

輕鬆學習 Python I 從基礎到應用，成為初級 Python 資料分析師

pandas：處理表格式資料的 Python 模組

郭耀仁

課程綱要

- pandas 解決了什麼問題
- pandas 基礎
- pandas 技巧

pandas 解決了什麼問題

實作練習：將開曼群島的上市公司股價挑選出來

<https://tw.stock.yahoo.com/d/i/rank.php?t=pri&e=tse&n=100>
(<https://tw.stock.yahoo.com/d/i/rank.php?t=pri&e=tse&n=100>).

Python 一直以來都非常適合資料處理，但她的分析能力很薄弱，pandas 的開發有助於補足 Python 資料分析的需求，讓使用者能夠在 Python 中執行完整的資料分析流程，而無需切換到 data-centric 的特定語言，如 R。

pandas 提供了新的資料結構

- Series 是具備索引的 ndarray
- DataFrame 能完美處理表格式資料 (tabular data)
- Panel 能為 DataFrame 加入第三個維度 (通常為時間)

pandas 主要的應用場景

- 表格式資料的讀取
- 豐富的資料清理與分析函數
- 視覺化：包裝了常用的 matplotlib.pyplot 圖形

pandas 基礎

什麼是 pandas?

Flexible and powerful data analysis / manipulation library for Python, providing labeled data structures similar to R data.frame objects, statistical functions, and much more.

pandas 的命名源自她的三個資料結構

- Panel
- DataFrame
- Series

創建 Series

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

numbers = [9, 23, 33, 91, 13]
players = ["Ron Harper", "Michael Jordan", "Scottie Pippen", "Dennis Rodman", "Luc Longley"]
ser = pd.Series(players)
ser
```

```
Out[1]: 0      Ron Harper
1      Michael Jordan
2      Scottie Pippen
3      Dennis Rodman
4      Luc Longley
dtype: object
```

```
In [2]: ser.index = numbers  
ser
```

```
Out[2]: 9          Ron Harper  
23      Michael Jordan  
33      Scottie Pippen  
91      Dennis Rodman  
13      Luc Longley  
dtype: object
```

創建 DataFrame

```
In [3]: df = pd.DataFrame()  
df["number"] = numbers  
df["player"] = players  
df
```

Out[3]:

	number	player
0	9	Ron Harper
1	23	Michael Jordan
2	33	Scottie Pippen
3	91	Dennis Rodman
4	13	Luc Longley

以 `df.<TAB>` 觀察屬性與方法

In [4]: `df.head()`

Out[4]:

	number	player
0	9	Ron Harper
1	23	Michael Jordan
2	33	Scottie Pippen
3	91	Dennis Rodman
4	13	Luc Longley

```
In [5]: df.tail(3)
```

```
Out[5]:
```

	number	player
2	33	Scottie Pippen
3	91	Dennis Rodman
4	13	Luc Longley

```
In [6]: df.index
```

```
Out[6]: RangeIndex(start=0, stop=5, step=1)
```



```
In [7]: df.describe()
```

```
Out[7]:
```

	number
count	5.000000
mean	33.800000
std	33.304654
min	9.000000
25%	13.000000
50%	23.000000
75%	33.000000
max	91.000000

選擇變數 Select

```
In [8]: df["player"]
```

```
Out[8]: 0      Ron Harper  
1  Michael Jordan  
2  Scottie Pippen  
3   Dennis Rodman  
4      Luc Longley  
Name: player, dtype: object
```

篩選觀測值 Filter

- `df.loc[]` 以標籤為準
- `df.iloc[]` 以位置為準
- 以布林陣列篩選

```
In [9]: df.index = ["PG", "SG", "SF", "PF", "C"]  
df.loc[["SG", "SF", "PF"], ["number", "player"]]
```

Out[9]:

	number	player
SG	23	Michael Jordan
SF	33	Scottie Pippen
PF	91	Dennis Rodman

```
In [10]: df.iloc[[1, 2, 3], [0, 1]]
```

Out[10]:

	number	player
SG	23	Michael Jordan
SF	33	Scottie Pippen
PF	91	Dennis Rodman

```
In [11]: is_trio = df["number"].isin([23, 33, 91])
print(is_trio)
df[is_trio]
```

```
PG    False
SG     True
SF     True
PF     True
C     False
Name: number, dtype: bool
```

Out[11]:

	number	player
SG	23	Michael Jordan
SF	33	Scottie Pippen
PF	91	Dennis Rodman

排序觀測值 Arrange

- `df.sort_index()`
- `df.sort_values()`

