

빅데이터 활용

<참고자료1. : 모두의 데이터분석 with 파이썬 / 길벗출판사>

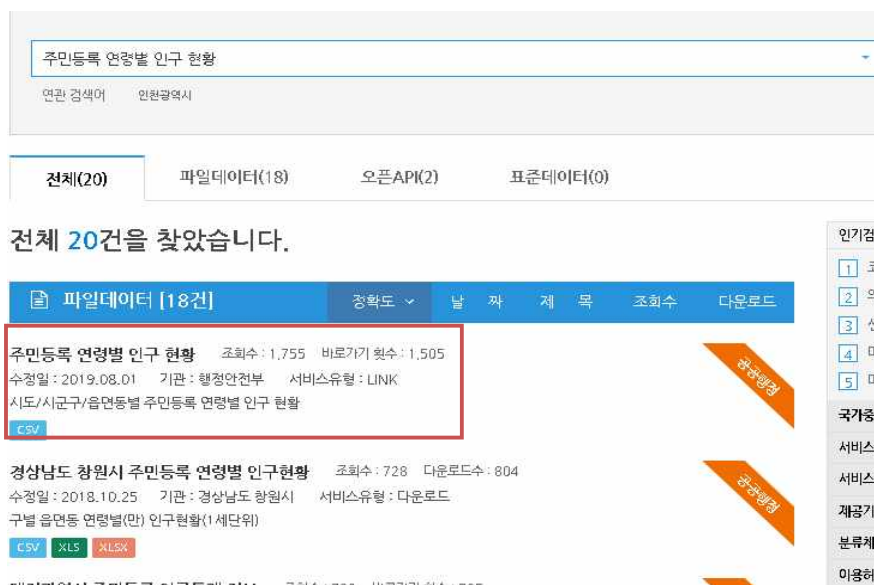
<참고자료2. : '빅데이터 활용' 숭실대학교 류현수 교수님>

[정부 데이터 다운로드 방법]

1. www.data.go.kr 접속
2. 검색 창에서 '주민등록 연령별 인구 현황' 검색



3. 검색 결과 화면에서 '주민등록 연령별 인구 현황' csv 클릭



4. [바로가기] 클릭

메뉴 / 네비게이션 / 쪽별네비게이션

주민등록 연령별 인구 현황

ENGLISH

시도/시군구/읍면동별 주민등록 연령별 인구 현황

마지막유형 : 텍스트 파일, 링크 건수 : 1 전체 행 수 : N/A 확장자 : CSV 다운로드 횟수(바로가기 횟수) : 1505

☐ 전체 선택 다운로드

※ 서비스 오류가 있을시 오류신고 버튼을 이용해주세요.

☐ LINK 주민등록 연령별 인구 현황

바로가기

닫기 추가 ★

주민등록 연령별 인구 현황			
업데이트 주기	수시	차기등록예정일	2020-07-19
비용부과유무	무료	비용부과기준 및 단위	없음
바로가기 횟수	1505		
등록일	2019-07-19	수정일	2019-08-01
이용허락범위	이용허락범위 제한 없음		

아래의 화면에서

5. '연령별 인구현황' 클릭
6. 구분 : '계' 선택
7. 만 연령 구분 : 0, 100세이상
8. 하단의 '전체 읍면동 현황' 선택
9. [검색] 클릭
10. 'CSV 파일 다운로드' <<< 몇 초가 지난 후 다운로드 완료

주민등록 인구통계

- ▶ 주민등록 인구 및 세대현황
- ▶ **연령별 인구현황**
- ▶ 주민등록 인구 기타현황

연령별 인구현황

통계표
그래프

행정구역

전국 ▼ 시·군·구 ▼ ⓘ

등록구분

전체 ▼ ⓘ

조회기간

☒ 월간 ☐ 연간 2020년 ▼ 01월 ▼ ~ 2020년 ▼ 01월 ▼ ⓘ

※ 매월 말일 작성 / 공표일시 : 매월 1일 12시 이후(공표일이 주말, 공휴일인 경우에는 다음 평일에 공표)

구분

☒ 계 ☐ 남·여 구분

정렬순서

행정기관코드 ▼ 오름차순 ▼

연령 구분 단위

1세 ▼

만 연령구분

0 ▼ 100이상 ▼

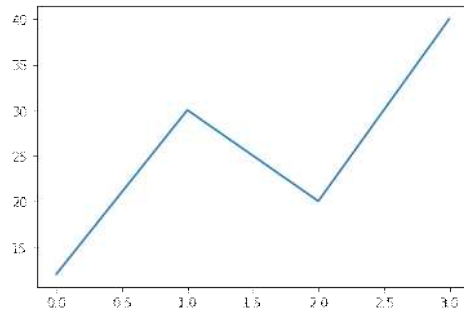
검색 🔍
초기화 ↺

● 현재화면
● 전체시군구현황
● 전체읍면동현황
📄 CSV 파일 다운로드 🔗
📄 xlsx 파일 다운로드 🔗

[주피터 노트북 코드]

- 라인 차트 그리기 연습

```
import matplotlib.pyplot as plt
plt.plot([12,30,20,40])
plt.show()
```



