

⋮ 문제해결을 위한 데이터 분석 및 시각화

데이터 분석을 위한 Pandas(3)_인구구조 데이터 분석 실습

한성대학교 노은희 교수

“미래로 향하는 새로운 이정표”



오늘의 학습

학습내용

- 인구구조 데이터 분석 실습



Python DataFrame 문자열 포함 확인 `str.contains()`

DataFrame에 저장한 데이터에 문자열이 검색 문자열이 포함되어 있는지 판단

pandas.Series 문자열 메서드인 `str.contains()`를 사용하면 지정한 문자열이 포함되어 있는지 확인

검색 문자열이 포함되어 있는 경우에는 `True`를 반환

str.contains() 사용 방법

```
import pandas as pd

df = pd.DataFrame({'name': ['김길동','홍사림','이지수','김지희','나진수','이채리'],
                  'age': [20,40,17,40,30,50],
                  'point': [60,20,70,70,80,55]}
)

print(df['name'].str.contains('김'))
```

str.contains() 사용 방법

str.contains() 사용 방법

contains에 지정한 '김' 문자열이 포함된 행은 True를 반환하고 문자열이 포함되지 않은 행은 False를 반환

반환된 결과값은 pandas.Series 형태로 반환

str.contains는 지정한 문자열이 완전히 일치하는 값이 아닌 부분 일치 하는 값을 찾는다.

```
1 import pandas as pd
2
3 df = pd.DataFrame({'name': ['김길동', '홍사림', '이지수', '김지희', '나진수', '이채리'],
4                    'age': [20, 40, 17, 40, 30, 50],
5                    'point': [60, 20, 70, 70, 80, 55]}
6                    )
7
8 print(df['name'].str.contains('김'))
```

```
0    True
1   False
2   False
3    True
4   False
5   False
Name: name, dtype: bool
```

```
print(df['name'].str.contains('김'))
```

str.contains() 사용 방법

DataFrame으로 추출하기 -> name컬럼에 '김' 문자열이 포함된 행의 값을 전부 표시

```
1 import pandas as pd
2
3 df = pd.DataFrame({'name': ['김길동', '홍사림', '이지수', '김지희', '나진수', '이채리'],
4                     'age': [20, 40, 17, 40, 30, 50],
5                     'point': [60, 20, 70, 70, 80, 55]})
6
7
8 df[df['name'].str.contains('김')]
```

	name	age	point
0	김길동	20	60
3	김지희	40	70

str.contains() 사용 방법

NaN 값이 존재하는 경우 -> 문자열에 검색할 컬럼에 NaN값이 존재하는 경우 True/False가 아닌 NaN반환

```
1 import pandas as pd
2
3 df = pd.DataFrame({'name': ['김길동', '홍사림', '이지수', 1, '나진수', '이채리'],
4                     'age': [20, 40, 17, 40, 30, 50],
5                     'point': [60, 20, 70, 70, 80, 55]})
6
7 print(df['name'].str.contains('김'))
```

```
0    True
1    False
2    False
3     NaN
4    False
5    False
Name: name, dtype: object
```

데이터를 수집하면 전산오류나 사람의 실수로 결측치(NaN)가 발생

str.contains() 사용 방법

pandas.Series로 결과를 받아 행을 취득하려고 하면 에러가 발생

```
1 import pandas as pd
2
3 df = pd.DataFrame({'name': ['김길동', '홍사림', '이지수', 1, '나진수', '이채리'],
4                   'age': [20, 40, 17, 40, 30, 50],
5                   'point': [60, 20, 70, 70, 80, 55]})
6
7
8 df[df['name'].str.contains('김')]
```

```
-----
ValueError                                Traceback (most recent call last)
C:\Users\Public\Documents\ESTsoft\CreatorTemp\ipykernel_18576\1308088444.py in <module>
      6         )
      7
----> 8 df[df['name'].str.contains('김')]

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in __getitem__(self, key)
    3446
    3447         # Do we have a (boolean) 1d indexer?
-> 3448         if com.is_bool_indexer(key):
    3449             return self._getitem_bool_array(key)
    3450

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\common.py in is_bool_indexer(key)
    137         # Don't raise on e.g. ["A", "B", np.nan], see
    138         # test_loc_getitem_list_of_labels_categorical_index_with_na
-> 139         raise ValueError(msg)
    140         return False
    141         return True
```