```
In [12]: df.sort_index()
```

Out[12]:

	number	player
C	13	Luc Longley
PF	91	Dennis Rodman
PG	9	Ron Harper
SF	33	Scottie Pippen
SG	23	Michael Jordan

```
In [13]: df.sort_values(by="number")
```

Out[13]:

	number	player
PG	9	Ron Harper
C	13	Luc Longley
SG	23	Michael Jordan
SF	33	Scottie Pippen
PF	91	Dennis Rodman

pandas 技巧

不那麼基礎的 pandas 觀念

- 資料的輸入輸出
- Apply
- 摘要
- 分組
- 遺漏值處理
- 轉置
- 合併

資料的輸入輸出

- 關聯式資料庫表格
- .csv 檔案
- array of JSON 的 .json 檔案
- Excel 試算表 .xls 與 .xlsx

```
In [14]: import requests

aqi_url = "https://opendata.epa.gov.tw/ws/Data/AQI/?$format=json"
r = requests.get(aqi_url, verify=False)
aqi = pd.read_json(r.text)
aqi.head()
```

```
/Users/kuoyaojen/anaconda3/lib/python3.7/site-packages/urllib3/connectionpool.py:847: InsecureRequestWarning: Unverified HTTPS request is being made. Adding
certificate verification is strongly advised. See: https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings
InsecureRequestWarning)
```

Out[14]:

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SO ₂
0				屏東縣	22.352222	120.377222					...	11			2019-04-02 14:00	
1	58	0.36	0.4	苗栗縣	24.616369	120.786028	0.9	11	11	65	...	30	18	細懸浮微粒	2019-04-02 14:00	3.4
2	53	0.21	0.2	彰化縣	23.843139	120.273117	0.5	3.8	4.3	77	...	14	16	細懸浮微粒	2019-04-02 14:00	2.6
3	61	0.13	0.1	臺南市	23.265066	120.124167	0.4	5.3	5.8	74	...	18	20	細懸浮微粒	2019-04-02 14:00	2.7
4	50	0.17	0.2	新北市	25.298562	121.536763	0.1	1.8	2	52	...	8	12		2019-04-02 14:00	0.3

5 rows × 23 columns

Apply

將函數映射至資料框

```
In [15]: import numpy as np

def get_na(x):
    """getting np.nan from "" and "ND" """
    if x == "" or x == "ND":
        return np.nan
    else:
        return x
```

```
In [16]: aqi["PM2.5"] = aqi["PM2.5"].apply(get_na)
aqi["AQI"] = aqi["AQI"].apply(get_na)
aqi.head()
```

Out[16]:

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SO
0	NaN			屏東縣	22.352222	120.377222					...	11			2019-04-02 14:00	
1	58	0.36	0.4	苗栗縣	24.616369	120.786028	0.9	11	11	65	...	30	18	細懸浮微粒	2019-04-02 14:00	3.4
2	53	0.21	0.2	彰化縣	23.843139	120.273117	0.5	3.8	4.3	77	...	14	16	細懸浮微粒	2019-04-02 14:00	2.6
3	61	0.13	0.1	臺南市	23.265066	120.124167	0.4	5.3	5.8	74	...	18	20	細懸浮微粒	2019-04-02 14:00	2.7
4	50	0.17	0.2	新北市	25.298562	121.536763	0.1	1.8	2	52	...	8	12		2019-04-02 14:00	0.3

5 rows × 23 columns

遺漏值處理

- pandas 主要以 `np.nan` 紀錄遺漏值
- `df.isna()`
- `df.dropna()`
- `df.fillna()`

```
In [17]: aqi["PM2.5"].isna()
```

```
Out[17]: 0      False
          1      False
          2      False
          3      False
          4      False
          5      False
          6      False
          7       True
          8      False
          9      False
         10      False
         11      False
         12      False
         13      False
         14      False
         15      False
         16      False
         17      False
         18      False
         19       True
         20      False
         21      False
         22      False
         23      False
         24      False
         25      False
         26      False
         27      False
         28      False
         29      False
          ...
         51      False
         52      False
         53      False
```

```
54    False
55    False
56    False
57    False
58    False
59    False
60    False
61    False
62    False
63    False
64    False
65    False
66    False
67    False
68    False
69    False
70    False
71     True
72     True
73    False
74    False
75    False
76    False
77     True
78    False
79    False
80    False
Name: PM2.5, Length: 81, dtype: bool
```

