

Problem-Based Learning

Python 3

Lecture: 09-Pandas

Online Python

- <https://www.programiz.com/python-programming/online-compiler/>

Lecture Notes

- <https://web.phy.ntnu.edu.tw/~hongyi/?url=notes>



串列

- 五個學生國英數成績如下，計算每個學生的平均成績

	國文	英文	數學
學生1	96	65	73
學生2	88	76	82
學生3	92	84	89
學生4	82	73	64
學生5	70	83	68

- 如何使用二維陣列(**Matrix**)同時存放文字(字串)和數字(整數或是浮點數)

Python list, Numpy array, Pandas DataFrame

	國文	英文	數學
學生1	96	65	73
學生2	88	76	82
學生3	92	84	89
學生4	82	73	64
學生5	70	83	68

✓ Python list:

```
[[96, 65, 73], [88, 76, 82], [92, 84, 89]]
```

✓ Numpy array:

```
[[96 65 73]
 [88 76 82]
 [92 84 89]]
```

✓ Pandas DataFrame:

	0	1	2 ← columns
index → 0	96	65	73
1	88	76	82
2	92	84	89
3	82	73	64
4	70	83	68

Numpy Array → Pandas DataFrame

- `import pandas as pd`
- `import numpy as np`
- `a = np.array(
 [[96, 65, 73],
 [88, 76, 82],
 [92, 84, 89],
 [82, 73, 64],
 [70, 83, 68]]
)`
- `d = pd.DataFrame(a)`

Python List → Pandas DataFrame

- `import pandas as pd`
- `a = [[96, 65, 73],
 [88, 76, 82],
 [92, 84, 89],
 [82, 73, 64],
 [70, 83, 68]]`
- `d = pd.DataFrame(a)`

Panda DataFrame

- `import pandas as pd`
- `import numpy as np`
- `a = [[96,65,73],
[88,76,82],
[92,84,89],
[82,73,64],
[70,83,68]]`

`# Python list to Pandas Dataframe`

- `d = pd.DataFrame(a)`
- `print(d)`

`# Python list to Numpy array`

- `b = np.array(a)`
- `print(b)`

`# Numpy array to Pandas Dataframe`

- `d = pd.DataFrame(b)`
- `print(d)`

Panda DataFrame : index & columns

- `d.index=["1", "2", "3", "4", "5"]`
- `d.columns=["國文", "英文", "數學"]`
- `print(d)`
- `print(d.shape)`
- `print(d.shape[0])`
- `print(d.shape[1])`

The diagram illustrates the structure of a DataFrame. The columns are labeled '國文', '英文', and '數學'. The rows are indexed from 1 to 5. Annotations include 'columns' pointing to the column headers, 'index' pointing to the row indices, 'column長度 shape(1)' in a box pointing to the number of columns, and 'index長度 shape(0)' in a box pointing to the number of rows.

	國文	英文	數學
1	96	65	73
2	88	76	82
3	92	84	89
4	82	73	64
5	70	83	68

Panda DataFrame : index & columns

- `import pandas as pd`
- `a = [[96,65,73],
[88,76,82],
[92,84,89],
[82,73,64],
[70,83,68]]`

另一種可行方式

- `ind = ["1", "2", "3", "4", "5"]`
 - `col = ["國文", "英文", "數學"]`
 - `d = pd.DataFrame(a, index=ind, columns=col)`
 - `print(d)`
- `ind = (x+1 for x in range(5))`

Panda DataFrame : 計算min, max, mean

axis=0

- `c = d['A'].min()`
- `c = d['A'].max()`
- `c = d['A'].mean()`

axis=1

- `c = d.mean(axis=1)`
- `print(c)`
- `print(c.index)`
- `print(c.values)`

	A	B	C	← axis=1
a	1	6	2	
b	2	7	3	
c	3	8	4	

↑
axis=0

a	3.0
b	4.0
c	5.0

↑ ↑
index values

Problem

- 計算各個學科平均分數
- 計算每一個學生的平均成績
- 找出每一個學生的最高分和最低分 and 其所對應的學科
- 提示：

`d.min(axis=?)`

`d.idxmin(axis=?)`

`d.max(axis=?)`

`d.idxmax(axis=?)`

`d.mean(axis=?)`