

Exercise 9: Which stocks move together?

In the previous exercise, you clustered companies by their daily stock price movements. So which company have stock prices that tend to change in the same way? Now inspect the cluster labels from your clustering to find out.

From the course *Transition to Data Science*. [Buy the entire course for just \\$10](#) for many more exercises and helpful video lectures.

Step 1: Load the data (*written for you*).

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

fn = 'datasets/company-stock-movements-2010-2015-incl.csv'
stocks_df = pd.read_csv(fn, index_col=0)
```

Step 2: Run your code from the previous exercise (*filled in for you*).

```
In [2]: companies = list(stocks_df.index)
movements = stocks_df.values

from sklearn.preprocessing import Normalizer
from sklearn.cluster import KMeans
from sklearn.pipeline import make_pipeline

normalizer = Normalizer()
kmeans = KMeans(n_clusters=14)
pipeline = make_pipeline(normalizer, kmeans)
pipeline.fit(movements)

Out[2]: Pipeline(steps=[('normalizer', Normalizer(copy=True, norm='l2')), ('kmeans', KMeans(
    n_clusters=14, n_init=10, n_jobs=1, precompute_distances='auto',
    random_state=None, tol=0.0001, verbose=0))])
```

Step 3: Use the `.predict()` method of the pipeline to predict the labels for movements.

```
In [3]: labels = pipeline.predict(movements)
```

Step 4: Align the cluster labels with the list of company names `companies` by creating a DataFrame `df` with `labels` and `companies` as columns.

```
In [23]: df = pd.DataFrame({'labels': labels, 'companies': companies})
```

Step 5: Now display the DataFrame, sorted by cluster label. To do this, use the `.sort_values()` method of `df` to sort the DataFrame by the `'labels'` column.

```
In [24]: df.sort_values('labels')
```

```
Out[24]:
```

	companies	labels
41	Philip Morris	0
40	Procter Gamble	0
25	Johnson & Johnson	0
27	Kimberly-Clark	0
9	Colgate-Palmolive	0
24	Intel	1
22	HP	1
33	Microsoft	1
51	Texas instruments	1
50	Taiwan Semiconductor Manufacturing	1
11	Cisco	1
20	Home Depot	2
58	Xerox	2
30	MasterCard	2
31	McDonalds	2
13	DuPont de Nemours	2
23	IBM	2
32	3M	2
47	Symantec	2
14	Dell	2
17	Google/Alphabet	3
0	Apple	3
48	Toyota	4
34	Mitsubishi	4
45	Sony	4
21	Honda	4
7	Canon	4
43	SAP	5
19	GlaxoSmithKline	5
39	Pfizer	5
49	Total	5
6	British American Tobacco	5
42	Royal Dutch Shell	5
52	Unilever	5
46	Sanofi-Aventis	5
37	Novartis	5
1	AIG	6
55	Wells Fargo	7
26	JPMorgan Chase	7
15	Ford	7
16	General Electrics	7

EXERCISE 9: WHICH STOCKS MOVE TOGETHER?

18	Goldman Sachs	7
3	American express	7
5	Bank of America	7
29	Lookheed Martin	8
36	Northrop Grumman	8
4	Boeing	8
35	Navistar	9
12	Chevron	9
10	ConocoPhillips	9
8	Caterpillar	9
53	Valero Energy	9
57	Exxon	9
44	Schlumberger	9
38	Pepsi	10
28	Coca Cola	10
54	Walgreen	11
2	Amazon	12
59	Yahoo	12
56	Wal-Mart	13

Great job! Using k-means, you've discovered which companies stock prices move together on the stock exchange.

In []: