

넬카말카 세미나 발표

CSV 파일 시각화

1 데이터 시각화

데이터 시각화가 무엇인가?

2 PYTHON으로 시각화

파이썬을 이용한 데이터 시각화

3 코드 설명

데이터를 시각화한 파이썬 코드 설명

4 후기

느낀 점, 깨달은 점

5 마무리

질의응답 & 마무리

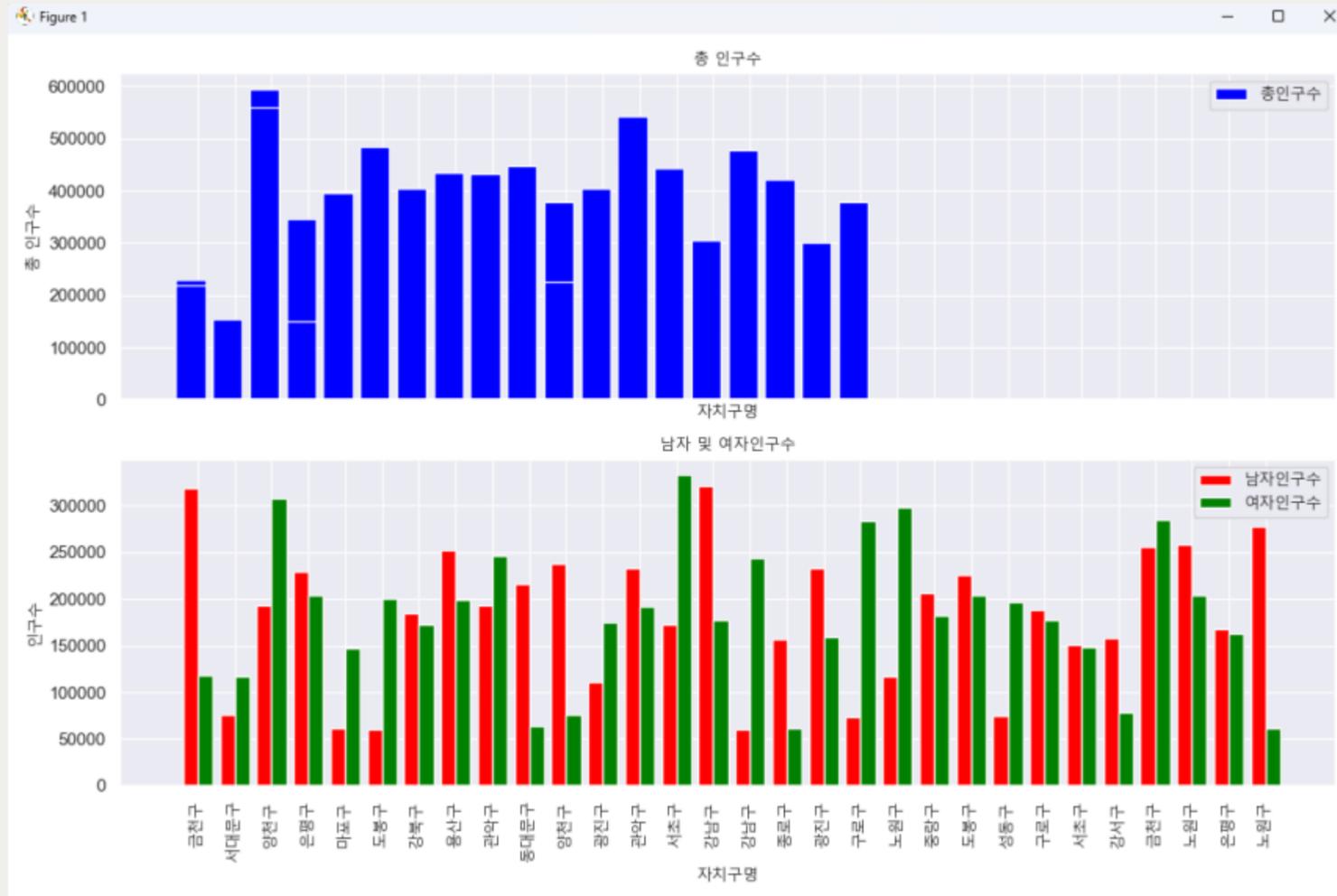


데이터 시각화란?

