

```
In [1]: # 필요 라이브러리 로드
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#출력 그림 크기 조절
sns.set(rc={'figure.figsize':(12,12)})
plt.style.use('ggplot')
from matplotlib import font_manager, rc
font_path = "C:/Windows/Fonts/NGULIM.TTF"
font = font_manager.FontProperties(fname=font_path).get_name()
rc('font', family=font)

import warnings
warnings.filterwarnings("ignore")
```

전처리한 각 데이터 로드

```
In [2]: df_1 = pd.read_csv('1번데이터_전처리.csv', encoding = 'cp949')
df_1
```

```
Out[2]:
```

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt
0	2018-01	서귀포시	남원읍	52.0	9570	42437700	1270773
1	2018-01	서귀포시	대륜동	38.0	21666	57612600	1676850
2	2018-01	서귀포시	대정읍	89.0	10185	38885550	1164122
3	2018-01	서귀포시	대천동	37.0	20280	53858550	1593709
4	2018-01	서귀포시	동홍동	49.0	45936	118701000	3501286
...
1661	2021-06	제주시	일도2동	87.0	84360	147438200	4402149
1662	2021-06	제주시	조천읍	141.0	27732	63927750	1911187
1663	2021-06	제주시	한경면	71.0	8031	27060150	809898
1664	2021-06	제주시	한림읍	112.0	25653	82746990	2476292
1665	2021-06	제주시	화북동	84.0	66088	110750050	3306029

1666 rows × 7 columns

```
In [3]: df_2 = pd.read_csv("방문인구병합.csv", encoding = "cp949")
df_2 = df_2.drop("Unnamed: 0", axis=1)
df_2
```

```
Out[3]:
```

	y_m	city	location	visit_pop_cnt	visit_pop_cnt_1f	visit_pop_cnt_sf
0	2018-01	서귀포시	남원읍	7650.915819	469.700347	1686.010590
1	2018-01	서귀포시	대륜동	7964.610507	371.723546	420.833558
2	2018-01	서귀포시	대정읍	4994.300798	410.511223	356.262995
3	2018-01	서귀포시	대천동	4671.393555	234.238089	225.574534

	y_m	city	location	visit_pop_cnt	visit_pop_cnt_lf	visit_pop_cnt_sf
4	2018-01	서귀포시	동홍동	3881.022202	270.258431	353.699388
...
1801	2021-06	제주시	조천읍	13478.916454	377.112397	30.738787
1802	2021-06	제주시	추자면	1043.921835	424.469189	41.810643
1803	2021-06	제주시	한경면	3886.111826	603.810148	13.838063
1804	2021-06	제주시	한림읍	11498.611893	607.927826	28.297822
1805	2021-06	제주시	화북동	2410.348416	157.539409	10.453663

1806 rows × 6 columns

```
In [4]: df_3 = pd.read_csv('3번 데이터_전처리.csv', encoding = 'cp949')
df_3
```

Out[4]:

	y_m	city	location	제주도 민_여	외국인 거주_여	제주도 민_남	외국인거 주_남	제주도민 _60이상	제주도민 _60미만	total_pop
0	2018-01	제주시	건입동	4715	82	5085	146	2478	7550	10028
1	2018-02	제주시	건입동	4724	82	5090	142	2484	7554	10038
2	2018-03	제주시	건입동	4711	81	5072	141	2494	7511	10005
3	2018-04	제주시	건입동	4718	86	5070	143	2484	7533	10017
4	2018-05	제주시	건입동	4730	90	5070	148	2504	7534	10038
...
1801	2021-02	서귀포시	효돈동	2616	21	2673	44	1748	3606	5354
1802	2021-03	서귀포시	효돈동	2609	20	2653	41	1752	3571	5323
1803	2021-04	서귀포시	효돈동	2607	19	2660	40	1761	3565	5326
1804	2021-05	서귀포시	효돈동	2605	19	2661	41	1766	3560	5326
1805	2021-06	서귀포시	효돈동	2597	18	2660	39	1764	3550	5314

1806 rows × 10 columns

```
In [5]: df_4 = pd.read_csv('4번 데이터_전처리.csv', encoding = 'cp949')
df_4
```

