



## Motivation & Key Idea

### 프로젝트 배경 및 목표

정부는 2027년까지 지하철역사 초미세먼지를 24 $\mu\text{g}/\text{m}^3$ 로 줄이기 위해 총 3천억 원을 투입할 것이라 밝혔다.

본 프로젝트는 서울시 지하철역사 대기질에 영향을 미치는 변수를 분석함으로써 역사별 미세먼지 저감을 위한 원인별 맞춤 대응방안을 제안한다. 이를 통해 지하철역사의 공기질을 개선하고, 효율적인 예산 활용 및 이용자에게 쾌적한 환경 제공을 목표로 한다.

### 프로젝트 핵심 아이디어

첫째, 변수선택법을 수행한다.

본 프로젝트에서 분석을 위해 수집한 48개의 변수 중에서 종속변수인 "역사 내 초미세먼지 농도"에 유의미한 변수를 선별하여 활용한다.

둘째, 회귀분석을 수행한다.

변수선택법을 통해 선별된 독립변수들을 이용하여 회귀분석을 실시하고, 종속변수에 미치는 영향을 확인한다.

셋째, 군집분석을 수행한다.

회귀분석에서 유의미한 영향을 미치는 독립변수들을 바탕으로 역사들에 대해 군집분석을 진행한다. 분석 결과를 통해 군집별 특성을 파악하고 미세먼지 저감을 위한 맞춤 대응 방안을 마련한다.

## Approach & Model

### Regression

선정된 독립변수가 역사 내 초미세먼지 농도에 미치는 영향 예측

	coef	std err	t	P> t	[0.025	0.975]
const	112.9631	3.629	31.13	0	105.851	120.075
전동차_초미세먼지농도	0.4005	0.004	94.056	0	0.392	0.409
역사외부_초미세먼지농도	0.2892	0.004	68.878	0	0.281	0.297
역사외부_미세먼지농도	0.3191	0.005	70.551	0	0.31	0.328
구별_녹지면적	-0.1558	0.002	-79.153	0	-0.16	-0.152
충승차수	0.0942	0.002	61.444	0	0.091	0.097
역사_배기	-0.2411	0.004	-65.445	0	-0.248	-0.234
인구밀도	-0.436	0.004	-107.243	0	-0.444	-0.428
노선명	-0.1717	0.005	-36.197	0	-0.181	-0.162
최고습도	0.0335	0.001	33.793	0	0.032	0.035
에스컬레이터수_합계	-0.1059	0.002	-60.661	0	-0.109	-0.103
준공년도	-13.9451	0.48	-29.074	0	-14.885	-13.005

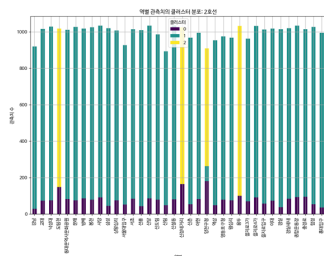
<그림 1. OLS Regression 결과>

각 독립변수의 회귀계수는 해당 변수가 한 단위 증가할 때 지하철역사의 초미세먼지 농도가 얼마큼 변화하는지를 해석하여 정책 활용에 사용하였다. *PyCaret* 을 사용하여 최적의 모델을 도출한 결과, Random Forest Regressor에서 MSE는 118.63, R-squared는 0.777로 나타났다.

### Clustering

군집별로 초미세먼지 농도에 영향을 주는 요인 판단

변수들이 서로 다른 단위와 범위를 가지므로 수치형 독립변수들 10개를 표준화하였다. *Elbow Method*를 이용해 군집의 개수를 3개로 선택해 *K-means Clustering*을 실시하였다.



<그림 2. 역별 클러스터 분포: 2호선>

동일한 역임에도 데이터가 일자에 따라 다른 군집에 속하여, 각 역별로 15% 이상의 관측치가 클러스터 0에 할당된 역은 클러스터 0에 속한다고 분류하였다. 클러스터 0와 1, 또는 0과 2에 중복 할당된 역은 중복된 요인에 대해 대응방안을 마련해야 함을 시사한다.

클러스터링의 성능을 평가하기 위해 실루엣 분석을 진행하였으며, *Silhouette Score*는 0.681로 나타났다.

## Datasets

역사 내 미세먼지 농도에 영향을 미칠 것으로 보이는 변수 선정

- 역사의 노후화 정도, 외부 대기질 농도, 유동인구, 날씨 데이터 등
  - 역사 내부 변수, 역사 외부 변수 데이터로 분류
- (데이터 출처: 공공데이터포털 및 서울열린데이터광장, 서울교통공사)

데이터 전처리: 데이터 병합, 결측치 처리, 범주형 변수 라벨링

변수 선택: *Stepwise Variable Selection*

- 전동차\_초미세먼지농도, 역사외부\_초미세먼지농도, 역사외부\_미세먼지농도, 구별\_녹지면적, 충승차수, 역사\_배기, 인구밀도, 노선명, 최고습도, 에스컬레이터수\_합계, 준공년도

T.air(역, 일평균)	역명	자치구	...	노선명	최고습도	에스컬레이터수_합계	준공년도
28	서울대입구	관악구	...	2	55.0955	8	1983
33.7	하계	노원구	...	7	95.45531	6	1996
29.9	잠실	송파구	...	8	9.706226	11	1980
21.2	버티고개	중구	...	6	26.99051	4	2001
17.9	종합운동장	송파구	...	2	9.706226	5	1980

<그림 3. 데이터셋 Overview>

## Results

*OLS Regression* 및 *Clustering* 결과를 바탕으로  
군집별 초미세먼지 대응 방안 마련

### Cluster 0

미세먼지 농도 높음. 외부 초미세먼지 농도가 높아 유입 통제 및 역사 외부의 공기질 개선 필요  
→ 역사 출입구 방풍문 설치 및 구별 녹지면적 확보

### Cluster 1

미세먼지 농도 중간, 전동차 내부 미세먼지 농도가 나빠져 지 역별 녹지 면적이 작고, 오래된 준공연도  
→ 외부 녹지 확보, 터널 본선 환기설비 개선, 노후된 환기시설 정비, 터널 내 자갈바닥 콘크리트 교체

### Cluster 2

미세먼지 농도 양호, 상대적으로 신설된 역과 높은 구별 녹지 면적, 그러나 최고 습도가 높음  
→ 역사 내 대형 제습기를 통한 습도 조절 및 공기청정