

02

행, 열 선택/ 추가/ 삭제 (pandas)

[row/column 선택, 추가, 삭제]

00:11

데이터 프레임의 인덱스로 특정 데이터를 추출하는 방법을

이해하는 것이 중요합니다!

1.2 row/column 선택, 추가, 삭제

1.2.1 row/column 선택

df['컬럼명'] : 하나의 컬럼을 Series 형식으로 출력
 df[['컬럼명1', '컬럼명2']] : 여러 개의 컬럼을 DataFrame 형식으로 출력
 df.loc[row_list, col_list] : 행과 열을 지정하여 출력 (모든 행 또는 열을 출력할 때는 콜론 : 사용)
 df.iloc[row위치, col위치] : 행과 열 위치를 지정하여 출력 (index 접근)

loc : 레이블별 선택을 위한 순전히 레이블 위치 기반 인덱서
 iloc : 위치별 선택을 위한 순수한 정수 위치 기반 인덱서

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: iris_data = pd.read_csv('../data/iris.csv')
iris_data.head(5)
```

```
Out[2]:
```

sepal length	sepal width	petal length	petal width	target
--------------	-------------	--------------	-------------	--------

[loc, iloc 예제 코드 실습]

04:03

1.2 row/column 선택, 추가, 삭제

1.2.1 row/column 선택

df['컬럼명'] : 하나의 컬럼을 Series 형식으로 출력
 df[['컬럼명1', '컬럼명2']] : 여러 개의 컬럼을 DataFrame 형식으로 출력
 df.loc[row_list, col_list] : 행과 열을 지정하여 출력 (모든 행 또는 열을 출력할 때는 콜론 : 사용)
 df.iloc[row위치, col위치] : 행과 열 위치를 지정하여 출력 (index 접근)

loc : 레이블별 선택을 위한 순전히 레이블 위치 기반 인덱서
 iloc : 위치별 선택을 위한 순수한 정수 위치 기반 인덱서

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: iris_data = pd.read_csv('../data/iris.csv')
iris_data.head(5)
```

```
Out[2]:
```

sepal length	sepal width	petal length	petal width	target
--------------	-------------	--------------	-------------	--------

[새로운 column 추가, 삭제]

08:35

10 5.4 3.7

1.2.2 column 추가 / 삭제

컬럼 추가 : df['컬럼명'] = 선언할 값
컬럼 삭제 : df.drop(columns = ['컬럼명1', '컬럼명2'], inplace=True)

```
In [9]: iris_data['temp'] = range(0,150)
iris_data['temp2'] = "iris"
iris_data['count'] = 1
iris_data.head()
```

Out [9]:

	sepal length	sepal width	petal length	petal width	target	temp	temp2	count
0	5.1	3.5	1.4	0.2	Iris-setosa	0	iris	1
1	4.9	3.0	1.4	0.2	Iris-setosa	1	iris	1
2	4.7	3.2	1.3	0.2	Iris-setosa	2	iris	1
3	4.6	3.1	1.5	0.2	Iris-setosa	3	iris	1
4	5.0	3.6	1.4	0.2	Iris-setosa	4	iris	1

```
In [10]: iris_data.drop(columns=['temp', 'temp2'], inplace=True)
iris_data
```

[새로운 row 추가, 삭제]

12:23

149 5.9 3.0 5.1 1.8 Iris-virginica 1

150 rows x 6 columns

1.2.3 row 추가 / 삭제

행 추가 : df.append(df와 컬럼명이 같은 데이터프레임)
행 삭제 : df.drop(인덱스 리스트)
결측 데이터 삭제 : df.dropna(inplace=True)
중복 데이터 삭제 : df.drop_duplicates()

```
In [11]: import numpy as np
temp = pd.DataFrame([[1,1,1,1,1,1],
                    [2,2,2,2,np.nan],
                    [1,np.nan,np.nan,np.nan,np.nan,np.nan]],
                    columns=iris_data.columns)
```

```
In [12]: iris_data = iris_data.append(temp)
iris_data
```

Out [12]:

	sepal length	sepal width	petal length	petal width	target	count
0	5.1	3.5	1.4	0.2	Iris-setosa	1.0

다음 시간에는 iloc와 loc를 이용하여 데이터 프레임에서 조건에 맞게 데이터 프레임을 선택하고 사용하는 방법에 대해 배워보겠습니다

다음 강의에서 봐요~