```
In [18]: # how many missing values are there  
aqi["PM2.5"].isna().sum()
```

```
Out[18]: 6
```



```
In [19]: aqi.dropna()
```

Out[19]:

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
1	58	0.36	0.4	苗栗縣	24.616369	120.786028	0.9	11	11	65	...	30	18	細懸浮微粒	2019-04-02 14:00	3.4
2	53	0.21	0.2	彰化縣	23.843139	120.273117	0.5	3.8	4.3	77	...	14	16	細懸浮微粒	2019-04-02 14:00	2.6
3	61	0.13	0.1	臺南市	23.265066	120.124167	0.4	5.3	5.8	74	...	18	20	細懸浮微粒	2019-04-02 14:00	2.7
4	50	0.17	0.2	新北市	25.298562	121.536763	0.1	1.8	2	52	...	8	12		2019-04-02 14:00	0.9
5	74	0.3	0.4	雲林縣	23.753506	120.251825	1.7	7.4	9.1		...	19	25	細懸浮微粒	2019-04-02 14:00	3.4
6	32	-		臺東縣	23.045083	121.161933	0.5	5	5.4	42	...	5	5		2019-04-02 14:00	2.2
8	67	0.31	0.3	金門縣	24.432133	118.312256	0.7	10	11	69	...	24	22	細懸浮微粒	2019-04-02 14:00	3.4
9	63	0.29	0.3	連江縣	26.160469	119.949875	0.3	7.2	7.5	63	...	20	21	細懸浮微粒	2019-04-02 14:00	3.8
10	68	0.39	0.4	南投縣	23.968842	120.967903	0.9	9	9.8	53	...	28	22	細懸浮微粒	2019-04-02 14:00	1.9
11	70	0.42	0.7	高雄市	22.608711	120.312017	2.7	18	21	53	...	12	23	細懸浮微粒	2019-04-02 14:00	1.9
12	45	0.66	0.8	新北市	25.017000	121.516306	10	29	40	28	...	14	14		2019-04-02 14:00	2
13	117	0.37	0.4	南投縣	23.756389	120.677306	0.1	9.2	9.4	78	...	45	42	細懸浮微粒	2019-04-02 14:00	2.5
14	55	1.55	2.0	桃園市	24.953278	121.221667	33	44	77	20	...	12	17	細懸浮微粒	2019-04-02 14:00	3
15	42	1.43	1.5	新北市	25.072611	121.493806	65	45	109	-	...	10	13		2019-04-02 14:00	
16	31	0.31	0.3	宜蘭縣	24.632203	121.792928	1.6	7.3	8.9	50	...	14	10		2019-04-02 14:00	1.8

[illegible]

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
47	80	0.3	0.5	彰化縣	24.131672	120.469061	1.5	6.2	7.7	73	...	22	27	細懸浮微粒	2019-04-02 14:00	
48	100	0.42	0.9	彰化縣	24.066000	120.541519	0.4	12	12	69	...	37	35	細懸浮微粒	2019-04-02 14:00	3.5
49	91	0.53	0.7	臺中市	24.162197	120.616917	1.2	19	20	70	...	35	32	細懸浮微粒	2019-04-02 14:00	5.4
50	98	0.63	0.8	臺中市	24.151958	120.641092	3.4	22	26	69	...	45	35	細懸浮微粒	2019-04-02 14:00	4.9
51	91	0.54	0.8	臺中市	24.099611	120.677689	1.1	15	16	71	...	35	32	細懸浮微粒	2019-04-02 14:00	4.5
52	73	0.5	0.7	臺中市	24.225628	120.568794	2.4	20	22	63	...	38	24	細懸浮微粒	2019-04-02 14:00	6.5
53	84	0.63	0.6	臺中市	24.256586	120.741711	3.5	17	20	74	...	46	29	細懸浮微粒	2019-04-02 14:00	2.5
54	95	0.43	0.4	苗栗縣	24.382942	120.758833	2.5	17	20	54	...	41	33	細懸浮微粒	2019-04-02 14:00	4.5
55	74	0.47	0.6	苗栗縣	24.565269	120.820200	3.1	15	18	42	...	29	25	細懸浮微粒	2019-04-02 14:00	3.7
56	46	0.34	0.5	苗栗縣	24.696969	120.898572	1.8	11	12	45	...	17	14		2019-04-02 14:00	1.5
57	63	0.54	0.5	新竹市	24.805619	120.972075	2.3	17	19	50	...	30	20	細懸浮微粒	2019-04-02 14:00	3.9
58	42	0.52	0.5	新竹縣	24.740644	121.088903	1.9	21	23	18	...	19	13		2019-04-02 14:00	3.6
59	56	0.46	0.5	新竹縣	24.900142	121.038653	1.8	15	17	60	...	23	17	細懸浮微粒	2019-04-02 14:00	4.9
60	65	0.5	0.6	桃園市	24.863869	121.216350	4	25	29	39	...	19	21	細懸浮微粒	2019-04-02 14:00	3.8
61	44	0.48	0.5	桃園市	24.952786	121.203986	3.1	20	23	44	...	13	13		2019-04-02 14:00	3.5
62	58	0.44	0.4	桃園市	25.035503	121.082761	2.7	14	17	58	...	24	18	細懸浮微粒	2019-04-02 14:00	3.6
63	50		0.4	桃園市	25.060344	121.201811					...	12	15		2019-04-02 14:00	
64	55	0.56	0.5	桃園市	24.995368	121.304383	3.4	21	25	32	...	16	17	細懸浮微粒	2019-04-02 14:00	4.9

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
65	43	0.79	1.0	臺北市	25.063200	121.513311	24	39	63	-	...	8	13		2019-04-02 14:00	2.2
67	50	0.53	0.6	臺北市	25.020608	121.529556	5.7	26	31	30	...	11	15		2019-04-02 14:00	2.2
68	35	0.5	0.5	臺北市	25.046503	121.507972	6.3	28	35	29	...	6	11		2019-04-02 14:00	2.2
69	47	0.66	0.7	臺北市	25.062361	121.526528	9.3	32	41	28	...	8	15		2019-04-02 14:00	4.1
70	54		0.3	臺北市	25.105417	121.515389				55	...	7	13	臭氧八小時	2019-04-02 14:00	
73	52	0.4	0.5	新北市	25.068950	121.481028	2.4	15	17	42	...	10	16	細懸浮微粒	2019-04-02 14:00	1
74	40	0.45	0.5	新北市	25.037972	121.432500	4.2	18	22	43	...	9	11		2019-04-02 14:00	1.9
75	50	0.58	0.6	新北市	25.012972	121.458667	4.8	25	30	33	...	14	15		2019-04-02 14:00	1.9
76	56	0.54	0.6	新北市	24.982528	121.451861	3.4	22	25	40	...	15	18	細懸浮微粒	2019-04-02 14:00	1.9
78	43	0.21	0.3	新北市	25.179667	121.689881	1.1	5.4	6.5	56	...	8	12		2019-04-02 14:00	0.8
79	60	0.37	0.4	新北市	25.067131	121.642300	4.1	17	21	34	...	16	19	細懸浮微粒	2019-04-02 14:00	1.7
80	37	0.28	0.6	基隆市	25.129167	121.760056	1.6	6.2	7.7	50	...	6	11		2019-04-02 14:00	1.8

73 rows × 23 columns



