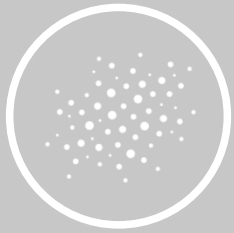


Team : DBNIS
미세먼지와 범죄율 간 상관관계 분석

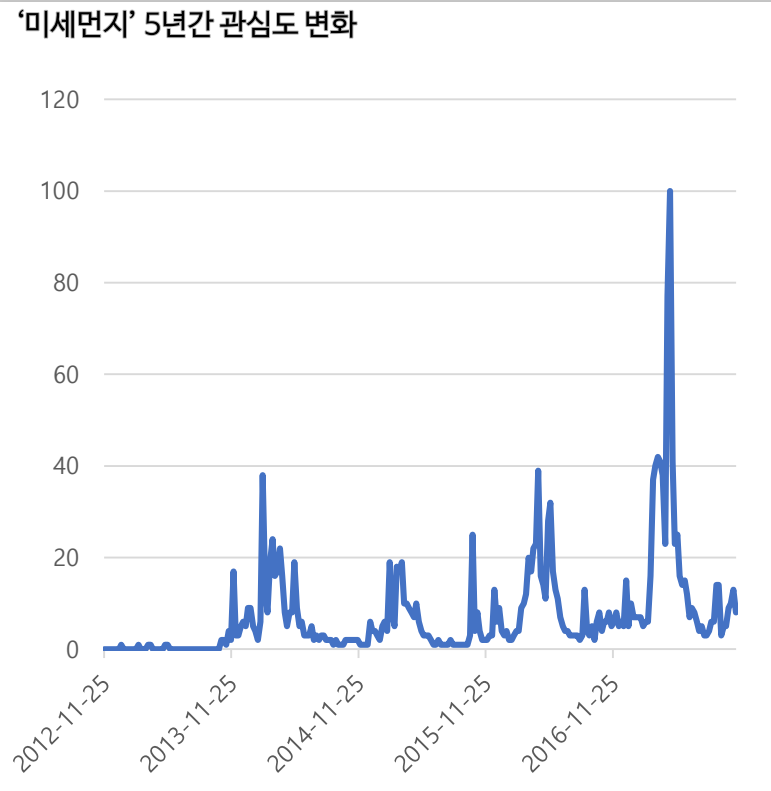


DBNIS
미세먼지와 범죄율 간 상관관계 분석

- 주제 선정
- 분석계획
- 데이터분석
- 결론 및 한계점



주제 선정



출처 : 구글트렌드

최근 관심도가 높아지고 있는 미세먼지를 주제로 선정했다.
“미세먼지가 높은 곳은 범죄율 또한 높을 것이다”라는 가설을 세우고, 이를 증명하기 위하여 분석을 진행한다.
Sensing Dataset의 sensorParser데이터를 활용하기 위해 분석 지역을 대구로 선정하였다.



RMySQL을 사용하여 sensorParcer테이블에서 lat값이 null이 아닌 데이터를 추출

대구의 인구수 데이터와 범죄율 데이터 수집 (3년간)

수집된 데이터와 sensorParcer데이터를 병합 (날짜 기준)

분석 및 결과 시각화



RMySQL을 사용하여 sensorParcer테이블에서 lat값이 null이 아닌 데이터를 추출

- > timestamp형식이 초마다 등록되어 사용기준을 정하지 못함
- > sensorParcer데이터가 아닌 16년도의 인구수, 미세먼지, 범죄율간 상관관계를 비교하는 것으로 변경

대구의 인구수 데이터와 범죄율 데이터 수집 (3년간)

- > 시간 관계 상, 3년이 아닌 1년간 데이터만을 분석
- > kosis사이트 이용, 대구시 동·읍·면/연령별 주민등록인구(내국인) 수집
- > 범죄율은 SBS가 2016년 경찰자료를 분석하여 만든 범죄지도 참고

수집된 데이터와 sensorParcer데이터를 병합 (날짜 기준)

