



PENERAPAN CLUSTERING PERUSAHAAN SEKTOR INDUSTRI

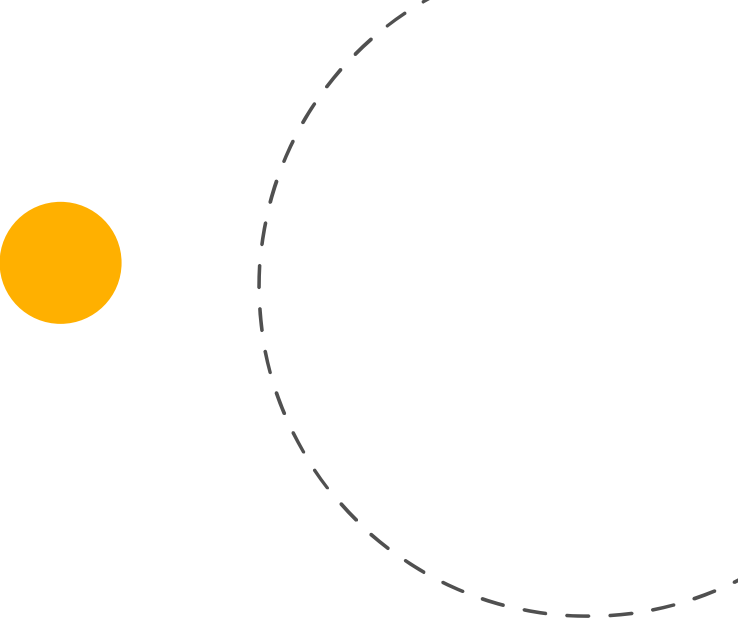


Hajar Hanifah

Business Understanding



Pandemi Covid-19 yang terjadi sejak tahun 2020 memberikan dampak negatif pada multisektor di dunia termasuk Indonesia. Indonesia mengalami pertumbuhan ekonomi negatif pada triwulan I dan II di tahun 2020. Sektor Industri merupakan salah satu sektor yang terdampak cukup besar semenjak terjadinya pandemi Covid-19.



- 
- Penelitian dilakukan dengan mengambil data dari emiten di sektor Industri di Bursa Efek Indonesia pada Quarter ke IV tahun 2021.
 - Tujuan penelitian adalah untuk melihat kelompok perusahaan mana saja yang terdampak dan mengalami penurunan, kenaikan, dan stabil.
 - Pengolahan data menggunakan teknik clustering.
- 
- 

Clustering

Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum.



Data Understanding

Data

Data yang diambil merupakan data statistical financial ratio Q4 2021 untuk emiten di sektor Industri.

Dataset

Data di dapatkan dari website IDX - Laporan Statistic Financial Data Ratio.

<https://www.idx.co.id/data-pasar/laporan-statistik/digital-statistic-beta/financial-data-ratio?>

Data Understanding

Emiten	Pihak yang melakukan penawaran umum, yaitu penawaran efek yang dilakukan oleh emiten untuk menjual efek kepada masyarakat berdasarkan tata cara yang diatur dalam peraturan undang-undang yang berlaku.
Sales Growth	Kenaikan jumlah penjualan dari tahun ke tahun atau dari waktu ke waktu.
Return of Asset (ROA)	Indikator untuk menunjukkan seberapa untung sebuah perusahaan dibandingkan dengan total asetnya.
Debt to Equity Ratio (DER)	Rasio hutang terhadap ekuitas atau rasio keuangan yang membandingkan jumlah hutang dengan ekuitas

Rumus

01 Sales Growth

(penjualan periode 2021 – penjualan periode 2020)

penjualan periode 2020

02 ROA

Laba setelah pajak

Total Asset

03 DER

Total Uang

Total Ekuitas

Data Preparation

Mencari Sales Growth

Mencari sales growth dengan membandingkan penjualan tahun sekarang dengan tahun sebelumnya (rumus terdapat di slide sebelumnya)

Menghapus Data Tertentu

Data yang dibutuhkan dalam analisa adalah data ROA, DER dan sales growth. Untuk data yang tidak dibutuhkan akan dihapus dari DataFrame.

Membandingkan Saham 2020 dan 2021

Terdapat beberapa emiten baru yang di tahun 2020 tidak ada, oleh karena itu kami memutuskan untuk mengambil data yang ada di kedua file Laporan Keuangan 2020 dan 2021 pada emiten di sektor industri.

Step K-Mean Clustering

Elbow Point

Menentukan distance / jumlah cluster dalam satu kumpulan data dengan mencari sudut sikunya.

Mengukur Jarak

mencari data point mana yang terdekat dengan centroid

Grouping

Dilakukan untuk menentukan objek berada di grup mana

Mereposisikan ulang centroid

Reposisikan centroid di titik tengah

Financial Data and Ratio Q4 Sektor Industri – 2021

emiten	sales 2021	sales 2020	sales growth	DER	ROA
AMFG	3,482.42	2,628.14	0.33	1.32	0.06
AMIN	126.41	89.53	0.41	1.14	-0.05
APII	196.58	191.18	0.03	0.49	0.05
ARKA	49.52	40.49	0.22	3.97	-0.02
ARNA	1,885.43	1,612.59	0.17	0.51	0.21
CAKK	192.80	148.83	0.30	0.88	0.03
CCSI	309.02	172.84	0.79	0.31	0.11
CTTH	63.36	76.58	-0.17	2.28	-0.05
HEXA	2,558.36	1,521.05	0.68	1.38	0.12
IKBI	1,607.30	959.71	0.67	0.67	0.01
IMPC	1,583.67	1,233.36	0.28	0.72	0.08
INTA	443.78	569.57	-0.22	-2.86	-0.32
JECC	1,265.57	1,095.40	0.16	1.58	-0.04
KBLI	1,105.12	1,328.94	-0.17	0.16	0.02
KBLM	895.40	657.97	0.36	0.38	-
KIAS	402.08	285.99	0.41	0.18	-0.01
KOBX	1,281.46	525.17	1.44	2.79	0.06
KOIN	1,779.88	894.80	0.99	8.53	-0.02
KPAL	30.56	30.56	0.00	3.15	-0.02
KRAH	179.50	179.50	0.00	16.33	-0.05
MARK	832.15	344.47	1.42	0.67	0.31
MLIA	3,190.36	2,759.79	0.16	0.91	0.08
SCCO	3,621.64	3,128.37	0.16	0.08	0.04
SINI	253.72	196.26	0.29	4.09	0.02
SKRN	258.01	357.83	-0.28	1.68	0.01
SPTO	1,582.87	1,347.45	0.17	0.54	0.06

