

Water Treatment Plant

This data set was pulled from [UCI](#). It also contains the following information for each sample:

- 0 Date (date of sample)
- 1 Q-E (input flow to plant)
- 2 ZN-E (input Zinc to plant)
- 3 PH-E (input pH to plant)
- 4 DBO-E (input Biological demand of oxygen to plant)
- 5 DQO-E (input chemical demand of oxygen to plant)
- 6 SS-E (input suspended solids to plant)
- 7 SSV-E (input volatile suspended solids to plant)
- 8 SED-E (input sediments to plant)
- 9 COND-E (input conductivity to plant)
- 10 PH-P (input pH to primary settler)
- 11 DBO-P (input Biological demand of oxygen to primary settler)
- 12 SS-P (input suspended solids to primary settler)
- 13 SSV-P (input volatile suspended solids to primary settler)
- 14 SED-P (input sediments to primary settler)
- 15 COND-P (input conductivity to primary settler)
- 16 PH-D (input pH to secondary settler)
- 17 DBO-D (input Biological demand of oxygen to secondary settler)
- 18 DQO-D (input chemical demand of oxygen to secondary settler)
- 19 SS-D (input suspended solids to secondary settler)
- 20 SSV-D (input volatile suspended solids to secondary settler)
- 21 SED-D (input sediments to secondary settler)
- 22 COND-D (input conductivity to secondary settler)
- 23 PH-S (output pH)
- 24 DBO-S (output Biological demand of oxygen)
- 25 DQO-S (output chemical demand of oxygen)
- 26 SS-S (output suspended solids)
- 27 SSV-S (output volatile suspended solids)
- 28 SED-S (output sediments)
- 29 COND-S (output conductivity)
- 30 RD-DBO-P (performance input Biological demand of oxygen in primary settler)
- 31 RD-SS-P (performance input suspended solids to primary settler)
- 32 RD-SED-P (performance input sediments to primary settler)
- 33 RD-DBO-S (performance input Biological demand of oxygen to secondary settler)
- 34 RD-DQO-S (performance input chemical demand of oxygen to secondary settler)
- 35 RD-DBO-G (global performance input Biological demand of oxygen)
- 36 RD-DQO-G (global performance input chemical demand of oxygen)
- 37 RD-SS-G (global performance input suspended solids)
- 38 RD-SED-G (global performance input sediments)

This dataset is good for clustering.

You can use the following code to load the data into google colab:

```
import pandas as pd
url = "https://raw.githubusercontent.com/the-codingschool/TRAIN/main/water%20treatment%20plant/water.csv"
df = pd.read_csv(url,
                 names = ["Date", "Q-E", "ZN-E", "PH-E", "DBO-E", "DQO-E", "SS-E", "SSV-E", "SED-E",
                          "COND-E", "PH-P", "DBO-P", "SS-P", "SSV-P", "SED-P", "COND-P", "PH-D", "DBO-D",
                          "DQO-D", "SS-D", "SSV-D", "SED-D", "COND-D", "PH-S", "DBO-S", "DQO-S", "SS-S",
                          "SSV-S", "SED-S", "COND-S", "RD-DBO-P", "RD-SS-P", "RD-SED-P", "RD-DBO-S",
                          "RD-DQO-S", "RD-DBO-G", "RD-DQO-G", "RD-SS-G", "RD-SED-G"])
```