

서울시,

2022.12.16 다변량통계분석 프로젝트
대기오염에 따른 서울시 지역 군집화

대기오염의 주범은?





다들 건강통계 분석의 전설

UNTITLED.ipynb

시빅데이터융합경영학과 이태범 20202658

시빅데이터융합경영학과 한윤지 20202672

시빅데이터융합경영학과 최준용 20182832



Time table
목차

- ◆ 01 프로젝트 배경과 주제
- ◆ 02 데이터 소개
- ◆ 03 데이터 탐색



Time table
목차

- ◆ 04 데이터 분석
- ◆ 05 분석결과 도출
- ◆ 06 활용방안



서울시 대기오염의 주범을 찾아라!

한국보건사회연구원 조사결과 위험에 대한 불안수준 1위 : 대기오염

한국은 경제협력개발기구(OECD) 회원국 중 초미세먼지 오염 농도 1위를 차지해 최악의 대기오염 국가




- 미세먼지 수치를 확인하며 하루를 시작할 만큼 큰 걱정거리로 자리잡음
- 한국보건사회연구원에서 위험에 대한 불안 수준을 측정한 결과 대기오염이 1위
- 현재 서울시에서는 권역별로 대기오염정책을 시행하고 있음



우리나라에서 특히 대기오염이 심한 서울시에서 비슷한 대기오염 유형을 보이는 지역을 **군집화**한 뒤 각 지역에 맞는 해결책을 제안



서울열린데이터광장

-  측정소별_일별_평균대기오염도_2020.csv
-  측정소별_일별_평균대기오염도_2021.csv
-  측정소별_일별_평균대기오염도_2022.csv

측정일시	측정소명	이산화질소농도	오존농도	일산화탄소농도	아황산가스농도	미세먼지농도	초미세먼지농도
20200101	강남구	0.037	0.003	0.7	0.003	31.0	23.0
...



대기측정소 데이터 소개

- 도시대기측정소(25개소)
- 도로변측정소(15개소)
- 도시배경·경계·입체 측정소(10개소)



atmospheric monitoring station

대기 측정소

- 서울시 대기 측정소 50곳
- 도시대기측정소(25개소)
도로변측정소(12개소)
도시배경·경계·입체 측정소(10개소)



변수정의 데이터 소개

변수생성

측정소별 평균값



월별 대기오염도

1월~12월



계절별 대기오염도

Spring, Summer, Fall, Winter

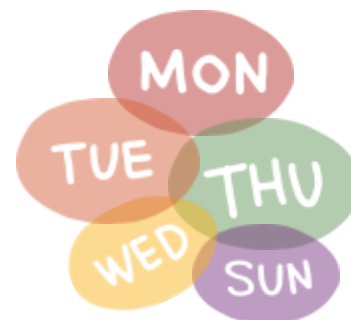
년도별 대기오염도

2020년, 2021년, 2022년



요일별 대기오염도

요일별 대기오염도 평균



상하반기 대기오염도

1-6월, 7-12월

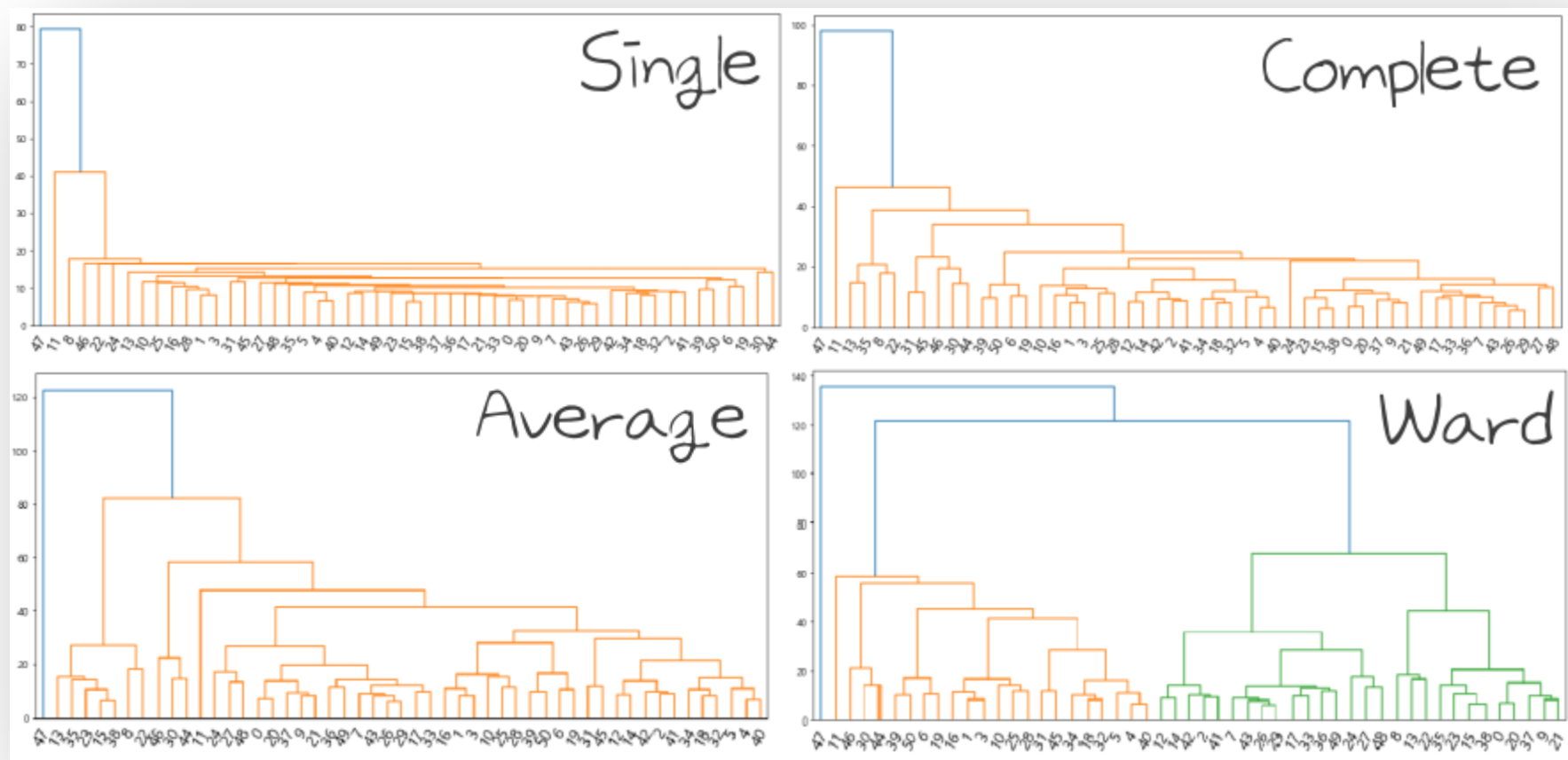
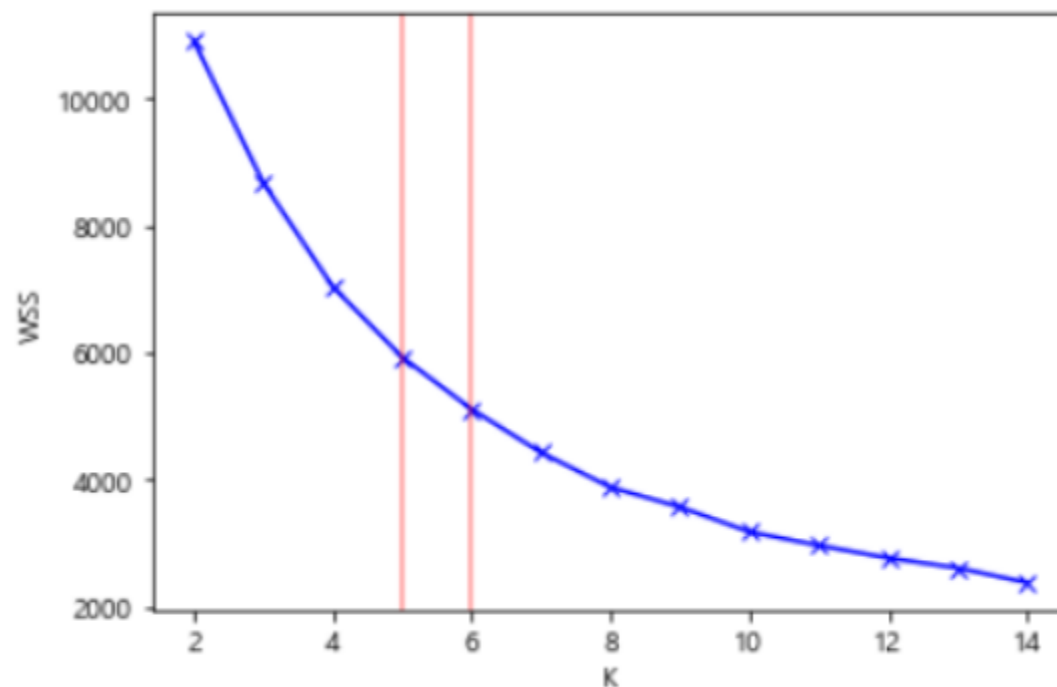


Linkage Method

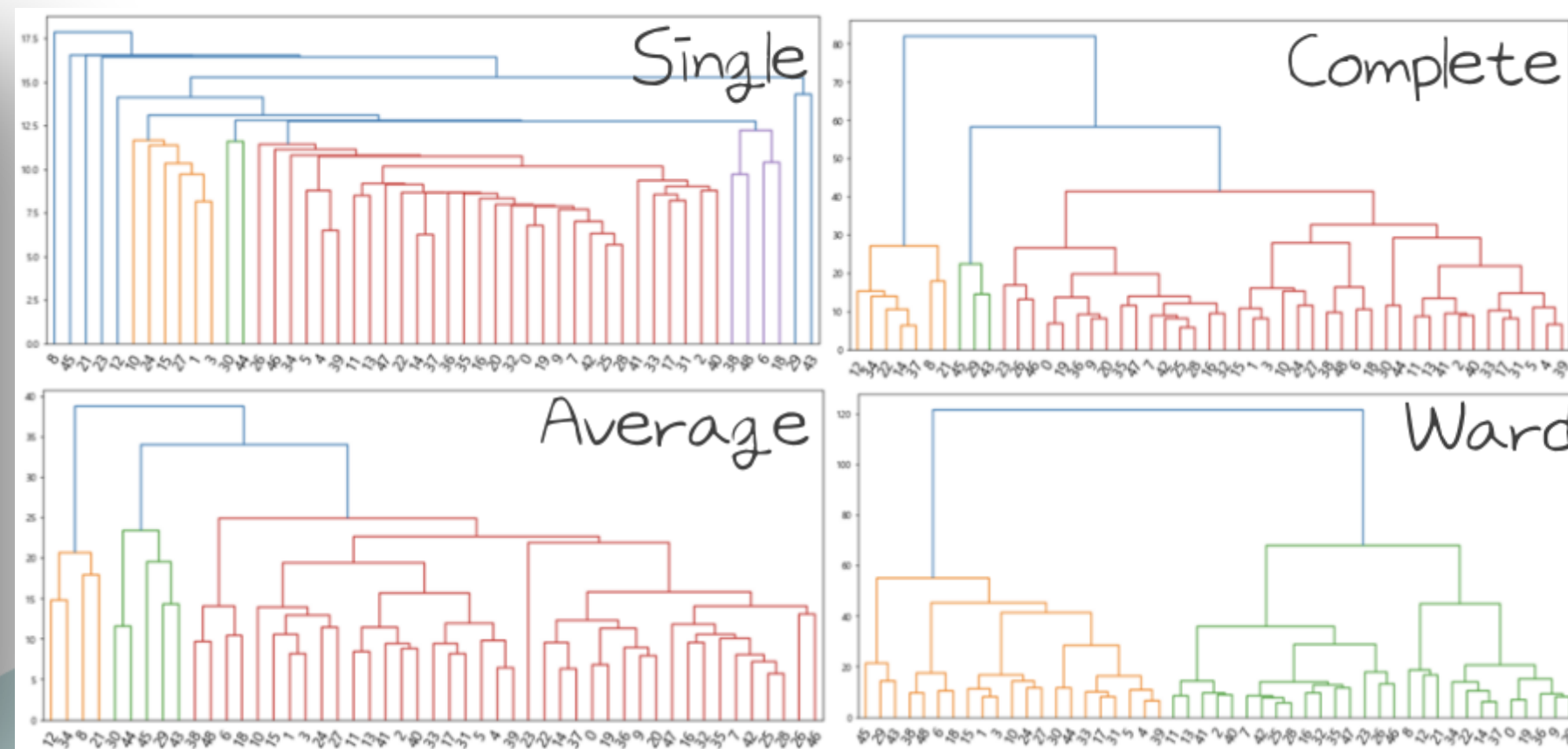
47번, 11번 Index인 항동, 궁동 제거

K-Means Clustering

4-6개 사이의 군집이 적합



Before



After



군집의 분포가 고른 **Ward** 방법을 이용하여 6개로 군집화

Cluster 0 '강남구', '광진구', '도봉구', '마포구', '마포아트센터', '서대문구', '올림픽공원', '은평구', '자연사박물관'

Cluster 1 '강남대로', '강변북로', '강북구', '강서구', '구로구', '도산대로', '동작구', '서초구', '세곡', '신촌로', '양천구', '영등포로', '종로', '청계천로'

Cluster 2 '강동구', '관악구', '금천구', '노원구', '동대문구', '서울숲', '성동구', '성북구', '송파구', '영등포구', '용산구', '종로구', '중구', '중랑구', '행주', '홍릉로'

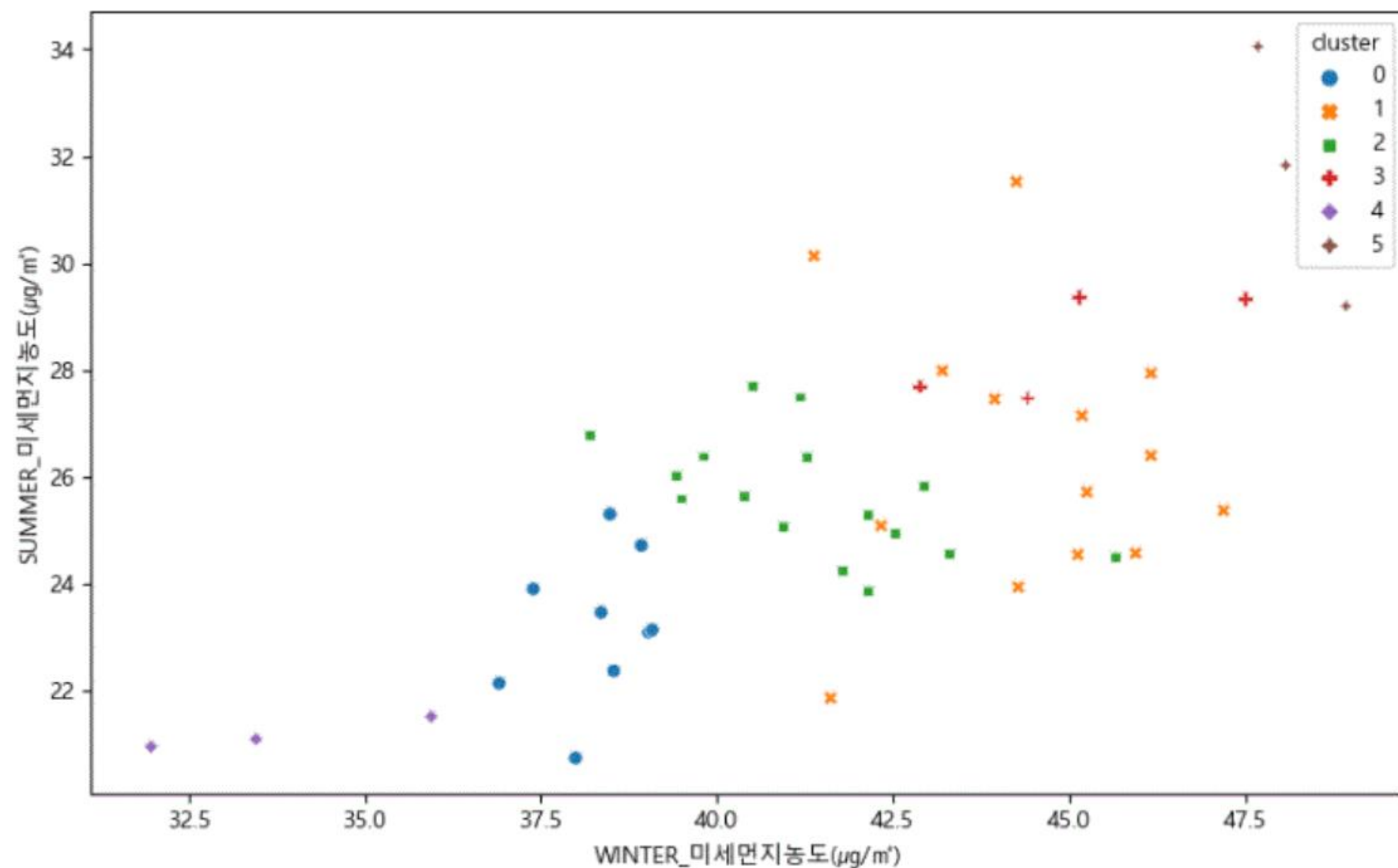
Cluster 3 '공항대로', '동작대로', '정릉로', '화랑로'

Cluster 4 '관악산', '남산', '북한산'

Cluster 5 '시흥대로', '천호대로', '한강대로'



군집의 분포가 고른 **Ward** 방법을 이용하여 6개로 군집화



군집화가 잘 진행 됨!