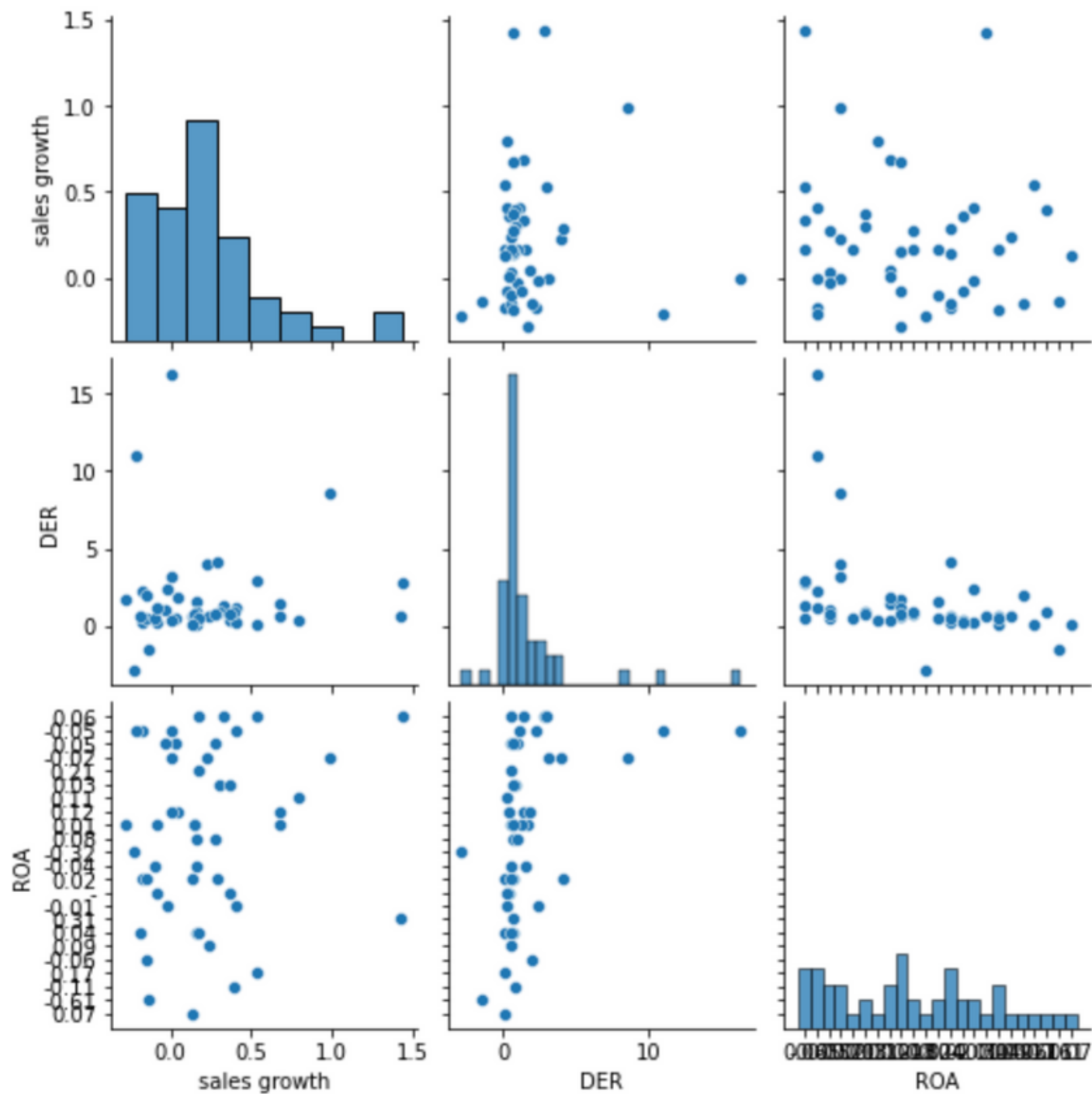
TOTO	1,303.30	1,146.93	0.14	0.65	0.02
UNTR	57,822.37	46,466.48	0.24	0.59	0.09
VOKS	1,223.14	1,431.34	-0.15	1.91	-0.06
ASGR	1,999.20	2,348.38	-0.15	0.49	0.02
BLUE	76.26	49.45	0.54	0.13	0.17
DYAN	226.98	163.18	0.39	0.86	-0.11
ICON	113.36	98.75	0.15	0.55	0.01
INDX	3.68	4.00	-0.08	0.19	-
JTPE	595.97	738.47	-0.19	0.65	0.04
KONI	86.97	63.32	0.37	0.70	0.03
LION	193.96	216.67	-0.10	0.51	-0.04
MDRN	57.95	67.77	-0.14	-1.55	-0.61
MFMI	108.39	104.66	0.04	1.82	0.12
SOSS	1,032.82	1,066.69	-0.03	1.02	0.05
TIRA	175.84	190.84	-0.08	1.21	0.01
ABMM	10,065.03	6,596.19	0.53	2.92	0.06
ASII	167,402.00	130,349.00	0.28	0.73	0.05
BHIT	12,400.93	10,818.20	0.15	0.73	0.01
BMTR	10,478.74	8,967.44	0.17	0.46	0.04
BNBR	1,569.74	1,979.03	-0.21	10.94	-0.05
MLPL	7,421.28	7,578.50	-0.02	2.33	-0.01
EMTK	9,597.63	8,518.51	0.13	0.12	0.07
TFAS	489.88	485.56	0.01	0.37	0.12

Deployment

Bahasa pemrograman yang digunakan adalah Python, adapun code dapat diakses di :

<https://github.com/Hajarhanifah/big-data/blob/main/clustering/clustering-sektor-industri-hajar-hanifah.ipynb>

