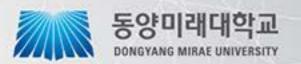


동양미래대학교

2주차 3차시

2차원 데이터를 위한 데이터 타입



# DataFrame



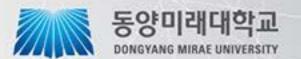


# DataFrame



### 학습개요

- 1/2차원 데이터
- 2/ Pandas의 DataFrame의 구조
- 3/ DataFrame 생성 실습

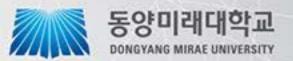


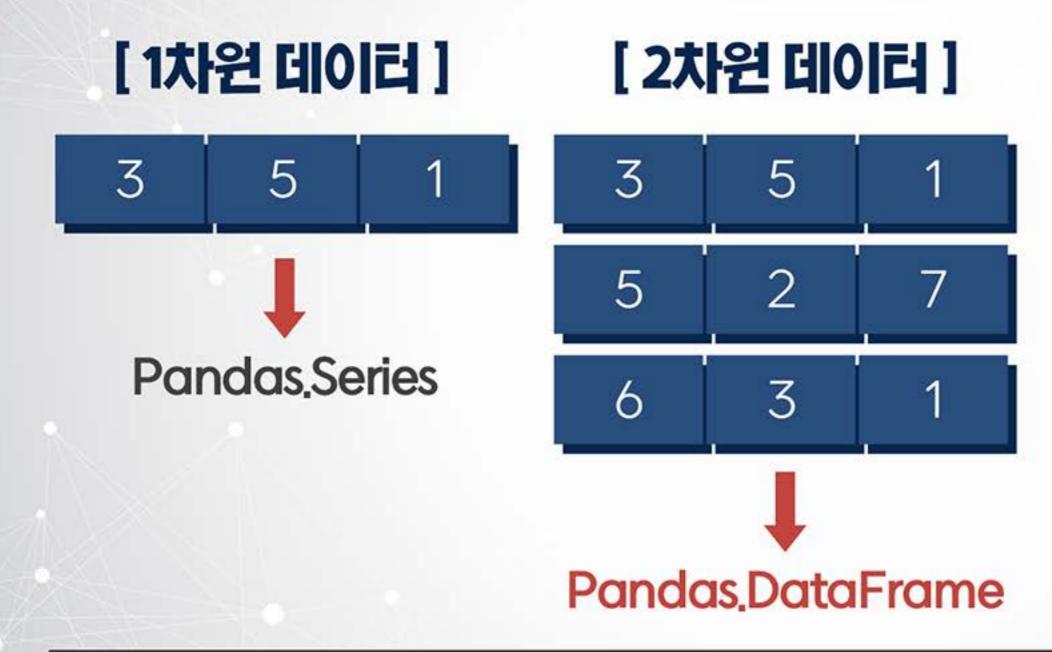
# 1차원형태의데이터

Series 데이터 타입

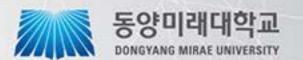
Series는 데이터의 저장과 처리, 분석을 위한 다양한 기능들을 함께 제공하기 때문에 매우 유용하게 사용할 수 있습니다.

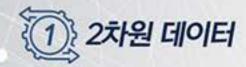








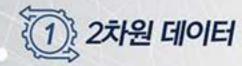




# 2차원 네이터

(데이터 저장·분석에 적합한 구조)





### → 데이터 저장·관리 방법은?

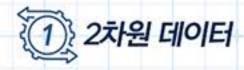
### ✓ 파워포인트 활용

	서울	대전	대구	부산	광주
2018	950	150	85	180	74
2019	945	151	88	187	80
2020	938	145	92	192	81

### ✓ 엑셀 활용

4	Α	В	С	D	E	F
1		서울	대전	대구	부산	광주
2	2018	950	150	85	180	74
3	2019	945	151	88	187	80
4	2020	938	145	92	192	81
5						





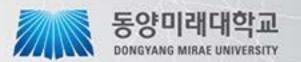
	서울	대전	대구	부산	광주
2018	950	150	85	180	74
2019	945	151	88	187	80
2020	938	145	92	192	81

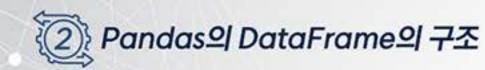
1차원데이터로이루어진 2차원데이터라고할수가있음

1차원데이터 1차원데이터









	서울
2018	950
2019	945
2020	938
1차원 Pandas	



각 도시별로 5개의 시리즈 변수를 만들어서 데이터를 저장할 수가 있습니다.