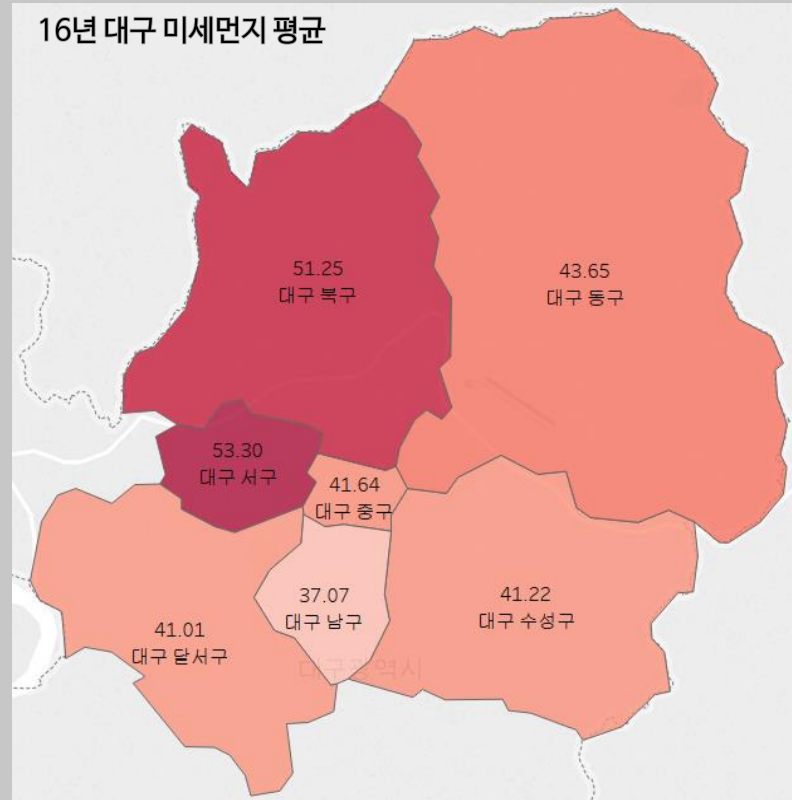
분석 및 결과 시각화



대구의 인구수 데이터와 범죄율 데이터 수집 (16년)

