

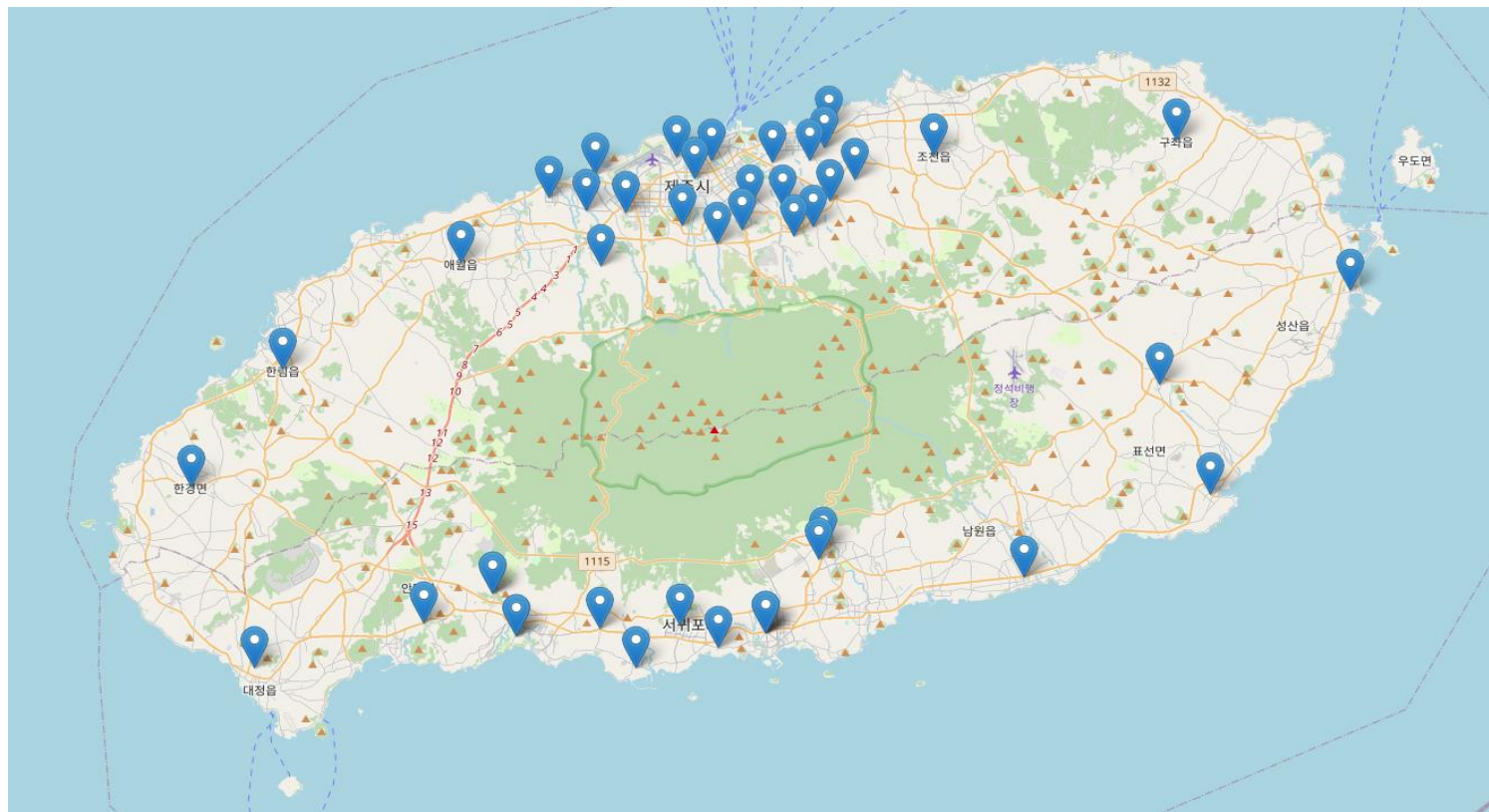
날씨 데이터를 활용한 태양광 발전량 예측

4강. 탐색적 데이터 분석 I



1. 태양광 발전소 현황 전처리
2. 태양광 발전소 위도 경도 구하기
3. 태양광 발전소 위치 시각화

제주도의 태양광 발전소 현황을 지도에 시각화



위의 그림과 같이 태양광 발전소의 위치를 지도 상에
표기하기 위해서 어떤 데이터를
어디서 구할 수 있을까요?



발전소 위치 정보
설비용량
사업개시일



제주도 태양광 발전소 현황 데이터 받기

DATA 공공데이터포털
. GO . KR

데이터찾기국가데이터맵데이터요청데이터활용정보공유이용안내

데이터 상세

 URL 복사

제주특별자치도_태양광발전소현황
제주특별자치도에 위치한 태양광 발전소와 관련한 데이터로 행정시, 읍면동, 허가일자, 상호, 설비용량(KW), 상태, 사업개시일 정보를 제공합니다.

 0  0  관심

파일데이터

오픈API

추천데이터

공공데이터활용지원센터는 공공데이터포털에 개방되는 3단계 이상의 오픈 포맷 파일데이터를 오픈 API(RestAPI 기반의 JSON/XML)로 자동변환하여 제공합니다.
오픈 API를 활용하기 위해서는 공공데이터포털 회원 가입 및 활용신청이 필요하며, 활용 관련 문의는 공공데이터활용지원센터로 연락주시기 바라며,
데이터 자체에 대한 문의는 아래 제공기관의 관리부서 전화번호로 연락주시기 바랍니다.
파일데이터는 로그인 없이 다운로드를 통해 이용하실 수 있습니다.

버튼을 클릭

CSV 제주특별자치도_태양광발전소현황

다운 받는 파일은 csv 포맷입니다

 다운로드  오류신고 및 문의

CSV 파일이란?

위키피디아에서는 다음과 같이 소개 합니다.

CSV (파일 형식)

문서 [토론](#)

위키백과, 우리 모두의 백과사전.

CSV([영어](#): comma-separated values)는 몇 가지 필드를 [쉼표](#)(,)로 구분한 [텍스트](#) 데이터 및 텍스트 파일이다. 확장자는 .csv이며 [MIME 형식](#)은 text/csv이다. **comma-separated variables**라고도 한다.

오래전부터 [스프레드시트](#)나 [데이터베이스](#) 소프트웨어에서 많이 쓰였으나 세부적인 [구현](#)은 소프트웨어에 따라 다르다. 그것들을 추가한 형태가 2005년 10월 RFC 4180에서 Informational([IESG](#)의 외부에서 결정된 유용한 정보의 제공)로 사양이 문서화됐다.

```
연도,제조사,모델,설명,가격
1997,Ford,E350,"ac, abs, moon",3000.00
1999,Chevy,"Venture ""Extended Edition""",,,4900.00
1999,Chevy,"Venture ""Extended Edition, Very Large""",,5000.00
1996,Jeep,Grand Cherokee,"MUST SELL!air, moon roof, loaded",4799.00
```



연도	제조사	모델	설명	가격
1997	Ford	E350	ac, abs, moon	3000.00
1999	Chevy	Venture "Extended Edition"		4900.00
1999	Chevy	Venture "Extended Edition, Very Large"		5000.00
1996	Jeep	Grand Cherokee	MUST SELL! air, moon roof, loaded	4799.00



제주도 태양광 발전소 현황 데이터 읽기

Encoding을 'cp949'로 설정해야
한글이 깨지지 않고 정상적으로 출력됩니다.

```
# 공공데이터포털에서 제주특별자치도_태양광발전소현황 파일 다운로드  
# https://www.data.go.kr/data/3082724/fileData.do  
df = pd.read_csv('제주특별자치도_태양광발전소현황_20220630.csv', encoding='cp949')  
df
```

	행정시	읍면동	허가일자	상호	설비용량(KW)	상태	사업개시일	데이터기준일자
0	제주시	구좌읍	1998-06-10	행원풍력발전단지(1차)	3480.0	사업개시	2000-04-05	2022-06-30
1	제주시	회천동	2002-12-09	파낙스에너지(주)	1000.0	사업개시	2002-12-18	2022-06-30
2	제주시	한경면	2003-04-14	한경풍력 1단계	6000.0	사업개시	2004-02-28	2022-06-30
3	제주시	한경면	2005-04-04	신창풍력발전단지	1700.0	사업개시	2006-03-03	2022-06-30
4	제주시	구좌읍	2005-08-30	제주월정풍력발전소	1500.0	사업개시	2006-07-20	2022-06-30
...	...	...	...	...	...	...	...	...
2143	제주시	한경면	2022-06-08	민재 태양광발전소	99.6	인허가	NaN	2022-06-30
2144	제주시	한림읍	2022-06-08	효일4호 태양광발전소	99.0	인허가	NaN	2022-06-30
2145	제주시	한림읍	2022-06-08	효일3호 태양광발전소	99.0	인허가	NaN	2022-06-30
2146	제주시	한림읍	2022-06-08	효일5호 태양광발전소	99.0	인허가	NaN	2022-06-30
2147	서귀포시	성산읍	2022-06-21	한라2호소수력발전소	210.0	인허가	NaN	2022-06-30

2148 rows x 8 columns



결측 값 확인하기

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2148 entries, 0 to 2147
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   행정시          2148 non-null   object
1   읍면동          2148 non-null   object
2   허가일자        2148 non-null   object
3   상호            2148 non-null   object
4   설비용량(KW)    2148 non-null   float64
5   상태            2148 non-null   object
6   사업개시일      1571 non-null   object
7   데이터기준일자  2148 non-null   object
dtypes: float64(1), object(7)
memory usage: 134.4+ KB
```

사업개시일 컬럼의 경우
결측 값을 확인 할 수 있습니다.

2143	제주시	한경면	2022-06-08	민재 태양광발전소	99.6	인허가	NaN	2022-06-30
2144	제주시	한림읍	2022-06-08	효일4호 태양광발전소	99.0	인허가	NaN	2022-06-30
2145	제주시	한림읍	2022-06-08	효일3호 태양광발전소	99.0	인허가	NaN	2022-06-30
2146	제주시	한림읍	2022-06-08	효일5호 태양광발전소	99.0	인허가	NaN	2022-06-30
2147	서귀포시	성산읍	2022-06-21	한라2호소수력발전소	210.0	인허가	NaN	2022-06-30

결측 값의 경우 인허가 상태임을 확인할 수 있습니다.



결측 값 제거

```
# 사업개시일이 있는 데이터만 골라낸다.
# 사업개시일이 없는 데이터는 아직 인허가가 진행중인 것이므로 drop
df = df.dropna(axis=0)
df
```

사업 개시일이 없는 데이터는 삭제

	행정시	읍면동	허가일자	상호	설비용량(KW)	상태	사업개시일	데이터기준일자
0	제주시	구좌읍	1998-06-10	행원풍력발전단지(1차)	3480.00	사업개시	2000-04-05	2022-06-30
1	제주시	회천동	2002-12-09	파낙스에너지(주)	1000.00	사업개시	2002-12-18	2022-06-30
2	제주시	한경면	2003-04-14	한경풍력 1단계	6000.00	사업개시	2004-02-28	2022-06-30
3	제주시	한경면	2005-04-04	신창풍력발전단지	1700.00	사업개시	2006-03-03	2022-06-30
4	제주시	구좌읍	2005-08-30	제주월정풍력발전소	1500.00	사업개시	2006-07-20	2022-06-30
...	...	...	...	...	...	...	...	...
2050	서귀포시	안덕면	2021-04-19	썬 태양광발전소	182.00	사업개시	2022-01-03	2022-06-30
2051	서귀포시	안덕면	2021-04-19	청춘 태양광발전소	662.48	사업개시	2022-01-03	2022-06-30
2055	서귀포시	표선면	2021-04-21	제주덕암 태양광발전소	99.18	사업개시	2022-01-03	2022-06-30
2083	서귀포시	안덕면	2021-08-03	태양7호태양광발전소	198.88	사업개시	2021-12-10	2022-06-30
2091	제주시	구좌읍	2021-09-15	홍산태양광발전소	999.63	사업개시	2022-04-25	2022-06-30

1571 rows × 8 columns



상태 이상치 수정

```
# 사업개시일이 기록되었는데 상태가 '공사진행', '인허가' 인 경우
df[df['상태'] != '사업개시']
```

	행정시	읍면동	허가일자	상호	설비용량(KW)	상태	사업개시일	데이터기준일자
6	제주시	한경면	2006-08-11	탐라해상풍력발전소	30000.0	공사진행	2017-09-16	2022-06-30
29	서귀포시	성산읍	2008-02-28	성산풍력 1단계	12000.0	공사진행	2009-03-30	2022-06-30
583	제주시	한경면	2016-11-16	성용유화발전	750.0	인허가	2016-11-16	2022-06-30

아래 홈페이지를 보면 '탐라해상풍력발전소'는 17년 9월에 건설을 마치고 사업을 시작한 것으로 보인다.

<http://tamra-owp.co.kr/2019/sub0201.php>

성산풍력 1단계도 준공 기사를 확인할 수 있다. <https://jejunews.com/news/articleView.html?idxno=279219>

따라서 상태를 '사업개시'로 수정하자.

데이터를 살펴보고 데이터의 이상 유무를 판단하는 것도 중요합니다.

허가일자와 사업개시일이 기록되었는데 상태가 '공사진행', '인허가'인 경우 모두 사업개시로 수정해도 될 듯 합니다.

그 외에도 운영중임을 뒷받침하는 내용을 홈페이지나 인터넷 뉴스 기사로 확인할 수도 있습니다.



날짜 타입으로 타입 변경

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 1571 entries, 0 to 2091  
Data columns (total 7 columns):  
#   Column          Non-Null Count  Dtype    
--  --            --           --   --     
0   행정시          1571 non-null   object   
1   읍면동          1571 non-null   object   
2   상호           1571 non-null   object   
3   설비용량(KW)    1571 non-null   float64  
4   상태            1571 non-null   object   
5   사업개시일      1571 non-null   object   
6   데이터기준일자  1571 non-null   object   
dtypes: float64(1), object(6)  
memory usage: 130.5+ KB
```



```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1571 entries, 0 to 1570  
Data columns (total 7 columns):  
#   Column          Non-Null Count  Dtype    
--  --            --           --   --     
0   행정시          1571 non-null   object   
1   읍면동          1571 non-null   object   
2   상호           1571 non-null   object   
3   설비용량(KW)    1571 non-null   float64  
4   상태            1571 non-null   object   
5   사업개시일      1571 non-null   datetime64[ns]  
6   데이터기준일자  1571 non-null   object   
dtypes: datetime64[ns](1), float64(1), object(5)  
memory usage: 86.0+ KB
```

더 이상 결측 값이 존재하지 않습니다.
사업개시일 컬럼이 문자열로 되어 있습니다.
날짜 타입으로 변경해줍니다.



상태 이상치 수정

df

	행정시	읍면동	상호	설비용량(KW)	상태	사업개시일	데이터기준일자
0	제주시	구좌읍	행원풍력발전단지(1차)	3480.00	사업개시	2000-04-05	2022-06-30
1	제주시	회천동	파낙스에너지(주)	1000.00	사업개시	2002-12-18	2022-06-30
2	제주시	한경면	한경풍력 1단계	6000.00	사업개시	2004-02-28	2022-06-30
3	제주시	한경면	신창풍력발전단지	1700.00	사업개시	2006-03-03	2022-06-30
4	제주시	한경면	오복산전	59.87	사업개시	2006-07-11	2022-06-30
...	...	...	...	...	...	...	...
1566	서귀포시	남원읍	다연솔라	999.60	사업개시	2022-05-31	2022-06-30
1567	제주시	한경면	태광발전소	99.60	사업개시	2022-06-02	2022-06-30
1568	제주시	한경면	맑음에너지	99.60	사업개시	2022-06-02	2022-06-30
1569	제주시	한경면	더썬에너지	99.60	사업개시	2022-06-02	2022-06-30
1570	서귀포시	남원읍	미소3호태양광발전소	999.60	사업개시	2022-06-07	2022-06-30

1571 rows x 7 columns

그런데, 분명 태양광 발전소 파일을 받았는데
상호를 보니 풍력도 섞여 있습니다.
신재생 에너지를 다 포함하고 있는 듯 합니다.