탐색적요인분석

Factor 1 : 미세먼지 오염

미세먼지농도, 초미세먼지농도

Factor 2 : 자동차 환경오염

이산화질소농도, 일산화탄소농도, 아황산가스농도

Factor 3 : 오존오염

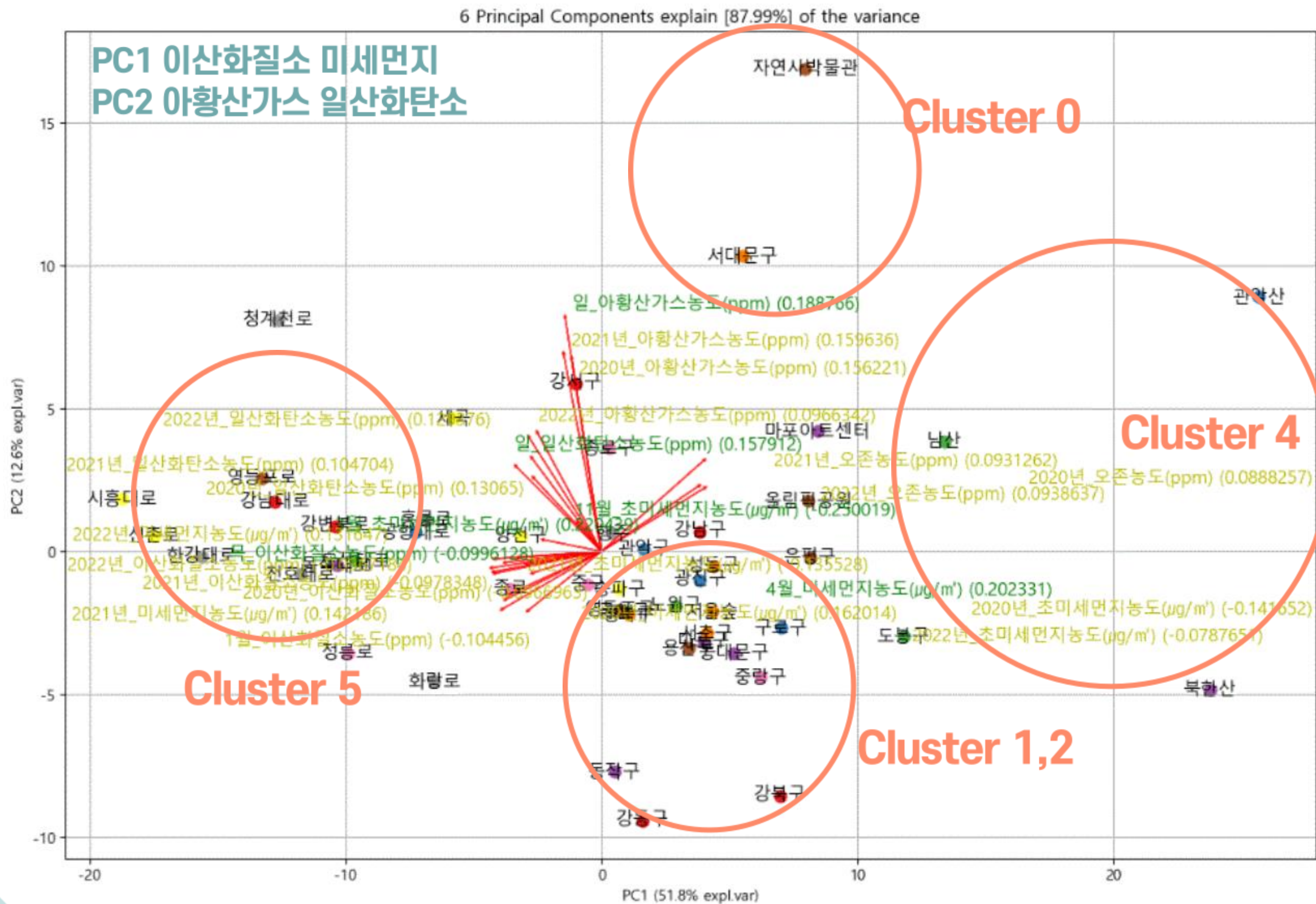
오존농도

	Factor 1	Factor 2	Factor 3
이산화질소농도(ppm)	0.293223	0.715414	-0.347297
오존농도(ppm)	0.042262	-0.277577	0.957202
일산화탄소농도(ppm)	0.326494	0.726757	-0.239410
아황산가스농도(ppm)	0.196419	0.505673	-0.077910
미세먼지농도($\mu\text{g}/\text{m}^3$)	0.728332	0.233180	0.014407
초미세먼지농도($\mu\text{g}/\text{m}^3$)	0.914719	0.397968	-0.021383



