




```

import pandas as pd
df = pd.read_csv('코로나확진자_발생별(국내발생+해외유입).csv', encoding='euc-kr', thousands=',')
df = df[1:]
df['국내발생(명)'] = df['국내발생(명)'].str.replace(',', '')
df['해외유입(명)'] = df['해외유입(명)'].str.replace(',', '')
df['사망(명)'] = df['사망(명)'].str.replace(',', '')
df['국내발생(명)'] = df['국내발생(명)'].str.replace('-', '0').astype('int64')
df['해외유입(명)'] = df['해외유입(명)'].str.replace('-', '0').astype('int64')
df['사망(명)'] = df['사망(명)'].str.replace('-', '0').astype('int64')
df['일자'] = df['일자'].astype('datetime64[ns]')
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('코로나확진자 (국내발생 + 해외유입)')
plt.plot(df['일자'], df['국내발생(명)'], 'b.-', label='국내발생(명)')
plt.plot(df['일자'], df['해외유입(명)'], 'r.-', label='해외유입(명)')
plt.xlabel('일자')
#plt.ylabel('y축')
plt.legend()
plt.savefig('코로나확진자(국내발생 + 해외유입).png')
plt.show()

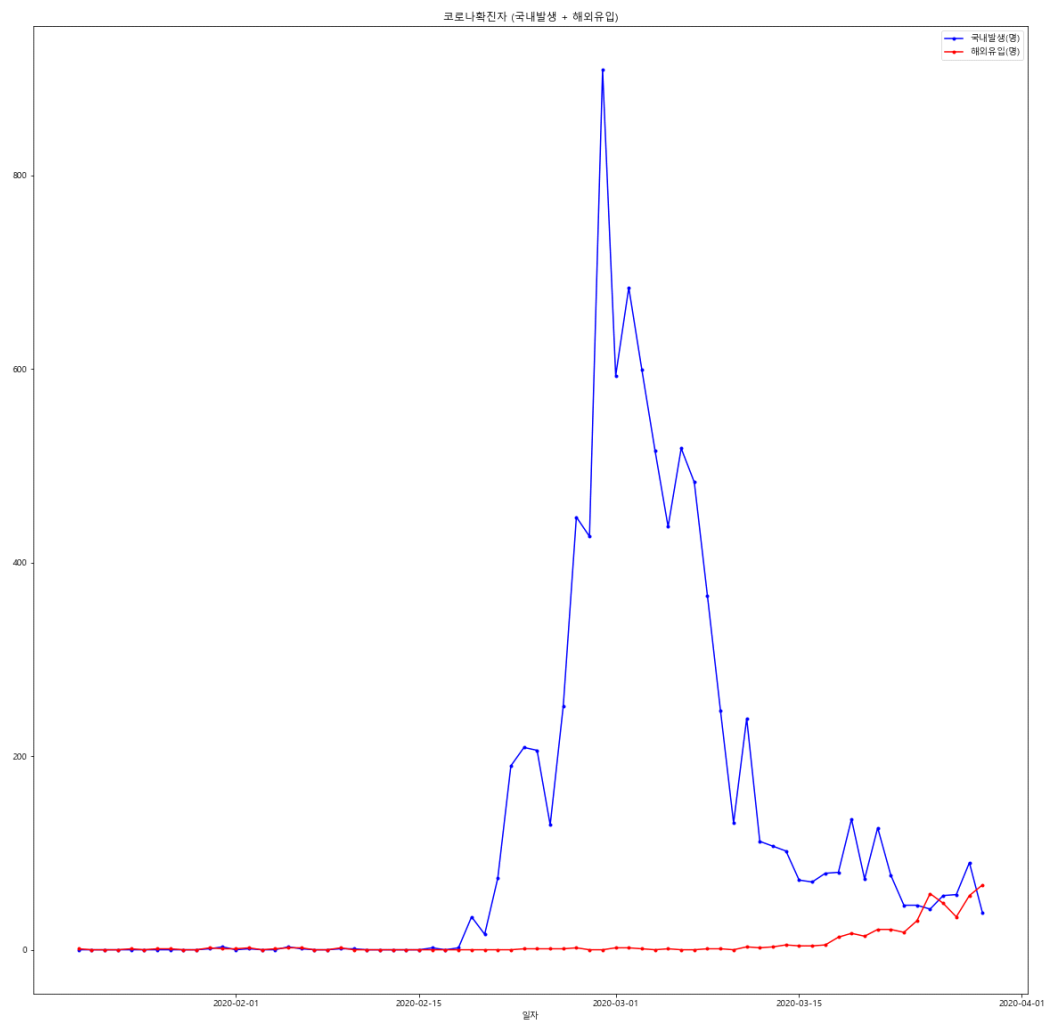
```

```
In [276]: import pandas as pd
df = pd.read_csv('코로나확진자_발생별(국내발생+해외유입).csv', encoding='euc-kr', thousands = ',')
df = df[1:]
df['국내발생(명)'] = df['국내발생(명)'].str.replace(',', '')
df['해외유입(명)'] = df['해외유입(명)'].str.replace(',', '')
df['사망(명)'] = df['사망(명)'].str.replace(',', '')
df['국내발생(명)'] = df['국내발생(명)'].str.replace('-', '0').astype('int64')
df['해외유입(명)'] = df['해외유입(명)'].str.replace('-', '0').astype('int64')
df['사망(명)'] = df['사망(명)'].str.replace('-', '0').astype('int64')
df['일자'] = df['일자'].astype('datetime64[ns]')
df
```

Out [276]:

	일자	계(명)	국내발생(명)	해외유입(명)	사망(명)
1	2020-01-20	1	0	1	0
2	2020-01-21	0	0	0	0
3	2020-01-22	0	0	0	0
4	2020-01-23	0	0	0	0
5	2020-01-24	1	0	1	0
...
1025	2022-11-09	62468	62426	42	59
1026	2022-11-10	55357	55307	50	52
1027	2022-11-11	54519	54470	49	40
1028	2022-11-12	54328	54270	58	46
1029	2022-11-13	48465	48412	53	48

1029 rows × 5 columns



•

•

•

-
-
- 해석

코로나 19 라는 단어는 기존에 있었던 ‘코로나’라는 질병에 바이러스 특성 상 기존 병원체와 달리 변형되어 2019 년부터 중국에서 발생하기 시작했다는 점에서 코로나 19 라는 이름이 붙여지게 되었다. 허나, 우리가 필요한 것은 국내 코로나 감염 증감 변화율이므로 국내에 처음 퍼지기 시작한 2020 년을 기준으로 그 전후의 코로나 확진자 변화율과 비교하여 코로나가 국내에 언제부터 만연하기 시작했는지, 그리고 그의 영향은 무엇인지를 알 수 있는, 다른 데이터들의 기초가 된다.

이미 우리는 코로나 19 가 2020 년부터 우리 생활에 침투하기 시작했다는 것을 알지만, 객관적인 지표를 봤을 때 역시 2020 년 2 월부터 서서히 증가하다가 2020 년 3 월에 그 수가 급진적으로 증가함을 알 수 있다. 그 이후에도 계속해서 그 확진자 수는 늘어나고 있음을 통해 코로나가 국내에 2020 년 초에 유입되어 계속하여 퍼지는 것을 알 수 있다.

또한, 이 지표는 확진자 수나 그 비율에 집중하지 않아도 된다. 단지 코로나 19 가 국내에 어느 년도부터 발생하고 만연한다는 ‘사실’자체를 증명하려고 하기 때문이다

-



```

import pandas as pd
df = pd.read_csv('최근 5년(2017년~2021년) 10세 단위별 우울증 환자수.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0], inplace=True)
df = df.drop(df.index[0])
df = df.drop(df.index[-1])
df = df.drop(df.index[-1])
#df.set_index(df['일자'], inplace=True)
df['일자'] = df.index
df['일자'] = df['일자'].str.replace('년', '-01-01')
df['일자'] = df['일자'].astype('datetime64[ns]')
df['전체'] = df['전체'].astype('int64')
df['10대 미만'] = df['10대 미만'].astype('int64')
df['10대'] = df['10대'].astype('int64')
df['20대'] = df['20대'].astype('int64')
df['30대'] = df['30대'].astype('int64')
df['40대'] = df['40대'].astype('int64')
df['50대'] = df['50대'].astype('int64')
df['60대'] = df['60대'].astype('int64')
df['70대'] = df['70대'].astype('int64')
df['80대 이상'] = df['80대 이상'].astype('int64')
df.iloc[1:]
df['청년층'] = df['20대'] + df['30대']
df['노년층'] = df['60대'] + df['70대'] + df['80대 이상']
df = df.iloc[1:]
df['10대 증감률'] = 0.0
df['청년층 증감률'] = 0.0
df.loc['2020년', '10대 증감률'] = ((df.loc['2020년', '10대'] - df.loc['2019년', '10대']) / df.loc['2019년', '10대']) * 100
df.loc['2021년', '10대 증감률'] = ((df.loc['2021년', '10대'] - df.loc['2020년', '10대']) / df.loc['2020년', '10대']) * 100
df.loc['2020년', '청년층 증감률'] = ((df.loc['2020년', '청년층'] - df.loc['2019년', '청년층']) / df.loc['2019년', '청년층'])
df.loc['2021년', '청년층 증감률'] = ((df.loc['2021년', '청년층'] - df.loc['2020년', '청년층']) / df.loc['2020년', '청년층'])
df2 = df.iloc[2:]
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('10대 우울증 환자 증감률 (%)')
plt.plot(df2['일자'], df2['10대 증감률'], 'b.-', label='10층 우울증 환자 증감률 (%)')
plt.legend()
plt.savefig('10대 우울증 환자 증감률 (%).png')
plt.show()

```



```

import pandas as pd
df = pd.read_csv('최근 5년(2017년~2021년) 10세 단위별 우울증 환자수.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0],inplace=True)
df = df.drop(df.index[0])
df = df.drop(df.index[-1])
df = df.drop(df.index[-1])
#df.set_index(df['일자'], inplace=True)
df['일자'] = df.index
df['일자'] = df['일자'].str.replace('년', '-01-01')
df['일자'] = df['일자'].astype('datetime64[ns]')
df['전체'] = df['전체'].astype('int64')
df['10대 미만'] = df['10대 미만'].astype('int64')
df['10대'] = df['10대'].astype('int64')
df['20대'] = df['20대'].astype('int64')
df['30대'] = df['30대'].astype('int64')
df['40대'] = df['40대'].astype('int64')
df['50대'] = df['50대'].astype('int64')
df['60대'] = df['60대'].astype('int64')
df['70대'] = df['70대'].astype('int64')
df['80대 이상'] = df['80대 이상'].astype('int64')
df.iloc[1:]
df['청년층'] = df['20대'] + df['30대']
df['노년층'] = df['60대'] + df['70대'] + df['80대 이상']
df = df.iloc[1:]
df['10대 증감률'] = 0.0
df['청년층 증감률'] = 0.0
df.loc['2020년', '10대 증감률'] = ((df.loc['2020년', '10대'] - df.loc['2019년', '10대']) / df.loc['2019년', '10대']) * 100
df.loc['2021년', '10대 증감률'] = ((df.loc['2021년', '10대'] - df.loc['2020년', '10대']) / df.loc['2020년', '10대']) * 100
df.loc['2020년', '청년층 증감률'] = ((df.loc['2020년', '청년층'] - df.loc['2019년', '청년층']) / df.loc['2019년', '청년층']) * 100
df.loc['2021년', '청년층 증감률'] = ((df.loc['2021년', '청년층'] - df.loc['2020년', '청년층']) / df.loc['2020년', '청년층']) * 100
df2 = df.iloc[2:]
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('청년층 우울증 환자 증감률 (%)')
plt.plot(df2['일자'], df2['청년층 증감률'], 'b.-', label='청년층 우울증 환자 증감률 (%)')
plt.legend()
plt.savefig('청년층 우울증 환자 증감률 (%).png')
plt.show()

```

```

In [266]: import pandas as pd
df = pd.read_csv('최근 5년(2017년~2021년) 10세 단위별 우울증 환자수.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0], inplace=True)
df = df.drop(df.index[0])
df = df.drop(df.index[-1])
df = df.drop(df.index[-1])
#df.set_index(df['일자'], inplace=True)
df['일자'] = df.index
df['일자'] = df['일자'].str.replace('년', '-01-01')
df['일자'] = df['일자'].astype('datetime64[ns]')
df['전체'] = df['전체'].astype('int64')
df['10대 미만'] = df['10대 미만'].astype('int64')
df['10대'] = df['10대'].astype('int64')
df['20대'] = df['20대'].astype('int64')
df['30대'] = df['30대'].astype('int64')
df['40대'] = df['40대'].astype('int64')
df['50대'] = df['50대'].astype('int64')
df['60대'] = df['60대'].astype('int64')
df['70대'] = df['70대'].astype('int64')
df['80대 이상'] = df['80대 이상'].astype('int64')
df.iloc[1:]
df['청년층'] = df['20대'] + df['30대']
df['노년층'] = df['60대'] + df['70대'] + df['80대 이상']
df = df.iloc[1:]
df

```

Out[266]:

	전체	10대 미만	10대	20대	30대	40대	50대	60대	70대	80대 이상	일자	청년층	노년층
2018년	764681	1243	43029	100598	94371	106764	130723	136019	123639	58324	2018-01-01	194969	317982
2019년	811862	1493	48099	122039	106756	113422	131282	137497	122168	61098	2019-01-01	228795	320763
2020년	848430	1350	47774	148136	118277	116660	127385	139883	119221	63670	2020-01-01	266413	322774
2021년	933481	1940	57587	177166	140270	130718	130059	148039	118611	66570	2021-01-01	317436	333220

```

In [267]: import pandas as pd
df = pd.read_csv('최근 5년(2017년~2021년) 10세 단위별 우울증 환자수.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0], inplace=True)
df = df.drop(df.index[0])
df = df.drop(df.index[-1])
df = df.drop(df.index[-1])
#df.set_index(df['일자'], inplace=True)
df['일자'] = df.index
df['일자'] = df['일자'].str.replace('년', '-01-01')
df['일자'] = df['일자'].astype('datetime64[ns]')
df['전체'] = df['전체'].astype('int64')
df['10대 미만'] = df['10대 미만'].astype('int64')
df['10대'] = df['10대'].astype('int64')
df['20대'] = df['20대'].astype('int64')
df['30대'] = df['30대'].astype('int64')
df['40대'] = df['40대'].astype('int64')
df['50대'] = df['50대'].astype('int64')
df['60대'] = df['60대'].astype('int64')
df['70대'] = df['70대'].astype('int64')
df['80대 이상'] = df['80대 이상'].astype('int64')
df.iloc[1:]
df['청년층'] = df['20대'] + df['30대']
df['노년층'] = df['60대'] + df['70대'] + df['80대 이상']
df = df.iloc[1:]
df['10대 증감률'] = 0.0
df['청년층 증감률'] = 0.0
df.loc['2020년', '10대 증감률'] = ((df.loc['2020년', '10대'] - df.loc['2019년', '10대']) / df.loc['2019년', '10대']) * 100
df.loc['2021년', '10대 증감률'] = ((df.loc['2021년', '10대'] - df.loc['2020년', '10대']) / df.loc['2020년', '10대']) * 100
df.loc['2020년', '청년층 증감률'] = ((df.loc['2020년', '청년층'] - df.loc['2019년', '청년층']) / df.loc['2019년', '청년층'])
df.loc['2021년', '청년층 증감률'] = ((df.loc['2021년', '청년층'] - df.loc['2020년', '청년층']) / df.loc['2020년', '청년층'])
df2 = df.iloc[2:]
df2

```

Out [267]:

	전체	10대 미만	10대	20대	30대	40대	50대	60대	70대	80대 이상	일자	청년층	노년층	10대 증감률	청년층 증감률
2020년	848430	1350	47774	148136	118277	116660	127385	139883	119221	63670	2020-01-01	266413	322774	-0.675690	16.441793
2021년	933481	1940	57587	177166	140270	130718	130059	148039	118611	66570	2021-01-01	317436	333220	20.540461	19.151843

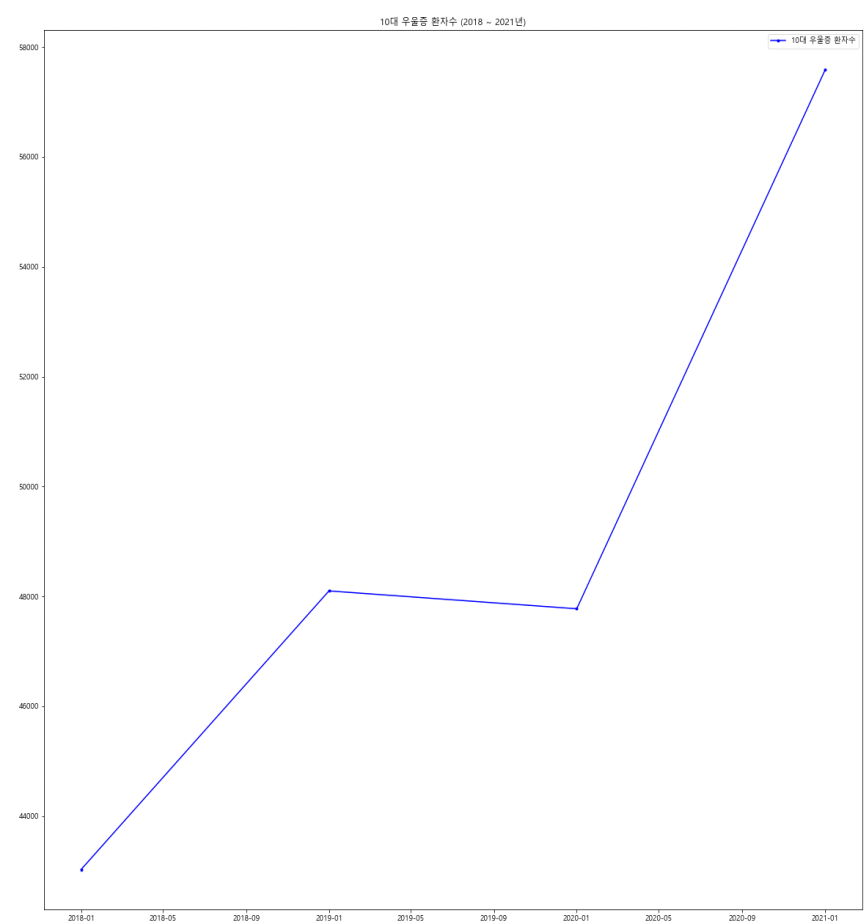
- 위의 쿼리가 필요한 이유

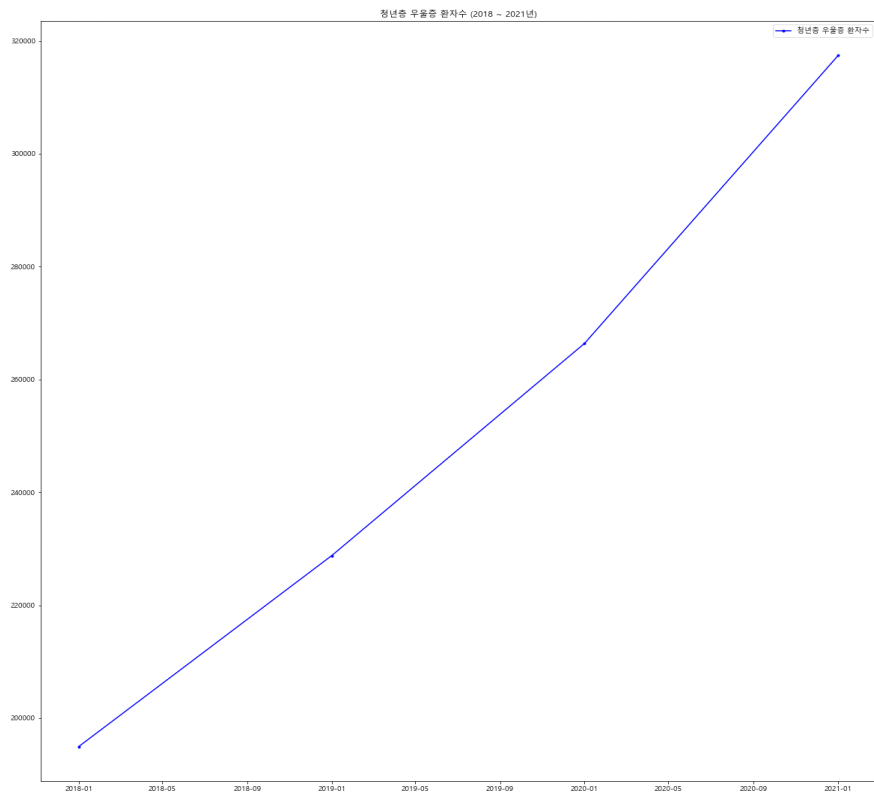
코로나가 우리나라에 녹아들기 시작한 2020 년 3 월(봄 즈음)을 기점으로 그 전과 그
후의 우울증 환자 증감

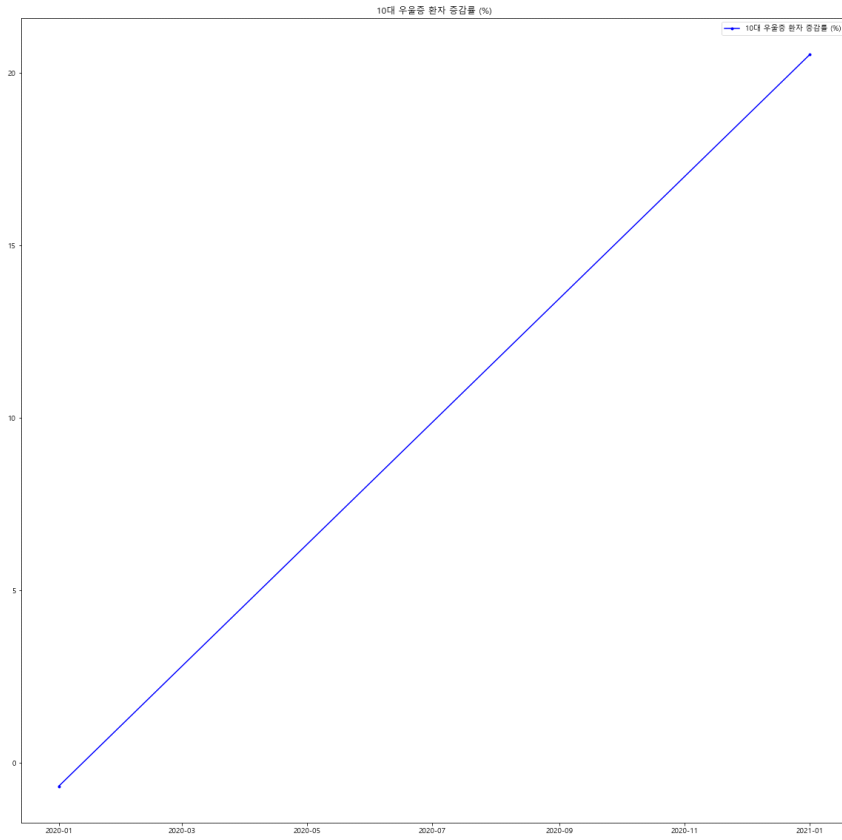
정도를 알고 싶기 때문에, 쉽게는 그래프의 기울기를 기준으로 비교할 수
있지만(경사도), 수치로 보기 위해

