

21년 서울시 코로나 확진자 분석

이한길



프로젝트 개요

목표 설정

- 서울시 코로나19 확진자 현황 데이터를 분석하여 유의미한 정보 도출
- 데이터 분석을 위한 데이터 정제, 데이터의 특성 분석, 데이터 시각화

프로젝트 수행 과정

1) 데이터 수집 :

- 서울시 공공 데이터를 활용하여 21년도 1월~10월 각 도시별 확진자 데이터 수집
- 서울시 행정구역 시군구 정보(좌표계) 데이터 수집

2) 데이터 전처리 :

- 유효한 데이터 만 사용하기 위해 csv 파일 확인 후 필요한 컬럼만 추출
- 지역별 확진자 시각화를 위해 위도,경도 데이터를 추출

3) 데이터 분석 및 시각화 :

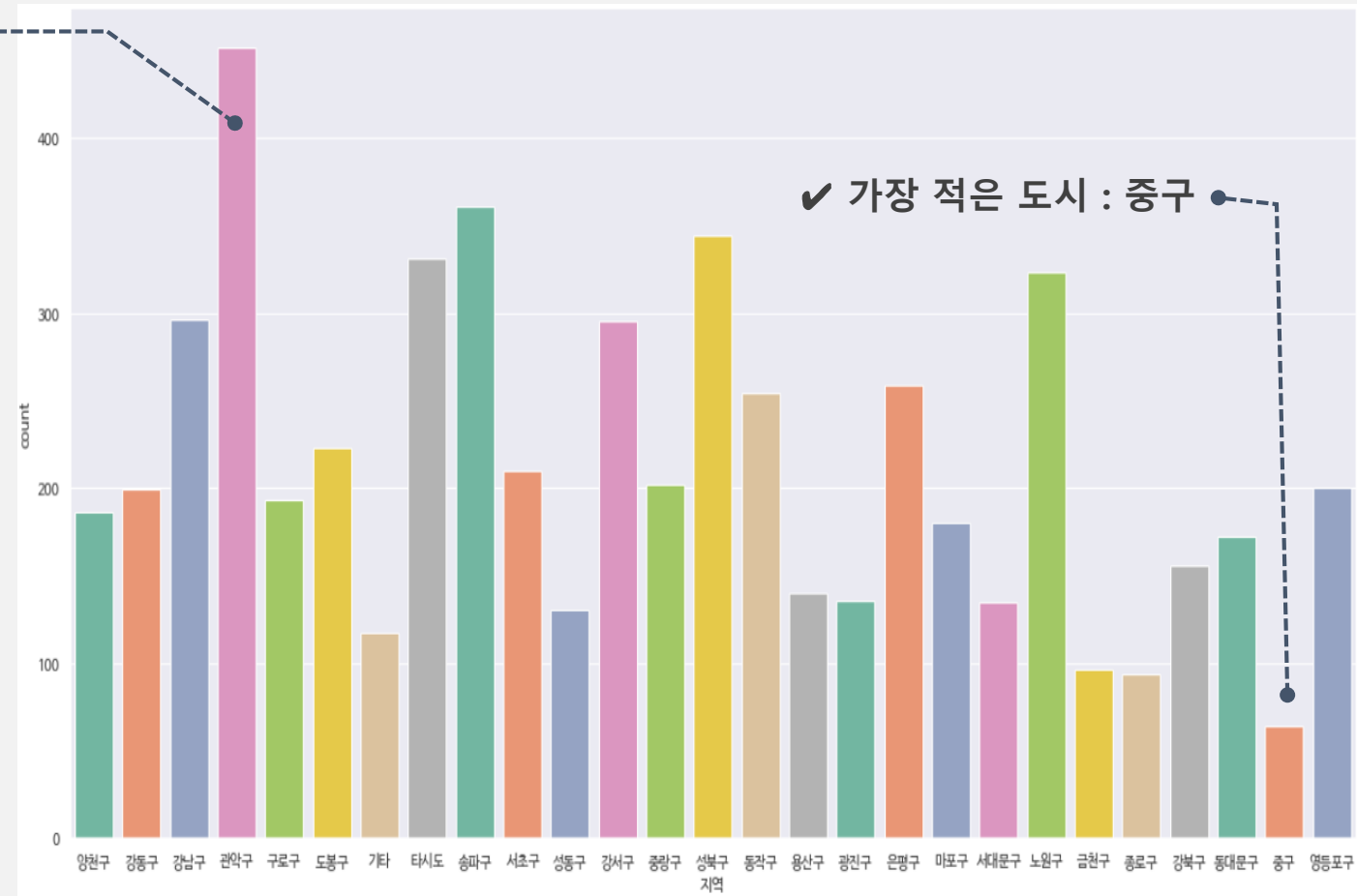
- 월별 확진자 수 파악, 확진자 수가 많은 특정월의 일 평균 확진자 수 확인
- 각 도시별 확진자 확인, 산출된 데이터의 시각화



확진자 데이터 시각화 및 분석

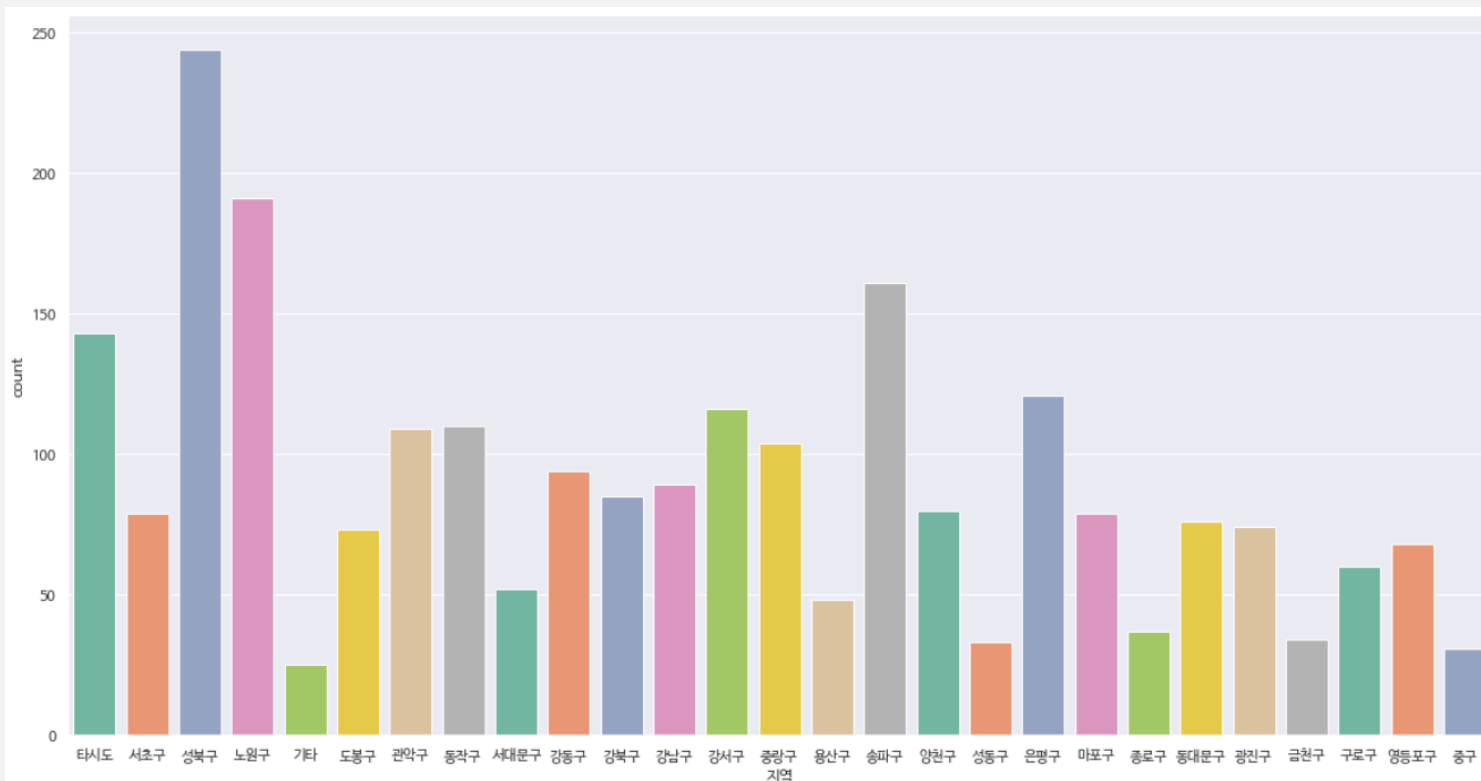
전처리한 데이터를 통해
각 도시별 확진자 파악

가장 많은 도시 : 관안구





확진자 데이터 시각화 및 분석



- 확진자가 가장 많은 달은 8월달로 확인
- 8월달 가장 많은 확진자가 발생한 곳은 성북구

```
[173] # value_counts()는 각 데이터를 세어서 내림차순으로 정리  
# 확진자가 가장 많은 달 확인  
corona_data['month'].value_counts()
```

```
8    2416  
9    1304  
6     460  
10    425  
3     391  
7     281  
5     228  
4     156  
2      80  
1       7  
Name: month, dtype: int64
```

```
[21] # 확진자가 많았던 8월의 지역별 데이터 확인  
corona_out_region[corona_data['month'] == '8']
```

	확진일	지역	접촉력	상태	month	day
1271	8.31.	타시도	감염경로 조사중	퇴원	8	31
1272	8.25.	타시도	성북구 사랑제일교회 관련	NaN	8	25
1273	8.26.	타시도	타시도 확진자 접촉	퇴원	8	26
1274	8.27.	타시도	기타 확진자 접촉	퇴원	8	27
1504	8.30.	타시도	감염경로 조사중	퇴원	8	30



서울시 지도 시각화를 위한 데이터 준비

지역별 확진자 지도 시각화

도시별 확진자의 시각화를 위해 folium 패키지 활용
서울시 행정구역 좌표계를 저장 후 확인

✓ 서울 지역의 확진자를 지도에 출력하여 확인 합니다.
(지도 출력을 위한 folium 패키지 활용)

```
[24] # 지도 출력을 위한 라이브러리 folium을 import 합니다.  
import folium
```

```
# Map 함수를 사용하여 지도를 출력합니다.  
map_osm = folium.Map(location=[37.529622, 126.984307], zoom_start=11)  
  
map_osm
```

```
[25] # CRS에 행정구역 좌표계 저장  
CRS=pd.read_csv("/content/drive/MyDrive/Colab Notebooks/Covid_Data/서울시 행정구역 시군구 정보 (좌표계_ WGS1984).csv")
```

```
[26] CRS.head()
```

	순번	시군구코드	시군구명_한글	시군구명_영문	ESRI_PK	위도	경도
0	1	11320	도봉구	Dobong-gu	0	37.665861	127.031767
1	2	11380	은평구	Eunpyeong-gu	1	37.617612	126.922700
2	3	11230	동대문구	Dongdaemun-gu	2	37.583801	127.050700
3	4	11590	동작구	Dongjak-gu	3	37.496504	126.944307
4	5	11545	금천구	Geumcheon-gu	4	37.460097	126.900155



좌표계 데이터 전처리



```
# 중구의 위도, 경도를 중심으로 시각화
m= folium.Map(location=[37.557945,126.99419], zoom_start=11)

for region in set(corona_seoul['지역']):
    count = len(corona_seoul[corona_seoul['지역']==region])
    CRS_region = CRS[CRS['시군구명_한글'] == region]
    #반복문을 돌며 위도,경도 데이터 저장 (.iloc 를 이용하여 필요한 데이터만[위도,경도] 가져옵니다)
    w = CRS_region.iloc[0]['위도']
    g = CRS_region.iloc[0]['경도']

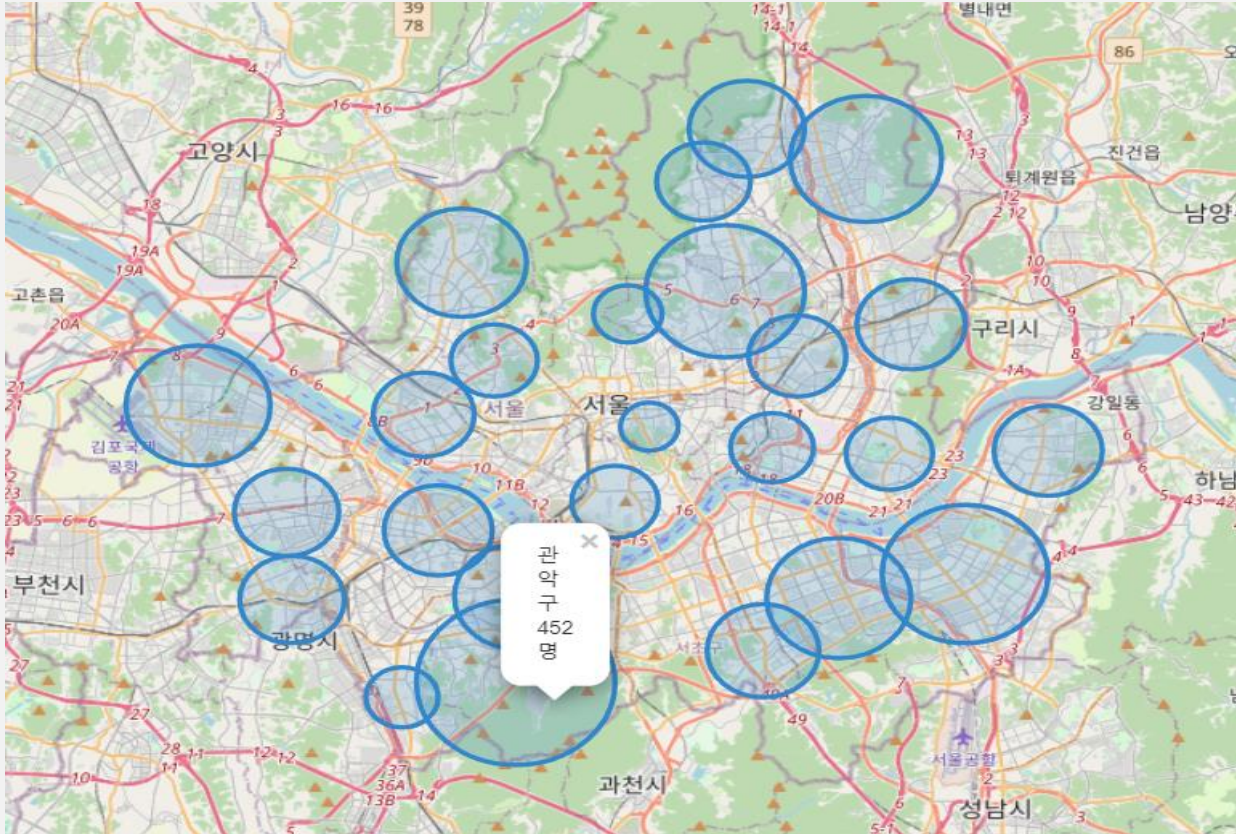
    folium.CircleMarker([w,g], # 위도,경도 표시
                        radius=count/10+10, # 범위
                        color='#3186cc', # 선 색상
                        fill_color='#3186cc', # 면 색상
                        popup=' '.join((region,str(count),'명'))
                        ).add_to(m)

m
```

좌표계 데이터를 전처리 하고,
해당 데이터를 활용해 지도에 시각화



서울시 지도 시각화



✓ 21년 1~10월 서울시에서 관악구의 확진자가 가장 많았으며 관악구의 확진자 수는 452명

✓서울시에서 확진자가 가장 많았던 달은 8월 이고, 8월의 서울시의 각 도시별 일평균 확진자 수는 약 78명

✓ folium 패키지의 CircleMarker를 활용하여 서울시의 확진자 데이터를 확인