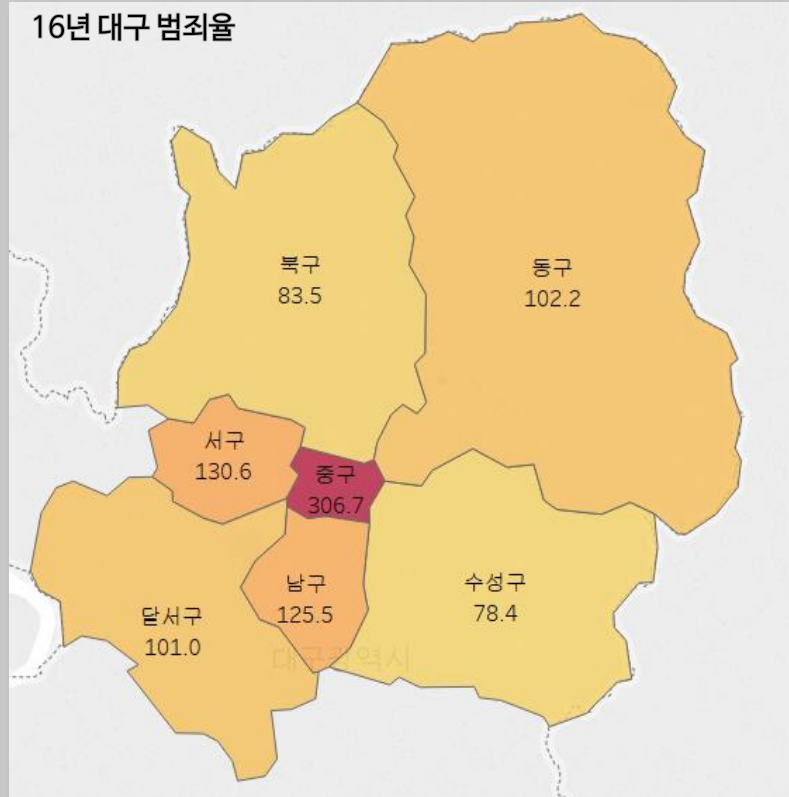
태블로를 이용하여 가설 설정 및 분석

결과 시각화



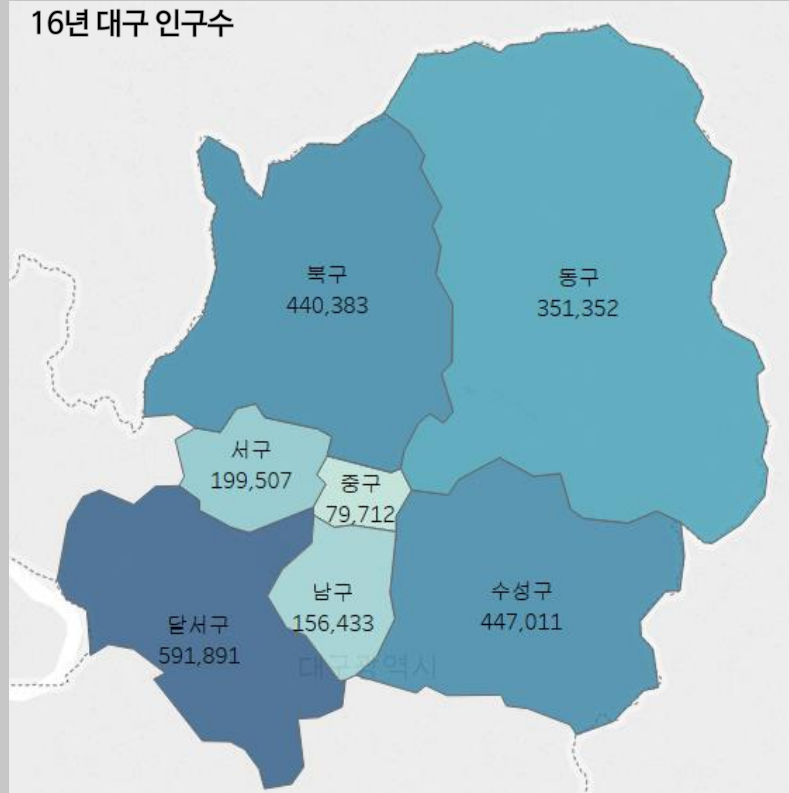
출처: KOSIS, 16년 미세먼지 (pm10) 월별 도시별 대기오염도

미세먼지의 경우 대구 서구가 가장 높았으며, 대구 남구가 가장 낮았다.
서구와 북구의 경우 공단이 위치하고 있기 때문에 미세먼지 농도가 높은 것으로 보인다.
“미세먼지가 높은 곳은 범죄율 또한 높을 것이다”라는 가설에 따라, 서구와 북구의 범죄율이 높게 나타날 것으로 예측된다.