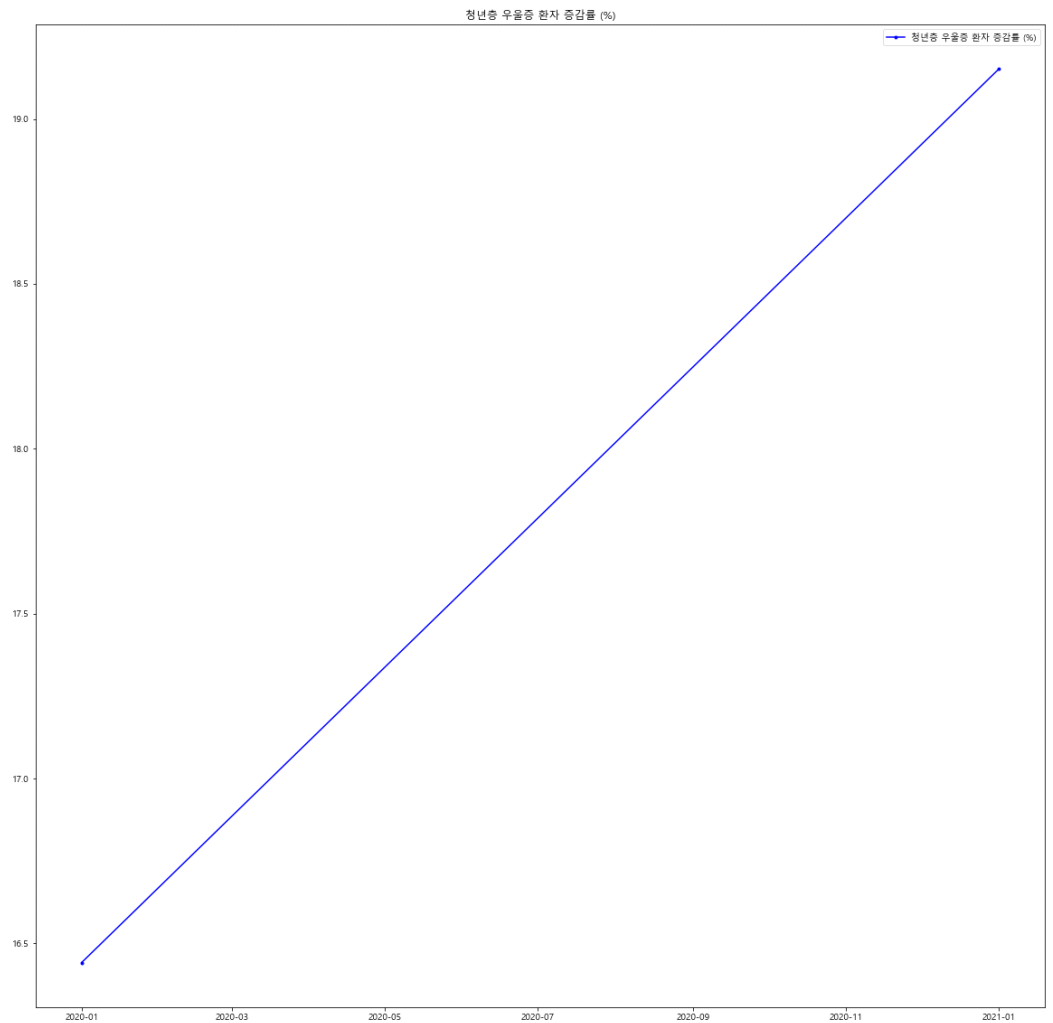
전년도와 비교하는(차이) 과정을 한 번 더 거치는 과정이 필요하다. 또한, 코로나 블루의
여파를 더 많이 체감한

다고 알려져 있는(나이가 비교적 어린 연령층인)10 대와 청년층을 기준으로 했다.









- 해석

코로나가 국내에서 시작된 2020 년 3 월 전과 그 이후를 보면 차이를 확연히 보인다.

10 대들은 2020 년

3 월을 기점으로 2021 년 1 월까지 우울증 환자의 수가 급증하고, 기율기의 경사만 봐도
보다 가파른 것을

알 수 있다. 기율기가 가파를수록, 증감률이 큰 것으로 해석이 되는데, 쉽게 보기 위해
따로 구한 증감률

그래프를 봐도 2020 년 3 월을 시작으로 전년도에 비해 그 환자의 수의 비율이 늘어난
것을 볼 수 있다.

이에 따라, 온전히 코로나 블루의 결과로 보기에는 성급한 결론일 수 있지만, 코로나가
국내에 유입된 후

우울증 환자가 급증했다는 점에서 코로나 블루의 영향이 있었음에 타당성을 부여할 수
있다.

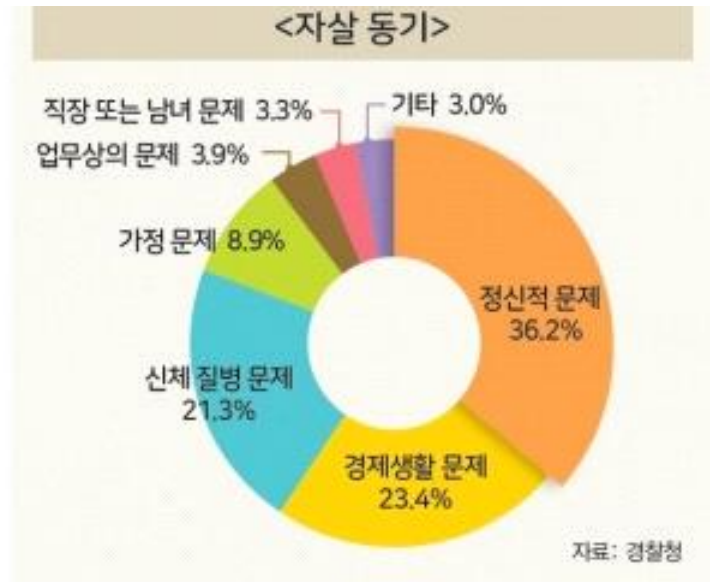
다음으로, 20-30 대의 청년층의 우울증 환자 수의 그래프를 봐도, 코로나 19 이후
전년도에 비해 그 증감

률이 늘어난 것을 발견할 수 있다. 그렇지만, 10 대의 그래프와 달리 급증하는 식의
모습이 보이지 않기 때문에

순수 코로나 블루의 여파라고 보기엔 다소 어렵지 않을까 하는 결론이다.

3. 코로나 블루로 전국의 자살률의 증감 변화

- 배경 지식



우리가 사용할 결론 데이터는, 가장 많은 비율을 차지하는 정신적 문제로 인한 자살과 경제 생활 문제로

인한 자살, 이렇게 두 개로 새로운 데이터를 분석할 것이다. 정신적 문제에는 대표적으로 우울증이 있으며,

흔히 전 연령대에 걸쳐 나타나는 현상이나, 경제 생활 문제는 경제 생활을 하는 연령대인 청년층, 중&장년층에게 나타나는 현상이다.

- 필요한 쿼리

2018 년부터 2021 년까지의 10 대 청소년들의 자살발생 수와 20 대와 30 대를 청년층으로 통합하여

그 자살 발생 수, 40~59 세를 중장년층으로 통합하여 그 자살 발생 수에 대한 쿼리를 구한다. 또한 국내의

전체적인 자살 발생 수에 대한 쿼리를 구하여 전연령대를 통합하여 분석할 수 있도록 한다.

```

import pandas as pd
df5 = pd.read_csv('자살률_성_연령별_20221115163012.csv', encoding='euc-kr', thousands = ',')
df5 = df5.loc[0:11,:]
df5 = df5.iloc[:, 1:]
for i in range(1998, 2022):
    df5.drop(str(i) + '.1', axis=1, inplace=True)
df5 = df5.drop(df5.index[0])
df5 = df5.transpose()
df5.rename(columns=df5.iloc[0], inplace=True)
df5['연도'] = df5.index
df5 = df5.drop(df5.index[0])
df = df5.iloc[20:]
df['계'] = df['계'].astype(int)
df['0-9세'] = df['0-9세'].astype(int)
df['10-19세'] = df['10-19세'].astype(int)
df['20-29세'] = df['20-29세'].astype(int)
df['30-39세'] = df['30-39세'].astype(int)
df['40-49세'] = df['40-49세'].astype(int)
df['50-59세'] = df['50-59세'].astype(int)
df['60-69세'] = df['60-69세'].astype(int)
df['70-79세'] = df['70-79세'].astype(int)
df['80세이상'] = df['80세이상'].astype(int)
df['연령미상'] = df['연령미상'].astype(int)
df['청년층'] = df['20-29세'] + df['30-39세']
df['중장년층'] = df['40-49세'] + df['50-59세']
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('10대 자살자 수')
plt.plot(df['연도'], df['10-19세'], 'b.-', label='10-19세')
plt.legend()
plt.savefig('10대 자살자 수.png')
plt.show()

```

```

import pandas as pd
df5 = pd.read_csv('자살률_성_연령별_20221115163012.csv', encoding='euc-kr', thousands = ',')
df5 = df5.loc[0:11,: ]
df5 = df5.iloc[:, 1:]
for i in range(1998, 2022):
    df5.drop(str(i) + '.1', axis=1, inplace=True)
df5 = df5.drop(df5.index[0])
df5 = df5.transpose()
df5.rename(columns=df5.iloc[0],inplace=True)
df5['연도'] = df5.index
df5 = df5.drop(df5.index[0])
df = df5.iloc[20:]
df['계'] = df['계'].astype(int)
df['0-9세'] = df['0-9세'].astype(int)
df['10-19세'] = df['10-19세'].astype(int)
df['20-29세'] = df['20-29세'].astype(int)
df['30-39세'] = df['30-39세'].astype(int)
df['40-49세'] = df['40-49세'].astype(int)
df['50-59세'] = df['50-59세'].astype(int)
df['60-69세'] = df['60-69세'].astype(int)
df['70-79세'] = df['70-79세'].astype(int)
df['80세이상'] = df['80세이상'].astype(int)
df['연령미상'] = df['연령미상'].astype(int)
df['청년층'] = df['20-29세'] + df['30-39세']
df['중장년층'] = df['40-49세'] + df['50-59세']
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('청년층 자살자 수')
plt.plot(df['연도'], df['청년층'], 'b.-', label='청년층 자살자 수')
plt.legend()
plt.savefig('청년층 자살자 수.png')
plt.show()

```

```

import pandas as pd
df5 = pd.read_csv('자살률_성_연령별_20221115163012.csv', encoding='euc-kr', thousands = ',')
df5 = df5.loc[0:11,:]
df5 = df5.iloc[:, 1:]
for i in range(1998, 2022):
    df5.drop(str(i) + '.1', axis=1, inplace=True)
df5 = df5.drop(df5.index[0])
df5 = df5.transpose()
df5.rename(columns=df5.iloc[0],inplace=True)
df5['연도'] = df5.index
df5 = df5.drop(df5.index[0])
df = df5.iloc[20:]
df['계'] = df['계'].astype(int)
df['0-9세'] = df['0-9세'].astype(int)
df['10-19세'] = df['10-19세'].astype(int)
df['20-29세'] = df['20-29세'].astype(int)
df['30-39세'] = df['30-39세'].astype(int)
df['40-49세'] = df['40-49세'].astype(int)
df['50-59세'] = df['50-59세'].astype(int)
df['60-69세'] = df['60-69세'].astype(int)
df['70-79세'] = df['70-79세'].astype(int)
df['80세 이상'] = df['80세 이상'].astype(int)
df['연령미상'] = df['연령미상'].astype(int)
df['청년층'] = df['20-29세'] + df['30-39세']
df['중장년층'] = df['40-49세'] + df['50-59세']
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('중장년층 자살자 수')
plt.plot(df['연도'], df['중장년층'], 'b.-', label='중장년층 자살자 수')
plt.legend()
plt.savefig('중장년층 자살자 수.png')
plt.show()