- csv 모듈로 파일 불러오기 연습

```
import csv
f=open('a.csv')
d = csv.reader(f)

for k in d:
    print(k)
```

```
['행정구역', '2020년1월_계_총인구수', '2020년1월_계_연령_
'2020년1월_계_3세', '2020년1월_계_4세', '2020년1월_계_5.
월_계_9세', '2020년1월_계_10세', '2020년1월_계_11세', '20
계_15세', '2020년1월_계_16세', '2020년1월_계_17세', '2020
21세', '2020년1월_계_22세', '2020년1월_계_23세', '2020년1
세', '2020년1월_계_28세', '2020년1월_계_29세', '2020년1
세', '2020년1월_계_34세', '2020년1월_계_35세', '2020년1
세', '2020년1월_계_40세', '2020년1월_계_41세', '2020년1
세', '2020년1월_계_46세', '2020년1월_계_47세', '2020년1
세', '2020년1월_계_52세', '2020년1월_계_53세', '2020년1
세', '2020년1월_계_58세', '2020년1월_계_59세', '2020년1
세', '2020년1월_계_64세', '2020년1월_계_65세', '2020년1
세', '2020년1월_계_70세', '2020년1월_계_71세', '2020년1
세', '2020년1월_계_76세', '2020년1월_계_77세', '2020년1
세', '2020년1월_계_82세', '2020년1월_계_83세', '2020년1
세']
```

- 특정 데이터 구간 가져오기

```
import csv
f = open('a.csv')
b = csv.reader(f)
next(b)
temp=[]

for k in b:
    if '월계2동' in k[0] :
        for i in range(3,104) :
            print(k[i])
```

```
114
162
178
185
201
198
192
209
215
205
179
197
213
192
192
209
```

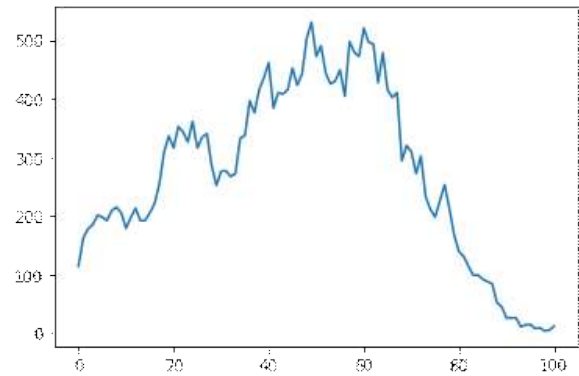
- csv모듈로 가져온 파일 차트로 표시

```
import csv
import matplotlib.pyplot as plt

f = open('a.csv')
d = csv.reader(f)
next(d)
temp=[]

for k in d:
    if '월계2동' in k[0] :
        for i in range(3,104) :
            temp.append(int(k[i]))

plt.plot(temp)
plt.show()
```

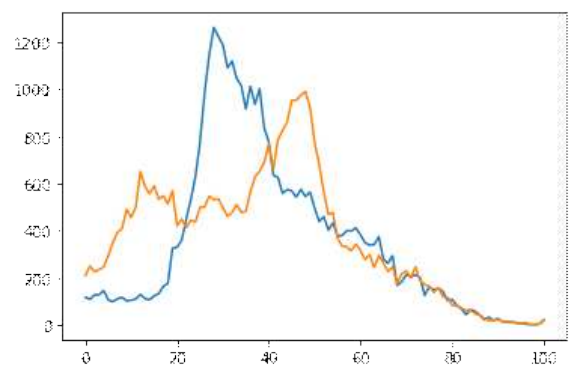


- 완성

```
import csv
import matplotlib.pyplot as plt
f = open('a.csv')
d = csv.reader(f)
next(d)
temp=[]
temp2=[]

for k in d:
    if '역삼1동' in k[0] :
        for i in range(3,104) :
            temp.append(int(k[i].replace(',','')))
    if '역삼2동' in k[0] :
        for i in range(3,104) :
            temp2.append(int(k[i].replace(',','')))

plt.plot(temp)
plt.plot(temp2)
plt.show()
```



```

import csv
import matplotlib.pyplot as plt

f = open('a.csv') # 파일을 열기 명령 f로 지정
d = csv.reader(f) # f 명령으로 연 파일을 읽어 d변수에 지정
next(d) # 첫 줄 생략
temp=[ ] # 빈 리스트 생성
temp2=[ ] # 빈 리스트 생성

for k in d : # d변수 안에 있는 데이터를 반복
    if '역삼1동' in k[0] : # 만약 반복문 변수 k의 첫번째 데이터가 '역삼1동' 이라면
        for i in range(3,104) : # 0세부터 100세 까지 0세 데이터는 4번째 즉, [0123] 순 3을 입력했음
            temp.append(int(k[i].replace(',','')))
            # temp라는 변수에 결과를 집어넣는다. replace 바꾸기 ,를 공백(none)으로

    if '역삼2동' in k[0] : # 만약 반복문 변수 k의 첫번째 데이터가 '역삼2동' 이라면
        for i in range(3,104) : # 0세부터 100세 까지 0세 데이터는 4번째 즉, [0123] 순 3을 입력했음
            temp2.append(int(k[i].replace(',','')))
            # temp라는 변수에 결과를 집어넣는다. replace 바꾸기 ,를 공백(none)으로

plt.plot(temp , label='y1') #라인차트의 데이터 temp로 지정, 라벨링 y1
plt.plot(temp2 , label='y2') #라인차트의 데이터 temp2로 지정, 라벨링 y2
plt.legend( ) # 범례 표시
plt.show( ) #차트 표시

```