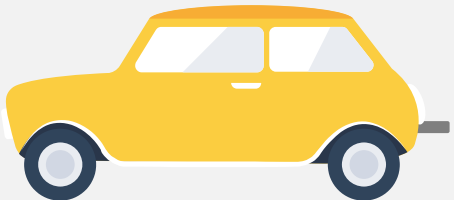


미세먼지 보고서

With bs4, Selenium, Plotly, Django....

김은배



목차

1. 활용 데이터

- 구성성분
- 미세먼지농도
- 측정소 정보
- 연도별 뉴스 수

2. 작업도구 · 과정

- 주요 작업도구
- 이외의 작업도구
- 작업과정

3. 데이터 분석

- 결측치 분석
- 각 데이터 분석

4. 평가 및 개선

- 평가
- 개선해야할 점



1장

활용 데이터



활용 데이터 설명

활용 데이터	설명
미세먼지 구성성분	환경부 측정 전국 미세먼지 구성 성분
미세먼지 농도	국립 환경과학원 측정 2015~2022 전국인구 가중평균 미세먼지 농도 & 17개 시·도 연평균 미세먼지(PM2.5)농도
측정소 위치	국가 배경 및 도시대기 측정망(소) 위치
미세먼지 관련 뉴스기사 수	2015~2023* 연도별 미세먼지 관련 뉴스기사 수

* 2023년은 1~10월까지 측정



2장

작업 도구



주요 작업 도구

Beautiful Soup

HTML 또는 XML 파일의 데이터를 추출하는 파이썬 라이브러리이다. 주로 웹 스크래핑을 할 때 사용된다.

Selenium

웹 애플리케이션 자동화 테스트를 위한 프레임워크다. 파이썬을 포함한 유명 프로그래밍 언어들에서 사용할 수 있다.

Plotly

상호작용적인 시각화가 가능한 파이썬 그래픽 라이브러리이다. 40개 이상의 차트 타입을 제공하는 오픈소스이다.

Django

파이썬으로 작성된 오픈소스 웹 프레임워크로, 웹 개발의 번거로운 요소를 새로 개발할 필요 없이 내장된 기능으로 빠른 개발을 할 수 있다.

BeautifulSoup



이외의 작업 도구

folium

Python 지도 시각화 라이브러리이다. folium을 사용하여 interactive한 지도를 생성하거나 마커 추가, 원으로 범위 등을 표시할 수 있다.

geokakao

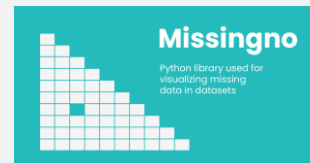
주소 정보를 포함하는 DataFrame이 주어졌을 때, 좌표 정보를 알아내어 데이터프레임에 다시 추가해주는 Kakao 지오코딩 패키지이다.

Pandas

데이터 처리와 분석을 위한 라이브러리로 행과 열로 이루어진 데이터 객체를 다룰 수 있다. 대용량의 데이터를 처리하는데 매우 편리하다.

Missingno

Pandas 데이터프레임에서 누락된 데이터를 찾고 시각화하는 기능을 제공한다.



