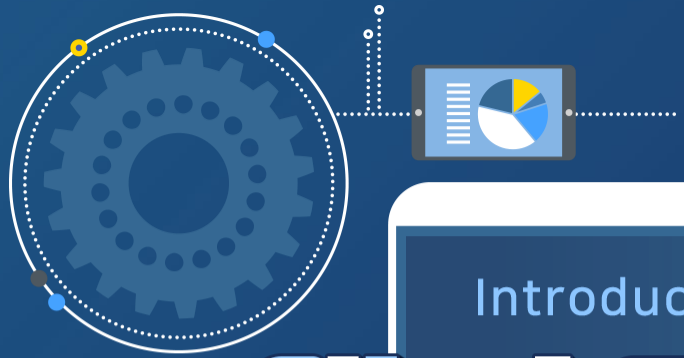
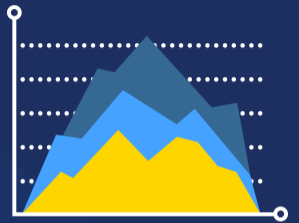




Introduction To Data Analysis

데이터분석입문

Lecture 16. 인구 데이터 준비와 지역의 연령별 인구 데이터 분석

인공지능소프트웨어학과 **강환수 교수**

학습개요

- 인구 데이터 준비하기
  - 행정안전부(https://www.mois.go.kr/)
- 우리 동네 연령별 인구 구조 시각화하기

학습목표

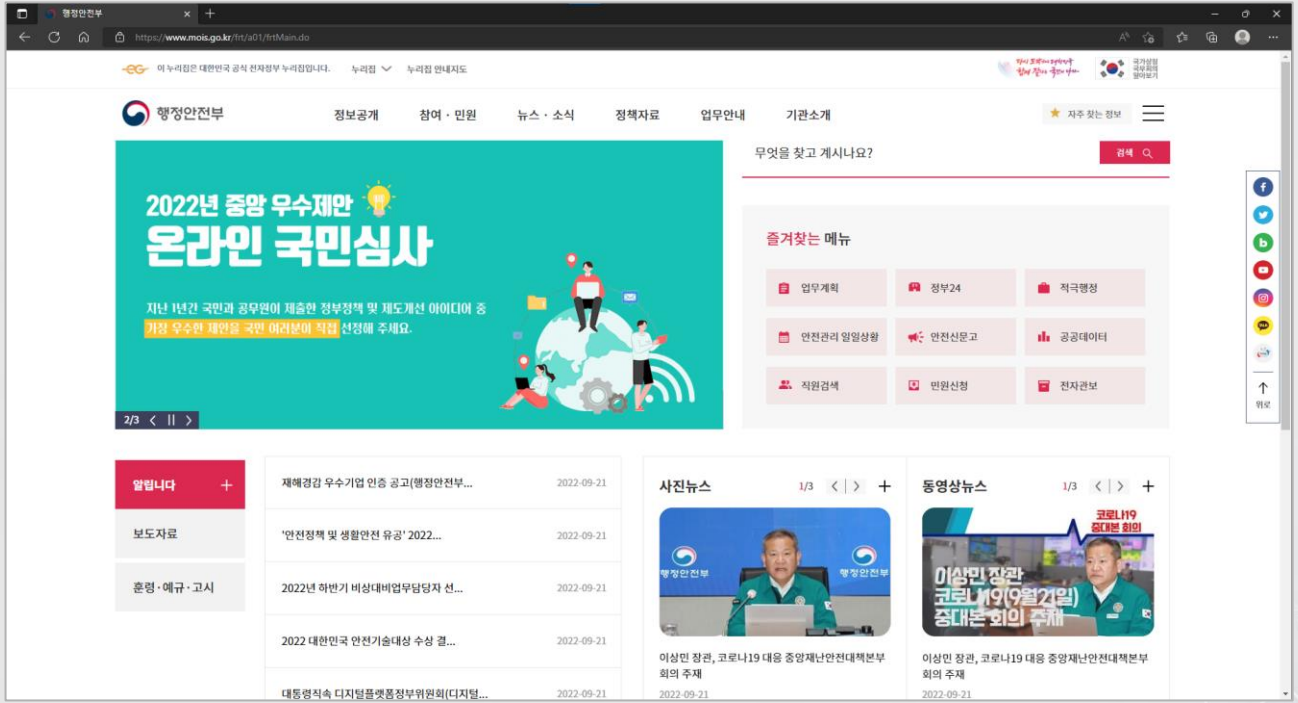
- 행정안전부에서 인구 데이터를 내려 받을 수 있다.
  - age.xls, gender.xls
- 특정 지역의 연령별 인구 수 그래프를 그릴 수 있다.
- 표준입력으로 받은 지역의 연령별 인구 수 그래프를 그릴 수 있다.

LESSON 01

인구 데이터 준비하기

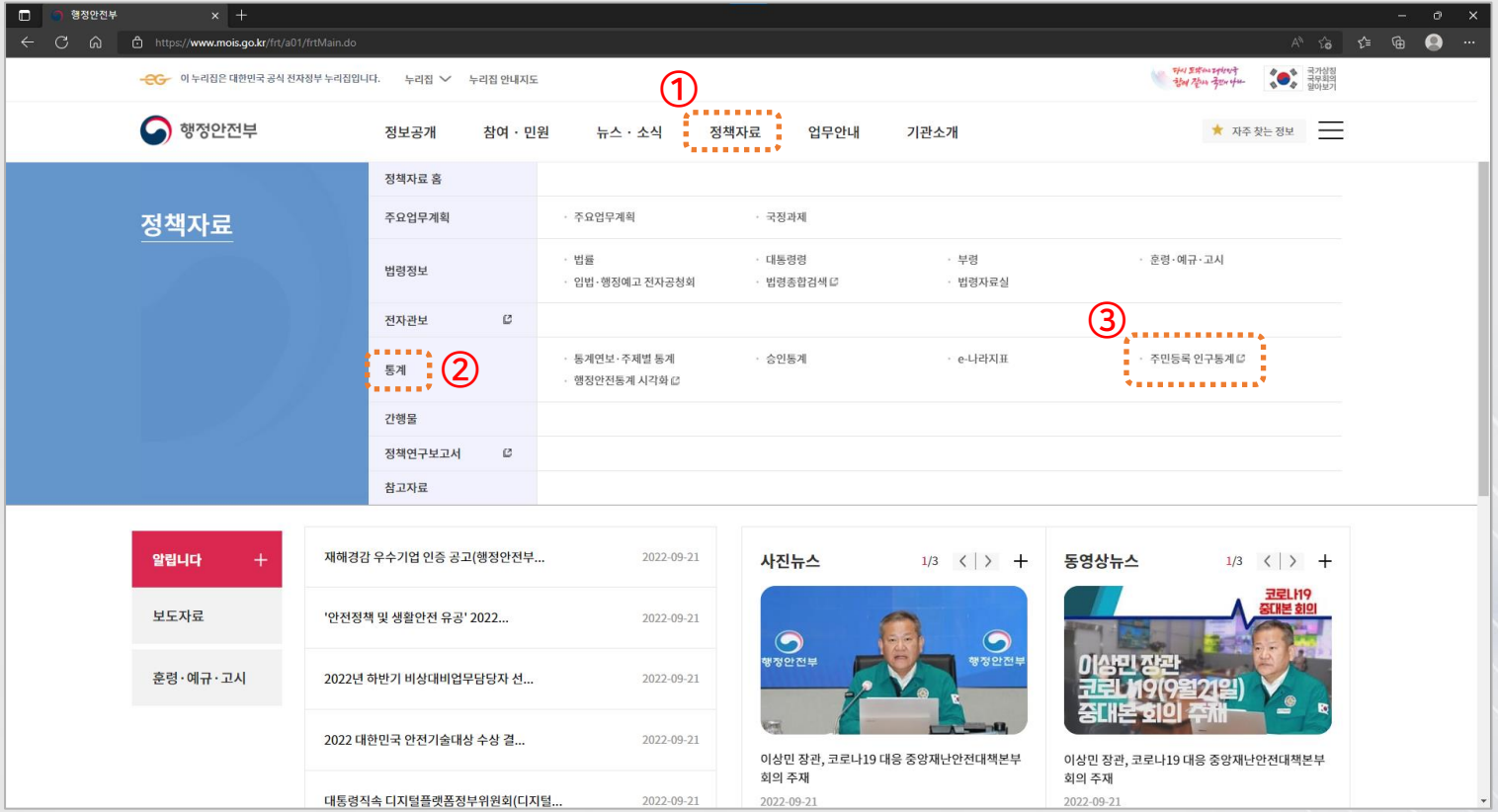
인구 공공데이터 내려받기 (1/8)

