

20221797 원영진 6주차 워크시트

목표설정: 출생율과 취업률, 주택비율의 상관관계를 알아보자

데이터 수집:

- 합계출산율 (https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1428)
- 인구 천명당 주택수 (https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1228#quick_02;)
- 고용센터 구인, 구직 및 취업현황 (https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1505#quick_02;)

```
In [ ]: #필요한 라이브러리 읽어오기
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from matplotlib import font_manager, rc

path = "c:/windows/Fonts/malgun.ttf"
font_name = font_manager.FontProperties(fname=path).get_name()
rc('font', family=font_name)
```

데이터 정제

```
In [248... 출생율_rate = pd.read_excel('합계 출산율.xls')
주택수_rate = pd.read_excel('인구 천명당 주택수.xls')
취업률_rate = pd.read_excel('구인, 구직 및 취업현황.xls')
```

```
In [249... 출생율_rate.head()
```

Out[249]:

	Unnamed: 0	1970	1971	1972	1973	1974	1975	1976	1977	1978	...	2011
0	출생아 수	1,007.0	1,025.0	953.00	966.00	923.00	874.00	796.0	825.00	751.00	...	471.300
1	합계 출산율	4.530	4.540	4.12	4.07	3.77	3.43	3.0	2.99	2.64	...	1.244
2	출처: 통계청, 『2021년 출생통계(확정), 국가승인통계 제10103호 출생통계』		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
3	주석: * 합계 출산율 (TFR, Total Fertility Rate): 여성 1명이 ...		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN

4 rows × 52 columns



In [250...

주택수_rate.head()

Out[250]:

	Unnamed: 0	2015년 인구수	2015년 주택수	2015년 주택수/인구 천명	2016년 인구수	2016년 주택수	2016년 주택수/인구 천명	2017년 인구수	2017년 주택수	2017년 주택수/인구 천명	2018년 인구수	2018년 주택수
0	전국 계	51,069	19,559	383.0	51,270	19,877	387.7	51423	20,313	395.0	51,630	20,818
1	수도권	25,274	9,017	356.8	25,390	9,161	360.8	25,519	9,336	365.8	25,713	9,588
2	서울	9,904	3,633	366.8	9,806	3,644	371.6	9,742	3,672	376.5	9,674	3,682
3	부산	3,449	1,370	397.3	3,440	1,376	399.8	3,417	1,396	408.6	3,395	1,413
4	대구	2,466	943	382.6	2,461	966	392.6	2,453	988	402.9	2,444	996



In [251...

취업률_rate.head()

Out[251]:

	Unnamed: 0	2001	2002	2003	2004	2005	2006	2007	2008
0	구인	615,306	642,032	489,168	608,117	705,302	769,796	1,008,448	1,112,341
1	구직	1,025,813	873,296	932,826	1,112,943	1,377,236	1,607,454	1,967,073	2,093,557
2	구인배수	0.60	0.74	0.52	0.55	0.51	0.48	0.51	0.53
3	취업자수	237,368	191,979	180,450	223,757	317,170	398,033	501,256	544,122
4	취업률	23.1	22.0	19.3	20.1	23.0	24.8	25.5	26.0

5 rows × 22 columns

In [252...]

```
출생율_rate.drop(출생율_rate.iloc[:,46], axis = 1, inplace = True)
출생율_rate.drop([0,2,3], axis = 0, inplace = True)
출생율_rate.reset_index(drop = True, inplace = True)
출생율_rate.head()
```

Out[252]:

	2015	2016	2017	2018	2019	2020
0	1.239	1.172	1.052	0.977	0.918	0.84

In [253...]

```
주택수_rate = 주택수_rate[['2015년 주택수/인구천명', '2016년 주택수/인구천명', '2017년 주택수/인구천명', '2018년 주택수/인구천명', '2019년 주택수/인구천명', '2020년 주택수/인구천명']]
주택수_rate.head()
```

Out[253]:

	2015년 주택수/인구천명	2016년 주택수/인구천명	2017년 주택수/인구천명	2018년 주택수/인구천명	2019년 주택수/인구천명	2020년 주택수/인구천명
0	383.0	387.7	395.0	403.2	411.6	418.2

In [254...]

```
취업률_rate.drop(취업률_rate.iloc[:,15], axis = 1, inplace = True)
취업률_rate.drop(columns = ['2021'], axis = 1, inplace = True)
취업률_rate = 취업률_rate[4:5]
취업률_rate.reset_index(drop = True, inplace = True)
취업률_rate.head()
```

Out[254]:

	2015	2016	2017	2018	2019	2020
0	29.9	29.8	27.1	25.7	25.1	21.8

In [307...]

```
출생율 = np.array(출생율_rate)
주택수 = np.array(주택수_rate/10)
취업률 = np.array(취업률_rate)

Total = pd.DataFrame({'연도별': ['2015년', '2016년', '2017년', '2018년', '2019년', '2020년'], '출생율': 출생율, '주택수': 주택수, '취업률': 취업률})
Total
```

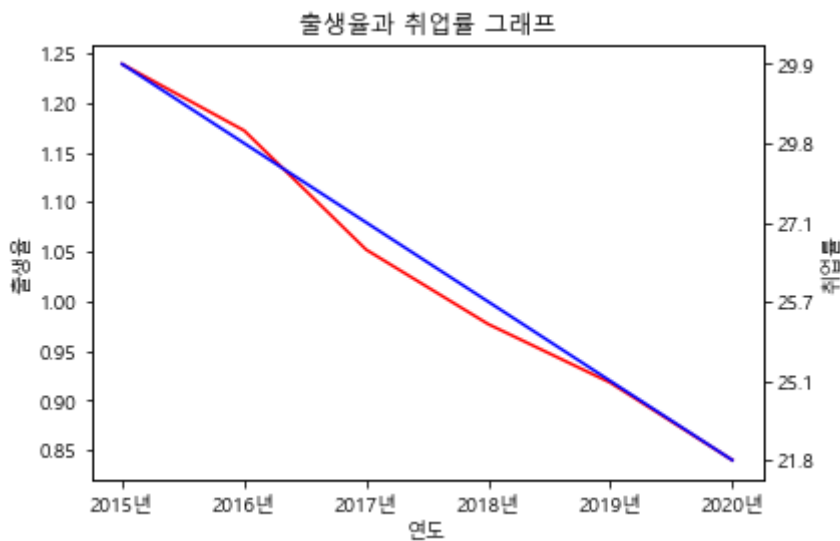
Out[307]:

	연도별	출생율	인구대비 주택비율	취업률
0	2015년	1.239	38.30	29.9
1	2016년	1.172	38.77	29.8
2	2017년	1.052	39.50	27.1
3	2018년	0.977	40.32	25.7
4	2019년	0.918	41.16	25.1
5	2020년	0.840	41.82	21.8

데이터 탐색(시각화)

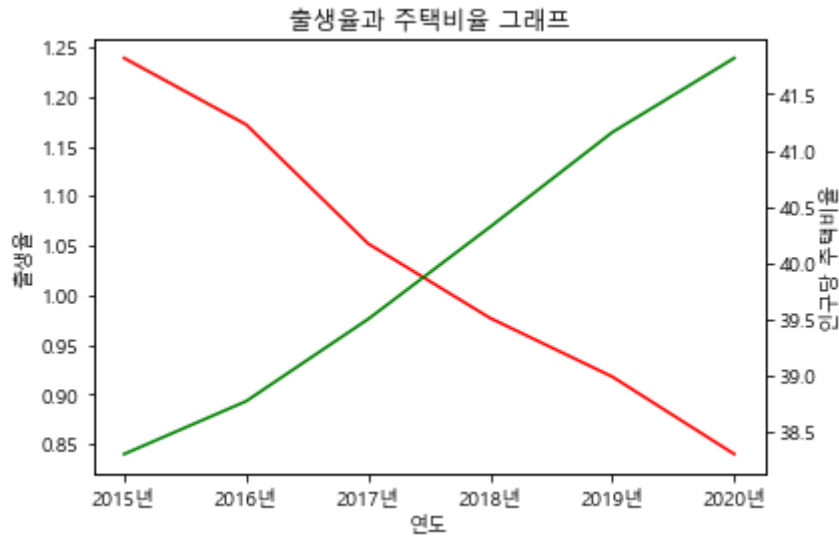
```
In [327... fig, ax1 = plt.subplots()
ax2 = ax1.twinx()
ax1.plot(Total['연도별'], Total['출생율'], color = "red")
ax2.plot(Total['연도별'], Total['취업률'], color = "blue")

plt.title("출생율과 취업률 그래프")
ax1.set_xlabel("연도")
ax1.set_ylabel("출생율")
ax2.set_ylabel("취업률")
plt.gca().invert_yaxis()
plt.show()
```



```
In [328... fig, ax1 = plt.subplots()
ax3 = ax1.twinx()
ax1.plot(Total['연도별'], Total['출생율'], color = "red")
ax3.plot(Total['연도별'], Total['인구대비 주택비율'], color = "green")

plt.title("출생율과 주택비율 그래프")
ax1.set_xlabel("연도")
ax1.set_ylabel("출생율")
ax3.set_ylabel("인구당 주택비율")
plt.show()
```