작업 과정



3장

데이터 분석

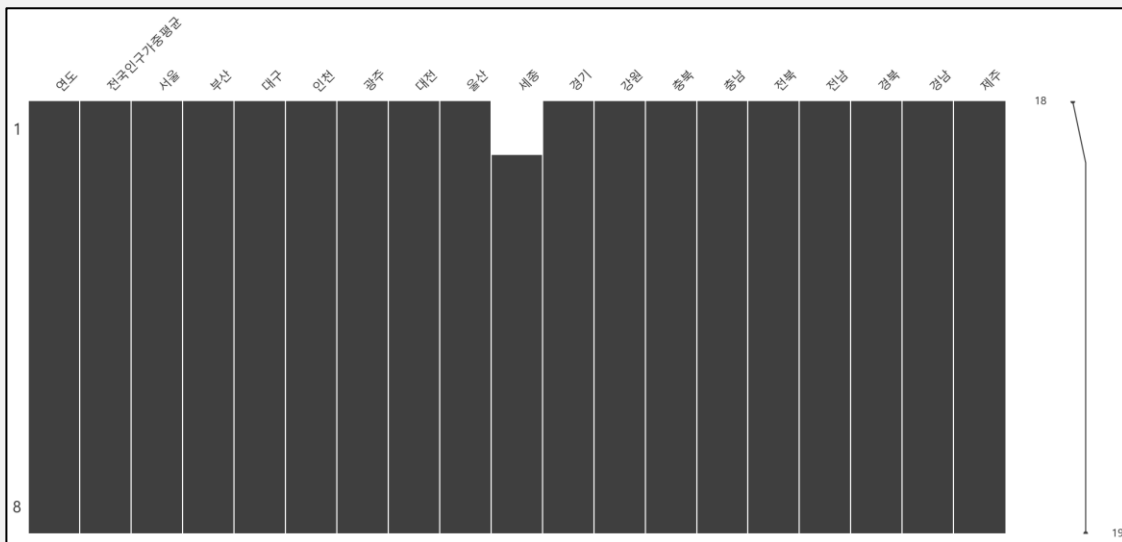


결측치 확인

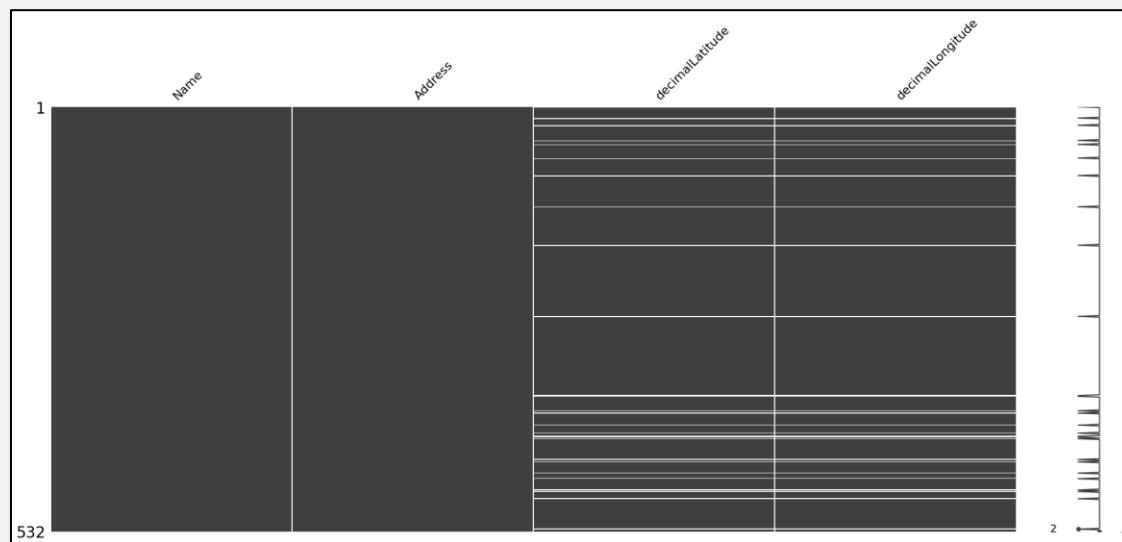
결측치(Missing Value) : 원래 값이 존재해야 할 위치에 값이 누락되어 있는 것

- 미세먼지 농도 데이터: 2015년 세종시 열 데이터 1개 누락 * → 세종시는 2016년부터 데이터 제공
- 측정소 위치 데이터: 위도 · 경도 열 각각 27개 누락

