1. **Community:** The name of the local community or settlement.
2. **Region:** The administrative region in Ghana.
3. **Latitude:** The geographical latitude coordinate of the community.
4. **Longitude:** The geographical longitude coordinate of the community.
5. **Water Quality:** This is the target variable that we are trying to predict. 0 = unsafe for drinking, while 1 = safe for drinking.
6. **Distance to the Nearest River:** This is in kilometers (km). This is the distance to the nearest river or major water source.
7. **Is Mining Zone:** Indicates whether the village/location is in, or within the vicinity of, a mining zone. 0 = not a mining zone, while 1 = located in a mining area.
8. **Contamination Level:** A number ranging from 0-10. 0 = no contamination, while 10 = extremely contaminated.
9. **Contamination Type:** This variable indicates the type of contaminant. This includes categories like “None”, “Heavy Metals”, “Bacteria”, etc.
10. **Water Source:** This variable indicates the primary water source. For example, “River”, “Borehole”, “Pipe-Borne”, etc.
11. **Population:** This is a numerical value that indicates the total, or estimated total, population of the community.
12. **Number of Children:** This is an estimation of the number of children, less than 18 years old, in the community.
13. **Average Daily Water Needs:** This is in liters. This is an estimation of the total daily water needs of the community.
14. **Urban/Rural:** This variable indicates whether the location of the community is within an urban or rural area.
15. **Average Household Income:** This is the estimated average household income within the community in Ghanaian Cedis.
16. **Education Level:** The average number of years of formal education within the community.
17. **Prevalence of Waterborne Diseases:** Do waterborne diseases such as cholera and dysentery exist within the community? 0 = no, while 1 indicates yes.
18. **Water Access Score:** A rating of how accessible water is. 1 = very poor, 10 = excellent.
19. **Accessibility:** The type of road access. For example, asphalt/tar road, dirt road, only accessible by boat, etc.
20. **Sanitation Facilities Available:** Whether sanitation facilities are available. 0 = none available, 1 = available. The sanitation facilities include toilets and other waste disposal facilities.
21. **Government Intervention Present:** Whether government intervention for water projects is available. 0 = none, 1 = government water/sanitation projects are active.
22. **NGO Presence:** Whether NGO are actively pursuing water/sanitation projects in the community. 0 = none, 1 = yes.
23. **Year Data Collected:** The year in which the data was collected.