



MAKE YOUR CUSTOMER SEGMENTATION

RFM ANALYSIS

Sedikit Lebih Dekat



I Made Krisna Jaya

Instagram : nana_krisnaa

LinkedIn : I Made Krisna Jaya

WhatsApp : 082146916936

E-mail : madekrisnaj@gmail.com



| 01


⋮ CASE OVERVIEW

LATAR BELAKANG

Dalam menjalankan bisnis, perusahaan e-commerce membutuhkan pengelolaan yang baik dalam hal penyetoran, marketing, penjualan, karyawan, produksi, inovasi dan lainnya. Adapun kesuksesan e-commerce juga bergantung pada engagement perusahaan terhadap pelanggannya. Dimana semakin tinggi tingkat customer engagement, semakin tinggi kesetiaan dari pelanggan kepada perusahaan. Sehingga semakin stabil kegiatan operasional, bahkan bisa mendapatkan laba lebih banyak.

Upaya perusahaan untuk menarik pelanggan dapat berupa:

- Pemberian bonus kepada pelanggan yang sudah lama berlangganan
- Diskon untuk pelanggan dengan pembelian terbanyak
- Promosi tahunan untuk semua pelanggan,
- dan lainnya.



“Customer engagement adalah interaksi yang terjadi antara perusahaan dengan pelanggannya, ditawarkan oleh perusahaan, dan dipilih pelanggan .”

—Paul Greenber

TUJUAN DAN MANFAAT

Proyek kali ini dilakukan untuk memberi referensi kepada perusahaan e-commerce dalam membuat strategi peningkatan customer engagement berdasarkan hasil penjualan pada periode Desember 2010 sampai dengan Desember 2011.

Adapun pemberian referensi yang diberikan berupa:

- Menganalisa customer churn dan retention rate selama periode Des 2010- Des 2011
- Menganalisa segmentasi customer berdasarkan analisa RFM
- Memberi rekomendasi ke perusahaan utk treatment customer berdasarkan segmentasinya

Model-model yang dibuat nantinya bermanfaat sebagai input/masukkan dan dasar pembuatan perencanaan bagi perusahaan dalam mentreatment pelanggannya.

ATTRIBUTE DATASET

01

InvoiceNo

Nomor faktur transaksi yang tercatat

02

StockCode

Kode identifikasi masing-masing unit produk

03

Description

Deskripsi unit produk

04

Quantity

Jumlah unit produk yang dibeli

05

InvoiceDate

Tanggal terjadinya transaksi

06

UnitPrice

Harga per unit produk

07

CustomerID

Kode identitas customer saat melakukan transaksi

08

Country

Negara tempat transaksi



| 02

⋮ DATA PREPROCESSING

1. DATA TYPE

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   InvoiceNo        541909 non-null object
1   StockCode       541909 non-null object
2   Description      540455 non-null object
3   Quantity        541909 non-null int64
4   InvoiceDate      541909 non-null object
5   UnitPrice       541909 non-null float64
6   CustomerID      406829 non-null float64
7   Country         541909 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 33.1+ MB
```

BEFORE

Masih terdapat tipe data yang kurang sesuai, yaitu attribute InvoiceDate yang bertipe object

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   InvoiceNo        541909 non-null object
1   StockCode       541909 non-null object
2   Description      540455 non-null object
3   Quantity        541909 non-null int64
4   InvoiceDate      541909 non-null datetime64[ns]
5   UnitPrice       541909 non-null float64
6   CustomerID      406829 non-null float64
7   Country         541909 non-null object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 33.1+ MB
```

AFTER

Tipe dari InvoiceDate sudah berubah menjadi datetime

2. MISSING VALUE

```
InvoiceNo      0.00
StockCode      0.00
Description     0.27
Quantity       0.00
InvoiceDate    0.00
UnitPrice      0.00
CustomerID     24.93
Country        0.00
dtype: float64
```

