

## [ 집계 ]

pandas 활용

< 컬럼 값들의 전체 합, 평균, 최소값, 최대값, 표준편차, 총건수 예) >

```
In [9]: import pandas as pd

#####
cctv_seoul = pd.read_csv("CCTV_in_Seoul.csv")

print(cctv_seoul.head())           #5개 행만

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

#####
```

	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
		1179.08				
		29477				
		485				
		2780				
		556.7289825399788				
		25				

요약정보

```
In [11]: print(cctv_seoul.describe())
```

	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



<그룹별 컬럼 값들의 합, 평균, 최소값, 최대값, 표준편차, 총건수 예) >

“폭력 발생” 컬럼의 전체 평균

```
In [13]: crime_seoul = pd.read_csv("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

2103.4193548387098
```

구별 컬럼 값 기준 그룹별 “폭력 발생” 컬럼 값들의 합, 평균, 최소값, 최대값, 표준편차, 총건수

```
In [16]: 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. 단, 구별
```

```
구별
강남구      2142.0
강동구      2712.0
강북구      2649.0
..중략
중구        1112.0
중랑구      2847.0
Name: 폭력 발생, dtype: float64
```

```
구별
강남구      4284
강동구      2712
강북구      2649
..중략
중구        2224
중랑구      2847
Name: 폭력 발생, dtype: int64
```

```
구별
강남구      1819
강동구      2712
강북구      2649
..중략
중구        869
중랑구      2847
Name: 폭력 발생, dtype: int64
```

```
구별
강남구      2465
강동구      2712
강북구      2649
..중략
중구        1355
```

```
중량구      2847
Name: 폭력 발생, dtype: int64
```

```
구별
강남구      456.790981
강동구      NaN
..중략
은평구      190.211724
종로구      185.969083
중구        343.653896
중랑구      NaN
```

```
Name: 폭력 발생, dtype: float64
```

```
구별
강남구      2
강동구      1
강북구      1
..중략
종로구      2
중구        2
중랑구      1
```

```
Name: 폭력 발생, dtype: int64
```