Pandas 외부데이터 가져오기

Pandas 엑셀데이터 일부가져오기

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
2	매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
3	매립처리량	일일(톤)	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6
4	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
5	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
6	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170
7	재활용처리량	일일(톤)	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42
8	음식물류발생량	연간(톤)	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648
9	음식물류발생량	일일(톤)	14	24	18	26	74	55	43	66	110	58	58	30	39	36	66	46

```
#엑셀 데이터 가져오기
import pandas as pd

data = pd.read_excel('부산쓰레기발생2017년.xlsx')
data
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
0	매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
1	매립처리량	일일(톤)	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6
2	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
3	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
4	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170
5	재활용처리량	일일(톤)	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42
6	음식물류발생량	연간(톤)	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648
7	음식물류발생량	일일(톤)	14	24	18	26	74	55	43	66	110	58	58	30	39	36	66	46

엑셀데이터 가져오기 – read_excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
2	매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
3	매립처리량	일일(톤)	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6
4	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
5	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
6	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170
7	재활용처리량	일일(톤)	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42
8	음식물류발생량	연간(톤)	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648
9	음식물류발생량	일일(톤)	14	24	18	26	74	55	43	66	110	58	58	30	39	36	66	46

```
#옵션 : header -시작 행수 (0부터 시작), parse_cols -열 명
data1 = pd.read_excel('부산쓰레기발생2017년.xlsx',
                      header = 0,
                      parse_cols = 'A, B, F')
data1
```

	항목	구분	영도구
0	매립처리량	연간(톤)	1219
1	매립처리량	일일(톤)	3
2	소각처리량	연간(톤)	8651
3	소각처리량	일일(톤)	24
4	재활용처리량	연간(톤)	24616
5	재활용처리량	일일(톤)	68
6	음식물류발생량	연간(톤)	9522
7	음식물류발생량	일일(톤)	26

정보 확인 – index/columns/values

```
#인덱스 정보 확인
data.index
```

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

```
#컬럼 정보 확인
data.columns
```

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

```
#열명에 공백이 있어 제거하고 열명을 변경
#열명 변경
c = data.columns
c = list(c)
c = [item.strip() for item in c ]

data.columns = c
data.columns
```

열명을 리스트로 만들어 공백을 제거

공백이 제거된 리스트를 이용하여 컬럼명 변경

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

```
#내용 확인
data.values
```

```
array([['매립처리량', '연간(톤)', 3034, 2323, 1653, 1219, 17393, 9195, 8042, 10498,
        5251, 3089, 7240, 9742, 1831, 2205, 9937, 2290],
       ['매립처리량', '일일(톤)', 8, 7, 5, 3, 48, 25, 22, 29, 14, 8, 20, 27, 5, 6,
        27, 6],
```

정보 확인 – info / describe

```
#데이터프레임 개요
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 18 columns):
항목      8 non-null object
구분      8 non-null object
중구      8 non-null int64
```

```
#통계적 개요
data.describe()
```

	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구
count	8.000000	8.000000	8.000000	8.00000	8.000000	8.000000	8.00000	8.000000	8.000000
mean	3885.500000	5370.875000	4301.500000	5516.12500	16733.625000	10762.625000	10236.62500	12368.125000	17784.125000
std	5154.228278	7695.575082	6407.015173	8697.50791	30757.419467	19850.254672	20081.97826	22528.407936	24790.055301
min	8.000000	7.000000	5.000000	3.00000	2.000000	0.000000	0.00000	0.000000	14.000000
25%	22.250000	27.750000	21.750000	25.50000	67.500000	23.000000	37.75000	56.750000	104.250000
50%	1536.000000	1190.500000	851.000000	643.50000	390.000000	105.000000	158.50000	136.000000	2714.000000
75%	6136.000000	9195.750000	6873.500000	8868.75000	19824.500000	11928.250000	9918.50000	13923.750000	33953.750000
max	13729.000000	21301.000000	17968.000000	24616.00000	88453.000000	56526.000000	57921.00000	63879.000000	64667.000000

정렬 – sort_values

```
#정렬
data.sort_values(by='구분', ascending=False )
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
1	매립처리량	일일(톤)	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6
3	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
5	재활용처리량	일일(톤)	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42
7	음식물류발생량	일일(톤)	14	24	18	26	74	55	43	66	110	58	58	30	39	36	66	46
0	매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
2	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
4	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170
6	음식물류발생량	연간(톤)	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648

열 보기

```
#하나의 열내용 보기
data['구분']
```

```
0    연간(톤)
1    일일(톤)
2    연간(톤)
3    일일(톤)
4    연간(톤)
5    일일(톤)
6    연간(톤)
7    일일(톤)
Name: 구분, dtype: object
```

```
#여러 열의 내용 보기
data[['항목', '구분', '영도구']]
```

	항목	구분	영도구
0	매립처리량	연간(톤)	1219
1	매립처리량	일일(톤)	3
2	소각처리량	연간(톤)	8651
3	소각처리량	일일(톤)	24
4	재활용처리량	연간(톤)	24616
5	재활용처리량	일일(톤)	68
6	음식물류발생량	연간(톤)	9522
7	음식물류발생량	일일(톤)	26