Terdapat dua attribute yang memiliki missing value yaitu Description sebesar 0,27% data kosong dan CustomerID sebesar 24.93%

Attribute CustomerID attribute penting dan attribute Description memiliki persentase missing value yang kecil oleh karena itu akan dilakukan penanganan missing value pada kedua attribute tersebut

2. MISSING VALUE (DESCRIPTION)

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
622	536414	22139	NaN	56	2010-12-01 11:52:00	0.0	NaN	United Kingdom
1970	536545	21134	NaN	1	2010-12-01 14:32:00	0.0	NaN	United Kingdom
1971	536546	22145	NaN	1	2010-12-01 14:33:00	0.0	NaN	United Kingdom
1972	536547	37509	NaN	1	2010-12-01 14:33:00	0.0	NaN	United Kingdom
1987	536549	85226A	NaN	1	2010-12-01 14:34:00	0.0	NaN	United Kingdom
...
535322	581199	84581	NaN	-2	2011-12-07 18:26:00	0.0	NaN	United Kingdom
535326	581203	23406	NaN	15	2011-12-07 18:31:00	0.0	NaN	United Kingdom
535332	581209	21620	NaN	6	2011-12-07 18:35:00	0.0	NaN	United Kingdom
536981	581234	72817	NaN	27	2011-12-08 10:33:00	0.0	NaN	United Kingdom
538554	581408	85175	NaN	20	2011-12-08 14:06:00	0.0	NaN	United Kingdom

1454 rows x 8 columns

```
0.0    1454
Name: UnitPrice, dtype: int64
```

2. MISSING VALUE (CUSTOMER ID)

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1443	536544	21773 DECORATIVE ROSE BATHROOM BOTTLE	1	2010-12-01 14:32:00	2.51	Guest	United Kingdom
1444	536544	21774 DECORATIVE CATS BATHROOM BOTTLE	2	2010-12-01 14:32:00	2.51	Guest	United Kingdom
1445	536544	21786 POLKADOT RAIN HAT	4	2010-12-01 14:32:00	0.85	Guest	United Kingdom
1446	536544	21787 RAIN PONCHO RETROSPOT	2	2010-12-01 14:32:00	1.66	Guest	United Kingdom
1447	536544	21790 VINTAGE SNAP CARDS	9	2010-12-01 14:32:00	1.66	Guest	United Kingdom

Handling missing value pada CustomerID dilakukan dengan cara mengisi nilai kosong dengan CustomerID 'Guest'

3. NOISE DATA

```
(data['UnitPrice']==0.0).value_counts()
```

```
False    539394  
Name: UnitPrice, dtype: int64
```

Pada pemahaman data sebelumnya, diketahui bahwa terdapat nilai 0, sehingga dilakukan penanganan dengan menghilangkan nilai 0 tersebut

	Quantity	UnitPrice	CustomerID
count	539394.000000	539394.000000	406789.000000
mean	9.845871	4.632614	15287.795830
std	215.412253	96.984656	1713.573064
min	-80995.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13954.000000
50%	3.000000	2.080000	15152.000000
75%	10.000000	4.130000	16791.000000
max	80995.000000	38970.000000	18287.000000



	Quantity	UnitPrice	CustomerID
count	530104.000000	530104.000000	397884.000000
mean	10.542037	3.907625	15294.423453
std	155.524124	35.915681	1713.141560
min	1.000000	0.001000	12346.000000
25%	1.000000	1.250000	13969.000000
50%	3.000000	2.080000	15159.000000
75%	10.000000	4.130000	16795.000000
max	80995.000000	13541.330000	18287.000000

4. DUPLICATE DATA

```
print('Banyaknya Duplikasi Data : {}'.format(data.duplicated().sum()))
```

Banyaknya Duplikasi Data : 5226

BEFORE

```
jumlahDuplikat = data.duplicated().sum()  
print('Banyaknya Duplikasi Data : {}'.format(data.duplicated().sum()))
```