* missingno 라이브러리 사용



미세먼지 농도 데이터



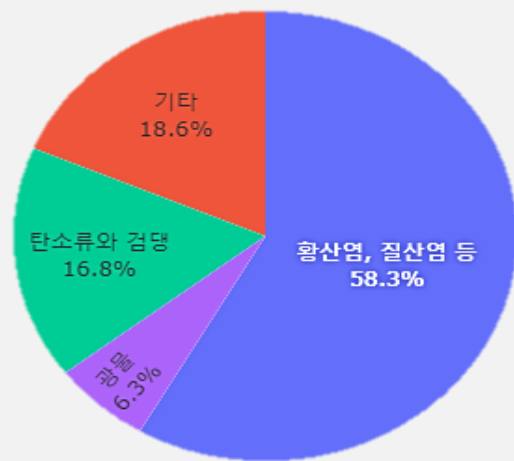
측정소 위치 데이터



데이터 분석- 미세먼지 구성성분

미세먼지의 성분: 직경에 따라 PM-10, PM-2.5(초미세먼지) 등으로 구분됨

- 자연·인위적으로 생성되나, 주로 문제되는 것은 인위적인 미세먼지
- 대기오염물질이 공기 중에서 반응하여 형성된 덩어리(황산염, 질산염 등), 석탄·석유 등 화석연료를 태우는 과정에서 발생하는 탄소류와 검댕, 지표면 흙먼지 등에서 생기는 광물로 구성
- 전국에서 측정된 미세먼지(PM2.5)의 구성 비율을 보면 황산염, 질산염이 50% 이상으로 가장 높고, 탄소류와 검댕이 20% 미만, 광물 5% 이상으로 구성



미세먼지 구성성분 파이 그래프

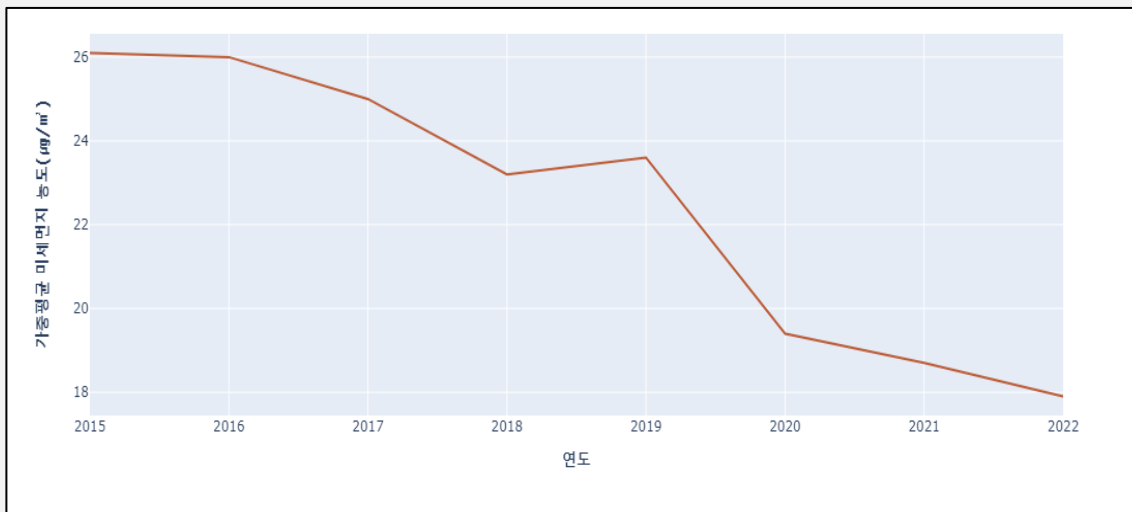


데이터 분석- 미세먼지 농도(PM2.5)

