

데이터 분석 시각화 프로젝트

서울시 범죄 데이터 분석 시각화

산업시스템공학과 2017112560 서준혁
산업시스템공학과 2017112508 최관호

진행 순서

INDEX

01.데이터 설명

//

- 주제 선정 이유
- 데이터셋 설명

//

02.EDA 목표

//

- 궁금한 점
- 연구 방향

//

03.통찰

//

- 시각화 과정
- 결론과 의미

//

04.고찰

//

- 아쉬운 점
- 더 알고 싶은 점

//

01

데이터 설명

서울시 범죄 데이터

데이터 설명 - 주제 선정 배경

주제: 서울시 범죄 발생 현황/ 유형조사에 기반한 특정 지역의 경력배치, 순찰 경로 지정

형식적인 거점 근무 때문에
정작 필요할 때 출동이
늦어지는 경우가 많은 것
같아!

범죄가 일어날 만한
장소와 시간에 집중
적으로 순찰을 돌 순
없을까?



서울시 범죄 데이터

데이터 설명 - 데이터셋

사용 데이터셋 - 범죄 발생 시간 및 유형 조사

A. 경찰청_범죄발생시간대 및 요일

-범죄 유형별, 시간대, 요일에 따른 범죄 발생 수

B. 경찰청_범죄발생장소별 통계

-범죄 유형별, 장소 별 범죄 발생 수

C. 서울시 5대 범죄 발생현황 통계

-서울시 행정 구 별, 범죄 유형 별 범죄 발생 수

서울시 범죄 데이터

데이터 설명 - 데이터셋

사용 데이터셋 - 강남구의 지역 별 특징 추출 및 분류

A. 공동주택현황_2020,하반기

- 강남구 아파트,다세대주택 등 위치 정보

B. 서울특별시 강남구 관광숙박업 인허가 정보

- 강남구 호텔, 모텔 등 위치 정보

C. 서울특별시 강남구 목욕장업 인허가 정보

- 강남구 목욕탕, 사우나 등 위치 정보

D. 강남구도시관리공단_공영주차장 정보 2020

- 강남구 공영주차장 위치 정보

E. 서울특별시 강남구 유흥주점 영업 인허가 정보

- 강남구 유흥주점 위치 정보

02

EDA 목표

서울시 범죄 데이터

EDA 목표 – 궁금한 점 (1)

A. 범죄가 가장 많이 일어나는 지역구는 어디일까?

- 범죄 발생 유형, 수를 통한 범죄 예방 필요 지역 선정

B. 범죄가 가장 많이 일어나는 장소는 어디일까?

- 단속하고자 하는 범죄만을 추출하여 장소 별 범죄 횟수 분석

C. 범죄 유형 별 범죄가 일어나는 장소에 특징이 있지 않을까?

- 범죄를 대분류로 나눈 후 장소 별 범죄 횟수 및 특징 분석

D. 범죄는 어느 요일에 가장 많이 일어날까?

- 요일 별 범죄 발생 횟수/ 유형 분석

서울시 범죄 데이터

EDA 목표 – 궁금한 점 (2)

E. 시간대 별 범죄율은 어떻게 변할까?

- 시간대 별 범죄 발생 횟수/유형 분석

F. 관할구 내 지역 별로 어떤 특징이 있을까?

- 범죄가 자주 일어나는 장소의 위치 분석 및 특정 장소들이 많이 분포한 지역 탐구

G. 기타

- 범죄 유형 별 상관관계, 시간과 요일의 상관관계, cctv 설치 유무 등

프로젝트 목적

분석 목적

WHAT?

“범죄가 많이 일어나는 행정 구의 관서가
범죄를 예방하기 위해
주목해야 할 장소와 시간은?”



HOW?

