

pandas 집계함수

In []:

```
import pandas as pd
cctv_seoul = pd.read_csv("data/CCTV_in_Seoul.csv")

cctv_seoul.head()      #5개 행만
```

Out[16]:

	기관명	소계	2013년도 이전	2014년	2015년	2016년
0	강남구	2780	1292	430	584	932
1	강동구	773	379	99	155	377
2	강북구	748	369	120	138	204
3	강서구	884	388	258	184	81
4	관악구	1496	846	260	390	613

In []:

```
print(cctv_seoul["소계"].mean())      # "소계" 열 값의 평균
print(cctv_seoul.소계.sum())          # "소계" 열 값의 합
print(cctv_seoul.소계.min())          # "소계" 열 값의 최소
print(cctv_seoul.소계.max())          # "소계" 열 값의 최대
print(cctv_seoul.소계.std())          # "소계" 열 값의 표준편차
print(cctv_seoul.소계.count())        # "소계" 열 값의 건수
```

1179.08
29477
485
2780
556.7289825399788
25

In []:

```
cctv_seoul.describe()
```

Out[18]:

	소계	2013년도 이전	2014년	2015년	2016년
count	25.000000	25.000000	25.000000	25.000000	25.000000
mean	1179.080000	764.760000	159.480000	205.200000	385.880000
std	556.728983	415.510257	104.132976	137.877059	182.479981
min	485.000000	238.000000	21.000000	30.000000	81.000000
25%	748.000000	464.000000	78.000000	103.000000	292.000000
50%	1015.000000	573.000000	142.000000	184.000000	377.000000
75%	1496.000000	1070.000000	218.000000	269.000000	467.000000
max	2780.000000	1843.000000	430.000000	584.000000	932.000000

In []:

```
crime_seoul = pd.read_csv("data/crime_in_Seoul_include_gu_name.csv")

#컬럼명에 공백이 포함되면 crime_seoul.폭력 발생 <-- 이렇게 쓸수 없음. 아래처럼 사용.
print( crime_seoul["폭력 발생"].mean()) # "폭력 발생" 열 값의 평균

print( crime_seoul.groupby("구별")["폭력 발생"].mean()) # "폭력 발생" 열 값의 평균. 단, 구별
print( crime_seoul.groupby("구별")["폭력 발생"].sum()) # "폭력 발생" 열 값의 합. 단, 구별
print( crime_seoul.groupby("구별")["폭력 발생"].min()) # "폭력 발생" 열 값의 최소값. 단, 구별
print( crime_seoul.groupby("구별")["폭력 발생"].max()) # "폭력 발생" 열 값의 최대값. 단, 구별
print( crime_seoul.groupby("구별")["폭력 발생"].std()) # "폭력 발생" 열 값의 표준편차. 단, 구별
print( crime_seoul.groupby("구별")["폭력 발생"].count()) # "폭력 발생" 열 값이 존재하는 행 갯수.
```

```
강남구      2142.0
강동구      2712.0
강북구      2649.0
관악구      3298.0
광진구      2625.0
구로구      3007.0
금천구      2054.0
노원구      2723.0
도봉구      1487.0
동대문구    2548.0
동작구      1910.0
마포구      2983.0
서대문구    2056.0
서초구      1199.5
성동구      1612.0
성북구      1104.5
송파구      3295.0
양천구      2858.0
영등포구    3572.0
용산구      2050.0
(이하 생략)
```

numpy 집계함수

In []:

```
import numpy as np

x = np.array([ 18, 5, 10, 23, 19, -8, 10, 0, 0, 5, 2, 15, 8,
               2, 5, 4, 15, -1, 4, -7, -24, 7, 9, -6, 23, -13])

print(x.sum(), np.sum(x)) # 합
print(x.mean(), np.mean(x)) # 평균
print(np.median(x)) # 중앙값
print(x.var(), np.var(x)) # 분산
print(x.std(), np.std(x)) # 표준편차

print(x.min(), np.min(x)) # 최소값
print(x.max(), np.max(x)) # 최대값
print(x.argmin(), np.argmin(x)) # 최소값의 인덱스
print(x.argmax(), np.argmax(x)) # 최대값의 인덱스
```

```
125 125
4.8076923076923075 4.8076923076923075
5.0
115.23224852071006 115.23224852071006
10.734628476137871 10.734628476137871
-24 -24
23 23
20 20
3 3
```

In []:

```
x = np.array([[ 1, 1],
               [ 2, 2]])
print(" 전체 합 ", x.sum())
print(x.sum(axis=0)) #열 합
print(x.sum(axis=1)) #행 합계

print(np.percentile(x, 0)) # 최소값
print(np.percentile(x, 25)) # 1/4
print(np.percentile(x, 50)) # 2/4
print(np.percentile(x, 75)) # 3/4
print(np.percentile(x, 100)) # 최대값
```