

공공 데이터 분석 (with folium)

- 전국 전통시장 표준 데이터

01. 라이브러리 불러오기

In [3]:

```
import pandas as pd
import folium

# 기본 encoding은 utf-8, 에러 발생시 : euc-kr로 설정
market = pd.read_csv('전국전통시장표준데이터.csv', encoding='euc-kr')
market.head()
```

Out[3]:

인덱스	시장명	시장유형	소재지도로명주소	소재지번주소	시장개설주기	위도	경도	점포수	취급품목	사용가능상품권	홈페이지주소	주차장보유여부	공중화장실보유여부	개설연도	전화번호	데이터기준일자
0	해도5시장(영산시장)	상설장	경상북도포항시남구대동143-1	Nan	매일	36.012702	129.366493	40.0	농산물생필품,수산물,잡화등	포항사랑상품권,온누리상품권	Nan	Y	N	1990	Nan	2020-02-25
1	상대종합시장	상설장	경상북도포항시남구양학천로76번길13	Nan	매일	36.019933	129.361108	41.0	농산물생필품,수산물,잡화등	포항사랑상품권,온누리상품권	Nan	Y	N	1986	Nan	2020-02-25

	시장명	시장유형	소재지도로명주소	소재지지번주소	시장개설주기	위도	경도	점포수	취급품목	사용가능상품권	홈페이지주소	주차장보유여부	공중화장실보유여부	개설연도	전화번호	데이터기준일자
2	동해종합시장	상설장	경상북도 포항시 남구 해동435-2	Nan	매일	36.025808	129.370949	110.0	농산물생필품, 수산물, 잡화등	포항사랑상품권, 온누리상품권	Nan	Y	Y	1981	Nan	2020-02-25
3	송도시장	상설장	경상북도 포항시 남구 송도동440-6	Nan	매일	36.032826	129.375281	56.0	농산물생필품, 수산물, 잡화등	포항사랑상품권, 온누리상품권	Nan	Y	Y	1982	Nan	2020-02-25
4	청림시장	상설장	경상북도 포항시 남구 청림동588-2	Nan	매일	35.996576	129.410356	13.0	농산물생필품, 수산물, 잡화등	포항사랑상품권, 온누리상품권	Nan	Y	N	1985	Nan	2020-02-25

In [4]:

```
### 데이터 살펴보기
print(market.columns)
print(market.shape)
print(market.info())
```

Index(['시장명', '시장유형', '소재지도로명주소', '소재지지번주소', '시장개설주기', '위도', '경도', '점포수', '취급품목', '사용가능상품권', '홈페이지주소', '공중화장실보유여부', '주차장보유여부', '개설연도', '전화번호',

```

'dealer_gjzilja', 'dealer_gjcode', 'dealer_gjname'],
      dtype='object')
(3489, 18)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3489 entries, 0 to 3488
Data columns (total 18 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   시장명        3489 non-null   object 
 1   시장유형      3489 non-null   object 
 2   소재지도로명주소  3197 non-null   object 
 3   소재지지번주소  3088 non-null   object 
 4   시장개설주기    3489 non-null   object 
 5   위도          3411 non-null   float64
 6   경도          3411 non-null   float64
 7   점포수        3406 non-null   float64
 8   취급품목      3275 non-null   object 
 9   사용가능상품권  1749 non-null   object 
 10  홈페이지주소    485 non-null   object 
 11  공중화장실보유여부 3489 non-null   object 
 12  주차장보유여부    3489 non-null   object 
 13  개설연도      1957 non-null   object 
 14  전화번호      1835 non-null   object 
 15  데이터기준일자  3489 non-null   object 
 16  제공기관코드    3489 non-null   object 
 17  제공기관명      3489 non-null   object 

dtypes: float64(3), object(15)
memory usage: 490.8+ KB
None

```

In [5]: `market.isnull().sum()`

```

Out[5]: 시장명          0
시장유형          0
소재지도로명주소  292
소재지지번주소    401
시장개설주기      0
위도            78
경도            78
점포수          83
취급품목          214
사용가능상품권    1740
홈페이지주소     3004
공중화장실보유여부 0
주차장보유여부    0
개설연도      1532
전화번호      1654
데이터기준일자    0
제공기관코드    0
제공기관명      0
dtype: int64

