서울시 범죄 현황 분석

2조

강호연, 유혁재, 김용현

START

목차





서울시 범죄 현황을 분석하는 이유?

[국감브리핑] '부자 동네'...서울 강남 3구 체감안전도 높 아

종로서 76.6점 1위...관악서 60.0점 최하위

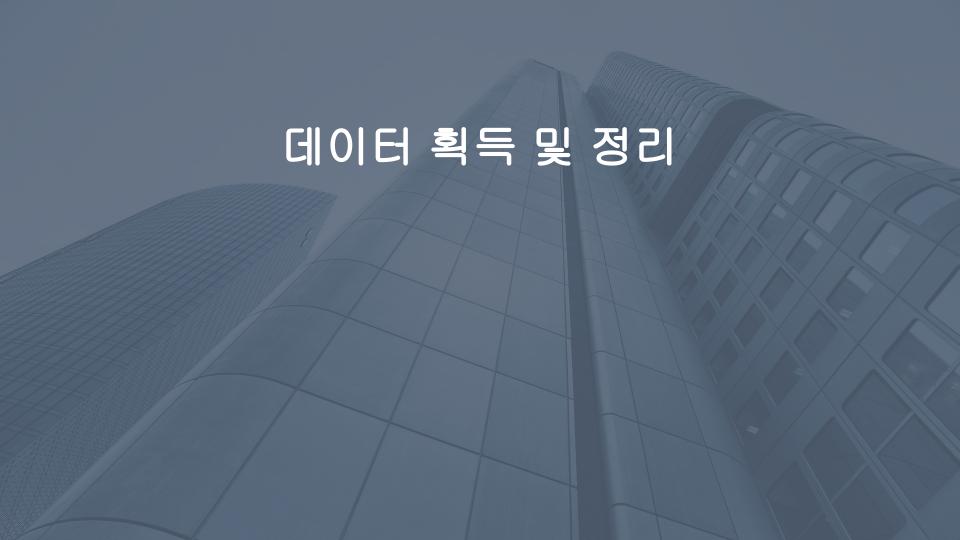
(서울=뉴스1) 성도현 기자 | 2014-10-20 09:54 송고





부유층이 많이 살아 '부자 동네'로 불리는 강남 3구(강남·서초·송파) 지역의 체감안 전도가 대체로 높은 것으로 나타났다.

- 체감 안전도가 높다는 것을 검증하기위해 실제 안전도가 높은지 분석



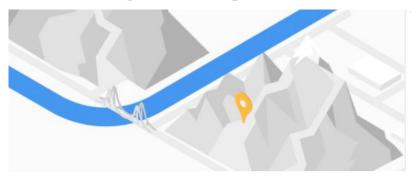
[기 데이터 정리

N	관서명	살인 발생	살인 검거	강도 발생	강도 검거	강간 발생	강간 검거	절도 발생	절도 검거	폭력 발생	폭력 검거
0	중부서	2	2	3	2	105	65	1395	477	1355	1170
1	종로서	3	3	6	5	115	98	1070	413	1278	1070
2	남대문서	1	0	6	4	65	46	1153	382	869	794
3	서대문서	2	2	5	4	154	124	1812	738	2056	1711
4	혜화서	3	2	5	4	96	63	1114	424	1015	861



- 한 구에 하나 혹은 두 군데의 경찰서
- 구 이름과 다른 경찰서

Google Maps API



Google Maps API

- geocoding api
- 위치에 대한 결과 중 주소와 위도, 경도 정보 제공

사용전 터미널에 pip install googlemaps 명령어

>pip install googlemaps



🖸 데이터 정리

```
In [3]: import googlemaps

In [9]: gmaps_key = "AlzaSyCcOwLK-2AQKdh7AkD9huUtR116AQu7QxM"
 gmaps = googlemaps.Client(key=gmaps_key)

In [14]: gmaps.geocode('서울중부경찰서', language='ko')
```

- Google maps를 이용해서 '서울중부경찰서'라는 단어 검색
- 경찰서의 주소와 위도, 경도 정보도 확인

☑ 데이터 정리

경찰서의 이름이 중부서, 수서서 처럼 되어있음검색에서 주소가 제대로 나오지 않음

관서명

0 중부서

종로서

2 남대문서

3 서대문서

4 혜화사

```
- 서울OO경찰서로 만들어야함
```

```
In [18]: station_name = []

for name in crime_anal_police['관서명']:
    station_name.append('서울' + str(name[:-1]) + '경찰서')

station_name
```

[데이터 정리

```
for name in station_addreess:
    tmp = name.split()

    tmp_gu = [gu for gu in tmp if gu[-1] == '구'][0]

    gu_name.append(tmp_gu)

crime_anal_police['구별'] = gu_name
crime_anal_police.head()
```

저장한 경찰서 주소에서 구만 따로 선택해서 구별이라는 컬럼에 저장

Out [27]:		관서명	살인 발생	살인 검거	강도 발생	강도 검거	강간 발생	강간 검거	절도 발생	절도 검거	폭력 발생	폭력 검거	구별
	0	중부서	2	2	3	2	105	65	1395	477	1355	1170	중구
	1	종로서	3	3	6	5	115	98	1070	413	1278	1070	종로구
	2	남대문서	1	0	6	4	65	46	1153	382	869	794	중구
	3	서대문서	2	2	5	4	154	124	1812	738	2056	1711	서대문구
9	4	혜화서	3	2	5	4	96	63	1114	424	1015	861	종로구

[] 데이터 정리

서울금천경찰서

현 청사로 이전하기 전엔 금천경찰서인데 관악구에 있는 것과, 1973년도에 지어져 현 청사 이전 전까지 쓰였던 매우 오래된 청사로 유명했었다.



금천경찰서의 경우 관악구에 위채해 있어서 예외 처리 (2018년도에 이전)

- 관서명: 관악구 → 금천구

In [32]: crime_anal_police.loc[crime_anal_police['관서명']=='금천서', ['구별']] = '금천구' crime_anal_police[crime_anal_police['관서명']=='금천서']

Out[32]:

-	관서명	살인 발생	살인 검거	강도 발생	강도 검거	강간 발생	강간 검거	절도 발생	절도 검거	폭력 발생	폭력 검거	구별
15	금천서	3	4	6	6	151	122	1567	888	2054	1776	금천구

[데이터 정리

 \times

관서명을 기초로 했기때문에 구별 컬럼에 구 이름이 두 번 나올 수 있음

```
In [83]: crime_anal_raw = pd.read_csv('../data/02. crime_in_Seoul_include_gu_name.csv', encoding='utf-8', index_col=0)

crime_anal = pd.pivot_table(crime_anal_raw, index='구별', aggfunc=np.sum)
crime_anal.head()
```

구별	강간 검거	강간 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
강남구	349	449	18	21	10	13	1650	3850	3705	4284
강동구	123	156	8	6	3	4	789	2366	2248	2712
강북구	126	153	13	14	8	7	618	1434	2348	2649
관악구	221	320	14	12	8	9	827	2706	2642	3298
광진구	220	240	26	14	4	4	1277	3026	2180	2625

[] 데이터 정리

- 추가로 각 범죄별 검거율 계산, 검거 건수는 검거율로 대체할 수 있어서 삭제

```
In [84]:

crime_anal['강간검거율'] = crime_anal['강간 검거']/crime_anal['강간 발생']*100
crime_anal['강도검거율'] = crime_anal['강도 검거']/crime_anal['강도 발생']*100
crime_anal['살인검거율'] = crime_anal['살인 검거']/crime_anal['살인 발생']*100
crime_anal['절도검거율'] = crime_anal['절도 검거']/crime_anal['절도 발생']*100
crime_anal['폭력검거율'] = crime_anal['폭력 검거']/crime_anal['폭력 발생']*100

del crime_anal['강간 검거']
del crime_anal['살인 검거']
del crime_anal['절도 검거']
del crime_anal['폭력 검거']

crime_anal.head()
```

- 발생이란 단어 삭제

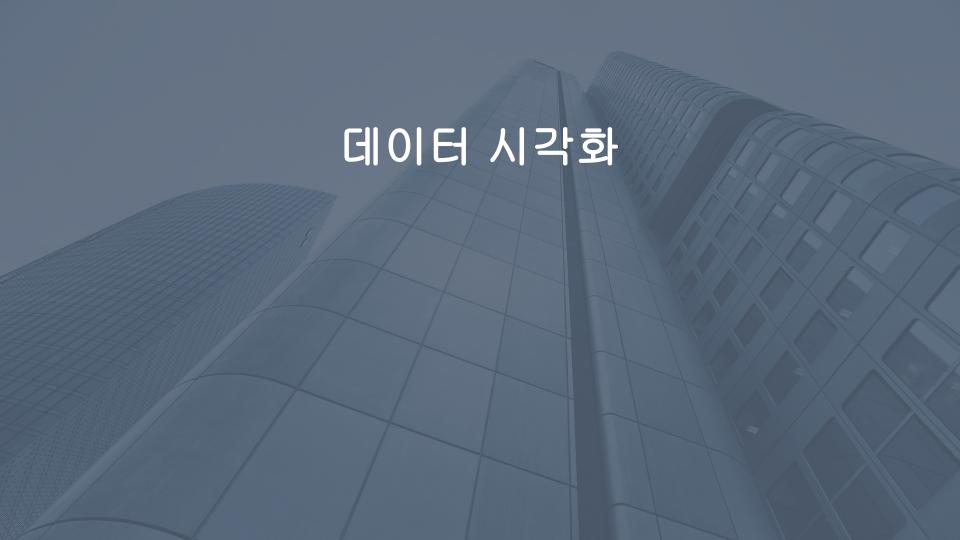


```
In [87]: from sklearn import preprocessing
        col = ['강간', '강도', '살인', '절도', '폭력']
        x = crime_anal[col].values
        min_max_scaler = preprocessing.MinMaxScaler()
        x_scaled = min_max_scaler.fit_transform(x.astype(float))
        crime_anal_norm = pd.DataFrame(x_scaled, columns = col, index = crime_anal.index)
        col2 = ['강간검거율', '강도검거율', '살인검거율', '절도검거율', '폭력검거율']
        crime_anal_norm[col2] = crime_anal[col2]
        crime anal norm.head()
```

구별	강간	강도	살인	절도	폭력	강간검거율	강도검거율	살인검거율	절도검거율	폭력검거율
강남구	1.000000	0.941176	0.916667	0.953472	0.661386	77.728285	85.714286	76.923077	42.857143	86.484594
강동구	0.155620	0.058824	0.166667	0.445775	0.289667	78.846154	100.000000	75.000000	33.347422	82.890855
강북구	0.146974	0.529412	0.416667	0.126924	0.274769	82.352941	92.857143	100.000000	43.096234	88.637222
관악구	0.628242	0.411765	0.583333	0.562094	0.428234	69.062500	100.000000	88.888889	30.561715	80.109157
광진구	0.397695	0.529412	0.166667	0.671570	0.269094	91.666667	100.000000	100.000000	42.200925	83.047619

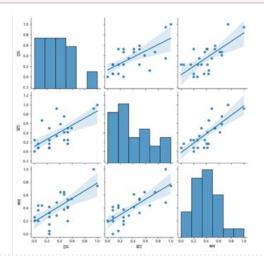
♡ 정규화

01. CCTV result.csv을 읽어서 인구수와 CCTV 개수 가져오기 In [34]: result_CCTV = pd.read_csv('../data/01. CCTV_result.csv', encoding='UTF-8', index col='구별') crime_anal_norm[['인구수', 'CCTV']] = result_CCTV[['인구수', '소계']] crime anal norm.head() 발생 건수의 합을 '범죄'라는 항목으로 통합 In [35]: col = ['강간','강도','살인','절도','폭력'] crime_anal_norm['범죄'] = np.sum(crime_anal_norm[col], axis=1) crime anal norm.head() 검거율 통합 In [36]: col = ['강간검거율','강도검거율','살인검거율','절도검거율','폭력검거율'] crime anal_norm['검거'] = np.sum(crime_anal_norm[col], axis=1) crime_anal_norm



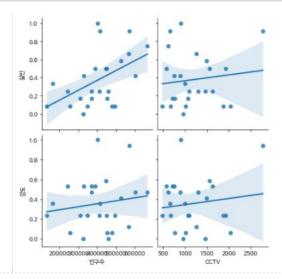
- 1) 그래프에 대한 한글 폰트 문제 해결 (사진은 생략)
- 2) pairplot으로 강도, 살인, 폭력 간의 상관관계 그래프 그리기

sns.pairplot(crime_anal_norm, vars=["강도", "살인", "폭력"], kind='reg', size=3) plt.show()



모든 그래프를 봤을 때 기울기가 양임을 알 수 있다.

3) x축을 인구수와 CCTV 개수, y축을 살인과 강도에 대해 조사



CCTV와 살인의 관계가 낮아보이지만, CCTV가 많이 없을 때 살인이나, 강도가 높은 수를 갖는 것을 볼 수 있다.

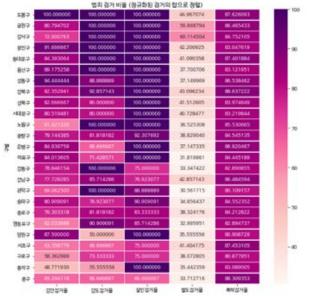
- 4) 검거율 그래프
- a) 각 구별로 검거율 데이터 확인

```
tmp_max = crime_anal_norm['검거'].max()
crime_anal_norm['검거'] = crime_anal_norm['검거'] / tmp_max * 100
crime_anal_norm_sort = crime_anal_norm.sort_values(by='검거', ascending=False)
crime_anal_norm_sort.head()
    구별
  도봉구 0.000000 0.235294 0.083333 0.000000 0.000000
                                                    100.000000
                                                                   100.0
                                                                                  44.967074 87.626093 348646.0
                                                                                                                 485 0.063725
                                                                                                                             100.000000
  금천구 0.141210 0.058824 0.083333 0.172426 0.134074
                                                     80.794702
                                                                   100.0
                                                                                            86.465433 255082.0
                                                                                                                1015 0.117973
                                                                                                                              97.997139
  광진구 0.397695 0.529412 0.166667 0.671570 0.269094
                                                                                  42.200925 83.047619 372164.0
                                                     91.666667
                                                                   100.0
                                                                                                                707 0.406888
                                                                                                                              96.375820
                                                     84.393064
                                                                   100.0
                                                                                  41.090358 87.401884 369496.0
                                                                                                                1294 0.298029
                                                                                                                              95.444250
  용산구 0.265130 0.529412 0.250000 0.169004 0.133128
                                                                            100 0 37 700706 83 121951 244203 0
                                                                   100.0
```

- 4) 검거율 그래프
 - a) 각 구별로 검거율 데이터 확인
- b) 검거율을 100으로 한정하여 정렬 후, heatmap에 그리기

1) 절도 검거율 낮음 2) 검거율이 우수한 구, 낮은 구 구별



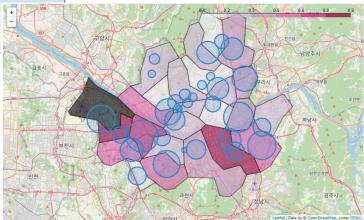


- 4) 검거율 그래프
 - a) 각 구별로 검거율 데이터 확인
- b) 검거율을 100으로 한정하여 정렬 후, heatmap에 그리기
- c) 발생 건수의 합으로 heatmap 정렬
 - 1) 강남, 양천, 영등포가 가장 높다.
 - 2) 송파, 서초구도 높은 편임
 - 3) 강남 **3**구가 안전하다고 보장 **X**

			별죄	비율 (정규화된	발생 건수로	정렬)		
	구성장	1000000	0941176	0916687	0953472	0661386	0894540	- 10
	양천구	0.806916	0823529	0.666687	1.000000	1000000	0.859423	
	영동포구	0.556196	1000000	1000000	0650359	0.493024	0739916	
	송파구	0.34005/8	0.470588	0750000	0744441	0.627524	0546522	
	관약구	0428242	0.411765	0543333	0562004		0.022710	- 08
	마포구	0553314	0529412	0500000	0510434	0.353748	6489762	-
	구로구	0.515850	0588235	0530000		0.959423	0.479756	
	서초구	0.838617	0235294	0.000000	0537904	0.215654	CARSETA	
	887	0.244957	0352941	0916667	0.356746	0.921589	0.880500	
	광진구	0.297695	6525412	0.166667	0671570	0.269094	0.406868	- 06
	노용구	0.273775	0117647	0.666667	0.386589	0.292268	0.347389	
51	홍로구	0.314121	0352941	0333333	0.383510	0.190589	0314899	
1	강복구	0.140974	0529412	0.416667	0126924	0.274769	0298949	
- 13	동대문구	0.204611	0470588	0250000	0.314061	0250887	0298029	
	용작구	6427,178	0235294	0.250000	0274376	0100024	0277414	-04
	용산구	0.265130	0529412	0250000	0.169004	0.133128	0.209335	
	87	0195965	0235294	0083333	0.508040	0174273	0239381	
	강동구	0155620	0.058824	0166667	2445775	0.289667	0223310	
	운영구	0.184438	0235294	0083333	0.291139	0.275715	0213964	
	성복구	0138329	0.000000	0250000	0.247007	0.170726	0161212	-02
	성동구	0069164	0235204	0166667	0196110	0.029558	0137359	
	급천구	0141210	0058824	0083333	0172426	0134074	0117973	
	서대문구	0149856	0.000000	0,000000	0.256244	0134547	0108129	
	도병구	0.000000	0235294	0083333	0.000000	0000000	0.063725	
		강간	25	설인	3/5	목적	智利	-00

지도 시각화 도구 Folium

강남 3구의 안정성을 얘기했는데 그렇게 범죄율, 검거율을 지도에 그리면서 데이터를 표현하는 것이 훨 시각적으로 도움이 된다.



서울시 범죄율에 대한 지도 시각화

구별로 구별하기 위해서는 경계선을 그려야 함. 서울시 구별 경계선을 그리기엔 문제가 발생함. 하지만 깃허브에 대한민국에 대한 json 파일을 얻을 수 있음. 이를 서울시 구별 데이터만 편집하여서 사용



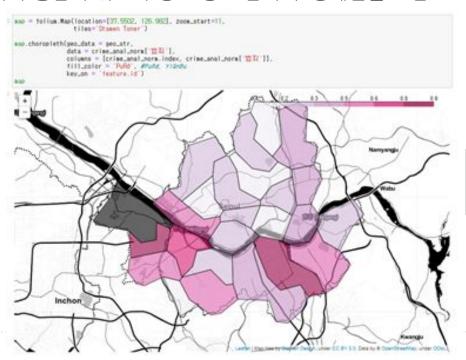
서울시 구별 json 데이터

서울시 범죄율에 대한 지도 시각화

Json 파일 로딩

import json
geo_path = '../data/02. skorea_municipalities_geo_simple.json'
geo_str = json.load(open(geo_path, encoding='utf-8'))

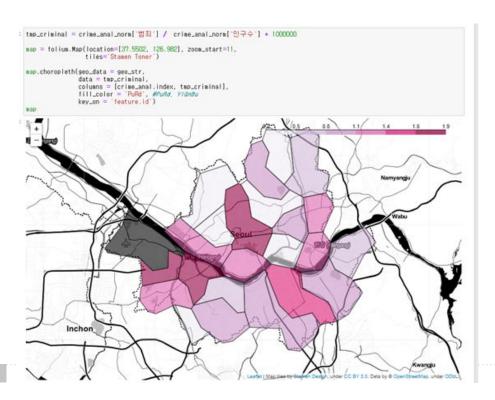
-> 서울시의 중심의 위도와 경도 정보 입력 후 경계선을 그림. -> 범죄 발생 건수로 표현



강남 3구와 주변 범죄 발생 건수 높음

서울시 범죄율에 대한 지도 시각화

3) 인구수 고려하기 = 인구 대비 범죄 발생 비율



- 인구대비 범죄 발생 건수가 강남이 1위는 아니지만 안전도 낮음
- 중구, 종로구 범죄율 상승 이유 ->거주인원 적고 관광지 때문 예상

- 경찰서별 검거율과 범죄 발생율을 동시에 표현하기
- 검거만 따로 모아두기

```
crime_anal_raw['lat'] = station_lat
crime_anal_raw['lng'] = station_lng

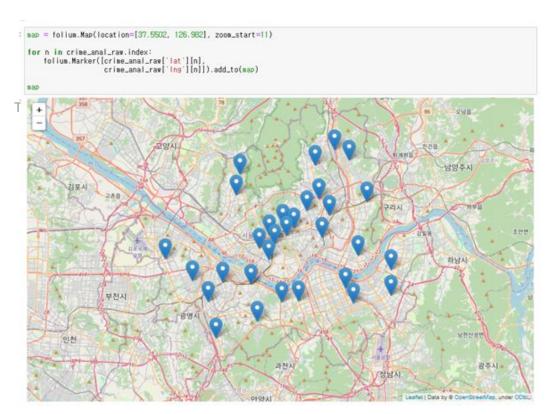
col = ['살인 검거', '강도 검거', '강간 검거', '절도 검거', '폭력 검거']
tmp = crime_anal_raw[col] / crime_anal_raw[col].max()

crime_anal_raw['검거'] = np.sum(tmp, axis=1)

crime_anal_raw.head()
```

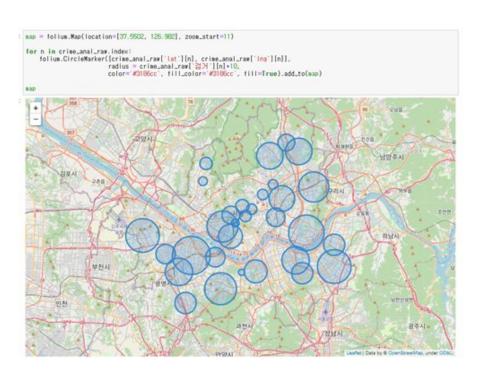
관서명	살인 발 생	살인 검 거	강도 <u>발</u> 생	강도 검 거	강간 발 생	강간 검 거	절도 발 생	절도 검 거	폭력 <u>발</u> 생	폭력 검 거	구별	lat	Ing	검거
중부서	2	2	3	2	105	65	1395	477	1355	1170	중구	37.563646	126.989580	1.275416
종로서	3	3	6	5	115	98	1070	413	1278	1070	종로구	37.575548	126.984747	1.523847
남대문 서	1	0	6	4	65	46	1153	382	869	794	중구	37.554758	126.973498	0.907372
서대문 서	2	2	5	4	154	124	1812	738	2056	1711	서대문 구	37.564744	126.966770	1.978299
혜화서	3	2	5	4	96	63	1114	424	1015	861	종로구	37.571853	126.998914	1.198382
1	0 중부서 1 종로서 2 남대문 서 3 서대문	전시형 생 0 중부서 2 1 종로서 3 2 남대문 1 3 서대문 2	변시병 생 거 0 중부서 2 2 1 종로서 3 3 2 남대문 서 1 0 3 서대문 2 2	변시성 생 거 생 0 중부서 2 2 3 1 종로서 3 3 6 2 남대문 1 0 6 3 서대문 2 2 5	변시병 생 거 생 거 0 중부서 2 2 3 2 1 종로서 3 3 6 5 2 남대문 서 1 0 6 4 3 서대문 2 2 5 4	변시성 생 거 생 거 생 거 생 0 중부서 2 2 3 2 105 1 종로서 3 3 6 5 115 2 남대문 서 1 0 6 4 65 3 서대문 2 2 5 4 154	판서병 생 거 생 거 생 거 0 중부서 2 2 3 2 105 65 1 종로서 3 3 6 5 115 98 2 남대문 서 1 0 6 4 65 46 3 서대문 2 2 5 4 154 124	반시형 생 거 생 거 생 거 생 거 생 0 중부서 2 2 3 2 105 65 1395 1 종로서 3 3 6 5 115 98 1070 2 남대문 서 1 0 6 4 65 46 1153 3 서대문 서 2 2 5 4 154 124 1812	반시당 생 거 생 거 생 거 생 거 생 거 생 거 생 거 생 거 생 거 생	반시형 생 거 생 거 생 거 생 거 생 거 생 여 생 여 생 여 생 이 중부서 2 2 3 2 105 65 1395 477 1355 1 종로서 3 3 6 5 115 98 1070 413 1278 12 남대문 서 1 0 6 4 65 46 1153 382 869 13 4대문 시 2 2 5 4 154 124 1812 738 2056	전시당 생 거 생 거 생 거 생 거 생 거 생 거 생 거 생 거 생 거 생	전시당 생 거 생 거 생 거 생 거 생 거 생 거 생 거 가 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다	전시당 생 거 생 거 생 거 생 거 생 거 생 거 생 거 가 다 다입. 0 중부서 2 2 3 2 105 65 1395 477 1355 1170 중구 37.563646 1 종로서 3 3 6 5 115 98 1070 413 1278 1070 종로구 37.575548 2 남대문 시 1 0 6 4 65 46 1153 382 869 794 중구 37.554758 3 서대문 시 2 2 5 4 154 124 1812 738 2056 1711 서대문 37.564744	전시당 생 거 생 거 생 거 생 거 생 거 생 거 생 거 가 나 가 다른 내용 내용 이 중부서 2 2 3 2 105 65 1395 477 1355 1170 중구 37.563646 126.989580 1 종로서 3 3 6 5 115 98 1070 413 1278 1070 종로구 37.575548 126.984747 2 남대문 시 0 6 4 65 46 1153 382 869 794 중구 37.554758 126.973498 3 서대문 시 2 2 5 4 154 124 1812 738 2056 1711 서대문 지 37.564744 126.966770

• 경찰서의 위도와 경도 정보 확인

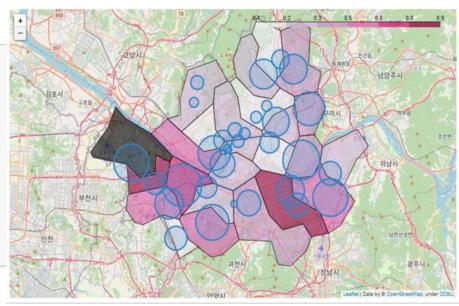


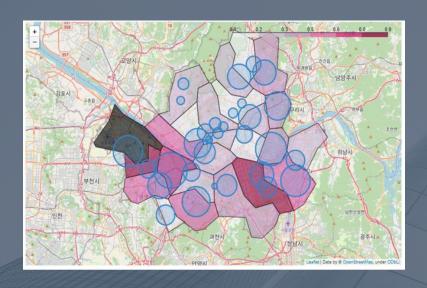
• 앞서 모아둔 검거에 10을 곱해서 원 넓이를 정하고 경찰서의 검거율을 원의 넓이로 정함

원이 클수록 검거율이 좋다고 할 수 있음.



● 원 안에 범죄 발생 건수를 붉은 정도로 비교하게 하면 검거율와 범죄 발생율 동시에 확인 가능





결론

범죄가 많이 일어날 수록 붉은색이고, 원이 클 수록 검거율이 높다.

서울 서부는 범죄율이 높지만 그만큼 검거율도 높다. 강북쪽은 검거율, 범죄율이 적은 것을 알 수 있다.

강남구는 색이 진한데 원의 크기가 영등포구의 원보다 작은데 이는 강남구가 안전하다고 보기 어렵다.

느낀점

평소 공부했던 파이썬은 별을 찍고, 계산기를 만드는 프로그래밍만 생각했다면, 이번 데이터 분석을 통해 그래프를 코드 몇줄로 쉽게 제작할 수 있고, 더 나아가서 그 데이터들을 시각화를 하여 보기 쉽게 구분하는 것을 보고 파이썬의 넓은 활용도를 직접적으로 코드를 쳐보면서 몸소 체험할 수 있는 좋은 경험이었다.

외부 라이브러리를 가져와서 쓰는 것에 익숙치 않아서 조금 두려움을 가졌고, 개념을 먼저 완벽히 공부하고 문제를 푸는 학교 공부와 달리 직접 체험하면서 배우는 책의 방식을 통해 프로그래밍의 새로운 공부 방법을 제시해주었다.

파이썬을 통한 데이터 분석으로 한 걸음 나아갈 수 있는 경험이었다. 프로젝트를 수행할 때 어려움을 겪으면서 문제해결 능력도 향상시킬 수 있었다. 강의를 들으며 따라가는 배움이 아닌 모르는 것이 생길 때마다 스스로 찾아보는 학습방법이 어색했지만 오히려 이 공부방법이 더 깊은 학구열을 자극할 수 있는 기회가 된 것 같았다.



Thanks for Watching