**CHAPTER -1**

**INTRODUCTION**

**1.1 PROJECT OVERVIEW**

The focus of municipalities has been on the supply of sufficient water quantities to the public with less attention paid to water quality. The deteriorating quality of raw water sources necessitates increased attention to water quality with professional scientists playing a central role at municipalities and water boards together with professional engineers. With many stringent regulations on the quality of drinking water and recreational water bodies, the young municipal chemist needs a handy manual to assist in the often neglected and complicated field of municipal water management.

Grounded in Science, Introduction to Municipal Water Quality Management not only links theory and regulations in practice but also offers simple numerical examples to better understand the rules and encourage a quantitative application to everyday problems. Developed from a series of lectures between 2015 and 2019, Introduction to Municipal Water Quality Management will give young professionals the confidence to analyse their results and apply their knowledge in a numerical fashion.

**1.2 PURPOSE**

A municipal water system is a public water supply system that includes a municipal water treatment plant, storage facilities such as water tanks, towers, and reservoirs, and a water pipe network for delivering treated water to residential and business customers.Municipalities require clean water supply. Several technological processes are used in the municipal water treatment process to develop functional municipal water systems.

Here's how municipal water systems ensure that homes and businesses in your area have safe drinking water.

When water is gathered, including supplies from streams and rivers, primary treatment takes place. Physical pollutants are removed from water during first treatment.

Secondary treatment involves the removal of fine particles and pollutants via coagulation, filtering, and other methods. The key distinction between primary and secondary wastewater treatment is that primary treatment focuses mostly on physical cleaning, whereas secondary treatment includes chemical treatments as well as microorganisms.

Tertiary treatment: pH adjustment, disinfection, and carbon treatments are applied in the final step, or tertiary treatment, at the municipal water plant.

**CHAPTER-2**

**IDEATION & PROPOSED SOLUTION**

**2.1 PROBLEM STATEMENT**

Continuously monitoring the quality of various waters is highly needed for safe drinking water practice and smart agriculture. The biological quality of water, which represents the presence and concentration of microorganisms in the water, is an important part of the task. The types of microorganisms in water can be a great number. Therefore, EPA recommends E. coli, which comes from human and animal wastes and is the most common form of fecal coliform, as the best indicator of health quality standards and are monitoring accordingly. By observing E. coli bacteria, the increase or decrease of many pathogenic bacteria can be estimated. Additionally, the presence of E.coli in a body of water may indicate that more harmful bacteria, viruses, or other microorganisms have contaminated that body of water. In Alabama, there are families using private wells as their primary water resource. Therefore, a device, which can be used by individuals without training for monitoring the quality of the water from the well in a real-time manner, is highly desirable. Alabama is rich in source water, such as Lake Martin and many rivers/creeks. The quality of source water is the key to ensure the quality of drinking water. Therefore, monitoring the quality of source water would be very critical. Additionally, the data from water quality monitoring is very critical to determine the source of pollution and would help local farmers use the land more efficient. Therefore, an inexpensive biosensor/technology that is suitable for field testing is urgently needed for monitoring quality of source water in a real-time manner. Due to the importance of water quality and the complexity of the problem, the EPA has implemented many technologies for monitoring water quality. However, all the technologies recommended by EPA are laboratory-based. That is, the water samples have to be delivered to a laboratory from the sources. The current EPA recommended technologies require the water samples to be delivered to a laboratory within 6 hours on ice. More importantly, the analysis of the water samples requires trained personnel and includes a 24-hour incubation period, which makes the analysis a time consuming process. In order to test the water quality in a real-time manner, it is believed that biosensors would be a strong candidate. Many types of biosensors have been developed or are under development. However, there is yet no biosensor suitable for in-field screening of water quality. This project was designed to develop a new device/methodology of monitoring water quality by detecting the concentration of E. coli in the water in a real-time manner. The device to be developed is based on the magnetostrictive particle (MSP) technology recently developed by the Principal Investigator. The device is ideal for in-field screening.

2.2 Empathy Map Canvas

2.3 IDEATION & BRAINSTORMING

**CHAPTER-3**

**REQUIREMENT ANALYSIS**

**3.1 FUNCTIONAL REQUIREMENTS**

**WATER SAMPLING**

Water used for processing fish, washing fish or making ice is supposed to meet drinking water standards if it is to be considered safe. Reason: contaminated water is the main cause for pathogen-loading of fish, posing a serious health hazard to its consumer.

