

04.공공 데이터 활용_인구

학습 목표

- 인구 공공 데이터를 활용하여 데이터를 분석 할 수 있다.
- 데이터 분석 결과 를 시각화 할 수 있다.

특정 지역 주민등록 인구 및 세대 현황 데이터 활용하여 분석및 시각화하기

- 인구 공공 데이터

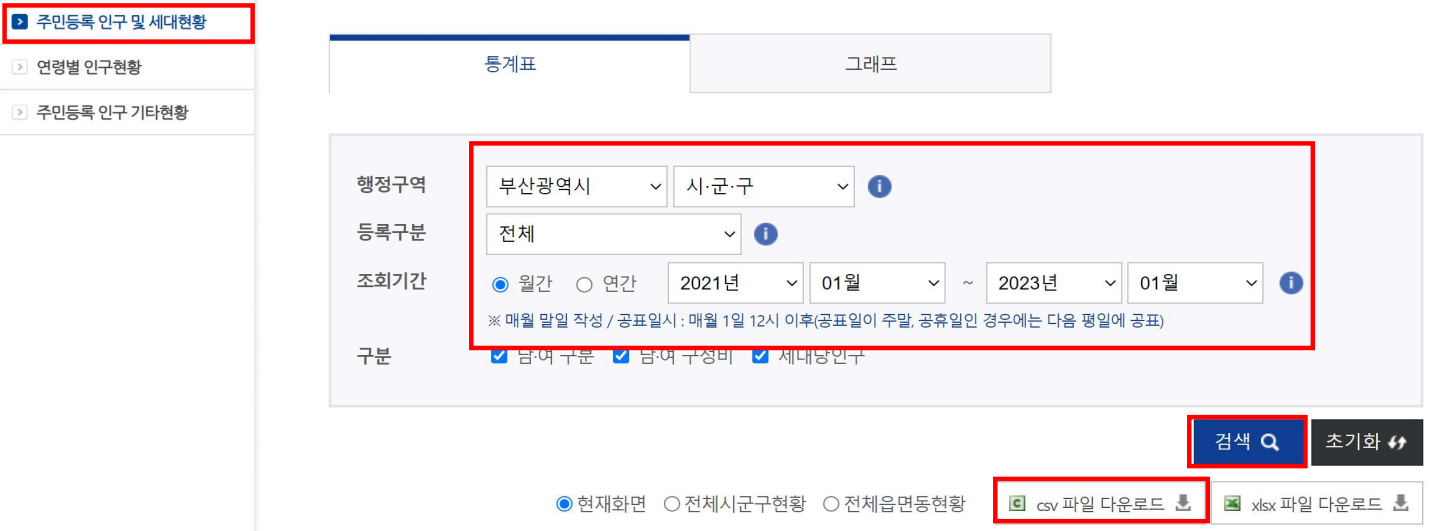
- . 행정안전부 홈페이지 에서 각종 공공 데이터 및 통계 자료를 무료로 다운로드 할 수 있도록 제공한다.
- . <https://mois.go.kr>
 - > 정책 자료 -> 통계 -> 주민등록 인구 통계



- * 주민등록 인구 및 세대 현황, 연령별 인구 현황, 주민등록 인구 기타 현황 등 인구 관련 다양한 데이터를 제공함
- * 행정구역, 등록구분, 조회기간 등 다양한 카테고리 별 검색을 지원함

- 데이터 다운로드 및 읽어오기

- . 주민등록 인구 및 세대 현황
- . 검색 조건 입력(부산광역시, 월간) -> 검색 -> CSV 파일 다운로드



. 파일 확인하기 (폴라리스 오피스 활용)

- * Excel 문서 편집기로 정리된 내역을 확인 가능 하지만 없는 경우 무료 문서 편집기 폴라리스오피스 를 사용할 수 있다.
- <https://www.polarisoffice.com/ko/>

	A	B	C	D	E	F	G
1	행정구역	2021년01월_총인구수	2021년01월_세대수	2021년01월_세대당 인구	2021년01월_남자 인구수	2021년01월_여자 인구수	2
2	부산광역시 (2600000000)	3,389,388	1,532,700	2.21	1,659,527	1,729,861	
3	부산광역시 부산광역시 중구 (2611000000)	41,549	23,923	1.74	20,269	21,280	
4	부산광역시 부산광역시 서구 (2614000000)	107,978	53,881	2	52,295	55,683	
5	부산광역시 부산광역시 동구 (2617000000)	88,763	46,081	1.93	43,462	45,301	
6	부산광역시 부산광역시 영도구 (2620000000)	113,048	54,905	2.06	55,697	57,351	
7	부산광역시 부산광역시 부산진구 (2623000000)	359,978	173,746	2.07	173,861	186,117	
8	부산광역시 부산광역시 동래구 (2626000000)	270,695	115,221	2.35	131,698	138,997	
9	부산광역시 부산광역시 남구 (2629000000)	267,351	117,885	2.27	130,792	136,559	
10	부산광역시 부산광역시 북구 (2632000000)	283,452	120,728	2.35	140,245	143,207	
11	부산광역시 부산광역시 해운대구 (2635000000)	401,710	171,528	2.34	193,341	208,369	
12	부산광역시 부산광역시 사하구 (2638000000)	311,052	138,852	2.24	154,893	156,159	
13	부산광역시 부산광역시 금정구 (2641000000)	231,959	107,415	2.16	113,238	118,721	
14	부산광역시 부산광역시 강서구 (2644000000)	138,610	58,512	2.37	71,620	66,990	
15	부산광역시 부산광역시 연제구 (2647000000)	208,900	92,694	2.25	100,563	108,337	
16	부산광역시 부산광역시 수영구 (2650000000)	176,742	84,460	2.09	83,514	93,228	
17	부산광역시 부산광역시 사상구 (2653000000)	212,173	97,687	2.17	107,191	104,982	
18	부산광역시 부산광역시 기장군 (2671000000)	175,428	75,182	2.33	86,848	88,580	

- * 파이썬 폴더 에 해당 자료 복사

>open() 함수 만 사용하여 읽어오기

* 모두 문자열로 읽어와 분석을 위한 전처리 과정이 복잡해진다.

* 파일을 불러오기 실패하였을 경우 encoding = 'cp949' 옵션을 추가
PC 환경에 따라 인코딩 방식을 작성하지 않아도 된다.

> 행 별로 데이터를 읽어 올수 있어 전처리 과정이 줄어들음

> DataFrame 형식으로 행과 열의 index 가 정리 되어 읽어오는것을 확인 할 수 있다.

한가람 IT 동상현 강사

- 평균 연령이 가장 높은 행정 구역 찾기
- . 데이터 다운로드
- . 주민등록 인구 기타 현황 - 지역 별 평균 연령
- > 최근 1달 기간 검색

연령별 인구현황

주민등록 인구 기타현황

- 주민등록기준 지역별 출생등록

- 주민등록기준 지역별 사망말소

- 지역별 세대원수별 세대수

- 지역별 평균연령

- 지역별 인구증감

- 지역별 아동, 청소년, 청년 인구현황

- 지역별 고령 인구현황

통계표

그래프

주민등록기준 지역별 출생등록 | 주민등록기준 지역별 사망말소 | 지역별 세대원수별 세대수 | 지역별 평균연령 | 지역별 인구증감 | 지역별 아동, 청소년, 청년 인구현황 | 지역별 고령 인구현황