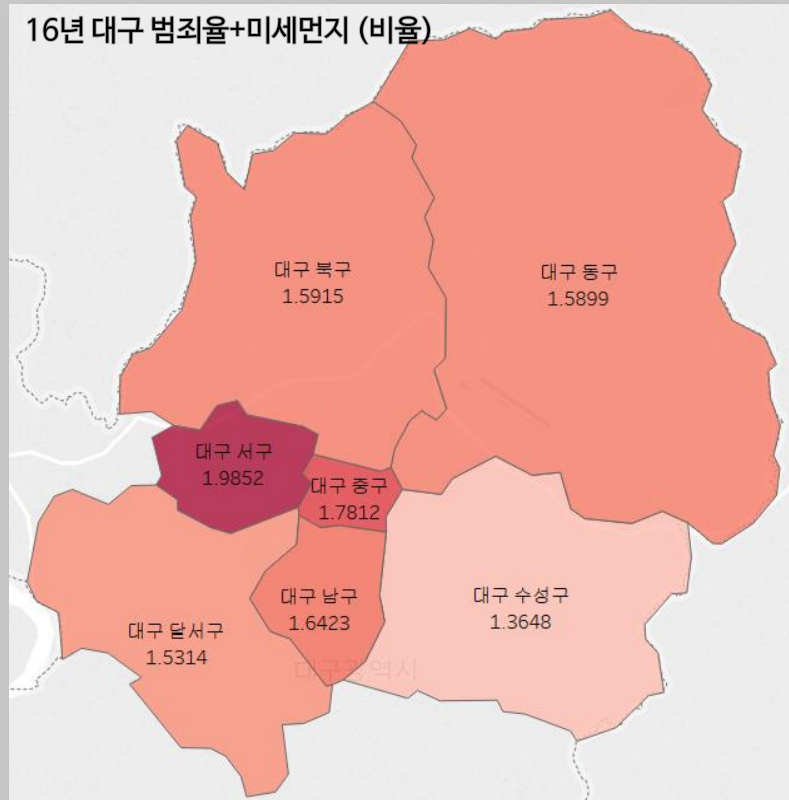
출처: 16년 범죄지도, KOSIS

범죄율의 경우 중구가 가장 높았으며 수성구가 가장 낮았다.
중구의 범죄율이 높은 원인은 변화가인 동성로가 위치한 도심 유흥지역이며, 유동인구의 수가 높기 때문으로 생각된다.
또한 Boxplot을 그려본 결과, 중구의 범죄율이 이상치로 나타났기에 이를 평균값으로 대체할 것이다.

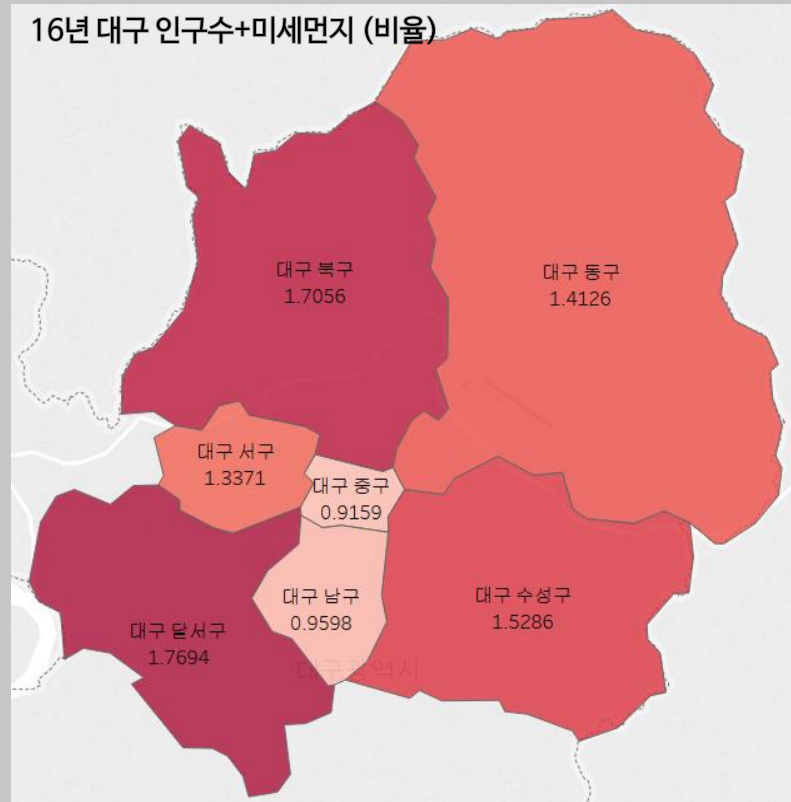


출처: KOSIS

인구수의 경우 달서구가 가장 높았으며 중구가 가장 낮았다. 달서구와 중구의 인구차이가 상당하여 이상치 검증을 위해 Boxplot을 그렸으나, 이상치가 발견되지 않았기에, 데이터 변경없이 분석을 그대로 진행했다.