Out[5]:

	y_m	city	location	패스트 푸드_결 제건수	패스트푸드_ 결제금액	간식_결 제건수	간식_결제금 액	농축 수산물_결 제건수	농축수산물_ 결제금액	마트/슈 퍼마켓_ 결제건 수	
0	2018-01	서귀포시	남원읍	3337.0	57030976.0	3073.0	36003854.0	1753.0	132565191.0	13786.0	7
1	2018-01	서귀포시	대륜동	4889.0	81574888.0	5266.0	62481472.0	2026.0	133506435.0	25909.0	1
2	2018-01	서귀포시	대정읍	5221.0	98015164.0	8310.0	88847534.0	1959.0	146060716.0	30433.0	1
3	2018-01	서귀포시	대천동	5136.0	80300156.0	5351.0	56156408.0	1722.0	116891485.0	23893.0	9
4	2018-01	서귀포시	동홍동	6908.0	126657868.0	7959.0	89540973.0	4031.0	263004249.0	55399.0	2
...
1759	2021-06	제주시	일도2동	10469.0	135801500.0	16303.0	186182590.0	8846.0	519191996.0	60791.0	1
1760	2021-06	제주시	조천읍	6678.0	113730242.0	8324.0	112097929.0	3773.0	307434356.0	25722.0	1
1761	2021-06	제주시	한경면	1406.0	26876799.0	1507.0	23045501.0	1040.0	96593409.0	8888.0	3
1762	2021-06	제주시	한림읍	4315.0	81433739.0	5256.0	74235134.0	2075.0	221504210.0	20129.0	7
1763	2021-06	제주시	화북동	7944.0	125227233.0	10916.0	129374585.0	8098.0	664687009.0	48143.0	1

1764 rows × 17 columns

```
In [6]: df_5 = pd.read_csv('기상데이터_전처리.csv', encoding = 'cp949')
df_5
```

```
Out[6]:
```

	y_m	location	풍속	기온	습도	강수
0	2018-01	건입동	2.593011	4.800134	66.350806	2.800000

	y_m	location	풍속	기온	습도	강수
1	2018-01	구좌읍	3.899194	4.218952	74.362903	2.870833
2	2018-01	남원읍	2.619220	6.256586	62.920699	1.891667
3	2018-01	노형동	2.991129	4.887231	68.990591	1.791667
4	2018-01	대륜동	1.831317	8.004301	66.793011	2.958333
...
1717	2021-06	표선면	2.315000	22.187778	89.306944	9.237500
1718	2021-06	한경면	4.363194	22.407500	93.737500	7.270833
1719	2021-06	한림읍	2.888889	22.537500	92.995833	6.758333
1720	2021-06	화북동	2.432778	23.299861	75.793056	5.058333
1721	2021-06	효돈동	1.550833	22.751111	86.093056	8.325000

1722 rows × 6 columns

```
In [7]: df_6 = pd.read_csv('코로나데이터_전처리.csv', encoding = 'cp949')
df_6
```

Out[7]:

	y_m	전국_누적확진자	전국_월별확진자	제주_누적확진자	제주_월별확진자
0	2020-03	9786	4974	9	6
1	2020-04	10765	1018	13	4
2	2020-05	11468	703	15	2
3	2020-06	12800	1334	19	4
4	2020-07	14305	1506	26	7
5	2020-08	19947	5642	45	19
6	2020-09	23812	3865	59	14
7	2020-10	26511	2616	59	0
8	2020-11	34201	7689	80	21
9	2020-12	60740	26541	416	336
10	2021-01	78203	17492	522	106
11	2021-02	89674	11468	570	48
12	2021-03	103084	13400	625	55
13	2021-04	122007	18921	707	82
14	2021-05	140338	18307	1030	323
15	2021-06	156961	16603	1264	234
16	2021-07	198339	41382	1732	468
17	2021-08	251416	53077	2602	870

각 데이터 병합

- 1번 데이터 기준으로 병합 진행
- 결측치 0으로 대치

In [8]:

```
# 1번 3번 데이터 병합
df_m = pd.merge(df_1,df_3, how = "left", on = ["y_m","city","location"])
df_m
```

Out[8]:

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt	제주도 민_여	외국 인거 주_여	제주도 민_남	외국 인거 주_남	
0	2018-01	서귀포시	남원읍	52.0	9570	42437700	1270773	9306.0	200.0	9806.0	428.0	59
1	2018-01	서귀포시	대륜동	38.0	21666	57612600	1676850	6637.0	95.0	6836.0	83.0	29
2	2018-01	서귀포시	대정읍	89.0	10185	38885550	1164122	10725.0	677.0	10360.0	841.0	59
3	2018-01	서귀포시	대천동	37.0	20280	53858550	1593709	6475.0	137.0	6685.0	107.0	29
4	2018-01	서귀포시	동홍동	49.0	45936	118701000	3501286	11569.0	642.0	11124.0	624.0	38
...
1661	2021-06	제주시	일도2동	87.0	84360	147438200	4402149	16569.0	200.0	16077.0	75.0	71
1662	2021-06	제주시	조천읍	141.0	27732	63927750	1911187	12422.0	242.0	13017.0	279.0	61
1663	2021-06	제주시	한경면	71.0	8031	27060150	809898	4531.0	100.0	4627.0	237.0	34
1664	2021-06	제주시	한림읍	112.0	25653	82746990	2476292	10341.0	1140.0	10891.0	2090.0	64
1665	2021-06	제주시	화북동	84.0	66088	110750050	3306029	12238.0	161.0	12062.0	169.0	48

1666 rows × 14 columns

```
In [9]: # 병합 데이터, 4번 데이터 병합
df_m = pd.merge(df_m, df_4, how = "left", on = ["y_m", "city", "location"])
df_m
```

Out[9]:

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt	제주도 민_여	외국 인거 주_여	제주도 민_남	...	농축 수산물_결 제건수
0	2018-01	서귀포시	남원읍	52.0	9570	42437700	1270773	9306.0	200.0	9806.0	...	1753.0
1	2018-01	서귀포시	대륜동	38.0	21666	57612600	1676850	6637.0	95.0	6836.0	...	2026.0
2	2018-01	서귀포시	대정읍	89.0	10185	38885550	1164122	10725.0	677.0	10360.0	...	1959.0
3	2018-01	서귀포시	대천동	37.0	20280	53858550	1593709	6475.0	137.0	6685.0	...	1722.0
4	2018-01	서귀포시	동홍동	49.0	45936	118701000	3501286	11569.0	642.0	11124.0	...	4031.0
...
1661	2021-06	제주시	일도2동	87.0	84360	147438200	4402149	16569.0	200.0	16077.0	...	8846.0
1662	2021-06	제주시	조천읍	141.0	27732	63927750	1911187	12422.0	242.0	13017.0	...	3773.0
1663	2021-06	제주시	한경면	71.0	8031	27060150	809898	4531.0	100.0	4627.0	...	1040.0
1664	2021-06	제주시	한림읍	112.0	25653	82746990	2476292	10341.0	1140.0	10891.0	...	2075.0
1665	2021-06	제주시	화북동	84.0	66088	110750050	3306029	12238.0	161.0	12062.0	...	8098.0

1666 rows × 28 columns

```
In [10]: # 병합 데이터, 기상청 데이터 병합
df_m = pd.merge(df_m, df_5, how = "left", on = ["y_m", "location"])
df_m
```

Out[10]:

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt	제주도 민_여	외국 인거 주_여	제주도 민_남	...	식품_? 제건=
0	2018-01	서귀포시	남원읍	52.0	9570	42437700	1270773	9306.0	200.0	9806.0	...	9114.
1	2018-01	서귀포시	대륜동	38.0	21666	57612600	1676850	6637.0	95.0	6836.0	...	3007.
2	2018-01	서귀포시	대정읍	89.0	10185	38885550	1164122	10725.0	677.0	10360.0	...	5072.
3	2018-01	서귀포시	대천동	37.0	20280	53858550	1593709	6475.0	137.0	6685.0	...	2819.
4	2018-01	서귀포시	동홍동	49.0	45936	118701000	3501286	11569.0	642.0	11124.0	...	5428.
...
1661	2021-06	제주시	일도2동	87.0	84360	147438200	4402149	16569.0	200.0	16077.0	...	28784.
1662	2021-06	제주시	조천읍	141.0	27732	63927750	1911187	12422.0	242.0	13017.0	...	27706.
1663	2021-06	제주시	한경면	71.0	8031	27060150	809898	4531.0	100.0	4627.0	...	5802.
1664	2021-06	제주시	한림읍	112.0	25653	82746990	2476292	10341.0	1140.0	10891.0	...	15129.
1665	2021-06	제주시	화북동	84.0	66088	110750050	3306029	12238.0	161.0	12062.0	...	14638.

