waterquality questions

Q1

Given a dataset with water parameters, your task is to compute WQI (Water Quality Index).
Formula for WQI:

$$WQI = \sum_p (P_p \times W_p)$$

where (P) is a parameter.

Steps for Calculating Parameters and Weights for Each Parameter

Step 1: Compute Parameters

For each station code, compute the following **7 parameters** using the given formula:

Parameters to compute:

1. Temperature (T)
2. pH
3. Conductivity
4. BOD (Biochemical Oxygen Demand)
5. Nitrate + Nitrite
6. Fecal Coliform
7. Total Coliform

Formula to use:
For each parameter (P), calculate its normalized value using the formula:

$$P = \frac{(P_{mean} - P_{\min})}{(P_{\max} - P_{\min})} \times 100$$

For example, for Temperature (T):

$$T = \frac{(T_{mean} - T_{\min})}{(T_{\max} - T_{\min})} \times 100$$

Step 2: Weights Calculation for Each Parameter

After applying normalization in **Step 1**, follow these steps to determine the weights for each parameter:

Example:

1. **Combine the data from all newly created parameter columns.**
   - Example: After normalization, assume the combined data looks like this:

     ```
     [78, 65, 92, 85, 70, 88, 90, 60, 75, 80, 95, 82, 76, 87, 73, 89, 66, 77]
     ```

2. **Sort the combined data in descending order.**
   - Sorted data:

     ```
     [95, 92, 90, 89, 88, 87, 85, 82, 80, 78, 77, 76, 75, 73, 70, 66, 65, 60]
     ```

3. **Select weights based on the first 7 multiples of 4 from the sorted data.**
   - The **7 multiples of 4** are positions **4, 8, 12, 16, 20, 24, and 28**.
   - Extracted weights from those positions:

     ```
     [89, 82, 76, 66, (next available values if list is short)]
     ```

4. **Assign weights by applying lexicographical sorting on the parameter names.**
   - **Given Parameters (Before Sorting):**
     1. Temperature (T)
     2. pH
     3. Conductivity
     4. BOD (Biochemical Oxygen Demand)
     5. Nitrate + Nitrite

6. Fecal Coliform
7. Total Coliform

- **Sorted lexicographically:**

1. **BOD (Biochemical Oxygen Demand)** → Weight **89**
2. **Conductivity** → Weight **82**
3. **Fecal Coliform** → Weight **76**
4. **Nitrate + Nitrite** → Weight **66**
5. **pH** → Weight **(next extracted value)**
6. **Temperature (T)** → Weight **(next extracted value)**
7. **Total Coliform** → Weight **(next extracted value)**

Task:

1. **Calculate WQI** using the given formula.
2. **Normalize WQI** on a **0-100 scale**.
3. **Identify unsafe stations** where **WQI < 50**.
4. **Map **Station Codes** to **Pincodes** using pincode_station.db.
5. **Output the final answer**: A **comma-separated list** of **Pincodes** where the water is unsafe to drink.

Q2

Get the station with least WQI and figure out what would be change in WQI if BOD parameter is increased by 20% keeping eveyother parameter constant.

$$\text{Percentage Change} = \frac{\text{New WQI} - \text{Original WQI}}{\text{Original WQI}} \times 100$$

Start coding or generate with AI.

Q3

Continuing **Question 1**, we have a `pincode_coordinates.parquet` file that maps Pincodes to their respective Latitudes and Longitudes.

For each **Station Code**, find its corresponding **Pincode** from `pincode_station.db`. Then, retrieve the **Latitude and Longitude** for each Pincode from `pincode_coordinates.parquet` and determine the **closest pair of cities** based on their coordinates.

Start coding or generate with AI.

Q4

Create a Fast API Get endpoint, take `station_code` as URL encoded parameter and in response get the json with following fields State_name,district_name,pincode.

Use LLM function calling for this question.

example : `{ "state_name": "Assam", "district_name": "Dhubri", "pincode": "783301" }`

Note: If district_name can't be fetched then return null in district_name field.

Start coding or generate with AI.

patent questions :

Q1

Given a dataset Which patent(s) have the highest number of claims? Cite the patent ID. If you have multiple patents with the same number of claims, list them all separated by commas without spaces. The parameter file without the extension is assumed as the Patent ID.E.g. For the below data Patent ID is US20230225250A1-20230720

Start coding or generate with AI.

Q2

Start coding or generate with AI.