ValueError: Cannot mask with non-boolean array containing NA / NaN values

str.contains() 사용 방법

NaN이 존재하는 경우에는 na로 NaN 결과를 치환해 주도록 설정

```
1 import pandas as pd
2
3 df = pd.DataFrame({'name': ['김길동', '홍사림', '이지수', 1, '나진수', '이채리'],
4                    'age': [20, 40, 17, 40, 30, 50],
5                    'point': [60, 20, 70, 70, 80, 55]})
6
7 # NaN이 존재하는 경우 False로 치환
8 print(df['name'].str.contains('김', na=False))
```

```
0    True
1    False
2    False
3    False
4    False
5    False
Name: name, dtype: bool
```

```
1 import pandas as pd
2
3 df = pd.DataFrame({'name': ['김길동', '홍사림', '이지수', 1, '나진수', '이채리'],
4                    'age': [20, 40, 17, 40, 30, 50],
5                    'point': [60, 20, 70, 70, 80, 55]})
6
7 # NaN이 존재하는 경우 True로 치환
8 print(df['name'].str.contains('김', na=True))
```

```
0    True
1    False
2    False
3     True
4    False
5    False
Name: name, dtype: bool
```

[실습]인구구조 분석

[실습 결과물]

[문제1] '2022년 2월 연령별 인구현황_월간'파일에서 우리동네 연령별 인구 분포 분석 후 엑셀 파일로 저장하기

202202_202202_연령별인구현황_월간.xlsx - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	※ 매월 말일자 통계 현황													
2														
3														
4	행정기관코드	행정기관	총 인구수	연령구간인구0세	1세	2세	3세	4세	5세	6세	7세	8세	9세	10세
5	1100000000	서울특별시	9,508,451	9,508,451	42,726	44,493	47,755	51,025	54,730	61,722	66,956	66,691	66,449	74,033
6	1111000000	서울특별시	144,575	144,575	469	478	565	544	606	748	842	817	810	966
7	1111051500	서울특별시	11,953	11,953	46	46	58	56	61	77	99	92	83	107
8	1111053000	서울특별시	9,362	9,362	42	24	43	36	50	69	79	67	63	73
9	1111054000	서울특별시	2,441	2,441	5	6	5	5	15	5	17	6	18	13
10	1111055000	서울특별시	9,541	9,541	28	17	37	30	44	56	52	62	73	89
11	1111056000	서울특별시	18,132	18,132	63	86	104	94	110	105	154	149	133	146
12	1111057000	서울특별시	8,219	8,219	42	44	38	47	64	86	63	79	76	89
13	1111058000	서울특별시	10,020	10,020	67	60	90	74	72	86	105	73	93	97
14	1111060000	서울특별시	4,063	4,063	15	5	14	13	14	21	20	21	20	24
15	1111061500	서울특별시	7,238	7,238	14	19	14	27	9	15	26	23	11	20
16	1111063000	서울특별시	5,072	5,072	7	7	6	6	8	12	13	6	8	14
17	1111064000	서울특별시	7,247	7,247	13	18	19	19	12	25	19	27	25	33
18	1111065000	서울특별시	16,409	16,409	38	38	43	42	45	64	70	90	80	103

삼선동 인구분포.xlsx - Excel

	A	B	C	D	E	F	G	H
1		서울특별시 성북구 삼선동						
2	행정기관코드	1129055500						
3	총 인구수	22,879						
4	연령구간인구수	22,879						
5	0세	69						
6	1세	62						
7	2세	64						
8	3세	83						
9	4세	78						
10	5세	106						
11	6세	123						
12	7세	123						
13	8세	118						
14	9세	150						

파일 원본

202202_202202_연령별인구현황_월간.csv

작업 후 저장파일

[파일준비] 행정안전부 홈페이지-> 정책자료->통계

행정안전부 > 정책자료 > 통계

mois.go.kr/frt/bbs/type001/commonSelectBoardList.do?bbsId=BBSMSTR_000000000013

ENGLISH 누리집 안내지도 어린이 정부혁신 보다는은 정부 국민의 나라 정치대로 대결민족 국가상징 국무회의 알아보기

내용을 입력하세요.

행정안전부 정보공개 참여·민원 뉴스·소식 **정책자료** 업무안내 기관소개

정책자료

정책자료 홈	
주요업무계획	<ul style="list-style-type: none">주요업무계획국정과제
법령정보	<ul style="list-style-type: none">법률입법·행정예고 전자공청회대통령령법령종합검색부령법령자료실훈령·예규·고시
전자관보	
통계 통계 페이지로 이동	<ul style="list-style-type: none">통계연보·주제별 통계승인통계e-나라지표주민등록 인구통계행정안전통계 시각화
간행물	
정책연구보고서	

https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardList.do?bbsId=BBSMSTR_000000000013

[파일준비]행정안전부 홈페이지-> 정책자료->통계->주민등록 인구통계

통계

통계연보·주제별 통계

승인통계

e-나라지표

주민등록 인구통계

행정안전통계 시각화

통계연보·주제별 통계

아래 주제별 통계를 클릭하시면 개별 통계관리 시스템으로 이동하여 보다 자세한 통계자료를 확인할 수 있습니다.

주제별 통계

정부조직통계 | 주민등록인구통계 | 자원봉사통계 | 지방재정통계 | 지방세통계 | 기록물통계 | 감정통계

전체 : 21건 1/3

제목

검색

전체보기

번호	제목	첨부	작성자	등록일	조회수
21	2021 행정안전통계연보('20.12.31.기준)		데이터정보화담당관	2021.08.25.	8929
20	2020 행정안전통계연보('19.12.31.기준)		정보통계담당관	2020.08.12.	20542
19	2019 행정안전통계연보('18.12.31.기준)		정보통계담당관	2019.08.27.	20984
18	2018 행정안전통계연보		정보통계담당관	2018.08.07.	29726
17	2017 재난안전통계연보		정보통계담당관	2017.12.28.	23536
16	2017 행정자치통계연보		정보통계담당관	2017.09.04.	29529

[파일준비] 주민등록 인구통계->[연령별 인구현황]

주인등록 인구통계 행정안전부

주인등록 인구통계

- 주인등록 인구 및 세대현황
- 연령별 인구현황
- 주인등록 인구 기타현황

연령별 인구현황

통계표 그래프

행정구역 전국 시·군·구 *i*

등록구분 전체 *i*

조회기간 ☒ 월간 ☐ 연간 2022년 02월 ~ 2022년 02월 *i*

※ 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

다운로드 받은 파일 확인하기

csv파일



202202_202202_연령별인구현황_월간.csv

202202_202202_연령별인구현황_월간.csv - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 개발 도구 파워 피벗 수행할 작업을 알려 주세요. 공유