1666 rows × 32 columns

```
In [11]: # 병합 데이터, 코로나 데이터 병합
df_m = pd.merge(df_m, df_6, how = "left", on = ["y_m"])
df_m
```

Out[11]:

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt	제주도 민_여	외국 인거 주_여	제주도 민_남	...	식당_이 제건수
0	2018-01	서귀포시	남원읍	52.0	9570	42437700	1270773	9306.0	200.0	9806.0	...	17810.
1	2018-01	서귀포시	대륜동	38.0	21666	57612600	1676850	6637.0	95.0	6836.0	...	23440.
2	2018-01	서귀포시	대정읍	89.0	10185	38885550	1164122	10725.0	677.0	10360.0	...	29023.
3	2018-01	서귀포시	대천동	37.0	20280	53858550	1593709	6475.0	137.0	6685.0	...	21592.
4	2018-01	서귀포시	동홍동	49.0	45936	118701000	3501286	11569.0	642.0	11124.0	...	38565.
...
1661	2021-06	제주시	일도2동	87.0	84360	147438200	4402149	16569.0	200.0	16077.0	...	72515.
1662	2021-06	제주시	조천읍	141.0	27732	63927750	1911187	12422.0	242.0	13017.0	...	43707.
1663	2021-06	제주시	한경면	71.0	8031	27060150	809898	4531.0	100.0	4627.0	...	10358.
1664	2021-06	제주시	한림읍	112.0	25653	82746990	2476292	10341.0	1140.0	10891.0	...	30330.
1665	2021-06	제주시	화북동	84.0	66088	110750050	3306029	12238.0	161.0	12062.0	...	57304.

1666 rows × 36 columns

```
In [12]: df_m = pd.merge(df_m, df_2, how = "left", on = ["y_m", "city", "location"])
df_m
```

Out[12]:

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt	제주도 민_여	외국 인거 주_여	제주도 민_남	...	;
--	-----	------	----------	----------	--------	------	---------	------------	-----------------	------------	-----	---

	y_m	city	location	area_cnt	em_cnt	em_g	pay_amt	제주도 민_여	외국 인거 주_여	제주도 민_남	...	;
0	2018-01	서귀포시	남원읍	52.0	9570	42437700	1270773	9306.0	200.0	9806.0	...	6.256
1	2018-01	서귀포시	대륜동	38.0	21666	57612600	1676850	6637.0	95.0	6836.0	...	8.004
2	2018-01	서귀포시	대정읍	89.0	10185	38885550	1164122	10725.0	677.0	10360.0	...	5.417
3	2018-01	서귀포시	대천동	37.0	20280	53858550	1593709	6475.0	137.0	6685.0	...	8.004
4	2018-01	서귀포시	동홍동	49.0	45936	118701000	3501286	11569.0	642.0	11124.0	...	5.771
...
1661	2021-06	제주시	일도2동	87.0	84360	147438200	4402149	16569.0	200.0	16077.0	...	23.299
1662	2021-06	제주시	조천읍	141.0	27732	63927750	1911187	12422.0	242.0	13017.0	...	22.025
1663	2021-06	제주시	한경면	71.0	8031	27060150	809898	4531.0	100.0	4627.0	...	22.407
1664	2021-06	제주시	한림읍	112.0	25653	82746990	2476292	10341.0	1140.0	10891.0	...	22.537
1665	2021-06	제주시	화북동	84.0	66088	110750050	3306029	12238.0	161.0	12062.0	...	23.299

1666 rows × 39 columns

```
In [13]: #결측치 0으로 대치
df_m = df_m.fillna(0)
```

```
In [14]: #최종 병합 파일
df_m.to_csv("전체데이터_병합.csv",encoding = "cp949",index=False)
```

시계열 분석에 활용할 데이터 추출

```
In [15]: #최종 병합 파일에서 시계열 데이터 추출
df_x = df_m.loc[:,["y_m", "location", "em_g"]]
df_x = df_x.rename(columns={"location" : "emd_nm"})
df_x.head()
```

```
Out[15]:
```

	y_m	emd_nm	em_g
0	2018-01	남원읍	42437700
1	2018-01	대륜동	57612600
2	2018-01	대정읍	38885550
3	2018-01	대천동	53858550
4	2018-01	동홍동	118701000

시계열 분석을 위해 알수없음 지역 연도별 비연속성 제거를 위해 결측값 0으로 채우기

```
In [16]: # 알수없음 지역 데이터 분리
df_nul = df_x[df_x["emd_nm"]=="알수없음"]
df_x = df_x[df_x.emd_nm != "알수없음"]
```

```
In [17]: # 알수없음 지역 데이터프레임 새로 생성
temp = pd.DataFrame(columns=["y_m", "emd_nm"])
list_y = df_x["y_m"].unique()
temp.loc[:, "y_m"] = list_y
temp.loc[:, "emd_nm"] = "알수없음"
temp = pd.merge(temp, df_nul, how="left", on = ["y_m", "emd_nm"])
temp = temp.fillna(0)
temp
```

```
Out[17]:
```

	y_m	emd_nm	em_g
0	2018-01	알수없음	8880900.0
1	2018-02	알수없음	7543950.0
2	2018-03	알수없음	9287150.0
3	2018-04	알수없음	9853350.0
4	2018-05	알수없음	10290950.0
5	2018-06	알수없음	10533050.0
6	2018-07	알수없음	11860200.0
7	2018-08	알수없음	12184350.0
8	2018-09	알수없음	9085850.0
9	2018-10	알수없음	9371850.0
10	2018-11	알수없음	9997100.0
11	2018-12	알수없음	8593600.0
12	2019-01	알수없음	5044050.0

	y_m	emd_nm	em_g
13	2019-02	알수없음	0.0
14	2019-03	알수없음	0.0
15	2019-04	알수없음	0.0
16	2019-05	알수없음	0.0
17	2019-06	알수없음	0.0
18	2019-07	알수없음	0.0
19	2019-08	알수없음	0.0
20	2019-09	알수없음	482950.0
21	2019-10	알수없음	550350.0
22	2019-11	알수없음	621300.0
23	2019-12	알수없음	828350.0
24	2020-01	알수없음	645900.0
25	2020-02	알수없음	337600.0
26	2020-03	알수없음	0.0
27	2020-04	알수없음	0.0
28	2020-05	알수없음	0.0
29	2020-06	알수없음	836200.0
30	2020-07	알수없음	1372799.0
31	2020-08	알수없음	2387650.0
32	2020-09	알수없음	2010300.0
33	2020-10	알수없음	2226900.0
34	2020-11	알수없음	2088750.0
35	2020-12	알수없음	2147470.0
36	2021-01	알수없음	2057700.0
37	2021-02	알수없음	1985100.0
38	2021-03	알수없음	2153800.0
39	2021-04	알수없음	1936900.0
40	2021-05	알수없음	1910750.0
41	2021-06	알수없음	2157850.0

```
In [18]: # 시계열 데이터셋 병합
df_x = pd.concat([df_x,temp],axis=0)
df_x
```

```
Out[18]:
```

	y_m	emd_nm	em_g
--	-----	--------	------

	y_m	emd_nm	em_g
0	2018-01	남원읍	42437700.0
1	2018-01	대륜동	57612600.0
2	2018-01	대정읍	38885550.0
3	2018-01	대천동	53858550.0
4	2018-01	동홍동	118701000.0
...
37	2021-02	알수없음	1985100.0
38	2021-03	알수없음	2153800.0
39	2021-04	알수없음	1936900.0
40	2021-05	알수없음	1910750.0
41	2021-06	알수없음	2157850.0

1676 rows × 3 columns

```
In [19]: df_x.to_csv("행정동_월별_배출량.csv", encoding = "cp949", index = False)
```