```

```

In [260]: import pandas as pd
df5 = pd.read_csv('자살률_성_연령별_20221115163012.csv', encoding='euc-kr', thousands = ',')
df5 = df5.loc[0:11,:]
df5 = df5.iloc[:, 1:]
for i in range(1998, 2022):
    df5.drop(str(i) + '.1', axis=1, inplace=True)
df5 = df5.drop(df5.index[0])
df5 = df5.transpose()
df5.rename(columns=df5.iloc[0],inplace=True)
df5['연도'] = df5.index
df5 = df5.drop(df5.index[0])
df = df5.iloc[20:]
df['계'] = df['계'].astype(int)
df['0-9세'] = df['0-9세'].astype(int)
df['10-19세'] = df['10-19세'].astype(int)
df['20-29세'] = df['20-29세'].astype(int)
df['30-39세'] = df['30-39세'].astype(int)
df['40-49세'] = df['40-49세'].astype(int)
df['50-59세'] = df['50-59세'].astype(int)
df['60-69세'] = df['60-69세'].astype(int)
df['70-79세'] = df['70-79세'].astype(int)
df['80세이상'] = df['80세이상'].astype(int)
df['연령미상'] = df['연령미상'].astype(int)
df['청년층'] = df['20-29세'] + df['30-39세']
df['중장년층'] = df['40-49세'] + df['50-59세']
df

```

Out [260]:

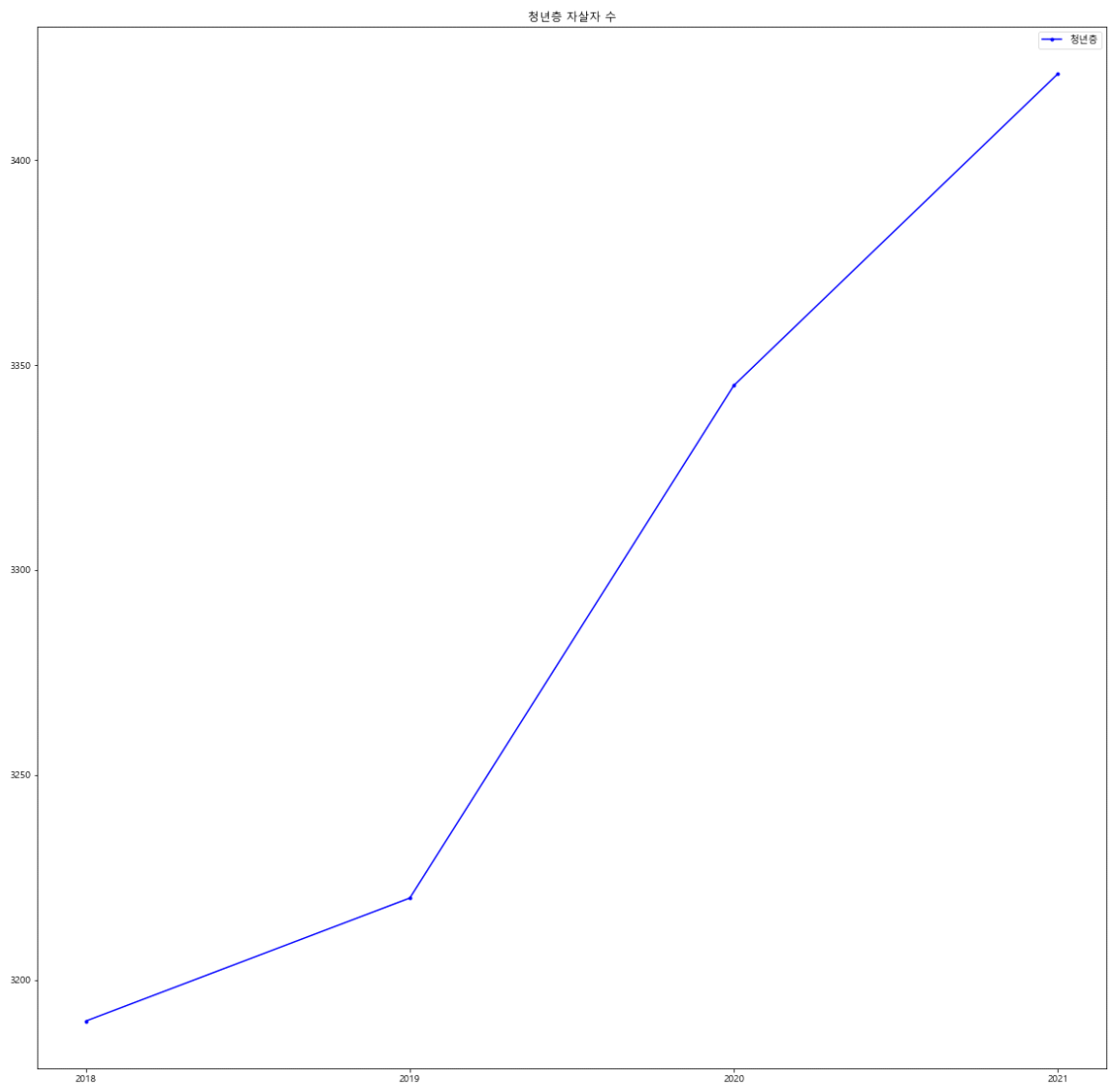
	계	0-9세	10-19세	20-29세	30-39세	40-49세	50-59세	60-69세	70-79세	80세이상	연령미상	연도	청년층	중장년층
2018	13670	1	300	1192	1998	2676	2812	1880	1664	1147	0	2018	3190	5488
2019	13799	2	298	1306	1914	2588	2837	2035	1620	1193	6	2019	3220	5425
2020	13195	2	315	1471	1874	2405	2606	1937	1398	1187	0	2020	3345	5011
2021	13352	1	338	1579	1842	2298	2569	1951	1535	1239	0	2021	3421	4867