범죄 데이터 분석

03

통찰

서울시 범죄 데이터

시각화 과정 - 전처리

<서울시 관서별 5대 범죄 발생/검거 현황>

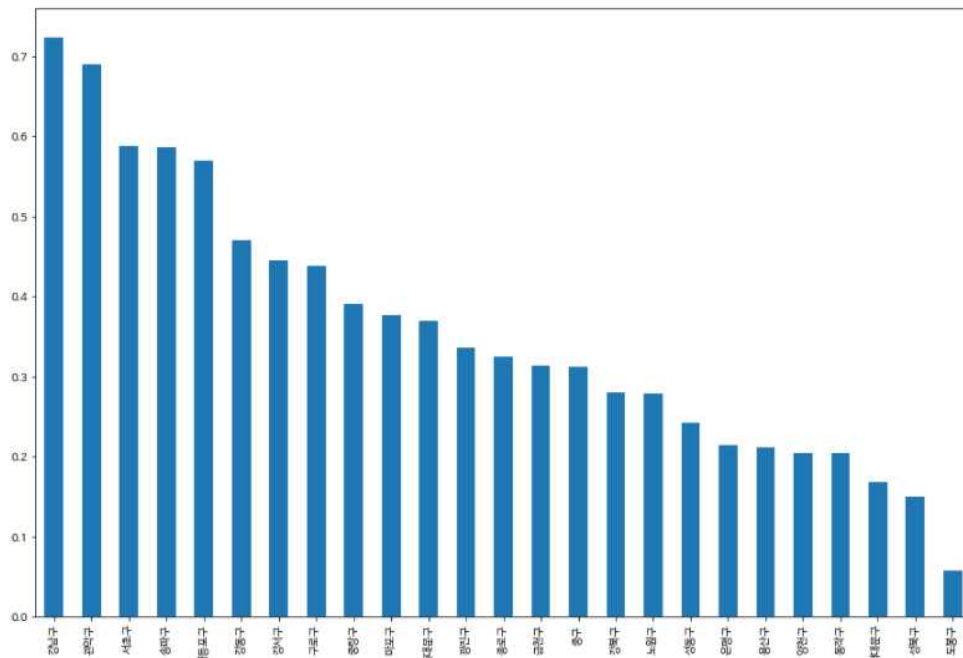
					강간발생 강도발생 살인발생 절도발생 폭력발생					강간발생 강도발생 살인발생 절도발생 폭력발생						
구분	죄종	발생검거	건수		관할구					관할구						
0	중부	살인	발생	1	강남구	666	5	5	2970	3652	강남구	1.000000	0.285714	0.333333	1.000000	1.000000
1	중부	살인	검거	1	강동구	180	15	6	1650	2161	강동구	0.156250	1.000000	0.416667	0.394495	0.385914
2	중부	강도	발생	4	강북구	176	3	6	1229	2422	강북구	0.149306	0.142857	0.416667	0.201376	0.493410
3	중부	강도	검거	4	강서구	306	3	8	2050	2555	강서구	0.375000	0.142857	0.583333	0.577982	0.548188
4	중부	강간	발생	127	관악구	408	10	13	2223	2667	관악구	0.552083	0.642857	1.000000	0.657339	0.594316