Scatter Plot

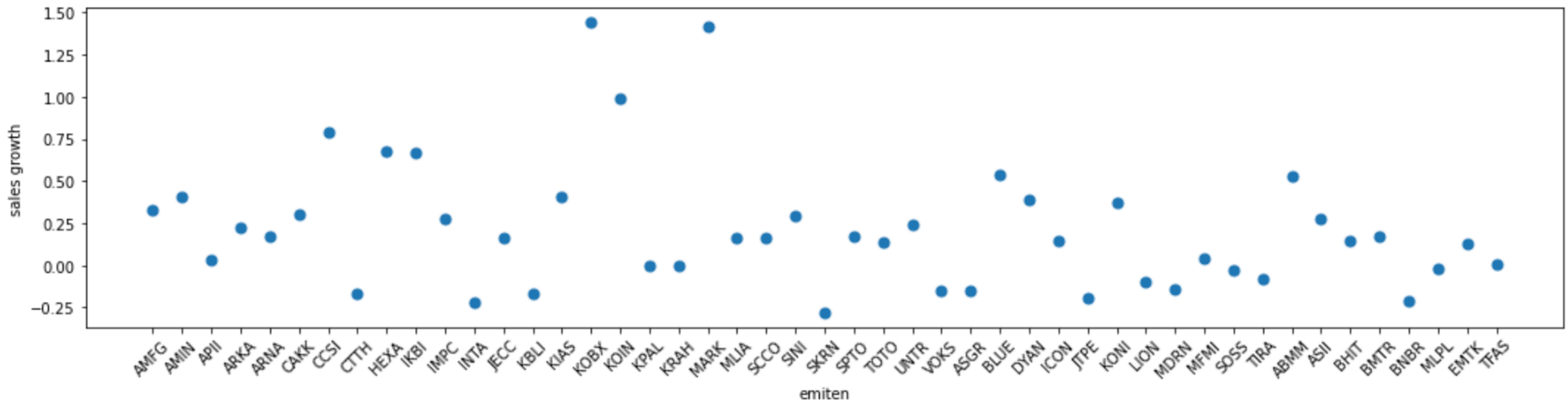


```
plt.figure(figsize=(20, 20))  
sns.pairplot(data, vars = data.columns[1:6]);
```

Scatter Plot

```
plt.figure(figsize=(18, 4))  
plt.xticks(rotation=45)  
sns.scatterplot(x='emiten', y='sales growth',  
               data=data, s=80)
```

<AxesSubplot:xlabel='emiten', ylabel='sales growth'>

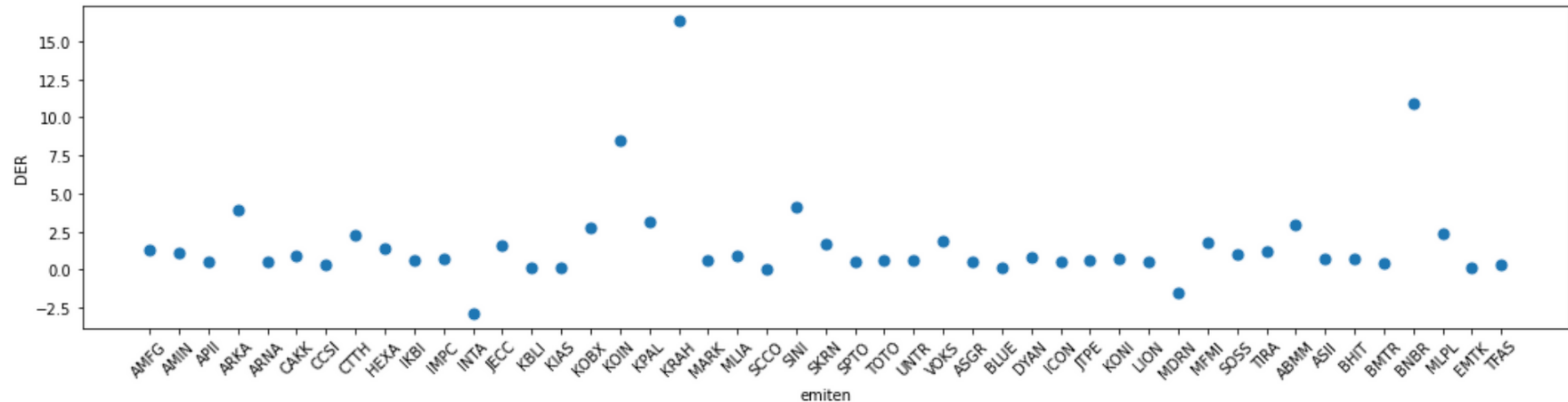


Scatter Plot

[54]:

```
plt.figure(figsize=(18, 4))
plt.xticks(rotation=45)
sns.scatterplot(x='emiten', y='DER',
               data=data, s=80)
```

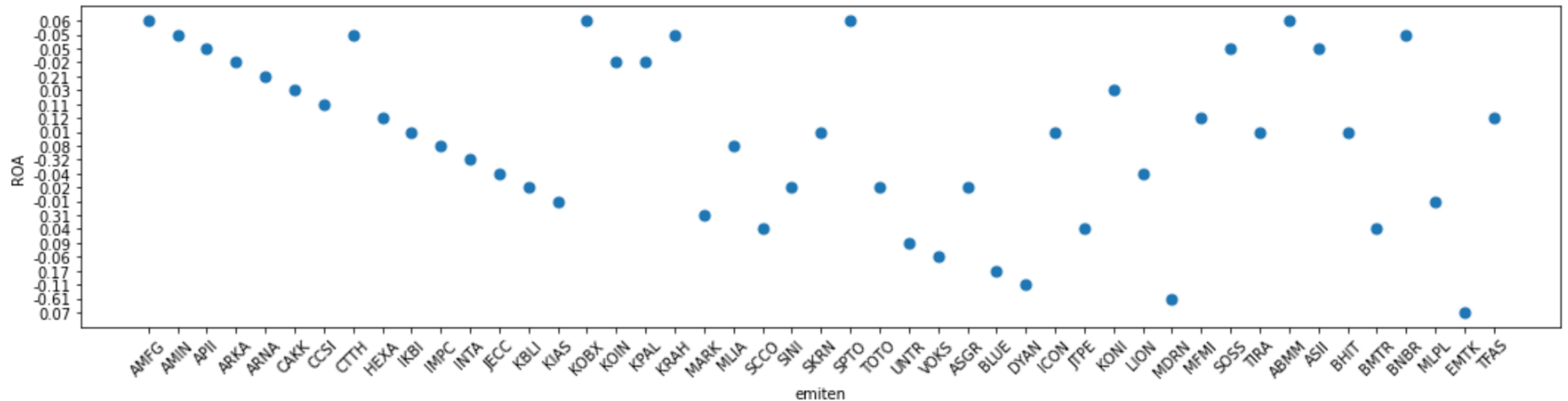
[54...] <AxesSubplot:xlabel='emiten', ylabel='DER'>



Scatter Plot

```
[55]: plt.figure(figsize=(18, 4))
plt.xticks(rotation=45)
sns.scatterplot(x='emiten', y='ROA',
               data=data, s=80)
```