제주도 신재생 에너지 발전시설 현황 받기

DATA 공공데이터포털
.GO.KR

데이터찾기국가데이터맵데이터요청데이터활용정보공유이용안내

데이터 상세

 URL 복사

제주특별자치도 신재생에너지발전시설현황
제주특별자치도 내에 소재하고 있는 신재생 에너지 발전시설의 허가일자, 상호, 설비용량(KW), 설치지역, 원동력종류, 사업개시일 등의 정보를 제공합니다.

 0  0  관심

파일데이터

오픈API

추천데이터

공공데이터활용지원센터는 공공데이터포털에 개방되는 3단계 이상의 오픈 포맷 파일데이터를 오픈 API(RestAPI 기반의 JSON/XML)로 자동변환하여 제공합니다.
오픈 API를 활용하기 위해서는 공공데이터포털 회원 가입 및 활용신청이 필요하며, 활용 관련 문의는 공공데이터활용지원센터로 연락주시기 바라며,
데이터 자체에 대한 문의는 아래 제반기관의 관리부서 전화번호로 연락주시기 바랍니다.
파일데이터는 로그인 없이 다운로드를 통해 이용하실 수 있습니다.

CSV

 제주특별자치도 신재생에너지발전시설현황

 다운로드  오류신고 및 문의

이번엔 공공데이터 포털에서 제주도 신재생 에너지 발전 시설 현황을 다운 받습니다.

제주도 신재생 에너지 발전시설 현황 읽기

```
df2 = pd.read_csv('제주특별자치도_신재생에너지발전시설현황_20200120.csv', encoding='cp949')
df2
```

	허가일자	상호	설비용량(KW)	설치행정시	설치지역	원동력종류	사업개시일	구분	데이터기준일자
0	2018-04-04	(유)나월(나월태양광발전소)	496.80	제주시	애월읍	태양광	2019-04-26	육상	2020-01-20
1	2014-10-07	(유)메가솔라 태양광발전소	1800.00	제주시	구좌읍	태양광	2015-07-01	없음	2020-01-20
2	2014-05-15	(유)무한에코에너지(서민행복8호태양광발전소)	516.00	서귀포시	남원읍	태양광	2014-10-16	없음	2020-01-20
3	2014-12-09	(유)용흥마을태양광발전소	99.00	서귀포시	강정동	태양광	2015-10-29	없음	2020-01-20
4	2014-10-14	(유)제주 늘해랑	2198.00	서귀포시	성산읍	태양광	2019-01-31	없음	2020-01-20
...	...	...	...	...	...	...	...	...	...
836	2018-09-19	후니3 태양광발전소	93.15	서귀포시	표선면	태양광	2019-03-04	육상	2020-01-20
837	2014-10-28	희운2 태양광발전소	99.00	서귀포시	성산읍	태양광	2015-06-22	없음	2020-01-20
838	2014-10-29	희운4 태양광발전소	99.28	서귀포시	성산읍	태양광	2019-07-11	없음	2020-01-20
839	2014-10-29	희운5 태양광발전소	99.28	서귀포시	성산읍	태양광	2019-07-11	없음	2020-01-20
840	2014-10-29	희운태양광발전소	99.28	서귀포시	성산읍	태양광	2019-07-11	없음	2020-01-20

841 rows x 9 columns

그런데, 여기서 하나의 문제가 발생합니다. 앞에서 전처리 했던 csv파일과 새로 읽어온 csv파일의 태양광 발전소의 수와 56개 차이가 나는 것을 확인할 수 있습니다. 두 파일 모두 공공데이터 포털에서 받았는데 어떤 파일을 신뢰해야 할까요?



