

レシピ32 文字順にスライシング (P85)

- `.loc` インデクサは通常インデックスの文字列ラベルでデータを選ぶ
- その場合、インデックスをソートしておく必要がある

(1) 読み込み、校名をインデックスにセット

```
In [40]: college = pd.read_csv('college.csv', index_col='INSTNM')
college.head(2)
```

```
Out[40]:
```

	CITY	STABBR	HBCU	MENONLY	WOMENONLY	RELAFFIL	SATVRMID	SATMTMID	DISTANCEONLY	UGDS	...	UGDS_2MOR	UGDS_NRA	UGDS_UNKN	PPTUG_EF	CURROPER	PCTPELL
INSTNM																	
Alabama A & M University	Normal	AL	1.0	0.0	0.0	0	424.0	420.0	0.0	4206.0	...	0.0000	0.0059	0.0138	0.0656	1	0.731
University of Alabama at Birmingham	Birmingham	AL	0.0	0.0	0.0	0	570.0	565.0	0.0	11383.0	...	0.0368	0.0179	0.0100	0.2607	1	0.341

2 rows × 26 columns

(2) 校名が'Sp'と'Su'の間の全大学を選ぶ (エラーになる)

```
In [37]: # college.loc['Sp':'Su']
```

```
KeyError: 'Sp'
```

```
In [43]: # cf
college.index.is_monotonic
```

```
Out[43]: False
```

(3) (2) はインデックスがソートされていないのでエラーとなった。よってソートする

```
In [44]: college = college.sort_index()
```

```
In [46]: # cf
college.index.is_monotonic
```

```
Out[46]: True
```

(4) (2) と同じコマンドを実行する

```
In [39]: college.loc['Sp':'Su']
```

```
Out[39]:
```

	CITY	STABBR	HBCU	MENONLY	WOMENONLY	RELAFFIL	SATVRMID	SATMTMID	DISTANCEONLY	UGDS	...	UGDS_2MOR	UGDS_NRA	UGDS_UNKN	PPTUG_EF	CURROPER	PCTPELL
INSTNM																	
Spa Tech Institute-Ipswich	Ipswich	MA	0.0	0.0	0.0	0	NaN	NaN	0.0	37.0	...	0.0000	0.0000	0.0541	0.4054	1	0.2656
Spa Tech Institute-Plymouth	Plymouth	MA	0.0	0.0	0.0	0	NaN	NaN	0.0	153.0	...	0.0000	0.0000	0.2484	0.3399	1	0.3716
Spa Tech Institute-Westboro	Westboro	MA	0.0	0.0	0.0	0	NaN	NaN	0.0	90.0	...	0.0000	0.0000	0.0222	0.5778	1	0.3409
Spa Tech Institute-Westbrook	Westbrook	ME	0.0	0.0	0.0	0	NaN	NaN	0.0	240.0	...	0.0000	0.0000	0.0042	0.2542	1	0.4350
Spalding University	Louisville	KY	0.0	0.0	0.0	1	490.0	440.0	0.0	1227.0	...	0.0302	0.0016	0.0326	0.2502	1	0.4442
...
Studio Academy of Beauty	Chandler	AZ	0.0	0.0	0.0	0	NaN	NaN	0.0	332.0	...	0.0392	0.0000	0.0090	0.0000	1	0.5855
Studio Jewelers	New York	NY	0.0	0.0	0.0	0	NaN	NaN	0.0	55.0	...	0.0000	0.0364	0.0000	0.6000	1	0.0451
Stylemaster College of Hair Design	Longview	WA	0.0	0.0	0.0	0	NaN	NaN	0.0	77.0	...	0.0130	0.0000	0.0000	0.0000	1	0.8036
Styles and Profiles Beauty College	Selmer	TN	0.0	0.0	0.0	0	NaN	NaN	0.0	31.0	...	0.0000	0.0000	0.0000	0.0000	1	0.8182
Styletrends Barber and Hairstyling Academy	Rock Hill	SC	0.0	0.0	0.0	0	NaN	NaN	0.0	45.0	...	0.0000	0.0000	0.0000	0.5556	0	1.0000

201 rows × 26 columns

(補足)

- このレシピでアルファベットの2文字で大学を選ぶのは簡単
- 文字DからSで始まる大学をすべて選ぶなら `college.loc['D':'T']` となる
- このようなスライシングは末尾を含むので、Tという大学も選ばれる
- ちなみにPythonのソートでは、大文字が小文字より先に、数字が大文字より先になる