```

전체 2882개 데이터중에 107개 정도가 결측치

전체 3489개 데이터중에 78개 정도가 결측치

In [6]:

```

market_name = market['시장명']
market_lati = market['위도']
market_long = market['경도']

market.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3489 entries, 0 to 3488
Data columns (total 18 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   시장명        3489 non-null   object 
 1   시장유형      3489 non-null   object 
 2   소재지도로명주소  3197 non-null   object 
 3   소재지지번주소  3088 non-null   object 
 4   시장개설주기    3489 non-null   object 
 5   위도          3411 non-null   float64
 6   경도          3411 non-null   float64
 7   점포수        3406 non-null   float64
 8   취급품목      3275 non-null   object 
 9   사용가능상품권  1749 non-null   object 
 10  홈페이지주소    485 non-null   object 
 11  공중화장실보유여부 3489 non-null   object 
 12  주차장보유여부    3489 non-null   object 
 13  개설연도      1957 non-null   object 
 14  전화번호      1835 non-null   object 
 15  데이터기준일자  3489 non-null   object 
 16  제공기관코드    3489 non-null   object 
 17  제공기관명      3489 non-null   object 

dtypes: float64(3), object(15)
memory usage: 490.8+ KB
None

```

#	Column	Non-Null Count	Dtype
0	시장명	3489	non-null object
1	시장유형	3489	non-null object
2	소재지도로명주소	3197	non-null object
3	소재지지번주소	3088	non-null object
4	시장개설주기	3489	non-null object
5	위도	3411	non-null float64
6	경도	3411	non-null float64
7	점포수	3406	non-null float64
8	취급품목	3275	non-null object
9	사용가능상품권	1749	non-null object
10	홈페이지주소	485	non-null object
11	공중화장실보유여부	3489	non-null object
12	주차장보유여부	3489	non-null object
13	개설연도	1957	non-null object
14	전화번호	1835	non-null object
15	데이터기준일자	3489	non-null object
16	제공기관코드	3489	non-null object
17	제공기관명	3489	non-null object

dtypes: float64(3), object(15)
memory usage: 490.8+ KB

02. 데이터 살펴보기

결측치를 제외한 행만 가져오기

```
In [7]: market['위도'].notnull()
```

```
Out[7]: 0      True
1      True
2      True
3      True
4      True
...
3484    True
3485    True
3486    True
3487    True
3488    True
Name: 위도, Length: 3489, dtype: bool
```

```
In [8]: market_1 = market.loc[market['위도'].notnull(), :]
market_1.shape
```

```
Out[8]: (3411, 18)
```

```
In [9]: market_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3411 entries, 0 to 3488
Data columns (total 18 columns):
 #   Column       Non-Null Count  Dtype  
--- 
 0   시장명       3411 non-null   object 
 1   시장유형     3411 non-null   object 
 2   소재지도로명주소  3136 non-null   object 
 3   소재지지번주소  3030 non-null   object 
 4   시장개설주기   3411 non-null   object 
 5   위도          3411 non-null   float64
 6   경도          3411 non-null   float64
 7   점포수        3328 non-null   float64
 8   취급품목      3197 non-null   object 
 9   사용가능상품권  1685 non-null   object 
```

```
10 홈페이지주소      461 non-null    object
11 공중화장실보유여부 3411 non-null    object
12 주차장보유여부   3411 non-null    object
13 개설연도        1891 non-null    object
14 전화번호        1777 non-null    object
15 데이터기준일자  3411 non-null    object
16 제공기관코드    3411 non-null    object
17 제공기관명     3411 non-null    object
dtypes: float64(3), object(15)
memory usage: 506.3+ KB
```

```
In [10]: market_1.columns
```

```
Out[10]: Index(['시장명', '시장유형', '소재지도로명주소', '소재지지번주소', '시장개설주기', '위도', '경도', '점포수',  
              '취급품목', '사용가능상품권', '홈페이지주소', '공중화장실보유여부', '주차장보유  
여부', '개설연도', '전화번호',  
              '데이터기준일자', '제공기관코드', '제공기관명'],  
              dtype='object')
```

```
In [11]: #### 필요한 정보(특징)만 가져오기
sel = ['시장명', '시장유형', '시장개설
        '사용가능상품권', '공중화장실보
        도',
        '데이터기준일자', '제공기관명']

market_2 = market_1[sel]
market_2
```

	시장명	시장 유형	시장 개설 주기	위도	경도	사용 가능 상품권	공증화 장실 보유여부	주차장 보유여부	개설 연도	데이터 기준일자	제공 기관명
0	해도5시장(형산시장)	상설장	매일	36.012702	129.366493	포항사랑 상품권, 온 누리상품권	Y	N	1990	2020-02-25	경상북도
1	상대종합시장	상설장	매일	36.019933	129.361108	포항사랑 상품권, 온 누리상품권	Y	N	1986	2020-02-25	경상북도
2	동해종합시장	상설장	매일	36.025808	129.370949	포항사랑 상품권, 온 누리상품권	Y	Y	1981	2020-02-25	경상북도
3	송도시장	상설장	매일	36.032826	129.375281	포항사랑 상품권, 온 누리상품권	Y	Y	1982	2020-02-25	경상북도
4	청림시장	상설장	매일	35.996576	129.410356	포항사랑 상품권, 온 누리상품권	Y	N	1985	2020-02-25	경상북도

	시장명	시장 유형	시장 개설 주기	위도	경도	사용 가능 상품권	공중화장실 보유여부	주차장 보유여부	개설연도	데이터 준기준일자	제공기관명
3484	수영팔도 시장	상설장	매일	35.168782	129.116897	NaN	Y	Y	NaN	2021-10-26	소상공인시장진흥공단
3485	현대종합 상가시장	상설장	매일	35.168408	129.120732	NaN	Y	Y	NaN	2021-10-26	소상공인시장진흥공단
3486	거성시장	상설장	매일	35.190006	129.068220	NaN	Y	N	NaN	2021-10-26	소상공인시장진흥공단
3487	거제시장	상설장	매일	35.181932	129.072793	NaN	Y	N	NaN	2021-10-26	소상공인시장진흥공단
3488	연동시장	상설장	매일	35.188843	129.092523	NaN	Y	Y	NaN	2021-10-26	소상공인시장진흥공단

3411 rows × 11 columns

03. 위치 정보를 이용한 시각화

In [12]:

```
market_name = market_2['시장명']
market_lat = market_2['위도']
market_long = market_2['경도']
```

In [13]:

```
import numpy as np

market_lat = list(market_lat)
market_long = list(market_long)
market_loc = np.array([market_lat, market_long]).T
print(market_loc.shape)
print(np.mean(market_lat), np.mean(market_long))
market_loc
```

(3411, 2)

36.20206678435737 127.85249836869701

Out[13]: array([[36.0127023 , 129.366493],
 [36.01993305, 129.36110774],
 [36.02580836, 129.3709491],

```
[...,
[ 35.19000574, 129.0682199 ],
[ 35.18193157, 129.0727929 ],
[ 35.18884338, 129.0925228 ]])
```

In [14]:

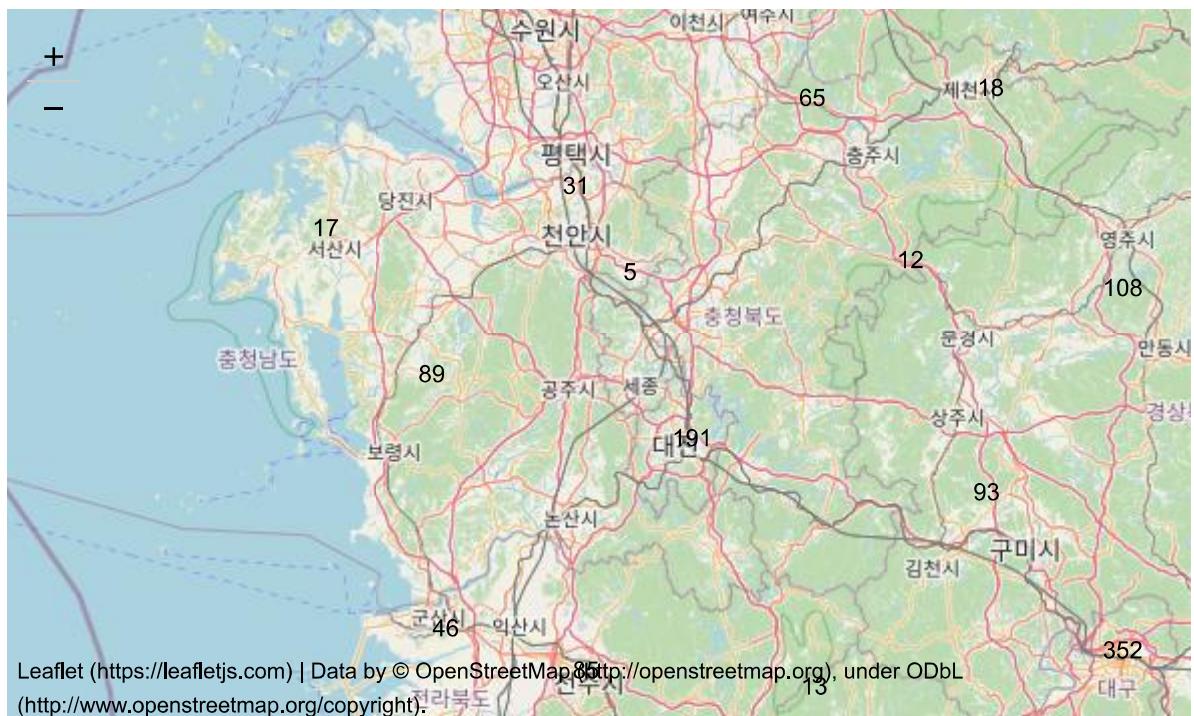
```
from folium import plugins
import os

market_map = folium.Map(location=[ np.mean( market_lat ),  
np.mean(market_long) ],  
zoom_start=8)

market_name = list(market_name)
plugins.MarkerCluster(market_loc, popups=market_name).add_to(market_map)

market_map.save(os.path.join('..', 'Market_location.html'))  
market_map
```

Out[14]:



In []: