

## Air quality

### Dataset:

Dataset Air quality of an Italian city. The dataset contains 9358 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multi-sensor device. The device was located on the field in a significantly polluted area, at road level, within an Italian city. Data were recorded from March 2004 to February 2005 (one year) representing the longest freely available recordings of on field deployed air quality chemical sensor devices responses. Ground Truth hourly averaged concentrations for CO, Non Metanic Hydrocarbons, Benzene, Total Nitrogen Oxides (NO<sub>x</sub>) and Nitrogen Dioxide (NO<sub>2</sub>) and were provided by a co-located reference certified analyzer. Evidences of cross-sensitivities as well as both concept and sensor drifts are present as described in De Vito et al., Sens. And Act. B, Vol. 129,2,2008 (citation required) eventually affecting sensors concentration estimation capabilities. Missing values are tagged with -200 value. "

Attributes of the dataset are 0 Date (DD/MM/YYYY)

1 Time (HH.MM. SS)

2 True hourly averaged concentration CO in mg/m<sup>3</sup> (reference analyzer)

3 PT08.S1 (tin oxide) hourly averaged sensor response (nominally CO targeted)

4 True hourly averaged overall Non Metanic Hydro Carbons concentration in microg/m<sup>3</sup> (reference analyzer)

5 True hourly averaged Benzene concentration in microg/m<sup>3</sup> (reference analyzer)

6 PT08.S2 (titania) hourly averaged sensor response (nominally NMHC targeted)

7 True hourly averaged NO<sub>x</sub> concentration in ppb (reference analyzer)

8 PT08.S3 (tungsten oxide) hourly averaged sensor response (nominally NO<sub>x</sub> targeted)

9 True hourly averaged NO<sub>2</sub> concentration in microg/m<sup>3</sup> (reference analyzer)

10 PT08.S4 (tungsten oxide) hourly averaged sensor response (nominally NO<sub>2</sub> targeted)

11 PT08.S5 (indium oxide) hourly averaged sensor response (nominally O<sub>3</sub> targeted)

12 Temperature in Â°C 13 Relative Humidity (%) 14 AH Absolute Humidity

Requirements:

1. Perform exploratory Data Analysis and come up with actionable insights from each plot and table. Do data wrangling if required.
2. Perform all the Hypothesis testing possible and conclude from each test.
3. Predict **Relative Humidity** of a given point of time based on the all other attributes affecting the change in RH. Do use Gridsearch Cross validation for any algorithms you use.

Please perform the entire analysis in **Python**