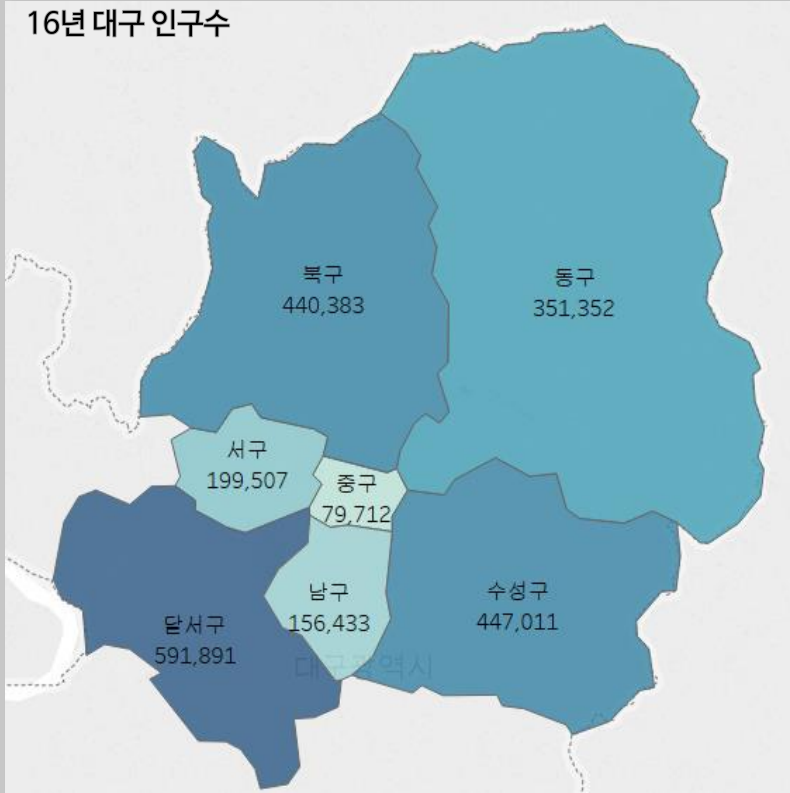
가장 큰 값을 가지는 데이터를 기준으로 항목별비율로 변경했다. 미세먼지의 경우 서구,범죄율의 경우 중구를 기준으로 잡았다. 서구와 북구의 범죄율이 높을것이다라는 예측과 달리, 서구와 중구의 범죄율이 높게나타났다.



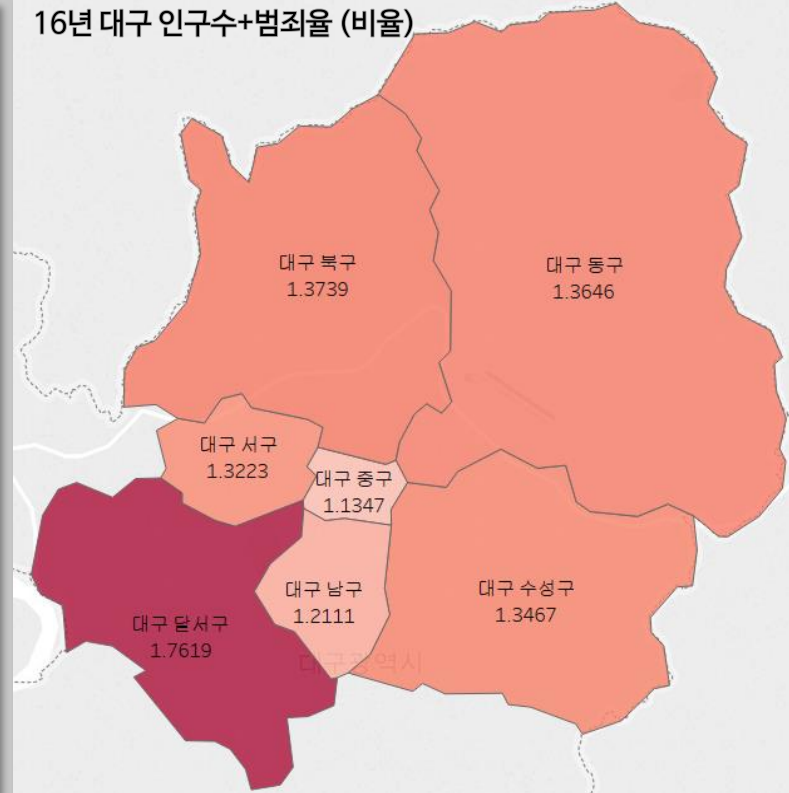
가장 큰 값을 가지는 데이터를 기준으로 항목별비율로 변경했다. 미세먼지의 경우 서구,인구수의 경우 달서구를 기준으로 잡았다.
인구수 - 미세먼지의 상관관계의 경우 북구와 달서구가 높게 나왔다.