[55...] <AxesSubplot:xlabel='emiten', ylabel='ROA'>



Outlier

```
print(f'Jumlah baris sebelum memfilter outlier: {len(data)}')

filtered_entries = np.array([False] * len(data))
for col in ['sales growth']:
    zscore = abs(stats.zscore(data[col]))
    filtered_entries = (zscore < 3) | filtered_entries

data = data[filtered_entries]

print(f'Jumlah baris setelah memfilter outlier: {len(data)}')
```

Jumlah baris sebelum memfilter outlier: 47
Jumlah baris setelah memfilter outlier: 45

```
data.describe()
```

	sales growth	DER
count	45.000000	45.000000
mean	0.165556	1.692889
std	0.286845	3.126935
min	-0.280000	-2.860000
25%	-0.030000	0.490000
50%	0.160000	0.730000
75%	0.300000	1.680000
max	0.990000	16.330000

K-Means – Elbow Point

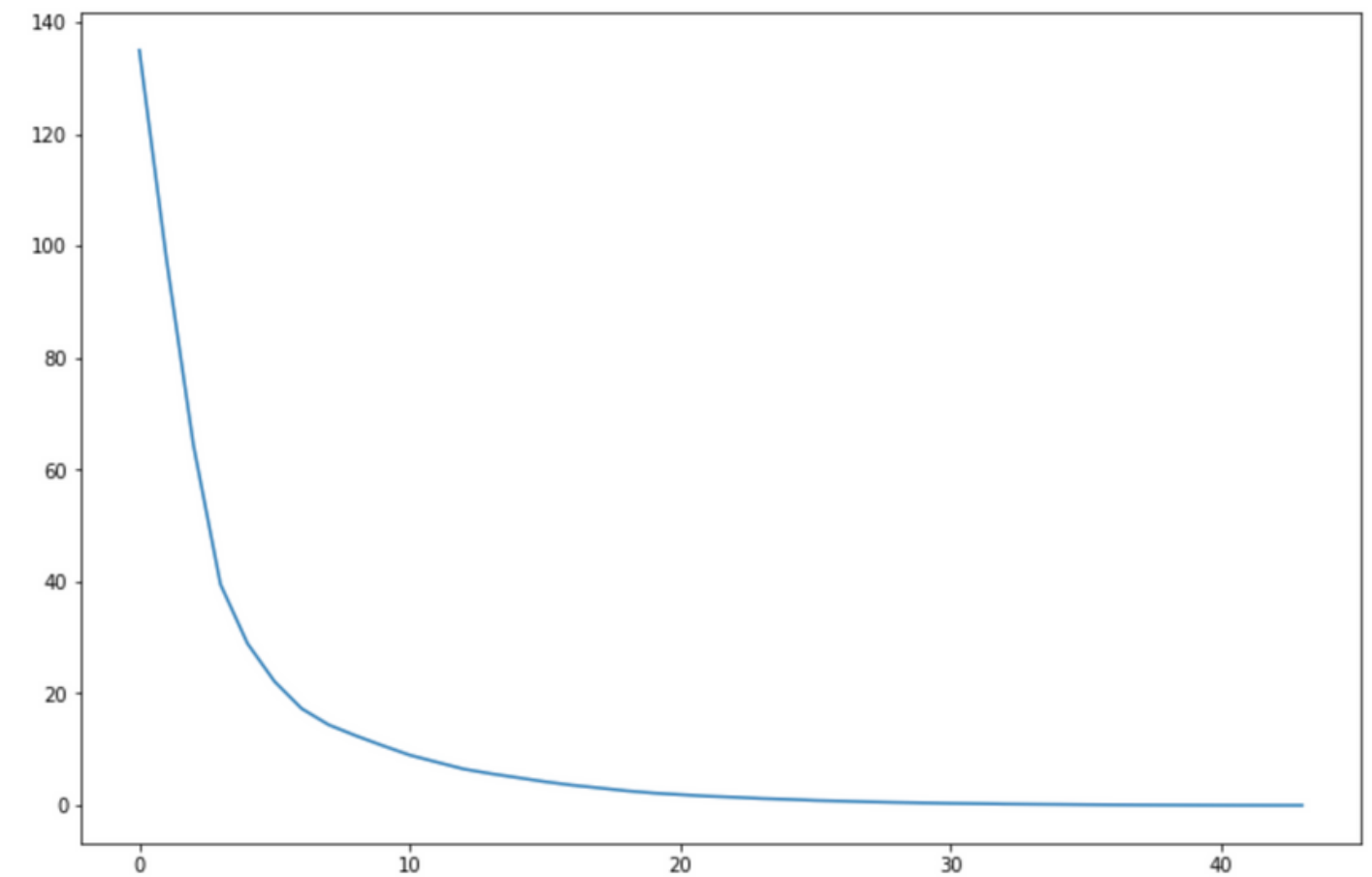
$$WSS = \sum_{i=1}^m (x_i - c_i)^2$$

Where x_i = data point and c_i = closest point to centroid

```
from sklearn.cluster import KMeans
inertia = []

for i in range(1, 45):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
    kmeans.fit(new_dataframe.values)
    inertia.append(kmeans.inertia_)

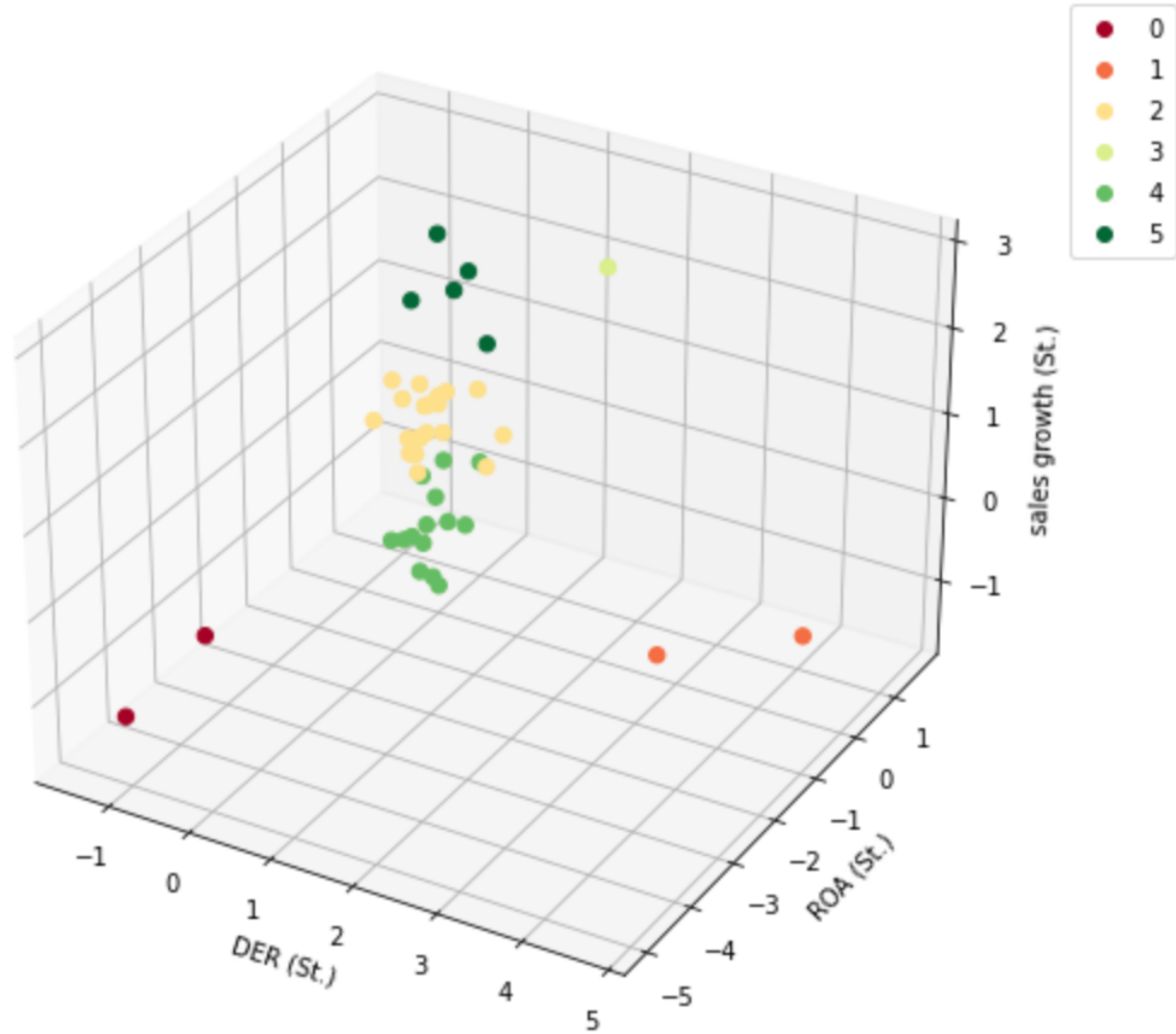
plt.figure(figsize=(12, 6))
plt.plot(inertia) #Elbow Method
```