행정구역

전국

시·군·구

조회기간

2023년

02월

~

2023년

02월

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

구분

☒ 계
 ☒ 남·여 구분

검색

초기화

☒ 현재화면
 ☐ 전체시군구현황
 ☐ 전체읍면동현황

. 데이터 확인

	A	B	C	D	E	F	G
1	행정구역	2023년01월_남자 평균연령	2023년01월_여자 평균연령	2023년01월_평균연령	2023년02월_남자 평균연령	2023년02월_여자 평균연령	2023년02월_평균연령
2	전국 (1000000000)	43.1	45.4	44.3	43.2	45.5	44.3
3	서울특별시 (1100000000)	43.2	44.8	44	43.2	44.8	44
4	부산광역시 (2600000000)	44.8	47.4	46.1	44.9	47.4	46.2
5	대구광역시 (2700000000)	43.2	46	44.7	43.3	46.1	44.7
6	인천광역시 (2800000000)	42.4	44.3	43.4	42.5	44.4	43.4
7	광주광역시 (2900000000)	41.3	43.6	42.5	41.3	43.7	42.5
8	대전광역시 (3000000000)	41.9	44.1	43	41.9	44.1	43
9	울산광역시 (3100000000)	42.1	44	43	42.2	44.1	43.1
10	세종특별자치시 (3600000000)	37.6	38.7	38.2	37.7	38.7	38.2
11	경기도 (4100000000)	41.6	43.3	42.5	41.7	43.4	42.5
12	강원도 (4200000000)	45.5	48.5	47	45.6	48.6	47.1
13	충청북도 (4300000000)	43.9	46.7	45.3	44	46.7	45.4
14	충청남도 (4400000000)	44	46.7	45.3	44.1	46.8	45.4
15	전라북도 (4500000000)	45.3	48.4	46.8	45.3	48.5	46.9
16	전라남도 (4600000000)	46.2	49.8	48	46.3	49.9	48.1
17	경상북도 (4700000000)	45.9	49.3	47.6	45.9	49.4	47.7
18	경상남도 (4800000000)	43.9	46.8	45.3	43.9	46.9	45.4
19	제주특별자치도 (5000000000)	42.3	44.7	43.5	42.4	44.7	43.5

. 파이썬으로 불러오기

```
import pandas as pd

result = pd.read_csv('202302_202302_주민등록인구기타현황(평균연령)_avgAge.csv', encoding='cp949')

print(result)
```

	행정구역	2023년02월_남자 평균연령	2023년02월_여자 평균연령	2023년02월_평균연령
0	전국 (1000000000)	43.2	45.5	44.3
1	서울특별시 (1100000000)	43.2	44.8	44.0
2	부산광역시 (2600000000)	44.9	47.4	46.2
3	대구광역시 (2700000000)	43.3	46.1	44.7
4	인천광역시 (2800000000)	42.5	44.4	43.4
5	광주광역시 (2900000000)	41.3	43.7	42.5
6	대전광역시 (3000000000)	41.9	44.1	43.0
7	울산광역시 (3100000000)	42.2	44.1	43.1
8	세종특별자치시 (3600000000)	37.7	38.7	38.2
9	경기도 (4100000000)	41.7	43.4	42.5
10	강원도 (4200000000)	45.6	48.6	47.1
11	충청북도 (4300000000)	44.0	46.7	45.4
12	충청남도 (4400000000)	44.1	46.8	45.4
13	전라북도 (4500000000)	45.3	48.5	46.9
14	전라남도 (4600000000)	46.3	49.9	48.1
15	경상북도 (4700000000)	45.9	49.4	47.7
16	경상남도 (4800000000)	43.9	46.9	45.4
17	제주특별자치도 (5000000000)	42.4	44.7	43.5

파이썬 데이터 분석 및 시각화

한가람 IT 동상현 강사

. 특정 컬럼 의 가장 높은 평균 연령 구하기

```
print(result['2023년02월_평균연령'].max())
```

48.1

> 높은 값의 행정 구역을 알고 싶은 것이므로 값의 index 를 찾아 이름과 값을 같이 표현한다.