잘라내기 붙여넣기 복사 서식 복사 클립보드 글꼴 맞춤 표시 형식 스타일 셀 편집

A1 행정구역

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	행정구역	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02	2022년02
2	서울특별시	9,508,451	9,508,451	42,726	44,493	47,755	51,025	54,730	61,722	66,956	66,691	66,449	74,033	71,822	73,949	69,246	73,318	79,4
3	서울특별시	144,575	144,575	469	478	565	544	606	748	842	817	810	966	900	1,002	942	992	1,
4	서울특별시	11,953	11,953	46	46	58	56	61	77	99	92	83	107	109	108	109	129	
5	서울특별시	9,362	9,362	42	24	43	36	50	69	79	67	63	73	68	72	59	76	
6	서울특별시	2,441	2,441	5	6	5	5	15	5	17	6	18	13	17	11	23	18	
7	서울특별시	9,541	9,541	28	17	37	30	44	56	52	62	73	89	59	80	79	71	
8	서울특별시	18,132	18,132	63	86	104	94	110	105	154	149	133	146	133	180	132	145	
9	서울특별시	8,219	8,219	42	44	38	47	64	86	63	79	76	89	94	101	96	100	
10	서울특별시	10,020	10,020	67	60	90	74	72	86	105	73	93	97	73	81	81	78	
11	서울특별시	4,063	4,063	15	5	14	13	14	21	20	21	20	24	38	29	35	32	
12	서울특별시	7,238	7,238	14	19	14	27	9	15	26	23	11	20	24	27	21	22	
13	서울특별시	5,072	5,072	7	7	6	6	8	12	13	6	8	14	15	16	20	14	
14	서울특별시	7,247	7,247	13	18	19	19	12	25	19	27	25	33	27	31	29	34	
15	서울특별시	16,409	16,409	38	38	43	42	45	64	70	90	80	103	102	117	116	110	
16	서울특별시	4,905	4,905	9	11	9	12	10	12	22	13	13	16	17	18	17	20	
17	서울특별시	7,875	7,875	11	19	19	23	18	19	21	18	21	24	35	28	38	32	
18	서울특별시	6,551	6,551	23	33	39	21	34	48	28	36	39	47	25	38	37	53	

202202_202202_연령별인구현황_월간

다운로드 받은 파일 확인하기

Xlsx 파일



202202_202202_연령별인구현황_월간.xlsx

202202_202202_연령별인구현황_월간.xlsx - Excel

파일홈삽입페이지 레이아웃수식데이터검토보기개발 도구파워 피벗수행할 작업을 알려 주세요.공유

붙여넣기

잘라내기

복사

서식 복사

클립보드

맑은 고딕

11

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가

가</

다운로드 받은 파일 이름 변경하기



202202_202202_연령별인구현황_월간.csv



파일명 간단하게 수정해서 사용하기
population.csv



202202_202202_연령별인구현황_월간.xlsx



Population.xlsx

다운로드 받은 파일 Pandas 라이브러리를 사용하여 불러오기

```
1 import pandas as pd
2 df = pd.read_csv('population.csv', encoding = 'cp949', low_memory=False, index_col = 0)
3 df.head(n=3)
```

	2022년02 월_계_총 인구수	2022년02 월_계_연 령구간인 구수	2022 년02 월_계 _0세	2022 년02 월_계 _1세	2022 년02 월_계 _2세	2022 년02 월_계 _3세	2022 년02 월_계 _4세	2022 년02 월_계 _5세	2022 년02 월_계 _6세	2022 년02 월_계 _7세	...	2022 년02 월_계 _91 세	2022 년02 월_계 _92 세	2022 년02 월_계 _93 세	2022 년02 월_계 _94 세	2022 년02 월_계 _95 세	2022 년02 월_계 _96 세	2022 년02 월_계 _97 세	2022 년02 월_계 _98 세	2022 년02 월_계 _99 세
행정구역																				
서울특별시 (1100000000)	9,508,451	9,508,451	42,726	44,493	47,755	51,025	54,730	61,722	66,956	66,691	...	8,025	6,853	5,356	4,119	2,851	1,867	1,357	1,202	9
서울특별시 종로구 (1111000000)	144,575	144,575	469	478	565	544	606	748	842	817	...	170	143	113	89	80	52	32	22	
서울특별시 종로구 청운 효자동 (1111051500)	11,953	11,953	46	46	58	56	61	77	99	92	...	9	10	8	7	4	5	3	3	

3 rows × 103 columns

다운로드 받은 파일 Pandas 라이브러리를 사용하여 불러오기

1. Pandas로 csv 파일과 excel 파일 불러오기(엑셀 파일)

```
1 import pandas as pd
2 df = pd.read_excel('population.xlsx')
3 df.head(n=3)
```

C:\ProgramData\Anaconda3\lib\site-packages\openpyxl\styles\stylesheet.py:226: UserWarning: Workbook contains no default style, apply openpyxl's default
warn("Workbook contains no default style, apply openpyxl's default")

※
매월
말일
자
통계
현황

	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 95	Unnamed: 96	Unnamed: 97	Unnamed: 98
0	NaN	NaN	2022년02월	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
1	NaN	NaN	계	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
2	행정기관 코드	행정기관	총 인구수	연령구간 인구수	0세	1세	2세	3세	4세	5세	...	91세	92세	93세

3 rows × 105 columns

엑셀에서 불필요한 행 삭제하기

The screenshot shows the Microsoft Excel interface with a file named 'population.xlsx'. The ribbon is set to '홈' (Home). A context menu is open over a selected range of rows (rows 4 to 26), showing options like '잘라내기(I)', '복사(C)', '붙여넣기 옵션:', '선택하여 붙여넣기(S)...', '삽입(I)', '삭제(D)', '내용 지우기(N)', '셀 서식(E)...', '행 높이(R)...', '숨기기(H)', and '숨기기 취소(U)'. The '삭제(D)' option is highlighted. A yellow text box is overlaid on the table with the text: '지우고자 하는 행을 선택후 -> 마우스 오른쪽 버튼을 클릭하여-> 삭제하기'.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1																		
2																		
3																		
4		총 인														1세	12세	13세
5		9,50														3,949	69,246	73,318
6		144														002	942	992
7		11,953	11,953	46	46	58	56	61	77	99	92	83	107	109	108	109	129	
8		9,362	9,362	42	24	43	36	50	69	79	67	63	73	68	72	59	76	
9		2,441	2,441	5	6	5	5	15	5	17	6	18	13	17	11	23	18	
10		9,541	9,541	28	17	37	30	44	56	52	62	73	89	59	80	79	71	
11		18,132	18,132	63	86	104	94	110	105	154	149	133	146	133	180	132	145	
12		8,219	8,219	42	44	38	47	64	86	63	79	76	89	94	101	96	100	
13		10,020	10,020	67	60	90	74	72	86	105	73	93	97	73	81	81	78	
14		4,063	4,063	15	5	14	13	14	21	20	21	20	24	38	29	35	32	
15		7,238	7,238	14	19	14	27	9	15	26	23	11	20	24	27	21	22	
16		5,072	5,072	7	7	6	6	8	12	13	6	8	14	15	16	20	14	
17	1111064000	서울특별시	7,247	7,247	13	18	19	19	12	25	19	27	25	33	27	31	29	34
18	1111065000	서울특별시	16,409	16,409	38	38	43	42	45	64	70	90	80	103	102	117	116	110
19	1111067000	서울특별시	4,905	4,905	9	11	9	12	10	12	22	13	13	16	17	18	17	20
20	1111068000	서울특별시	7,875	7,875	11	19	19	23	18	19	21	18	21	24	35	28	38	32
21	1111069000	서울특별시	6,551	6,551	23	33	39	21	34	48	28	36	39	47	25	38	37	53
22	1111070000	서울특별시	5,850	5,850	19	18	10	20	21	16	24	35	30	43	32	28	28	27
23	1111071000	서울특별시	9,697	9,697	27	27	17	19	19	32	30	20	24	28	32	37	22	31
24	1114000000	서울특별시	122,167	122,167	562	561	570	570	607	669	671	609	669	665	641	627	524	585
25	1114052000	서울특별시	2,241	2,241	13	14	8	16	16	13	25	14	25	19	18	8	10	14
26	1114054000	서울특별시	4,749	4,749	15	9	15	10	18	16	17	23	31	21	24	17	23	19

