

서울시 미세먼지 데이터를 활용한 분석 기획

: 유소년 보행 환경 개선에 대한 시나리오



〈OBAZ팀〉
김동환, 황윤선

목차



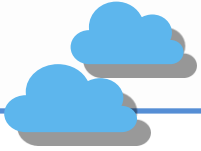
개요

분석 기획

분석 시나리오

결론

01 개요



(1) 기획 배경

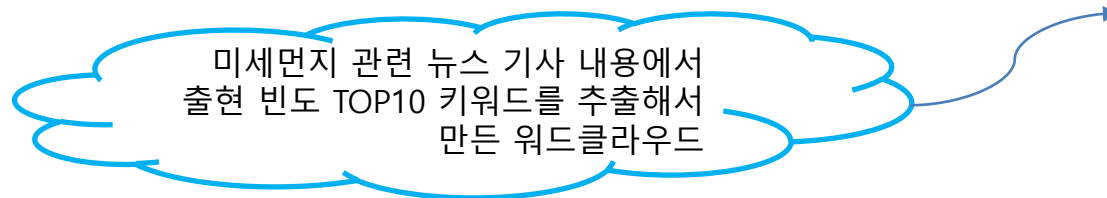
- '미세먼지'는 주기적으로 우리나라 대기환경 문제에 언급됨
- 2017년 상반기 미세먼지 역사상 2번째로 심각한 수준의 대기오염 발발
- 건강취약계층에 대한 이슈가 지속적으로 나타남

미세먼지, 어린이·임산부·노인에게 더욱 치명적... 왜?

대응 기준 강화했지만... 미세먼지에 둔감한 교육현장

(2) 필요성

- 건강취약계층 : 어린이, 노약자, 임산부 등
- 15세 미만 유소년들의 보행 및 교육환경에 대한 이슈



차량 발령 에어 지원 통학 사업 저감 어린이

미먼지

01 개요



(3) 현황 및 실태

- 미세먼지 예보등급 : 4단계 (좋음, 보통, 나쁨, 매우 나쁨)

| 구분 | | 등급($\mu\text{m}/\text{m}^3$) | | | |
|----------|-----------------|--------------------------------|-------|--------|--------|
| | | 좋음 | 보통 | 나쁨 | 매우나쁨 |
| 예보 물질 | 부유먼지 (PM10) | 0~30 | 31~80 | 81~150 | 151 이상 |
| | 미세먼지 (PM2.5) | 0~15 | 16~50 | 51~100 | 101 이상 |

* 출처 : 환경부 / * PM10 : 입자지름 < $10\mu\text{m}$, PM2.5 : 입자지름 < $2.5\mu\text{m}$

- 환경부 '미세먼지 대응매뉴얼'

| 단계 | 수준 | 기준 | 학교의 대응 |
|----|--------|---|-----------------------------|
| 1 | 고농도예보 | 익일 예보 평균 '나쁨' 이상 | . |
| 2 | 예비주의보 | 당일 예보 '나쁨' 이상, PM10 $100\mu\text{m}$ 또는 PM2.5 $50\mu\text{m}$ 이상 2시간 지속 | 야외수업 자제 |
| 3 | 주의보 발령 | PM10 $150\mu\text{m}$ 또는 PM2.5 $90\mu\text{m}$ 이상 2시간 지속 | 야외수업 단축 또는 야외수업 금지 |
| 4 | 경보 발령 | PM10 $300\mu\text{m}$ 또는 PM2.5 $180\mu\text{m}$ 이상 2시간 지속 | 등하교 시간 조정, 수업 단축 및 휴업 권고 |

01 개요



(3) 현황 및 실태

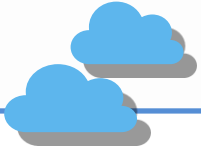
● 서울특별시 '미세먼지 감소 대응 정책6'

1. 노후 경유차 운행제한 제도의 확대
 - 범위를 전국 차량으로 늘림
2. 단속 장비의 확대
 - 13개소 -> 22개소 -> 61개소(2019년)
3. 서울, 인천 진입 버스의 CNG 전환 유도
 - 현재 시내버스 : 100% CNG 버스
 - 서울 진입 경기 및 인천 버스 일부 : 경유버스
4. 생활주변 오염원에 대한 비산먼지 저감대책 실행
 - 비산먼지 발생사업장으로 신고된 공사장 대상 특별점검, 분진흡입청소차 확보, 물청소차 가동률 확대
5. 제도 개선
 - 환경부 건의, 친환경 자동차 저변 확대 지원 등
6. 동북아 주요 도시 간 연대 및 협력 추진