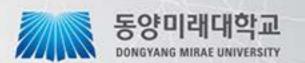


### 도시의 수가 계속 늘어나게 되면?

### 도시 50개 → 50개의 변수 필요

데이터 구조가 필요하게 됨

즉, DataFrame은여러개의시리즈의묶음임





## フPandas의 DataFrame

Series

1차원 데이터를 저장하는 값 밸류(Value), 인덱스(Index)

인덱스의라벨로표현





### フPandas의 DataFrame

✓ DataFrame

2차원 데이터를 저장하는 값, 인덱스(Index)







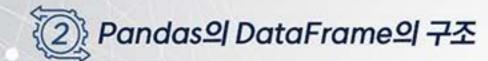
✓ DataFrame



	서울	대전	대구	부산	광주
2018	950	150	85	180	74
2019	945	151	88	187	80
2020	938	145	92	192	81

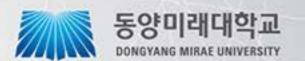






✓ 로우 인덱스와 컬럼 인덱스의 차이점 ①

	서울	대전	대구	부산	광주
2018	950	150	85	180	74
2019	945	151	88	187	80
2020	938	145	92	192	81

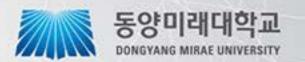


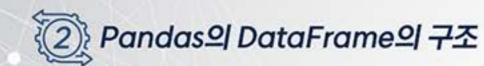


### フPandas의 DataFrame

- ✓ 로우 인덱스는 데이터 간에 순서가 존재함
  - ⊙ 숫자 인덱스나 문자 인덱스로 색인 가능함

이는 첫 번째 데이터만 선택한다거나 혹은 두 번째부터 열 번째까지에 해당하는 데이터를 슬라이싱(slicing) 해서 색인하거나 하는 등의 데이터 색인이 가능함을 의미합니다.





- ✓ 칼럼 인덱스는 데이터 간에 순서가 존재하지 않음
  - ⊙ 순서를 통해서 칼럼을 색인할 수 없음을 의미함



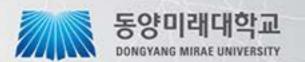


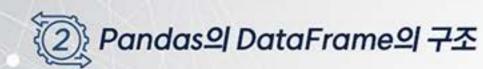
- ✓ 로우 인덱스와 컬럼 인덱스의 차이점 ②
  - ⑥ '축'이라고 하는 의미의 액시스(Axis)라고 하는 인자 제공

Axis = 1

Axis =0

	서울	대전	대구	부산	광주
2018	950	150	85	180	74
2019	945	151	88	187	80
2020	938	145	92	192	81





- ✓ 2차원 형태의 값
  - 데이터 타입의 제약 X

모든 데이터 타입을 하나의 값으로 표현해서 제약 없이 하나의 값으로 모든 데이터 타입을 저장할 수가 있습니다.





# DataFrame을 생성하는 함수

값(데이터)

인덱스

dtype

데이터 타입

copy





칼럼즈(Columns) 인자 존재

컬럼 인덱스(Column index)

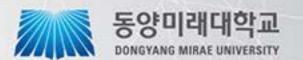
로우 인덱스(Row index)

DataFrame

값(데이터)

인덱스

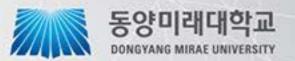
인자값들을지정해서 저희가 데이터 프레임을생성할 수가 있음





### フ DataFrame을 생성하는 함수

- ✓ 데이터 인자
  - ⊙ numpy(넘파이)의 ndarray(엔디 어레이)이나 사전 타입, 리스트, 혹은 다양한 2차원 형태를 다루는 데이터 타입 활용