행 보기

```
#하나의 행 내용보기
data[1:2]
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
1	매립처리량	일일(톤)	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6

```
#여러행 내용보기
data[2:5]
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
2	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
3	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
4	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170

슬라이싱을 이용한 행보기

```
#행 슬라이싱
data.loc[3]
```

```
항목       소각처리량
구분       일일(톤)
중구          38
서구          29
동구          23
영도구         24
부산진구         2
동래구          0
남구           0
북구           0
해운대구        87
사하구         89
금정구          1
강서구         13
연제구         40
수영구         33
사상구          0
기장군         37
Name: 3, dtype: object
```

```
data.loc[2:5]
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
2	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
3	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
4	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170
5	재활용처리량	일일(톤)	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42

슬라이싱 – loc / iloc

```
#일부분 슬라이싱

data.loc[2:3, ['항목', '구분', '영도구']]
```

	항목	구분	영도구
2	소각처리량	연간(톤)	8651
3	소각처리량	일일(톤)	24

- loc :
  라벨값 기반의 2차원 인덱싱
- iloc :
  순서를 나타내는 정수 기반의 2차원 인덱싱

```
#일부분 슬라이싱
data.iloc[2:4, [0,1,5]]
```

	항목	구분	영도구
2	소각처리량	연간(톤)	8651
3	소각처리량	일일(톤)	24

데이터 생성 – 조건에 맞는 행추출

```
#조건에 맞는 데이터 추출하여 데이터 분리

df1 = data[data['구분'] == '연간(톤)']
df1 = df1.sort_values('항목')
df1
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
0	매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
2	소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
6	음식물류발생량	연간(톤)	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648
4	재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170

```
df2 = data[data.구분 == '일일(톤)']
df2 = df2.sort_values('항목')
df2
```

	항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
1	매립처리량	일일(톤)	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6
3	소각처리량	일일(톤)	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
7	음식물류발생량	일일(톤)	14	24	18	26	74	55	43	66	110	58	58	30	39	36	66	46
5	재활용처리량	일일(톤)	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42

인덱스 변경 – set_index

	항목	구분	중구	서구	동구
0	매립처리량	연간(톤)	3034	2323	1653
2	소각처리량	연간(톤)	13729	10446	8297
6	음식물류발생량	연간(톤)	5154	8779	6399
4	재활용처리량	연간(톤)	9082	21301	17968

```
#인덱스 변경
df1 = df1.set_index('항목')
df1
```

set_index :
기존의 행 인덱스를 제거하고 데이터 열 중 하나를 인덱스로 설정

항목	구분	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
매립처리량	연간(톤)	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
소각처리량	연간(톤)	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
음식물류발생량	연간(톤)	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648
재활용처리량	연간(톤)	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170

열 삭제 – del/drop

```
#열 삭제
del df1['구분']
df1
```

항목	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
매립처리량	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290
소각처리량	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687
음식물류발생량	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648
재활용처리량	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170

```
df2 = df2.drop('구분', axis=1)
df2
```

항목	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군
매립처리량	8	7	5	3	48	25	22	29	14	8	20	27	5	6	27	6
소각처리량	38	29	23	24	2	0	0	0	87	89	1	13	40	33	0	37
음식물류발생량	14	24	18	26	74	55	43	66	110	58	58	30	39	36	66	46
재활용처리량	25	58	49	68	242	155	159	175	177	194	154	71	102	77	156	42

평균 열/행 추가

```
# 평균 열 추가
df1['평균'] = df1.mean(axis=1)
df1
```

항목	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군	평균
매립처리량	3034	2323	1653	1219	17393	9195	8042	10498	5251	3089	7240	9742	1831	2205	9937	2290	5933.8750
소각처리량	13729	10446	8297	8651	538	17	158	97	31924	32671	222	4788	14521	11987	216	13687	9496.8125
음식물류발생량	5154	8779	6399	9522	27119	20128	15548	24201	40043	21183	21273	10810	14122	13289	24161	16648	17398.6875
재활용처리량	9082	21301	17968	24616	88453	56526	57921	63879	64667	70923	56153	25977	37115	28135	57110	15170	43437.2500

```
#행 추가
df2.loc['평균'] = df2.mean(axis=0)
df2
```

항목	중구	서구	동구	영도구	부산진구	동래구	남구	북구	해운대구	사하구	금정구	강서구	연제구	수영구	사상구	기장군	평균
매립처리량	8.00	7.0	5.00	3.00	48.0	25.00	22.0	29.0	14.0	8.00	20.00	27.00	5.0	6.0	27.00	6.00	16.250000
소각처리량	38.00	29.0	23.00	24.00	2.0	0.00	0.0	0.0	87.0	89.00	1.00	13.00	40.0	33.0	0.00	37.00	26.000000
음식물류발생량	14.00	24.0	18.00	26.00	74.0	55.00	43.0	66.0	110.0	58.00	58.00	30.00	39.0	36.0	66.00	46.00	47.687500
재활용처리량	25.00	58.0	49.00	68.00	242.0	155.00	159.0	175.0	177.0	194.00	154.00	71.00	102.0	77.0	156.00	42.00	119.000000
평균	21.25	29.5	23.75	30.25	91.5	58.75	56.0	67.5	97.0	87.25	58.25	35.25	46.5	38.0	62.25	32.75	52.234375

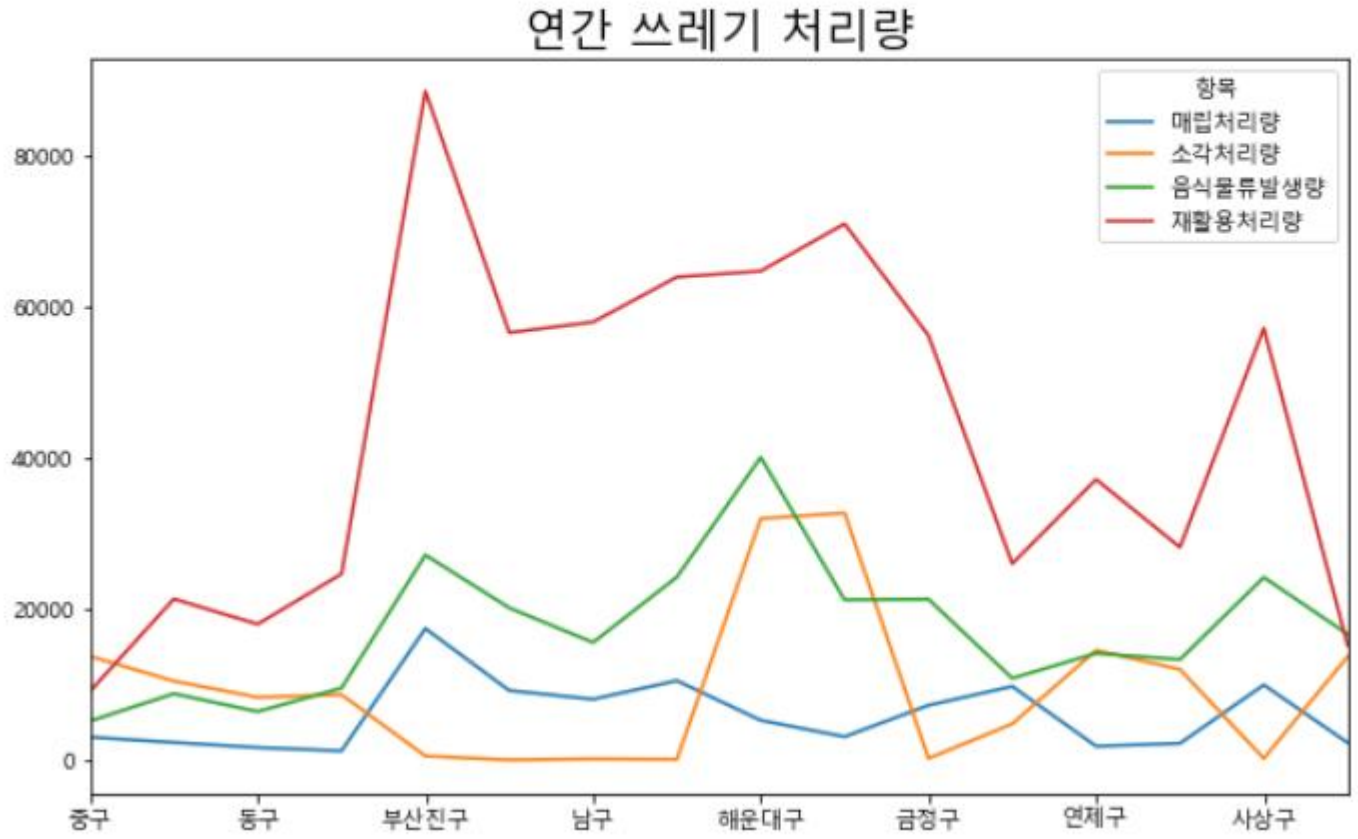
행열 전환 – T

```
#행과 열 전환
df1 = df1.T
df1
```

항목	매립처리량	소각처리량	음식물류발생량	재활용처리량
중구	3034.000	13729.0000	5154.0000	9082.00
서구	2323.000	10446.0000	8779.0000	21301.00
동구	1653.000	8297.0000	6399.0000	17968.00
영도구	1219.000	8651.0000	9522.0000	24616.00
부산진구	17393.000	538.0000	27119.0000	88453.00
동래구	9195.000	17.0000	20128.0000	56526.00
남구	8042.000	158.0000	15548.0000	57921.00
북구	10498.000	97.0000	24201.0000	63879.00
해운대구	5251.000	31924.0000	40043.0000	64667.00
사하구	3089.000	32671.0000	21183.0000	70923.00
금정구	7240.000	222.0000	21273.0000	56153.00
강서구	9742.000	4788.0000	10810.0000	25977.00
연제구	1831.000	14521.0000	14122.0000	37115.00
수영구	2205.000	11987.0000	13289.0000	28135.00
사상구	9937.000	216.0000	24161.0000	57110.00
기장군	2290.000	13687.0000	16648.0000	15170.00
평균	5933.875	9496.8125	17398.6875	43437.25