제주도 태양광 발전소 데이터의 신뢰성 확인

2019년 12월 기준으로 신뢰할 만한 제주도내 태양광 발전소의 수와 설비용량 정보를 확인하기 위해 아래 자료를 참조 하였습니다.



이 문서를 참고한 결과 2번째로 읽어온 csv의 데이터가 더 신뢰할 만한 데이터임을 확인할 수 있다.
이 pdf문서는 제주에너지공사에서 다운 받았는데,
이 문서 뿐만 아니라 제주도 태양광 발전량 일데이터도 제공하고 있습니다.
따라서, 811개소, 245KW를 기준으로 정리 합니다.

<표 V-34> 태양광 현황(총괄) (2019년 12월 기준)

구분	합 계		운전중		추진중	
	용량(kW)	개소	용량(kW)	개소	용량(kW)	개소
계	639,462	1,860	245,321	811	394,141	1,049

* 자료: 제주특별자치도 내부자료

제주도 신재생 에너지 발전시설 현황 받기

```
df2 = df2[df2['원동력종류']=='태양광']
df2.reset_index(drop=True, inplace=True)
df2
```

	허가일자	상호	설비용량(KW)	설치행정시	설치지역	원동력종류	사업개시일	구분	데이터기준일자
0	2018-04-04	(유)나월(나월태양광발전소)	496.80	제주시	애월읍	태양광	2019-04-26	육상	2020-01-20
1	2014-10-07	(유)메가솔라 태양광발전소	1800.00	제주시	구좌읍	태양광	2015-07-01	없음	2020-01-20
2	2014-05-15	(유)무한에코에너지(서민행복8호태양광발전소)	516.00	서귀포시	남원읍	태양광	2014-10-16	없음	2020-01-20
3	2014-12-09	(유)용흥마을태양광발전소	99.00	서귀포시	강정동	태양광	2015-10-29	없음	2020-01-20
4	2014-10-14	(유)제주 늘해랑	2198.00	서귀포시	성산읍	태양광	2019-01-31	없음	2020-01-20
...	...	...	...	...	...	...	...	...	...
806	2018-09-19	후니3 태양광발전소	93.15	서귀포시	표선면	태양광	2019-03-04	육상	2020-01-20
807	2014-10-28	희운2 태양광발전소	99.00	서귀포시	성산읍	태양광	2015-06-22	없음	2020-01-20
808	2014-10-29	희운4 태양광발전소	99.28	서귀포시	성산읍	태양광	2019-07-11	없음	2020-01-20
809	2014-10-29	희운5 태양광발전소	99.28	서귀포시	성산읍	태양광	2019-07-11	없음	2020-01-20
810	2014-10-29	희운태양광발전소	99.28	서귀포시	성산읍	태양광	2019-07-11	없음	2020-01-20

811 rows × 9 columns

```
df2['설비용량(KW)'].sum()
```

245321.82

문서에서 확인한 것 같이 811개소, 245KW입니다.



Geopy패키지로 위도 경도 변환

geopy는 '주소' 컬럼을 사용하고 '시'는 삭제해야 정상적으로 좌표값을 얻을 수 있다.

```
# 가입 없이 주소->좌표 변환
from geopy.geocoders import Nominatim

unique_addr=[]
unique_lat=[]
unique_lng=[]

def geocoding(address):
    geocoder = Nominatim(user_agent = 'South Korea', timeout=None)
    geo = geocoder.geocode(address)
    if geo==None:
        print('geo is null')
        return 0, 0
    else:
        return geo.latitude, geo.longitude

for addr in geo_pos:
    print(addr)
    lat, lng = geocoding(addr)
    if lat==0 or lng==0:
        continue
    unique_addr.append(addr)
    unique_lat.append(lat)
    unique_lng.append(lng)
    print(lat, lng)
```

한글 주소가 완벽하게 호환되지 않아
에러가 발생합니다.
에러가 발생할 경우 별도로 처리합니다.

