


# Assignment 2

데이터 탐색



# 과제 소개

- 미세먼지 농도와 바람과의 관계를 탐색해봅니다.
- 1시간 간격으로 측정된 미세먼지/바람 데이터가 주어집니다.
- 구름에서 채점되는 항목은 다음과 같습니다.
  - 이상치/결측치 제거
  - 상관관계 분석
  - 파생변수 생성 및 통계치 확인
- 분석의 편의를 위해 시각화 지침이 있는 코랩 파일이 주어집니다.
  - 시각화 내용은 채점하지 않습니다.

# 데이터셋

- 기상청 서울관측소에서 2017년부터 2019년까지 한 시간 단위로 측정한 데이터를 기반으로, 다음과 같은 항목이 있습니다.

| 변수명                | 설명            | 단위 (간격)   |
|--------------------|---------------|-----------|
| Measurement date   | 측정 일시         | 1시간       |
| Wind Speed         | 풍속            | m/s       |
| Wind Direction     | 풍향            | 각도 (16방위) |
| Atmo. Movement W/E | 서->동 방향 대기 이동 | km        |
| Atmo. Movement S/N | 남->북 방향 대기 이동 | km        |
| Visibility         | 시정 (가시거리)     | 10m       |

<weather.csv >

| 변수명              | 설명    | 단위 (간격)                  |
|------------------|-------|--------------------------|
| Measurement date | 측정 일시 | 1시간                      |
| PM10             | 미세먼지  | $\mu\text{g}/\text{m}^3$ |
| PM2.5            | 초미세먼지 | $\mu\text{g}/\text{m}^3$ |

<air\_pollution.csv >

# 1

## 파일 불러오기 & 결측치/이상치 제거

- 미세먼지 데이터와 기상 데이터를 Pandas로 불러옵니다.
- 관측일시 기준으로 merge해 사용할 데이터셋을 만듭니다.
- 미세먼지와 초미세먼지에서 이상치를 제거합니다.
  - PM10 또는 PM2.5값이  $500.0\mu\text{g}$  이상인 경우를 이상치로 봅니다.
- 기상 데이터에서 결측치를 제거합니다.
  - 하나라도 결측치가 있는 행은 분석에서 제외합니다.
- 마지막으로 인덱스를 초기화합니다.

## 2

# 상관관계 분석

- 미세먼지와 기상 데이터 간 상관관계를 파악합니다.
  - 시각화를 활용하면 좋습니다.
- 미세먼지 데이터를 열로, 기상 데이터를 행으로 하는 상관관계 데이터프레임을 출력해야 합니다.
- 문제를 푼 뒤 한번 생각해 보세요.
  - 선형적 상관관계를 계산하는 것이 부적합한 열이 있을까요?
  - 상관계수가 인과관계를 알려줄 수 있을까요?

# 3

## 파생변수 & 통계치 분석

- 초미세먼지를 단계 별로 분리한 열을 추가합니다.
- 환경부 초미세먼지 등급 기준을 이용합니다.
- 초미세먼지 등급 별로 다른 변수들의 통계치를 분석합니다.
  - PM2.5 등급 별 다른 수치들의 평균치를 출력하세요.

| 구분 | 좋음       | 보통          | 나쁨          | 매우 나쁨 |
|----|----------|-------------|-------------|-------|
| 기준 | 0이상 15미만 | 15 이상 35 미만 | 35 이상 75 미만 | 75 이상 |



# 시각화

- 시각화 지침이 담긴 코랩 파일이 아이캠퍼스에 있습니다.
- 구름의 Assignment 2-1번 문제를 푼 뒤, 해당 문제의 전처리 과정을 사용해서 시각화를 진행하세요.
- 시각화 내용은 채점되지 않습니다.



# 데이터셋 출처

- [Kaggle "Air Pollution in Seoul"](#)
- [기상청 종관기상관측](#)