

Pandas



Pandas DataFrame

- 2차원 자료구조인 DataFrame는 행과 열이 있는 테이블 데이터

딕셔너리를 DataFrame으로 형 변환

```
dt = {'일시' : ['2017-01-01', '2017-02-01'], '평균' : [4.1, 5.5], '최고' : [15.1, 17.7], '최저' : [-7.7, -5.7]}
dt
```

```
{'일시': ['2017-01-01', '2017-02-01'],  
 '최고': [15.1, 17.7],  
 '최저': [-7.7, -5.7],  
 '평균': [4.1, 5.5]}
```



```
#딕셔너리를 데이터 프레임으로 변경  
#키가 열이 되고 값이 행이 됨  
df = pd.DataFrame(dt)  
df
```

	일시	평균	최고	최저
0	2017-01-01	4.1	15.1	-7.7
1	2017-02-01	5.5	17.7	-5.7

데이터 프레임 정보 보기

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 30 entries, 0 to 29  
Data columns (total 4 columns):  
#   Column    Non-Null Count  Dtype  
---  ---  
0   일시      30 non-null    object  
1   평균      30 non-null    float64  
2   최고      30 non-null    float64  
3   최저      30 non-null    float64  
dtypes: float64(3), object(1)  
memory usage: 1.1+ KB
```



Pandas 외부데이터 가져오기

- csv 읽기

- `df = pd.read_csv('파일명')`
- 한글 파일명을 사용할 경우 :
`df = pd.read_csv('파일명', engine='python')`
- 파일 내용 한글이 깨지는 경우(euc-kr, cp949) :
`df = pd.read_csv('파일명', encoding='cp949')`

- csv 쓰기

- `df.to_csv('파일명')`

- 엑셀 읽기

- 한글 파일명을 사용할 경우 : `df = pd.read_excel('파일명')`



Pandas DataFrame 인덱스

- 열의 인덱스 : 데이터프레임.columns
- 행의 인덱스 : 데이터프레임.index

```
#열 인덱스
```

```
print(df.columns)
```

```
Index(['항목', '구분', '중구', '서구', '동구', '영도구', '부산진구', '동래구', '남구', '북구',  
      '해운대구', '사하구', '금정구', '강서구', '연제구', '수영구', '사상구', '기장군'],  
      dtype='object')
```

```
#행 인덱스
```

```
print(df.index)
```

```
RangeIndex(start=0, stop=8, step=1)
```



Pandas DataFrame 인덱스

- 데이터프레임의 인덱스 변경

- 열인덱스 명 변경 :

- 데이터프레임.column = [열명인덱스]

- 열항목 중 하나를 인덱스로 사용 :

- 데이터프레임.set_index(열명)

- 행인덱스로 사용 :

- 데이터프레임.reset_index()

#열 인덱스명 변경

```
df.columns = [item.strip() for item in df.columns]
df.columns
```

```
Index(['항목', '구분', '중구', '서구', '동구', '영도구',  
      '사하구', '금정구', '강서구', '연제구', '수영구',  
      dtype='object')
```

#행 인덱스 변경

```
df = df.set_index('항목')
df
```

	구분	중구	서구	동구	영도구	부산진구
항목						
매립처리량	연간(톤)	3034	2323	1653	1219	17393
매립처리량	일일(톤)	8	7	5	3	48
소각처리량	연간(톤)	13729	10446	8297	8651	538

#행 인덱스를 자료형으로 변경

```
df = df.reset_index()
df
```

	항목	구분	중구	서구	동구	영도구	부산진구
0	매립처리량	연간(톤)	3034	2323	1653	1219	17393
1	매립처리량	일일(톤)	8	7	5	3	48
2	소각처리량	연간(톤)	13729	10446	8297	8651	538

Pandas DataFrame 열/행 가져오기

- 열 가져오기 : 데이터프레임[열인덱스]
- 행 가져오기 : 데이터프레임[행인덱스]

```
#열가져오기
dfitem = df['항목']
print(type(dfitem))
print(dfitem)
```

```
<class 'pandas.core.series.Series'>
0    매립처리량
1    매립처리량
2    소각처리량
3    소각처리량
4    재활용처리량
5    재활용처리량
6    음식물류발생량
7    음식물류발생량
Name: 항목, dtype: object
```

```
#열가져오기
df2 = df[['항목', '구분', '금정구']]
print(type(df2))
df2
```

```
<class 'pandas.core.frame.DataFrame'>
      항목  구분  금정구
0  매립처리량  연간(톤)    7240
1  매립처리량  일일(톤)     20
```

```
#행 데이터 가져오기
df23 = df[2:4]
df23
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구
2	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	15
3	소각처리량	일일(톤)	38	29	23	24	2	0	

```
#행 데이터 가져오기
df23 = df[df['항목'] == '매립처리량']
df23
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구
0	매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042
1	매립처리량	일일(톤)	8	7	5	3	48	25	22

해결문제

- 부산쓰레기발생2017년.csv 파일에서 '금정구, 해운대구, 수영구'의 연간 매립처리량과 소각처리량, 음식물류발생량 정보를 추출하시오.

	항목	구분	금정구	해운대구	수영구
0	매립처리량	연간(톤)	7240	5251	2205
2	소각처리량	연간(톤)	222	31924	11987
4	재활용처리량	연간(톤)	56153	64667	28135
6	음식물류발생량	연간(톤)	21273	40043	13289

