

Pandas 모듈 II



JEONJU UNIVERSITY

Pandas

CSV 읽고 구조 알아보기

```

▶ import pandas as pd
import numpy as np

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
...
1304	3	0	Zabour, Miss. Hileni	female	14.5
1305	3	0	Zabour, Miss. Thamine	female	NaN



SITY

The Place for Superstars

Pandas

```
df.columns
```

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

```
[5] df.index
```

```
RangeIndex(start=0, stop=1309, step=1)
```

```
[4] df.shape
```

```
(1309, 14)
```


JEONJU UNIVERSITY
The Place for Superstars


Pandas

```
df.head()
```

	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

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


	pclass	survived	name	sex	age	sibsp	parch
1304	3	0	Zabour, Miss. Hileni	female	14.5	1	0
1305	3	0	Zabour, Miss. Thamine	female	NaN	1	0
1306	3	0	Zakarian, Mr. Mapriededer	male	26.5	0	0
1307	3	0	Zakarian, Mr. Ortin	male	27.0	0	0
1308	3	0	Zimmerman, Mr. Leo	male	29.0	0	0


JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[16] 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

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

	pclass	survived	age	sibsp	parch	fare	body
count	1309.000000	1309.000000	1034.000000	1309.000000	1309.000000	1308.000000	121.000000
mean	2.294882	0.381971	30.220019	0.498854	0.385027	33.295479	160.809917
std	0.837836	0.486055	14.147138	1.041658	0.865560	51.758668	97.696922
min	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	21.000000	0.000000	0.000000	7.895800	72.000000
50%	3.000000	0.000000	28.000000	0.000000	0.000000	14.454200	155.000000
75%	3.000000	1.000000	39.000000	1.000000	0.000000	31.275000	256.000000
max	3.000000	1.000000	80.000000	8.000000	9.000000	512.329200	328.000000



JEONJU UNIVERSITY
The Place for Superstars

Pandas


판다스 함수 사용

```
[12] df['pclass'].value_counts()

3    709
1    323
2    277
Name: pclass, dtype: int64
```

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

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



JEONJU UNIVERSITY
The Place for Superstars

Pandas

특정 변수만 뽑아 내기

```
#이름 열만 추출
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

```
[32] # 이름, 나이, 생존여부 열만 추출
df[['name', 'age', 'survived']]
```

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

1309 rows × 3 columns



The Place for Superstars

Pandas

```
[31] # 조건에 의해 특정행 추출
# index가 12~15인 값 추출
df.loc[12:15]
```

	pclass	survived	name	sex	age
12	1	1	Aubart, Mme. Leontine Pauline	female	24.
13	1	1	Barber, Miss. Ellen "Nellie"	female	26.
14	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.
15	1	0	Baumann, Mr. John D	male	NaN




The Place for Superstars

Pandas

```
# 조건에 의해 특정행 추출 후 특정열 추출
# 생존자의 이름 추출
df.loc[df['survived'] == 1, ['name', 'age']]
```

	name	age
0	Allen, Miss. Elisabeth Walton	29.0
1	Allison, Master. Hudson Trevor	NaN
5	Anderson, Mr. Harry	48.0
6	Andrews, Miss. Kornelia Theodosia	63.0
8	Appleton, Mrs. Edward Dale (Charlotte Lamson)	53.0
...
1261	Turkula, Mrs. (Hedwig)	63.0
1277	Vartanian, Mr. David	22.0
1286	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	38.0
1290	Wilkes, Mrs. James (Ellen Needs)	47.0
1300	Yasbeck, Mrs. Antoni (Selini Alexander)	15.0

500 rows × 2 columns




JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
# 조건에 의해 특정행 추출 후 특정열 추출
# 생존자의 이름, 나이 추출
df.loc[df['survived'] == 1][['name', 'age']]
```

	name	age
0	Allen, Miss. Elisabeth Walton	29.0
1	Allison, Master. Hudson Trevor	NaN
5	Anderson, Mr. Harry	48.0
6	Andrews, Miss. Kornelia Theodosia	63.0
8	Appleton, Mrs. Edward Dale (Charlotte Lamson)	53.0
...
1261	Turkula, Mrs. (Hedwig)	63.0
1277	Vartanian, Mr. David	22.0
1286	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	38.0
1290	Wilkes, Mrs. James (Ellen Needs)	47.0
1300	Yasbeck, Mrs. Antoni (Selini Alexander)	15.0

500 rows × 2 columns



JEONJU UNIVERSITY
The Place for Superstars

Pandas


```

# 조건에 의해 특정열 추출 후 특정행 추출
# 생존자의 이름 추출

df[['name', 'age']].loc[df['survived'] == 1]
```

	name	age
0	Allen, Miss. Elisabeth Walton	29.0
1	Allison, Master. Hudson Trevor	NaN
5	Anderson, Mr. Harry	48.0
6	Andrews, Miss. Kornelia Theodosia	63.0
8	Appleton, Mrs. Edward Dale (Charlotte Lamson)	53.0
...
1261	Turkula, Mrs. (Hedwig)	63.0
1277	Vartanian, Mr. David	22.0
1286	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	38.0
1290	Wilkes, Mrs. James (Ellen Needs)	47.0
1300	Yasbeck, Mrs. Antoni (Selini Alexander)	15.0

500 rows x 2 columns


JEONJU UNIVERSITY
The Place for Superstars

Pandas

```

[27] # 나이(age)가 30이상인 명단 추출

df.loc[df['age'] >= 30]
```

	pclass	survived	name	s
3	1	0	Allison, Mr. Hudson Joshua Creighton	ma
5	1	1	Anderson, Mr. Harry	ma
6	1	1	Andrews, Miss. Kornelia Theodosia	fema
7	1	0	Andrews, Mr. Thomas Jr	ma
8	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	fema
...
1286	3	1	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	fema
1287	3	0	Widegren, Mr. Carl/Charles Peter	ma
1290	3	1	Wilkes, Mrs. James (Ellen Needs)	fema
1298	3	0	Wittevrongel, Mr. Camille	ma
1301	3	0	Youseff, Mr. Gerious	mJU

477 rows x 14 columns


JEONJU UNIVERSITY
The Place for Superstars

Pandas

```
[41] # 이름이 A로 시작하는 행만 출력

df.loc[df['name'].str.startswith('A')]
# str은 문자열, startswith()는 문자열의 함수
```

	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	29.0
...
649	3	0	Assam, Mr. Ali	male	29.0
650	3	0	Attalah, Miss. Malake	female	17.0
651	3	0	Attalah, Mr. Sleiman	male	30.0
652	3	0	Augustsson, Mr. Albert	male	29.0
653	3	1	Ayoub, Miss. Banoura	female	17.0

75 rows x 14 columns



Pandas

```
[39] # 이름에 Miss 가 들어있는 행만 추출

df.loc[df['name'].str.find('Miss') >= 0]
# str은 문자열, find()는 문자열의 함수
```

	pclass	survived	name	sex	age	sibsp
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0	0
2	1	0	Allison, Miss. Helen Loraine	female	2.0	1
6	1	1	Andrews, Miss. Kornelia Theodosia	female	63.0	1
13	1	1	Barber, Miss. Ellen "Nellie"	female	26.0	0
18	1	1	Bazzani, Miss. Albina	female	32.0	0
...
1265	3	0	Van Impe, Miss. Catharina	female	10.0	0
1273	3	0	Vander Planke, Miss. Augusta Maria	female	18.0	2
1279	3	0	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0
1304	3	0	Zabour, Miss. Hileni	female	14.5	1
1305	3	0	Zabour, Miss. Thamine	female	NaN	1

260 rows x 14 columns



Pandas

```

[42] #조건이 여러개인 경우(and)
# 생존자 중 나이가 30이상인 사람
df[ (df['age'] >= 30) & (df['survived'] == 1) ]
    
```

	pclass	survived	name	sex	age
	5	1	Anderson, Mr. Harry	male	48.0
	6	1	Andrews, Miss. Kornelia Theodosia	female	63.0
	8	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0
	14	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0
	17	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.0
...
	1228	3	Stranden, Mr. Juho	male	31.0
	1233	3	Sundman, Mr. Johan Julian	male	44.0
	1261	3	Turkula, Mrs. (Hedwig)	female	63.0
	1286	3	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	female	38.0
	1290	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0

194 rows x 14 columns

Pandas

```

#조건이 여러개인 경우(or)
# 생존자 모두 와 나이가 30이상인 사람
df[ (df['age'] >= 30) | (df['survived'] == 1) ]
    
```

	pclass	survived	name	sex	age	sibsp
	0	1	Allen, Miss. Elisabeth Walton	female	29.0	0
	1	1	Allison, Master. Hudson Trevor	male	NaN	1
	3	1	Allison, Mr. Hudson Joshua Creighton	male	30.0	1
	5	1	Anderson, Mr. Harry	male	48.0	0
	6	1	Andrews, Miss. Kornelia Theodosia	female	63.0	1
...
	1287	3	Widegren, Mr. Carl/Charles Peter	male	51.0	0
	1290	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1
	1298	3	Wittevrongel, Mr. Camille	male	36.0	0
	1300	3	Yasbeck, Mrs. Antoni (Selini Alexander)	female	15.0	1
	1301	3	Youseff, Mr. Gerious	male	45.5	0

783 rows x 14 columns