분석한 결과

- 내가 세운 가설은 인구대비 주택비율이 늘어나면 출생율도 늘어나고, 취업률이 줄어들면 출생율도 줄어들거라는 가설을 세웠다. 이를 알아보기 위해 e-나라지표에서 합계출산율 / 인구 천명당 주택수 / 고용센터 구인, 구직 및 취업현황 공공데이터를 가지고 분석하였다.
- 인구 천명당 주택수가 2015년부터 매년 시행되었기 때문에 6년간의 데이터 변화를 살펴 보았다. 그러기 위해 먼저 각각의 xls 파일을 불러오고, 필요한 정보가 어디 있는지 확인 하였다. 필요한 정보를 추출하는 과정은 pandas 모듈을 이용하였다.
 - 출생율의 경우 먼저, 필요없는 연도와 행을 지운 다음, index 번호를 초기화 해주었다.
 - 주택수는 필요한 열만 가져와서 주택수_rate 변수에 다시 저장하였다.
 - 취업율의 경우 필요없는 연도를 먼저 삭제했다. 그 다음 필요한 행만 추출하여 취업률_rate에 다시 저장하고 index 번호를 초기화 해주었다.
- 위 과정을 마친 후 세 데이터를 하나의 frame으로 묶기 위하여 numpy 모듈을 이용해 각각의 데이터를 생성하여 pd.DataFrame으로 Total에 저장하였다.
- 데이터를 시각화하기 위해 matplotlib 모듈을 사용하였다. 각각의 데이터 값의 차이가 커 한 눈에 알아보기 쉽도록 그래프를 각각 '출생율과 취업률 그래프', '출생율과 주택비율 그래프'로 나누었고, y축은 2가지로 나눠 각각의 데이터를 표시하였다. 또한 색상을 서로 다르게 표현하여 한눈에 들어오도록 만들었다.
- 위 과정을 통해 출산율과 취업률, 주택비율의 상관관계를 조사한 결과 지금까지 취업률이 감소함에 따라 출생율도 감소하였다. 하지만 주택비율이 증가하더라도 출생율은 감소하는 것을 알 수 있다. 이를 볼 때 취업률은 출생율에 영향을 미치지만, 인구당 주택비율은 큰 영향을 못미치는 것을 볼 수 있다.
- 이러한 결과가 발생한 이유는 주택의 수는 증가하였지만, 집값 상승과 취업률 감소로 인해 집을 사기 힘든 환경에 처한 20 ~ 30대의 결혼 시기가 늦어지고, 출산 의욕도 감소시켜 이러한 결과가 발생한 것으로 해석된다.
 - 내가 세운 가설과 비교해보면 취업률이 줄어들면 출생율도 줄어들거라는 가설은 맞았지만, 주택비율이 늘어나면 출생율도 늘어날 것이라는 가설은 틀렸다.

- 출생율과 주택비율의 상관관계를 조사하는 것보다 출생율과 집값, 물가와 비교하는 것이 더 정확한 결과를 얻을 수 있을 것 같다.

In []: