# Assignment 1

You are required to use water quality check data as reference data to implement a regression model (such as Ridge or Lasso). Additionally, you need to utilize MLflow for logging, tracking, and ensuring model compliance.

## Data Collection and Preprocessing:

- Obtain a dataset related to water quality measurements.
- Clean and preprocess the data, handling any missing values or outliers.
- Perform exploratory data analysis (EDA) to understand the data distribution and relationships.

## Model Implementation:

- Split the data into training and testing sets.
- Implement two regression models: Ridge Regression and Lasso Regression.
- Compare the performance of both models using appropriate metrics (e.g., RMSE, $R^2$)

## MLflow Integration:

- Set up MLflow in your development environment.
- Log the following details using MLflow:
- Parameters used for model training (e.g., alpha values for Ridge and Lasso).
- Performance metrics of the models.
- Model artifacts (trained models)

## Model Compliance and Documentation:

- Ensure the models comply with the specified requirements and document the compliance process.
- Provide a detailed report on the implementation, including:
- Description of the dataset and preprocessing steps.
- Explanation of the regression models and their parameters.
- Results of the model comparisons.

- Screenshots or logs from MLflow showing the tracking and logging of experiments

**Deliverables:**

- A Jupyter Notebook or Python script containing the entire workflow from data preprocessing to model implementation and evaluation.
- A detailed report in PDF format.
- MLflow logs and screenshots demonstrating the tracking and logging process.