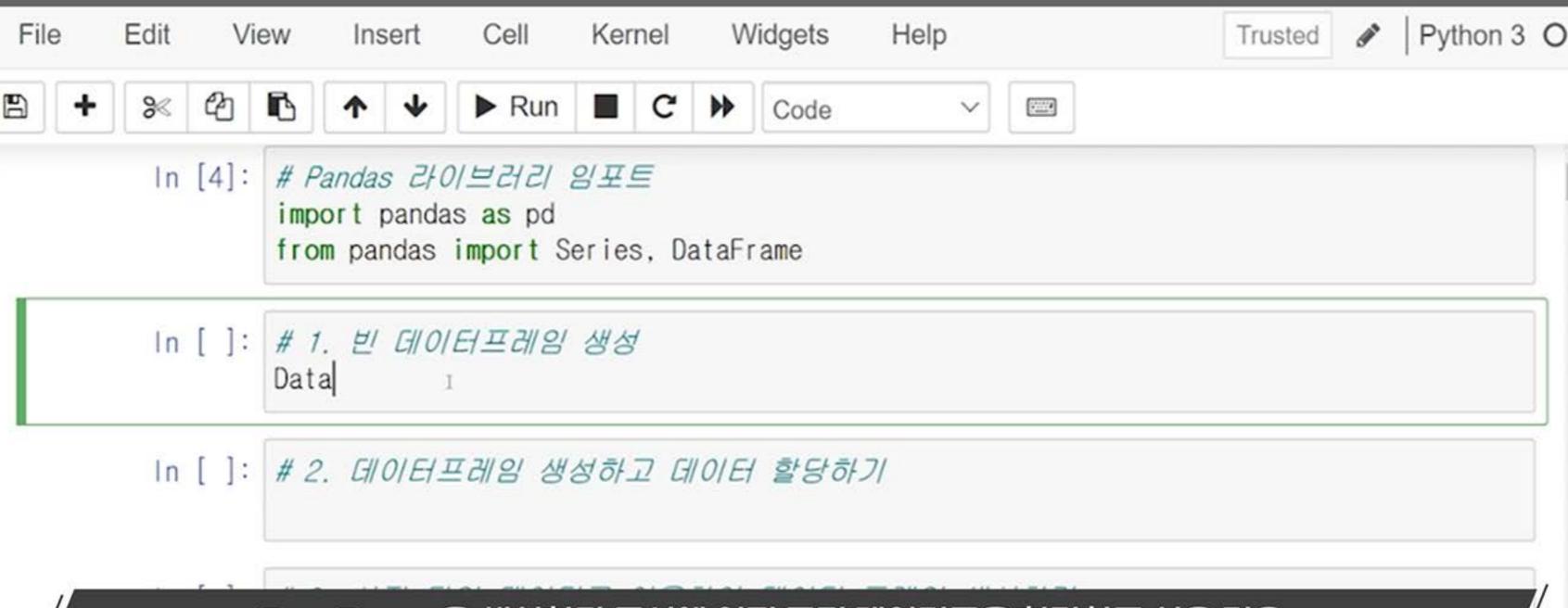


# DataFrame 생성실습

### ◎ 시리즈와 DataFrame을 각각 직접 import 함

```
Insert
File
     Edit
           View
                        Cell
                              Kernel
                                     Widgets
                                              Help
                                                                          Python 3 O
                                                                Trusted
      ▶ Run ■ C ▶ Code
                                                    In []: # Pandas 라이브러리 임포트
             import pandas as pd
             from pandas import Series, DataFrame
       In []: # 1. 빈 데이터프레임 생성
       In []: #2. 데이터프레임 생성하고 데이터 할당하기
       In []: #3. 사전 타입 데이터를 이용하여 데이터 프레임 생성하기
             인구통계 = { '서울 ': [950, 945, 938.5],
                   '대전':[50, 151, 145],
                   '대구':[85, 88, 92].
```

### ◎ DataFrame을 생성하기 위해 DataFrame이라고 하는 생성 함수를 실행함



DataFrame을 생성함과 동시에 어떤 특정 데이터들을 할당하고 싶은 경우 DataFrame 생성 함수에 인자를 활용하면 됩니다.