인구 데이터 수집 → 행정안전부(https://www.mois.go.kr/)



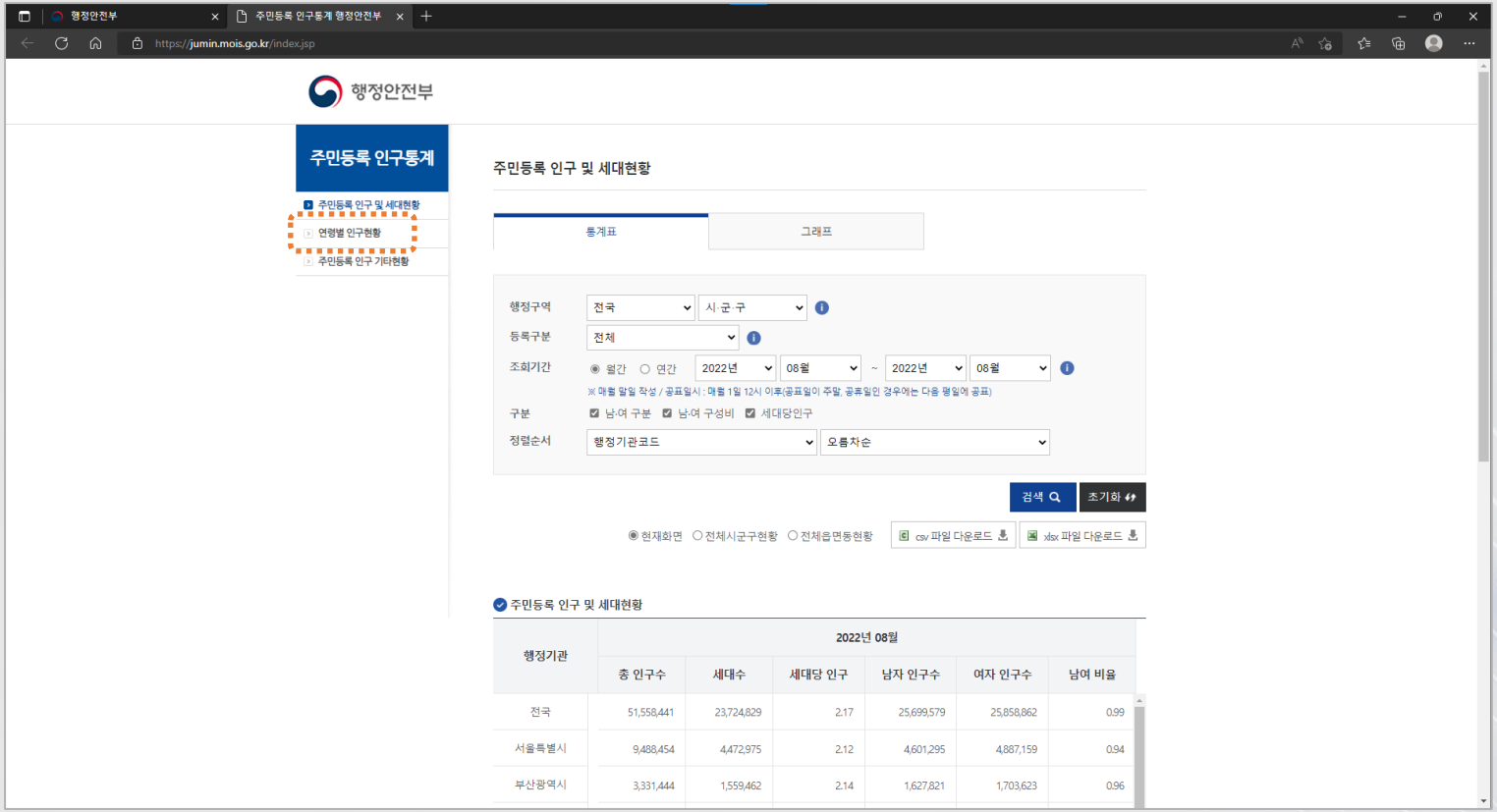
인구 공공데이터 내려받기 (2/8)

[정책자료] – [통계] – [주민등록 인구통계] 버튼 클릭



인구 공공데이터 내려받기 (3/8)

[연령별 인구현황] 버튼 클릭

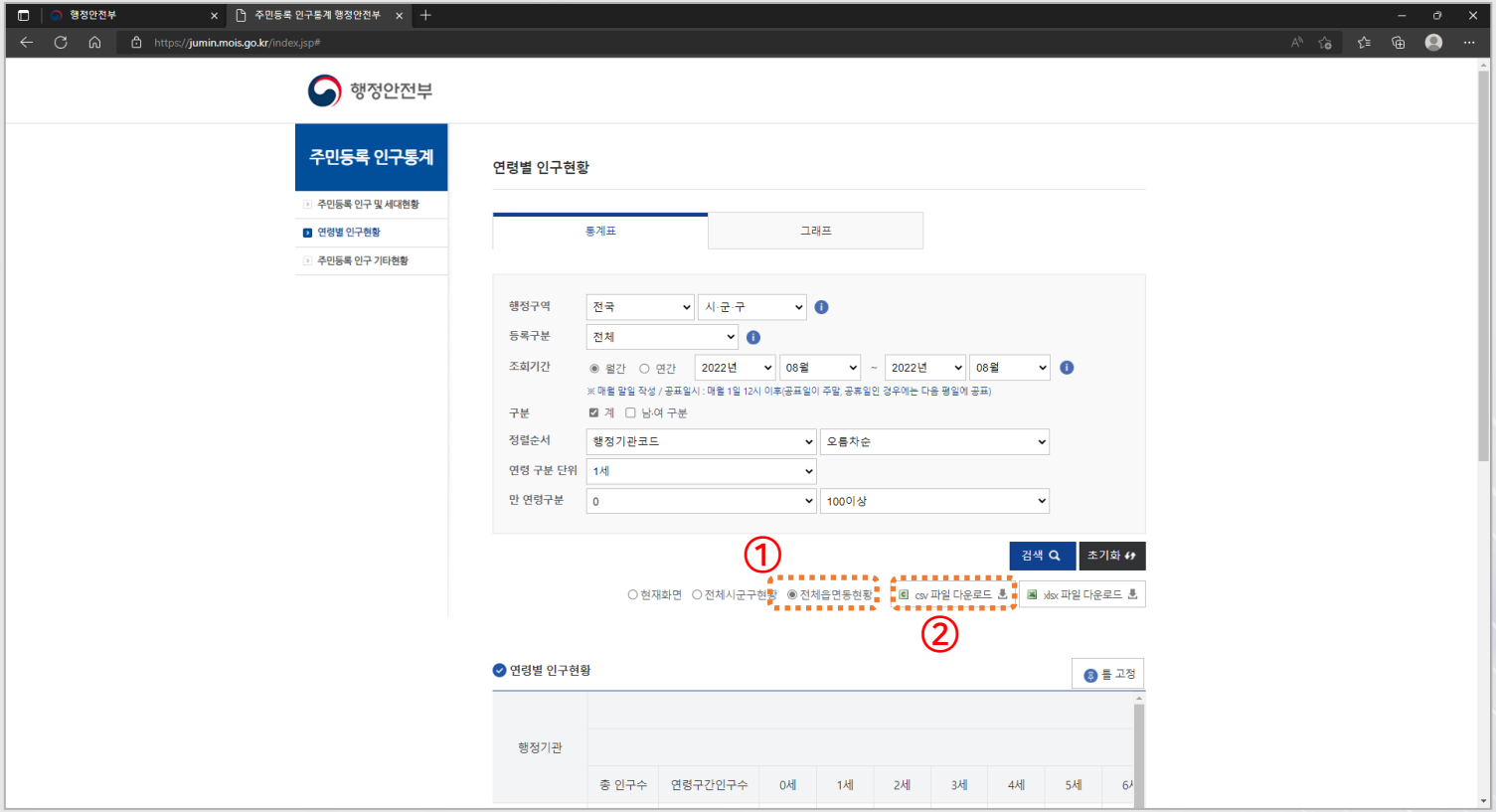


인구 공공데이터 내려받기 (4/8)

- [조회기간]: 2024년 08월 ~ 2024년 08월
- [남·여 구분] 체크 해제
- [연령 구분 단위]: 1세
- 만 연령구분: 0, 100이상
- [검색] 버튼 클릭

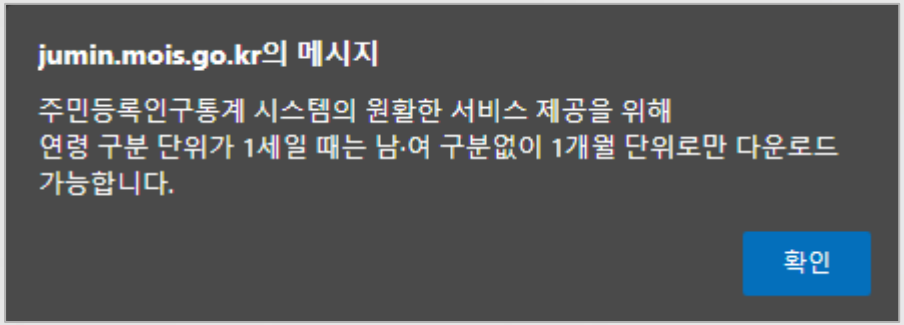
인구 공공데이터 내려받기 (5/8)

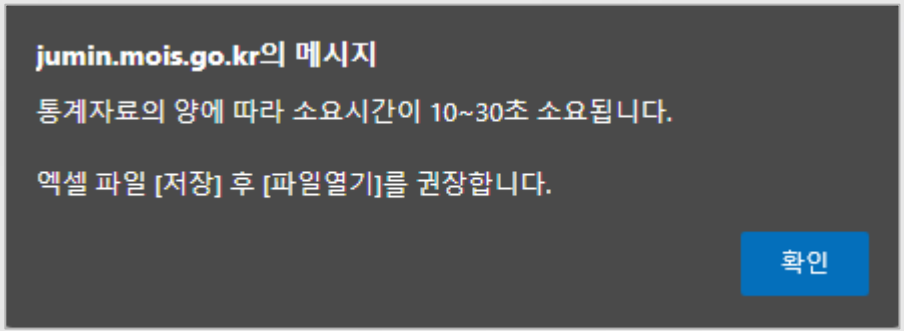
[전체읍면동현황] 체크 후, [CSV 파일 다운로드] 버튼 클릭



인구 공공데이터 내려받기 (6/8)

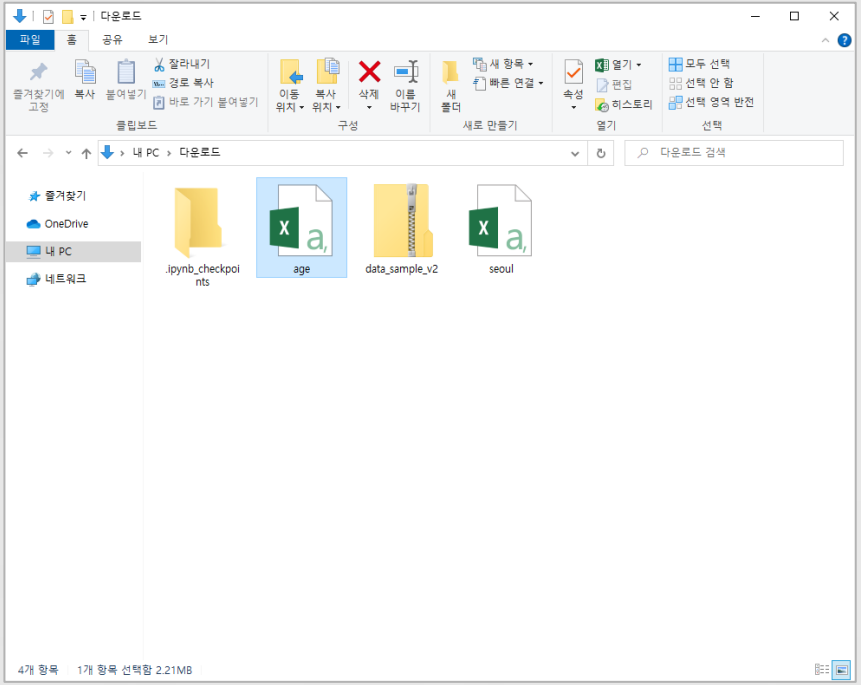
아래와 같이 경고 창이 뜨면 [확인] 버튼을 클릭하세요.





인구 공공데이터 내려받기 (7/8)

