

# Water Quality Dataset Summary and Analysis

Harsh Vishwakarma, Ayan kumar Singh

September 9, 2025

## 1 Dataset Overview

- **Shape of dataset:** 19 rows  $\times$  27 columns
- **Context:** Water quality measurements from multiple Yamuna River monitoring stations across different states.
- **Key Parameters:** Temperature, Dissolved Oxygen (DO), pH, Conductivity, Biochemical Oxygen Demand (BOD), Nitrate/Nitrite levels, Fecal Coliform, Total Coliform.

## 2 Column Names

| Column                                    | Description / Standard Range                 |
|---|--|
| STATION CODE                              | Unique station identifier                    |
| LOCATIONS                                 | Sampling location                            |
| STATE                                     | State name                                   |
| TEMPERATURE (°C) : Min/Max/Mean           | Water temperature                            |
| D.O. (mg/l) : Min/Max/Mean                | Dissolved Oxygen; standard > 4 mg/l          |
| pH : Min/Max/Mean                         | Standard range: 6.5–8.5                      |
| CONDUCTIVITY (µmhos/cm) : Min/Max/Mean    | Electrical conductivity                      |
| B.O.D. (mg/l) : Min/Max/Mean              | Biochemical Oxygen Demand; standard < 3 mg/l |
| NITRATE+NITRITE-N (mg/l) : Min/Max/Mean   | Nutrient levels                              |
| FECAL COLIFORM (MPN/100ml) : Min/Max/Mean | Standard < 2500 MPN/100ml                    |
| TOTAL COLIFORM (MPN/100ml) : Min/Max/Mean | Standard < 5000 MPN/100ml                    |

### 3 Sample Data (First 5 Rows)

| Station Code | Location                          | State            | Temp Min (°C) | Temp Max (°C) | Temp Mean (°C) |
|--------------|-----------------------------------|------------------|---------------|---------------|----------------|
| 1494         | Yamuna at U/S of Lakhwar Dam, U.P | Uttarakhand      | 18.0          | 21.0          | 19.7           |
| 1490         | Yamuna at U/S Dak Patthar, U.P    | Uttarakhand      | 17.0          | 21.0          | 18.7           |
| 1553         | Yamuna, U/S Paonta Sahib, H.P     | Himachal Pradesh | 15.0          | 28.0          | 22.3           |
| 1554         | Yamuna, D/S Paonta Sahib, H.P     | Himachal Pradesh | 17.0          | 29.0          | 23.7           |
| 1120         | Yamuna at Wazirabad, Delhi, CPCB  | Delhi            | —             | —             | —              |

### 4 Missing Values Summary

| Column                           | Missing Count |
|----------------------------------|---------------|
| TEMPERATURE (°C) : Min           | 11            |
| TEMPERATURE (°C) : Max           | 11            |
| TEMPERATURE (°C) : Mean          | 11            |
| D.O. (mg/l) : Min                | 7             |
| D.O. (mg/l) : Max                | 7             |
| pH : Min                         | 7             |
| pH : Max                         | 7             |
| CONDUCTIVITY (µmhos/cm) : Min    | 13            |
| B.O.D. (mg/l) : Min              | 7             |
| NITRATE+NITRITE-N (mg/l) : Min   | 2             |
| FECAL COLIFORM (MPN/100ml) : Min | 4             |
| TOTAL COLIFORM (MPN/100ml) : Min | 0             |

### 5 Summary Statistics (Selected Parameters)

| Parameter                       | Count | Mean  | Std Dev | Min   | 25%   | Max    |
|---------------------------------|-------|-------|---------|-------|-------|--------|
| Temperature Min (°C)            | 8     | 16.23 | 2.07    | 14.75 | 16.00 | 19.80  |
| Temperature Max (°C)            | 8     | 28.17 | 2.59    | 25.05 | 28.50 | 34.00  |
| DO Mean (mg/l)                  | 12    | 5.25  | 2.92    | 1.20  | 3.13  | 9.30   |
| pH Mean                         | 12    | 7.73  | 0.30    | 7.50  | 7.60  | 8.20   |
| BOD Mean (mg/l)                 | 12    | 8.93  | 10.95   | 1.10  | 1.45  | 38.20  |
| Fecal Coliform Mean (MPN/100ml) | 15    | 40827 | 77298   | 2250  | 4900  | 282000 |

|                                 |    |        |        |      |       |        |
|---------------------------------|----|--------|--------|------|-------|--------|
| Total Coliform Mean (MPN/100ml) | 19 | 2.92e6 | 7.05e6 | 6754 | 33000 | 2.38e7 |
|---------------------------------|----|--------|--------|------|-------|--------|

---

## 6 Observations

- Several stations exceed permissible limits for Fecal and Total Coliform.
- Dissolved Oxygen levels vary widely; some readings fall below the safe threshold.
- BOD values at certain stations indicate high organic pollution.
- Missing data is significant for temperature and conductivity parameters.