UT . [00, 00, 92],

◎ 2차원 데이터로 DataFrame 생성하기

```
Python 3 O
                                                         Trusted
✔ DataFrame 생성 함수를 만드는 것을 호출하면서
                                               5.....5
  data라고하는인자에 2차원 데이터를 담아 주면 됨
         DataFrame()
   Out [5]:
   In []: #2. 데이터프레임 생성하고 데이터 할당하기
         DataFrame(data = [[]])
   In []: #3. 사전 타입 데이터를 이용하여 데이터 프레임 생성하기
         인구통계 = { '서울': [950, 945, 938.5],
               '대전':[50, 151, 145],
               '대구':[85, 88, 92],
               '부산':[180, 187, 192],
               '광주':[74, 80, 80]
```

◎ 3개의 값을 가진 리스트 생성

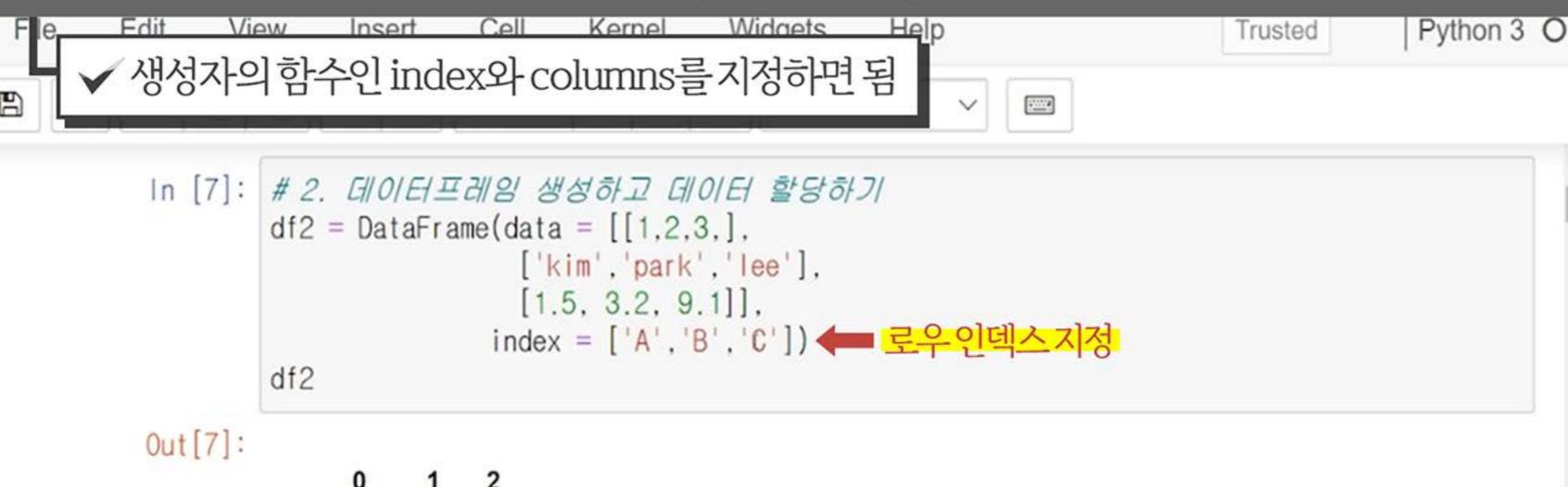
```
Python 3 O
✔ 3개의 값을 가진 리스트 생성할 경우 데이터가 할당된 DataFrame이 생성됨
   In [5]: # 1. 빈 데이터프레임 생성
          DataFrame()
   Out[5]:
   In []: #2. 데이터프레임 생성하고 데이터 할당하기
          DataFrame(data = [[1,2,3,],
                       ['kim', 'park', 'lee'],
                       [1.5, 3.2, 9.1]])
   In []: #3. 사전 타입 데이터를 이용하여 데이터 프레임 생성하기
          인구통계 = { '서울': [950, 945, 938.5],
                '대전':[50, 151, 145],
                '대구':[85, 88, 92].
```

'부사':[180 187 192]