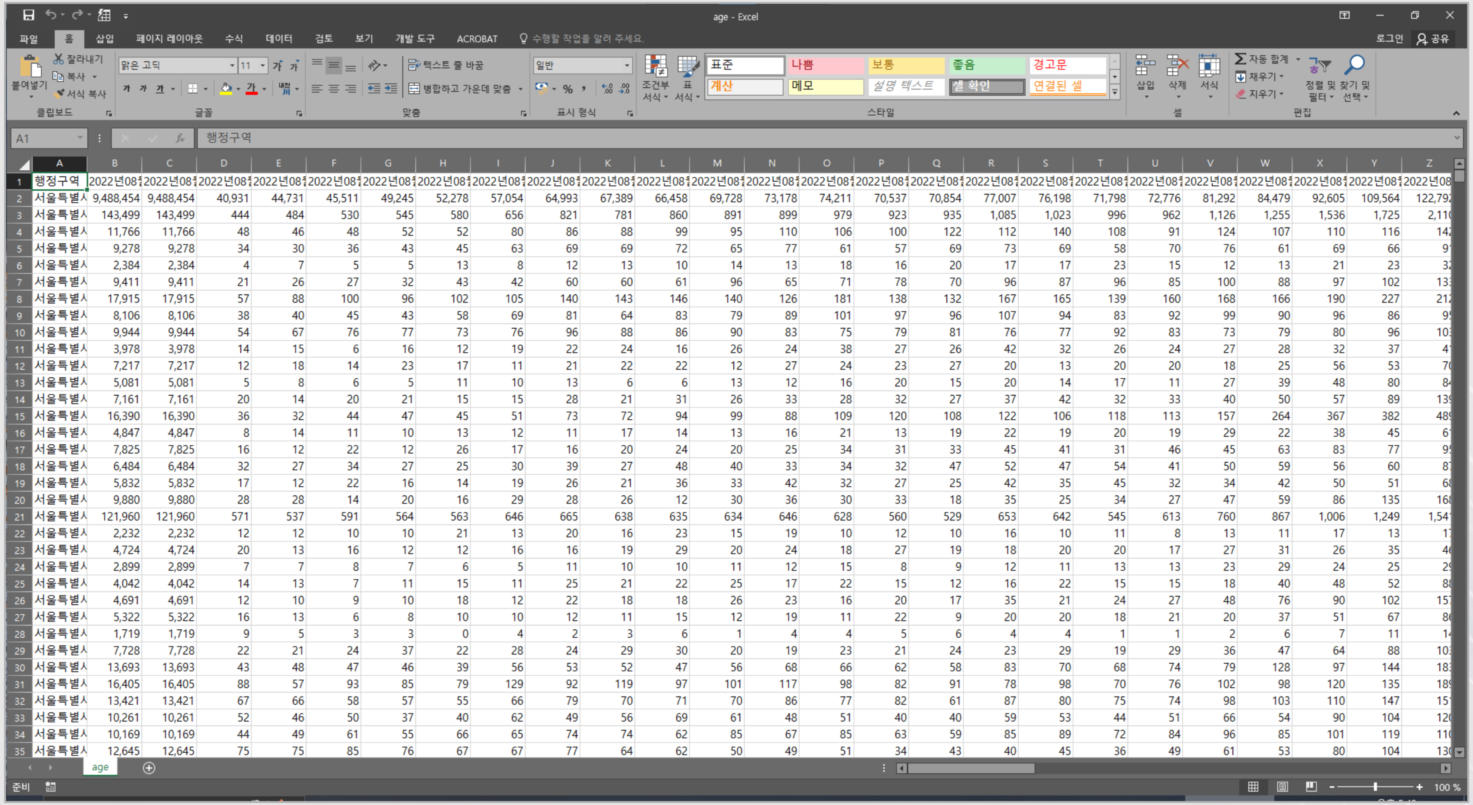
- 파일 이름을 “age.csv”로 변경합니다.
- 파일은 다운로드(Downloads) 폴더에 저장되어 있다고 가정하고 실습을 진행하겠습니다.



DMUAi 동양미래대학교 인공지능소프트웨어학과

인구 공공데이터 내려받기 (8/8)

age.csv 파일을 열면 전국의 읍면동, 연령별 인구를 확인할 수 있습니다.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	행정구역	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08	2022년08
2	서울특별시	9,488,454	9,488,454	40,931	44,731	45,511	49,245	52,278	57,054	64,993	67,389	66,458	69,728	73,178	74,211	70,537	70,854	77,007	76,198	71,798	72,776	81,292	84,479	92,605	109,564	122,792
3	서울특별시	143,499	143,499	444	484	530	545	580	656	821	781	860	891	899	979	923	935	1,085	1,023	996	962	1,126	1,255	1,536	1,725	2,110
4	서울특별시	11,766	11,766	48	46	48	52	52	80	86	88	99	95	110	106	100	122	112	140	108	91	124	107	110	116	142
5	서울특별시	9,278	9,278	34	30	36	43	45	63	69	69	72	65	77	61	57	69	73	69	58	70	76	61	69	66	91
6	서울특별시	2,384	2,384	4	7	5	5	13	8	12	13	10	14	13	18	16	20	17	17	23	15	12	13	21	23	32
7	서울특별시	9,411	9,411	21	26	27	32	43	42	60	60	61	96	65	71	78	70	96	87	96	85	100	88	97	102	133
8	서울특별시	17,915	17,915	57	88	100	96	102	105	140	143	146	140	126	181	138	132	167	165	139	160	168	166	190	227	212
9	서울특별시	8,106	8,106	38	40	45	43	58	69	81	64	83	79	89	101	97	96	107	94	83	92	99	90	96	86	95
10	서울특별시	9,944	9,944	54	67	76	77	73	76	96	88	86	90	83	75	79	81	76	77	92	83	73	79	80	96	103
11	서울특별시	3,978	3,978	14	15	6	16	12	19	22	24	16	26	24	38	27	26	42	32	26	24	27	28	32	37	41
12	서울특별시	7,217	7,217	12	18	14	23	17	11	21	22	22	12	27	24	23	27	20	13	20	20	18	25	56	53	70
13	서울특별시	5,081	5,081	5	8	6	5	11	10	13	6	6	13	12	16	20	15	20	14	17	11	27	39	48	80	84
14	서울특별시	7,161	7,161	20	14	20	21	15	15	28	21	31	26	33	28	32	27	37	42	32	33	40	50	57	89	139
15	서울특별시	16,390	16,390	36	32	44	47	45	51	73	72	94	99	88	109	120	108	122	106	118	113	157	264	367	382	489
16	서울특별시	4,847	4,847	8	14	11	10	13	12	11	17	14	13	16	21	13	19	22	19	20	19	29	22	38	45	61
17	서울특별시	7,825	7,825	16	12	22	12	26	17	16	20	24	20	25	34	31	33	45	41	31	46	45	63	83	77	95
18	서울특별시	6,484	6,484	32	27	34	27	25	30	39	27	48	40	33	34	32	47	52	47	54	41	50	59	56	60	87
19	서울특별시	5,832	5,832	17	12	22	16	14	19	26	21	36	33	42	32	27	25	42	35	45	32	34	42	50	51	68
20	서울특별시	9,880	9,880	28	28	14	20	16	29	28	26	12	30	36	30	33	18	35	25	34	27	47	59	86	135	168
21	서울특별시	121,960	121,960	571	537	591	564	563	646	665	638	635	634	646	628	560	529	653	642	545	613	760	867	1,006	1,249	1,541
22	서울특별시	2,232	2,232	12	12	10	10	21	13	20	16	23	15	19	10	12	10	16	10	11	8	13	11	17	13	17
23	서울특별시	4,724	4,724	20	13	16	12	12	16	16	19	29	20	24	18	27	19	18	20	20	17	27	31	26	35	46
24	서울특별시	2,899	2,899	7	7	8	7	6	5	11	10	10	11	12	15	8	9	12	11	13	13	23	29	24	25	29
25	서울특별시	4,042	4,042	14	13	7	11	15	11	25	21	22	25	17	22	15	12	16	22	15	15	18	40	48	52	88
26	서울특별시	4,691	4,691	12	10	9	10	18	12	22	18	18	26	23	16	20	17	35	21	24	27	48	76	90	102	157
27	서울특별시	5,322	5,322	16	13	6	8	10	10	12	11	15	12	19	11	22	9	20	20	18	21	20	37	51	67	86
28	서울특별시	1,719	1,719	9	5	3	3	0	4	2	3	6	1	4	4	5	6	4	4	1	1	2	6	7	11	14
29	서울특별시	7,728	7,728	22	21	24	37	22	28	24	29	30	20	19	23	21	24	23	29	19	29	36	47	64	88	103
30	서울특별시	13,693	13,693	43	48	47	46	39	56	53	52	47	56	68	66	62	58	83	70	68	74	79	128	97	144	183
31	서울특별시	16,405	16,405	88	57	93	85	79	129	92	119	97	101	117	98	82	91	78	98	70	76	102	98	120	135	189
32	서울특별시	13,421	13,421	67	66	58	57	55	66	79	70	71	70	86	77	82	61	87	80	75	74	98	103	110	147	151
33	서울특별시	10,261	10,261	52	46	50	37	40	62	49	56	69	61	48	51	40	40	59	53	44	51	66	54	90	104	120
34	서울특별시	10,169	10,169	44	49	61	55	66	65	74	74	62	85	67	85	63	59	85	89	72	84	96	85	101	119	110
35	서울특별시	12,645	12,645	75	75	85	76	67	67	77	64	62	50	49	51	34	43	40	45	36	49	61	53	80	104	130

인구 데이터 살펴보고 질문하기

- 어느 동네에 영유아가 가장 많을까?
- 어느 동네가 가장 고령화되었을까?
- 우리 동네에 가장 많이 살고 있는 연령은 몇 살일까?
- 내 또래 사람들이 가장 많이 사는 지역은 어디일까?

