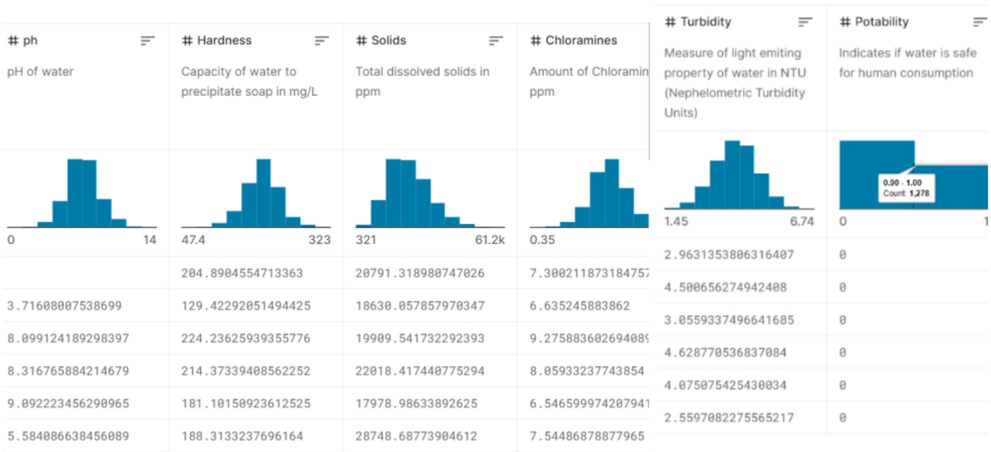
Project Title: Water Quality Analysis
Phase 1: Project Definition and Design Thinking

<u>Project Definition:</u>

The project involves analyzing water quality data to assess the suitability of water for specific purposes, such as drinking. The objective is to identify potential issues or deviations from regulatory standards and determine water potability based on various parameters. This project includes defining analysis objectives, collecting water quality data, designing relevant visualizations, and building a predictive model.



<u>Design Thinking:</u>
<u>Analysis Objectives:</u>

Objective 1: Analyzing water quality data:

Analyzing water quality data is a multifaceted process crucial for safeguarding public health and the environment. In-depth analysis can identify contamination sources, assess risks, and inform mitigation strategies. Continuous monitoring and transparent reporting are essential to track water quality over time. Assessing potability in water quality data entails a comprehensive examination of physical, chemical, and microbiological parameters. Data must adhere to regulatory standards, with particular attention to indicators of contamination like coliform bacteria and chemical concentrations such as heavy metals and disinfection byproducts. Detection of waterborne pathogens and monitoring taste, odor, and appearance are essential. Temperature and turbidity extremes warrant investigation, and historical data analysis reveals trends.

Objective 2: Identifying deviations from standards

In a water quality project, identifying deviations from established standards is paramount for ensuring the safety and compliance of the water supply. Begin by referencing regulatory standards and collecting comprehensive data covering various parameters. Analyze the data, utilizing statistical techniques and graphical representations to pinpoint values exceeding permissible limits. Special attention should be given to critical parameters affecting water quality. Investigate potential causes, considering seasonal variations, and assess associated risks.

Objective 3: Understanding parameter relationships.
In a water quality project, comprehending parameter relationships is pivotal for deciphering the intricate dynamics of water systems. Gathering diverse data on physical, chemical, and microbiological parameters provides the foundation. Statistical analyses unveil correlations, illuminating how variations in one parameter impact others. Visualization aids in pattern recognition, while causality investigations reveal direct influences. Over time, these relationships can shift, warranting ongoing monitoring.

Data Collection:
The following is a general overview of the data collection process for water quality analysis:

1. Identify Sampling Locations:
Determine where to collect water samples. This could be from natural water bodies like rivers, lakes, or oceans, or from man-made sources like water treatment plants or industrial discharge points.

2. Select Parameters:
1. ph: pH of 1. water (0 to 14).
2. Hardness: Capacity of water to precipitate soap in mg/L.
3. Solids: Total dissolved solids in ppm.
4. Chloramines: Amount of Chloramines in ppm.
5. Sulfate: Amount of Sulfates dissolved in mg/L.
6. Conductivity: Electrical conductivity of water in μS/cm.
7. Organic_carbon: Amount of organic carbon in ppm.
8. Trihalomethanes: Amount of Trihalomethanes in μg/L.
9. Turbidity: Measure of light emiting property of water in NTU.
10. Potability: Indicates if water is safe for human consumption. Potable -1 and Not potable -0

3. Choose Sampling Equipment:
Depending on the parameters you're measuring, you may need various equipment, such as pH meters, conductivity meters, turbidity meters, sampling bottles, thermometers, and more.

Visualization Strategy:

1. Data Integration: Grafana supports a wide range of data sources, including databases, time series databases like InfluxDB, Prometheus, and cloud platforms like AWS and Azure. You can integrate your water quality data source with Grafana, making it a centralized hub for your data.
2. Dashboard Creation: Grafana provides a user-friendly interface for creating dashboards. You can design customized dashboards to display water quality data, parameter distributions, correlations, and other relevant information.
3. Visualization Options: Grafana offers various visualization options, including time series graphs, bar charts, pie charts, gauges, and more.
4. Data Exploration: Grafana allows users to interactively explore data by zooming in on specific time periods, selecting regions of interest, or filtering based on parameters. This feature aids in in-depth analysis.
5. Integration with Other Tools: Grafana can be integrated with data storage, processing, and analysis tools to create a comprehensive data analysis pipeline for your water quality project.

Predictive Modeling:

Predicting water potability is an important task that involves classification since it typically falls into categories like "potable" or "non-potable." Here's a general approach to deciding on machine learning algorithms and features:

1. Algorithms: Consider classification algorithms like Logistic Regression, Random Forest, or Support Vector Machines for predicting water potability.
2. Features: Include water quality parameters such as pH levels, turbidity, hardness, chloride concentration, sulfate levels, and any other relevant attributes.
3. Data Preprocessing: Prioritize data cleaning to handle missing values and outliers, ensuring the dataset's quality.
4. Model Evaluation: Use cross-validation techniques to assess the performance of selected algorithms and choose the one that performs best.
5. Continuous Monitoring: After deployment, continuously monitor the model's accuracy and retrain it periodically with updated data to maintain its predictive power.

Summarized Questions:

1.Can we predict the pH level of water based on other water quality parameters such as hardness, sulfate, and organic carbon?
2.Can we predict the concentration of chloramines in water based on turbidity and temperature?
3.What factors contribute most to the variation in chloramines levels, and can we create a predictive model?
4. How does electrical conductivity relate to the overall water quality and potability?
5.Can we predict the potability of water (safe for human consumption or not) based on a combination of all the provided water quality parameters?