◎ 3개의 값을 가진 리스트 생성

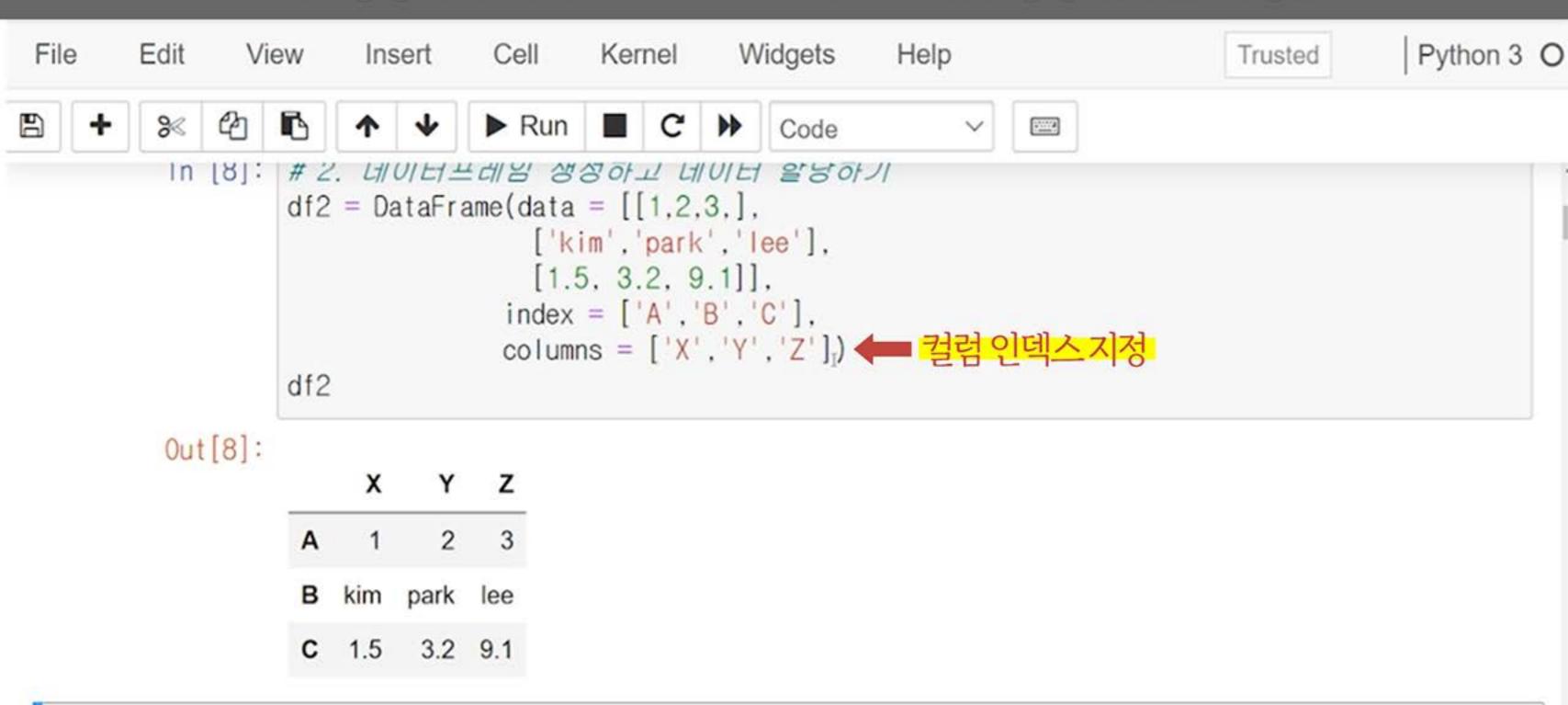


### ◎ 로우 인덱스와 칼럼 인덱스를 같이 지칭하고 싶은 경우

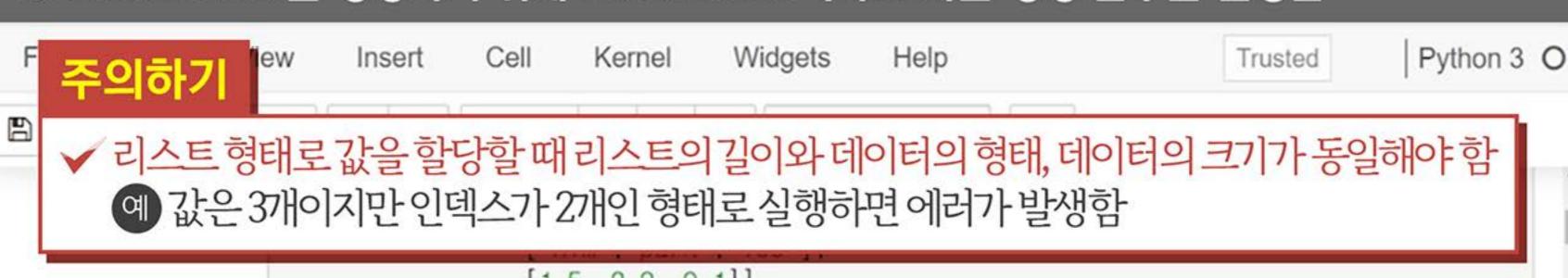


```
0
kim park lee
1.5
    3.2 9.1
```

### ◎ DataFrame을 생성하기 위해 DataFrame이라고 하는 생성 함수를 실행함



◎ DataFrame을 생성하기 위해 DataFrame이라고 하는 생성 함수를 실행함

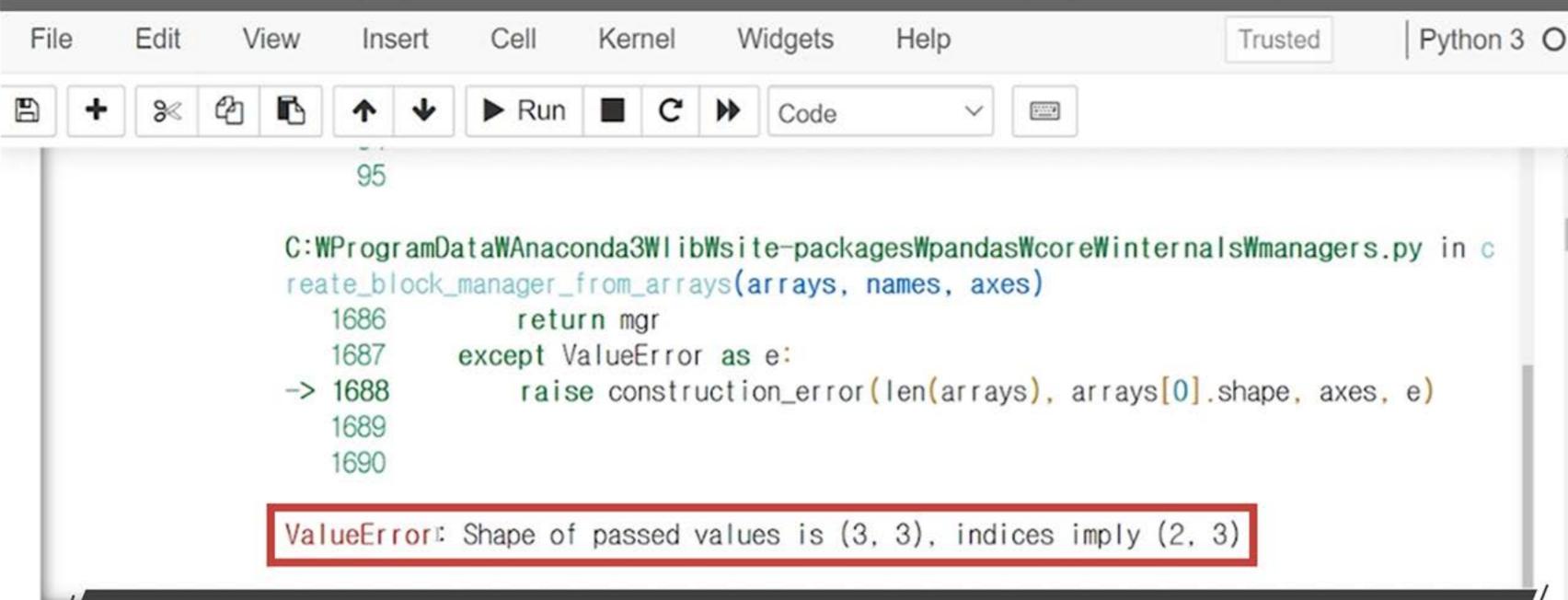


```
[1.5, 3.2, 9.1]].
                index = ['A', 'B', 'C'],
                columns = ['X', 'Y', 'Z']
df2
```

Out[8]:

```
X
kim
    park lee
1.5
     3.2 9.1
```

### ◎ DataFrame을 생성하기 위해 DataFrame이라고 하는 생성 함수를 실행함



데이터를 생성한 후 로우 인덱스와 칼럼 인덱스를 지정할 때에는 반드시 데이터의 모양과 개수를 맞춰 주어야 합니다.