우리 동네 인구 구조 시각화하기

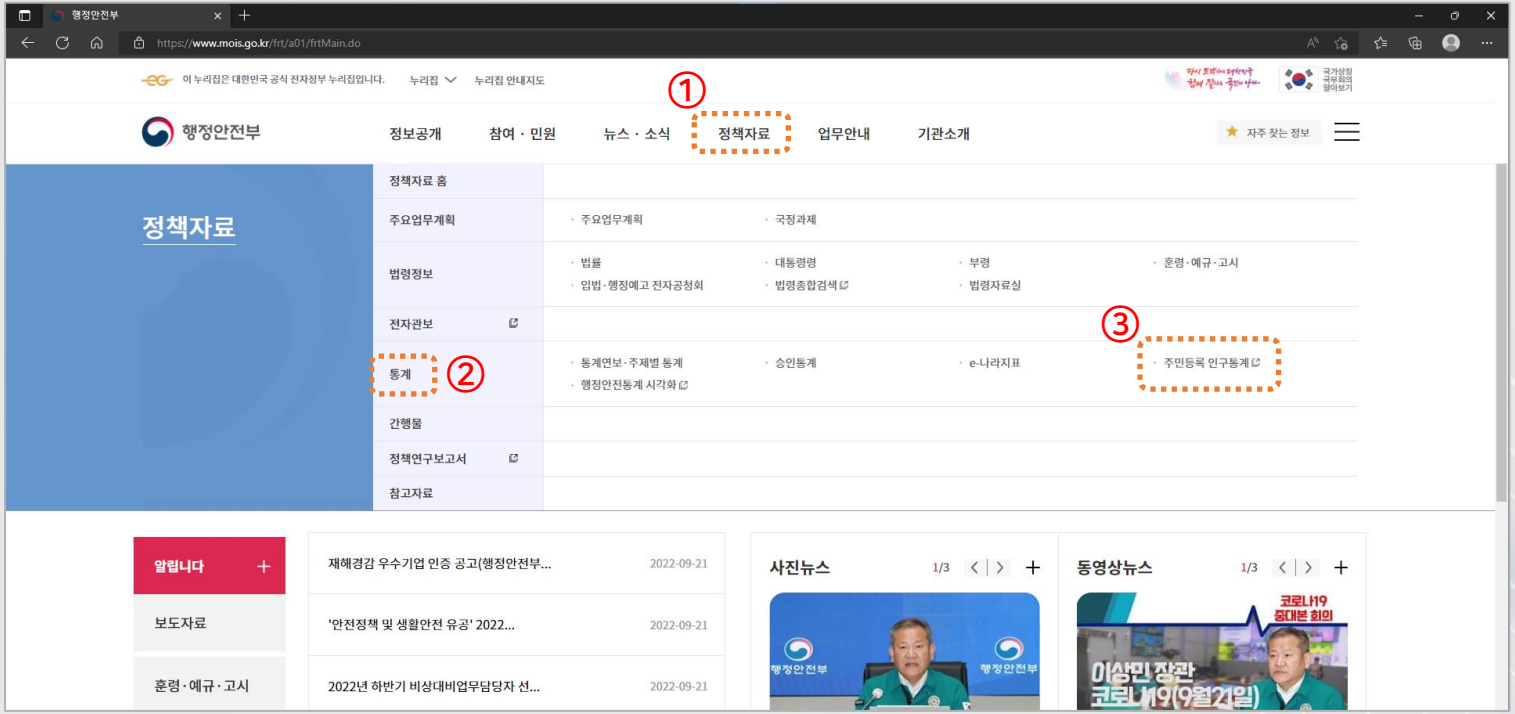
우리 동네의 인구 구조를 시각화 하려면 어떤 단계(또는 절차)를 거치면 될까요?

- 알고리즘(Algorithm) 설계하기

  - Step 1) 인구 데이터 파일을 읽어온다.
  - Step 2) 전체 데이터에서 한 줄씩 반복해서 읽어온다.
  - Step 3) 우리 동네에 대한 데이터인지 확인한다.
  - Step 4) 우리 동네일 경우 0세부터 100세 이상까지의 인구수를 순서대로 저장한다.
  - Step 5) 저장된 연령별 인구수 데이터를 시각화한다.

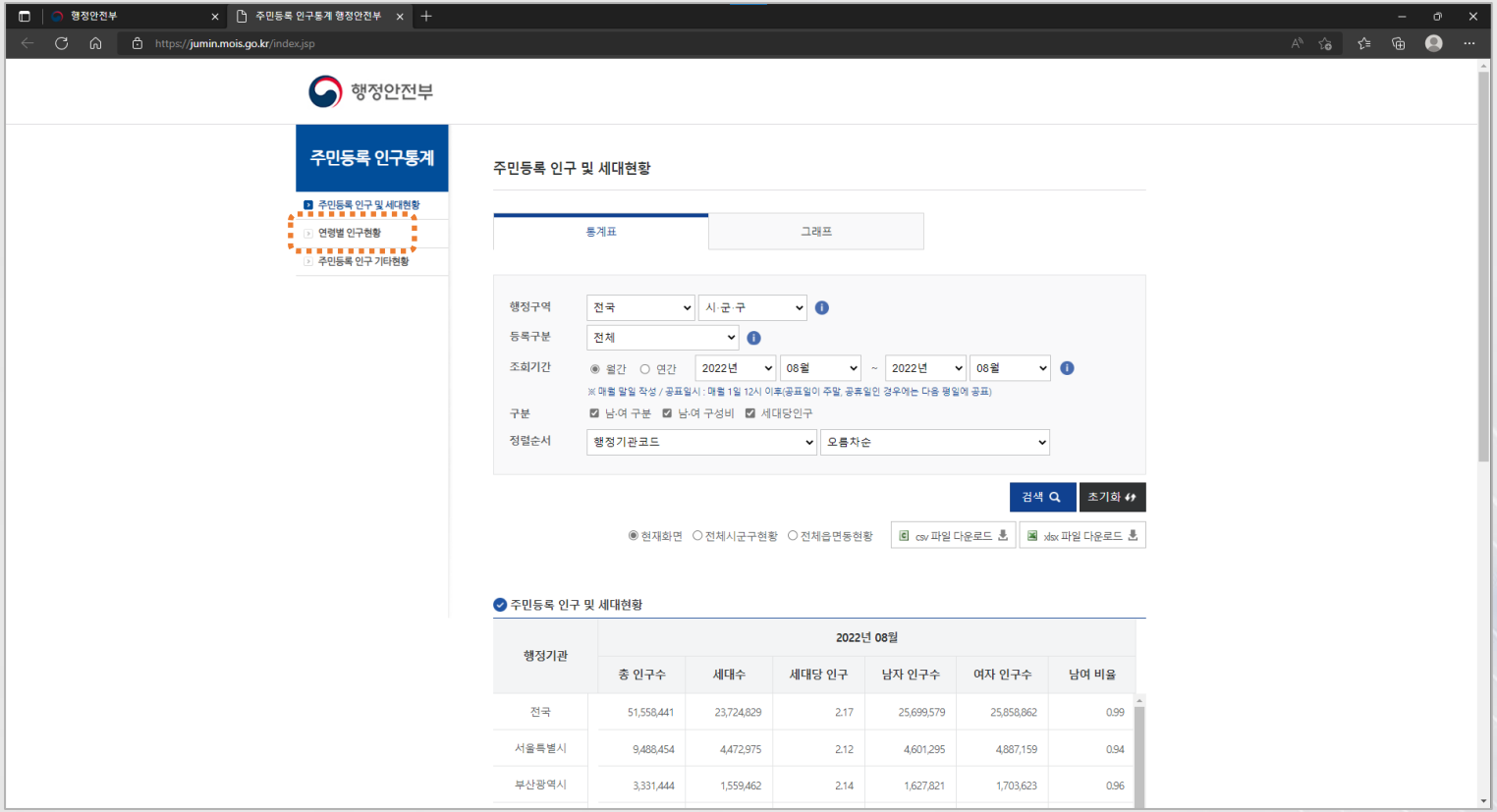
남·여 인구 공공데이터 내려받기 (1/8)

