With K-Means

Mapping Class From Clusters

Middle Spending
Active Customers

0

New Customers

3

Low Spending
Old Customers

1

Best Customers

2

Customer Segemntation

Recommendation

Best Customers

- Penawaran Product Baru atau Product yang mungkin belum dibeli sebelumnya
- Memberikan reward untuk customer relationship management

Low Spending Pasif Customers

- Melakukan follow up dengan marketing campaign terkait bundling product dan memberikan rekomendasi product yang paling diminati.
- Memberikan penawaran potongan harga agar customer spent lebih banyak.

Middle Spending Active Customers

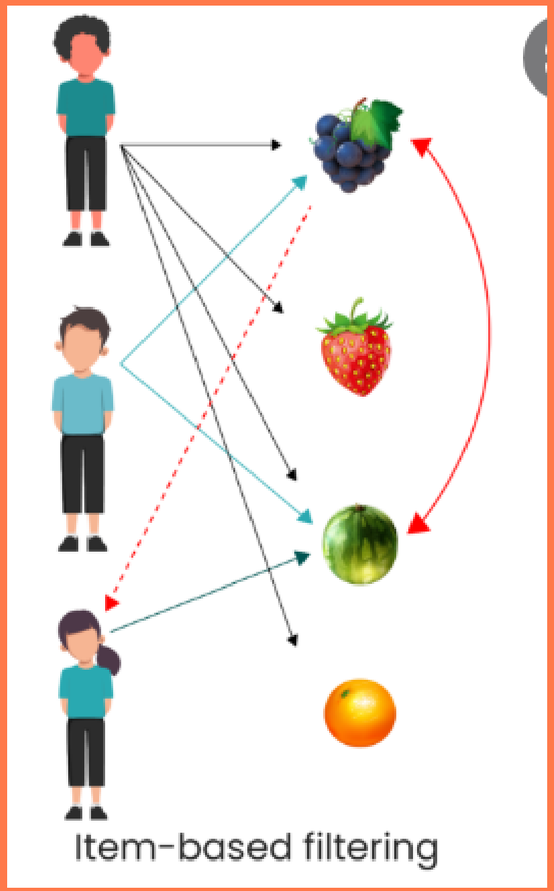
- Memberikan penawaran potongan harga agar menarik customer untuk bertransaksi lagi.
- Memberikan poin sebagai reward dan dapat ditukarkan dengan product atau potongan harga

New Customers

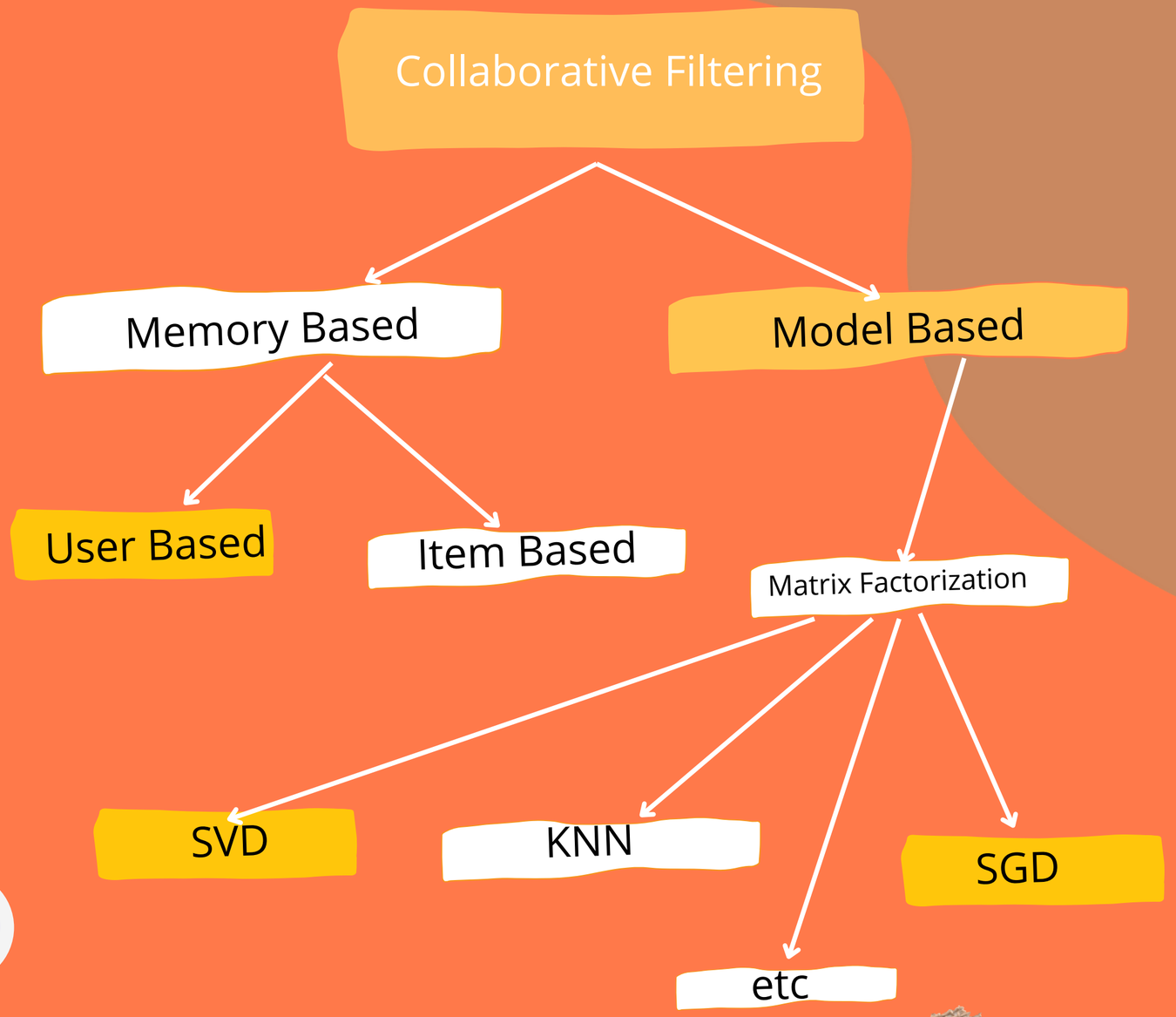
- Memberikan Rekomendasi Product berdasarkan product yang diminati customer lainnya.
- Memberikan poin agar customer bertransaksi lebih intens.

