參考資料

## D14 用 pandas 撰寫樞紐分析表





在前天學到畫圖,x 軸為索引(index),y 軸為欄位(column),此時 x 軸與 y 軸被索引以及欄位所限制,

## 想畫出各種不同的圖做分析,就必須用到欄位轉索引或是索引轉欄位。

索引轉欄位、欄位轉索引

以下表資料為例,欄位為 subject、type 組成,索引為 year、visit 組成 • 欄位轉索引:將一欄位(column)轉成一索引(index),使用.stack()即可,可以將 type 這個欄位 轉成了索引,所以索引變成了 year、visit、type。

• 注意 .stack() 會由最外層的欄位開始轉換,原欄位為 subject、type,會先由 type 轉換過去索

欄位轉索引

subject Bob Guido Sue visit year

引,如果在做一次才會把 subject 也轉換過去索引,如左表。

1.2 -1.7 0.3 0.4 0.8 0.2 2013 -1.3 -0.2 -1.9 -0.3 -0.7 -1.4 2014 0.7 -0.9 1.5 -0.5 -0.6 -0.1

0.6 -0.6 -0.2 0.2 0.3 -1.9

df.stack() subject Bob Guido Sue year visit 2013 1.2 0.3 0.8 -1.7 0.4 0.2 Temp -1.3 -1.9 -0.7 -0.2 -0.3 -1.4 Temp 0.7 1.5 -0.6 2014 1 -0.9 -0.5 -0.1 Temp

0.6 -0.2 0.3 Temp -0.6 0.2 -1.9 df.stack().stack() year visit type subject -0.4 -1.3 Guido -1.60.0 Temp Bob -0.7 Guido -1.4Sue -0.1 Bob 1.1 1.5 1.9 Temp Bob -1.02014 1 -3.3Bob 0.4 -0.1 Sue Temp Bob 0.7 Guido -1.0-1.0-0.7Bob 0.4 -0.0 1.3 Temp Bob -0.4 Guido 1.0 dtype: float64

索引轉欄位

0.7 -0.9 1.5 -0.5 -0.6 -0.1 0.6 -0.6 -0.2 0.2 0.3 -1.9

以下表資料為例,欄位為 subject、type 組成,索引為 year、visit 組成

引轉成了欄位,所以欄位變成了 subject、type 、visit。

subject Bob

visit

year

2013

df.unstack()

type

subject Bob

year

Guido

visit 1 2 1 2 1 2 1 2 1 2 1 2

• 索引轉欄位:將一索引(index)轉成一欄位(column),使用.unstack()即可,可以將 visit 這個索

• 注意與 .stack() 相同會由最外層開始轉換,原索引為 year、visit,會先由 visit 轉換過去欄位

Guido

Temp HR Temp HR Temp

1.2 -1.7 0.3 0.4 0.8 0.2

**2** -1.3 -0.2 -1.9 -0.3 -0.7 -1.4

Temp

Temp

Sue

**2014** -3.3 -0.7 0.7 1.3 0.4 0.4 -1.0 -0.4 -0.1 -0.0 -1.0 1.0 欄位名稱轉為欄位值 數據分析的時候經常要規<mark>寬數據變成長數據</mark>,有點像你們用 excel 做透視跟逆透視的過程。

例如下表格(紅框),要將欄位轉成欄位值,也就是說將 Name、Course、Age 轉成欄位值,如下

• value\_vars:需要轉換的列名,如果剩下的列全部都要轉換,就不用寫了。

2013 -0.4 -0.1 0.0 1.9 -1.3 1.1 -0.7 2.3 -1.6 1.5 -1.4 -1.0

圖,使用.melt() 就可以做到。

• id\_vars:不需要被轉換的列名

參數

2 Shiela Graduate 21 df.melt()

variable

Name

Name

Name

Course

Course Age

value

John

Bob

Shiela

Masters

Course Graduate

Course Graduate

John Masters 27

Bob Graduate



df.melt(value\_vars='Name')

variable value

Name

Name John

Name Shiela

• .pivot()函數根據給定的索引/列值重新組織給定的 DataFrame,接下來以右表為例做介紹

fff bbb baa zzz

Bob

# 重新組織資料

• index:新資料的索引名稱

• values:新資料的值名稱

• columns:新資料的欄位名稱

參數

索引轉成 fff 欄位

值轉成 baa 欄位

索引轉成f ff 欄位

知識點回顧

參考資料

Syntax:

Parameters:

Name

index

Example:

欄位轉成 bbb 欄位

值轉成 baa、zzz 欄位

欄位轉成 bbb 欄位

只轉換 Name 欄位

2 one R 3 two 4 two

0 one

5 two

df.pivot (index = 'fff', columns = 'bbb', values = 'baa')

w3resource.com

在做資料分析時很常要重新組織資料,在裡面最靈活好用的就是 .pivot() 函數

values = 'baa' columns = 'bbb' index = 'fff'

DataFrame

DataFrame index = 'fff' values = [ 'baa', 'zzz' ] columns = 'bbb' w3resource.com 索引轉欄位 .unstack()、欄位轉索引 .stack(),注意都是由最外層開始轉換。

• value\_vars:需要轉換的列名,如果剩下的列全部都要轉換,就不用寫了。

df.pivot (index = 'fff', columns = 'bbb', values = ['baa', 'zzz'])

### pivot 網站:<u>Pandas DataFrame:pivot() function</u>

DataFrame.pivot(self, index=None, columns=None, values=None)

Column to use to make new frame's index. If None, uses existing index.

4.stack和unstack預設旋轉軸的級別將會成果結果中的最低級別(最內層)

欄位名稱轉為欄位值.melt(),其中參數

重新組織資料.pivot(),其中參數

• index:新資料的索引名稱

• values:新資料的值名稱

DataFrame - pivot() function

Description

multiple values will result in a MultiIndex in the columns.

● □columns:新資料的欄位名稱

• id\_vars:不需要被轉換的列名

Column to use to make new frame's columns. columns Column(s) to use for populating new frame's values. If not specified, all remaining columns will be used values and the result will have hierarchically indexed columns. Returns: DataFrame Returns reshaped DataFrame. Raises: ValueError- When there are any index, columns combinations with multiple values. DataFrame.pivot\_table when you need to aggregate.

2.unstack:將資料的行"旋轉"為列

3.stack和unstack預設操作為最內層

5.stack和unstack為一組逆運算操作

stack & unstack 網站:<u>【Python】pandas 軸旋轉 stack 和 unstack 用法詳解</u> 摘要 前面給大家分享了pandas做資料合併的兩篇[pandas.merge] ☑和[pandas.cancat] ☑的用法。今天這篇主要講的 是pandas的DataFrame的軸旋轉操作,stack和unstack的用法。 首先,要知道以下五點: 1.stack:將資料的列"旋轉"為行

The pivot() function is used to reshaped a given DataFrame organized by given index / column values. This function does not support data aggregation,

Type/Default Value

string or object

string or object

the previous

string, object or a list of Optional

Required /

**Optional** 

Optional

Required

你就懂了。 1. 建立DataFrame,行索引名為state,列索引名為number import pandas as pd import numpy as np data = pd.DataFrame(np.arange(6).reshape((2,3)),index=pd.Index(['Ohio','Colorado'],name='sta ,columns=pd.Index(['one','two','three'],name='number')) data melt

為了簡化表中數據的分析,我們可以使用Python中的Pandas將數據重塑為更計算機友好的形式。Panda

第一點和第二點以及第五點比較好懂,可能乍看第三點和第四點會不太理解,沒關係,看看具體下面的例子,

s.melt()是創建的功能之一 。Pandas.melt()取消將DataFrame從寬格式轉換為長格式。melt () 函數很有用,可以將DataFrame壓縮為一種格式,其中一列或多列是標識符變量,而所有其他列(被 視為測量變量)都不會旋轉到行軸,僅留下兩個非標識符列,變量和值。 用法:

**值名稱[標量,默認為"值"]**:用於"值"列的名稱。

pandas.melt (frame , id\_vars = None , value\_vars = None ,

var\_name = None, value\_name ='value', col\_level = None)

網站:<u>Python Pandas.melt()使用及代碼示例</u>

- 參數: ■ 框架: DataFrame id\_vars [元組,列表或ndarray,可選]: 使用標識符變量的列 。value\_vars [元組,列表或ndarray,可選]:要取消透視的列。如果未指定,則使用未。設置為id\_vars的所有列

col\_level [INT或字符串,可選]:如果列是多指標,則使用此級別進行融合。

下一步:閱讀範例與完成作業

VAR\_NAME [標量]:用於"變量"列的名稱如果為無,則使用frame.columns.name或"可變的"。