* idxmax() 로 가장 높은 값의 index 를 찾는다.

```
print(result['2023년02월_평균연령'].idxmax())
```

14

> iloc[] 에 가장 높은 평균연령 index 로 출력을 하면 평균 연령이 가장 높음 지역을 확인 할 수 있다.

```
print(result.iloc[result['2023년 02월_평균연령'].idxmax()])
```

행정구역	전라남도	(4600000000)
2023년02월_남자_평균연령		46.3
2023년02월_여자_평균연령		49.9
2023년02월_평균연령		48.1
Name: 14, dtype: object		

- 남 / 여 의 연령 차이가 가장 큰 행정 자치구역 구하기

. 여자 평균연령 컬럼 과 남자 평균 연령 컬럼 의 차이를 data 리스트 에 할당

```
data = result['2023년02월_여자 평균연령'] - result['2023년02월_남자 평균연령']
```

. 가장 차이가 많이 나는 index 구하기

```
print(data.idxmax())
```

14

해당 index 의 행 정보 표현하여 남녀 연령 차이가 가장 큰 행정 구역 검출하기

```
print(result.iloc[data.idxmax()])
```

행정구역	전라남도	(4600000000)
2023년02월_남자_평균연령		46.3
2023년02월_여자_평균연령		49.9
2023년02월_평균연령		48.1
Name: 14, dtype: object		

- 부산광역시 의 해당 기간동안 월별 행정 구역 별 주민등록 인구 및 세대 현황 구하기

. 데이터 구조 파악하기 (202301 202309 주민등록인구및세대현황 월간.csv)

> 총 인구 수는 두번째 열 부터 시작 하여 다음 월의 인구수는 6칸 뒤에 나타난다.

> 행정구역별 인구 수는 세번째 행 (실제 데이터의 두번째 행) 부터 확인 할 수 있다.

A	B	C	D	E	F	G	H
행정구역	2021년01월_총인구수	2021년01월_세다	2021년01월_	2021년01월_	2021년01월_	2021년01월_	2021년02월_총인구수
부산광역시 (2600000000)	3,389,388	1,532,700	2.21	1,659,527	####	0.96	3,387,761
부산광역시 부산광역시 중구 (2611000000)	41,549	23,923	1.74	20,269	21,280	0.95	41,551
부산광역시 부산광역시 서구 (2614000000)	107,978	53,881	2	52,295	55,683	0.94	107,824
부산광역시 부산광역시 동구 (2617000000)	88,763	46,081	1.93	43,462	45,301	0.96	88,639
부산광역시 부산광역시 영도구 (2620000000)	113,048	54,905	2.06	55,697	57,351	0.97	112,809
부산광역시 부산광역시 부산진구 (2623000000)	359,978	173,246	2.07	173,861	186,117	0.93	360,025
부산광역시 부산광역시 동래구 (2626000000)	270,695	115,221	2.35	131,698	138,997	0.95	270,678
부산광역시 부산광역시 남구 (2629000000)	267,351	117,885	2.27	130,792	136,559	0.96	266,989
부산광역시 부산광역시 북구 (2632000000)	283,452	120,728	2.35	140,245	143,207	0.98	282,799
부산광역시 부산광역시 해운대구 (2635000000)	401,710	171,528	2.34	193,341	208,369	0.93	401,280
부산광역시 부산광역시 사하구 (2638000000)	311,052	138,852	2.24	154,893	156,159	0.99	310,740
부산광역시 부산광역시 금정구 (2641000000)	231,959	107,415	2.16	113,238	118,721	0.95	231,743
부산광역시 부산광역시 강서구 (2644000000)	138,610	58,512	2.37	71,620	66,990	1.07	139,088
부산광역시 부산광역시 연제구 (2647000000)	208,900	92,694	2.25	100,563	108,337	0.93	208,743
부산광역시 부산광역시 수영구 (2650000000)	176,742	84,460	2.09	83,514	93,228	0.9	176,703
부산광역시 부산광역시 사상구 (2653000000)	212,173	97,687	2.17	107,191	104,982	1.02	212,038
부산광역시 부산광역시 기장군 (2656000000)	175,428	75,182	2.33	86,848	88,580	0.98	176,112

. Pandas 모듈을 활용하여 데이터 전처리 하기

- > columns 함수 : 컬럼의 목록을 가져올 수 있음
- > iloc 함수 : 행 번호로 값을 선택 할 수 있음
- * 월별 행정구역별 거주자 인구 를 가져 올 수 있다.
- > 기능 적용 을 위한 데이터 확인하기

```
print(result.iloc[1,0])
```