● 서울특별시교육청 '학교 미세먼지 종합대책'

1. 학교 미세먼지 관리 기반 구축
2. 학교 미세먼지 저감 환경조성
 - 환경부 미세먼지 대응매뉴얼을 한 단계씩 상향 시행
3. 미세먼지 관련 교육활동 강화
 - 마스크 사용법에 대한 교육자료 배포
4. 유관기관 협력체계 강화

01 개요



(4) 문제 정의

- 빅데이터를 활용한 서울시 유소년 미세먼지 노출 환경 개선
 - ▶ 학교 앞 유소년 보행 구역의 미세먼지 환경 개선을 위한 대책이 미약
 - ▶ 교육기관 등하교 시간을 조정하거나 수업을 단축하는 등의 일시적 방편은 근본적인 대응방안이 될 수 없음



(5) 목적

- 교육기관별 미세먼지 환경 문제 파악 및 대응방안 강구
- 미세먼지 분석 및 시각화 결과를 유소년 미세먼지 환경 개선에 활용
- IoT 기기 활용을 통한 데이터 수집 및 구역별 맞춤형 환경 개선안 제안

02 분석 기획



(1) 가설 설정

가설1

- 학교 앞 유소년 보행구역에서, 유소년의 주 야외 노출 시간대인 등하교 시간대에는 타 시간대에 비해 높은 미세먼지 농도를 가질 것이다.

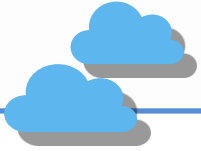
가설2

- 학교 앞 유소년 보행구역에서, 등하교 시간대에 발생하는 미세먼지의 주요 원인은 주*정차량 및 통행 차량일 것이다.