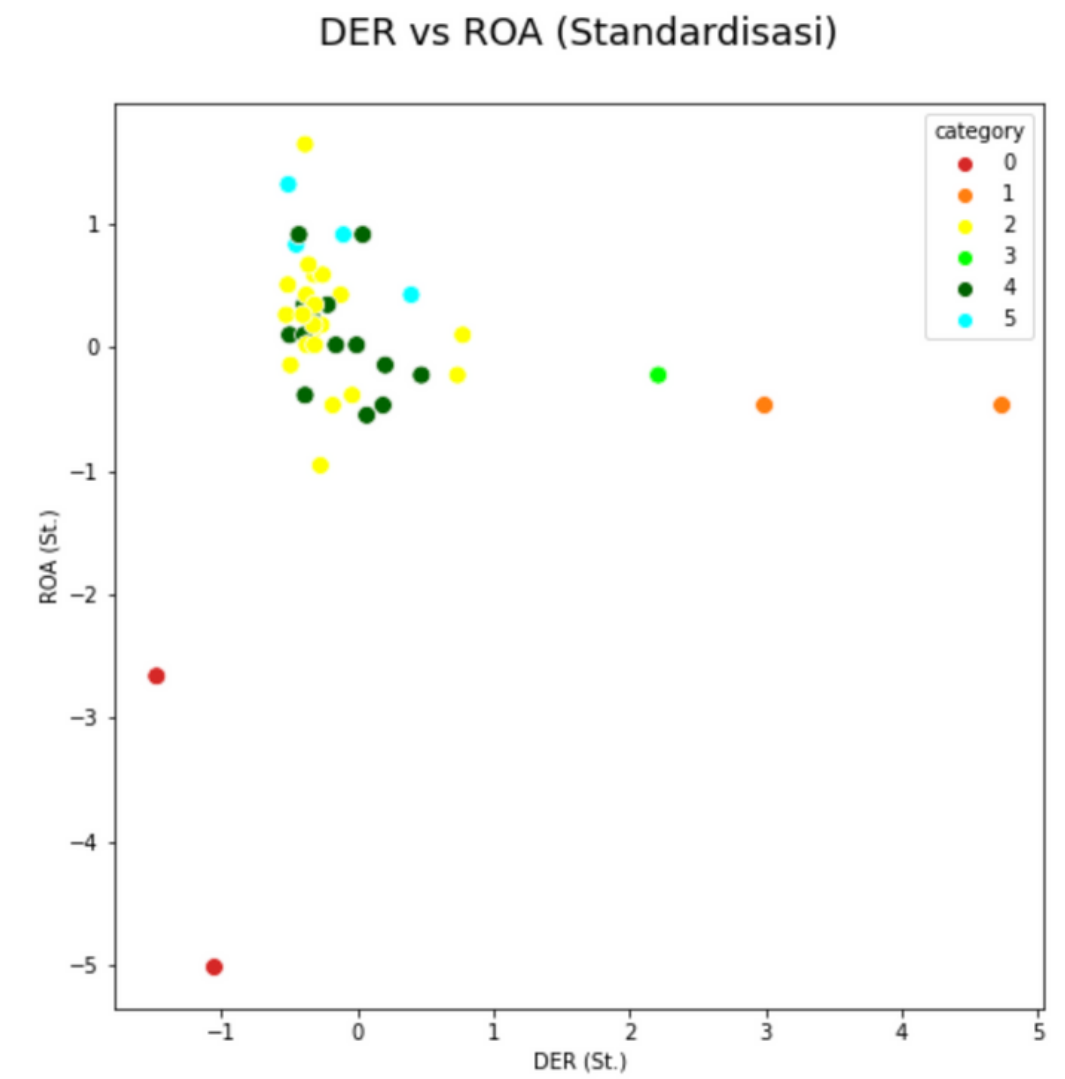
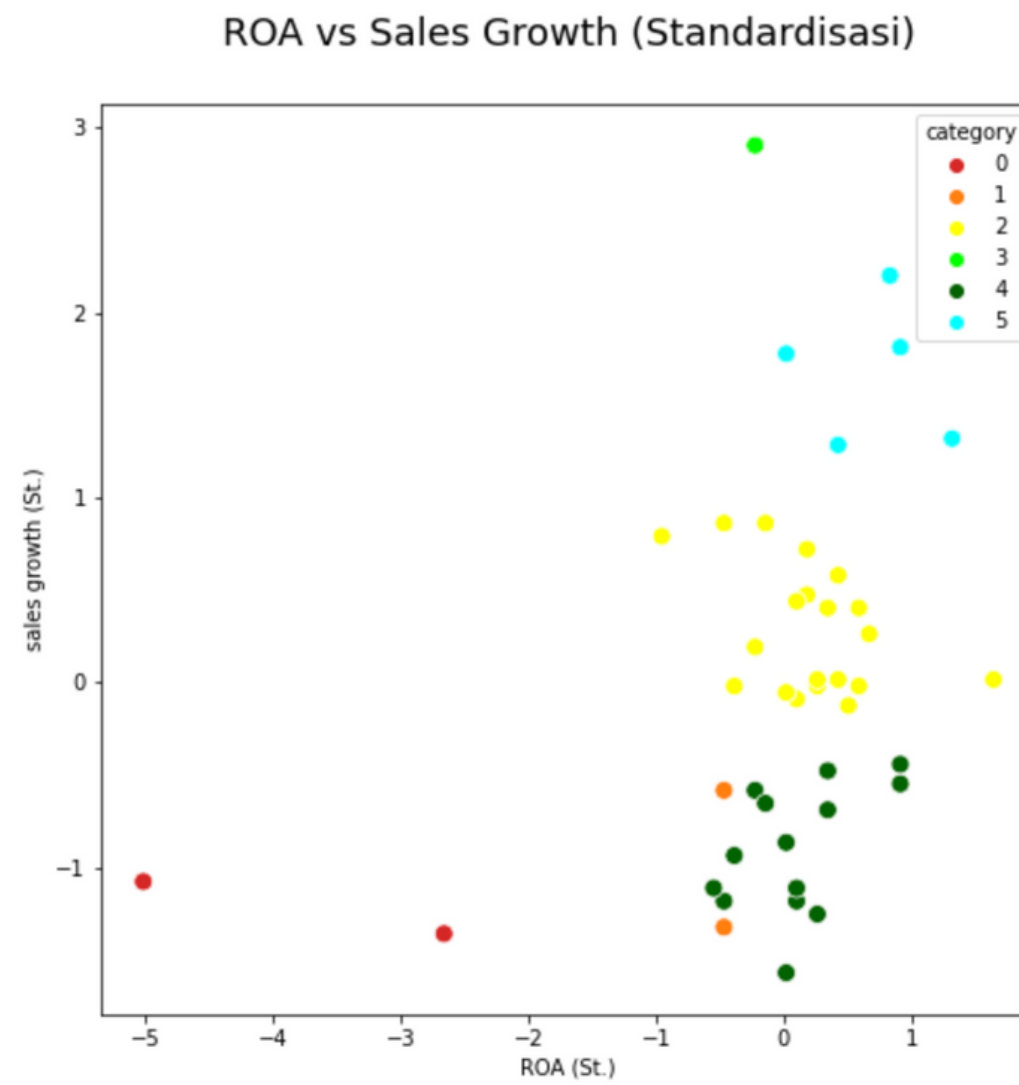
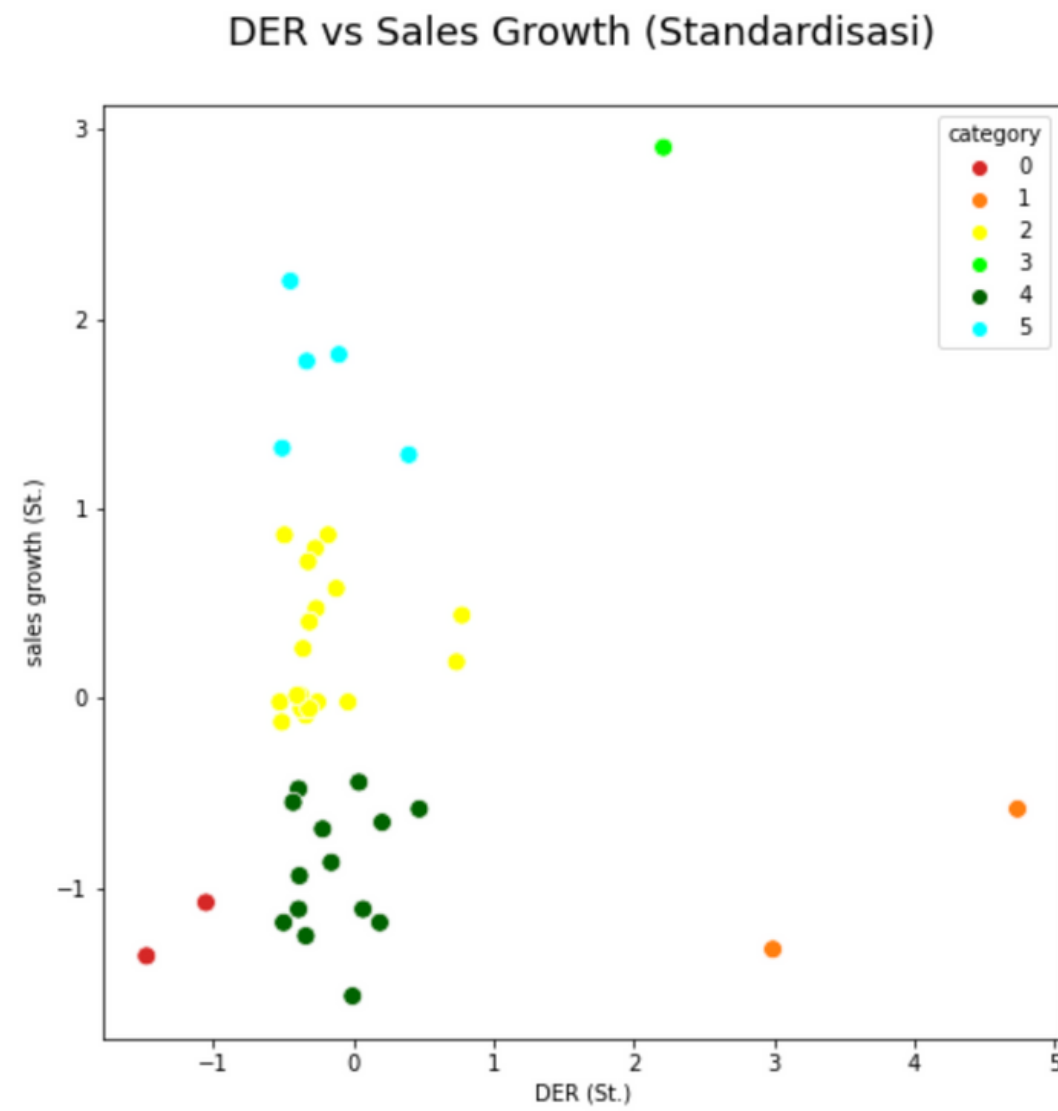
+ Code + Markdown

Diketahui titik siku : 6

Scatter Plot Clustering



Scatter Plot Clustering



Data Clustering

	emiten	stock name	category	sales growth (St.)	DER (St.)	ROA (St.)
0	AMFG	Asahimas Flat Glass Tbk	2.0	0.579766	-0.120598	0.423833
1	AMIN	PT Ateliers Mecaniques D Indonesie Tbk.	2.0	0.861814	-0.178813	-0.468922
2	APII	PT Arita Prima Indonesia Tbk.	4.0	-0.477915	-0.389033	0.342673
3	ARKA	PT Arkha Jayanti Persada Tbk.	2.0	0.191949	0.736453	-0.225443
4	ARNA	Arwana Citramulia Tbk	2.0	0.015669	-0.382565	1.641226
5	CAKK	Cahayaputra Asa Keramik Tbk	2.0	0.473998	-0.262901	0.180354
6	CCSI	Communication Cable Systems Indonesia Tbk	5.0	2.201543	-0.447248	0.829630
7	CTTH	Citatah Tbk	4.0	-1.183035	0.189881	-0.468922
8	HEXA	Hexindo Adiperkasa Tbk	5.0	1.813726	-0.101193	0.910790
9	IMPC	Impack Pratama Industri Tbk	5.0	1.778470	-0.330818	0.018035
10	INTA	Intraco Penta Tbk	2.0	0.403486	-0.314648	0.586152
11	JECC	Jembo Cable Company Tbk	0.0	-1.359316	-1.472476	-2.660228
12	KBLI	KMI Wire & Cable Tbk	2.0	-0.019587	-0.036510	-0.387762
13	KIAS	Keramika Indonesia Assosiasi Tbk	4.0	-1.183035	-0.495760	0.099195
14	KOIN	Kokoh Inti Arebama Tbk	2.0	0.861814	-0.489292	-0.144284
15	KPAL	Steadfast Marine Tbk	3.0	2.906663	2.211229	-0.225443
16	KRAH	PT Grand Kartech Tbk	4.0	-0.583683	0.471253	-0.225443
17	MARK	PT Mark Dynamics Indonesia Tbk.	1.0	-0.583683	4.733872	-0.468922
18	MLIA	Mulia Industrindo Tbk	2.0	-0.019587	-0.253199	0.586152
19	SCCO	Supreme Cable Manufacturing & Commerce Tbk	2.0	-0.019587	-0.521634	0.261514
20	SINI	Singaraja Putra Tbk	2.0	0.438742	0.775263	0.099195
21	SKRN	Superkrane Mitra Utama Tbk	4.0	-1.570852	-0.004168	0.018035
22	SPTO	Surya Pertiwi Tbk	2.0	0.015669	-0.372862	0.423833

Centroid

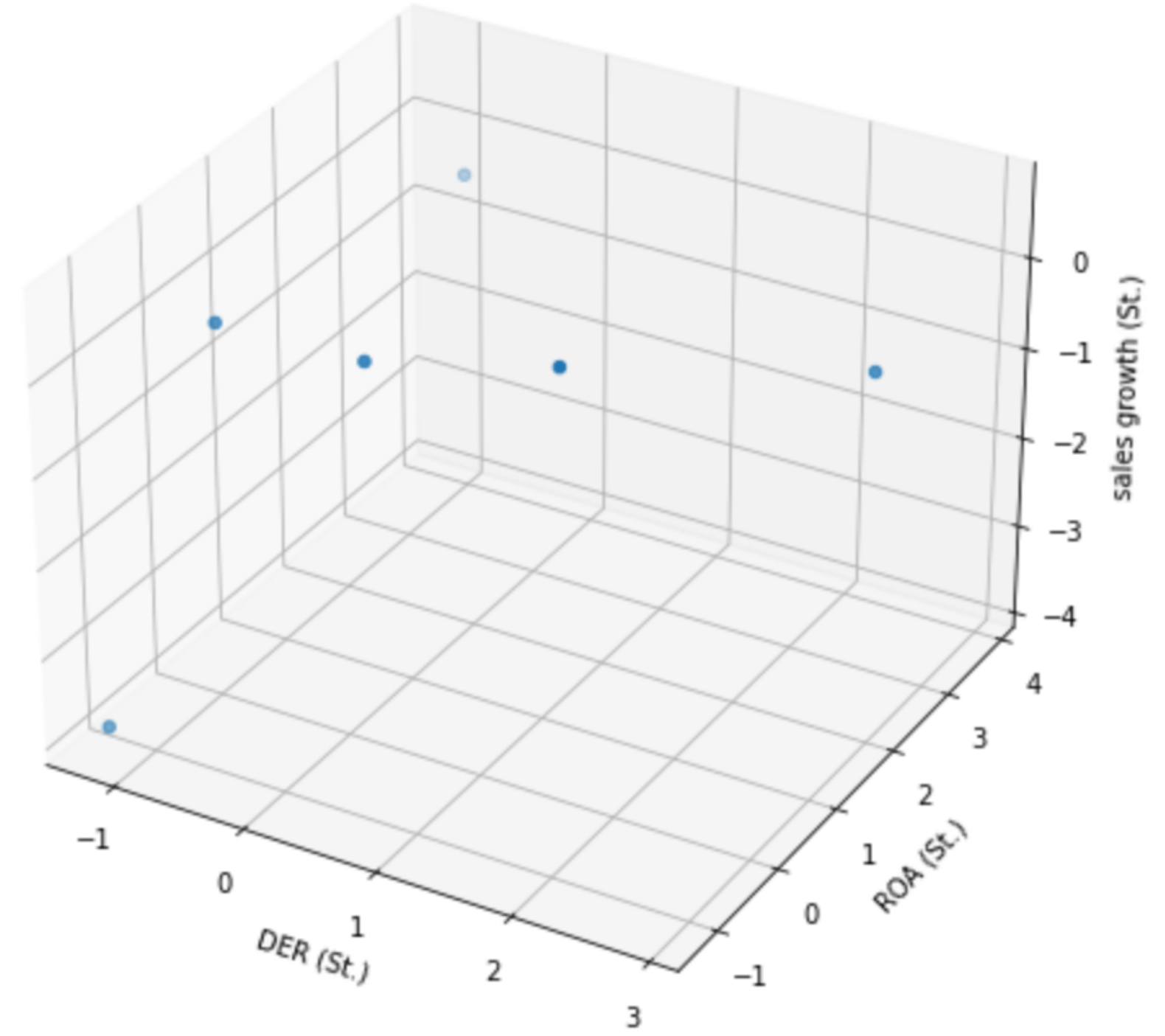
```
kmeans.cluster_centers_
```

```
array([[ 0.26917693, -0.21931691,  0.19581341],  
       [-0.90098723, -0.13006958,  0.08760074],  
       [-0.95387127,  3.86226622, -0.46892159],  
       [-1.21829143, -1.26063846, -3.83704106],  
       [ 2.90666316,  2.21122907, -0.22544307],  
       [ 1.67975359, -0.19757106,  0.69977529]])
```

```
from mpl_toolkits.mplot3d import Axes3D
```

```
fig = plt.figure()  
ax = fig.add_subplot(111, projection = '3d')  
x = [0.07796577, 3.3825386, -0.19978728, -0.22183117, -0.32525515]  
y = [1.27721216, -0.22442569, -1.39490532, 0.15549501, -0.03911061]  
z = [-0.22773712, 0.35006475, -0.89274149, 4.44060221, 0.14030381]
```

```
ax.set_xlabel("DER (St.)")  
ax.set_ylabel("ROA (St.)")  
ax.set_zlabel("sales growth (St.)")  
ax.scatter(x, y, z)  
plt.show()
```