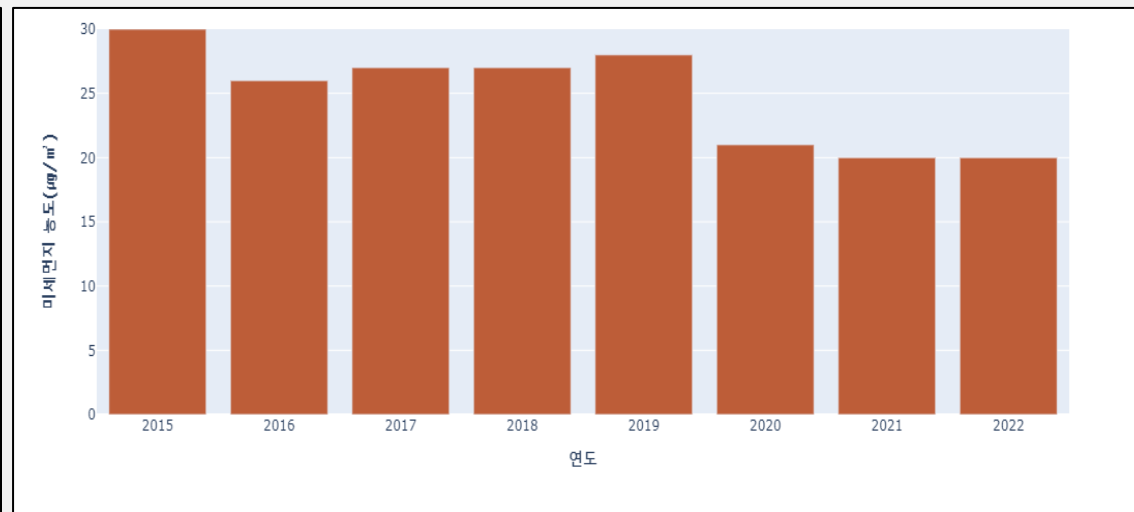
미세먼지 농도: 2015~2022 전국인구 가중평균 미세먼지 & 농도 17개 시 · 도 연평균 미세먼지 농도

- 17개 시·도의 미세먼지 농도를 인구 가중평균하여 합산한 값은 2022년 $17.9\mu\text{g}/\text{m}^3$ 으로 2015년 이후 가장 낮은 수치를 보임.
- 전국 대부분 지역에서 미세먼지 농도가 2020년보다 감소.
- 2022년 지역별로 서울, 인천, 경기, 충북, 충남, 지역의 미세먼지 농도가 높고, 부산, 강원, 전남, 경남, 제주 지역이 $15\mu\text{g}/\text{m}^3$ 이하로 낮은편.

* 세종시는 2015년 이후부터 미세먼지 농도 제공



2015~2022 전국인구 가중평균 미세먼지 농도



2015~2022 17개 시 · 도 연평균 미세먼지 농도 (충북)



데이터 분석- 측정망(소) 위치

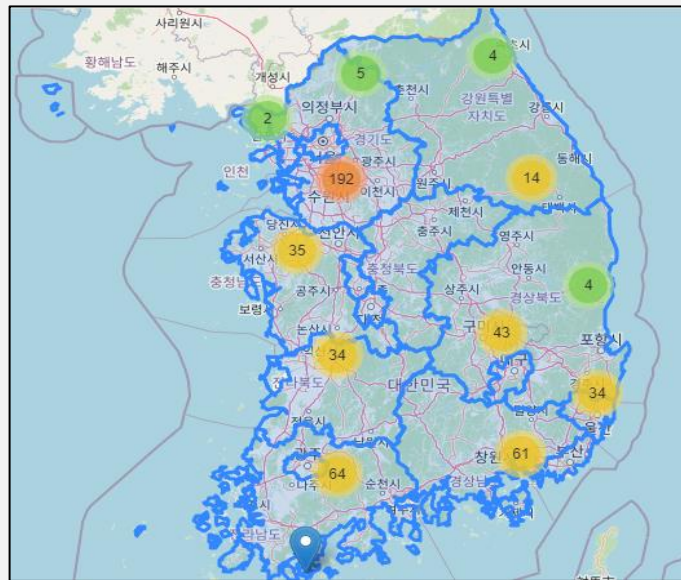
측정망(소) 위치: 국가배경 농도 측정망 & 도시대기 측정망 데이터

- 국가배경: 국가의 배경농도를 측정하고 외국으로의 오염물질 유입, 유출상태를 파악하기 위한 측정소. 총 11개 (7개 시 · 군)
- 도시대기: 도심(거주)지역의 평균 대기질 농도 파악을 위한 측정소, 총 532개 (지도 상에 505개, 누락 데이터 27개)

* geokakao 라이브러리 사용



국가배경 농도 측정망(소)



도시 대기 측정망(소)

측정소	위치
3공단	경북 포항시 남구 대송면 철강산단로130번길 29 3공단 배수지
강진읍	전라남도 강진군 강진읍 동성로 72 강진읍 측정소
고덕동	경기도 팔택시 고덕면 고덕국제2로 111 종덕초등학교 2층 옥상
과천동	경기 과천시 상하별로 17 과천시 환경사업소
광명동(폐쇄)	경기 광명시 광명3동 광명3동 주민센터
금천구	서울 금천구 금하로21길 20 시흥5동 주민센터
녹산동	부산 강서구 녹산산단 382로 49번길 39 부산환경공단 녹산사업소
독곶리	충남 서산시 대산을 평산1로(독곶리) 한국수자원공사 대산산업용수센터 저수동 옥상
북정동	경기 성남시 수정구 성남대로 1416번길 22 북정정수장
송정1동(폐쇄)	광주 광산구 광산로 70 송정1동 주민센터
웅지동	경남 창원시 의창구 웅지로 239번길 19-4 (웅지동 주민센터)
웅호동	부산광역시 남구 이기대공원로 11 부산환경공단 남부사업소 북쪽 (웅호동)
월내동	전남 여수시 여수산단로 1201(월내동) 여수산단 월내 폐수종말처리장
유촌동	광주광역시 서구 천변우하로 203 광주보건환경연구원 신청사 옥상 (유촌동)
인제읍	강원특별자치도 인제군 인제읍 비룡로44번길 93 인제군기상관측소
장유동	경남 김해시 장유동 능동로 149 (장유건강지원센터)
장호원읍	경기 이천시 장호원읍 오날리 27 재래시장 공영주차장
장흥동	경북 포항시 남구 철강로 331 공단정수장 공단정수장
재송동	부산광역시 해운대구 센텀동로 191 동부하수처리장 (재송동)
중방동	경북 경산시 중방동 708-5 남매로 158 경산시보건소 정신보건센터 2층
중앙로	강원특별자치도 춘천시 중앙로길 135(중앙로 3가 67-1) 춘천시보건소 3층 옥상

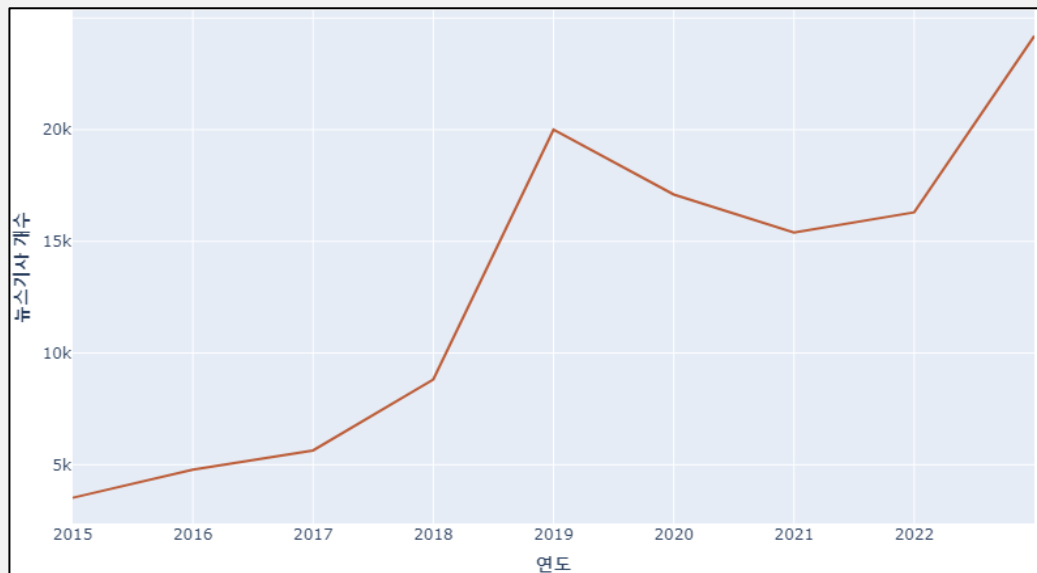
도시 대기 측정망 누락 데이터 테이블



데이터 분석- 연도별 미세먼지 관련 뉴스기사 수

연도별 미세먼지 관련 뉴스기사 수 : 2015 ~ 2023 연도별 미세먼지 관련 뉴스기사 수 (2023년은 10월까지)

