

Chapter 6 インデックスアラインメント

レシピ45 Indexオブジェクトの精査(P120)

- SeriesやDataFrameの各軸には、値にラベルを付けるIndexオブジェクトがある
- Indexオブジェクトの型は様々だが、同じ振る舞いをする
- MultilIndexを除けば、IndexオブジェクトはすべてPythonの集合とNumpyのndarrayを実装し
- その機能を備えた1次元データ構造になる

- このレシピでは、大学データセットのカラムインデックスを調べ、その機能を試す

(1) collegeデータセットを読み込み、カラムインデックスを変数に代入

```
In [2]: import pandas as pd
college = pd.read_csv('college.csv')
columns = college.columns
columns

Out[2]: Index(['INSTNM', 'CITY', 'STABBR', 'HBCU', 'MENONLY', 'WOMENONLY', 'RELAFFIL',
              'SATVRMID', 'SATMTMID', 'DISTANCEONLY', 'UGDS', 'UGDS_WHITE',
              'UGDS_BLACK', 'UGDS_HISP', 'UGDS_ASIAN', 'UGDS_AIAN', 'UGDS_NHPI',
              'UGDS_2MOR', 'UGDS_NRA', 'UGDS_UNKN', 'PPTUG_EF', 'CURROPER', 'PCTPELL',
              'PCTFLOAN', 'UG25ABV', 'MD_EARN_WNE_P10', 'GRAD_DEBT_MDN_SUPP'],
              dtype='object')
```

(2) values属性をつかって、もとのNumpy配列にアクセスする

```
In [3]: columns.values

Out[3]: array(['INSTNM', 'CITY', 'STABBR', 'HBCU', 'MENONLY', 'WOMENONLY',
              'RELAFFIL', 'SATVRMID', 'SATMTMID', 'DISTANCEONLY', 'UGDS',
              'UGDS_WHITE', 'UGDS_BLACK', 'UGDS_HISP', 'UGDS_ASIAN', 'UGDS_AIAN',
              'UGDS_NHPI', 'UGDS_2MOR', 'UGDS_NRA', 'UGDS_UNKN', 'PPTUG_EF',
              'CURROPER', 'PCTPELL', 'PCTFLOAN', 'UG25ABV', 'MD_EARN_WNE_P10',
              'GRAD_DEBT_MDN_SUPP'], dtype=object)
```

(3) スカラー、リスト、スライスの整数位置をインデックスにして要素を洗濯する

```
In [4]: columns[5]

Out[4]: 'WOMENONLY'

In [5]: columns[[1,8,10]]

Out[5]: Index(['CITY', 'SATMTMID', 'UGDS'], dtype='object')

In [11]: # cf
         columns[-1]

Out[11]: 'GRAD_DEBT_MDN_SUPP'

In [12]: # cf
         columns[-7]

Out[12]: 'PPTUG_EF'

In [7]: columns[-7:-4]

Out[7]: Index(['PPTUG_EF', 'CURROPER', 'PCTPELL'], dtype='object')
```

(4) インデックスは、SeriesやDataFrameと同じメソッドを使う

```
In [13]: columns.min(), columns.max(), columns.isnull().sum()

Out[13]: ('CITY', 'WOMENONLY', 0)
```

(5) Indexオブジェクトに基本算術演算と比較演算を適用する

```
In [14]: columns + '_A'

Out[14]: Index(['INSTNM_A', 'CITY_A', 'STABBR_A', 'HBCU_A', 'MENONLY_A', 'WOMENONLY_A',
              'RELAFFIL_A', 'SATVRMID_A', 'SATMTMID_A', 'DISTANCEONLY_A', 'UGDS_A',
              'UGDS_WHITE_A', 'UGDS_BLACK_A', 'UGDS_HISP_A', 'UGDS_ASIAN_A',
              'UGDS_AIAN_A', 'UGDS_NHPI_A', 'UGDS_2MOR_A', 'UGDS_NRA_A',
              'UGDS_UNKN_A', 'PPTUG_EF_A', 'CURROPER_A', 'PCTPELL_A', 'PCTFLOAN_A',
              'UG25ABV_A', 'MD_EARN_WNE_P10_A', 'GRAD_DEBT_MDN_SUPP_A'],
              dtype='object')

In [15]: columns > 'G'

Out[15]: array([ True, False,  True,  True,  True,  True,  True,  True,  True,
              False,  True,  True,  True,  True,  True,  True,  True,  True,  True,
              True,  True,  True, False,  True,  True,  True,  True,  True])
```

(6) Indexオブジェクトの値を直接変更しようとするとエラーになる。Indexは変更不能オブジェクト

```
In [16]: columns[1] = 'city'

-----
TypeError                                Traceback (most recent call last)
<ipython-input-16-0697dab364a3> in <module>
----> 1 columns[1] = 'city'

~/opt/anaconda3/lib/python3.8/site-packages/pandas/core/indexes/base.py in __setitem__(self, key, value)
    4275     @final
    4276     def __setitem__(self, key, value):
-> 4277         raise TypeError("Index does not support mutable operations")
    4278
    4279     def __getitem__(self, key):

TypeError: Index does not support mutable operations
```

(補足) 割愛