```
In [20]: filled_val = 999
        aqi.fillna(filled_val)
```

Out[20]:

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
0	999			屏東縣	22.352222	120.377222					...	11			2019-04-02 14:00	
1	58	0.36	0.4	苗栗縣	24.616369	120.786028	0.9	11	11	65	...	30	18	細懸浮微粒	2019-04-02 14:00	3.4
2	53	0.21	0.2	彰化縣	23.843139	120.273117	0.5	3.8	4.3	77	...	14	16	細懸浮微粒	2019-04-02 14:00	2.6
3	61	0.13	0.1	臺南市	23.265066	120.124167	0.4	5.3	5.8	74	...	18	20	細懸浮微粒	2019-04-02 14:00	2.7
4	50	0.17	0.2	新北市	25.298562	121.536763	0.1	1.8	2	52	...	8	12		2019-04-02 14:00	0.9
5	74	0.3	0.4	雲林縣	23.753506	120.251825	1.7	7.4	9.1		...	19	25	細懸浮微粒	2019-04-02 14:00	3.4
6	32	-		臺東縣	23.045083	121.161933	0.5	5	5.4	42	...	5	5		2019-04-02 14:00	2.2
7	62			澎湖縣	23.569031	119.566158					...	999	20	細懸浮微粒	2019-04-02 14:00	
8	67	0.31	0.3	金門縣	24.432133	118.312256	0.7	10	11	69	...	24	22	細懸浮微粒	2019-04-02 14:00	3.4
9	63	0.29	0.3	連江縣	26.160469	119.949875	0.3	7.2	7.5	63	...	20	21	細懸浮微粒	2019-04-02 14:00	3.8
10	68	0.39	0.4	南投縣	23.968842	120.967903	0.9	9	9.8	53	...	28	22	細懸浮微粒	2019-04-02 14:00	1.9
11	70	0.42	0.7	高雄市	22.608711	120.312017	2.7	18	21	53	...	12	23	細懸浮微粒	2019-04-02 14:00	1.7
12	45	0.66	0.8	新北市	25.017000	121.516306	10	29	40	28	...	14	14		2019-04-02 14:00	2
13	117	0.37	0.4	南投縣	23.756389	120.677306	0.1	9.2	9.4	78	...	45	42	細懸浮微粒	2019-04-02 14:00	2.5
14	55	1.55	2.0	桃園市	24.953278	121.221667	33	44	77	20	...	12	17	細懸浮微粒	2019-04-02 14:00	3

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
15	42	1.43	1.5	新北市	25.072611	121.493806	65	45	109	-	...	10	13		2019-04-02 14:00	
16	31	0.31	0.3	宜蘭縣	24.632203	121.792928	1.6	7.3	8.9	50	...	14	10		2019-04-02 14:00	1.8
17	48	0.32	0.4	宜蘭縣	24.747917	121.746394	1	5.9	6.9	55	...	10	15		2019-04-02 14:00	2.5
18	49	0.22	0.2	臺北市	25.182722	121.529583	0.2	0.4	0.6	50	...	4	10		2019-04-02 14:00	1.5
19	32	0.21	0.3	花蓮縣	23.971306	121.599769	0.6	3.2	3.8	66	...	999	2		2019-04-02 14:00	1.4
20	31	0.29	0.3	臺東縣	22.755358	121.150450	2.2	3.5	5.6	35	...	2	7		2019-04-02 14:00	0.9
21	61	0.15	0.2	屏東縣	21.958069	120.788928	0.5	1.5	2	57	...	6	7	臭氧八小時	2019-04-02 14:00	1.1
22	110	0.44	0.5	屏東縣	22.523108	120.561175	0.7	12	12	87	...	39	39	細懸浮微粒	2019-04-02 14:00	3.6
23	82	0.43	0.5	屏東縣	22.673081	120.488033	0.6	12	13	79	...	19	28	細懸浮微粒	2019-04-02 14:00	3
24	66	0.3	0.5	高雄市	22.565833	120.337736	4	15	19	57	...	12	22	細懸浮微粒	2019-04-02 14:00	2.4
25	67	0.33	0.5	高雄市	22.605386	120.307564	3.6	16	20	57	...	12	22	細懸浮微粒	2019-04-02 14:00	1.4
26	100	0.36	0.5	高雄市	22.632567	120.288086	1.8	5.9	7.7	78	...	24	35	細懸浮微粒	2019-04-02 14:00	2
27	64	0.28	0.5	高雄市	22.674861	120.292917	1.4	6.6	8	68	...	12	21	細懸浮微粒	2019-04-02 14:00	2.2
28	88	0.37	0.5	高雄市	22.733667	120.328289	1.8	8.4	10	73	...	6	31	細懸浮微粒	2019-04-02 14:00	2.6
29	77	0.27	0.5	高雄市	22.479500	120.411750	0.7	4.6	5.3	70	...	13	26	細懸浮微粒	2019-04-02 14:00	2.3