행정안전부(https://www.mois.go.kr/) – [정책자료] – [통계] – [주민등록 인구통계] 버튼 클릭



남·여 인구 공공데이터 내려받기 (2/8)

[연령별 인구현황] 버튼 클릭

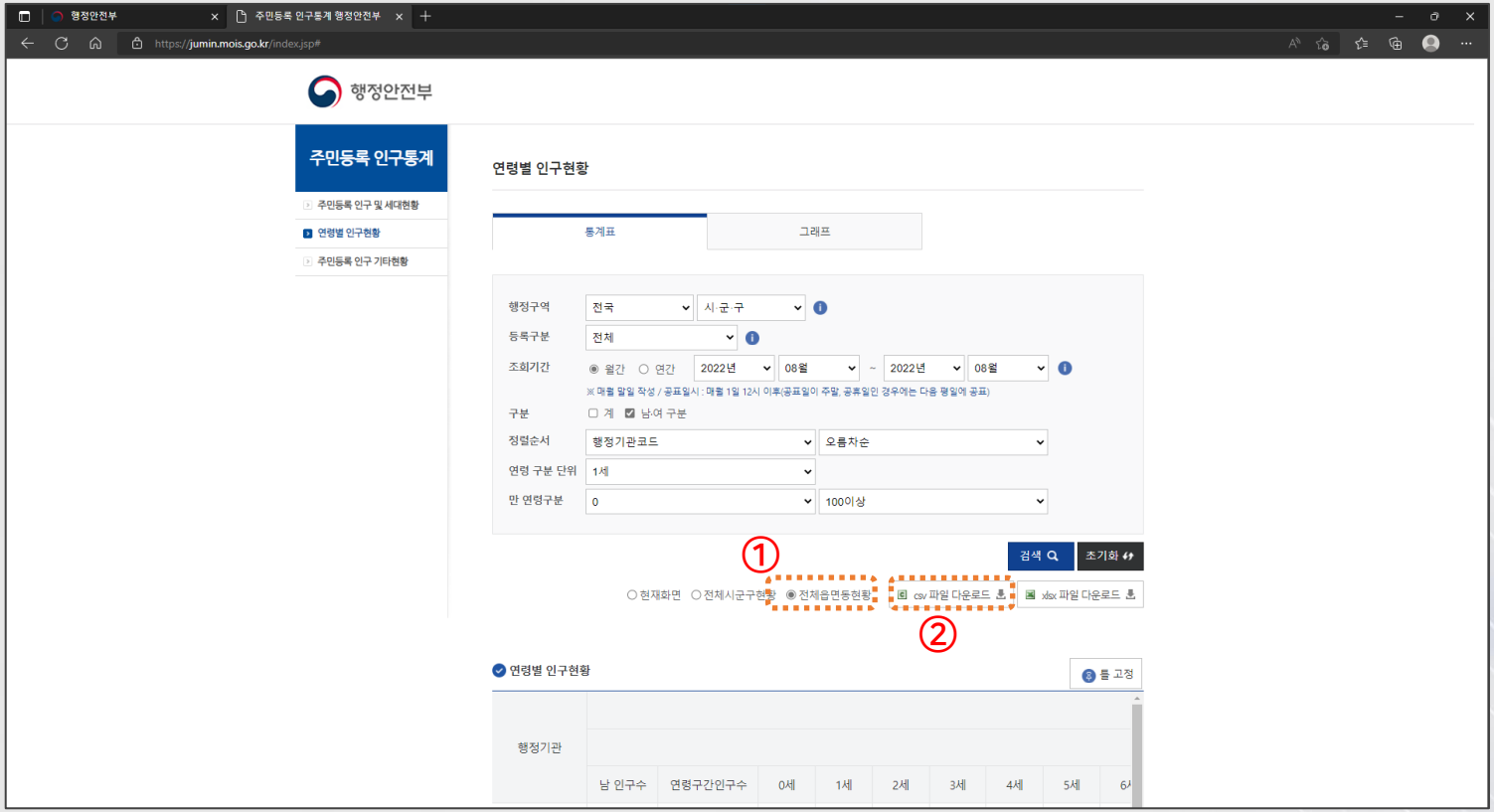


남·여 인구 공공데이터 내려받기 (3/8)

- [조회기간]: 2024년 08월 ~ 2024년 08월
- [계] 체크 해제
- [연령 구분 단위]: 1세
- 만 연령구분: 0, 100이상
- [검색] 버튼 클릭

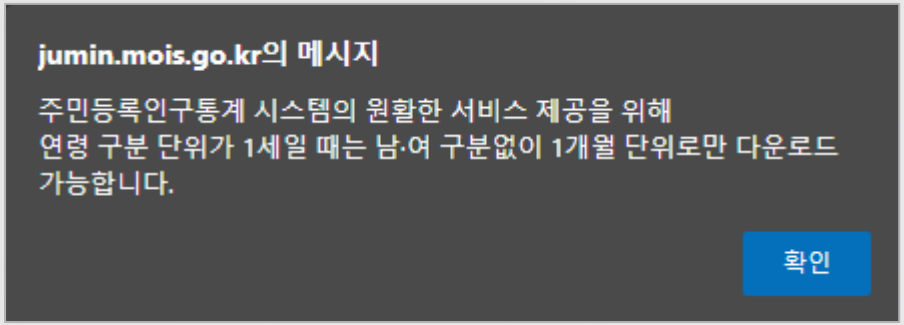
남·여 인구 공공데이터 내려받기 (4/8)

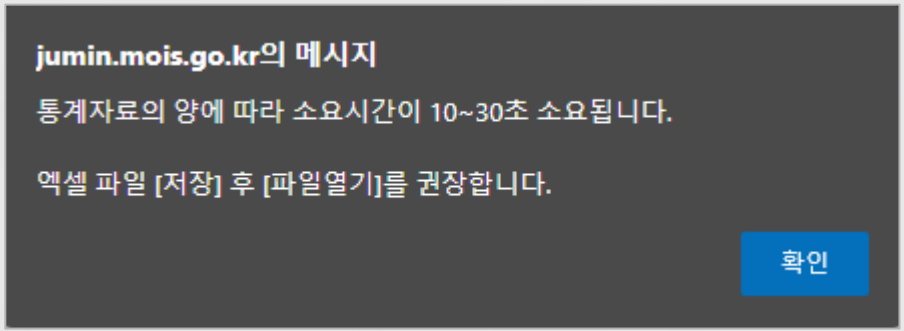
[전체읍면동현황] 체크 후, [CSV 파일 다운로드] 버튼 클릭



남·여 인구 공공데이터 내려받기 (5/8)

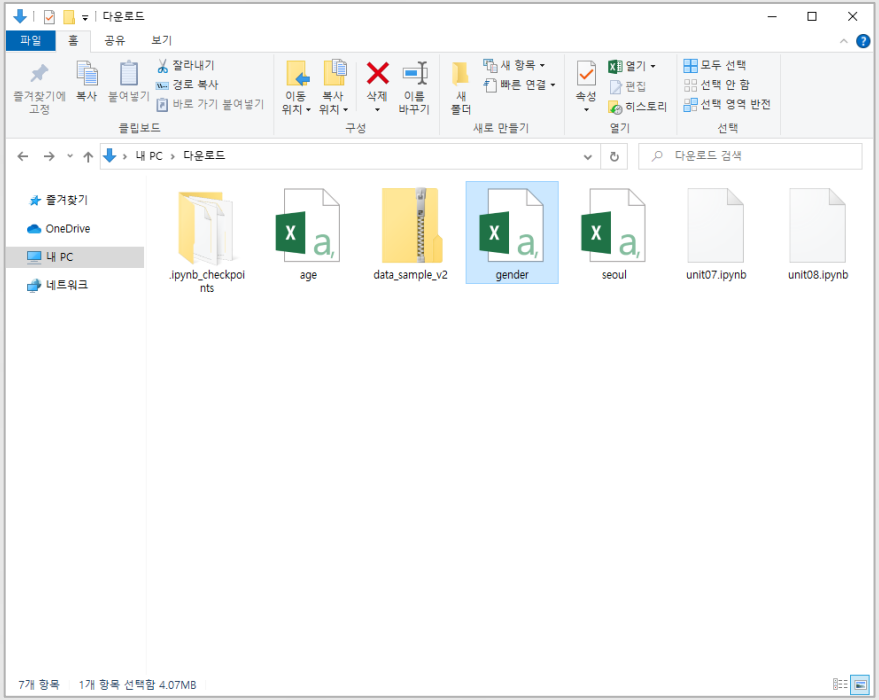
아래와 같이 경고 창이 뜨면 [확인] 버튼을 클릭하세요.





남·여 인구 공공데이터 내려받기 (6/8)

- 파일 이름을 “gender.csv”로 변경합니다.
- 파일은 다운로드(Downloads) 폴더에 저장되어 있다고 가정하고 실습을 진행하겠습니다.



남·여 인구 공공데이터 내려받기 (7/8)

gender.csv 파일을 열어 봅니다

- B~CZ열: 남성 인구수 데이터
- DA~GY열: 여성 인구수 데이터

	CW	CX	CY	CZ	DA	DB	DC	DD	DE
1	2022년08월_남_97세	2022년08월_남_98세	2022년08월_남_99세	2022년08월_남_100세 이상	2022년08월_여_총인구수	2022년08월_여_연령구간인구수	2022년08월_여_0세	2022년08월_여_1세	2022년08월_여_2세
2	301	240	246	414	4,887,159	4,887,159	19,816	21,718	22,053
3	12	9	8	15	74,091	74,091	212	236	260
4	3	1	0	1	6,374	6,374	24	22	20
5	2	2	1	1	5,154	5,154	17	16	21
6	0	1	0	2	1,254	1,254	2	4	3
7	0	1	0	0	4,935	4,935	10	14	11
8	2	1	2	2	9,570	9,570	21	46	51
9	0	0	0	1	4,349	4,349	20	19	27
10	0	0	0	0	5,365	5,365	27	32	42
11	0	0	0	0	2,148	2,148	6	4	3
12	0	1	1	4	2,937	2,937	5	11	7
13	0	0	1	1	2,380	2,380	2	5	3
14	0	0	0	0	3,767	3,767	14	8	9
15	1	1	3	1	8,501	8,501	16	19	19
16	2	0	0	0	2,257	2,257	1	5	3
17	1	0	0	0	3,799	3,799	8	3	11
18	1	0	0	2	3,297	3,297	18	13	16
19	0	0	0	0	3,027	3,027	10	2	10
20	0	1	0	0	4,977	4,977	11	13	4
21	7	7	3	14	62,651	62,651	299	255	289
22	0	0	0	0	1,156	1,156	5	5	3
23	0	1	0	1	2,281	2,281	11	6	8

남·여 인구 공공데이터 내려받기 (8/8)

인구수 데이터와 인덱스(Index)의 관계

		남성 연령별 인구수							여성 연령별 인구수						
열 이름	지역명	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상
인덱스	0	1	2	3	4	…	102	103	104	105	106	107	…	205	206

		남성 연령별 인구수							여성 연령별 인구수						
열 이름	지역명	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상
인덱스	0	1	2	3	4	…	102	103	-103	-102	-101	-100	…	-2	-1

LESSON 02

우리 동네 인구 구조 시각화

우리 동네 인구 구조 시각화하기 (1/8)

기준 폴더 준비

- [x] CSV 파일, age.csv 준비
- [x] 소스 파일 lecture-16-17-18-age.ipynb 파일 생성해 코딩

인구 데이터 파일(age.csv)을 읽어와서 한 줄씩 출력하는 코드 작성

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

for row in data:
    print(row)

f.close()
```

우리 동네 인구 구조 시각화하기 (2/8)

in 연산자를 사용하여 우리 동네 이름이 포함된 지역명 찾기

```
print('신도림' in '서울특별시 구로구 신도림동(1153051000)')
print('1153' in '서울특별시 구로구 신도림동(1153051000)')
print('()' in '서울특별시 구로구 신도림동(1153051000)')
```

```
True
True
False
```

“A in B”는 A가 B 안에 존재하면 참(True), 존재하지 않으면 거짓(False)

우리 동네 인구 구조 시각화하기 (3/8)

in 연산자를 활용하여 우리 동네 인구수 데이터 출력하기

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

for row in data:
    if '신도림' in row[0]:
        print(row)

f.close()
```

```
['서울특별시 구로구 신도림동(1153051000)', '35,964', '35,964', '274', '270', '319', '304', '321', '377', '426', '403', '374', '391', '408', '384', '370', '400', '354', '322', '332', '311', '328', '353', '359', '422', '355', '370', '429', '451', '456', '441', '454', '486', '501', '472', '561', '619', '636', '612', '619', '664', '701', '687', '760', '691', '712', '637', '628', '602', '626', '636', '663', '600', '616', '579', '567', '511', '462', '487', '509', '472', '426', '479', '477', '484', '453', '457', '412', '370', '450', '368', '340', '338', '229', '278', '249', '248', '265', '161', '141', '134', '135', '154', '121', '107', '97', '94', '71', '69', '58', '53', '35', '30', '23', '14', '16', '13', '11', '8', '8', '4', '4', '2', '4']
```

우리 동네 인구 구조 시각화하기 (4/8)

인구수 데이터를 저장하기에 앞서 헤더(Header)를 살펴봅니다.

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
print(header)

# for row in data:
#     if '신도림' in row[0]:
#         print(row)

f.close()
```

인구수 데이터는
row[3]부터 저장되어 있겠군요.

```
['행정구역', '2021년11월_계_총인구수', '2021년11월_계_연령구간인구수', '2021년11월_계_0세', '2021년11월_계_1세', '2021년11월_계_2세', '2021년11월_계_3세', '2021년11월_계_4세', '2021년11월_계_5세', '2021년11월_계_6세', '2021년11월_계_7세', '2021년11월_계_8세', '2021년11월_계_9세', '2021년11월_계_10세', '2021년11월_계_11세', '2021년11월_계_12세', '2021년11월_계_13세', '2021년11월_계_14세', '2021년11월_계_15세', '2021년11월_계_16세', '2021년11월_계_17세', '2021년11월_계_18세', '2021년11월_계_19세', '2021년11월_계_20세', '2021년11월_계_21세', '2021년11월_계_22세', '2021년11월_계_23세', '2021년11월_계_24세', '2021년11월_계_25세', '2021년11월_계_26세', '2021년11월_계_27세', '2021년11월_계_28세', '2021년11월_계_29세', '2021년11월_계_30세', '2021년11월_계_31세', '2021
```

우리 동네 인구 구조 시각화하기 (5/8)

문자열 자료형을 정수 자료형으로 변환하기

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)

result = []
for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(int(i))

f.close()
print(result)
```

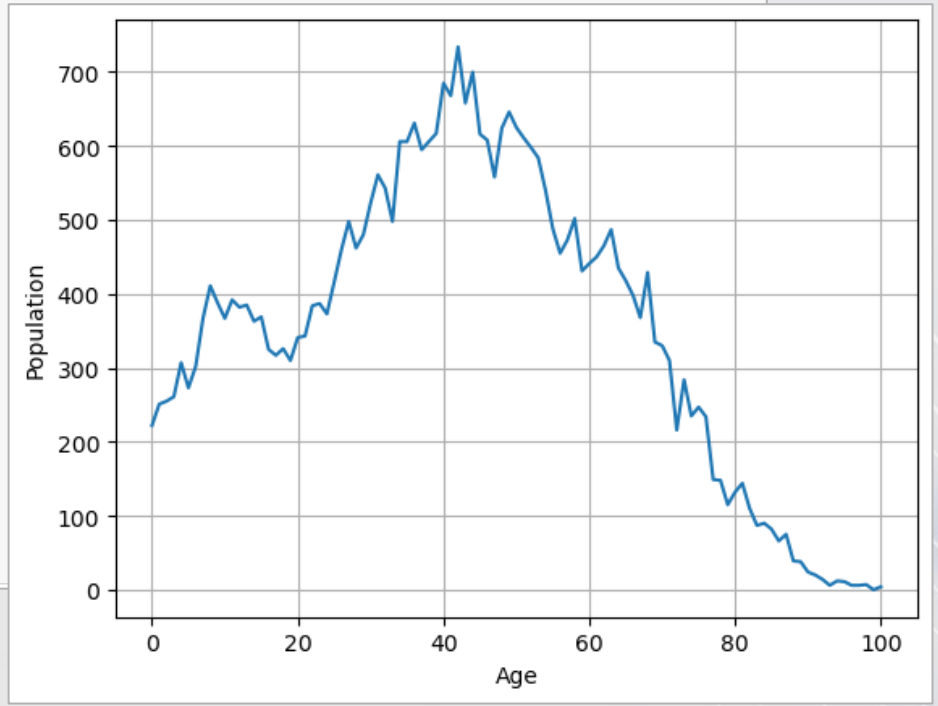
작은 따옴표 (‘’)가 사라졌습니다.

```
[274, 270, 319, 304, 321, 377, 426, 403, 374, 391, 408, 384, 370, 400, 354, 322, 332, 311, 328, 353, 359, 422, 355, 370, 429, 45
1, 456, 441, 454, 486, 501, 472, 561, 619, 636, 612, 619, 664, 701, 687, 760, 691, 712, 637, 628, 602, 626, 636, 663, 600, 616, 5
79, 567, 511, 462, 487, 509, 472, 426, 479, 477, 484, 453, 457, 412, 370, 450, 368, 340, 338, 229, 278, 249, 248, 265, 161, 141,
134, 135, 154, 121, 107, 97, 94, 71, 69, 58, 53, 35, 30, 23, 14, 16, 13, 11, 8, 8, 4, 4, 2, 4]
```

우리 동네 인구 구조 시각화하기 (6/8)

우리 동네의 연령별 인구수 데이터를 시각화하기

```
import csv
import matplotlib.pyplot as plt

f = open('age.csv')
data = csv.reader(f)

header = next(data)

result = []
for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(int(i))

f.close()

plt.plot(result)
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```



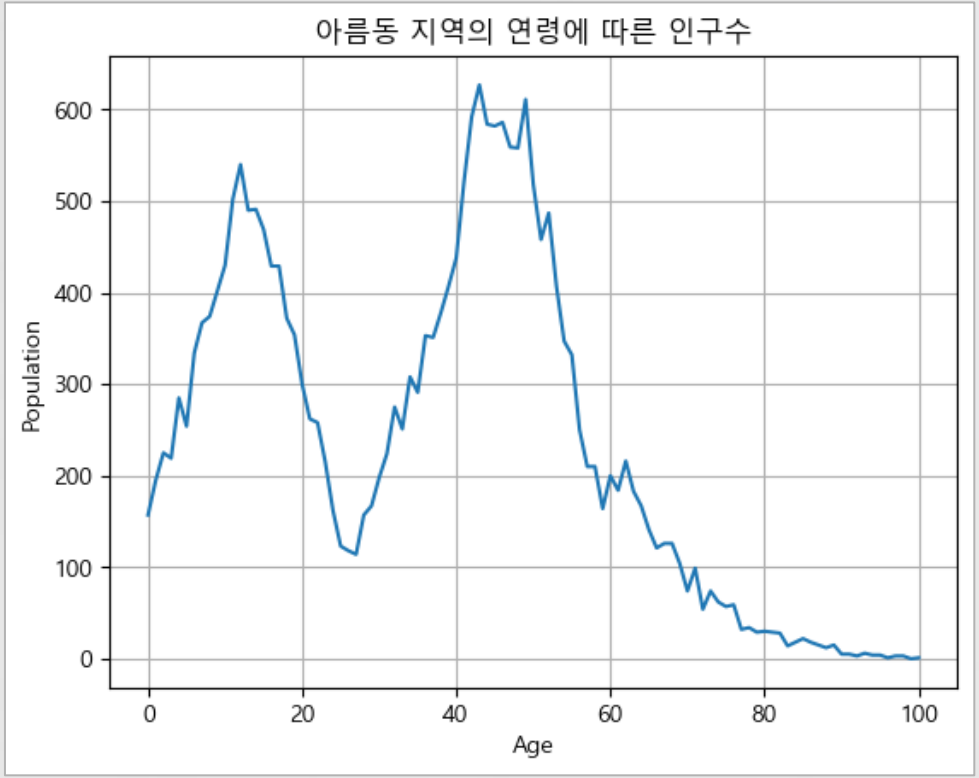
우리 동네 인구 구조 시각화하기 (7/8)

지역명을 입력 받아서 연령별 인구수 데이터를 시각화하기

```
import csv
import matplotlib.pyplot as plt

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

result = []
name = input('인구 구조가 알고 싶은 지역의 이름(읍면동 단위)을 입력해주세요 : ')
for row in data :
    if name in row[0] :
        for i in row[3:] :
            result.append(int(i))

f.close()

plt.rc('font', family='Malgun Gothic')
plt.rc('axes', unicode_minus=False)

plt.plot(result)
plt.title(f'{name} 지역의 연령에 따른 인구수')
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```

우리 동네 인구 구조 시각화하기 (8/8)

지역명을 입력 받아서 연령별 인구수 데이터를 시각화하기



SUMMARY

학습정리

행정안전부에서 인구 데이터를 내려 받기

- age.xls, gender.xls

특정 지역의 연령별 인구 수 그래프

- ‘지역명’ in row[0]
- 0세부터 100세 이상: row[3:]

표준입력으로 받은 지역의 연령별 인구 수 그래프

```
result = []
name = input('인구 구조가 알고 싶은 지역의 이름(읍면동 단위)을 입력해주세요 : ')
for row in data :
    if name in row[0] :
        for i in row[3:] :
            result.append(int(i))

f.close()
```