- 관서 별 범죄 발생만을 추출
- 해당 지역 관할구로 그룹화
- 범죄 유형 별 같은 단위로 비교하기 위해 정규화

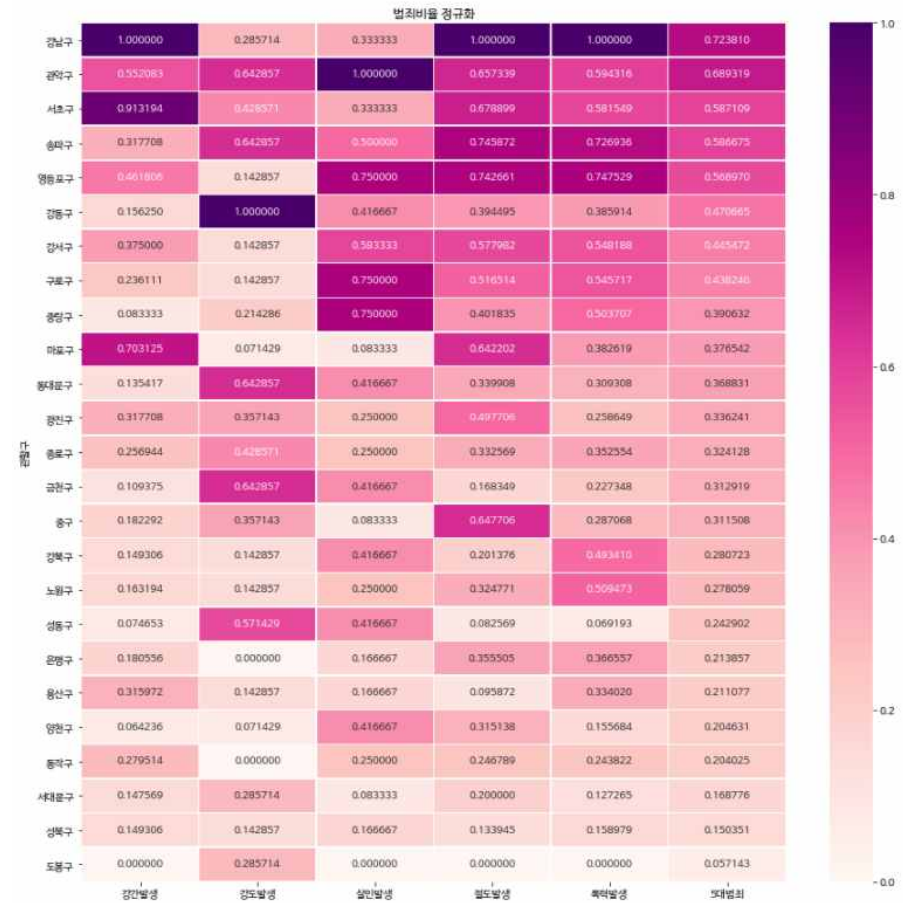
서울시 범죄 데이터

시각화 과정 - 결과

A. 범죄가 가장 많이 일어나는 지역구는 어디일까?



<서울시 관서별 5대 범죄 발생/검거 현황>



서울시 범죄 데이터 시각화 과정 - 전처리

<서울시 장소별 범죄 발생량>

범죄대분류	범죄중분류	아파트, 연세대	노상	백화점	슈퍼마켓	대형할인매장	상점	시장, 노점	숙박업소, 목욕탕	유점점	사무실	공사장, 광산	참고	역, 지하철				
0	살인, 인신	95	95	0	59	0	1	1	0	2	0	21	12	8	4	1	1	
1	살인, 인신	90	102	3	126	0	0	1	0	3	4	16	25	14	4	0	4	
2	살인, 인신	138	182	2	423	0	26	154	1	36	10	89	101	33	3	3	0	3
3	살인, 인신	882	954	0	249	0	0	4	0	32	19	1447	281	58	4	2	12	3
4	살인, 인신	68	63	0	30	0	0	0	0	2	2	47	36	9	0	0	0	0
5	살인, 인신	902	790	11	3324	9	66	158	13	175	75	671	1639	411	33	6	6	222



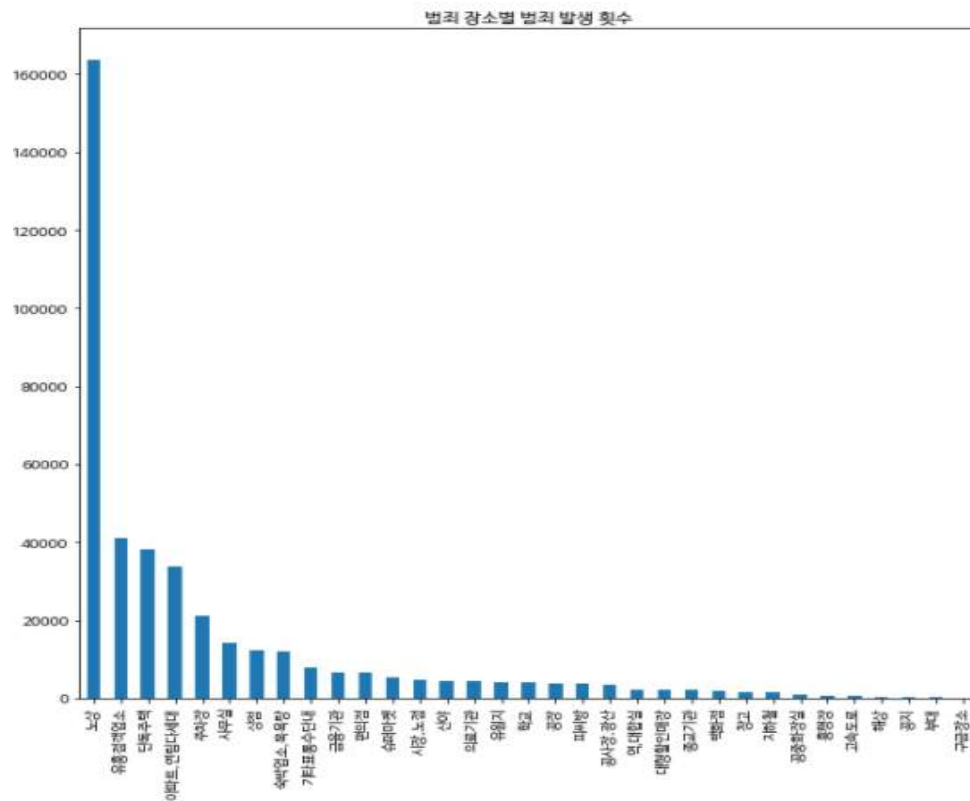
범죄대분류	아파트, 연세대	대학, 다문화	고층노출	노상	백화점	슈퍼마켓	편의점	대형할인매장	상점	시장, 노점	숙박업소, 목욕탕	유점점	흥개, 소	사무실	공사장	참고	역, 대학
강력범죄	2670	2787	17	4469	11	99	322	16	277	145	2454	2187	586	69	16	33	243
절도범죄	14333	19515	177	49527	1828	4718	4579	1966	9520	2998	6736	13952	6247	2412	2384	1424	1172
폭력범죄	16774	15930	273	109581	139	595	1617	93	2554	1616	2836	24774	7288	1239	877	163	859