...
51	91	0.54	0.8	臺中市	24.099611	120.677689	1.1	15	16	71	...	35	32	細懸浮微粒	2019-04-02 14:00	4.5
52	73	0.5	0.7	臺中市	24.225628	120.568794	2.4	20	22	63	...	38	24	細懸浮微粒	2019-04-02 14:00	6.5

	AQI		CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
53	84	0.63	0.6		臺中市	24.256586	120.741711	3.5	17	20	74	...	46	29	細懸浮微粒	2019-04-02 14:00	2.5
54	95	0.43	0.4		苗栗縣	24.382942	120.758833	2.5	17	20	54	...	41	33	細懸浮微粒	2019-04-02 14:00	4.5
55	74	0.47	0.6		苗栗縣	24.565269	120.820200	3.1	15	18	42	...	29	25	細懸浮微粒	2019-04-02 14:00	3.7
56	46	0.34	0.5		苗栗縣	24.696969	120.898572	1.8	11	12	45	...	17	14		2019-04-02 14:00	1.5
57	63	0.54	0.5		新竹市	24.805619	120.972075	2.3	17	19	50	...	30	20	細懸浮微粒	2019-04-02 14:00	3.9
58	42	0.52	0.5		新竹縣	24.740644	121.088903	1.9	21	23	18	...	19	13		2019-04-02 14:00	3.6
59	56	0.46	0.5		新竹縣	24.900142	121.038653	1.8	15	17	60	...	23	17	細懸浮微粒	2019-04-02 14:00	4.9
60	65	0.5	0.6		桃園市	24.863869	121.216350	4	25	29	39	...	19	21	細懸浮微粒	2019-04-02 14:00	3.8
61	44	0.48	0.5		桃園市	24.952786	121.203986	3.1	20	23	44	...	13	13		2019-04-02 14:00	3.5
62	58	0.44	0.4		桃園市	25.035503	121.082761	2.7	14	17	58	...	24	18	細懸浮微粒	2019-04-02 14:00	3.6
63	50		0.4		桃園市	25.060344	121.201811					...	12	15		2019-04-02 14:00	
64	55	0.56	0.5		桃園市	24.995368	121.304383	3.4	21	25	32	...	16	17	細懸浮微粒	2019-04-02 14:00	4.9
65	43	0.79	1.0		臺北市	25.063200	121.513311	24	39	63	-	...	8	13		2019-04-02 14:00	2.2
66	999	0.38	0.5		臺北市	25.050000	121.578611	4.9	20	25	33	...	9			2019-04-02 14:00	2.5
67	50	0.53	0.6		臺北市	25.020608	121.529556	5.7	26	31	30	...	11	15		2019-04-02 14:00	2.2
68	35	0.5	0.5		臺北市	25.046503	121.507972	6.3	28	35	29	...	6	11		2019-04-02 14:00	2.1
69	47	0.66	0.7		臺北市	25.062361	121.526528	9.3	32	41	28	...	8	15		2019-04-02 14:00	4.1
70	54		0.3		臺北市	25.105417	121.515389				55	...	7	13	臭氧八小時	2019-04-02 14:00	

	AQI	CO	CO_8hr	County	Latitude	Longitude	NO	NO2	NOx	O3	...	PM2.5	PM2.5_AVG	Pollutant	PublishTime	SC
71	71		0.2	新北市	25.164500	121.449239					...	999	11	臭氧八小時	2019-04-02 14:00	
72	44		0.3	新北市	25.077197	121.376869					...	999	12		2019-04-02 14:00	
73	52	0.4	0.5	新北市	25.068950	121.481028	2.4	15	17	42	...	10	16	細懸浮微粒	2019-04-02 14:00	1
74	40	0.45	0.5	新北市	25.037972	121.432500	4.2	18	22	43	...	9	11		2019-04-02 14:00	1.9
75	50	0.58	0.6	新北市	25.012972	121.458667	4.8	25	30	33	...	14	15		2019-04-02 14:00	1.9
76	56	0.54	0.6	新北市	24.982528	121.451861	3.4	22	25	40	...	15	18	細懸浮微粒	2019-04-02 14:00	1.9
77	51		0.5	新北市	24.977222	121.537778					...	999	16	細懸浮微粒	2019-04-02 14:00	
78	43	0.21	0.3	新北市	25.179667	121.689881	1.1	5.4	6.5	56	...	8	12		2019-04-02 14:00	0.8
79	60	0.37	0.4	新北市	25.067131	121.642300	4.1	17	21	34	...	16	19	細懸浮微粒	2019-04-02 14:00	1.7
80	37	0.28	0.6	基隆市	25.129167	121.760056	1.6	6.2	7.7	50	...	6	11		2019-04-02 14:00	1.8

81 rows × 23 columns



摘要

```
In [21]: aqi["County"].unique()
```

```
Out[21]: array(['屏東縣', '苗栗縣', '彰化縣', '臺南市', '新北市', '雲林縣', '臺東縣', '澎湖縣',  
              '金門縣',  
              '連江縣', '南投縣', '高雄市', '桃園市', '宜蘭縣', '臺北市', '花蓮縣', '嘉義市',  
              '嘉義縣',  
              '臺中市', '新竹市', '新竹縣', '基隆市'], dtype=object)
```

```
In [22]: aqi["AQI"] = aqi["AQI"].astype(float)  
aqi["PM2.5"] = aqi["PM2.5"].astype(float)  
print(aqi["AQI"].max())  
print(aqi["PM2.5"].min())
```

```
127.0  
2.0
```

分組

```
In [23]: aqi.groupby("County").mean()["PM2.5"]
```

```
Out[23]: County
南投縣      41.333333
嘉義市      31.000000
嘉義縣      30.500000
基隆市       6.000000
宜蘭縣      12.000000
屏東縣      18.750000
彰化縣      24.000000
新北市      11.555556
新竹市      30.000000
新竹縣      21.000000
桃園市      16.000000
澎湖縣           NaN
臺中市      39.800000
臺北市       7.571429
臺南市      15.000000
臺東縣       3.500000
花蓮縣           NaN
苗栗縣      29.250000
連江縣      20.000000
金門縣      24.000000
雲林縣      28.666667
高雄市      17.916667
Name: PM2.5, dtype: float64
```

轉置

- `df.stack()`: 寬轉長格式
- `s.unstack()`: 長轉寬格式