Note : Product Reccomendation sangat disarankan untuk tahap selanjutnya

Collaborative Filtering



https://www.google.com/url?sa=i&url=https%3A%2F%2Ftezeract.ai%2Frecommendation-systems-its-types%2F&psig=AOvVaw2UFIaty9abhc-pBD3VQy78&ust=1667703118973000&source=images&cd=vfe&ved=0CA0QjRxqFwoTCNCTz7qElvsCFQAAAAAdAAAAAB AJ



Collaborative Filtering

Item-based

- Pivoting
- Correlation
- Sort by Highest Corr
- Recommendation

Collaborative Filtering

Item-based Filtering

Description InvoiceNo	4 PURPLE FLOCK DINNER CANDLES	50'S CHRISTMAS GIFT BAG LARGE	DOLLY GIRL BEAKER	I LOVE LONDON MINI BACKPACK	NINE DRAWER OFFICE TIDY	OVAL WALL MIRROR DIAMANTE	RED SPOT GIFT BAG LARGE	SET 2 TEA TOWELS I LOVE LONDON	SPACEBOY BABY GIFT SET	TRELLIS COAT RACK	...	ZINC HEARTS PLANT POT HOLDER	ZINC HERB GARDEN CONTAINER	ZINC M H DECOR
539993	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
540001	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
540002	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
540003	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
540004	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	

Pivot Table

Collaborative Filtering

Pearson Correleation

```
recommendations = df_items.corrwith(df_items['CERAMIC CAKE DESIGN SPOTTED MUG'])
recommendations.dropna(inplace=True)
recommendations = pd.DataFrame(recommendations, columns=['correlation']).reset_index()
recommendations
```

	Description	correlation
0	4 PURPLE FLOCK DINNER CANDLES	-0.001981
1	50'S CHRISTMAS GIFT BAG LARGE	-0.002191
2	DOLLY GIRL BEAKER	-0.001757
3	I LOVE LONDON MINI BACKPACK	-0.002031
4	NINE DRAWER OFFICE TIDY	-0.002187
...	...	...
3515	ZINC T-LIGHT HOLDER STAR LARGE	-0.004492
3516	ZINC T-LIGHT HOLDER STARS SMALL	-0.004255
3517	ZINC WILLIE WINKIE CANDLE STICK	-0.000099
3518	ZINC WIRE KITCHEN ORGANISER	-0.001471
3519	ZINC WIRE SWEETHEART LETTER TRAY	-0.001374

Collaborative Filtering

Sorting by Highest Correlation

	Description	correlation
600	CERAMIC CAKE DESIGN SPOTTED MUG	1.000000
601	CERAMIC CAKE DESIGN SPOTTED PLATE	0.342997
3073	STRAWBERRY FAIRY CAKE TEAPOT	0.335386
609	CERAMIC STRAWBERRY DESIGN MUG	0.323295
599	CERAMIC CAKE BOWL + HANGING CAKES	0.241313
...	...	...
1778	LUNCH BAG SPACEBOY DESIGN	-0.007337
184	ANTIQUE SILVER TEA GLASS ETCHED	-0.007430
1769	LUNCH BAG BLACK SKULL.	-0.007987
1776	LUNCH BAG RED RETROSPOT	-0.008051
1809	MAILOUT	-0.009004

Recommendation

	Description	correlation
601	CERAMIC CAKE DESIGN SPOTTED PLATE	0.342997
3073	STRAWBERRY FAIRY CAKE TEAPOT	0.335386
609	CERAMIC STRAWBERRY DESIGN MUG	0.323295
599	CERAMIC CAKE BOWL + HANGING CAKES	0.241313
3133	TEA TIME TEA SET IN GIFT BOX	0.221122

Collaborative Filtering

THANK YOU