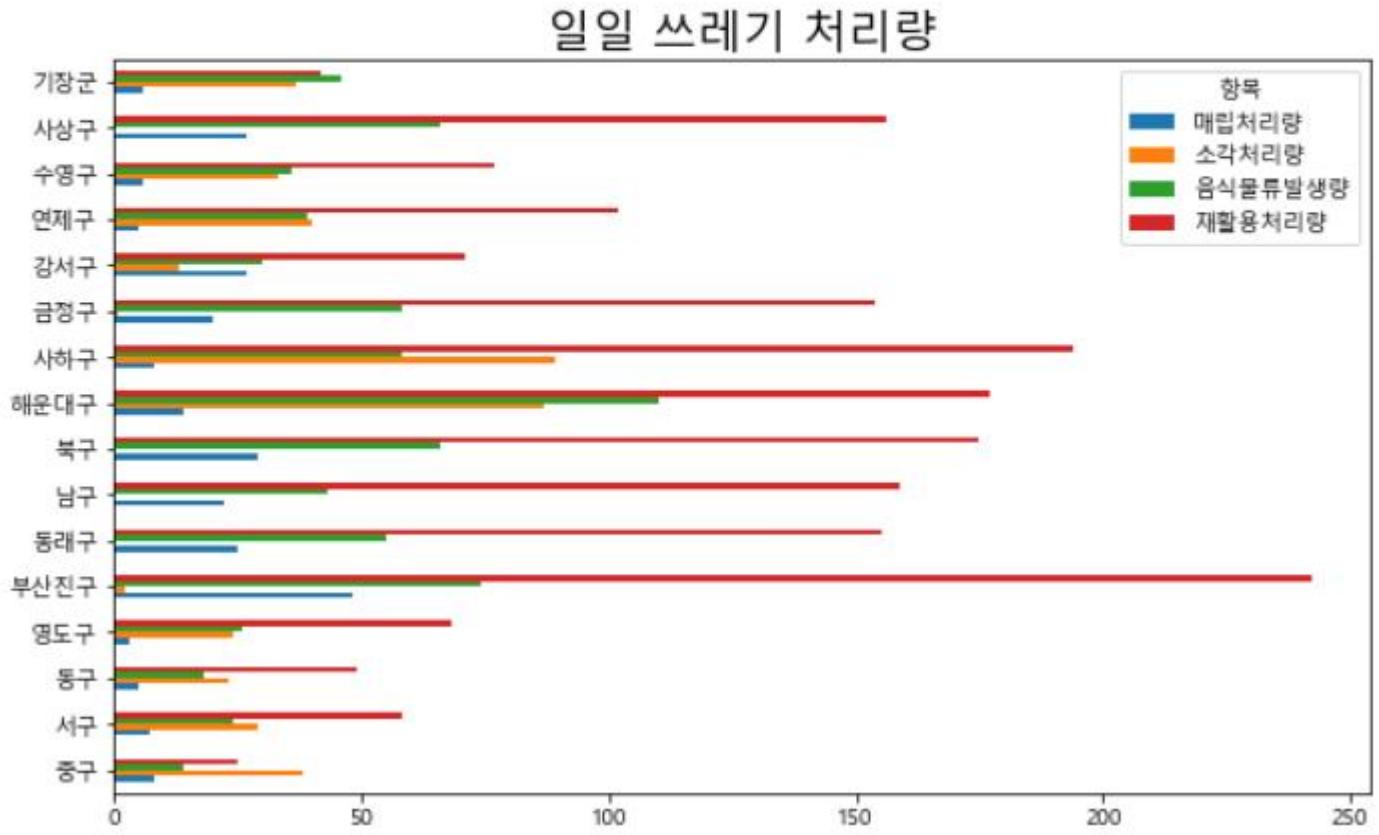
시각화 - plot

```
#그림 크기 지정
df1.plot(figsize=(10,6))
plt.title('연간 쓰레기 처리량', size=20)
plt.show()
```