WHO has issued guidelines for drinking water quality, a report in three volumes. Vol. 1 deals with guideline values, Vol. 2 deals with each contaminant and Vol. 3 gives information on how to handle water supplies in small rural communities. WHO recognizes that very stringent standards cannot be used universally and so a range of guideline values for more than 60 parameters have been elaborated. Most nations have their own guidelines or standards. The control exerted by local regulatory authorities may differ from place to place depending on the local situation. So how can acceptable water quality be defined? What can the harbour-master do to ensure quality? Ensuring the quality of the harbour basin when it is contiguous with estuarine or coastal waters is perhaps beyond the scope of the harbour-master except to ensure that activities in his harbour do not add to the pollution. However, he is duty-bound to ensure that the water used for drinking, cleaning fish, ice making and fish processing meets standards of potability set in his country.

Qualitative and quantitative measurements are needed from time to time to constantly monitor the quality of water from the various sources of supply. The harbour-master should then ensure appropriate water treatment within the fishery harbour complex as well as initiate remedial measures with the suppliers when water supply from outside is polluted.

Water sampling and analysis should be done by ISO-certified laboratories. Wherever laboratories available locally are not ISO-certified, it is advisable to get their quality assessed by an ISO-certified laboratory by carrying out collaborative tests to ensure that variation in the accuracy of results is sufficiently small. Unreliable results exacerbate problems of pollution when corrective action cannot be taken in time. Sampling and monitoring tests should be carried out by qualified technicians.

**3.1.1 BOREWELLS**

Contamination may arise from pollutants entering the water table some distance from the port or from sewage entering the borehole itself in the port area through cracked or corroded casings. In cases where overdrawing is evident (water is brackish), tests should be conducted at least monthly.

**3.1.2 MUNICIPAL MAIN**

Supply could be contaminated at source or through corroded pipelines leading to the fishery harbour. Mixing with sewage lines due to defective piping has been known to occur often. Complete tests should be carried out every half year, and the authorities should be informed when results indicate contamination.

**3.1.3 WATER TANKS AND RESERVOIRS**

Both types of structure are prone to bacterial growth if the residual chlorine levels in them are low or non-existent. Testing may not be necessary if periodic scrubbing is carried out. Bacteriological tests should be done at least half-yearly.

**3.1.4 HARBOUR BASIN WATER**

Typically, harbour basins are tested yearly. However, in areas where monsoons are very active, it may be advisable to test at the peak of the dry season when effluent point discharges tend to remain concentrated in the water body and again during the wet season when agriculture run-off may be considerable. Another critical period for harbours is the peak of the fishing season when the harbour is at its busiest and vessel-generated pollution is likely to be at its peak.

**3.1.5 TESTING PROCEDURES**

While the details of sampling, testing and analysis are beyond the scope of this handbook, what follows is a general description of the significance of water quality tests usually made.

Testing procedures and parameters may be grouped into physical, chemical, bacteriological and microscopic categories.

* Physical tests indicate properties detectable by the senses.
* Chemical tests determine the amounts of mineral and organic substances that affect water quality.
* Bacteriological tests show the presence of bacteria, characteristic of faecal pollution.

**3.1.6 PHYSICAL TESTS**

Colour, turbidity, total solids, dissolved solids, suspended solids, odour and taste are recorded.

Colour in water may be caused by the presence of minerals such as iron and manganese or by substances of vegetable origin such as algae and weeds. Colour tests indicate the efficacy of the water treatment system.

Turbidity in water is because of suspended solids and colloidal matter. It may be due to eroded soil caused by dredging or due to the growth of micro-organisms. High turbidity makes filtration expensive. If sewage solids are present, pathogens may be encased in the particles and escape the action of chlorine during disinfection.

Odour and taste are associated with the presence of living microscopic organisms; or decaying organic matter including weeds, algae; or industrial wastes containing ammonia, phenols, halogens, hydrocarbons. This taste is imparted to fish, rendering them unpalatable. While chlorination dilutes odour and taste caused by some contaminants, it generates a foul odour itself when added to waters polluted with detergents, algae and some other wastes.

**3.1.7 CHEMICAL TEST**

pH, hardness, presence of a selected group of chemical parameters, biocides, highly toxic chemicals, and B.O.D are estimated.

pH is a measure of hydrogen ion concentration. It is an indicator of relative acidity or alkalinity of water. Values of 9.5 and above indicate high alkalinity while values of 3 and below indicate acidity. Low pH values help in effective chlorination but cause problems with corrosion. Values below 4 generally do not support living organisms in the marine environment. Drinking water should have a pH between 6.5 and 8.5. Harbour basin water can vary between 6 and 9.

B.O.D.: It denotes the amount of oxygen needed by micro-organisms for stabilization of decomposable organic matter under aerobic conditions. High B.O.D. means that there is less of oxygen to support life and indicates organic pollution.

**3.1.8 BACTERIOLOGICAL TEST**

For technical and economic reasons, analytical procedures for the detection of harmful organisms are impractical for routine water quality surveillance. It must be appreciated that all that bacteriological analysis can prove is that, at the time of examination, contamination or bacteria indicative of faecal pollution, could or could not be demonstrated in a given sample of water using specified culture methods. In addition, the results of routine bacteriological examination must always be interpreted in the light of a thorough knowledge of the water supplies, including their source, treatment, and distribution.

Whenever changes in conditions lead to deterioration in the quality of the water supplied, or even if they should suggest an increased possibility of contamination, the frequency of bacteriological examination should be increased, so that a series of samples from well chosen locations may identify the hazard and allow remedial action to be taken. Whenever a sanitary survey, including visual inspection, indicates that a water supply is obviously subject to pollution, remedial action must be taken, irrespective of the results of bacteriological examination. For unpiped rural supplies, sanitary surveys may often be the only form of examination that can be undertaken regularly.

The recognition that microbial infections can be waterborne has led to the development of methods for routine examination to ensure that water intended for human consumption is free from excremental pollution. Although it is now possible to detect the presence of many pathogens in water, the methods of isolation and enumeration are often complex and time-consuming. It is therefore impractical to monitor drinking water for every possible microbial pathogen that might occur with contamination. A more logical approach is the detection of organisms normally present in the faeces of man and other warm-blooded animals as indicators of excremental pollution, as well as of the efficacy of water treatment and disinfection. The presence of such organisms indicates the presence of faecal material and thus of intestinal pathogens. (The intestinal tract of man contains countless rod-shaped bacteria known as coliform organisms and each person discharges from 100 to 400 billion coliform organisms per day in addition to other kinds of bacteria). Conversely, the absence of faecal commensal organisms indicates that pathogens are probably also absent. Search for such indicators of faecal pollution thus provides a means of quality control. The use of normal intestinal organisms as indicators of faecal pollution rather than the pathogens themselves is a universally accepted principle for monitoring and assessing the microbial safety of water supplies. Ideally, the finding of such indicator bacteria should denote the possible presence of all relevant pathogens.

Indicator organisms should be abundant in excrement but absent, or present only in small numbers, in other sources; they should be easily isolated, identified and enumerated and should be unable to grow in water. They should also survive longer than pathogens in water and be more resistant to disinfectants, such as chlorine. In practice, these criteria cannot all be met by any one organism, although many of them are fulfilled by coliform organisms, especially Escherichia coli as the essential indicator of pollution by faecal material of human or animal origin.

3.1.9 Investigative analysis

A harbour master's knowledge of the state of the environment in and around the fishing harbour goes a long way toward preventing outbreaks of contamination or disease with subsequent loss of resources and income. This is particularly so for the many small-to-medium fishing ports scattered around coastlines in developing countries, where, more often than not, environmental help and support from central bodies is meagre and very time-consuming.

The following is a true-life example of an investigative analysis carried out in an ASEAN country in a harbour that was experiencing problems with hygiene (coliform contaminated fish).

**3.1.10 TEST CASE**

The port in question is situated in the mouth of an estuary. The town's water supply cannot provide the port with potable water and the port draws groundwater from a series of boreholes in and around the port area. The port's storage infrastructure consists of only one elevated concrete tank which cannot be taken out of service for cleaning. Ice is supplied by outside contractors.

Current laboratory test results were examined and found to be too consistent to reflect natural changes in the environment, pointing a finger of suspicion at the laboratory's Quality Assurance. A new laboratory with I.S.O. certification was selected to carry out the new tests.

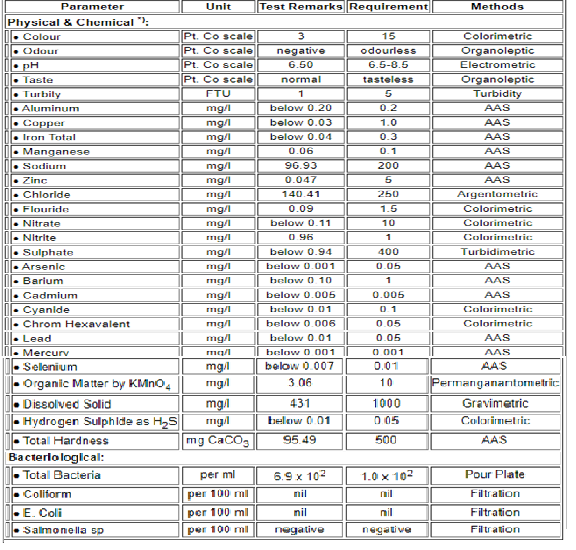
Water samples were taken by external technicians from the port's borehole, the auction hall's water taps, each and every one of the external ice suppliers and the harbour basin. A sample report from the laboratory is shown in Table 3.1.11.

In this table, the first column indicates the test parameter and the last column indicates the method used to determine the test result (sometimes, more than one method may be used to determine residuals).

The second column indicates how the parameters are measured, the third column gives the actual test result which may then be compared to the values in the fourth column. The values in the fourth column are national standards or limits set by Governments and may differ from country to country. The values in the third column should not exceed those in the fourth column.

Table 3.1.11 shows the recommended WHO standard limits for potable water.

Table 3.1.11: SAMPLE WATER ANALYSIS REPORT - PORT TAP WATER



\*) Standard Methods

A. Examination of the port's deep borehole test report revealed that whereas the iron and manganese levels were over the limit, indicating vegetable matter in the acquifer, the sodium and chloride levels were low, indicating that the pump was not overdrawing. Both the nitrate and nitrite levels were low indicating that sewage intrusion into the borehole casing was not a problem. The total bacterial count, however, was very high, indicating that the water has to be chlorinated to lower the count.

B. Examination of the auction hall's tap water test report (comparing them to the borehole water) indicates that the bacterial count is slightly lower but not enough to be considered sanitary and fit for drinking. The turbidity also dropped dramatically between borehole and tap, indicating deposition of solids inside the port's only storage tank. The nitrate level also drops as the nitrates are further converted to nitrites indicating bacteriological activity inside the overhead tank as well. As it turned out, chlorinating equipment was not installed.

C. Examination of the ice test reports reveals that both sodium and chlorides are over the limit indicating either leaking cans at the ice plants (dirty brine water enters the ice water during the chilling operation) or overdrawing at the plant's borehole. Closer examination also revealed that the nitrite levels are very high (indicating decomposed sewage) and that coliforms were present in the ice. This pointed a finger at the borehole of one particular plant, which in fact was found to be overdrawing water to meet an increase in demand. The presence of the coliforms also indicated that the ice plant's own chlorinating equipment was not functioning properly.

D. A close look at the river basin water indicated heavy contamination by sewage of the water course.

The conclusions to be drawn from the above exercise are that:

a) The most likely source of contamination was the ice supplied to the fishermen, which in turn contaminated the fish in the holds;

b) The port's own water supply and storage system was in need of an overhaul;

c) The port's river water was not to be used in any of the fish handling processes.

**3.2 NON-FUNCTIONAL REQUIREMENTS**

**3.2.1. Boiling**

The simplest method to purify water is to boil it for a good time. High temperatures cause the bacteria and virus to dissipate, removing all impurities from the water. In doing so, chemical additions cease to exist in the water. However, the dead micro-organisms and impurities settle at the bottom of the water, and boiling does not help eliminate all the impurities. You must strain the water through a microporous sieve to completely remove the impurities.