BiPlot

- 직관적으로 보았을 때 PC1이 0일 때를 기준으로 왼쪽은 차가 많이 다니는 곳 (ex. 대로변들, 다 끝에 '로'가 붙어있음) 오른쪽은 깨끗한 지역 (ex. 산, 박물관, 공원)
- PC1에는 이산화질소와 미세먼지가 영향을 미치고 PC2에는 아황산가스와 일산화탄소가 영향을 미침



Air pollution

최종분석 및 문제해결

측정소 군집별 대기 오염도

측정소별로 점수(좋음:6, 보통:4, 나쁨2, 매우나쁨:0)를 부여하고 군집별로 대기오염의 평균점수 비교

지수		(1)		(2)		(3)		(4)	(5)
지수구분		좋음		보통		나쁨		매우나쁨	
점수구분값	I _{LO}	0		51		101		251 이상	
	I _{HI}	50		100		250			
농도구분		BP _{LO}	BP _{HI}	BP _{LO}	BP _{HI}	BP _{LO}	BP _{HI}	BP _{LO}	BP _{HI}
아황산가스(ppm)	1hr	0	0.020	0.021	0.050	0.051	0.150	0.151	1
이산화질소(ppm)	1hr	0	0.030	0.031	0.060	0.061	0.200	0.201	2
일산화탄소(ppm)	1hr	0	2.00	2.01	9	9.01	15	15.01	50
오존(ppm)	1hr	0	0.030	0.031	0.090	0.091	0.150	0.151	0.600
미세먼지PM-10(μg/m³)	24hr	0	30	31	80	81	150	151	600
미세먼지PM-2.5(μg/m³)	24hr	0	15	16	35	36	75	76 이상	

	오존오염	자동차환경오염	미세먼지오염
Cluster0	5.081696	5.885650	4.917874
Cluster1	5.383851	5.757790	4.733540
Cluster2	5.260024	5.828582	4.824940
Cluster3	5.577295	5.611108	4.758937
Cluster4	4.728502	5.972719	5.089211
Cluster5	5.630918	5.627323	4.585507

군집화 결과

Cluster 0 : 비교적 **최근 교통량이 적고**, 강남구, 광진구, 마포구, 서대문구, 은평구 등으로 군집화되었다.

Cluster 1 : 강북구, 강서구와 같이 **사람이 많이 거주**하는 지역과 청계천, 종로 등의 지역으로 군집화되었다.

Cluster 2 : 관악구, 강동구, 서울숲같이 **유동인구는 많지만 거주하는 인구는 상대적으로 적은 지역**으로 군집화되었다.

Cluster 3 : 공항대로, 동작대로, 정릉로, 화랑로같이 **도로 위주**로 군집화되었다.

Cluster 4 : **산에 있는 측정소**로 군집화. 미세먼지가 제일 적고 자동차 오염이 제일 적으며, 오존 오염정도가 가장 높다는 특징을 지닌다.

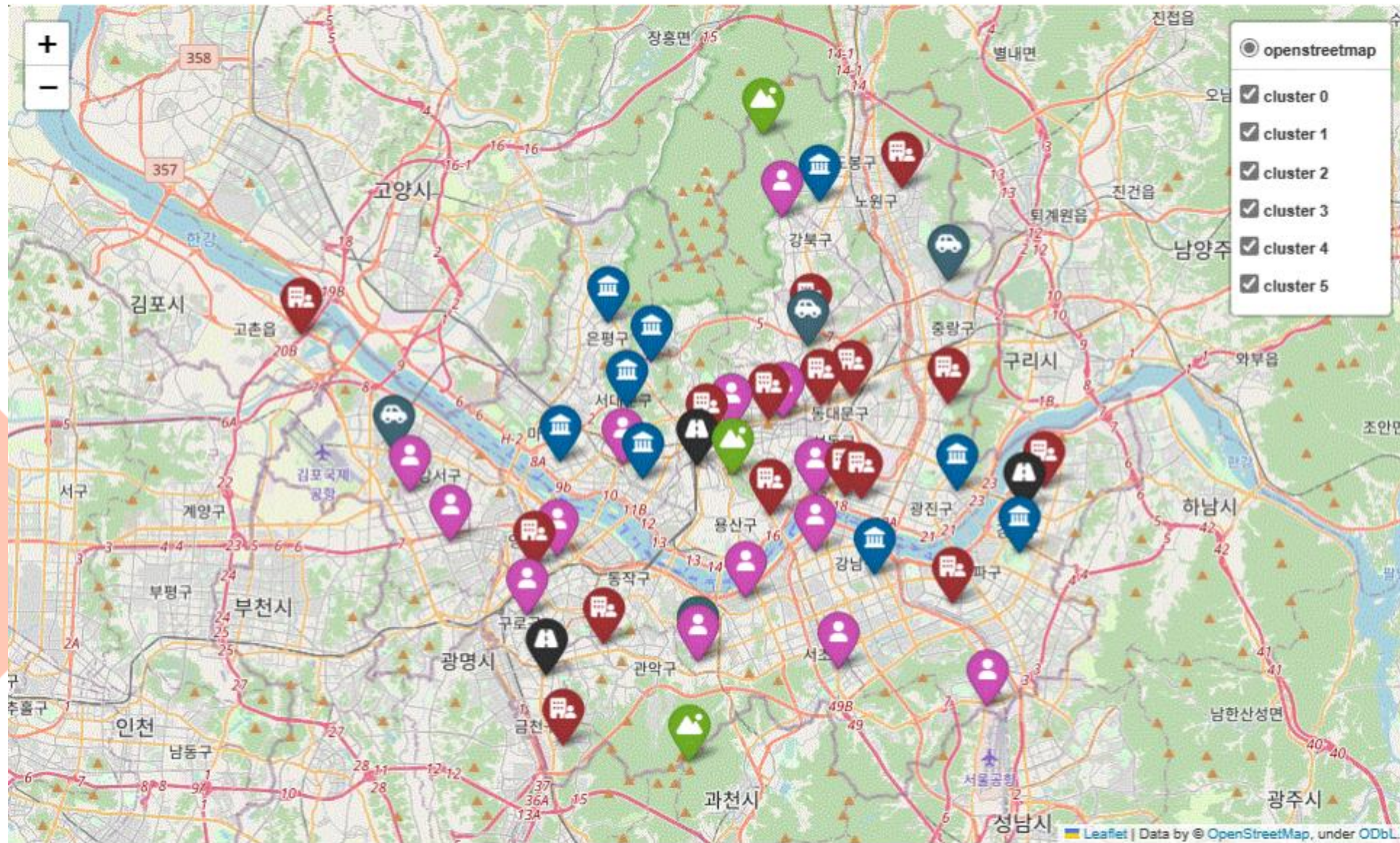
Cluster 5 : 미세먼지오염 점수가 가장 낮은 시흥대로, 천호대로, 한강대로로 **대로변 위주**로 군집화되었다.



Clustering Result-Map

최종분석 및 문제해결

군집화 결과 지도



Cluster 0 ●

올림픽공원, 마포아트센터, 강남구, 자연사박물관등이 포진 -> Landmark icon

Cluster 1 ●

비교적 조용한 지역과 청계천, 종로 등의 지역 -> 사람이 많이 거주하는 곳 User icon

Cluster 2 ●

유동인구는 많지만 거주하는 인구는 상대적으로 적은 지역 -> 회사들이 많은 곳 추측 Building icon

Cluster 3 ●

공항대로, 동작대로, 정릉로, 화랑로같이 도로 위주 -> 차가 많은 곳 Car icon

Cluster 4 ●

산에 있는 측정소 -> Mountain icon

Cluster 5 ●

시흥대로, 천호대로, 한강대로로 대로변 위주 -> Road icon



①도로청소경로지정

**서울시 대기환경정보 미세먼지 비상저감조치

◇ 주요 간선도로, 일반도로 청소 강화
- 도로 분진흡입 및 물청소확대(1일 1회→2회)

도로 청소

◇ 소방차 활용 거리물분사
- 소방서 주변 이면도로, 공원 등 물분사
※ 화재진압, 인명 구조·구급 활동에 지장이 없는 범위내에서 시행



- 군집별로 미세먼지가 많은 시간을 예측해서 시행하는 것이 매일 청소를 진행하는 것보다 효율적
- 미세먼지가 심하지 않은 시간에 시행하는 것은 의미 없음
- 현재는 일별 데이터밖에 없지만 시간대별로 조사를 해서 미세먼지가 심한 시간대에 해당 Cluster에 도로Map을 보고 경로를 지정 해주어 도로 청소를 진행하는 방향 (ex. 심한 군집은 2회, 심하지 않은 군집은 1회)



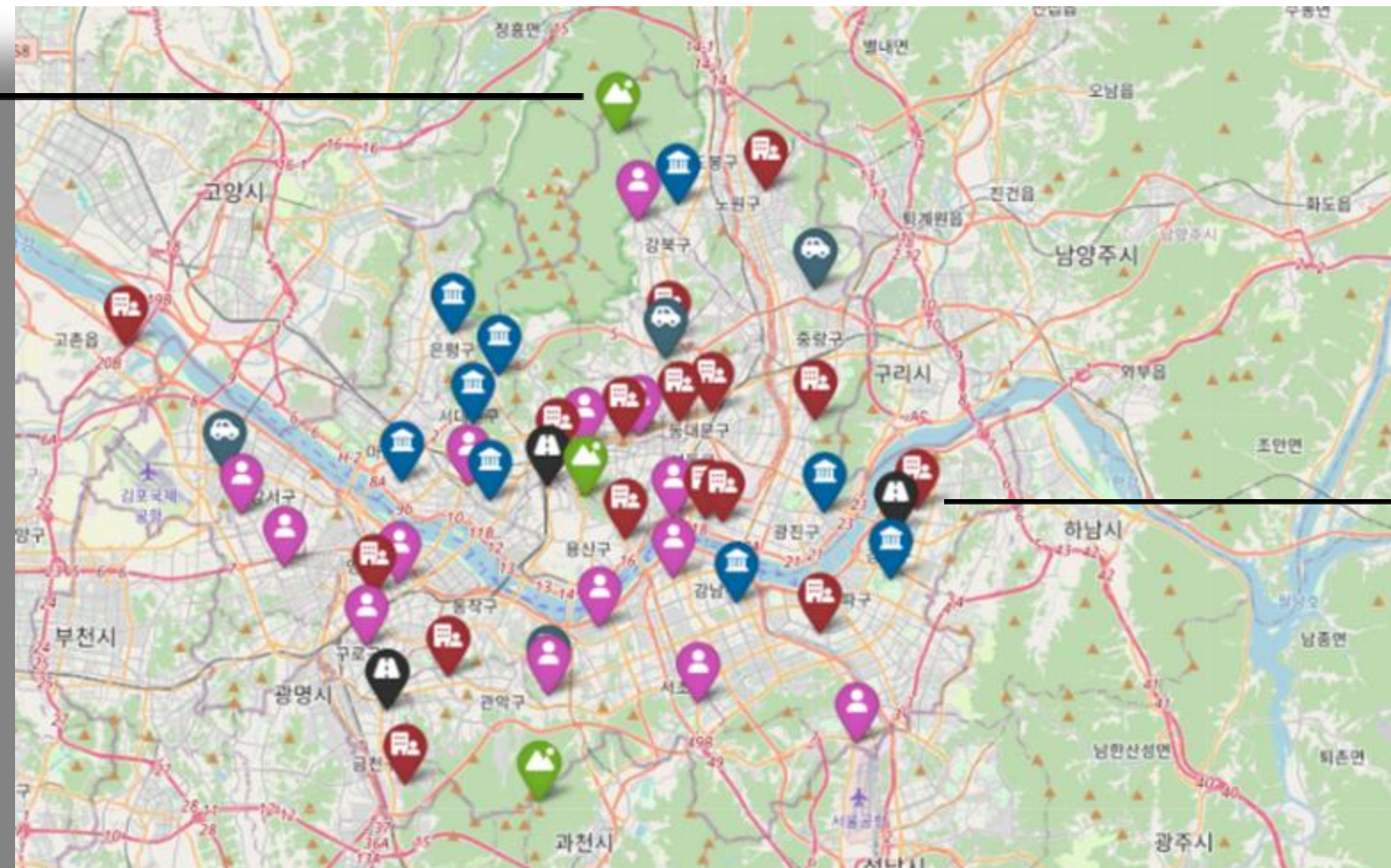
②군집별 대기환경정책

- 대기환경요인을 파악하여 군집화된 지역별로 다른 환경정책을 실시하여 대기오염 정도를 더 최소화시킬수 있다

Cluster 4



오존농도를 낮추기 위한 방안제시



미세먼지를 줄이기 위한 방안
미세먼지의 대응책
자동차 교통량 규제



Cluster 5



2022.12.16 다변량통계분석 프로젝트

Thank you!

감사합니다.