ICILTIL FED JEJ JJE1

'대구':[85, 88, 92],

◎ DataFrame의 로우 인덱스변경하기

```
Python 3 O
                                                                    Trusted
     ✔ DataFrame2(dt2)의 index라고하는속성을활용하여 할당함
In [11]: # 인덱스 변경
               df2.index = ['AA', 'BB', 'CC']
               df2
       Out[11]:
               AA
                   kim park lee
               BB
               CC 41.5
                      3.2 9.1
        In []: #3. 사전 타입 데이터를 이용하여 데이터 프레임 생성하기
               인구통계 = { '서울 ': [950, 945, 938.5],
                     '대전':[50, 151, 145].
```

◎ DataFrame의 칼럼 인덱스변경하기

```
Python 3 O
                                                       Help
                                                                             Trusted
      ✓ columns라고하는속성에 각각의 값을 할당함
13
                                                                .....
        In [12]: # 인덱스 변경
                 df2.index = ['AA', 'BB', 'CC']
                 df2
                 df2.columns = ['XX', 'YY', 'ZZ']
                 df2
        Out [12]:
                      XX
                          YY ZZ
                  AA
                     kim park lee
                 BB
                 CC 1.5 3.2 9.1
```

```
In []: #3. 사전 타입 데이터를 이용하여 데이터 프레임 생성하기
      인구통계 = {'서울': [950, 945, 938.5].
```

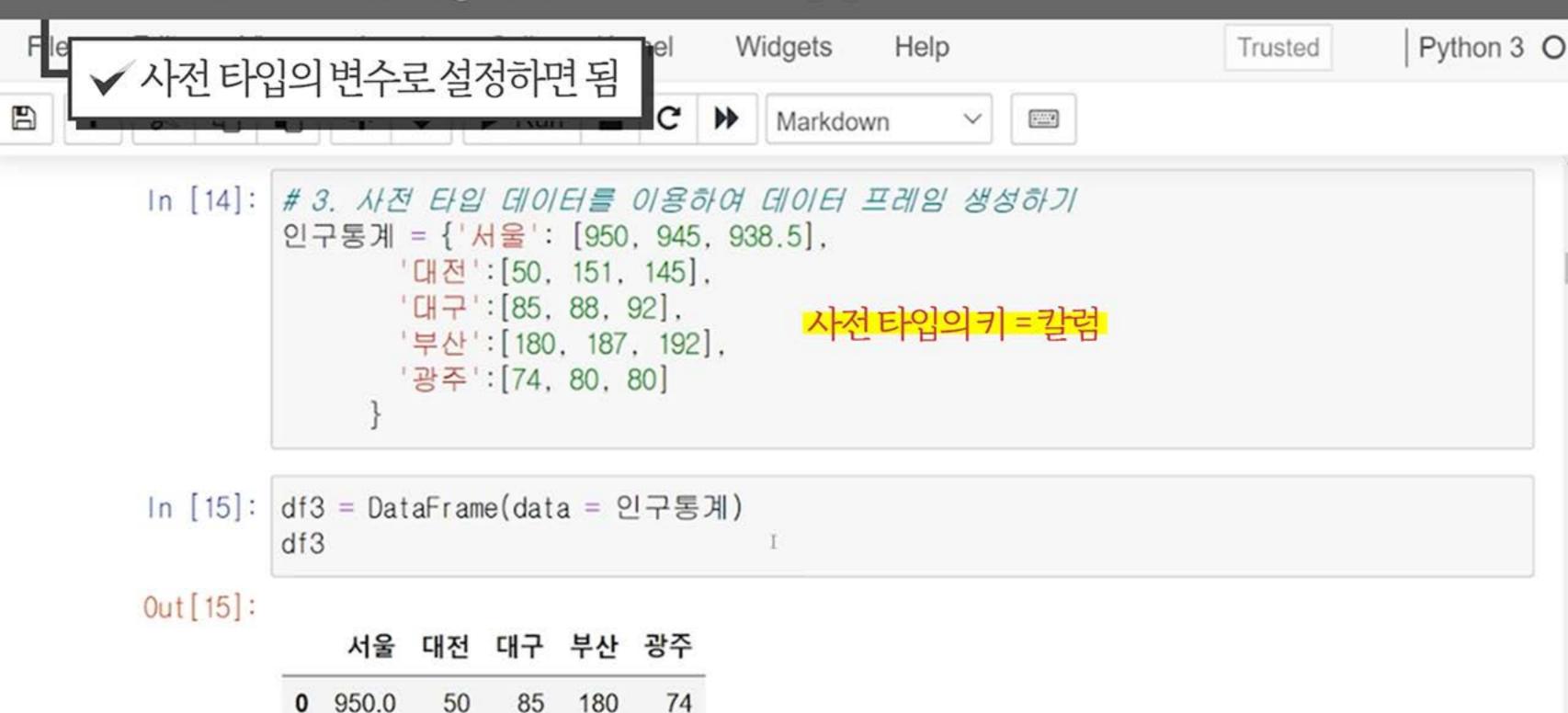
### ◎ 사전 타입 데이터를 이용하여 DataFrame 생성하기

```
View
      Edit
                   Insert
                           Cell
                                         Widgets
                                                                                 Python 3 O
                                 Kernel
                                                  Help
                                                                       Trusted
File
                  ↑ ↓ Run ■ C → Code
              6
                                                          $
               AA
                            3
                   kim park lee
               CC 1.5 3.2 9.1
        In []: #3. 사전 타입 데이터를 이용하여 데이터 프레임 생성하기
               인구통계 = {'서울': [950, 945, 938.5],
                     '대전':[50, 151, 145],
                     '대구':[85, 88, 92],
                     '부산':[180, 187, 192],
                     '광주':[74, 80, 80]
       In [ ]: d
```

◎ 사전 타입 데이터를 이용하여 DataFrame 생성하기

```
Widgets
                                                   Help
                                                                                     Python 3 O
                                                                          Trusted
✔ 사전 타입의 변수로 설정하면 됨
                                       1
                                                            (COO.)
                                           Code
                    '부산':[180, 187, 192],
                    '광주':[74, 80, 80]
   In [13]: df3 = DataFrame(data = 인구통계)
            df3
            NameError
                                                   Traceback (most recent call last)
            <ipython-input-13-fbd04ea41d1a> in <module>
            ----> 1 df3 = DataFrame(data = 인구통계)
                 2 df3
            NameError: name '인구통계' is not defined
```

### ◎ 사전 타입 데이터를 이용하여 DataFrame 생성하기

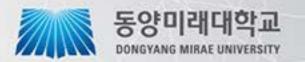


◎ 사전 타입 데이터를 이용하여 DataFrame 생성하기



### Out[16]:

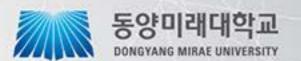
	서울	대전	대구	부산	광주
2018	950.0	50	85	180	74
2019	945.0	151	88	187	80
2020	938.5	145	92	192	80





## **B** DataFrame의 구조

- ✓ 로우 인덱스(Row Index)
- ✓ 칼럼 인덱스(Columns Index)
- ✓ 값(Data) → numpy의 ndarray 타입





### **DataFrame**

- ✓ 2차원 데이터를 다루는 매우 중요한 데이터 타입임
- ➡ 데이터를저장하고관리하기위해 많이사용하는형태의데이터





## ® DataFrame의 다양한 기능

- ✓ DataFrame의 주요 속성
- ✓ DataFrame의 주요 함수



# 로우 인덱스는 순서가 있고, 칼럼 인덱스는 순서가 없다.





## **銀 2차원 데이터를 위한 데이터 타입**

- 1/2차원 데이터
- 2/ Pandas의 DataFrame의 구조
- 3/ DataFrame 생성 실습



수고하셨습니다!