Banyaknya Duplikasi Data : 0

AFTER

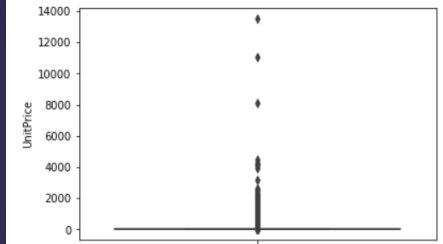
5. OUTLIER DATA

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
	1814	536544	DOT DOTCOM POSTAGE	1	2010-12-01 14:32:00	569.77	NaN	United Kingdom
	3041	536592	DOT DOTCOM POSTAGE	1	2010-12-01 17:06:00	607.49	NaN	United Kingdom
	6165	536876	DOT DOTCOM POSTAGE	1	2010-12-03 11:36:00	887.52	NaN	United Kingdom
	10812	537237	DOT DOTCOM POSTAGE	1	2010-12-06 09:58:00	863.74	NaN	United Kingdom
	11381	537240	DOT DOTCOM POSTAGE	1	2010-12-06 10:08:00	940.87	NaN	United Kingdom

	536834	581219	DOT DOTCOM POSTAGE	1	2011-12-08 09:28:00	1008.96	NaN	United Kingdom
	537254	581238	DOT DOTCOM POSTAGE	1	2011-12-08 10:53:00	1683.75	NaN	United Kingdom
	539368	581439	DOT DOTCOM POSTAGE	1	2011-12-08 16:30:00	938.59	NaN	United Kingdom
	540908	581492	DOT DOTCOM POSTAGE	1	2011-12-09 10:03:00	933.17	NaN	United Kingdom
	541540	581498	DOT DOTCOM POSTAGE	1	2011-12-09 10:26:00	1714.17	NaN	United Kingdom

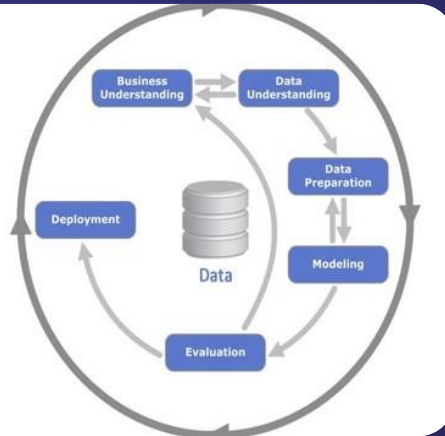
151 rows x 8 columns

<matplotlib.axes._subplots.AxesSubplot at 0x7f4cd670a250>



Pada attribute UnitPrice terdapat nilai ekstrim yang cenderung terdapat pada StockCode “DOT”

CRISP-DM Process Diagram



ALUR PROJECT

1. Business Understanding dan Pencarian Dataset
2. Data Understanding: explore data, Analisa awal dataset, quality control dataset, mencari masalah pada data seperti duplicated data, missing value, outliers, dan tipe data.
3. Data Preparation : Cleaning data dari masalah-masalah yang ada pada raw data, Pre-processing dataset seperti handling null values, duplicated data, noise data, modifikasi data type. Dilanjutkan dengan EDA (Exploratory Data Analysis) dan scaling.
4. Modeling: melakukan modeling unsupervised clustering K-Means. Analisis RFM
5. Evaluations: Silhouette Score
6. Deployment : membuat kesimpulan dan pemberian rekomendasi



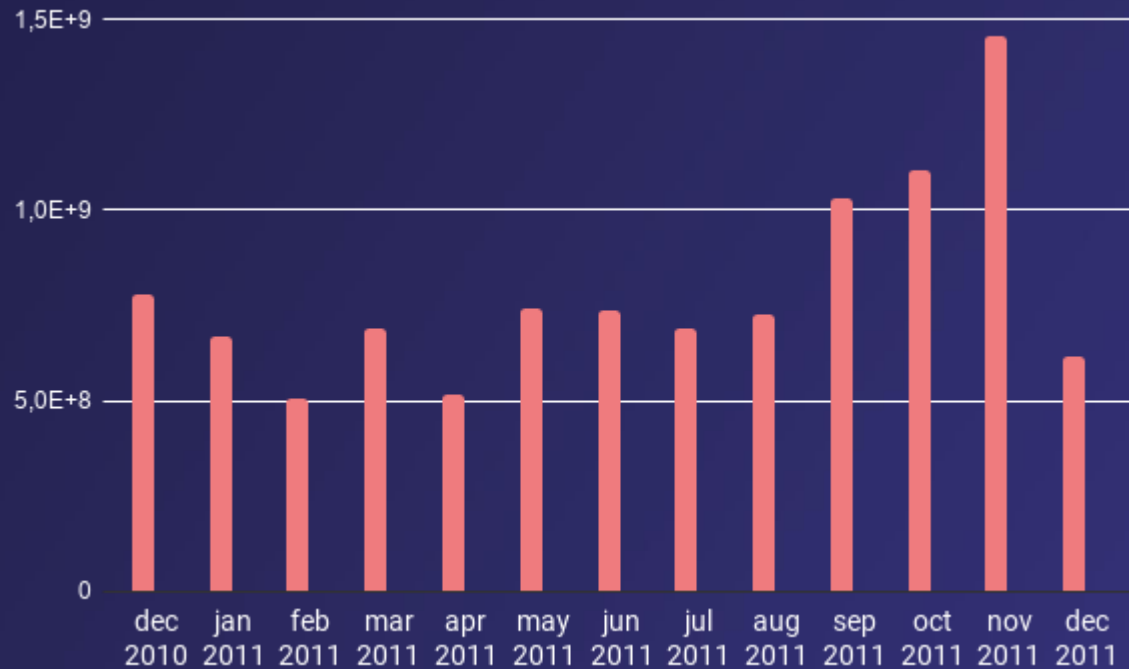
03 EXPLORATORY DATA ANALYSIS



522,727

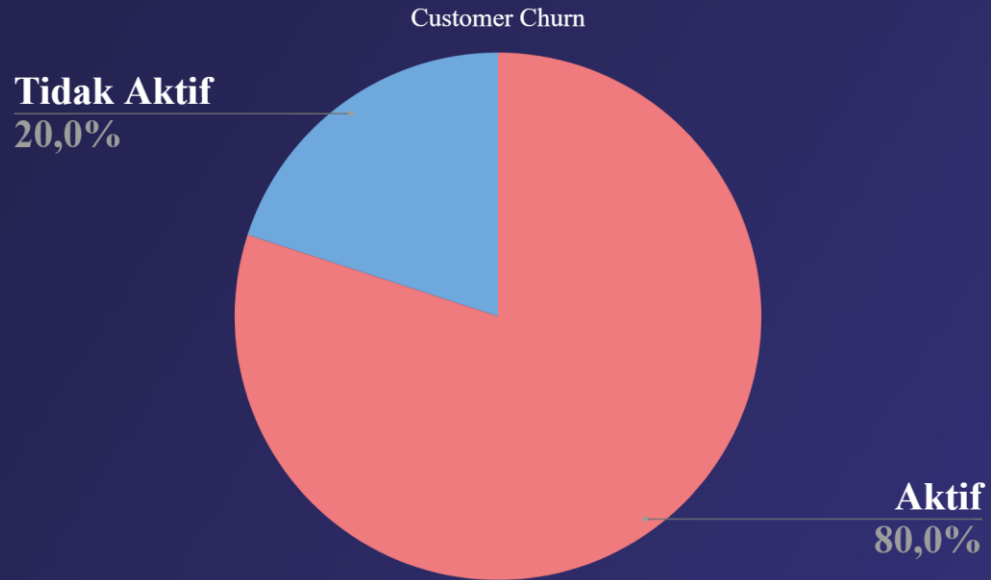
Jumlah Transaksi

TOTAL TRANSAKSI PER BULAN



Tertinggi : November 2011 (1.453.295.980) | Terendah : Februari 2011 (508.081.540)

CUSTOMER AKTIF DAN TIDAK AKTIF



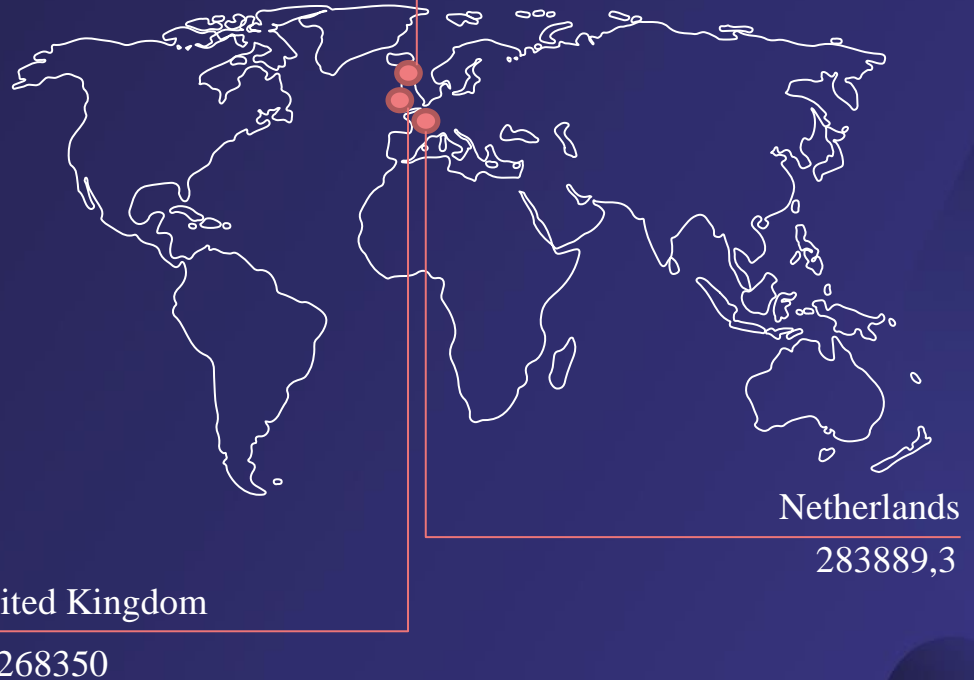
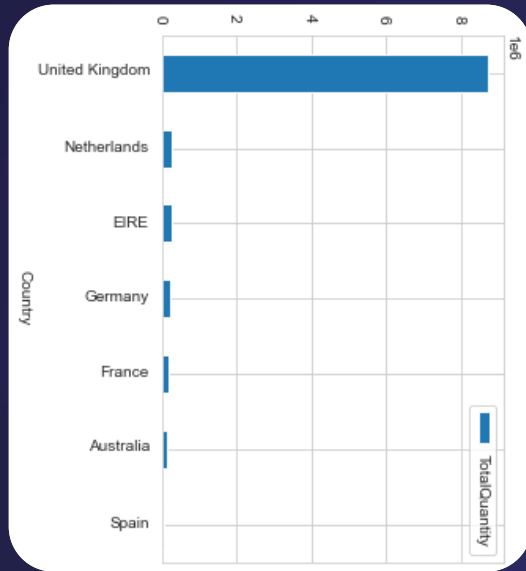
10 ITEM PENJUALAN TERTINGGI

StockCode	Description	TotalTransaction
23084	RABBIT NIGHT LIGHT	34478.4
22086	PAPER CHAIN KIT 50'S CHRISTMAS	28955.5
22197	POPCORN HOLDER	14188.8
85123A	WHITE HANGING HEART T-LIGHT HOLDER	14130.8
22423	REGENCY CAKESTAND 3 TIER	13774.1
22910	PAPER CHAIN KIT VINTAGE CHRISTMAS	12929.7
23355	HOT WATER BOTTLE KEEP CALM	11760.1
85099B	JUMBO BAG RED RETROSPOT	11639.3
79321	CHILLI LIGHTS	11485.8
23344	JUMBO BAG 50'S CHRISTMAS	11063.0

10 ITEM PENJUALAN TERENDAH

StockCode	Description	TotalTransaction
90021	LASER CUT MULTI STRAND NECKLACE	1
85036b	CHOCOLATE 1 WICK MORRIS BOX CANDLE	1
90020	FILIGREE DIAMANTE CHAIN	1
84550	CROCHET LILAC/RED BEAR KEYRING	1
85049c	ROMANTIC PINKS RIBBONS	1
35609A	PINK FEATHER CHRISTMAS DECORATION	1
20703	BLUE PADDED SOFT MOBILE	1
21120	*Boombox Ipod Classic	1
84510e	SET OF 4 GREEN CAROUSEL COASTERS	1
m	Manual	1

PERSEBARAN TRANSAKSI





04

RFM ANALYSIS . . .

WHAT IS RFM ANALYSIS?

- RFM analisis membantu perusahaan untuk memutuskan customer yang terpilih untuk menerima penawaran dan item promosi
- Sebuah proses perusahaan untuk menemukan langkah meningkatkan tingkat konsumsi customer
- Perusahaan sering menggunakan RFM untuk menentukan insentif bagi target customer yang hamper hilang
- RFM dapat membantu perusahaan menjaga relation dengan customer

KEUNGGULAN RFM

- Perusahaan memiliki data yang dapat digunakan untuk target marketing
- Budget marketing akan lebih efisien dan tepat sasaran terhadap customer
- Targeting customer yang spesifik dapat meningkatkan keuntungan dan mengurangi biaya
- Perusahaan dapat menentukan pemberian insentif yang tepat bagi customer untuk meningkatkan pembelian
- RFM cepat dan mudah untuk diinterpretasikan



RECENCY

Jarak antara tanggal sekarang dengan waktu terakhir transaksi



FREQUENCY

Jumlah pesanan setiap customer



MONETARY

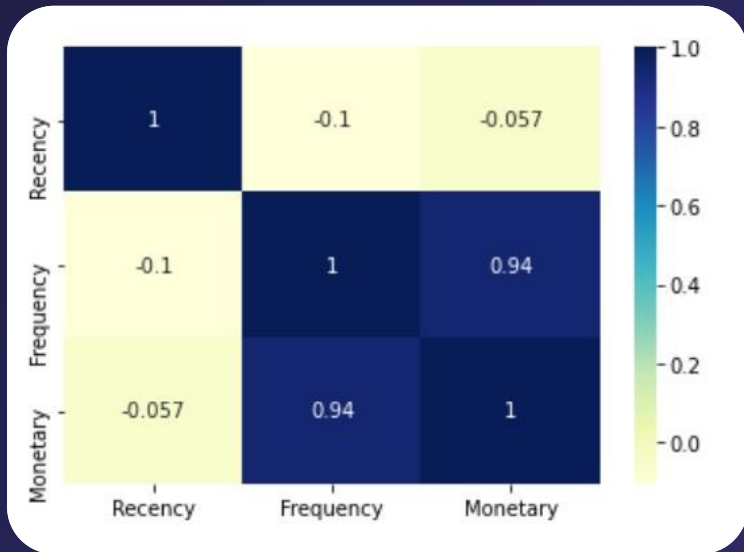
Jarak antara tanggal sekarang dengan waktu terakhir transaksi

