

Pandas 모듈




JEONJU UNIVERSITY

Pandas

Pandas 버전확인

```
▶ import pandas  
pandas.__version__
```

↳ '1.1.3'


Pandas

Pandas Series 객체

리스트로 Series 생성

```

[3] import pandas as pd
import numpy as np

data = pd.Series([0.25, 0.5, 0.75, 1.0])

data

0    0.25
1    0.50
2    0.75
3    1.00
dtype: float64


[4] data.values

array([0.25, 0.5 , 0.75, 1.  ])

[5] data.index

RangeIndex(start=0, stop=4, step=1)

```



UNIVERSITY
The Place for Superstars


Pandas

대괄호 표기법

```

▶ data[1]
↳ 0.5

```

인덱스 사용

```

[8] data = pd.Series([0.25, 0.5 , 0.75, 1.  ],
                    index=['a', 'b', 'c', 'd'])

data['b']

0.5

```



JEONJU
UNIVERSITY
The Place for Superstars

Pandas

특수한 딕셔너리 : 딕셔너리로 Series 생성

```
[13] population_dict = {'California' : 38332521,
                        'Texas' : 26448193,
                        'New York' : 19651127,
                        'Florida' : 19552860,
                        'Illinois' : 12882135}

population = pd.Series(population_dict)

population

California    38332521
Texas         26448193
New York     19651127
Florida      19552860
Illinois     12882135
dtype: int64
```

```
[14] population['Texas']

26448193
```



JEONJU UNIVERSITY
The Place for Superstars

Pandas

Series 객체 구성

```
[16] pd.Series([2,4,6])

0    2
1    4
2    6
dtype: int64
```

```
[17] pd.Series(5, index=[100,200,300])


100    5
200    5
300    5
dtype: int64
```

```
[18] pd.Series({'a': 2, 'b': 1, 'c': 3})

2    a
1    b
3    c
dtype: object
```

```
[19] pd.Series({'a': 2, 'b': 1, 'c': 3}, index=[3,2])

3    c
2    a
dtype: object
```



JEONJU UNIVERSITY
The Place for Superstars

Pandas

Pandas DataFrame 객체

배열


```
[21] area_dict = {'California' : 423967,
                 'Texas' : 695662,
                 'New York' : 141297,
                 'Florida' : 170312,
                 'Illinois' : 149995}

area = pd.Series(area_dict)

area
```

California	423967
Texas	695662
New York	141297
Florida	170312
Illinois	149995

dtype: int64



JEONJU
UNIVERSITY
The Place for Superstars

Pandas

```
states = pd.DataFrame({'population' : population, 'area' : area})
states
```


	population	area
California	38332521	423967
Texas	26448193	695662
New York	19651127	141297
Florida	19552860	170312
Illinois	12882135	149995

```
[24] states.index

Index(['California', 'Texas', 'New York', 'Florida', 'Illinois'], dtype='object')
```

```
[25] states.columns

Index(['population', 'area'], dtype='object')
```



UNIVERSITY
The Place for Superstars

Pandas

```
[27] states['area']
```

California	423967
Texas	695662
New York	141297
Florida	170312
Illinois	149995

Name: area, dtype: int64

```
[28] states['population']
```

California	38332521
Texas	26448193
New York	19651127
Florida	19552860
Illinois	12882135

Name: population, dtype: int64



JEONJU UNIVERSITY
The Place for Superstars

Pandas

DataFrame 객체 구성하기

Series로 생성

```
[33] pd.DataFrame(population, columns=['population'])
```

	population
California	38332521
Texas	26448193
New York	19651127
Florida	19552860
Illinois	12882135



JEONJU UNIVERSITY
The Place for Superstars

Pandas

딕셔너리의 리스트로 생성


```
[34] data = [{'a':i, 'b':2 * i} for i in range(3)]
      pd.DataFrame(data)
```

	a	b
0	0	0
1	1	2
2	2	4

일부 키가 누락되면 NaN으로 채워짐

```
[35] pd.DataFrame([{'a':1, 'b':2}, {'b':3, 'c':4}])
```

	a	b	c
0	1.0	2	NaN
1	NaN	3	4.0




JEONJU UNIVERSITY
The Place for Superstars

Pandas

Series의 딕셔너리로 구성

```
[36] pd.DataFrame({'population':population, 'area':area})
```

	population	area
California	38332521	423967
Texas	26448193	695662
New York	19651127	141297
Florida	19552860	170312
Illinois	12882135	149995



JEONJU UNIVERSITY
The Place for Superstars

Pandas

행과 열의 이름 지정

```
[37] pd.DataFrame(np.random.rand(3,2),
                  columns=['foo', 'bar'],
                  index=['a', 'b', 'c'])
```

	foo	bar
a	0.699237	0.794865
b	0.701263	0.427741
c	0.719097	0.122213

JEONJU UNIVERSITY
The Place for Superstars

Pandas

➤ DataFrame

- 행 : index
- 열 : columns

열(변수)

	1열	2열	3열	4열	5열
	이름	성별	영어	수학	학교
1행	뷔	남자	80	69	방탄고
2행	RM	남자	100	78	방탄고
3행	리사	여자	57	90	블핑고
4행	제니	여자	70	100	블핑고

행

NJU UNIVERSITY
The Place for Superstars

Pandas

```
df = pd.DataFrame({'국어' : [98,88,80],
                  '영어' : [80,59,78],
                  '수학' : [69,78,90]},
                  index = ['홍길동', '김길동', '이길동'])
```

df

	국어	영어	수학
홍길동	98	80	69
김길동	88	59	78
이길동	80	78	90



JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
df = pd.DataFrame([[98,88,80],
                  [80,59,78],
                  [69,78,90]],
                  index = ['홍길동', '김길동', '이길동'],
                  columns=['국어', '영어', '수학'])
```

df

	국어	영어	수학
홍길동	98	88	80
김길동	80	59	78
이길동	69	78	90



JEONJU UNIVERSITY
The Place for Superstars

Pandas

▶ 다음 표를 데이터프레임으로 생성하시오

이름	집	나이	키
남예진	수원	20	150
김희정	인천	10	180
박종서	대전	25	200
박건우	대구	50	140
서예인	부산	80	170

JEONJU UNIVERSITY
The Place for Superstars

Pandas 데이터블러오기

```
df = pd.read_csv('titanic.csv')
```

df

	pclass	survived	name	sex	age
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0
1	1	1	Allison, Master. Hudson Trevor	male	NaN
2	1	0	Allison, Miss. Helen Loraine	female	2.0
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.0
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0

JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[41] df.columns

Index(['pclass', 'survived', 'name', 'sex', 'age', 'sibsp', 'parch', 'ticket',
      'fare', 'cabin', 'embarked', 'boat', 'body', 'home.dest'],
      dtype='object')
```

```
[42] df.shape

(1309, 14)
```

```
[45] df['survived'].value_counts()

0    809
1    500
Name: survived, dtype: int64
```




JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[46] df.head()
```

	pclass	survived	name	sex	age	sibsp
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0	0
1	1	1	Allison, Master. Hudson Trevor	male	NaN	1
2	1	0	Allison, Miss. Helen Loraine	female	2.0	1
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.0	1
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0	1



JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[47] df.tail()
```

	pclass	survived	name	sex	age	sibsp
1304	3	0	Zabour, Miss. Hileni	female	14.5	
1305	3	0	Zabour, Miss. Thamine	female	NaN	
1306	3	0	Zakarian, Mr. Mapriededer	male	26.5	
1307	3	0	Zakarian, Mr. Ortin	male	27.0	
1308	3	0	Zimmerman, Mr. Leo	male	29.0	

JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
df.info()
```


```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   pclass      1309 non-null   int64
1   survived    1309 non-null   int64
2   name        1309 non-null   object
3   sex         1309 non-null   object
4   age         1034 non-null   float64
5   sibsp       1309 non-null   int64
6   parch       1309 non-null   int64
7   ticket      1309 non-null   object
8   fare        1308 non-null   float64
9   cabin       295 non-null    object
10  embarked    1307 non-null   object
11  boat        486 non-null    object
12  body        121 non-null    float64
13  home.dest   745 non-null    object
dtypes: float64(3), int64(4), object(7)
memory usage: 143.3+ KB
```

JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[49] df.describe()
```

	pclass	survived	age	sibsp
count	1309.000000	1309.000000	1034.000000	1309.000000
mean	2.294882	0.381971	30.220019	0.498854
std	0.837836	0.486055	14.147138	1.041658
min	1.000000	0.000000	1.000000	0.000000
25%	2.000000	0.000000	21.000000	0.000000
50%	3.000000	0.000000	28.000000	0.000000
75%	3.000000	1.000000	39.000000	1.000000
max	3.000000	1.000000	80.000000	8.000000


JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
df['name']
```

```
0      Allen, Miss. Elisabeth Walton
1      Allison, Master. Hudson Trevor
2      Allison, Miss. Helen Loraine
3      Allison, Mr. Hudson Joshua Creighton
4      Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
...
1304      Zabour, Miss. Hileni
1305      Zabour, Miss. Thamine
1306      Zakarian, Mr. Mapriededer
1307      Zakarian, Mr. Ortin
1308      Zimmerman, Mr. Leo
Name: name, Length: 1309, dtype: object
```



JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
df[['name', 'sex', 'age']]
```

	name	sex	age
0	Allen, Miss. Elisabeth Walton	female	29.0
1	Allison, Master. Hudson Trevor	male	NaN
2	Allison, Miss. Helen Loraine	female	2.0
3	Allison, Mr. Hudson Joshua Creighton	male	30.0
4	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0
...
1304	Zabour, Miss. Hileni	female	14.5
1305	Zabour, Miss. Thamine	female	NaN
1306	Zakarian, Mr. Mapriededer	male	26.5
1307	Zakarian, Mr. Ortin	male	27.0
1308	Zimmerman, Mr. Leo	male	29.0

1309 rows × 3 columns



JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[53] df[12:15]
```

	pc class	survived	name	sex	age
12	1	1	Aubart, Mme. Leontine Pauline	female	24.0
13	1	1	Barber, Miss. Ellen "Nellie"	female	26.0
14	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0

```
[54] df.loc[12:15]
```

	pc class	survived	name	sex	age
12	1	1	Aubart, Mme. Leontine Pauline	female	24.0
13	1	1	Barber, Miss. Ellen "Nellie"	female	26.0
14	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0
15	1	0	Baumann, Mr. John D	male	NaN



JEONJU UNIVERSITY
The Place for Superstars