

[← Go Back to Data Analysis & Visualization](#)[☰ Course Content](#)

## Session Problem Statement #1 - Air Pollution Data Exploration

### Dimensionality Reduction - Air Pollution Dataset

#### Context:

Air pollution is all around us. Indoors, outdoors, in cities, and in the countryside. It affects us all, whether we realize it or not. The rise in permissible concentrations of various pollutants observed over the last few decades contributes to rising pollution levels. As a result, affecting the weather conditions in an unfavorable manner and the formation of smog phenomena. In fact, the affected air quality level in the atmosphere has a negative impact on individual health and may cause an economic imbalance. Diseases caused by rising pollution levels are one of the major issues confronting urban settlements.

In this case study, we will explore and visualize the Air Pollution dataset, which contains information about air molecules and pollutants found in the air.

#### Objective:

The objective of this problem is to reduce the number of features by using dimensionality reduction techniques like PCA and t-SNE and extracting insights.

#### Data Dictionary:

- This dataset contains about 13 months of data on major pollutants and meteorological levels of a city.
- The data shows average readings for Nitric Oxide(NO), Nitrogen Dioxide(NO2), Oxides of Nitrogen, Ozone, Particulate Matter (PM10 and PM2.5), Sulphur Dioxide, Benzene, Toulene, and P\_Xylene. Measured in Micrograms per Cubic Meter of Air (ug/m3). Also shows few generic features WindSpeed, VerticalWindSpeed, BarPressure, Weather,Temp, and WindDirection.
- The PD denotes the Previous Day. In this dataset, we have 5 columns that are Previous Date values of correponsing columns in the dataset. They are PD\_PM10, PD\_NO2, PD\_SO2, PD\_CO, PD\_PM2.5.
- The Date is the feature at which the data was recorded.

[← Previous](#)[Next →](#)