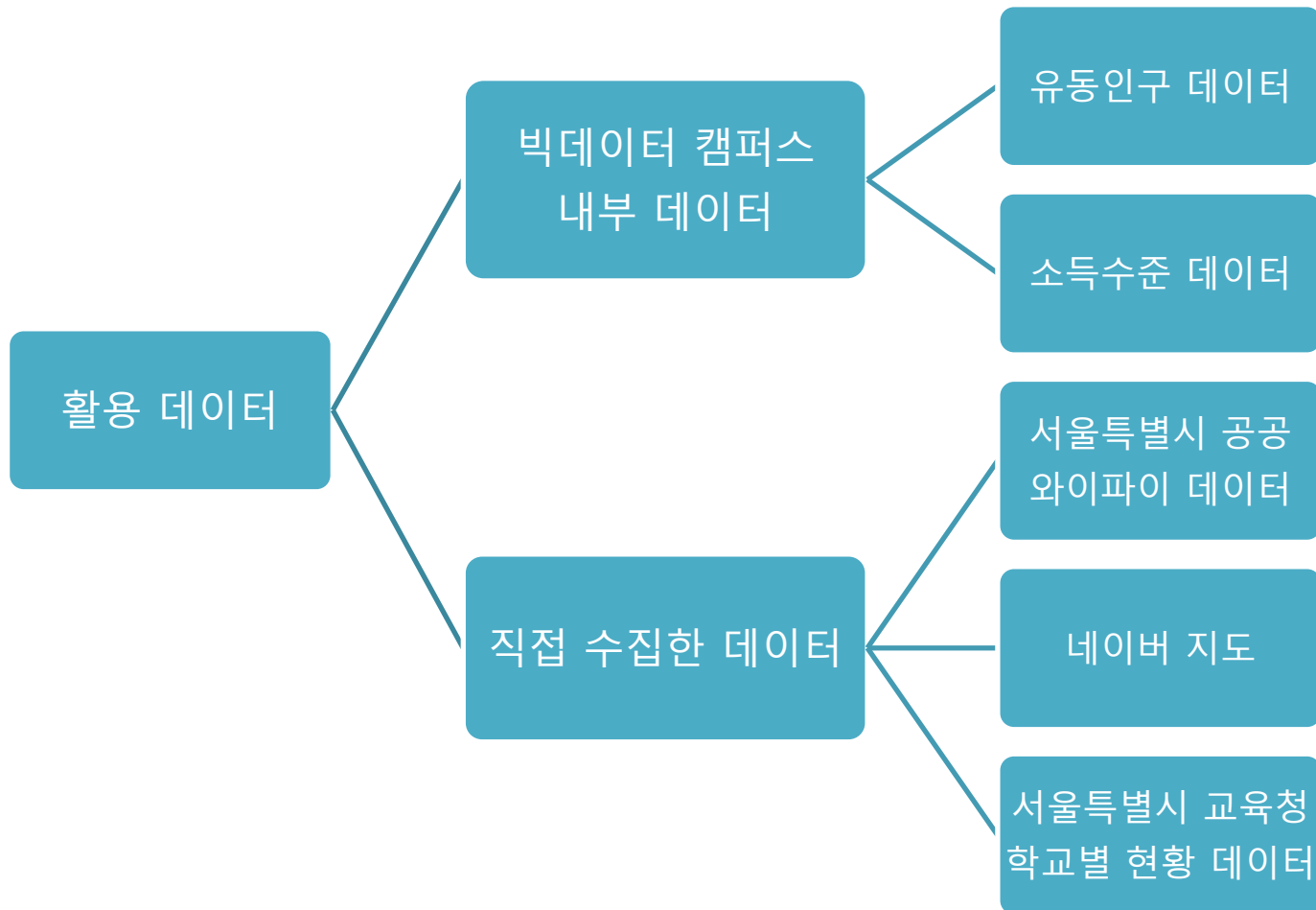
가설3

- 지역별 평균측정치보다 특히 높거나 낮은 학교의 경우, 인근 환경(지리적, 인적 등)이 미세먼지 농도에 영향을 미칠 것이다.

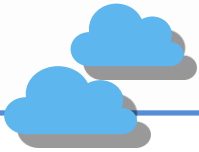
02 분석 기획



(2) 데이터 수집



02 분석 기획



(2) 데이터 수집

★ 서울시 빅데이터 캠퍼스

● 유동인구 데이터

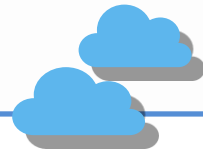
- 보유기관 : 빅데이터 캠퍼스 (제공기관 : KT)
- 수집목적 : 서울시 내 IoT기기 위치 선정에 위함
- 속 성 : 1개 변수 (등교시간대와 하교시간대의 동별 유동인구 평균)

● 소득수준 데이터

- 보유기관 : 빅데이터 캠퍼스 (제공기관 : 나이스지니데이터㈜)
- 수집목적 : 서울시 내 IoT기기 위치 선정에 위함
- 속 성 : 1개 변수 (행정동별 소득수준)

| id | 학교명 | 동 | 소득수준 | 유동인구_등교 | 유동인구_하교 | 유동인구 |
|----|--------------|------|------|---------|---------|--------|
| 1 | 서울갈현초등학교 | 갈현2동 | 3551 | 2243 | 2477 | 2360 |
| 2 | 서울갈현초등학교 | 갈현2동 | 3551 | 2243 | 2477 | 2360 |
| 3 | 서울갈현초등학교 | 갈현2동 | 3551 | 2243 | 2477 | 2360 |
| 4 | 성균관유치원 | 갈현2동 | 3551 | 2243 | 2477 | 2360 |
| 5 | 선정중학교 | 갈현2동 | 3551 | 2243 | 2477 | 2360 |
| 6 | 그리새몬테소리 어린이집 | 갈현2동 | 3551 | 2243 | 2477 | 2360 |
| 7 | 서울녹번초등학교 | 녹번동 | 3608 | 148237 | 279943 | 214090 |
| 8 | 은정유치원 | 대조동 | 3451 | 33190 | 56160 | 44675 |
| 9 | 서울신사초등학교 | 신사동 | 3332 | 162413 | 285333 | 223873 |

02 분석 기획



(2) 데이터 수집

★ 직접 수집한 데이터 - 교육기관별 데이터 수집 기준 수립 및 수집

● 서울특별시 공공와이파이

- 보유기관 : 서울특별시
- 수집목적 : 송수신을 고려하여 공공와이파이 근처에 IoT 기기 위치 선정을 위한 지도를 통해 공공와이파이 근처의 환경 또한 파악 가능
- 속 성 : 15개 변수
구, 와이파이 주소, 학교명, 환경변수 (산, 공원, 천) 등



와이파이 스팟 예시(1)

(출처 : 서울시공공와이파이 홈페이지)

이 레코드의 경우 버스정류장은 1,
주도로는 1의 값을 가진다.



와이파이 스팟 예시(2)

(출처 : 서울시공공와이파이 홈페이지)

이 레코드의 경우 공원은 1,
근접측정은 1의 값을 가진다.

