EDATHON – 2026

Dataset Overview

Dataset Name

Pune Smart City Environmental Sensor Dataset (2019)

Data Source

- Pune Smart City Development Corporation Limited (PSCDCL)
- IISc Bangalore
- Year: **2019**

Objective

The dataset captures **environmental and atmospheric sensor readings** across multiple locations in Pune city.

This dataset enables:

- Monitoring **urban air quality**
- Identifying **pollution hotspots**
- Supporting **data-driven policy decisions**
- Improving **citizen health & living conditions**

Dataset Structure

Number of Records

103,205 rows

Number of Features

28 columns

Each row represents **sensor readings at a specific location and timestamp**.

Column Description (Feature Dictionary)

1.) Location Information

Column	Description
NAME	Monitoring station / area name
Attitude	Latitude of sensor
Longitude	Longitude of sensor

2.) Air Pollutants (Core AQI Drivers)

Column	Description
PM2_MAX, PM2_MIN	PM2.5 particulate matter
PM10_MAX, PM10_MIN	PM10 particulate matter
NO_MAX, NO_MIN	Nitric Oxide
NO2_MAX, NO2_MIN	Nitrogen Dioxide
SO2_MAX, SO2_MIN	Sulfur Dioxide
CO_MAX, CO_MIN	Carbon Monoxide
CO2_MAX, CO2_MIN	Carbon Dioxide
OZONE_MAX, OZONE_MIN	Ozone

3.) Atmospheric & Environmental Parameters

Column	Description
HUMIDITY	Relative humidity (%)
TEMPRATURE_MAX, TEMPRATURE_MIN	Temperature (°C)
AIR_PRESSURE	Atmospheric pressure
UV_MAX, UV_MIN	Ultraviolet radiation
LIGHT	Ambient light intensity
SOUND	Noise levels (dB)

4.) Time Information

Column	Description
LASTUPDATEDATETIME	Timestamp of data capture

EDA Problem Satatement

These questions ensure students understand the dataset before visualization.

Q1.

- How many records and features are present in the dataset?
- What percentage of data is missing in each column?

Q2.

- Are there any sensor readings with zero or unrealistic values?
- Which parameters show maximum data quality issues?

Q3.

- How many unique monitoring locations are present?
- Which locations have the highest number of readings?

Q4.

- What is the distribution of key air pollutants (PM2.5, PM10, NO_2, CO)?

Q5.

- What is the overall distribution of environmental factors like temperature, humidity, sound, and light?

Q6.

- Which pollutants show the highest variability across Pune?

Q7.

- How does PM2.5 vary with temperature and humidity?

Q8.

- Is there a relationship between traffic-related pollution (CO, NO_2) and sound levels?

Q9.

- How do PM10 levels change with air pressure?

Q10.

- Which pollutants are most strongly correlated with each other?

Q11.

- Do atmospheric parameters show any strong correlation with pollutant levels?

Q12.

- Which 5 parameters should be prioritized for AQI monitoring based on correlation strength?

Q13.

- Which locations consistently show higher pollution levels?

Q14.

- How does pollution differ between:
  - Railway stations
  - Bus stops
  - IT hubs
  - Residential areas

Q15.

- Are there pollution hotspots in Pune based on latitude and longitude?

Q16.

- How does air quality vary over time (hourly / daily)?

Q17.

- Are there specific times of the day when pollution peaks?

Q18.

- Is there a noticeable difference in pollution between weekdays and weekends?

Q19.

- Rank the top 5 most polluted locations based on PM2.5 and PM10.

Q20.

- Which pollutants exceed safe limits most frequently?

Q21.

- Which locations show the most stable vs unstable pollution levels?

Q22.

- What are the top 3 environmental risks identified from the data?

Q23.

- If city authorities want to take immediate action, which locations should be prioritized and why?

Q24.

- Which environmental parameter acts as an early warning indicator for poor air quality?

Q25.

- Can we cluster locations based on pollution behavior?