

7. Giới thiệu Panel

Panel được sử dụng không nhiều, nhưng vẫn quan trọng đối với dữ liệu 3 chiều. Các panel data được lấy từ kinh tế lượng và chịu trách nhiệm một phần về cái tên pandas: pan(el)-da(ta)-s. Tên của 3 trục này nhằm tạo ra một số ý nghĩa ngữ nghĩa để mô tả các hoạt động liên quan đến panel data, và đặc biệt là phân tích các bảng dữ liệu kinh tế lượng. Tên của 3 trục được dùng để đưa ra một số ý nghĩa ngữ nghĩa để mô tả các hoạt động liên quan đến panel data. chúng là –

Items – trục 0, mỗi item tương ứng với một dataframe chứa bên trong.

major_axis - trục 1, đó là chỉ mục (hàng) của mỗi dataframe.

minor_axis - trục 2, đó là các cột của mỗi dataframe.

Pandas cung cấp phương thức **pandas.Panel(data, items, major_axis, minor_axis, dtype, copy)** để có thể tạo được Panel từ nhiều kiểu dữ liệu khác nhau.

Trong đó:

‘data’ có thể nhận các kiểu dữ liệu sau: ndarray, series, map, lists, dict, hằng số và cả dataframe khác.

‘items’ là axis = 0.

‘major_axis’ là axis = 1.

‘minor_axis’ là axis = 2.

‘dtype’ là kiểu dữ liệu mỗi cột.

‘copy’ nhận giá trị True/False để khởi tạo dữ liệu có chia sẻ memory hay không.

1. Tạo một panel rỗng.

```
>>> import pandas as pd
>>> p = pd.Panel()
>>> print p
<class 'pandas.core.panel.Panel'>
Dimensions: 0 (items) x 0 (major_axis) x 0 (minor_axis)
Items axis: None
Major_axis axis: None
Minor_axis axis: None
>>>
```

2. Từ mảng 3D ndarrays.

```
>>> import pandas as pd
>>> import numpy as np
>>> data = np.random.rand(2,3,4)
>>> p = pd.Panel(data)
>>> print p
<class 'pandas.core.panel.Panel'>
Dimensions: 2 (items) x 3 (major_axis) x 4 (minor_axis)
Items axis: 0 to 1
Major_axis axis: 0 to 2
Minor_axis axis: 0 to 3
>>> p.to_frame()
      0      1
major minor
0  0  0.562182 0.288014
   1  0.729975 0.099355
   2  0.730366 0.438827
   3  0.175517 0.357904
1  0  0.992136 0.664511
   1  0.179336 0.241909
   2  0.572047 0.880752
   3  0.667717 0.977959
2  0  0.406752 0.934063
   1  0.623932 0.505132
   2  0.900456 0.007313
   3  0.219703 0.171776
>>>
```

3. Khởi tạo dữ dict của các dataframe

```
>>> import pandas as pd
>>> import numpy as np
>>> data = {'Item1': pd.DataFrame(np.random.randn(3, 2)), 'Item2': pd.DataFrame(np.random.randn(3, 1))}
>>> p = pd.Panel(data)
>>> print p
<class 'pandas.core.panel.Panel'>
Dimensions: 2 (items) x 3 (major_axis) x 2 (minor_axis)
Items axis: Item1 to Item2
Major_axis axis: 0 to 2
Minor_axis axis: 0 to 1
>>> p.to_frame()
      Item1      Item2
major minor
0  0 -0.575876 -0.049086
   1 -0.230539 -1.313769
   2  0.134019  1.081346
>>>
```

Ở đây giá trị **Major_axis** = max[row of item1, row of item2] và **Minor_axis** = max[column of item1, column of item2].

Tuy nhiên khi biểu diễn dưới dạng frame thì lấy giá trị min (Major_axis = min[row of item1, row of item2] và Minor_axis = min[column of item1, column of item2].)

Chọn (Selecting) dữ liệu từ panel. Dưới đây là ba cách biến đổi panel về dataframe.

Có thể chọn (selecting) dữ liệu từ panel ứng với mỗi item.

```
>>> import pandas as pd
>>> import numpy as np
>>> data = {'Item1': pd.DataFrame(np.random.randn(4, 3)),
...        'Item2': pd.DataFrame(np.random.randn(4, 2))}
>>> p = pd.Panel(data)
>>> print p['Item1']
      0      1      2
0  0.257905  1.714854 -0.097306
1  1.008341 -0.289634  0.110710
2  0.534798  0.901090 -0.203789
3 -1.536464 -0.498765 -0.489292
>>> print p['Item2']
      0      1      2
0 -0.719140  0.726088 NaN
1 -1.118916  1.394300 NaN
2  0.426197  1.543967 NaN
3 -0.378640  0.434465 NaN
>>>
```

Sử dụng `major_axis` để có được dữ liệu của panel ứng với trục `axis = 1`.

```
>>> import pandas as pd
>>> import numpy as np
>>> data = {'Item1': pd.DataFrame(np.random.randn(4, 3)),
...        'Item2': pd.DataFrame(np.random.randn(4, 2))}
>>> p = pd.Panel(data)
>>> print p.major_xs(1)
      Item1      Item2
0 -1.270247  1.510719
1 -0.220837  1.669597
2 -0.240013      NaN
>>>
```

Sử dụng `minor_axis` để có được dữ liệu của panel ứng với trục `axis = 2`.

```
>>> import pandas as pd
>>> import numpy as np
>>> data = {'Item1': pd.DataFrame(np.random.randn(4, 3)),
...        'Item2': pd.DataFrame(np.random.randn(4, 2))}
>>> p = pd.Panel(data)
>>> print p.minor_xs(1)
      Item1      Item2
0 -1.212101  0.256905
1  0.831389 -0.453344
2  0.246443  0.572504
3  1.612013  0.321815
>>>
```

Kết Luận

Trong bài này chúng tôi đã giới thiệu đến các bạn cấu trúc dữ liệu thứ 3 trong pandas là Panel, cách khởi tạo cũng như truy xuất dữ liệu từ một panel. Đây là cấu trúc mảng 3D, mặc dù ít được sử dụng nhưng nó có vai trò quan trọng trong việc phân tích các bảng dữ liệu kinh tế lượng.