16년 대구 인구수



16년 대구 인구수+범죄율 (비율)

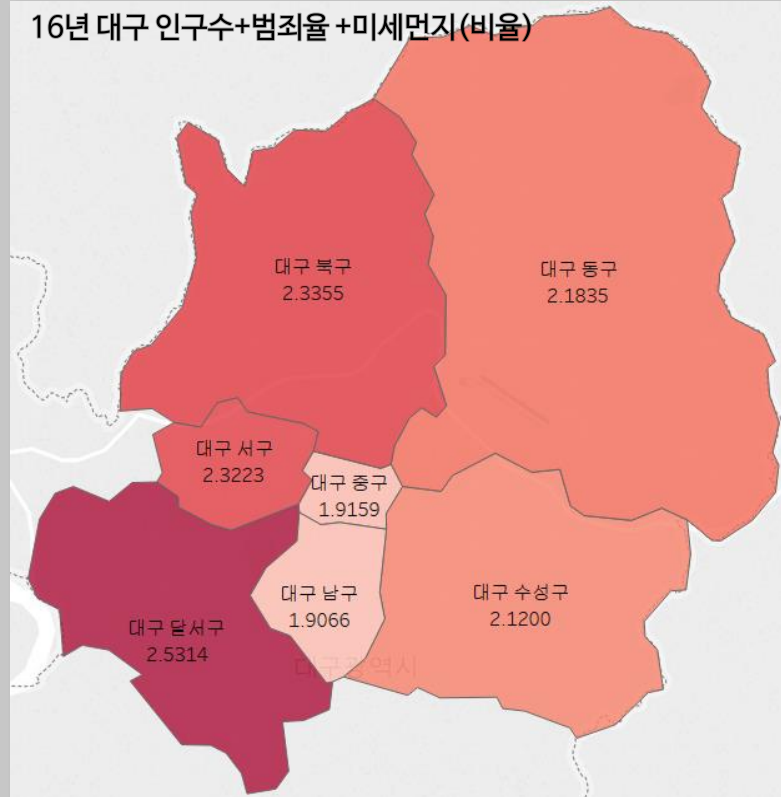


출처 : KOSIS

가장 큰 값을 가지는 데이터를 기준으로 항목별비율로 변경했다. 인구수의 경우 달서구, 범죄율의 경우 중구를 기준으로 잡았다. 인구수 - 범죄율의 비율 경우 달서구가 높게 나왔으며, 중구가 가장 낮게 나왔다. 이는 달서구와 중구의 인구수와 높은 관련이 있다. 인구수와 범죄율은 상관관계가 있음을 알 수 있다.



16년 대구 인구수+범죄율+미세먼지(비율)



인구수와 범죄율, 미세먼지의 비율을 종합적으로 평가한 결과 달서구-북구-서구-동구-수성구-중구-남구 순으로 결과가 나왔다.
달서구와 높은 인구수와 중구의 낮은 인구수 차이로 인한 결과로 생각된다.

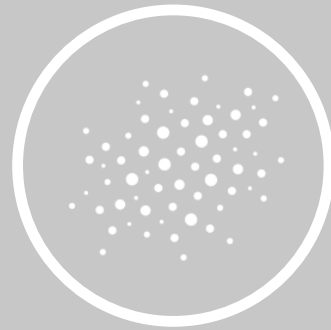


대구의 경우 “미세먼지가 높은 곳은 범죄율 또한 높을 것이다”라는 가설의 증명은 실패하였지만 범죄율은 인구밀도와 높은 상관관계가 있다는 것을 확인하였다.



하지만 인구수가 많은 달서구를 제외하면 서구와 북구의 순위가 높게 나타났기 때문에
인구수가 비슷한 지역이나 유동인구수로 비교한다면,
미세먼지와 범죄율은 양의 상관관계를 가질 수 있을 것이라 생각된다.

또한 1년치 데이터가 아니라 몇년간의 데이터를 분석한다면,
또 다른 유의미한 결과를 얻을 수 있을 것이라 생각된다.



Team : DBNIS
감사합니다