부산광역시 부산광역시 중구 (2611000000)

```
print(result.iloc[1,1])
```

41,549

```
print(result.iloc[1])
```

행정구역	부산광역시 부산광역시 중구 (2611000000)
2021년01월_총인구수	41,549
2021년01월_세대수	23,923
2021년01월_세대당 인구	1.74
2021년01월_남자 인구수	20,269

2023년01월_세대수	23,965
2023년01월_세대당 인구	1.65
2023년01월_남자 인구수	19,188
2023년01월_여자 인구수	20,439
2023년01월_남여 비율	0.94

Name: 1, Length: 151, dtype: object

> 1행 0번 컬럼의 데이터 : 첫번째 행정구역 명칭

> 1행 1번 컬럼의 데이터 :
첫번째 행정구역 첫월 총 인구수

> 1행의 모든 데이터 출력

```
print(result.iloc[1,0])  
print(result.iloc[1,range(1,len(result.columns),6)])
```

부산광역시 부산광역시 중구 (2611000000)

2021년01월_총인구수	41,549
2021년02월_총인구수	41,551
2021년03월_총인구수	41,370
2021년04월_총인구수	41,311
2021년05월_총인구수	41,209
2021년06월_총인구수	40,561
2021년07월_총인구수	40,666
2021년08월_총인구수	40,713
2021년09월_총인구수	40,708
2021년10월_총인구수	40,676
2021년11월_총인구수	40,556
2021년12월_총인구수	40,524
2022년01월_총인구수	40,537
2022년02월_총인구수	40,582
2022년03월_총인구수	40,596
2022년04월_총인구수	40,585
2022년05월_총인구수	40,537
2022년06월_총인구수	40,455
2022년07월_총인구수	40,342
2022년08월_총인구수	40,297
2022년09월_총인구수	40,117
2022년10월_총인구수	39,936
2022년11월_총인구수	39,840
2022년12월_총인구수	39,689
2023년01월_총인구수	39,627

> 첫번째 행정 구역의 명칭 과

> 1행 1열(첫번째 행정구역 의 첫월 인구수)
부터 끝까지 6칸씩 건너뛰는 데이터 표현
(컬럼의 이름과 데이터 가 같이 표현된다.)

> 모든 행정 구역 표시를 위해 반복문 을 통해 사전에 '행정구역명칭': '월별 인구수 리스트'

```
import pandas as pd
result = pd.read_csv('202101_202301_주민등록인구및세대현황_월간.csv',encoding = 'cp949')
data = {}
for i in range(1,len(result)):
    data[result.iloc[i,0]] = result.iloc[1,range(1,len(result.columns),6)]
print(data)
```

2022년04월_총인구수	40,589	
2022년05월_총인구수	40,537	
2022년06월_총인구수	40,455	
2022년07월_총인구수	40,342	
2022년08월_총인구수	40,297	
2022년09월_총인구수	40,117	
2022년10월_총인구수	39,936	
2022년11월_총인구수	39,840	
2022년12월_총인구수	39,689	
2023년01월_총인구수	39,627	
Name: 1, dtype: object, '부산광역시 부산광역시 서구 (2614000000)': 2021년01월_총인구수		41,549
2021년02월_총인구수	41,551	
2021년03월_총인구수	41,370	
2021년04월_총인구수	41,311	
2021년05월_총인구수	41,209	
2021년06월_총인구수	40,561	
2021년07월_총인구수	40,666	
2021년08월_총인구수	40,713	
2021년09월_총인구수	40,708	
2021년10월_총인구수	40,676	

* data[result.iloc[i,0]] : 디셔너리의 키 로 생성 할 행정 자치 구역 명칭, ([1,0] 의 위치값부터 순차적으로 Key 에 저장
result.iloc[1,range(1,len(result.columns),6)] : 해당 Key 에 등록 할 1열 부터 6칸씩 건너뛰면서 가져올 월별 인구 데이터