

빅데이터를이용한통계그래픽스

TEAM PROJECT

[자료 분석 방향]

4조

김송희 양보연 이하은 정유진 주선미 최희원

Data Analysis

Introduction

1. 변수 파악 및 정리(2p, 4p [첨부])
 - 기본적인 통계량 계산, 범주형 연속형 구분 및 카테고리 설정, 변수 제거 및 추가
2. 중복 관측치 정리 및 Missing 처리
 - 1일 4개 관측치 중 여러 변수값 중복 및 AQI NA → 원인 분석 및 처리
3. 오염물질 단위 통일
4. 관측 자료 이해
 - 주된 Max_Value 관측 시간 및 장소 확인

Main

1. 날짜
 - 관측 간격 확인
 - 년/월/일별 및 계절별 각 오염물질 평균 농도 구하고 비교
 - 오염물질 농도가 급격히 증가/감소한 시점 알아보기
2. 오염물질
 - 네(NO2, O3, SO2, CO) 오염물질 조사
 - First_max_value, AQI 확인
 - 오염물질의 증감 기울기 확인 후 미래의 오염물질 농도 예측해보기
 - 대기 오염 줄이기 위한 방안 마련
 - 이상치 관측 시 원인 분석
3. 위치 및 지역
 - 코드화된 실제 지명 확인
 - 주/군/시별 각 오염물질 평균 농도와 비율 구하고 비교
 - 지역별 관측 특징 분석
 - 동/서/남/북으로 나누고 지역별 관측 특징 분석
 - 지역의 산업 특성과 오염물질 관계 분석

Summary

1. 추가 논의 및 정리 방향
 - 지도에 시각화 및 추가 자료 검색 예정
 - 가설 세우기
2. 보고서 및 발표
 - 팀원 모두 분석에 참여, 특징을 잘 나타내는 코드와 그림으로 제출