제주 애월읍
33.45079 126.3752



Geopy패키지로 위도 경도 변환

geopy에서 한글 주소로 검색되지 않는 목록은 영문으로 입력할 경우 위경도를 얻어올 수 있다.

```
ko_addr = ['서귀포 안덕면', '서귀포 남원읍', '서귀포 도순동']
eng_addr = ["Andeok-myeon, Seogwipo-si", 'Namwon-eup, Seogwipo-si', '29beon-gil, Seogwipo-si']

index = 0
for addr in eng_addr:
    print(ko_addr[index])
    lat, lng = geocoding(addr)
    if lat==0 or lng==0:
        continue
    unique_addr.append(ko_addr[index])
    unique_lat.append(lat)
    unique_lng.append(lng)
    print(lat, lng)
    index+=1
```

```
서귀포 안덕면
33.2572438 126.3525278
서귀포 남원읍
33.2817661 126.71760849622656
서귀포 도순동
33.3854321 126.8000815
```



변환된 위도 경도 데이터프레임에 저장하기

```
df_unique = pd.DataFrame({'주소': unique_addr,  
                           '경도(x)': unique_lng,  
                           '위도(y)': unique_lat})
```

df_unique

	주소	경도(x)	위도(y)
0	제주 애월읍	126.375200	33.450790
1	제주 구좌읍	126.810550	33.516620
2	서귀포 강정동	126.481362	33.232835
3	서귀포 성산읍	126.916280	33.435961
4	제주 한경면	126.210660	33.330500
5	서귀포 표선면	126.831045	33.326671
6	제주 회천동	126.614790	33.495600
7	서귀포 대정읍	126.249599	33.233025



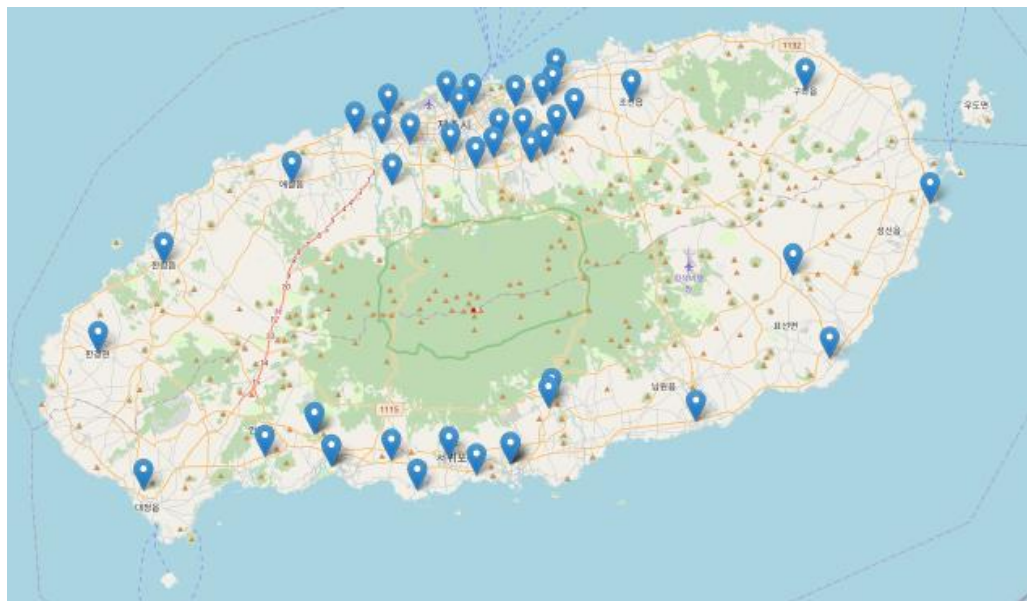
Folium으로 지도 위에 마커 표시하기

```
import pandas as pd
import folium

jeju_map = folium.Map(location=[33.38, 126.55], zoom_start=11)

for i in range(len(pos_df)):
    pos_list = [pos_df.iloc[i][1], pos_df.iloc[i][0]]
    folium.Marker(pos_list, popup=pos_df.index[i]).add_to(jeju_map)

jeju_map.save('jeju.html')
```



수고하셨습니다

× ×
× ×
× ×