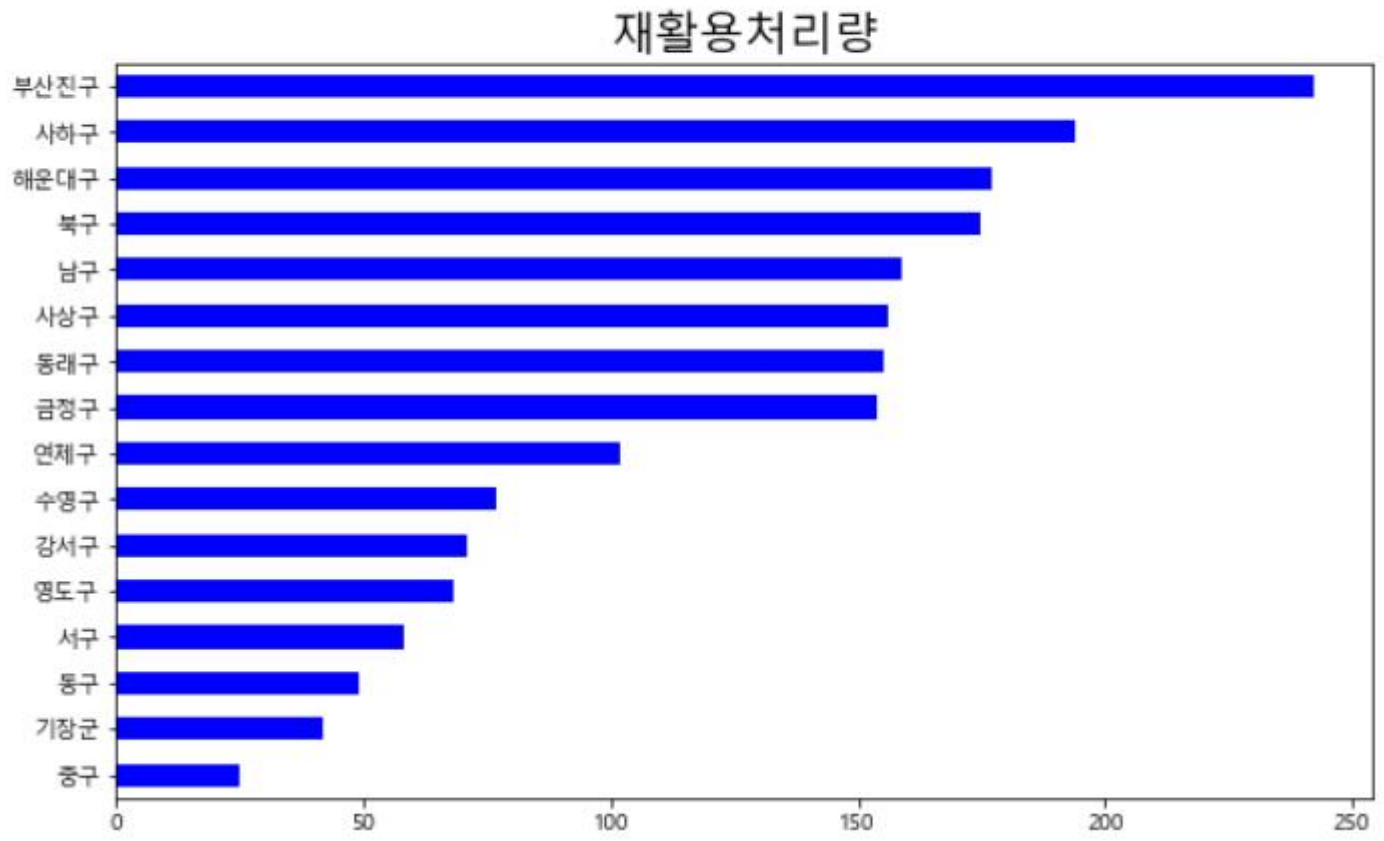
시각화 – bar/barh

```
#막대 그래프
df2.plot(kind='barh', figsize=(10,6))
plt.title('일일 쓰레기 처리량', size=20)
plt.show()
```



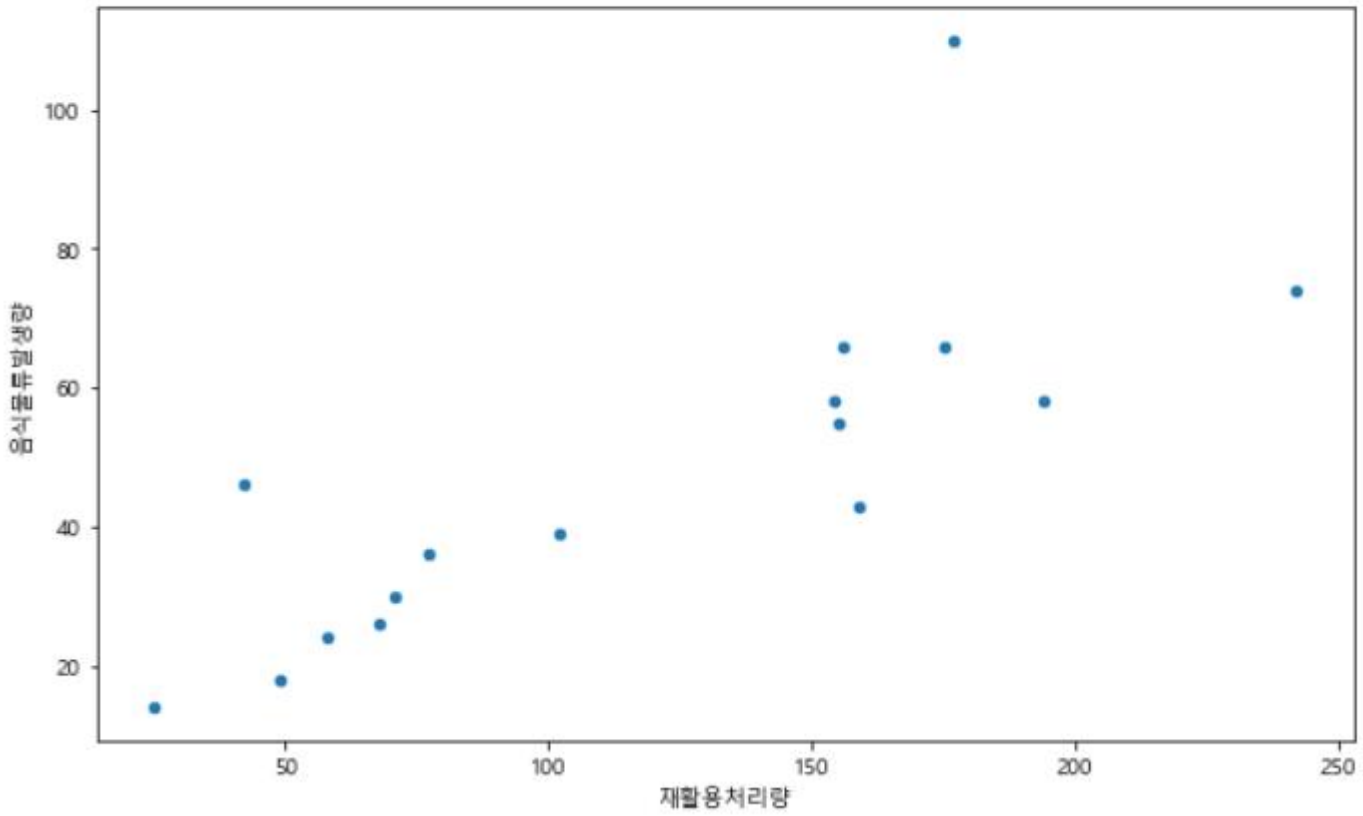
시각화 – bar/barh

```
df2['재활용처리량'].sort_values().plot(kind='barh',color='blue' , figsize=(10,6))
plt.title('재활용처리량', size=20)
plt.show()
```