Cluster Description

```
#Highly Negatively Impacted
df[df['category'] == 0].describe()
```

	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER
count	2.0	2.000000	2.000000	2.000000	2.000000	2.000000
mean	0.0	-1.218291	-1.260638	-3.837041	0.100000	1.700000
std	0.0	0.199438	0.299583	1.664265	0.084853	0.169706
min	0.0	-1.359316	-1.472476	-5.013854	0.040000	1.580000
25%	0.0	-1.288803	-1.366557	-4.425447	0.070000	1.640000
50%	0.0	-1.218291	-1.260638	-3.837041	0.100000	1.700000
75%	0.0	-1.147779	-1.154720	-3.248635	0.130000	1.760000
max	0.0	-1.077267	-1.048801	-2.660228	0.160000	1.820000

```
: #Negatively Impacted
df[df['category'] == 1].describe()
```

	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER
count	2.0	2.000000	2.000000	2.000000	2.000000	2.000000
mean	1.0	-0.953871	3.862266	-0.468922	0.070000	1.620000
std	0.0	0.523525	1.232636	0.000000	0.127279	1.004092
min	1.0	-1.324059	2.990661	-0.468922	-0.020000	0.910000
25%	1.0	-1.138965	3.426464	-0.468922	0.025000	1.265000
50%	1.0	-0.953871	3.862266	-0.468922	0.070000	1.620000
75%	1.0	-0.768777	4.298069	-0.468922	0.115000	1.975000
max	1.0	-0.583683	4.733872	-0.468922	0.160000	2.330000

Cluster Description

```
#Neutral Impacted
df[df['category'] == 2].describe()
```

	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER
count	21.0	21.000000	21.000000	21.000000	21.000000	21.000000
mean	2.0	0.269177	-0.219317	0.195813	0.127619	1.541429
std	0.0	0.341022	0.344622	0.516253	0.322442	2.993878
min	2.0	-0.125355	-0.521634	-0.955879	-0.280000	-2.860000
25%	2.0	-0.019587	-0.372862	0.018035	-0.150000	0.460000
50%	2.0	0.191949	-0.314648	0.180354	0.150000	0.590000
75%	2.0	0.473998	-0.253199	0.423833	0.240000	1.140000
max	2.0	0.861814	0.775263	1.641226	0.990000	10.940000

```
#Positively Impacted
df[df['category'] == 3].describe()
```

	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER
count	1.0	1.000000	1.000000	1.000000	1.0	1.00
mean	3.0	2.906663	2.211229	-0.225443	0.0	3.15
std	NaN	NaN	NaN	NaN	NaN	NaN
min	3.0	2.906663	2.211229	-0.225443	0.0	3.15
25%	3.0	2.906663	2.211229	-0.225443	0.0	3.15
50%	3.0	2.906663	2.211229	-0.225443	0.0	3.15
75%	3.0	2.906663	2.211229	-0.225443	0.0	3.15
max	3.0	2.906663	2.211229	-0.225443	0.0	3.15

Cluster Description

```
#Highly Positively Impacted
df[df['category'] == 4].describe()
```

	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER
count	14.0	14.000000	14.000000	14.000000	13.000000	13.000000
mean	4.0	-0.900987	-0.130070	0.087601	0.203846	2.110769
std	0.0	0.344606	0.295318	0.448904	0.280373	4.480825
min	4.0	-1.570852	-0.495760	-0.550081	-0.150000	-1.550000
25%	4.0	-1.165407	-0.387416	-0.205153	-0.030000	0.180000
50%	4.0	-0.900987	-0.186898	0.058615	0.170000	0.700000
75%	4.0	-0.601311	0.062940	0.322384	0.410000	1.380000
max	4.0	-0.442659	0.471253	0.910790	0.680000	16.330000

```
df[df['category'] == 5].describe()
```

	category	sales growth (St.)	DER (St.)	ROA (St.)	sales growth	DER
count	5.0	5.000000	5.000000	5.000000	5.000000	5.000000
mean	5.0	1.679754	-0.197571	0.699775	0.290000	1.052000
std	0.0	0.382491	0.366584	0.495671	0.302572	0.690051
min	5.0	1.284886	-0.505463	0.018035	-0.170000	0.670000
25%	5.0	1.320142	-0.447248	0.423833	0.280000	0.720000
50%	5.0	1.778470	-0.330818	0.829630	0.280000	0.730000
75%	5.0	1.813726	-0.101193	0.910790	0.390000	0.860000
max	5.0	2.201543	0.396867	1.316588	0.670000	2.280000



Terima Kasih

Hajar Hanifah

email : hajar.hanifah@gmail.com

github : github.com/hajarhanifah

kaggle : [hajarhanifah](https://www.kaggle.com/hajarhanifah)