- 2015년 이후, 미세먼지 관련 뉴스 수가 지속적으로 증가하다, 2020년 큰 폭으로 감소
- 이는 코로나로 인한 사회 경제적 활동의 감소, 정부 정책, 양호한 기상상황 등의 복합적 요인에 따른 미세먼지 감소*가 그 이유로 보임
- 하지만, 2021년 이후 다시 국내 코로나 규제의 완화, 외부요인(해외 공장 가동 재개)*에 대한 우려가 심화돼 뉴스기사 수가 증가한 것으로 예상
- 2023년 10월 현재 이미 2019년 수치를 이미 능가



연도별 미세먼지 관련 뉴스기사(2015~2023)

참조

- * 환경부, 자주하는 질문
- * “코로나블루” 사라지고 잿빛 도시된 중국.. 그 먼지 한국 온다.”
- * “코로나19 규제 완화에 세종시 미세먼지 오염 다시 심해지나”



4장

평가 및 개선

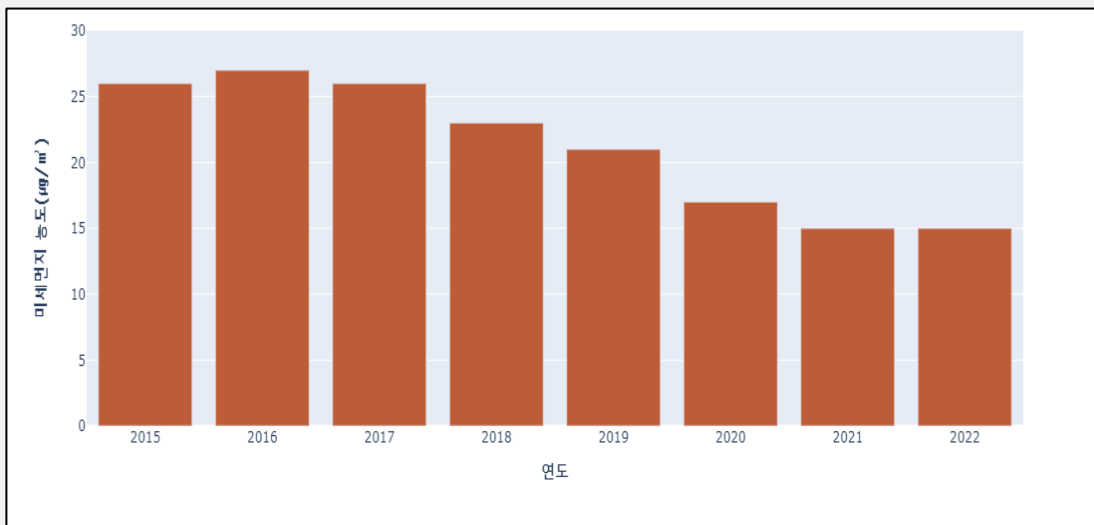


평가 - 부산 · 충청권 미세먼지 농도 비교

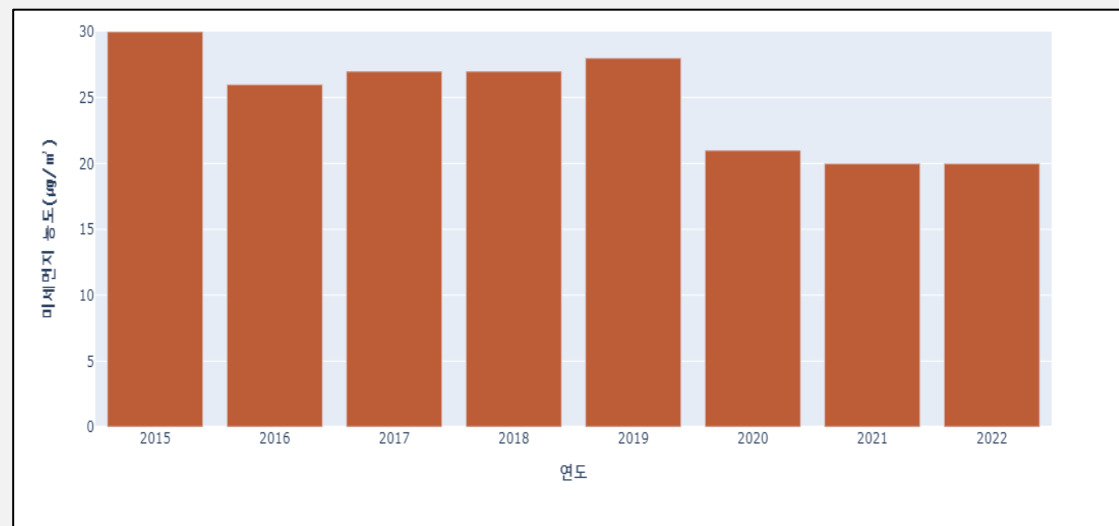
2015~2022 부산 · 충청권 미세먼지 농도 비교

- 거의 매해 충청권인 충북의 미세먼지 농도가 부산, 강원, 전남, 경남, 제주 지역에 비해 높게 관측됨
- 충청권의 경우 중국 북동부에서 발원해 북서풍을 타고 날아온 황사 등이 차령산맥에 가로막혀 초미세먼지 농도가 높아지는 것으로 분석됨
- 또한 초미세먼지를 발생시키는 항만, 석탄화력발전소 등이 위치하고 있으며, 특히 전국 59기의 석탄화력 발전소 중 29기가 충남 서해안 지역에 집중됨*.

* “초미세먼지 짙은 충청권... 농도 전국 최상위권, 충청투데이”



2015~2022 도시별 미세먼지 농도(부산)



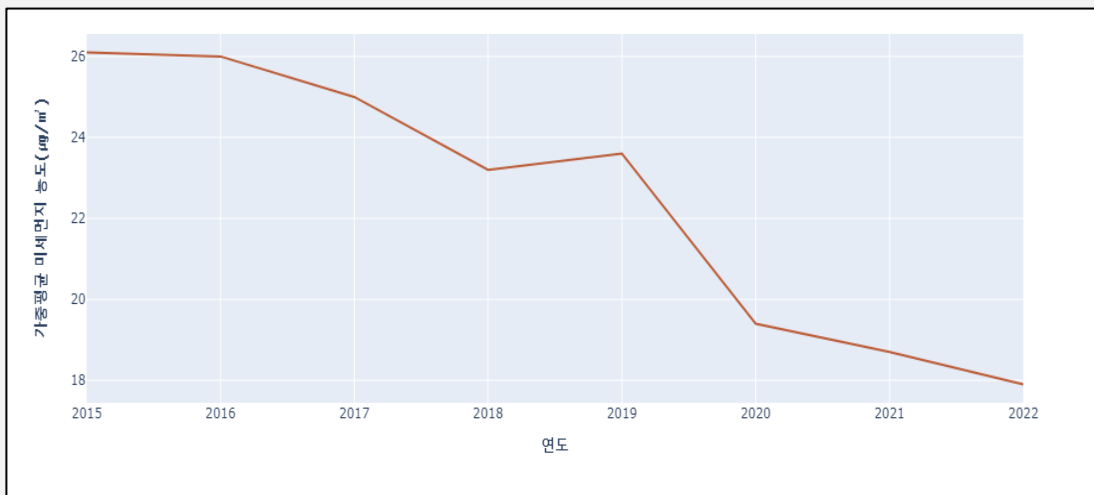
2015~2022 도시별 미세먼지 농도(충북)