시각화 - scatter

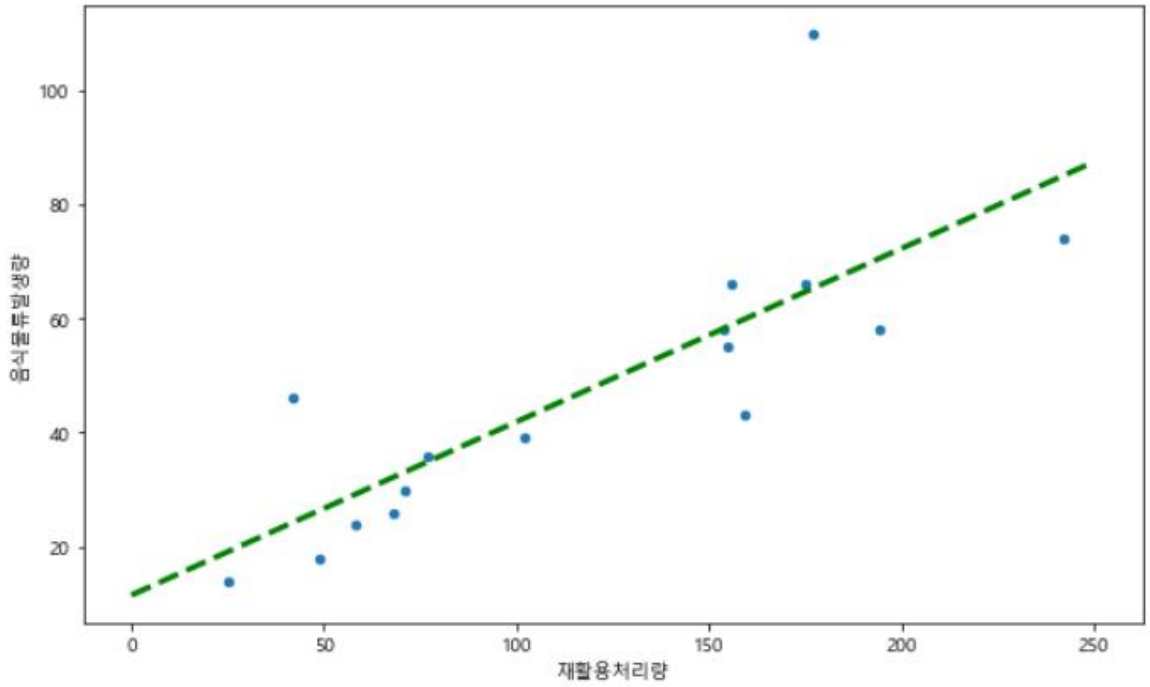
```
#산점도 그래프
df2.plot.scatter(x='재활용처리량', y='음식물류발생량', figsize=(10,6))
plt.show()
```



시각화 - scatter

```
#산점도 추세선
import numpy as np
fp1 = np.polyfit(df2['재활용처리량'], df2['음식물류발생량'], 1)
f1 = np.poly1d(fp1)
fx = np.linspace(0, 250, 10)
```

```
df2.plot.scatter(x='재활용처리량', y='음식물류발생량', figsize=(10,6))
plt.plot(fx, f1(fx), ls='dashed', lw=3, color='g' )
plt.show()
```