02 분석 기획



(2) 데이터 수집

★ 직접 수집한 데이터 – 교육기관별 데이터 수집 기준 수립 및 수집

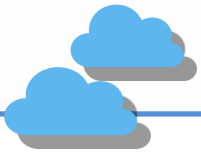
● 네이버 지도

- 보유기관 : 네이버
- 수집목적 : 서울시 내 IoT 기기 위치 선정을 위함
- 속 성 : 운동장 여부, 버스정류장 등

| id | 학교명 | 지역구분_한강 | 구 | 와이파이주소 | 주도로이름 | 주도로 | 공원 | 근접측정 | 산 | 천 | 버스정류장 | 주거지역 |
|----|--------------|---------|-----|-----------------------------|---------|-----|----|------|---|---|-------|------|
| 1 | 서울갈현초등학교 | 강북 | 은평구 | 갈현2동 주민센터(연서로27길 31-6) | 연서로27길 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 2 | 서울갈현초등학교 | 강북 | 은평구 | 갈현초 후문 CCTV(연서로29길 35) | 연서로29길 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 서울갈현초등학교 | 강북 | 은평구 | 갈현초 정문 CCTV(연서로27길 25-1) | 연서로27길 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 4 | 성균관유치원 | 강북 | 은평구 | 갈현초 정문 CCTV(연서로27길 25-1) | 연서로27길 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 5 | 선정중학교 | 강북 | 은평구 | 갈현동 거리_3(갈현2동 327-21) | 서오릉로20길 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 6 | 그리새몬테소리 어린이집 | 강북 | 은평구 | 길마 어린이공원(연서로 21길 15) | 연서로 | 0 | 1 | 0 | 0 | 0 | 2 | 1 |
| 7 | 서울녹번초등학교 | 강북 | 은평구 | 녹번초교(녹번로 1길 1) | 녹번로1길 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| 8 | 은정유치원 | 강북 | 은평구 | 대조동 주민센터 교육장(연서로 20길 24-3) | 연서로 20길 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 9 | 서울신사초등학교 | 강북 | 은평구 | 드림스타트센터 민원실(증산로 15길 401-18) | 증산로 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 10 | 은화어린이집 | 강북 | 은평구 | 박석고개공원(통일로 89길 18-29) | 통일로 89길 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 11 | 은지어린이집 | 강북 | 은평구 | 새마을금고 건너편(갈현로 1길 24) | 갈현로1길 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 12 | 구립개나리아린이집 | 강북 | 은평구 | 새장골공원(연서로37가길 10-12) | 연서로37가길 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 13 | 꽃들어린이집 | 강북 | 은평구 | 서울재활체육센터(갈현로 11길 30) | 갈현로11길 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 14 | 하늘빛어린이집 | 강북 | 은평구 | 시내공원(응암어린이)(은평로11길 10) | 은평로11길 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 15 | 새롬어린이집 | 강북 | 은평구 | 시루메공원(증산로5길 10-1) | 증산로5길 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

공공와이파이 및 네이버 지도를 통해 수집한 데이터 일부 (총 324개 교육기관(324 row))

02 분석 기획



(2) 데이터 수집

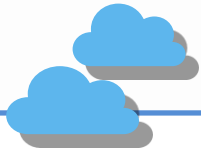
★ 직접 수집한 데이터 - 교육기관별 데이터 수집 기준 수립 및 수집

● 서울특별시 교육청 학교별 현황

- 보유기관 : 서울특별시 교육청
- 수집목적 : 서울시 내 IoT기기의 개수 배분 방식 수립을 위함
- 속 성 : 3개 변수
구 별 유치원, 초등학교, 중학교 개수

| 구 | 구_유치원 | 구_초등학교 | 구_중학교 | 계 |
|-----|-------|--------|-------|-------|
| 노원구 | 7364 | 31048 | 20968 | 59380 |
| 송파구 | 6027 | 32093 | 20261 | 58381 |
| 양천구 | 5168 | 25237 | 19798 | 50203 |
| 강서구 | 5643 | 27960 | 12962 | 46565 |
| 강남구 | 4213 | 24007 | 17025 | 45245 |
| 은평구 | 4600 | 23918 | 13364 | 41882 |
| 성북구 | 5440 | 23306 | 11732 | 40478 |
| 강동구 | 4698 | 21620 | 12481 | 38799 |
| 서초구 | 3097 | 22817 | 12793 | 38707 |
| 구로구 | 3568 | 19363 | 9992 | 32923 |