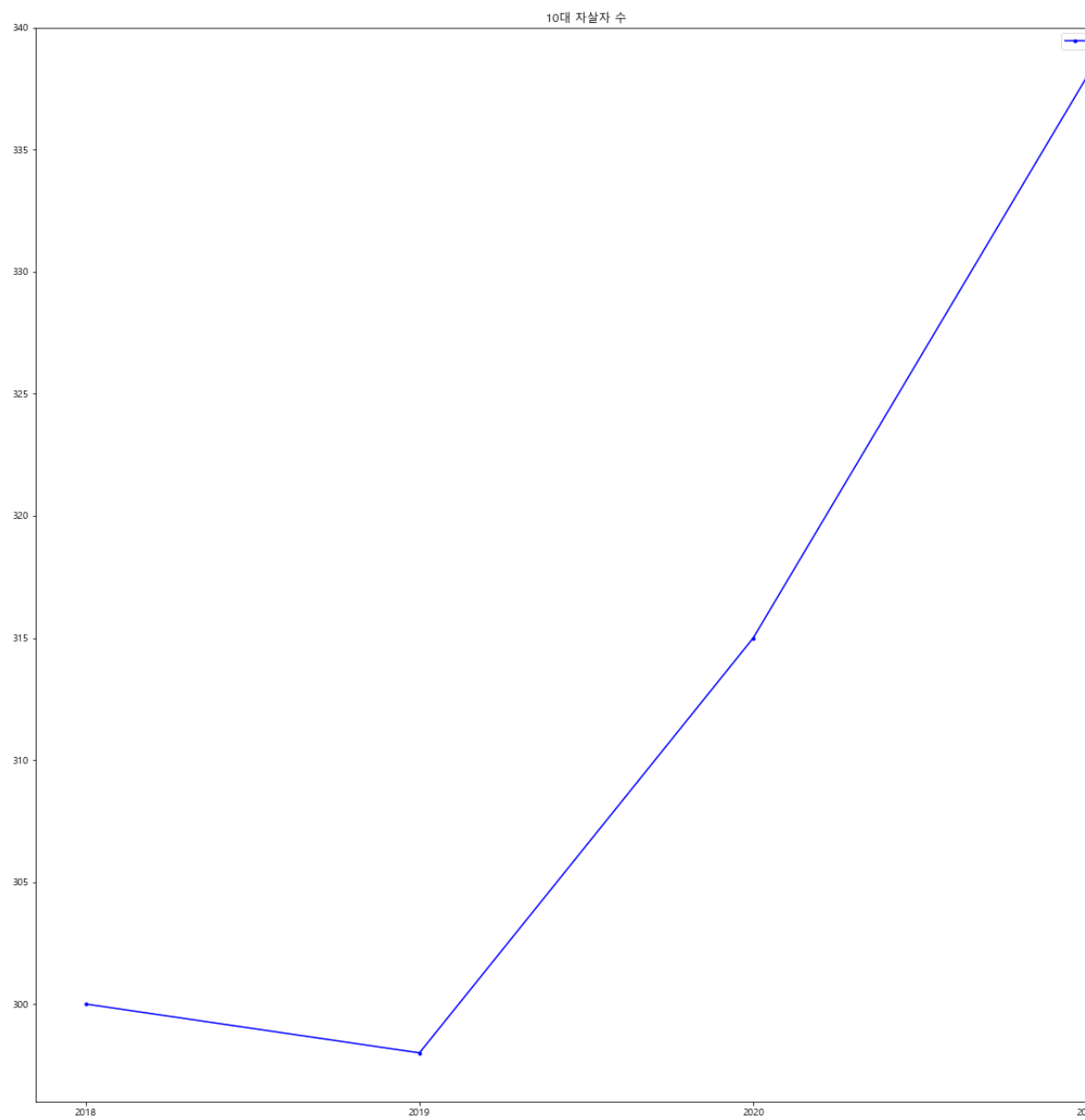
- 위의 쿼리가 필요한 이유

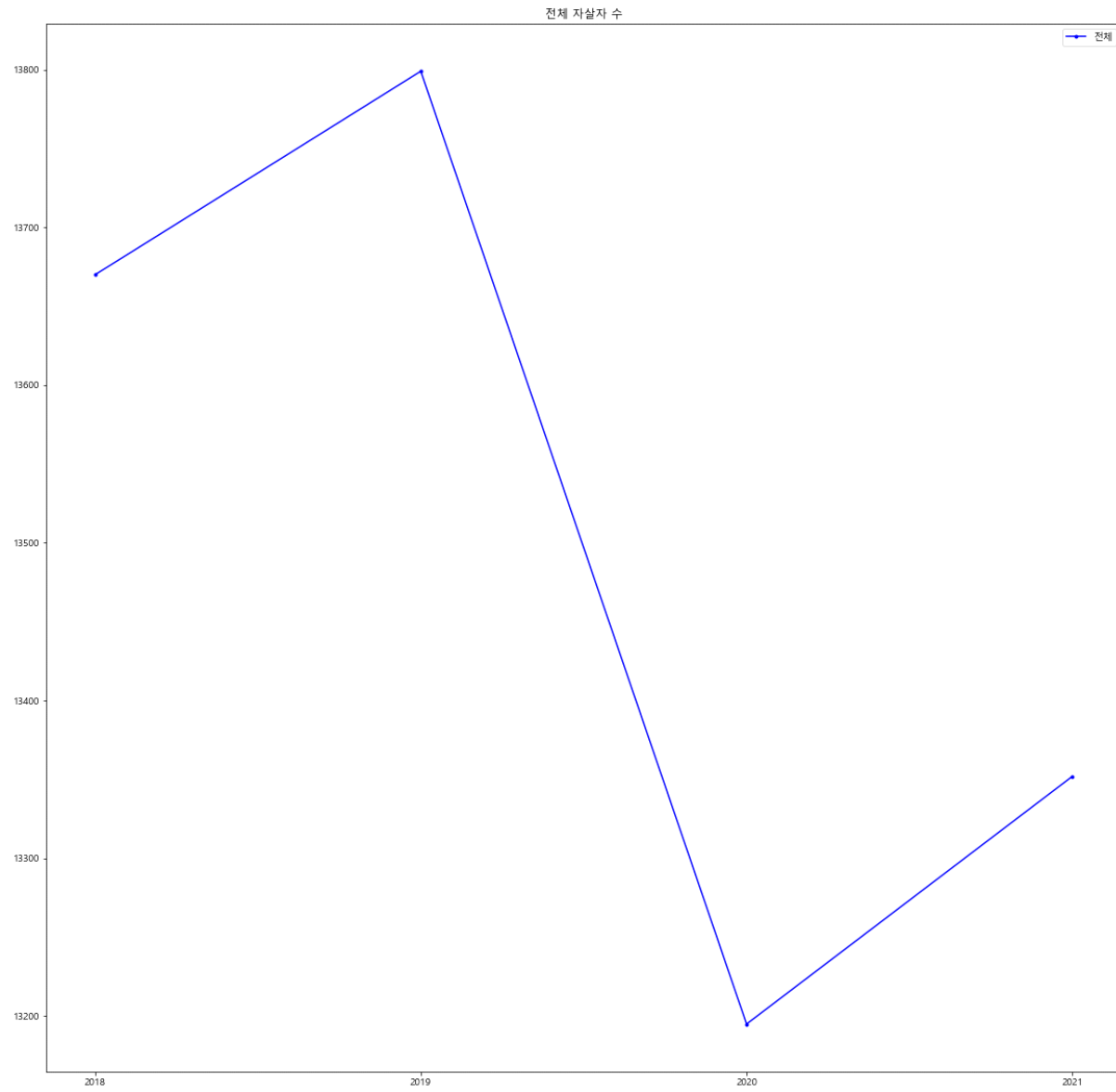
앞에서 구한 우울증 환자의 수와 자살의 가장 큰 원인인 정신적 문제를 연결하여

그것이 자살률에까지 영향을 끼치는지 확인해보기 위함이다. 우리의 예상과 다르다면,

그 원인은 무엇인지도 파악하기 위함이라고도 할 수 있다.







- 해석

10 대를 포함한 비교적 나이가 적은 연령층은 자살률이 코로나 19 이후로 늘었으나, 코로나 19 첫 해인 2020 년 국내 전체적인 자살률은 전년과 견줘 감소한 것으로 보아 추가적인 원인 규명이 필요하다 판단하여 다른 자료들을 찾아보았다. 이에 따라, 앞서 언급한 코로나 19 에 따른 블루 현상, 즉 우울증현상에 대한 자료들도 추가로 모을 수 있었는데, 코로나 19 초기 발생 년도인 2020 년은 우울현상이 전년도에 비해 더 높게 나타나기도 하지만, 코로나 19 라는 세계적인, 즉 공통적인 팬데믹 기간동안 초기에는 대규모 재정 정책 덕에, 그리고 처음 겪는 대규모 현상인 만큼 위기 상황에 대응하느라 자살을 행할 여력이 없다는 점, 그리고 국민적 단합이 발휘돼 자살 사망이 감소한다는 점을 고려해보면 이해 가능한 데이터이다. 하지만, 이 시기가 지나고 향 후 2~3 년 간 자살 사망률이 증가할 것을 우려하는 전문가들의 견해가 있었다. 이러한 자료들을 보니, 우울증 환자들이 점점 늘어나는 것도 설명이 가능했다. 또한 조사한대로 향후 2~3 년간 우울증 비율과 자살률이 늘어난다면, 코로나 19 의 여파의 영향으로도 어느정도 타당성을 부여할 수 있지 않을까? 물론 자살이라는 것은 복합적인 원인으로 발생하는 것이고, 긍정적이지 않은 사고인만큼 전문가들이 걱정하는 일이 일어나지 않기를 바라야 하겠지만 말이다.

4. 코로나 19 에 따른 실업률 및 취업률

- 필요한 쿼리

위의 바탕 지식의 그래프에서 설명했듯이, 청년과 중장년층의 자살 원인으로는 경제생활이 있었다. 경제 활동과 직결된 실업률 및 취업률을 통해 그 자살률과 우울증의 타당성을 부여할 수 있다는 점을 고려하여, 20 대 취업률과 전체적인 실직률 쿼리를 구해야 한다.

최근 5 년 간 20 대 취업률 쿼리, 최근 5 년간 전체 연령의 실업률 쿼리를 필요로 한다.

```
import pandas as pd
df = pd.read_csv('연령별_경제활동인구_총괄_20221115164116.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0], inplace=True)
df = df.drop(df.index[0])
df['일자'] = df.index
df['일자'] = df['일자'].astype('datetime64[ns]')
a = df.loc[(df['연령계층별'] == '취업자 (천명)')].index
a
df = df.loc[a, :]
cols = ['15세 이상 전체', '15 - 19세', '20 - 29세', '30 - 39세', '40 - 49세', '50 - 59세', '60세이상', '15 - 64세', '15 - 24세']
for col in cols:
    df[col] = df[col].astype('int64')
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('20대 취업자 수 (천명)')
plt.plot(df['일자'], df['20 - 29세'], 'b.-', label='취업자 수 (천명)')
plt.xlabel('일자')
plt.legend()
plt.savefig('20대 취업자 수 (천명).png')
plt.show()
```

```
import pandas as pd
df = pd.read_csv('연령별_경제활동인구_총괄_20221115164116.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0], inplace=True)
df = df.drop(df.index[0])
df['일자'] = df.index
df['일자'] = df['일자'].astype('datetime64[ns]')
a = df.loc[(df['연령계층별'] == '실업자 (천명)')].index
df = df.loc[a, :]
cols = ['15세 이상 전체', '15 - 19세', '20 - 29세', '30 - 39세', '40 - 49세', '50 - 59세', '60세이상', '15 - 64세', '15 - 24세']
for col in cols:
    df[col] = df[col].astype('int64')
df.dtypes
import matplotlib.pyplot as plt
import pandas as pd

plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('실업자 수 (천명)')
plt.plot(df['일자'], df['15세 이상 전체'], 'b.-', label='실업자 수 (천명)')
plt.legend()
plt.savefig('실업자 수.png')
plt.show()
```

```

In [241]: import pandas as pd
df = pd.read_csv('연령별_경제활동인구_총괄_20221115164116.csv', encoding='euc-kr', thousands = ',')
df = df.transpose()
df.rename(columns=df.iloc[0],inplace=True)
df = df.drop(df.index[0])
df['일자'] = df.index
df['일자'] = df['일자'].astype('datetime64[ns]')
a = df.loc[(df['연령계층별'] == '실업자 (천명)')].index
df = df.loc[a, :]
cols = ['15세 이상 전체', '15 - 19세', '20 - 29세', '30 - 39세', '40 - 49세', '50 - 59세', '60세이상']
for col in cols:
    df[col] = df[col].astype('int64')
df

```

Out [241]:

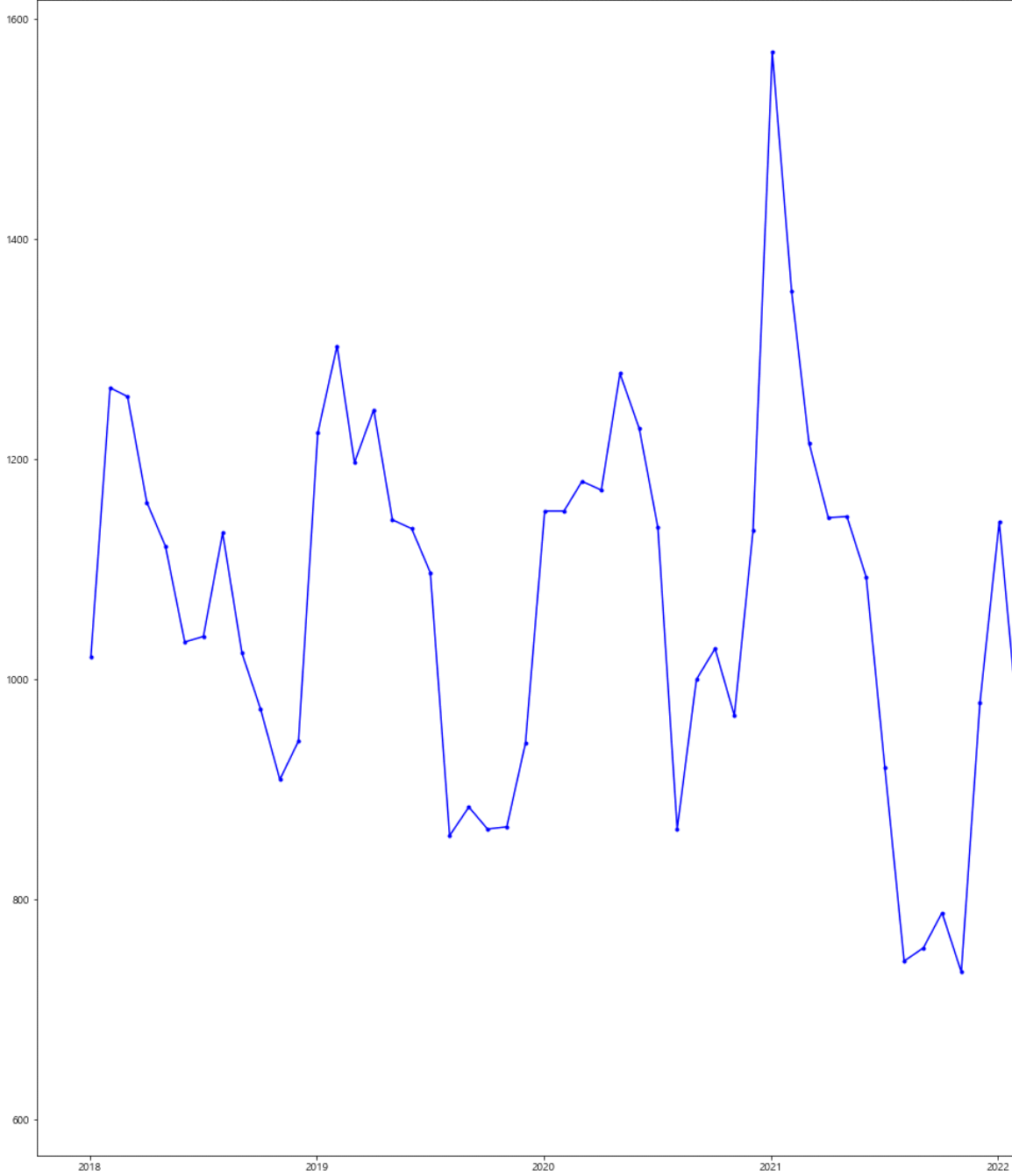
	연령계층별	15세 이상 전체	15 - 19세	20 - 29세	30 - 39세	40 - 49세	50 - 59세	60세이상	15 - 64세	15 - 24세	15 - 29세
2013.01.3	실업자 (천명)	851	33	275	197	133	121	92	791	153	306
2013.02.3	실업자 (천명)	991	29	337	208	145	142	132	904	171	330
2013.03.3	실업자 (천명)	884	19	320	191	159	124	72	854	172	301
2013.04.3	실업자 (천명)	827	28	306	185	151	111	45	809	167	282
2013.05.3	실업자 (천명)	799	23	278	184	145	111	58	780	137	265
...
2022.06.3	실업자 (천명)	888	7	293	168	148	138	134	824	122	306
2022.07.3	실업자 (천명)	836	20	275	167	137	118	119	778	114	291
2022.08.3	실업자 (천명)	615	8	221	139	100	85	62	587	77	216
2022.09.3	실업자 (천명)	704	6	252	141	113	102	90	670	91	253
2022.10.3	실업자 (천명)	692	4	231	148	116	100	92	655	79	243