- 필요하지 않은 선거범죄, 병역범죄 등은 제외하고 범죄 대분류로 그룹화

서울시 범죤 데이터

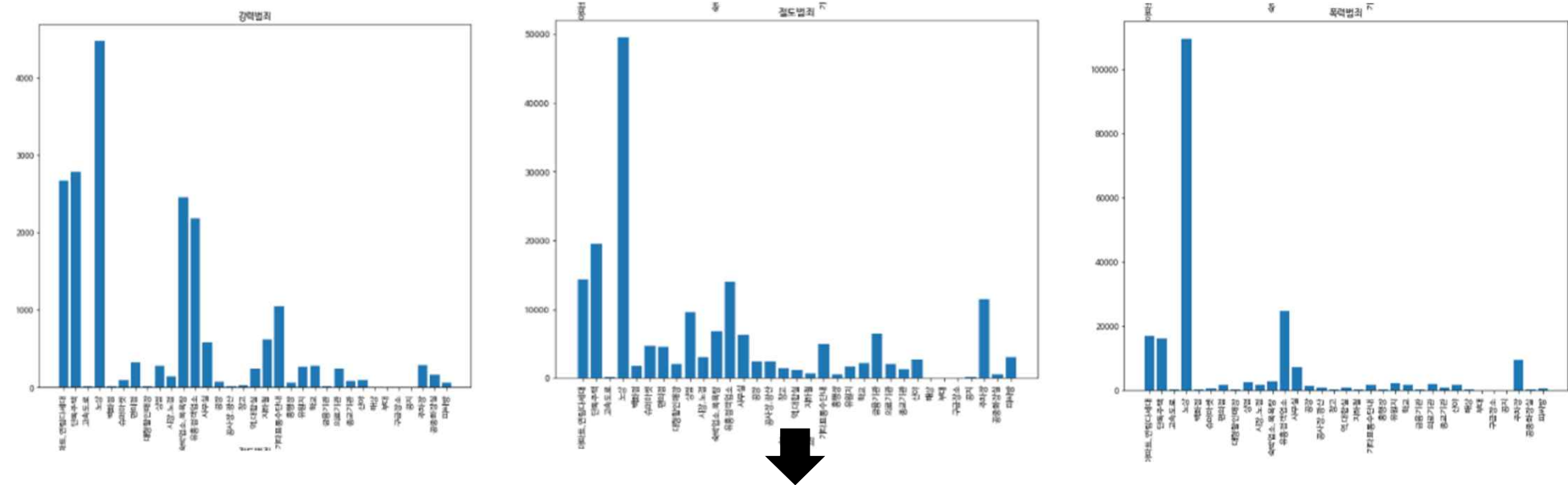
시각화 과정 - 결과

B. 범죄가 가장 많이 일어나는 장소는 어디일까?



<서울시 장소별 범죄 발생량>

C. 범죄 유형 별 범죄가 일어나는 장소에 특징이 있지 않을까?



노상, 아파트/주택을 제외한 범죄 유형 별 특징적인 장소 추출



서울시 범죄 데이터

시각화 과정 - 전처리

<서울시 시간 및 요일 별 범죄 발생량>

	00:00-02:59	03:00-05:59	06:00-08:59	09:00-11:59	12:00-14:59	15:00-17:59	18:00-20:59	21:00-23:59	일	월	화	수	목	금	토
범죄대분류															
강력범죄	2920	3316	1977	1799	1966	2447	3089	4215	4081	3519	3607	3634	3445	3880	4310
절도범죄	13088	14167	13094	21501	27167	28553	27229	20568	25466	26239	25671	26061	26098	28156	29266
폭력범죄	31290	26880	17572	23112	23919	28939	39618	57944	45295	38034	38553	38905	39086	41198	46842

요일과 시간으로 분리

범죄대분류	강력범죄	절도범죄	폭력범죄
월	3519	26239	38034
화	3607	25671	38553
수	3634	26061	38905
목	3445	26098	39086
금	3880	28156	41198
토	4310	29266	46842
일	4081	25466	45295

범죄대분류	강력범죄	절도범죄	폭력범죄
월	0.187283	0.053947	0.058924
화	0.218497	0.156579	0.098887
수	0.000000	0.166316	0.119437
목	0.502890	0.707895	0.359219
금	1.000000	1.000000	1.000000
토	0.735260	0.000000	0.824364
일	0.085549	0.203421	0.000000

범죄대분류	강력범죄	절도범죄	폭력범죄
00:00-02:59	2920	13088	31290
03:00-05:59	3316	14167	26880
06:00-08:59	1977	13094	17572
09:00-11:59	1799	21501	23112
12:00-14:59	1966	27167	23919
15:00-17:59	2447	28553	28939
18:00-20:59	3089	27229	39618
21:00-23:59	4215	20568	57944

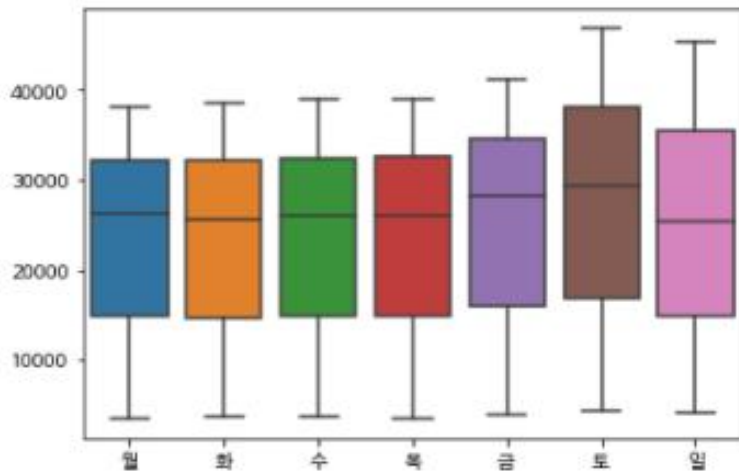
범죄대분류	강력범죄	절도범죄	폭력범죄
00:00-02:59	0.463990	0.000000	0.339790
03:00-05:59	0.627897	0.069770	0.230556
06:00-08:59	0.073675	0.000388	0.000000
09:00-11:59	0.000000	0.544003	0.137224
12:00-14:59	0.069123	0.910378	0.157213
15:00-17:59	0.268212	1.000000	0.281557
18:00-20:59	0.533940	0.914387	0.546072
21:00-23:59	1.000000	0.483673	1.000000

➡ 요일별, 시간별 유형별 범죄 발생의 특징을 찾기 위해 정규화 작업 진행

서울시 범죄 데이터

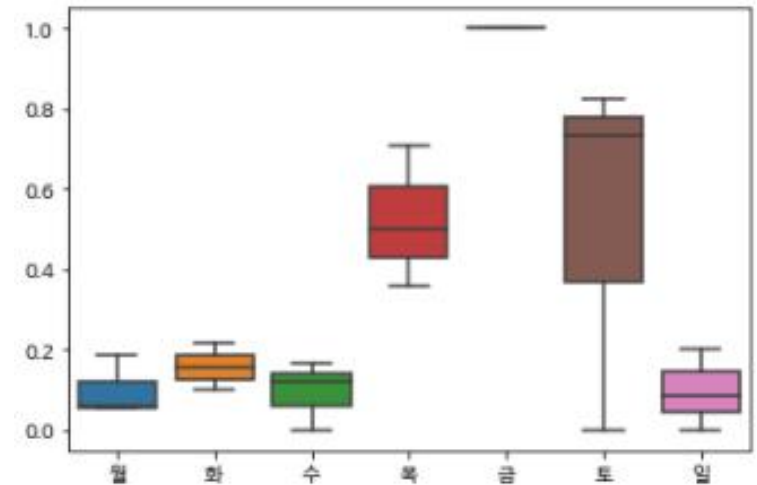
시각화 과정 - 결과

D. 범죄는 어느 요일에 가장 많이 일어날까? - 정규화의 필요성



<서울시 요일 별 범죄 발생량>

- 정규화 하기 전 요일 별 범죄 발생 비교.
데이터 수가 너무 많고 유형 별 범죄 횟수가 너무
많이 차이 나기 때문에 한눈에 비교하기가 어렵다



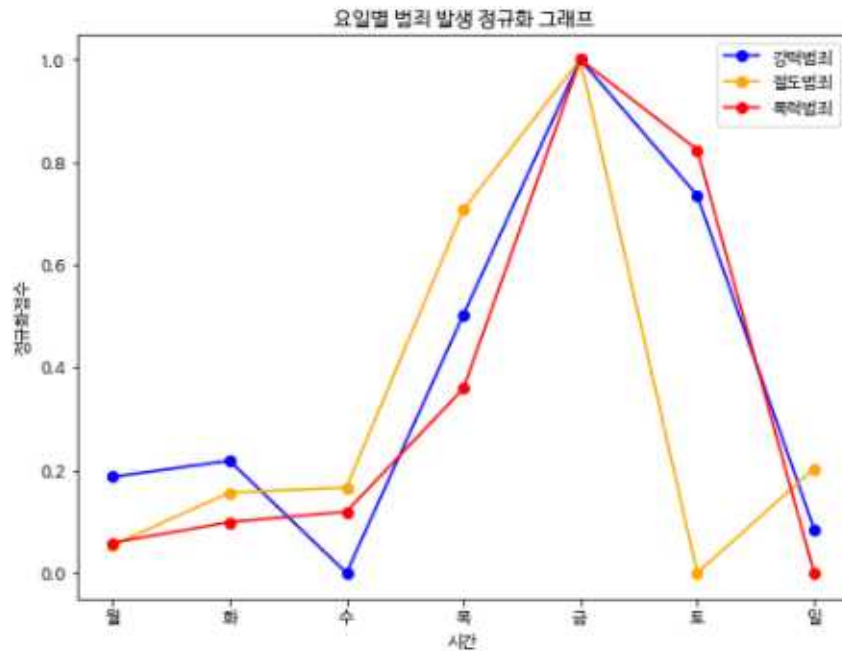
<서울시 요일 별 범죄 발생량>

- 정규화 작업 진행 후 요일 별 범죄 발생 비교.
요일 별 범죄 발생율을 한눈에 비교 할 수 있고,
유형 별 범죄 횟수를 같은 단위에 두고 비교 가능하다.

서울시 범죄 데이터

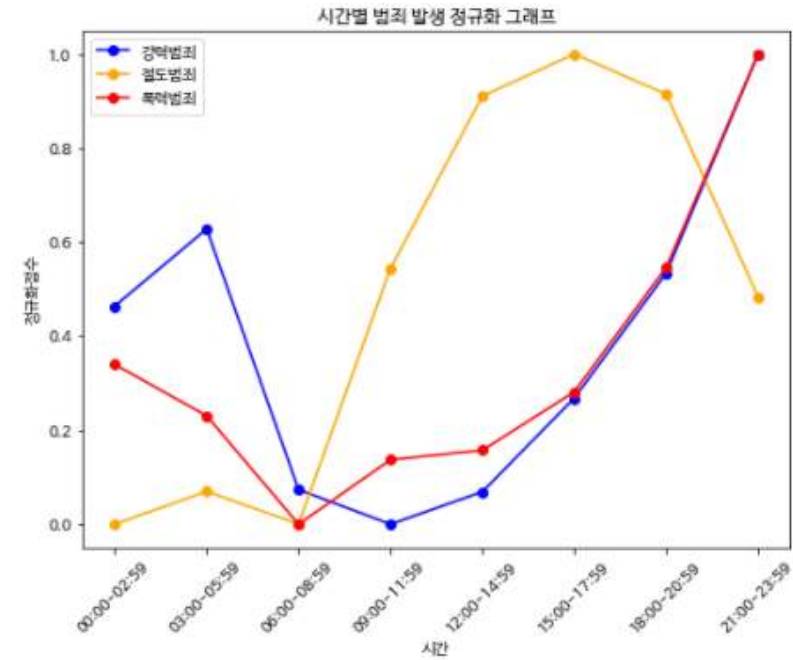
시각화 과정 - 결과

D.요일 별 가장 많이 일어나는 범죄는 무엇일까?



<서울시 요일 별 범죄 발생량>

E.시간 별 가장 많이 일어나는 범죄는 무엇일까?



<서울시 시간 별 범죄 발생량>

서울시 범죄 데이터

시각화 과정 - 전처리

<서울시 범죄 발생 주요 장소 위치>

	구분	도로명주소
0	아파트,다세대	강남구 압구정로 151
1	아파트,다세대	강남구 압구정로11길 17
2	아파트,다세대	강남구 압구정로 113
3	아파트,다세대	강남구 압구정로2길 20
4	아파트,다세대	강남구 강남대로112길 41
...
273	아파트,다세대	강남구 학동로82길 6
274	아파트,다세대	강남구 언주로103길 14
275	아파트,다세대	역삼로19길 25
276	아파트,다세대	강남구 테헤란로5길 25
277	아파트,다세대	강남구 삼성로11

	구분	도로명주소	lat	lng
0	아파트,다세대	강남구 압구정로 151	37.526025	127.024660
1	아파트,다세대	강남구 압구정로11길 17	37.525182	127.019999
2	아파트,다세대	강남구 압구정로 113	37.524113	127.018523
3	아파트,다세대	강남구 압구정로2길 20	37.520011	127.019613
4	아파트,다세대	강남구 강남대로112길 41	37.506295	127.027638

	구분	도로명주소	lat	lng
0	아파트,다세대	강남구 압구정로 151	37.526025	127.024660
1	아파트,다세대	강남구 압구정로11길 17	37.525182	127.019999
2	아파트,다세대	강남구 압구정로 113	37.524113	127.018523
3	아파트,다세대	강남구 압구정로2길 20	37.520011	127.019613
4	아파트,다세대	강남구 강남대로112길 41	37.506295	127.027638
...
553	유흥업소	서울특별시 강남구 테헤란로 435 (삼성동)	37.506595	127.054586
554	유흥업소	서울특별시 강남구 언주로93길 27 (역삼동)	37.503133	127.039911
555	유흥업소	서울특별시 강남구 테헤란로 411 (삼성동)	37.505458	127.050855
556	유흥업소	서울특별시 강남구 논현로87길 지하 41 (역삼동)	37.498788	127.033412
557	유흥업소	서울특별시 강남구 논현로87길 41 (역삼동)	37.498821	127.033535

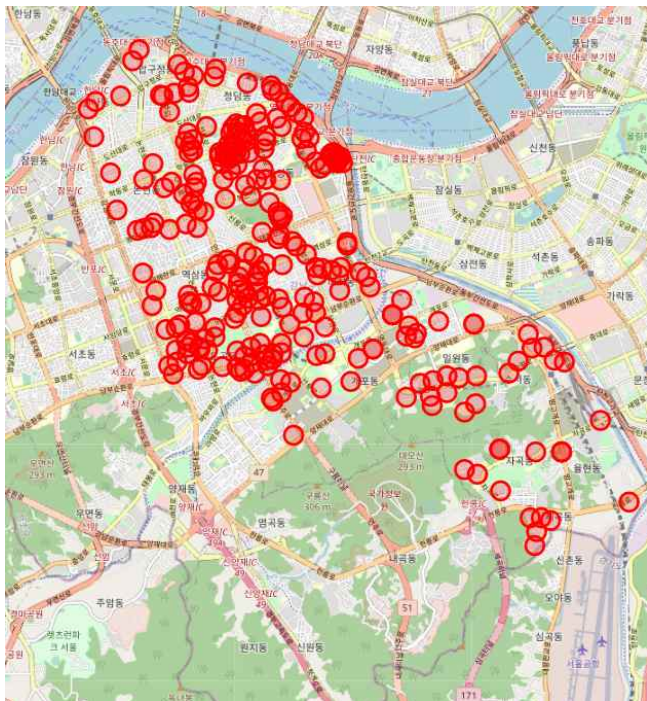
- 각 장소 데이터에 대해 도로명 주소를 갖는 컬럼만을 사용

- Google maps API 를 통해 위도,경도 컬럼을 추가

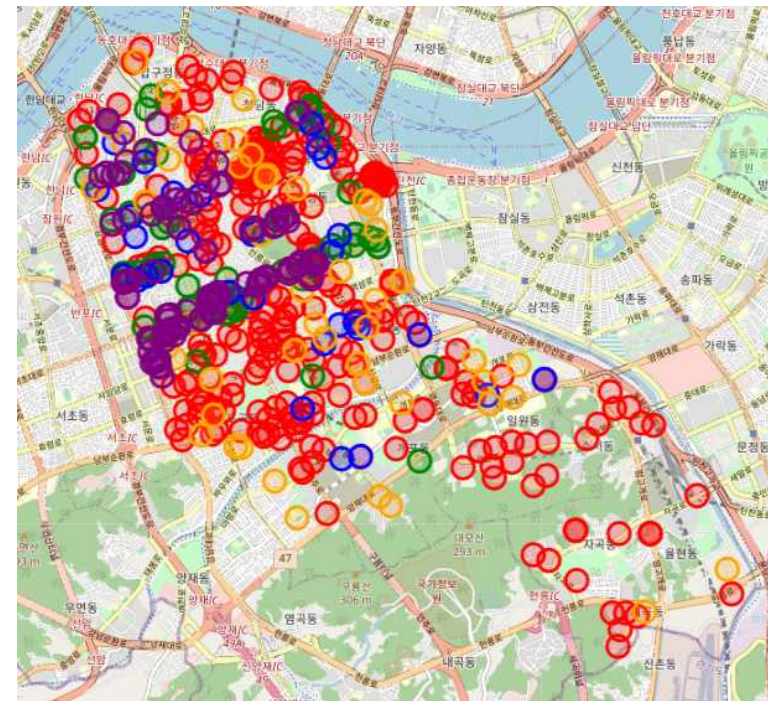
- 앞서 언급했던 5가지 특징적인 장소에 대해서 같은 작업 반복 후 데이터프레임을 통합

서울시 범죄 데이터 시각화 과정 - 결과

F. 관할구 내 지역 별로 어떤 특징이 있을까?



<강남구 내 아파트 위치 마커>

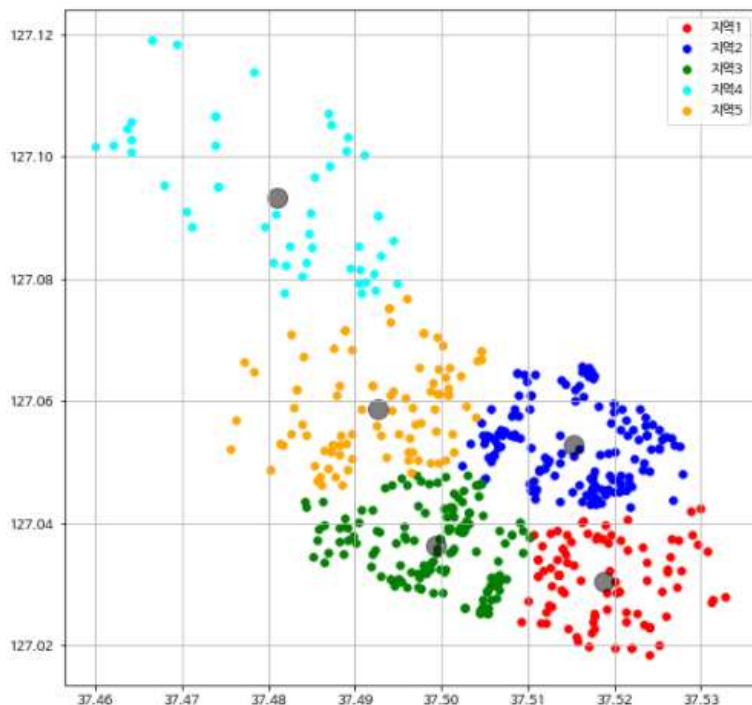


<5가지 장소에 대한 위치 마커>

서울시 범죄 데이터

시각화 과정 - 결과

F. 관할구 내 지역 별로 어떤 특징이 있을까? - 군집화

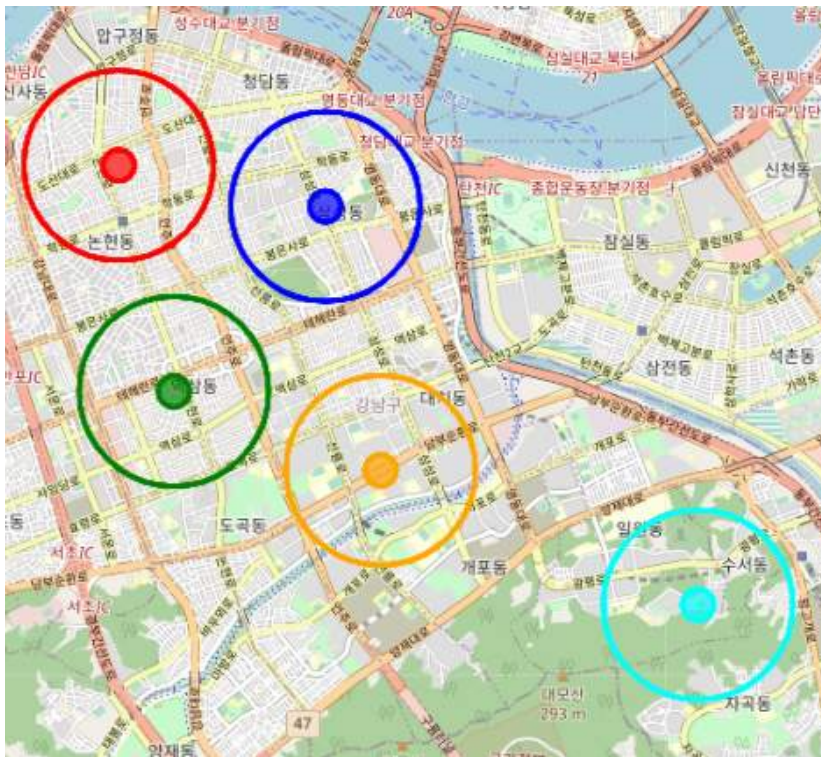


- 위치 마커 만으론 지역 별 특징을 알기 어려움
- 임의의 중심점 5 곳을 설정
- 각 중심점 과 마커 간 거리에 따라 5개의 그룹으로 분리
- 각 그룹의 특징 분석

서울시 범죄 데이터

시각화 과정 - 결과

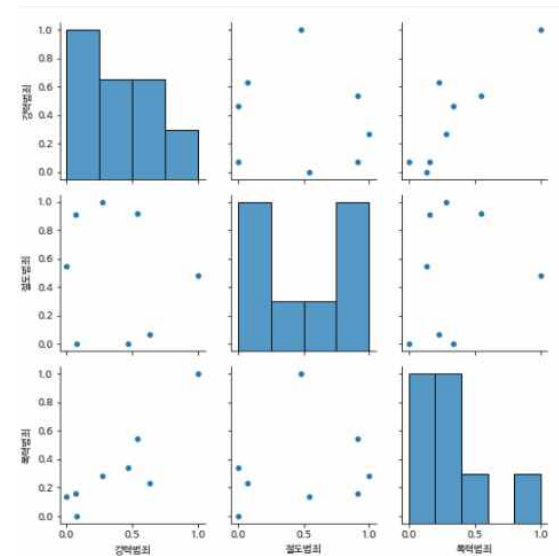
F. 관할구 내 지역 별로 어떤 특징이 있을까? - 군집화



clustering	구분	개수
지역1	아파트,다세대	35
	유흥업소	32
	숙박업소	21
	목욕탕	9
	주차장	8
지역2	아파트,다세대	80
	숙박업소	26
	유흥업소	24
	주차장	14
	목욕탕	7
지역3	아파트,다세대	54
	유흥업소	54
	숙박업소	35
	목욕탕	8
	주차장	8
지역4	아파트,다세대	42
	주차장	7
	목욕탕	2
지역5	아파트,다세대	67
	주차장	14
	목욕탕	8
	숙박업소	3

서울시 범죄 데이터

시각화 과정 - 기타



<각 유형별 범죄와의 상관관계>

영 아니올시다....?

서울시 범죄 데이터

시각화 과정 - 최종 결론

분석 목적 - 범죄가 많이 일어나는 행정 구의 관서가 범죄를 예방하기 위해 주목해야 할 장소와 시간은?

예) 강남구는 금요일 늦은 밤에 어느 지역을 중심으로 순찰을 돌아야 할까?

- 1) 금요일의 범죄 발생율은 전체 요일 중 가장 높다
- 2) 21시~00시 사이에 강력범죄, 폭력범죄의 발생율이 가장 높다
- 3) 강력범죄와 폭력범죄 발생의 특징적인 장소는 유흥업소, 숙박업소, 목욕탕이다
- 4) 지역 3에는 유흥업소, 숙박업소, 목욕탕이 많다
- 5) 그 뒤를 잇는 지역은 지역 1이다

 안전한 '불금' 을 즐기기 위해선 역삼동과 논현동 일대를 집중단속 하는 것이 좋다

04
고찰

프로젝트 고찰

- 단순 통계 데이터의 사용
- 시간, 장소의 상관관계를 연결하지 못함
- 범죄 유형 별 상관관계는 유의미한 정보가 아니었음
- 뚜렷한 특징을 보여주는 분류가 아님
- 장소, 시각의 관계를 분명하게 나타내는 알고리즘을 구현하고자 함

참고 문헌 및 사이트

- 구글 API 및 지도 마커 사용법: <https://velog.io/@konh2e/Googlemaps-Geocoding-API-%EC%82%AC%EC%9A%A9%EB%B2%95>
- K-Means clustering - 파이썬 데이터 사이언스 핸드북 (제이크 벤더플래스, 위키북스)
- 서울시 5대 범죄 현황 분석 - <https://truman.tistory.com/210>

데이터분석시각화 프로젝트

감사합니다 Q&A

최종발표