02 분석 기획



(3) 위치 선정

★ 100개의 교육기관(IoT기기 위치) 선정

[1단계] 현황 파악

- 당시 예정되어 있던 미세먼지 측정 IoT기기 : 총 100대 (제약조건)
- 실제 기기 설치가 어려운 주변 환경을 고려하여 2순위 설치 기관까지 선정
- 2회에 걸쳐 50군데씩 총 100개의 설치 스팟 선정



[2단계] 데이터 수집

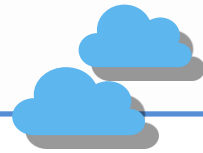
- 대상 교육기관 : 어린이집, 유치원, 초등학교, 중학교로 선정
- 데이터 송수신을 고려, 서울시 공공와이파이가 설치 스팟에 근접한 교육기관 선정
- 빅데이터캠퍼스 내부데이터를 활용하거나 외부에서 데이터를 직접 수집함



[3단계] IoT 설치 교육기관(위치) 선정

- 1차 50개 : 서울특별시의 전 방위를 고려하여 25개 자치구별로 2개씩 선정
- 2차 50개 : 미세먼지 환경에 영향을 받는 유소년 수를 고려하여 자치구별 유소년 교육기관 수에 비례해서 개수를 배분

02 분석 기획



(3) 위치 선정

★ 1차 50개 위치 선정

● 서울시 25개 자치구별 2개씩 선정

- 목 적 : 서울시 전 지역 내 고르게 분포시켜 지역별 차이를 분석할 수 있게 하기 위함
- 선정 기준 : 가설2의 '주정차 차량 및 통행차량'과 관련이 있는
주거지역, 버스정류장, 운전면허장, 주유소, 주도로의 5개 변수에
우선순위를 두어 선정
(주도로, 버스정류장, 주유소, 주거지역, 운전면허장 순)

- ① 주도로 여부 : 미세먼지 발생에 직접적인 원인을 제공하는 배기가스 고려
- ② 버스정류장 여부 : 출발할 때와 멈출 때에 배기가스를 다량 방출하는 버스의 영향 고려
- ③ 주유소 여부 : 주정차 차량 고려
- ④ 주거지역 : 많은 차량통행 고려
- ⑤ 운전면허장 : 많은 차량통행 고려

02 분석 기획



(3) 위치 선정

★ 2차 50개 위치 선정

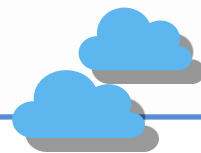
● 1차 선정 후, 교육기관 수에 비례하여 구별로 개수 배분

- 목적 : 더 많은 유소년에게 영향을 줄 수 있으므로 교육기관 수에 비례하여 추가 배분
- 선정기준 : 가설3을 검증하기 위해 '지리적, 인적 요소'를 고려 반영할 수 있도록
유동인구, 소득수준, 산, 천의 4개 변수에 우선순위를 부여하여 선정

- ① 유동인구 : 사람들이 움직일 때 비산먼지가 많이 발생하므로 유동인구 변수를 고려
- ② 소득수준 : 해당 교육기관 일대의 전반적인 소득수준 고려
- ③ 산 : 지리적 환경 대변
- ④ 천 : 지리적 환경 대변

| 구 | 개수 | 구 | 개수 | 구 | 개수 |
|-----|----|------|----|------|----|
| 노원구 | 4 | 영등포구 | 2 | 성동구 | 1 |
| 송파구 | 3 | 관악구 | 2 | 서대문구 | 1 |
| 강서구 | 3 | 광진구 | 2 | 강북구 | 1 |
| 성북구 | 3 | 동작구 | 2 | 금천구 | 1 |
| 양천구 | 3 | 동대문구 | 2 | 용산구 | 1 |
| 은평구 | 3 | 중랑구 | 2 | 종로구 | 1 |
| 강남구 | 2 | 도봉구 | 2 | 중구 | 1 |
| 강동구 | 2 | 마포구 | 2 | | |
| 구로구 | 2 | 서초구 | 2 | | |

02 분석 기획



(4) 분석 방법론

- ★ 가설1 : “학교 앞 유소년 보행구역에서,
등하교 시간대에는 타 시간대에 비해 높은 미세먼지 농도를 가질 것이다.”