데이터 시각화는 차트, 플롯, 인포그래픽, 애니메이션과 같이 일반적인 그래픽을 사용하여 데이터를 표현하는 것입니다.

데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정을 말합니다.

Python으로 시각화



전체 코드

```
seoul.py > ...
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 import matplotlib.font_manager as fm
5
6 data = pd.read_csv('seoul3.csv', encoding='utf-8')
7 data['date'] = pd.to_datetime(data['date'], format='%Y%m')
8
9 font_path = 'C:/Windows/Fonts/malgun.ttf'
10 fontprop = fm.FontProperties(fname=font_path)
11
12 print(data.head())
13
14 sns.set(style='darkgrid')
15
16 fig, ax = plt.subplots(2, 1, figsize=(12, 8), sharex=True)
17
18 width0 = 0.4
19 y = range(len(data))
20
21 ax[0].bar(data['atdrc_nm'], data['tot_popltn_co'], label='총인구수', color='blue', align='center')
22 ax[0].bar
23
24 ax[0].set_xticks([i + width0 / 2 for i in y])
25 ax[0].set_xticklabels(data['atdrc_nm'], fontproperties=fontprop, rotation=90)
26
27 ax[0].set_title('총 인구수', fontproperties=fontprop)
28 ax[0].set_ylabel('총 인구수', fontproperties=fontprop)
29 ax[0].set_xlabel('자치구명', fontproperties=fontprop)
30
31 ax[0].set_title('총 인구수', fontproperties=fontprop)
32 ax[0].set_ylabel('총 인구수', fontproperties=fontprop)
33 ax[0].set_xlabel('자치구명', fontproperties=fontprop)
34
35 width = 0.4
36 x = range(len(data))
37
38 ax[1].bar(x, data['male_popltn_co'], width=width, label='남자인구수', color='red', align='center')
39 ax[1].bar([i + width for i in x], data['female_popltn_co'], width=width, label='여자인구수', color='green', align='center')
40
41 ax[1].set_xticks([i + width / 2 for i in x])
42 ax[1].set_xticklabels(data['atdrc_nm'], fontproperties=fontprop, rotation=90)
43
44 ax[1].set_title('남자 및 여자인구수', fontproperties=fontprop)
45 ax[1].set_ylabel('인구수', fontproperties=fontprop)
46 ax[1].set_xlabel('자치구명', fontproperties=fontprop)
47
48 for a in ax:
49     a.legend(prop=fontprop)
50
51 plt.tight_layout()
52 plt.show()
```

코드 설명

```
seoul.py > ...  
1 import pandas as pd  
2 import matplotlib.pyplot as plt  
3 import seaborn as sns  
4 import matplotlib.font_manager as fm  
5
```

파이썬 라이브러리 부분입니다.

pandas는 파이썬으로 데이터를 조작하거나 분석하기 위해 사용되는 라이브러리입니다.

일반적으로 데이터 전처리, 분석, 변환 작업등에 주로 사용됩니다.

Matplotlib는 데이터 시각화를 위한 라이브러리입니다.

seaborn은 Matplotlib을 기반으로 하는 시각화 라이브러리입니다.

데이터 시각화에서 그래픽화 하기 위해 필요합니다.

Matplotlib의 pyplot 은 간단한 인터페이스 제공합니다.

코드 설명

```
data = pd.read_csv('seoul3.csv', encoding='utf-8')
data['date'] = pd.to_datetime(data['date'], format='%Y%m')

font_path = 'C:/Windows/Fonts/malgun.ttf'
fontprop = fm.FontProperties(fname=font_path)
```

data 변수에 pandas 라이브러리로 데이터가 들어가있는 csv 불러옵니다
불러온것에 date 칼럼에서 20200101 로 되어 있는 부분을 format 하여 형식을 변환해줍니다.

font_path는 한글이 깨질 경우를 대비해 폰트 경로를 설정해줍니다
Matplotlib 의 서브 모듈인 font_manager을 여기서 사용합니다.

코드 설명

```
sns.set(style='darkgrid')

fig, ax = plt.subplots(2, 1, figsize=(12, 8), sharex=True)

width0 = 0.4
y = range(len(data))

ax[0].bar(data['atdrc_nm'], data['tot_popltn_co'], label='총인구수', color='blue', align='center')
ax[0].bar

ax[0].set_xticks([i + width0 / 2 for i in y])
ax[0].set_xticklabels(data['atdrc_nm'], fontproperties=fontprop, rotation=90)

ax[0].set_title('총 인구수', fontproperties=fontprop)
ax[0].set_ylabel('총 인구수', fontproperties=fontprop)
ax[0].set_xlabel('자치구명', fontproperties=fontprop)
```

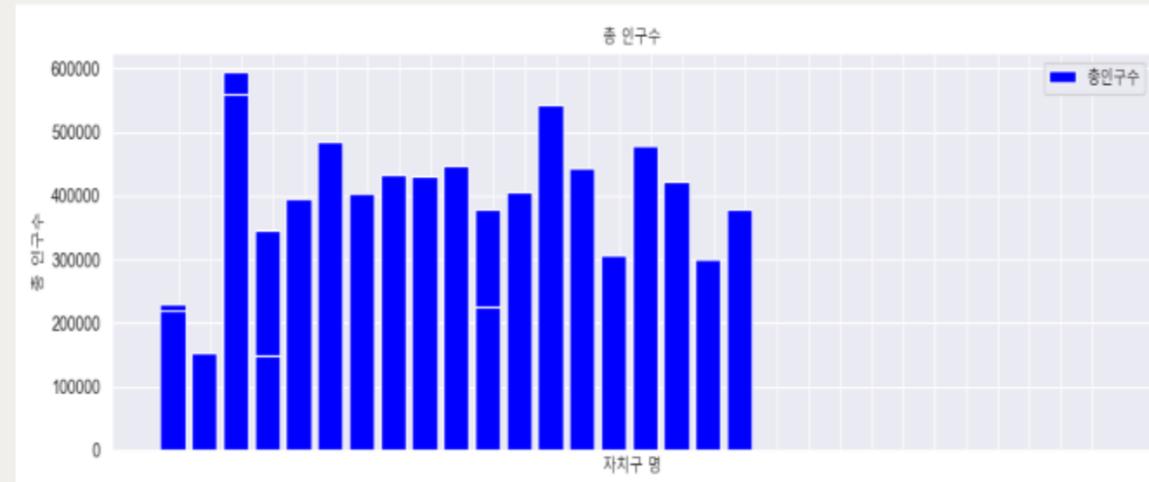


차트 중에 자치구 별 총 인구수를 나타내주는 코드입니다.

코드 설명

```
28
29 width = 0.4
30 x = range(len(data))
31
32 ax[1].bar(x, data['male_popltn_co'], width=width, label='남자인구수', color='red', align='center')
33 ax[1].bar([i + width for i in x], data['female_popltn_co'], width=width, label='여자인구수', color='green', align='center')
34
35 ax[1].set_xticks([i + width / 2 for i in x])
36 ax[1].set_xticklabels(data['atdrc_nm'], fontproperties=fontprop, rotation=90)
37
38 ax[1].set_title('남자 및 여자인구수', fontproperties=fontprop)
39 ax[1].set_ylabel('인구수', fontproperties=fontprop)
40 ax[1].set_xlabel('자치구명', fontproperties=fontprop)
41
42 for a in ax:
43     a.legend(prop=fontprop)
44
45 plt.tight_layout()
46 plt.show()
```

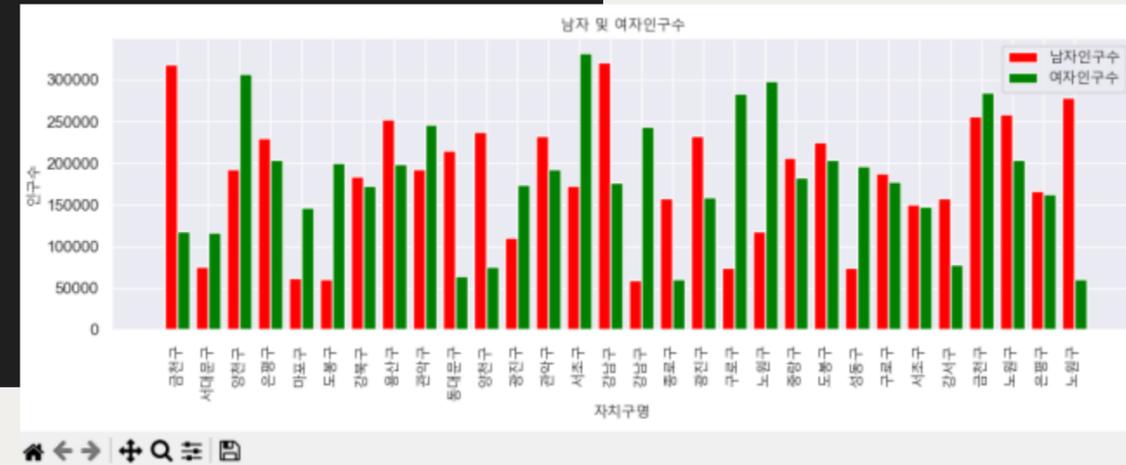


차트 중에 자치구 별 남자 및 여자 인구수를 나타내주는 코드입니다.

코드 설명

```
kr_daily2.py > ...
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 import matplotlib.font_manager as fm
5
6 data = pd.read_csv('kr_daily.csv')
7
8 data['date'] = pd.to_datetime(data['date'], format='%Y%m%d')
9
10 font_path = 'C:/Windows/Fonts/malgun.ttf'
11 fontprop = fm.FontProperties(fname=font_path)
12
13 sns.set(style='darkgrid')
14
15 fig, ax = plt.subplots(figsize=(12, 8))
16
17 ax.plot(data['date'], data['confirmed'], label='확진자 수', color='blue')
18
19 ax.plot(data['date'], data['death'], label='사망자 수', color='red')
20
21 ax.plot(data['date'], data['released'], label='격리 해제된 사람 수', color='green')
22 ax.set_title('코로나바이러스감염증-19 타임라인', fontproperties=fontprop)
23 ax.set_ylabel('인구 수', fontproperties=fontprop)
24 ax.set_xlabel('날짜', fontproperties=fontprop)
25
26 # 범례 추가
27 ax.legend(prop=fontprop)
28
29 # 그래프 보여주기
30 plt.show()
31
```



날짜 별 코로나바이러스감염증-19 타임라인을 나타내는 코드입니다.

후기

파이썬 코드로 데이터를 시각화해보고 많은 것을 배웠습니다.
미래에도 많은 데이터들을 시각화 해야할 일들이 많을 것입니다.
이를 미리 대비해 배워두어 나중에 실무에서 응용할 수 있을 것으로 예상됩니다.

사실 누구든 쉽게 도전할수 있는 부분이라고 생각합니다. 데이터를 수치화 한다는 것에
어려운 코드가 필요하지 않기 때문인데요,

여기에서 그치지않고
지금까지 발표했던 내용에서 나아가 더 세부적인 부분을 공부해서
숫자로된 수치들만 시각화해보지 않고
문자열 같은 부분도 시각화해보고 싶은 생각이 있습니다.



감사합니다.