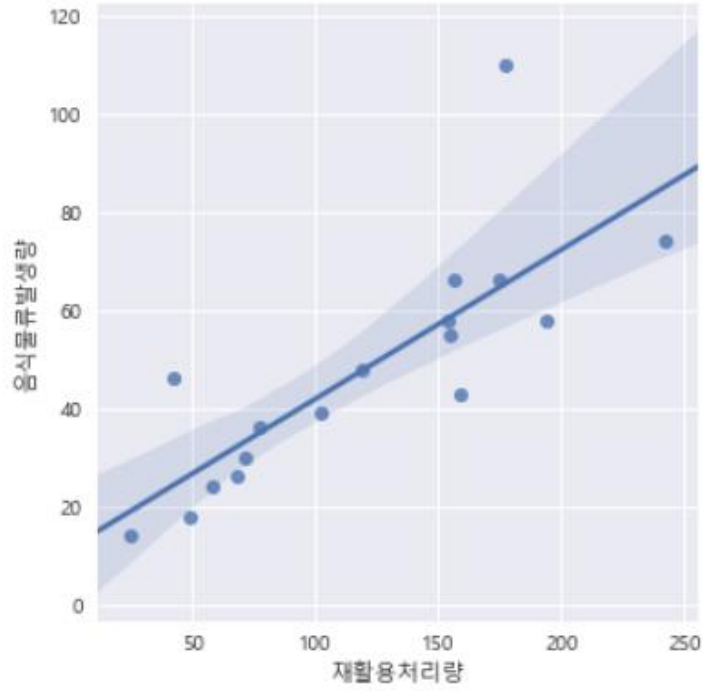


seaborn 시각화 - lmplot

```
#seaborn 라이브러리를 이용한 시각화
import seaborn as sns

plt.figure(figsize=(10,6))
sns.lmplot(x='재활용처리량', y='음식물류발생량', data=df2)
plt.show()
```

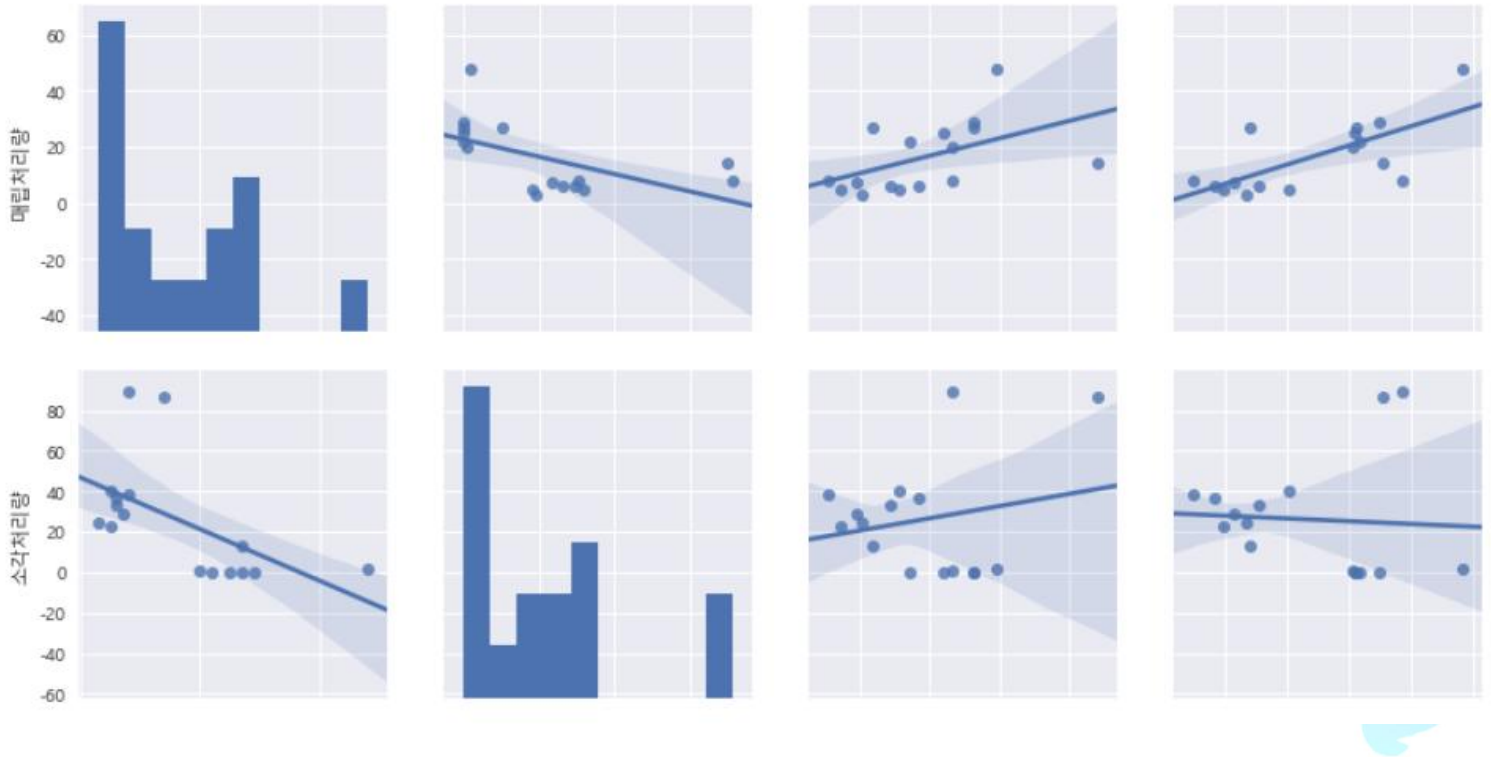
```
<matplotlib.figure.Figure at 0x247be822ac8>
```



seaborn 시각화 - pairplot

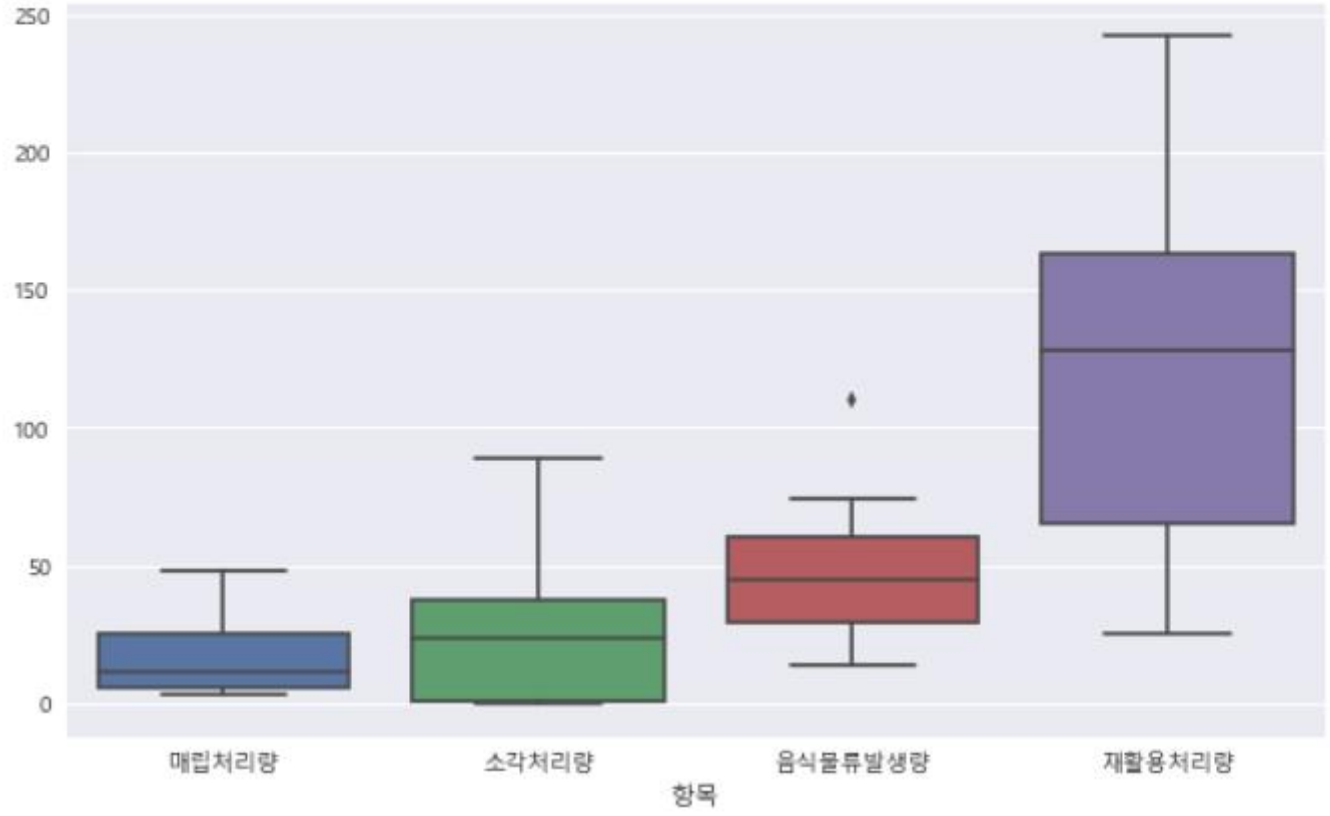
```
plt.figure(figsize=(10,6))
sns.pairplot(df2, kind='reg', size=3)
plt.show()
```

```
<matplotlib.figure.Figure at 0x247ca4ffb38>
```



seaborn 시각화 - boxplot

```
#boxplot
plt.figure(figsize=(10,6))
sns.boxplot(data=df2)
plt.show()
```



Folium 시각화

- 설치 : pip install folium

```
#지도 시각화
#pip install folium
#https://github.com/southkorea/southkorea-maps
import folium
map_osm = folium.Map(location=[35.166804, 129.083479], zoom_start=11)

for item in busan.index:
    lat = busan.loc[item,'위도']
    long = busan.loc[item,'경도']
    item = item.strip()
    folium.CircleMarker( [lat,long],
                        radius=df2.loc[item,'재활용처리량']/10 ,
                        popup=df2.loc[item,'재활용처리량'] ,
                        color='crimson',
                        fill =True).add_to(map_osm)
map_osm
```