```
In [24]: wide_df = aqi[["SiteName", "AQI", "PM2.5"]]  
wide_df = wide_df.set_index("SiteName", drop=True)  
wide_df.head()
```

Out[24]:

	AQI	PM2.5
SiteName		
屏東(琉球)	NaN	11.0
苗栗(後龍)	58.0	30.0
彰化(大城)	53.0	14.0
臺南(北門)	61.0	18.0
富貴角	50.0	8.0

```
In [25]: wide_df.stack()
```

```
Out[25]: SiteName
屏東(琉球)    PM2.5    11.0
苗栗(後龍)    AQI      58.0
              PM2.5    30.0
彰化(大城)    AQI      53.0
              PM2.5    14.0
臺南(北門)    AQI      61.0
              PM2.5    18.0
富貴角        AQI      50.0
              PM2.5     8.0
麥寮          AQI      74.0
              PM2.5    19.0
關山          AQI      32.0
              PM2.5     5.0
馬公          AQI      62.0
金門          AQI      67.0
              PM2.5    24.0
馬祖          AQI      63.0
              PM2.5    20.0
埔里          AQI      68.0
              PM2.5    28.0
復興          AQI      70.0
              PM2.5    12.0
永和          AQI      45.0
              PM2.5    14.0
竹山          AQI     117.0
              PM2.5    45.0
中壢          AQI      55.0
              PM2.5    12.0
三重          AQI      42.0
              PM2.5    10.0
              ...
桃園          AQI      55.0
              PM2.5    16.0
```

大同	AQI	43.0
	PM2.5	8.0
松山	PM2.5	9.0
古亭	AQI	50.0
	PM2.5	11.0
萬華	AQI	35.0
	PM2.5	6.0
中山	AQI	47.0
	PM2.5	8.0
士林	AQI	54.0
	PM2.5	7.0
淡水	AQI	71.0
林口	AQI	44.0
菜寮	AQI	52.0
	PM2.5	10.0
新莊	AQI	40.0
	PM2.5	9.0
板橋	AQI	50.0
	PM2.5	14.0
土城	AQI	56.0
	PM2.5	15.0
新店	AQI	51.0
萬里	AQI	43.0
	PM2.5	8.0
汐止	AQI	60.0
	PM2.5	16.0
基隆	AQI	37.0
	PM2.5	6.0

Length: 154, dtype: float64

```
In [26]: wide_df.stack().unstack()
```

Out[26]:

	AQI	PM2.5
SiteName		
屏東(琉球)	NaN	11.0
苗栗(後龍)	58.0	30.0
彰化(大城)	53.0	14.0
臺南(北門)	61.0	18.0
富貴角	50.0	8.0
麥寮	74.0	19.0
關山	32.0	5.0
馬公	62.0	NaN
金門	67.0	24.0
馬祖	63.0	20.0
埔里	68.0	28.0
復興	70.0	12.0
永和	45.0	14.0
竹山	117.0	45.0
中壢	55.0	12.0
三重	42.0	10.0
冬山	31.0	14.0
宜蘭	48.0	10.0
陽明	49.0	4.0
花蓮	32.0	NaN
臺東	31.0	2.0
恆春	61.0	6.0
潮州	110.0	39.0
屏東	82.0	19.0
小港	66.0	12.0
前鎮	67.0	12.0
前金	100.0	24.0

AQI PM2.5		
SiteName		
左營	64.0	12.0
楠梓	88.0	6.0
林園	77.0	13.0
...
大里	91.0	35.0
沙鹿	73.0	38.0
豐原	84.0	46.0
三義	95.0	41.0
苗栗	74.0	29.0
頭份	46.0	17.0
新竹	63.0	30.0
竹東	42.0	19.0
湖口	56.0	23.0
龍潭	65.0	19.0
平鎮	44.0	13.0
觀音	58.0	24.0
大園	50.0	12.0
桃園	55.0	16.0
大同	43.0	8.0
松山	NaN	9.0
古亭	50.0	11.0
萬華	35.0	6.0
中山	47.0	8.0
士林	54.0	7.0
淡水	71.0	NaN
林口	44.0	NaN
菜寮	52.0	10.0
新莊	40.0	9.0
板橋	50.0	14.0
土城	56.0	15.0
新店	51.0	NaN

	AQI	PM2.5
SiteName		
萬里	43.0	8.0
汐止	60.0	16.0
基隆	37.0	6.0

81 rows × 2 columns

```
In [27]: long_df = pd.DataFrame(wide_df.stack())
long_df = long_df.reset_index()
long_df.columns = ["SiteName", "AirQualityIndex", "Value"]
long_df.head()
```

Out[27]:

	SiteName	AirQualityIndex	Value
0	屏東(琉球)	PM2.5	11.0
1	苗栗(後龍)	AQI	58.0
2	苗栗(後龍)	PM2.5	30.0
3	彰化(大城)	AQI	53.0
4	彰化(大城)	PM2.5	14.0

合併

利用 `pd.merge()` 進行 SQL 風格的合併。

```
In [28]: left = long_df.copy()
right = aqi[["SiteName", "Latitude", "Longitude"]]
pd.merge(left, right, on="SiteName")
```

Out[28]:

	SiteName	AirQualityIndex	Value	Latitude	Longitude
0	屏東(琉球)	PM2.5	11.0	22.352222	120.377222
1	苗栗(後龍)	AQI	58.0	24.616369	120.786028
2	苗栗(後龍)	PM2.5	30.0	24.616369	120.786028
3	彰化(大城)	AQI	53.0	23.843139	120.273117
4	彰化(大城)	PM2.5	14.0	23.843139	120.273117
5	臺南(北門)	AQI	61.0	23.265066	120.124167
6	臺南(北門)	PM2.5	18.0	23.265066	120.124167
7	富貴角	AQI	50.0	25.298562	121.536763
8	富貴角	PM2.5	8.0	25.298562	121.536763
9	麥寮	AQI	74.0	23.753506	120.251825
10	麥寮	PM2.5	19.0	23.753506	120.251825
11	關山	AQI	32.0	23.045083	121.161933
12	關山	PM2.5	5.0	23.045083	121.161933
13	馬公	AQI	62.0	23.569031	119.566158
14	金門	AQI	67.0	24.432133	118.312256
15	金門	PM2.5	24.0	24.432133	118.312256
16	馬祖	AQI	63.0	26.160469	119.949875
17	馬祖	PM2.5	20.0	26.160469	119.949875
18	埔里	AQI	68.0	23.968842	120.967903
19	埔里	PM2.5	28.0	23.968842	120.967903
20	復興	AQI	70.0	22.608711	120.312017
21	復興	PM2.5	12.0	22.608711	120.312017
22	永和	AQI	45.0	25.017000	121.516306
23	永和	PM2.5	14.0	25.017000	121.516306
24	竹山	AQI	117.0	23.756389	120.677306
25	竹山	PM2.5	45.0	23.756389	120.677306

	SiteName	AirQualityIndex	Value	Latitude	Longitude
26	中壢	AQI	55.0	24.953278	121.221667
27	中壢	PM2.5	12.0	24.953278	121.221667
28	三重	AQI	42.0	25.072611	121.493806
29	三重	PM2.5	10.0	25.072611	121.493806
...
124	桃園	AQI	55.0	24.995368	121.304383
125	桃園	PM2.5	16.0	24.995368	121.304383
126	大同	AQI	43.0	25.063200	121.513311
127	大同	PM2.5	8.0	25.063200	121.513311
128	松山	PM2.5	9.0	25.050000	121.578611
129	古亭	AQI	50.0	25.020608	121.529556
130	古亭	PM2.5	11.0	25.020608	121.529556
131	萬華	AQI	35.0	25.046503	121.507972
132	萬華	PM2.5	6.0	25.046503	121.507972
133	中山	AQI	47.0	25.062361	121.526528
134	中山	PM2.5	8.0	25.062361	121.526528
135	士林	AQI	54.0	25.105417	121.515389
136	士林	PM2.5	7.0	25.105417	121.515389
137	淡水	AQI	71.0	25.164500	121.449239
138	林口	AQI	44.0	25.077197	121.376869
139	菜寮	AQI	52.0	25.068950	121.481028
140	菜寮	PM2.5	10.0	25.068950	121.481028
141	新莊	AQI	40.0	25.037972	121.432500
142	新莊	PM2.5	9.0	25.037972	121.432500
143	板橋	AQI	50.0	25.012972	121.458667
144	板橋	PM2.5	14.0	25.012972	121.458667
145	土城	AQI	56.0	24.982528	121.451861
146	土城	PM2.5	15.0	24.982528	121.451861
147	新店	AQI	51.0	24.977222	121.537778
148	萬里	AQI	43.0	25.179667	121.689881
149	萬里	PM2.5	8.0	25.179667	121.689881
150	汐止	AQI	60.0	25.067131	121.642300

	SiteName	AirQualityIndex	Value	Latitude	Longitude
151	汐止	PM2.5	16.0	25.067131	121.642300
152	基隆	AQI	37.0	25.129167	121.760056
153	基隆	PM2.5	6.0	25.129167	121.760056

154 rows × 5 columns

延伸閱讀

- [pandas: powerful Python data analysis toolkit \(http://pandas.pydata.org/pandas-docs/stable/\)](http://pandas.pydata.org/pandas-docs/stable/).