

Pandas行列の選択

2021/12/01

- `DataFrame = pd.DataFrame(Series or ndarray or list )`
- 1次元: `df.tolist()` → リスト化
- 2次元: `df.to_dict()` → 辞書化
- スライス利用→インデックス演算子は不要。ex. `1:3`
- 複数の羅列→インデックス演算子で囲む ex. `[1, 3]`
- インデックス(行)は列のパラメータ省略可。列は行パラ省略不可
- 整数(`iloc`)とカラム名(`loc`)は混在不可、どちらかに寄せる

No.1

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

`df['COL_1']`(Series)←省略した記法
`df[['COL_1']]`(DataFrame)←特殊？

`df.loc[:, 'COL_1']`(Series)
`df.iloc[:, 1]`(Series)
 (↑この2つは`[[ ]]`形式だとErrorになる)

No.2

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

`df.iloc[1]`(Series)
`df.iloc[[1]]`

`df.loc['ROW_1']`(Series)
`df.loc[['ROW_1']]`

No.3

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

`df[['COL_1', 'COL_3']]`

`df.iloc[:, [1,3]]` <-- 行の「:」は省略不可
`df.loc[:, ['COL_1', 'COL_3']]`

`df.filter(items=['COL_1', 'COL_3':])`

No.4

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

`df.iloc[[1,3],:]`
`df.iloc[[1,3]]` <-- 列の「:」は省略可能

`df.loc[['ROW_1', 'ROW_3'],:]`
`df.loc[['ROW_1', 'ROW_3']]`

No.5

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

`df.iloc[:,1:4]`
`df.loc[:, 'COL_1': 'COL_3']`

No.6

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[1:4,:]
```

```
df.iloc[1:4]
```

```
df.loc['ROW_1':'ROW_3',:]
```

```
df.loc['ROW_1':'ROW_3']
```

No.7

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[1,1]
```

```
df.loc['ROW_1','COL_1']
```

NG) 行列番号と行列名の同時混在は不可

No.8

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[1:4,[1,2]] <-- [1:4, 1:3]でもいいはず
```

```
df.iloc[[1,2,3],[1,2]]
```

```
df.loc['ROW_1':'ROW_3',['COL_1','COL_2']]
```

```
df.loc[['ROW_1','ROW_2','ROW_3'], ['COL_1','COL_2']]
```

No.9

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[1:4,[0,2]]
```

```
df.loc['ROW_1':'ROW_3',['COL_0','COL_2']]
```

No.10

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[[1,3],1:4]
```

```
df.loc[['ROW_1','ROW_3'],'COL_1':'COL_3']
```

No.11

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.index
```

No.12

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.columns
```

Vo.13

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[[1,3]].index
df.loc[['ROW_1','ROW_3']].index
```

Vo.14

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

```
df.iloc[:,[1,3]].columns
df.loc[:,['COL_1','COL_3']].columns
```

Vo.15

	0	1	2	3	4
COL_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

行番号を取得

```
df.index.get_loc('ROW_1')
```

Vo.16

	0	1	2	3	4
COL_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

列番号を取得

```
df.columns.get_loc('COL_1')
df.columns.get_loc('COL_3')
```

Vo.17

	(カラムをセットする)				
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

読み込み後(上書きされる)

```
df.columns = ['col_0','col_1' ~]
            = list('ABC')      ← ['A','B','C']と同じ
            = ['col_' + str(i) for i in range(3)]
```

読み込み時

```
pd.read_csv(~, names=[ 'col_0', 'col_1',~])
```

既存のカラムをリセットする

(リセットは不可だが、以下の方法がある)

```
df.columns = range( (df.shape)[1] )
```

↓indexをセットする

Vo.18

	COL_0	COL_1	COL_2	COL_3	COL_4
	00	01	02	03	04
	10	11	12	13	14
	20	21	22	23	24
	30	31	32	33	34
	40	41	42	43	44
	50	51	52	53	54

リセットするときは

df.reset_index

元のindexは残る、(drop=True)で除外可

読み込み後(既存の列名からセット)

df.set_index('col_0') 上書き: inplace=Trueを追加

読み込み後(新規でセット)

df.index = ['row_1', 'row_2', ~]

読み込み時(既存の列名からセット)

pd.read_csv(～, index_col='col_0')

読み込み時(新規でセット)

pd.read_csv(～, index=['row_1', 'row_2', ~])

indexに名前をつける

df.index.name = '~'

Vo.19

	COL_3	COL_0	COL_4	COL_1	COL_2
ROW_3	33	30	34	31	32
ROW_2	23	20	24	21	22
ROW_1	13	10	14	11	12
ROW_4	43	40	44	41	42
ROW_5	53	50	54	51	52
ROW_0	03	00	04	01	02

行・列の並び替え

```
df.reindex(index=['ROW_3', 'ROW_2', ~],
           columns=['COL_3', 'COL_0', ~])
```

→列だけの並び替えであれば df[['COL_3', 'COL_0', ~]]でもよい

Vo.20		C1 ←集約カラム		C2 ↓集約関数		
idx1	idx2	COL_0	COL_1	COL_2	COL_3	COL_4
R1	ROW_0	00	01	02	03	04
	ROW_1	10	11	12	13	14
	ROW_2	20	21	22	23	24
R2	ROW_3	30	31	32	33	34
	ROW_4	40	41	42	43	44
	ROW_5	50	51	52	53	54

MultiIndex(2階層)を1階層にまとめる(カラム&インデックス)

カラムを1階層にまとめる

level1 = df.columns.get_level_values(0)

level2 = df.columns.get_level_value(1)

df.columns = level1 + '_' + level2

インデックスを1階層に設定

df.reset_index()

Vo.21

	COL_0	COL_1	COL_2	COL_3	COL_4
ROW_0	00	01	02	03	04
ROW_1	10	11	12	13	14
ROW_2	20	21	22	23	24
ROW_3	30	31	32	33	34
ROW_4	40	41	42	43	44
ROW_5	50	51	52	53	54

行名や列名の名前変更

```
df.rename(columns={'bf_name': 'af_name', ~},
          index={'bf_name': 'af_name', ~})
```