H_0 : 등하교 시간대의 미세먼지 농도 = 타 시간대의 미세먼지 농도

H_1 : 등하교 시간대의 미세먼지 농도 > 타 시간대의 미세먼지 농도

➡ 정규성 검정과 등분산성 검정 후, T검정을 수행하여 농도에 차이가 있는지 확인

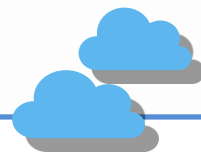
- ★ 가설2 : “등하교 시간대에 발생하는 미세먼지의 주요 원인은 주,정차량 및 통행차량일 것이다.”

방법1. 변수들 간 상관관계 시각화를 통해 변수별 영향력 파악

방법2. 미세먼지 측정량을 종속변수(y), 나머지 변수들을 독립변수(x)로 하는
다중회귀분석 수행

➡ 회귀분석의 기본 가정(선형성, 오차의 정규성, 독립성, 등분산성)을 확인
변수 선택법을 이용하여 모델 선정
최종 모델의 유의성검정 및 각 회귀계수의 유의성을 T-test를 이용하여 검정
유의한 변수들 중, 미세먼지 수치에 상대적으로 큰 영향을 미치는 독립변수를
찾기 위해 각 변수의 SSR/SST값 확인

02 분석 기획



(4) 분석 방법론

★ 가설3 : “지역별 평균측정치보다 특히 높거나 낮은 학교의 경우,
인근 환경(지리적, 인적 등)이 미세먼지 농도에 영향을 미칠 것이다.”

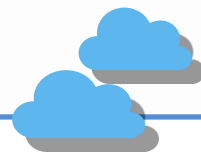
군집분석 : 교육기관들을 변수에 따라 군집으로 형성

군집 별 주요 변수 파악 후, 맞춤 환경 개선안 도출 가능

- { H-clustering : 개체들을 가까운 집단부터 차근차근 묶어가는 방식
- { K-means : 전체 데이터의 영역을 특정 기준에 의해 동시에 구분하는 방식

➡ 변수의 개수가 많고 유사한 개체들이 결합되는 과정 파악이 필요하므로
H-clustering 진행

03 분석 시나리오



Case1) '버스정류장 여부' 변수의 영향도가 가장 큰 군집

해당 군집에 속한 교육기관의 지역구 자치단체에서,
유소년 보행 환경에 영향을 끼치는 버스 정류 매연을 감소시킬 수 있는 방안 추진

Ex. 도로변 펜스(fence) 설치 등

Case2) '어린이집' 과 '유치원' 이 속한 비율이 높은 군집

- 해당 군집에 속한 교육기관이
유소년 등하교 버스, 학교 근처 주정차 부모 차량 등의
'1분 시동 끄기 캠페인'을 실시하는 방안 추진
- 모든 버스와 자동차의 시동을 1분만 꺼도 미세먼지 생성의
주 원인인 자동차 배기가스량을 획기적으로 줄일 수 있을 것으로 기대

04 결론



활용방안 및 기대효과

★ 활용방안

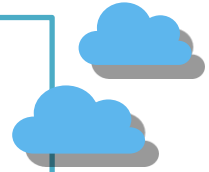
1. 유사한 미세먼지 환경을 가진 교육기관 군집을 형성
2. 형성된 교육기관 군집 별 미세먼지 환경 지수 개발
3. 형성된 교육기관 군집 별 환경 개선안 검토
4. 미세먼지 발생 주요 원인 해결과 관련된 캠페인 실시

★ 차후 기술 융합 연구 관련

- IoT를 활용한 데이터 수집에 대한 아이디어 제공
: 데이터 송수신 관련 공공 와이파이 활용

● 활용 데이터

- 유동인구 데이터 : 서울시 빅데이터 캠퍼스(www.bigdata.seoul.go.kr)
- 소득수준 데이터 : 서울시 빅데이터 캠퍼스(www.bigdata.seoul.go.kr)
- 서울특별시 공공 와이파이 : www.wifi.seoul.go.kr
- 네이버 지도 : www.map.naver.com
- 교육청 학교별 현황 데이터 : www.sen.go.kr
- SGIS 행정구역 코드



● 분석 툴

- R
- Excel
- Naver DataLab

● 참고문헌

- '초미세먼지로 인한 어린이 환경성 질환 영향 연구' - 한국환경정책평가연구원
- '건강취약계층 보호를 위한 고농도 미세먼지 대응메뉴얼' - 환경부
- '서울시 먼지오염 특성연구' - 보건환경연구원
- '교통량과 대기오염물질' - 보건환경연구원
- '서울시 간선도로변 대기환경특성' - 보건환경연구원