

# 데이터분석 및 시각화 프로젝트 설명서

2017112560 서준혁

2017112508 최관호

주제: 서울시 범죄 발생 현황/ 유형조사를 통한 (특정 지역) 경찰서의 경력배치, 순찰 경로 지정

## (1). 데이터 설명

서울시 구 별 범죄 발생 현황, 분석범죄 발생 유형 분석을 통해 범죄 예방 필요 지역 선정을 하기위해 범죄 발생 통계 데이터와 범죄 장소에 대한 데이터를 수집하여 사용함

| 범주 |        | 데이터 |                | 출처  | 파일명               |
|----|--------|-----|----------------|---|-------------------|
| 1  | 범죄 발생량 | 1-1 | 관서별 범죄발생량      | <a href="https://data.seoul.go.kr/dataList/316/C/2/datasetView.do">https://data.seoul.go.kr/dataList/316/C/2/datasetView.do</a> | seoul_crime.csv   |
|    |        | 1-2 | 장소별 범죄발생량      | <a href="https://www.data.go.kr/data/3074463/fileData.do">https://www.data.go.kr/data/3074463/fileData.do</a>                   | crime_area.csv    |
|    |        | 1-3 | 시간 및 요일별 범죄발생량 | <a href="https://www.data.go.kr/data/3074459/fileData.do">https://www.data.go.kr/data/3074459/fileData.do</a>                   | crime_time.csv    |
| 2  | 범죄 장소  | 2-1 | 장소1_아파트        | <a href="http://data.gangnam.go.kr/index.jsp">http://data.gangnam.go.kr/index.jsp</a>   | apt.csv           |
|    |        | 2-2 | 장소2_숙박업소       | <a href="http://data.gangnam.go.kr/index.jsp">http://data.gangnam.go.kr/index.jsp</a>   | accommodation.csv |
|    |        | 2-3 | 장소3_목욕탕        | <a href="http://data.gangnam.go.kr/index.jsp">http://data.gangnam.go.kr/index.jsp</a>   | public_bath.csv   |
|    |        | 2-4 | 장소4_주차장        | <a href="http://data.gangnam.go.kr/index.jsp">http://data.gangnam.go.kr/index.jsp</a>   | parking.csv       |
|    |        | 2-5 | 장소5_유흥업소       | <a href="http://data.gangnam.go.kr/index.jsp">http://data.gangnam.go.kr/index.jsp</a>   | salon.csv         |

## 1.범죄 발생량

서울시의 관서별/장소별/시간 및 요일별 범죄 발생량 데이터를 사용함.

### 1-1. 관서 별 범죄 발생량

-관서별 범죄 발생량 데이터에는 총 4개의 변수, 310행으로 구성 되어있음.



→요일 데이터프레임 생성 및 MinMaxScaler()를 활용한 정규화 작업 진행

→시간 데이터프레임 생성 및 MinMaxScaler()를 활용한 정규화 작업 진행

## 2.범죄 장소 데이터

-강남구의 아파트/ 숙박업소/ 목욕탕/ 주차장/ 유흥업소 도로명주소 데이터를 사용

→ 위 5가지 장소가 범죄 발생의 특징적인 장소라고 판단하였음

-각 장소 데이터는 도로명 주소를 갖는 컬럼만을 이용

→ google maps API를 통해 위도, 경도 컬럼을 추가

| 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 |

→ 5가지 장소에 대해서 같은 작업 진행

→ concat()을 활용해 위도, 경도 데이터 통합하여 사용

## (2).EDA 목표

서울시 범죄 발생량 데이터와 장소 데이터에 대하여 아래의 주제로 탐색적 자료분석을 하고자 함.

-관할구별/장소별/시간별/요일별 발생하는 범죄의 유형

-서울시에서 범죄가 가장 많이 발생하는 관할구

-범죄가 발생하는 주요 장소/범죄 유형

-범죄가 발생하는 주요 시간 /범죄 유형

-범죄가 발생하는 주요 요일/범죄 유형

-관서에서 주목해야 할 범죄 예방 장소

-기타 범죄 유형과 시간, 요일 등의 상관관계

## (3).통찰

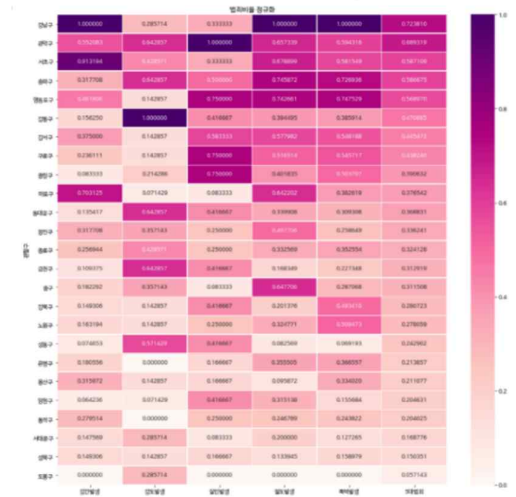
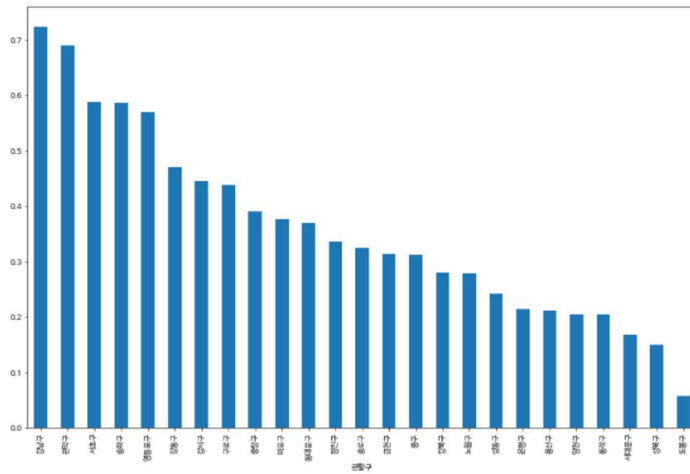
### 1.관할구별 5대 범죄 발생량

-관할구별 5대 범죄 발생현황은 아래와 같으며 다음과 같은 특징이 있음.

-서울시에서 5대범죄가 많이 발생하는 관할구는 [아래 그림]

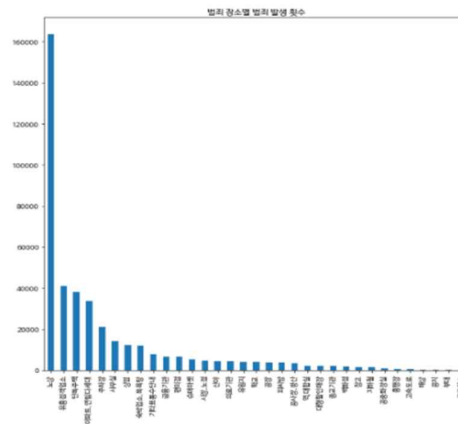
→ 강남구, 관악구, 서초구... 순으로 5대 범죄가 많이 발생

→ 강남구에서 강간, 절도, 폭력범죄가 가장 많이 발생



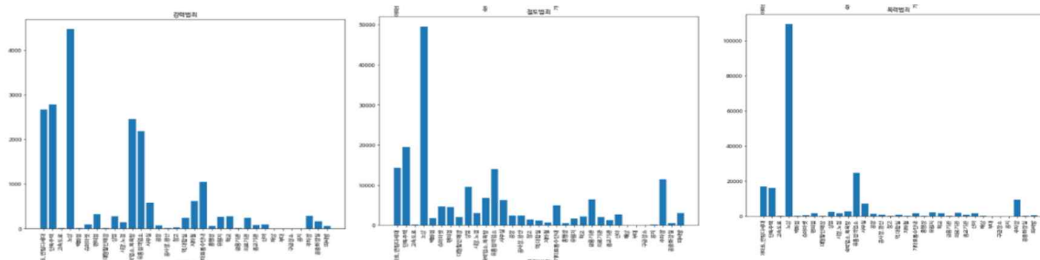
## 2. 범죄 장소별 범죄 발생량

- 서울시에서 범죄가 많이 일어나는 장소는 [아래 그림]
- 노상에서 범죄가 가장 많이 발생하며
- 그 뒤로 유흥업소>단독주택>아파트,연립다세대>주차장>사무실>상점>숙박업소>목욕탕 순임



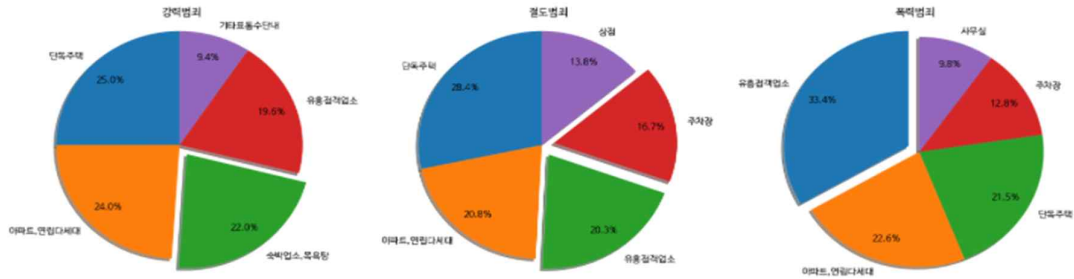
## 3. 범죄유형별 장소 분석

- 강력범죄는 노상>단독주택>아파트,연립다세대>숙박업소,목욕탕>유흥접객업소 순임
- 절도범죄는 노상>단독주택>아파트,연립다세대>유흥접객업소>주차장 순임
- 폭력범죄는 노상>유흥접객업소>아파트,연립다세대>단독주택>주차장 순임



-노상, 단독주택, 아파트 연립다세대를 제외하고 **특징적인 장소**를 분석하면

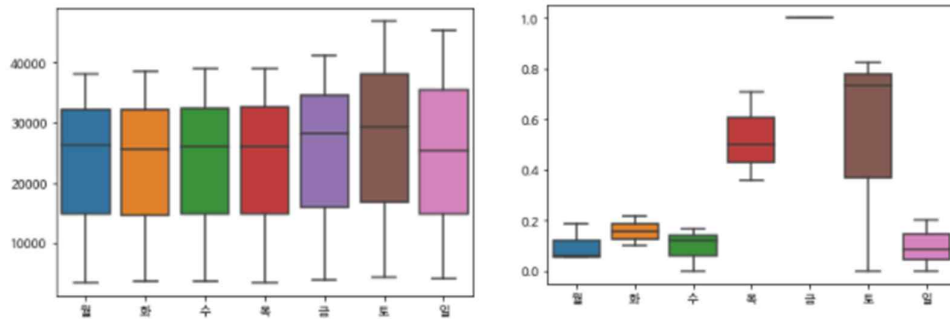
-각각 숙박업소,목욕탕/ 유흥접객업소,주차장/ 유흥접객업소에서 범죄가 많이 발생



#### 4. 요일별 범죄 발생량

-발생횟수 boxplot과 정규화 후 그린 boxplot

-정규화 한 후 범죄 발생량 비교가 편리함



-상세한 비교를 위해 lineplot을 이용

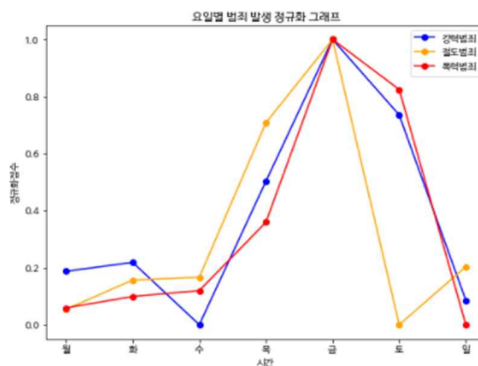
-범죄 유형별 line의 색을 다르게 표현

-수요일에서 금요일까지 범죄 발생이 증가하는 추이

-금요일에서 일요일까지 범죄 발생이 감소하는 추이

-모든 범죄가 금요일에 가장 많이 발생

-절도범죄는 토요일에 가장 적게 발생

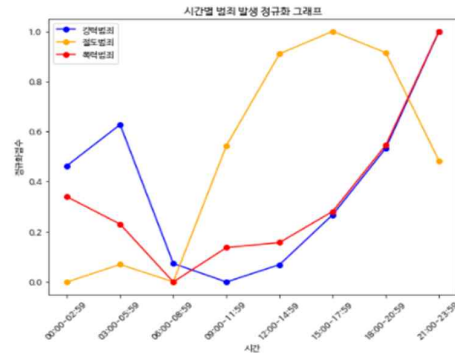


#### 5. 시간대별 범죄 발생량

-시간대별 범죄 발생 추이를 확인하기 위해 lineplot 을 이용

-절도범죄는 새벽에 발생량이 적으며 06:00시부터 18:00시까지 증가하고 24:00로가며 감소하는 추이

- 폭력범죄는 00:00시에서 08:00시까지 감소했다가 24:00로가며 계속 증가하는 추이
- 강력범죄는 00:00~06:00에 발생량이 있으며 12:00시까지 감소했다가 24:00까지 계속 증가하는 추이
- 모든 범죄가 03:00~09:00 구간에서 감소하며, 09:00~18:00구간에서 증가

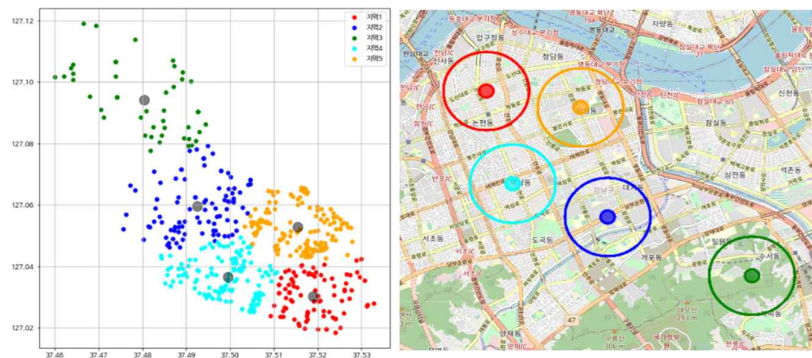


- 범죄 유형과 시간대별 범죄를 확인하기 위해 heatmap이용
- 총 범죄는 21:00~23:59>18:00~20:59>15:00~17:59 순으로 많이 발생
- 강력범죄와 폭력범죄는 21:00~23:59에 가장 많이 발생
- 절도범죄는 15:00~17:59에 가장 많이 발생



## 6. 장소 시각화 및 구역별 범죄 장소 분석

- Kmeans-clustering을 이용해 5개 구역으로 군집화
- folium모듈을 이용해 kmeans의 중심점을 지도에 시각화



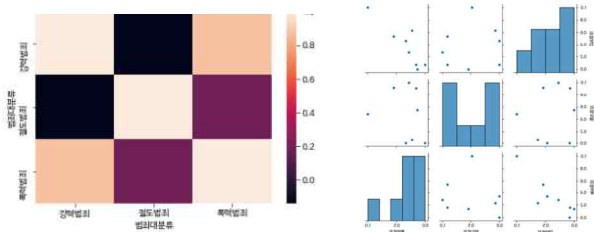
- 지도상에서 지역1,2,3,4,5는 각각 논현동, 삼성동, 역삼동, 대치동, 일원동 일대에 해당

지역1,4,5에 해당하는 논현동, 대치동, 일원동 일대에 유흥업소, 숙박업소가 비교적 많이 분포하여 5대 범죄를 더 유의해야함.

| clustering |         | 개수 |     |         |    |
|------------|---------|----|-----|---------|----|
|            |         | 구분 |     |         |    |
| 지역1        | 아파트,다세대 | 34 | 지역3 | 아파트,다세대 | 40 |
|            | 유흥업소    | 32 |     | 주차장     | 6  |
|            | 숙박업소    | 21 |     | 목욕탕     | 2  |
|            | 목욕탕     | 9  | 지역4 | 아파트,다세대 | 56 |
|            | 주차장     | 8  |     | 유흥업소    | 55 |
| 지역2        | 아파트,다세대 | 69 |     | 숙박업소    | 35 |
|            | 주차장     | 15 |     | 목욕탕     | 8  |
|            | 목욕탕     | 8  |     | 주차장     | 8  |
|            | 숙박업소    | 3  | 지역5 | 아파트,다세대 | 79 |
|            |         |    |     | 숙박업소    | 26 |
|            |         |    |     | 유흥업소    | 23 |
|            |         |    |     | 주차장     | 14 |
|            |         |    |     | 목욕탕     | 7  |

### 7. 기타

- 범죄 유형 간 상관관계가 있는지, 범죄 발생 시간과 요일에는 어떤 상관관계가 있는지 고민이 많이 되었음



→ 범죄 유형 간 상관관계를 분석한 다양한 plot. 별 의미는 찾지 못하였다.

### (4)고찰

첫째, 범죄와 범죄장소의 상관관계를 밝혀 예방을 위한 구체적인 순찰 경로를 지정하고 싶었으나 데이터의 수집 단계에서 위치, 범죄, 시간을 모두 아우르는 데이터를 찾지 못했다. 범죄 발생 통계, 범죄 장소와 관련된 데이터를 따로 수집하여 시간별, 요일별, 장소별 범죄 발생 추이와 범죄의 유형별 특징을 분석할 수 있었으나, 데이터셋이 단순한 통계 데이터였기 때문에 유의미한 상관관계를 도출하지 못한 점을 한계로 느낀다.

둘째, 범죄 발생에는 시간,장소 이외에 다양한 요소들의 내용을 담지 못했다(cctv 위치, 동네의 분위기 등). 범죄 발생의 특징을 분석에 범죄장소를 시각화 하면서 데이터를 잘 수집하면 범죄 발생 장소와 시간과의 상관관계를 정확하게 분석할 수 있다고 생각했다. 마지막으로, 강남구 내의 건물 유형분포 간 엄청난 차이가 나지 않아 지역 간 뚜렷한 특징을 찾을 수 없었다.

현재는 단순히 범죄가 많이 발생한 장소와 시간을 시각화 한 것에 불과하지만, 시각화 프로젝트가 끝난 이후에는 범죄의 다양한 요소를 고려하고 범죄 유형과 시간, 장소, 지역의 구체적인 상관관계를 나타내는 머신러닝 알고리즘을 공부하여 정말 경찰이 사용할 수 있는 유의미한 결과를 도출하는 것이 목표이다.