CREATE RFM TABLE

```
rfm= rfm_data.groupby('CustomerID').agg({'InvoiceDate': lambda date: (PRESENT - date.max()).days,  
    'InvoiceNo': lambda num: len(num),  
    'TotalPrice': lambda price: price.sum()})
```

	Recency	Frequency	Monetary_Value
CustomerID			
12346.0	325	1	77183.60
12747.0	2	103	4196.01
12748.0	0	4397	31650.78
12749.0	3	199	4090.88
12820.0	3	59	942.34

RFM CORRELATION





RFM QUARTILE

```
segmented_rfm['R_quartile'] = segmented_rfm['Recency'].apply(RScore, args=('Recency',quantiles))
segmented_rfm['F_quartile'] = segmented_rfm['Frequency'].apply(FMScore, args=('Frequency',quantiles))
segmented_rfm['M_quartile'] = segmented_rfm['Monetary_Value'].apply(FMScore, args=('Monetary_Value',quantiles))
segmented_rfm.head()
```

```
segmented_rfm.head()
```

	Recency	Frequency	Monetary_Value	R_quartile	F_quartile	M_quartile
CustomerID						
12346.0	325	1	77183.60	4	4	1
12747.0	2	103	4196.01	1	1	1
12748.0	0	4397	31650.78	1	1	1
12749.0	3	199	4090.88	1	1	1
12820.0	3	59	942.34	1	2	2

RFM SEGMENTATION

	Recency	Frequency	Monetary_Value	R_quartile	F_quartile	M_quartile	RFM_Segment
CustomerID							
12346.0	325	1	77183.60	4	4	1	441
12747.0	2	103	4196.01	1	1	1	111
12748.0	0	4397	31650.78	1	1	1	111
12749.0	3	199	4090.88	1	1	1	111
12820.0	3	59	942.34	1	2	2	122

RFM SCORE

	Recency	Frequency	Monetary_Value	R_quartile	F_quartile	M_quartile	RFM_Segment	RFM_Score
CustomerID								
12346.0	325	1	77183.60	4	4	1	441	9
12747.0	2	103	4196.01	1	1	1	111	3
12748.0	0	4397	31650.78	1	1	1	111	3
12749.0	3	199	4090.88	1	1	1	111	3
12820.0	3	59	942.34	1	2	2	122	5
...
18281.0	180	7	80.82	4	4	4	444	12
18282.0	7	12	178.05	1	4	4	144	9
18283.0	3	719	2039.58	1	1	1	111	3
18287.0	42	70	1837.28	2	2	1	221	5
Guest	0	130033	1482339.76	1	1	1	111	3

Untuk melihat RFM Score secara lebih jelas dapat dilihat pada link berikut
<https://public.tableau.com/profile/gadih.nurlissalmi#!/vizhome/FinalProjectDSShift/Sheet1>

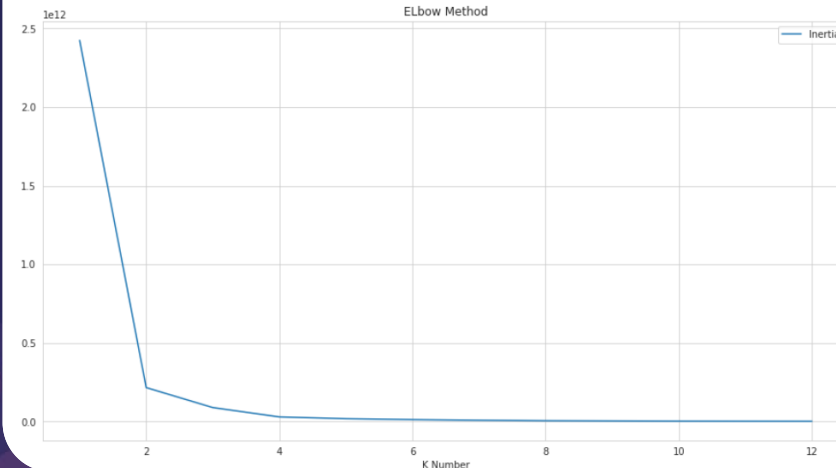
K-MEANS CLUSTERING

Selain dengan metode quartile, klasterisasi dilakukan dengan menggunakan algoritma K-Means Clustering dan melakukan evaluasi model dengan Silhouette Score.

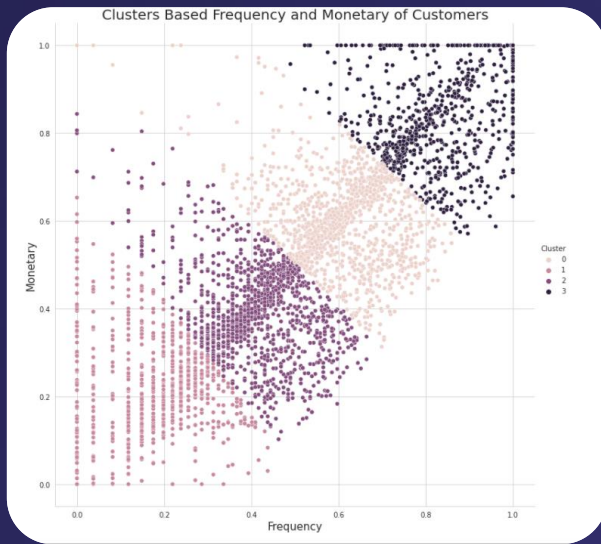
K yang digunakan sebagai hyperparameter adalah 4

```
inertia.plot(kind='line', figsize=(15,8))  
plt.title('Elbow Method')  
plt.xlabel('K Number')
```

```
Text(0.5, 0, 'K Number')
```



SILHOUETTE SCORE



```
from sklearn.metrics import silhouette_score, calinski_harabasz_score, davies_bouldin_score

silhouette_scores = []
silhouette_score(RFM_Table_scaled[['Frequency', 'Monetary_Value']], RFM_Table_scaled['Cluster'])
```

0.37307728002637885

CUSTOMER SEGMENTATION

Berdasarkan nilai Silhouette Score diperoleh cluster optimal adalah membagi segmentasi menjadi 4 cluster.

Klaster yang digunakan yaitu : Pelanggan Terbaik, Transaksi Besar, Pelanggan Hampir Hilang , dan Pelanggan yang Hilang

```
print("Pelanggan Terbaik: ",len(segmented_rfm[segmented_rfm['RFM_Segment']=='111']))
print("Transaksi Besar: ",len(segmented_rfm[segmented_rfm['M_quartile']==1]))
print('Pelanggan Hampir Hilang: ', len(segmented_rfm[segmented_rfm['RFM_Segment']=='134']))
print('Pelanggan yang Hilang: ',len(segmented_rfm[segmented_rfm['RFM_Segment']=='444']))
```

```
Pelanggan Terbaik: 408
Transaksi Besar: 980
Pelanggan Hampir Hilang: 22
Pelanggan yang Hilang: 341
```

BUSINESS STRATEGY

05

Setelah melakukan beberapa Analisa, kami memberikan rekomendasi sebagai berikut:

Perbaikan dalam data collecting

Terutama untuk
CustomerID

Potensi di Belanda dan Irelandia

Meningkatkan potensi
penjualan di negara tsb

Evaluasi Produk

- Meningkatkan penjualan produk yg banyak diminati
- Mengurangi yang penjualannya paling rendah

Adapun berdasarkan segmentasi yang telah dihasilkan, berikut merupakan strategi bisnis yang dapat digunakan untuk meningkatkan penjualan

Pelanggan Terbaik

Tidak ada insentif, produk baru, program loyalitas

Transaksi Besar

Pemasaran produk yang memiliki harga tinggi

Pelanggan Hampir Hilang

Pemberian insentif harga yang agresif

Pelanggan Hilang

Melakukan tindakan promosi normal

TERIMA KASIH

