

```
In [1]: import pandas as pd

from sklearn.cluster import KMeans

from sklearn.pipeline import make_pipeline

df = pd.read_csv('stock.csv', skiprows=1)

companies = df.iloc[:, 0]

movements = df.iloc[:, 1:964]

# Import Normalizer
from sklearn.preprocessing import Normalizer

# Create a normalizer: normalizer
normalizer = Normalizer()

# Create a KMeans model with 10 clusters: kmeans
kmeans = KMeans(n_clusters=10)

# Make a pipeline chaining normalizer and kmeans: pipeline
pipeline = make_pipeline(normalizer, kmeans)

# Fit pipeline to the daily price movements
pipeline.fit(movements)

# Predict the cluster labels: labels
labels = pipeline.predict(movements)

# Create a DataFrame aligning labels and companies: df
df = pd.DataFrame({'labels': labels, 'companies': companies})

# Display df sorted by cluster label
print(df.sort_values('labels'))
```

	labels	companies
0	0	AIG
38	1	Pfizer
41	1	Royal Dutch Shell
43	1	Schlumberger
12	1	DuPont de Nemours
11	1	Chevron
31	1	3M
9	1	ConocoPhillips
48	1	Total
5	1	British American Tobacco
52	1	Valero Energy
56	1	Exxon
7	1	Caterpillar
3	2	Boeing
35	2	Northrop Grumman
28	2	Lockheed Martin
47	3	Toyota
44	3	Sony
53	3	Walgreen
33	3	Mitsubishi
29	3	MasterCard
14	3	Ford
10	3	Cisco
6	3	Canon
20	3	Honda
19	3	Home Depot
30	4	McDonalds
37	4	Pepsi
40	4	Philip Morris
27	4	Coca Cola
57	5	Xerox
25	5	JPMorgan Chase
17	5	Goldman Sachs
4	5	Bank of America
2	5	American express
54	5	Wells Fargo
15	5	General Electrics
50	6	Texas instruments
49	6	Taiwan Semiconductor Manufacturing
13	6	Dell
21	6	HP
58	6	Yahoo
34	6	Navistar
23	6	Intel
55	7	Wal-Mart
42	8	SAP
16	8	Google/Alphabet
45	8	Sanofi-Aventis
46	8	Symantec
1	8	Amazon
32	8	Microsoft
18	8	GlaxoSmithKline
36	8	Novartis
51	8	Unilever
22	8	IBM
26	9	Kimberly-Clark
8	9	Colgate-Palmolive
24	9	Johnson & Johnson
39	9	Procter Gamble

```
In [3]: # Perform the necessary imports

from scipy.cluster.hierarchy import linkage, dendrogram

import matplotlib.pyplot as plt

import numpy as np

companies = list(companies)

movements = np.array(movements)

# Import normalize
from sklearn.preprocessing import normalize

# Normalize the movements: normalized_movements
normalized_movements = normalize(movements)

# Calculate the linkage: mergings
mergings = linkage(normalized_movements, method='complete')

# Plot the dendrogram
#dendrogram(mergings, labels=companies, above_threshold_color='#bcbddc', leaf_rotation=90, leaf_font_size=6)

dendrogram(mergings, labels=companies, leaf_rotation=90, leaf_font_size =6)

plt.show()
```

```
In [7]: from scipy.cluster.hierarchy import fcluster

mergings = linkage(movements, method='complete')

labels = fcluster(mergings, 15, criterion='distance')

print(labels)

pairs = pd.DataFrame({'labels': labels, 'companies': companies})

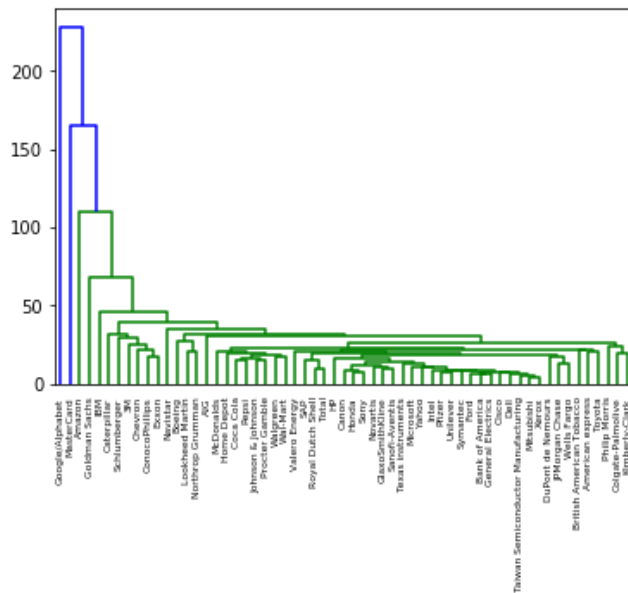
print('\n', pairs.sort_values('labels'))

dendrogram(mergings, labels=companies, leaf_rotation=90, leaf_font_size =6)

plt.show()
```

[29 33 23 9 18 25 18 6 26 1 18 3 21 18 18 18 35 32 18 14 18 19 31 18
 11 20 27 10 7 34 15 4 18 18 30 8 18 10 18 11 28 16 17 5 18 18 18 24
 16 18 18 18 22 12 20 13 2 18 18]

	labels	companies
9	1	ConocoPhillips
56	2	Exxon
11	3	Chevron
31	4	3M
43	5	Schlumberger
7	6	Caterpillar
28	7	Lookheed Martin
35	8	Northrop Grumman
3	9	Boeing
37	10	Pepsi
27	10	Coca Cola
24	11	Johnson & Johnson
39	11	Procter Gamble
53	12	Walgreen
55	13	Wal-Mart
19	14	Home Depot
30	15	McDonalds
41	16	Royal Dutch Shell
48	16	Total
42	17	SAP
44	18	Sony
38	18	Pfizer
45	18	Sanofi-Aventis
36	18	Novartis
46	18	Symantec
33	18	Mitsubishi
32	18	Microsoft
49	18	Taiwan Semiconductor Manufacturing
58	18	Yahoo
51	18	Unilever
14	18	Ford
50	18	Texas instruments
23	18	Intel
6	18	Canon
20	18	Honda
18	18	GlaxoSmithKline
10	18	Cisco
15	18	General Electrics
4	18	Bank of America
57	18	Xerox
13	18	Dell
21	19	HP
25	20	JPMorgan Chase
54	20	Wells Fargo
12	21	DuPont de Nemours
52	22	Valero Energy
2	23	American express
47	24	Toyota
5	25	British American Tobacco
8	26	Colgate-Palmolive
26	27	Kimberly-Clark
40	28	Philip Morris
0	29	AIG
34	30	Navistar
22	31	IBM
17	32	Goldman Sachs
1	33	Amazon
29	34	MasterCard
16	35	Google/Alphabet



```
In [8]: from sklearn.manifold import TSNE

# Create a TSNE instance: model
model = TSNE(learning_rate=50)

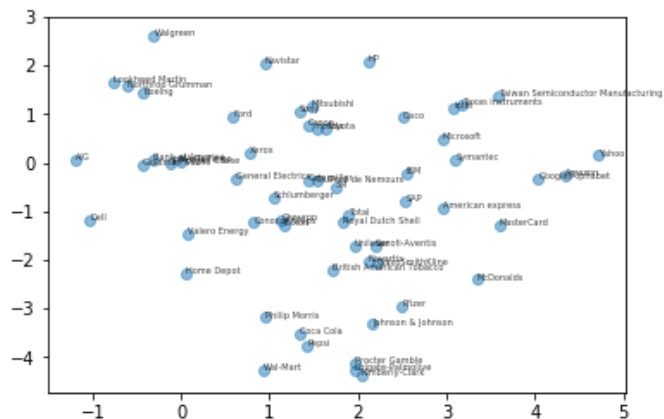
# Apply fit_transform to normalized_movements: tsne_features
tsne_features = model.fit_transform(normalized_movements)

# Select the 0th feature: xs
xs = tsne_features[:,0]

# Select the 1th feature: ys
ys = tsne_features[:,1]

# Scatter plot
plt.scatter(xs, ys, alpha=0.5)

# Annotate the points
for x, y, company in zip(xs, ys, companies):
    plt.annotate(company, (x, y), fontsize=5, alpha=0.75)
plt.show()
```



```
In [9]: from sklearn.decomposition import PCA
```

```
model = PCA()
```

```
model.fit(movements)
```

```
transformed = model.transform(movements)
```

```
print(transformed)
```

```
[[-4.93092609e+00 -9.16709808e-01 -1.01966495e+00 ... 1.63861006e-02
 2.34476152e-02 1.52655666e-15]
 [ 4.39554135e+01 1.69348558e+01 8.50040502e+01 ... 3.82284357e-03
 6.34618304e-03 -5.27355937e-15]
 [-2.93912026e+00 1.47423723e+00 -6.76319243e-01 ... -4.89200696e-02
 1.61862266e-02 1.83186799e-15]
 ...
 [-3.29495208e+00 -3.14918742e-01 -9.20745678e-01 ... -2.04801434e-02
 2.00771121e-07 2.31759056e-15]
 [-1.11493066e+01 -3.87495016e+00 -1.79776671e+00 ... 3.02024907e+00
 -4.82023017e-01 2.51187959e-15]
 [-9.53495780e+00 -3.23186980e+00 -4.48927270e-01 ... -5.58762268e-02
 -2.87600575e-02 2.16493490e-15]]
```