다운로드 받은 파일 Pandas 라이브러리를 사용하여 불러오기

엑셀파일에서 3행을 제거후 엑셀 파일 다시 불러오기

```
1 import pandas as pd
2 df = pd.read_excel('population.xlsx')
3 df.head(n=3)
```

	행정기관코드	행정기관	총 인구수	연령구간 인구수	0세	1세	2세	3세	4세	5세	...	91세	92세	93세	94세	95세	96세	97세	98세	99세	100세 이상
0	1100000000	서울특별시	9,508,451	9,508,451	42,726	44,493	47,755	51,025	54,730	61,722	...	8,025	6,853	5,356	4,119	2,851	1,867	1,357	1,202	900	1,708
1	1111000000	서울특별시종로구	144,575	144,575	469	478	565	544	606	748	...	170	143	113	89	80	52	32	22	26	37
2	1111051500	서울특별시종로구	11,953	11,953	46	46	58	56	61	77	...	9	10	8	7	4	5	3	3	2	2

파일에서 필요없는 열 삭제하기

2. 엑셀 파일에서 필요없는 열 삭제하기

```
1 import pandas as pd
2 df = pd.read_excel('population.xlsx')
3 del df['행정기관코드'], df['총 인구수'], df['연령구간인구수'] # 행정기관코드, 총 인구수, 연령구간인구수 삭제하기
4 df.head(n=5)
```

	행정기관	0세	1세	2세	3세	4세	5세	6세	7세	8세	...	91세	92세	93세	94세	95세	96세	97세	98세	99세	100세 이상
0	서울특별시	42,726	44,493	47,755	51,025	54,730	61,722	66,956	66,691	66,449	...	8,025	6,853	5,356	4,119	2,851	1,867	1,357	1,202	900	1,708
1	서울특별시 종로구	469	478	565	544	606	748	842	817	810	...	170	143	113	89	80	52	32	22	26	37
2	서울특별시 종로구 청 운효자동	46	46	58	56	61	77	99	92	83	...	9	10	8	7	4	5	3	3	2	2
3	서울특별시 종로구 사 직동	42	24	43	36	50	69	79	67	63	...	13	16	10	9	6	7	4	2	2	1
4	서울특별시 종로구 삼 청동	5	6	5	5	15	5	17	6	18	...	4	6	2	3	3	0	2	0	2	2

5 rows × 102 columns

궁금한 지역의 연령별 인구 추출하기

3. 궁금한 지역 이름 입력받고 해당 지역의 인구 추출하기

```
1 import pandas as pd
2 df = df = pd.read_excel('population.xlsx', index_col = 1)
3
4 name = input('원하는 지역의 이름을 입력=>')    #input()함수를 통해 입력
5
6 a = df.index.str.contains(name)
7
8 df1 = df[a]
9 df1
```

원하는 지역의 이름을 입력=>삼선동

		행정기관코드	총 인구 수	연령구간인구 수	0세	1세	2세	3세	4세	5세	6세	...	91세	92세	93세	94세	95세	96세	97세	98세	99세	100세 이상
행정기관																						
서울특별시	성북구 삼선동	1129055500	22,879	22,879	69	62	64	83	78	106	123	...	26	14	13	14	7	5	3	0	4	10

1 rows × 104 columns



[개인 실습하기]

[문제 풀기] 본인이 궁금한 지역(살고 있는 지역)의 연령별 인구 추출하기

전체열과 전체행 보이기

row 생략 없이 출력 `pd.set_option('display.max_rows', None)`
col 생략 없이 출력 `pd.set_option('display.max_columns', None)`

0	서울특별시 (1100000000)	9,508,451	9,508,451	42,726	44,493	47,755	51,025	54,730	61,722	66,956	...	5,901	5,042	4,031	3,178	2,227	1,479	1,042	898	685
1	서울특별시 종로구 (1111000000)	144,575	144,575	469	478	565	544	606	748	842	...	117	104	81	68	58	34	20	14	14
2	서울특별시 종로구 청운 효자동 (1111051500)	11,953	11,953	46	46	58	56	61	77	99	...	6	8	6	4	3	3	1	2	1
3	서울특별시 종로구 사직 동 (1111053000)	9,362	9,362	42	24	43	36	50	69	79	...	10	10	5	4	4	4	0	2	1
4	서울특별시 종로구 삼청 동 (1111054000)	2,441	2,441	5	6	5	5	15	5	17	...	2	4	2	2	0	0	1	0	1
...
3861	제주특별자치 도 서귀포시 서홍동 (5013058000)	11,192	11,192	79	72	89	98	117	143	167	...	5	3	8	1	1	2	0	0	2

데이터 정제

분석하기 좋은 데이터 : 분석하기 좋은 데이터는 데이터 집합을 분석하기 좋은 상태로 만들어 놓은 것

데이터 분석단계에서 데이터 정리가 중요(데이터 분석 작업의 70% 이상이 데이터 정리작업)

분석하기 좋은 데이터 조건

- 데이터 분석 목적에 맞는 데이터를 모아 새로운 표(Table)을 만들기
- 측정한 값은 행(row)을 구성
- 변수는 열(column)로 구성

데이터프레임의 행과 열 바꾸기

데이터프레임의 행과 열을 바꾸기 위해 `df.transpose()` 사용

행정기관	0세	1세	2세	3세	4세	5세	6세	7세	8세	9세	...	91세	92세	93세	94세	95세	96세	97세	98세	99세	100세 이상
서울특별시 성북구 삼선동	69	62	64	83	78	106	123	123	118	150	...	26	14	13	14	7	5	3	0	4	10

1 rows × 101 columns

```
1 ### 데이터를 행과 열 바꾸기
```

```
1 filt = df1.transpose()           #행과 열 바꾸기
2 filt.head(n=filt)
```

행정기관	서울특별시 성북구 삼선동
행정기관코드	1129055500
총 인구수	22,879
연령구간인구수	22,879
0세	69
1세	62
2세	64
3세	83
4세	78
5세	106
6세	123
7세	123

엑셀로 저장하고 확인하기

`filt.to_excel('삼선동 인구분포.xlsx')`

	A	B	C	D	E	F	G	H
1		서울특별시 성북구 삼선동						
2	행정기관코드	1129055500						
3	총 인구수	22,879						
4	연령구간인구수	22,879						
5	0세	69						
6	1세	62						
7	2세	64						
8	3세	83						
9	4세	78						
10	5세	106						
11	6세	123						
12	7세	123						
13	8세	118						
14	9세	150						

[참고] CSV 파일에서 iloc로 열과 행 추출하기

```
1 import pandas as pd
2 df = pd.read_csv('age.csv', encoding = 'cp949', low_memory=False)
3 filt = df.iloc[0:5,0:10]
4 filt
```

	행정구역	2022년02월_계_ 총인구수	2022년02월_계_연 령구간인구수	2022년02월 _계_0세	2022년02월 _계_1세	2022년02월 _계_2세	2022년02월 _계_3세	2022년02월 _계_4세	2022년02월 _계_5세	2022년02월 _계_6세
0	서울특별시 (1100000000)	9,508,451	9,508,451	42,726	44,493	47,755	51,025	54,730	61,722	66,956
1	서울특별시 종로구 (1111000000)	144,575	144,575	469	478	565	544	606	748	842
2	서울특별시 종로구 청운호 자동(1111051500)	11,953	11,953	46	46	58	56	61	77	99
3	서울특별시 종로구 사직동 (1111053000)	9,362	9,362	42	24	43	36	50	69	79
4	서울특별시 종로구 삼청동 (1111054000)	2,441	2,441	5	6	5	5	15	5	17