118 rows × 12 columns

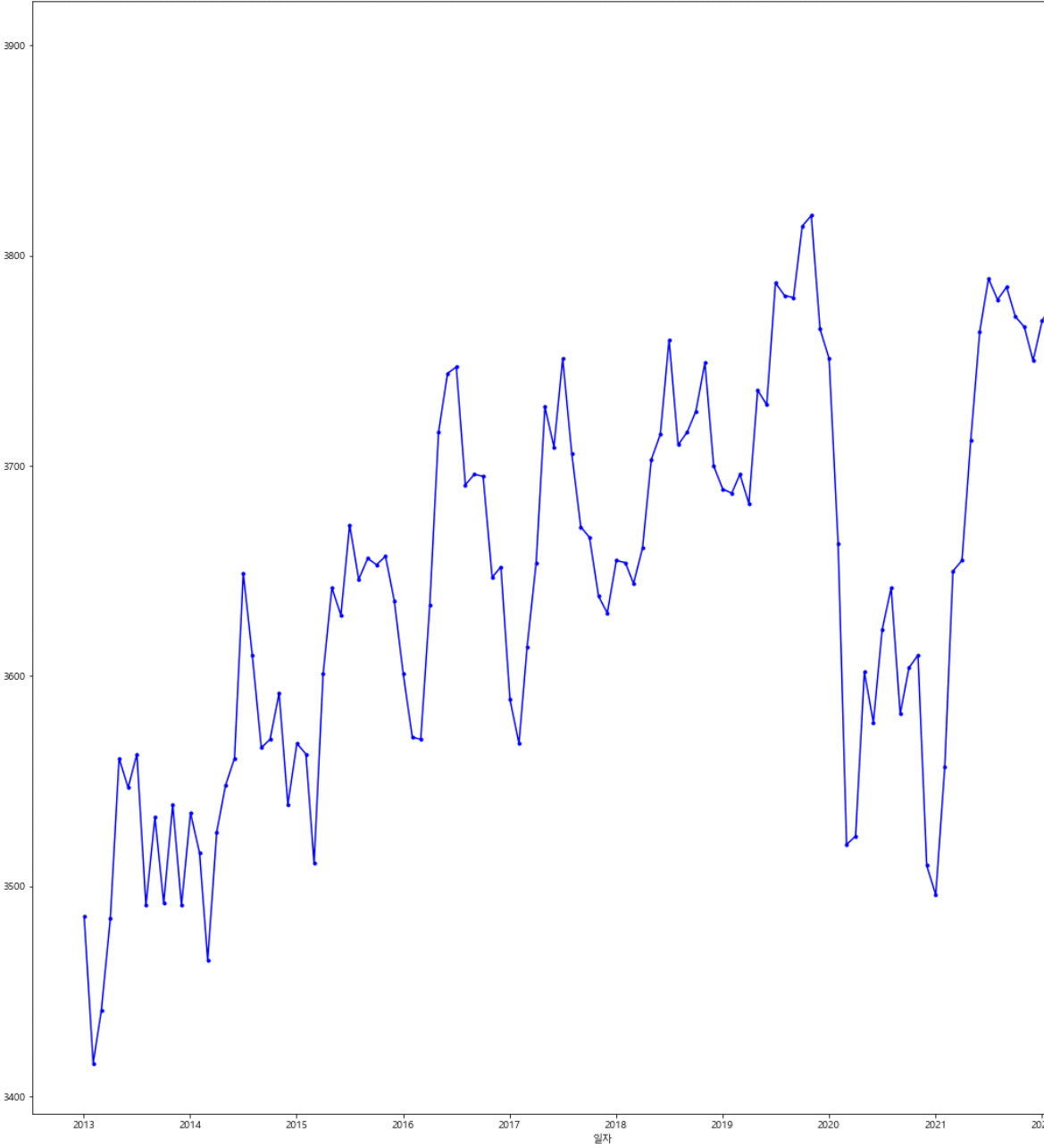
- 위의 쿼리가 필요한 이유

보통 취업은 20 대 때 처음으로 많이들 하여 사회 초년생이라는 단어가 20 대에 보통 쓰이는 것이다. 코로나 19 로 평균적으로 경제생활을 시작할 20 대의 취업률과 전체적인 실직률로 코로나 19 의 여파를 확인하고, 이를 자살률과 우울증과 연결시킬 수 있다고 생각하여 조사하였다.

실업자 수 (천명)



20대 취업자 수 (천명)



- 해석

사실 코로나 19 에 의해서 가장 많이 변화를 보인 건, 경제활동 측면이라고 할 수 있다. 코로나 19 는 생활패턴 자체를 변화시킨 장본인이 될 수도 있었기에, 이와 같은 결과를 낳는 건 어찌보면 당연한 일이라고 할 수 있다. 생계와 직결된 문제인만큼, 코로나 19 의 어려움이 사회에 전체적으로 영향을 미친 대부분은 이러한 경제적 측면이라고 봐도 무방하다.

- 실업자 수의 그래프를 보면, 코로나가 국내에 터지고 그 초기인 2020 년보다 오히려 2021 년 실업자 수가 최근 5 년 간 최고점을 찍은 것을 볼 수 있다. 이는 다른 그래프 선의 형상과 다르게 급하게 올라간 모양을 보아, 시기까지 고려해보았을 때 코로나 19 가 그 영향을 지대하게 미쳤다는 점을 알 수 있다. 그렇다면, 왜 2020 년보다 코로나가 우리 생활에 스며든 지 1 년 후인 2021 년에 실직자 수가 급증했을까?

여러 자료를 찾아보니, 2020 년의 실직률이 예상보다 높지 않은 이유는, 코로나 19 충격에 의한 상관관계일 가능성이 있다는 자료가 있었다. 다수가 코로나 19 는 일시적일 현상일 뿐이라 안심하게 생각한 탓도 있었고, 실제로 그 해 우리나라의 코로나 예방율이 다른 나라보다 우수하여 다른 나라와 비교하였을 때 그 코로나 감염자 수의 증가비율이 낮아지는 추세를 겪기도 하였으니, 기업 입장에서선 선불리 판단하지 않은 것이 아닐까. 하지만, 예상보다 길어지는 코로나 19 팬데믹 현상에 그에 따른 정부 정책이나, 자동화의 발전, 그리고 비대면에 적합한 서비스가 점점 多등장하면서 어느정도 코로나에 익숙해진 사회인 2021 년은 그 실직률이 갑자기 높아진 것으로 추정된다.

- 본래라면 실직률이 높아진 만큼, 새로운 인력이 필요하기에 고용률, 특히 20 대 고용률이 높아야 하지만, 이 경우는 인간 대체 하는 서비스가 증가하여 실직률이 증가한 탓이 더 크므로, 고용률 역시 부진했다. 고용률이 낮으니 장기 구직자가 늘어나고 이는 우리 사회의 청년들에게 암담한 현실을 보여주기도 하였으며, 이것이 우울증상으로 이어지는 현상이 적지 않았다.

위의 실직률 그래프와 다르게 청년 고용률의 그래프는 코로나 19 의 초반 해인 2020 년부터 확연한 차이를 보였다. 이는, 위기 시기 노동시장 참여에 대한 청년층의 취약성을 나타내는데, 청년층의 고용 비중이 높은 여가 및 여행 산업의 감소로 자연스레 그에 필요한 인력을 구하는 일 또한 필요성이 적어지게 되어 나타난 현상이라고도 볼 수 있다.

또한 기업은 이러한 경제위기 시 신규 고용을 축소하는 방식으로 대응할 가능성이 크다는 염려가 있었는데, 그 예상이 맞아떨어졌다고도 볼 수 있다. 비교적 비전문직이나 비정규직이 많은 비중을 차지하고 있는 청년층 중에서도 20 대층은(숙박업, 음식점업, 교육 서비스업, 여가 서비스업) 코로나 19 의 즉각적이고도 직접적인 위기에 노출되어 그 영향이 바로 보였다.

그래프를 잘 살펴보면, 2020 년 총 세 차례의 유행기가 있는데, 그 유행기가 시작될 때마다 고용률이 낮아지는 것을 확인할 수 있다. 1 차 유행기인 2020 년 3-4 월, 2 차 유행기인 그 해 8-9 월, 마지막으로 급진적인 확진자 수를 기록한 12 월부터 다음 해 2021 년 초의 그래프를 보면 알 수 있다. 실제로 그래프를 봐도, 2020 년 초와 2021 년 초에 급격히 감소하는 청년 고용률을 볼 수 있다. 또한, 1 차 유행과 3 차 유행에 비해 비교적 낮은 확진자 수를 기록한 2020 년 중후반에는 청년

고용률 역시 급격한 차이보다는 확진자 수의 추세와 비슷한 경향을 보이고 있음을 확인할 수 있다. 이로써, 청년 고용률 자체가 코로나 19 에 영향을 직격으로 받고 있음이 적나라하게 드러난다.

5. 코로나 19 로 인한 가정 폭력 사건율의 변화

코로나 19 로 집에 있는 시간이 늘어나면서, 과연 가정 폭력의 비율 변화는 어떤 양상을 보일까 하는 생각이 들었다. 집 안에 있는 시간이 많으니 증가하였을까, 아님 감소한 변화를 보여서 추가적인 원인이나 조사할 것들, 추정할 것들이 있을까 조사해보고 싶었다.

- 배경지식

가정 보호 사건: '가정 구성원'사이에서 일어난 가정폭력범죄이나, 형사처벌이 아닌 '가정폭력범죄의 처벌 등에 관한 특례법 상 보호처분을 하는 절차

일반 가정폭력범죄와 다른 점: 피해자의 고소가 있어야 공소를 제기할 수 있는
가정폭력범죄에서 피해자의

고소가 없거나 취소된 경우, 피해자의 의사에 반해 공소를 제기할 수 없는 가정
폭력범죄에서 피해자가 처벌을 희망하지 않는다는 명시적 표시가 있거나 처벌을
희망하는 의사표시가 철회된 경우

로, 즉, 같은 사건이지만 피해자의 의사에 따라 처리과정이 다르다는 것을 뜻한다.

- 필요한 쿼리

가정폭력도 엄연한 폭력 사건이므로, 가정 폭력 관련 송치, 구속 인원을 조사해보면 된다.

년도 별 가정폭력 가해자 구속 데이터, 년도 별 가정폭력 보호 조치 데이터, 코로나 국내 유입 전부터 코로나 국내 유입 초반 년도인 2020 년까지의 세부 데이터의 쿼리가 필요하다.

또한, 년도 별 가정폭력 가해 건수에 대한 쿼리가 추가로 필요한데, 이는 검거건수+가정보호사건 송치 의 쿼리를 구하며, 피해 건수에 대한 쿼리는 년도 별로 응급조치 2 호+응급 조치 3 호+긴급 임시 조치+임시조치 신청 의 쿼리를 재구성하여 구할 수 있다.

```
In [271]: import pandas as pd
df = pd.read_csv('경찰청_가정폭력 검거 및 조치 현황_20211231.csv', encoding='euc-kr', thousands = ',')
df['연도'] = df['연도'].astype(str)
for idx, row in df.iterrows():
    df['연도'][idx] = str(df['연도'][idx]) + '-01-01'
df['연도'] = df['연도'].astype('datetime64[ns]')
df
```

Out[271]:

	연도	검거건수	검거총인원	구속인원	불구속인원	가정보호사건 송치
0	2017-01-01	38583	45264	384	44880	15979
1	2018-01-01	41905	43576	355	43221	14689
2	2019-01-01	50277	59472	490	58982	21228
3	2020-01-01	44459	52431	330	52101	19379
4	2021-01-01	46041	53985	410	53575	24867

```
In [273]: import pandas as pd
df = pd.read_csv('경찰청_가정폭력 피해자 보호조치 현황_20211231.csv', encoding='euc-kr', thousands = ',')
df['연도'] = df['연도'].astype(str)
for idx, row in df.iterrows():
    df['연도'][idx] = str(df['연도'][idx]) + '-01-01'
df['연도'] = df['연도'].astype('datetime64[ns]')
df
```

Out [273]:

	연도	응급조치 2호(보호기관)	응급조치 3호(의료기관)	긴급임시조치	임시조치 신청
0	2017-01-01	29124	5028	1183	4436
1	2018-01-01	38079	3861	1787	4833
2	2019-01-01	47351	3543	3447	5570
3	2020-01-01	54684	1959	2567	4003
4	2021-01-01	55028	1969	3865	6704

```
import pandas as pd
df = pd.read_csv('경찰청_가정폭력 검거 및 조치 현황_20211231.csv', encoding='euc-kr', thousands = ',')
df['연도'] = df['연도'].astype(str)
for idx, row in df.iterrows():
    df['연도'][idx] = str(df['연도'][idx]) + '-01-01'
df['연도'] = df['연도'].astype('datetime64[ns]')
df.dtypes
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus']=False
plt.rcParams["figure.figsize"] = (20, 20)
plt.title('경찰청 가정폭력 검거 건수')
plt.plot(df['연도'], df['검거건수'], 'b.-', label='검거건수')
plt.xlabel('일자')
plt.legend()
plt.savefig('경찰청 가정폭력 검거 건수.png')
plt.show()
```

```
In [294]: import pandas as pd
df = pd.read_csv('경찰청_가정폭력 피해자 보호조치 현황_20211231.csv', encoding='euc-kr', thousands = ',')
df['연도'] = df['연도'].astype(str)
for idx, row in df.iterrows():
    df['연도'][idx] = str(df['연도'][idx]) + '-01-01'
df['연도'] = df['연도'].astype('datetime64[ns]')
df = df.iloc[1:]
df['피해'] = df['긴급임시조치'] + df['응급조치 3호(의료기관)'] + df['응급조치 2호(보호기관)'] + df['임시조치']
df
```

Out [294]:

	연도	응급조치 2호(보호기관)	응급조치 3호(의료기관)	긴급임시조치	임시조치 신청	피해
1	2018-01-01	38079	3861	1787	4833	48560
2	2019-01-01	47351	3543	3447	5570	59911
3	2020-01-01	54684	1959	2567	4003	63213
4	2021-01-01	55028	1969	3865	6704	67566

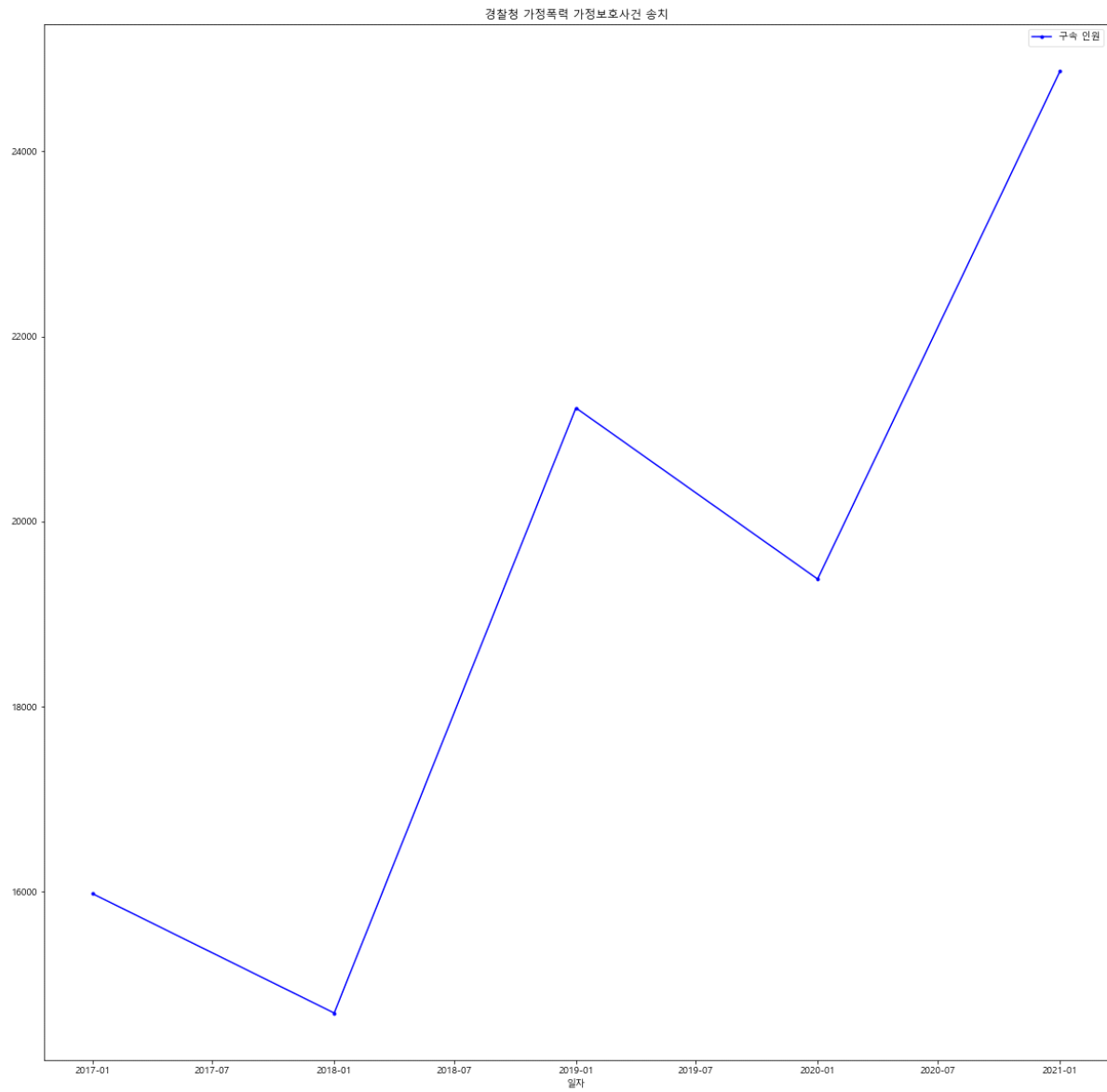
```
In [290]: import pandas as pd
df = pd.read_csv('경찰청_가정폭력 검거 및 조치 현황_20211231.csv', encoding='euc-kr', thousands = ',')
df['연도'] = df['연도'].astype(str)
for idx, row in df.iterrows():
    df['연도'][idx] = str(df['연도'][idx]) + '-01-01'
df['연도'] = df['연도'].astype('datetime64[ns]')
df['가해'] = df['검거건수'] + df['가정보호사건 송치']
df
```

Out [290]:

	연도	검거건수	검거총인원	구속인원	불구속인원	가정보호사건 송치	가해
0	2017-01-01	38583	45264	384	44880	15979	54562
1	2018-01-01	41905	43576	355	43221	14689	56594
2	2019-01-01	50277	59472	490	58982	21228	71505
3	2020-01-01	44459	52431	330	52101	19379	63838
4	2021-01-01	46041	53985	410	53575	24867	70908

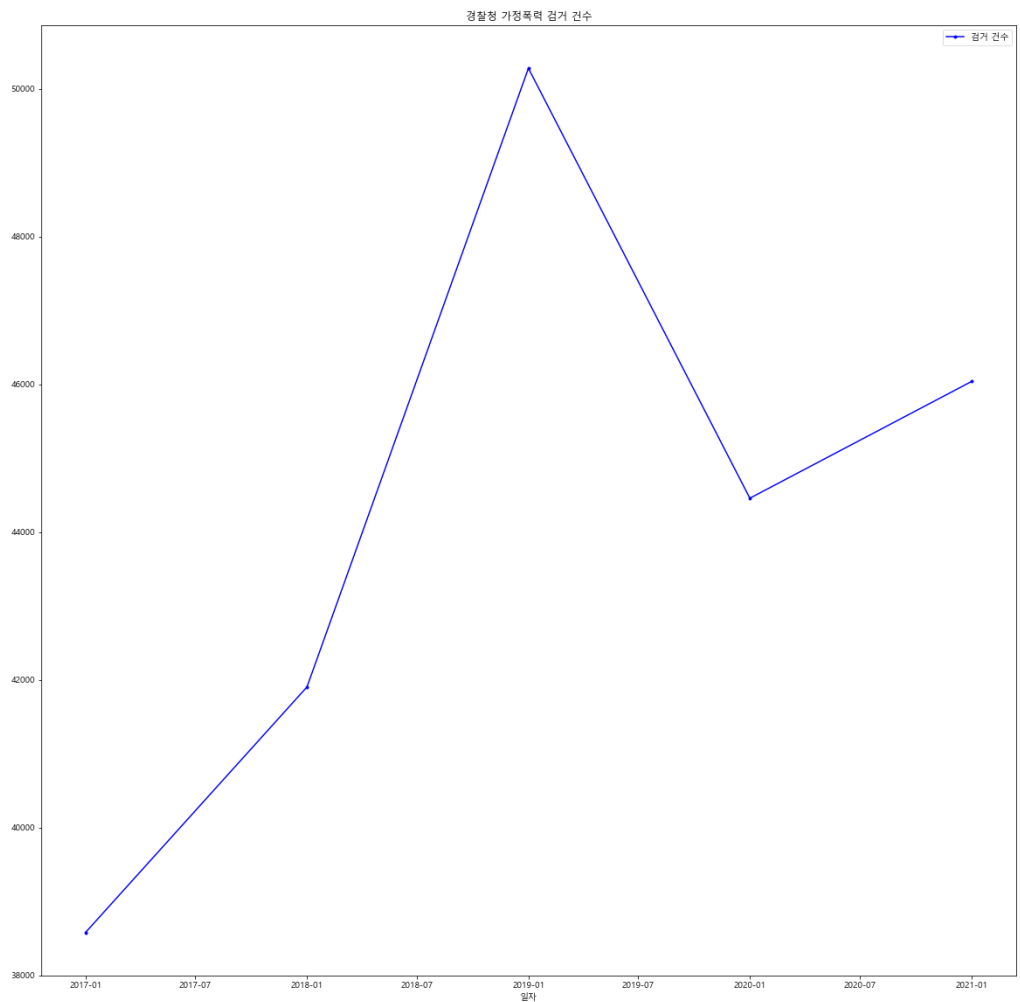
- 위의 쿼리가 필요한 이유

‘가정폭력’이라고 일컬어지는 만큼, ‘집에 있는 시간’이 그 중요점일 수도 있다는
 예상이다. 특히, 이렇게 코로나 19 로 가정에 있는 시간이 늘어나면서 어떠한
 상관관계가 있는지 알아보기 위함이다. 따라서 코로나 19 가 터지기 전 해들과,
 코로나 19 가 국내에 터지고 거리두기가 어느정도 풀렸을 때 까지인 2021 년의
 년도별로의 가정폭력 사건의 쿼리를 필요로 한다. 좀 더 정확한 데이터를 위해,
 코로나 19 의 초창기인 2020 년을 중점적으로 또 다른 데이터의 쿼리가 필요하다.



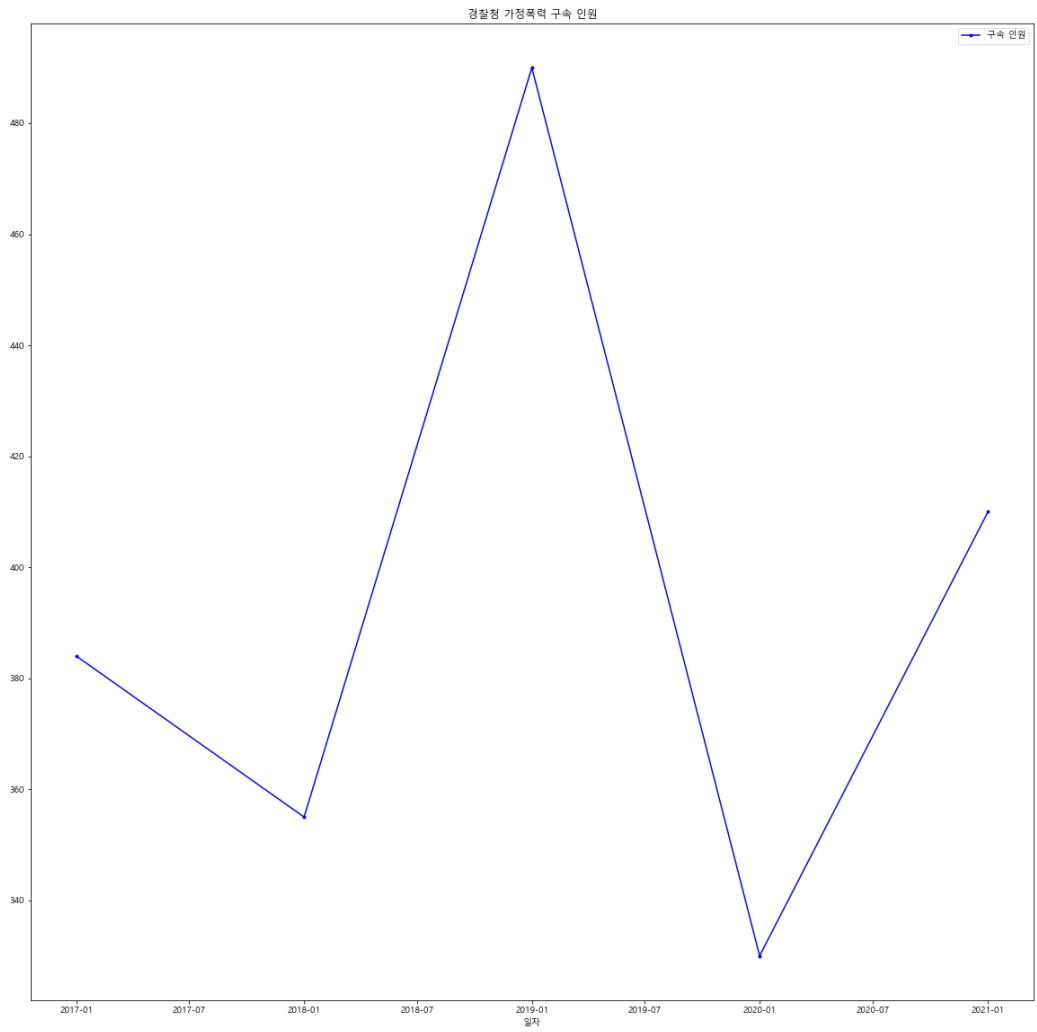
- 해석

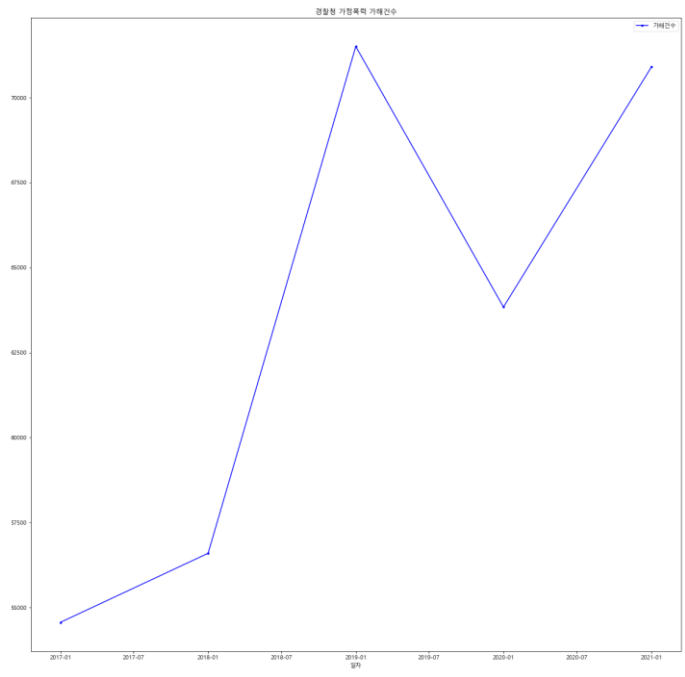
그래프에서도 보았듯이, 2020 년 1 월부터 2021 년 1 월까지 그 수가 증가한 것으로 볼 수 있다. 이는, 2018-2019 년에도 볼 수 있는 양상이므로 코로나 19 의 영향으로 결론짓기에는 다소 이른 감이 있다. 하지만, 감소하는 추세가 그 1 년동안 보이지 않았으므로 아예 영향이 없다고 보기에 어려움은 결론이다.

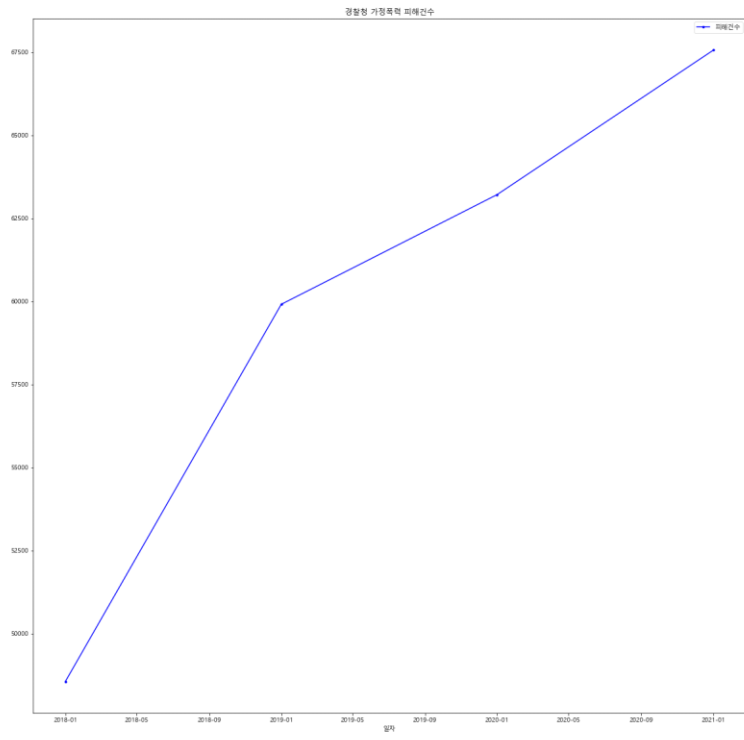


또한, 이것으로만 가정폭력의 여부를 다 알 수는 없다는 점 또한 고려해야한다. 경찰에 관여를 했다는 것은, 누군가가 최초 신고를 했다는 것이고, 따라서 오히려 코로나로 집에 있는 시간이 길어지면서 신고할 수 있는 가능성도 적어졌다는 (가해자와 함께 있는 시간이 길어짐 등) 가능성을 배제할 수 없다는 입장이다.

그것들을 다 고려하고 봤을 때도, 코로나 19 이후 가정폭력의 증감은 증가 쪽이라는 점을 확인해볼 수 있으나, 위에서도 말했다시피 원인을 코로나 19로 보기에다 어렵다는 점이다.







- 해석

이상하게도, 가해건수의 증감 추세와 피해건수의 증감 추세가 다른 것을 알 수 있다.

피해건수의 증감 그래프는 점점 늘어나고 있는 반면, 가해건수의 증감 그래프는 코로나 초기에 잠깐 감소하였다가 증가한 것을 볼 수 있는데, 이는 앞서 언급한 '신고가 어려운 상황'의 영향이 있을 수도 있음을 생각할 수 있다. 따라서 이 그래프가 온전히 가정폭력의 모든 데이터라고 볼 수는 어렵다는 것도 추측해볼 수 있다.

그렇다면 왜 피해 그래프의 양상과 가해 그래프 양상 자체가 반대(2020 년 초)가 될 수 있느냐 에 대한 답을 감히 개인적인 견해를 뒤섞자면, 피해그래프는 가정 폭력으로 인한 부상에 대한 그래프인데, 이는 피해자의 의중이 들어가지 않았기에 굳이 따지자면 가해 그래프보다는 좀 더 정확한 그래프라고 생각한다. 하지만, 가해 그래프는 앞서 말했듯이 피해자가 보복이 두려워, 혹은 다른 이유로 신고자체를 하지 않거나 신고를 하고도 무마하여 가해 기록 조차도 남기지 않으려는 사례가 있기에, 그리고 위에서도 표현했듯이 코로나로 장시간 집에 있는 시간이 길어진 것이 이 사례들에 힘을 실어줄 가능성이 있다 판단하였다.

소감

정한철: 주제에 필요한 데이터를 직접 공공데이터 사이트를 방문해서 찾고 그래프 그리기 유리한 형태로 가공해야 하므로 생각을 많이 하게 됐다. 처음에는 데이터를 가공하는 일이 쉽지 않았다. csv 파일마다 형식이 전부 다 달라서 그래프 그리기 유리한 형태로 가공해야 하는 데 어려움이 있을 때마다 인터넷 검색을 통해서 데이터프레임을 다루는 법을 익혔다. 데이터프레임을 다루는 작업이 많이 진행될수록 내 파이썬 실력이 늘고 있다는 느낌을 받았고 의미 있는 그래프를 그릴 때마다 데이터를 시각화해서 직관적으로 받아들여져서 매우 흥미로웠다. 다음 학기에는 빅데이터와 데이터 사이언스 수업을 신청하고 싶어졌다. 생각보다 코드를 짜야 하는 일과 보고서의 양이 많아서 다소 벅찼는데, 협업을 통해서 혼자 일할 때 보다 더 나은 결과가 발생할 수 있었던 거 같아 협업의 가치를 깨닫게 되었다.



고나연: 데이터에 따라 그 결과를

도출해내고 추가로 추정하는 일이 쉽지는 않다는 것을 느꼈다. 추측했던 결과와 그래프의 양상이 다르면 검색을 하면서 추가로 그 해석을 할 수 있게 되었다. 또한, 데이터에 있어서 기간 설정이 생각보다 중요하다는 것도 깨달았다. 계획서를 쓸 때까지만 해도, 단순히 공공데이터를 csv 로 다운받아 그대로 분석하면 되는 간단한 작업인 줄 알았으나, 최종 과제를 수행하며 데이터를 보기 쉽게 합치고 덜어내는(기간) 가공 단계에서 어려움을 겪고 나니, 계획서와 달라진 부분이 있어 아쉬운 점이 있었다. 허나, 그마저도 다른 식으로 새롭게 해석하고 알아가는 여러 과정에서 단순히 수업을 들었을 때의 흡수력과 직접 과제를 해 나가면서의 흡수력의 결이 다르다는 것도 느꼈다. 물론, 수업 때의 실습은 교수님의 수업 내용을 복습하는데 많은 도움이 되었지만, 과제는 주제부터 학생들이 직접 정하고 그것에 따라 수업 내용을 활용하는 단계이기 때문에 이 점에서 많은 도움이 되었다고 느꼈다. 또한, 팀을 이루어 활동하는 점에서, 분업과 협업의 중요성을 깨달았고, 생각보다 방대했던 양을 혼자서 아니기에 비교적 순조롭게 끝낼 수 있었던 점이 참 의미 있다고 생각했다.