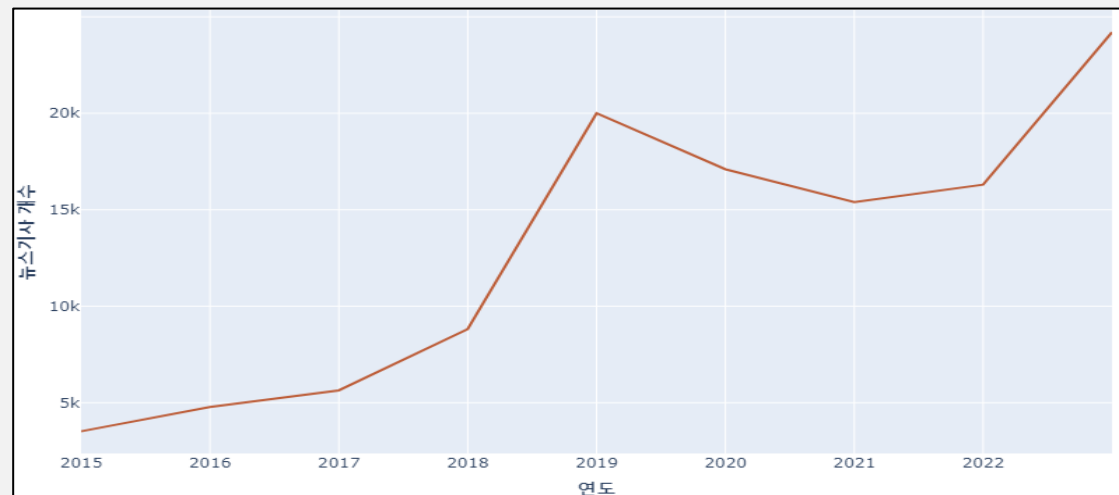
평가 - 가중평균 미세먼지 농도 & 뉴스기사 수 비교

전국인구 가중평균 미세먼지 농도 & 연도별 미세먼지 관련 뉴스기사 수 그래프 비교

- 실제 미세먼지 농도는 2019년 이후 지속적으로 감소함에도 연도별 미세먼지 관련 뉴스기사의 수는 2021년 이후 계속해서 증가하는 추세를 보임
- 2011년 초미세먼지 대기환경 기준 개정 후, 2015년 초미세먼지 대기환경 기준 시행(서울은 2002년 전국적으로는 2015년).
- 실제 수치는 감소 추세에 있음에도 초미세먼지 관측의 수치화 및 각종 규제 완화로 그 결과에 상관없이 언론의 우려 또는 관심이 증가한 것으로 보임



2015~2022 전국인구 가중평균 미세먼지 농도



연도별 미세먼지 관련 뉴스기사 수 (2015~2023)



평가 – 개선해야 할 점

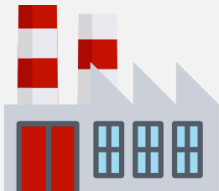
개선해야할 점: 결측치 발생 데이터 처리

1. 세종시 2015년 미세먼지 농도에 결측치 발생

- 세종시는 미세먼지 농도(PM2.5)를 2016년부터 제공
- 결측치를 잘못 대체할 시 데이터에 편향이 발생
- 결측치는 내버려두기, 제거, 중앙값, 평균값, 최빈값으로 대체 등의 방법 존재

2. geokakao 라이브러리 사용시 측정소 위도 · 경도 추출이 안되는 결측치 발생

- Google · Naver Geocoding API 사용
- 주소 내용 전처리



참조

- 환경부와 친해지구, *미세먼지, 너는 누구니? 미세먼지의 모든 것*, 2018.6.19. 11:11, <https://blog.naver.com/mesns/221302000466>
- *미세먼지 농도*, 국가지표체계, 2023-04-04, <https://www.index.go.kr/unify/idx-info.do?idxCd=8058>
- *측정소 정보*, 에어코리아, n.d. 2023-10-15 검색, https://www.airkorea.or.kr/web/stationInfo?pMENU_NO=93
- 환경부, “올해 미세먼지가 획기적으로 개선된 것은 코로나19의 영향이 아닌지?”, 2020-11-27, https://www.airkorea.or.kr/portal/web/board/12/1652/?pMENU_NO=144&page=2
- 천권필, “‘코로나 블루’ 사라지고 잿빛도시 된 중국 ... 그 먼지 한국 온다.”, 2023.03.07 17:00, <https://www.joongang.co.kr/article/25145331#home>
- 최준호, “코로나19 규제 완화에 세종시 먼지오염 다시 심해지나”, 2020.10.28 13:36:39, <https://www.inews365.com/mobile/article.html?no=641734>
- 김성준, “초미세먼지 짙은 충청권... 농도 전국 최상위권”, 2022.05.11, <https://www.cctoday.co.kr/news/articleView.html?idxno=2161852>



Q & A

미세먼지 보고서

<https://tinyurl.com/45kh7ne2>