**3.2.2. Water Purifier**

An electric water purifier is the most trusted form of water purification found in most houses today. A water purifier uses a multi-stage process involving . UV and UF filtration, carbon block, and modern water filtration technology that eliminates most of the chemicals and impurities, making it the purest drinking water

**3.2.3. Reverse Osmosis**

An RO Purifier proves to be one of the best methods of purifying water. Reverse Osmosis forces water through a semipermeable membrane and removes contam

inants. The TDS controller and mineraliser Technology, like the one found in an A. O. Smith RO UV Water Purifier, help retain the necessary nutrients while doing away with harmful impurities.

**3.2.4. Water Chlorination**

It is an older technique used usually during an emergency, wherein a mild bleach with approximately 5% chlorine is added to the water. This mixture works as an oxidant and quickly kills microorganisms, making water safe for consumption.

**3.2.5. Distillation**

Distillation is a water purification process involving collecting the condensed water after evaporation, ensuring that water is free of contaminants. However, this isn’t Effective as an RO filter  because it is time-consuming and eliminates minerals.

**3.2.6. Iodine Addition**

Iodine is a red chemical that is easily available as a tablet or a liquid. It is extremely powerful as it kills bacteria and viruses. However, it adds an unpleasant taste and can be fatal if taken in high doses. Therefore, it should only be used if you don’t have access to a better method of purification like an electric water purifier

**3.2.7. Solar Purification**

An alternative to the . uv filtratuin  is solar purification which involves treating water with the ultraviolet radiation of the sun. The process involves filling a plastic bottle with water, shaking it to activate the oxygen and leaving it horizontally in the sunlight. This effectively kills bacteria and viruses present in the water, making it safe for consumption.

**3.2.8. Clay Vessel Filtration**

Way before people had access to an RO or UV Purifier, they used clay pots which purified muddy water, by blocking out the mud and allowing pure, potable water to pass through. This method is still used in some rural regions.

**3.2.9. UV Radiation**

Water is exposed to a UV Light that kills microorganisms, thereby preventing it from breeding further. But if not coupled with an RO Filter, UV Radiation alone cannot remove impurities and heavy metals.

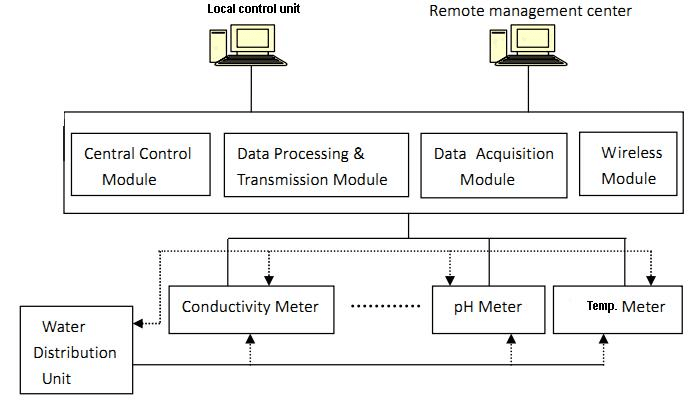
**3.2.10. Desalination**

This method is used when water with a certain level of salinity needs to be filtered. This process is helpful.

**CHAPTER-4**

**PROJECT DESIGN**

**4.1 DATA FLOW DIAGRAMS**



**4.2 SOLUTION & TECHNICAL ARCHITECTURE**

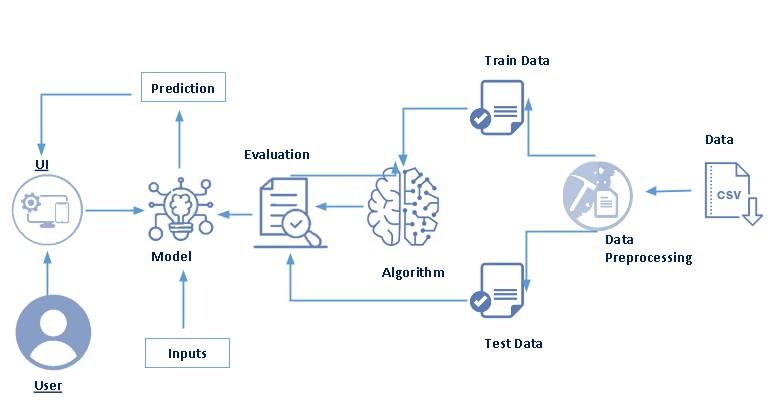
 Water quality is a critical issue that affects both human health and the environment. Our project aims to assess water quality in a specific region by examining various physical, chemical and biological parameters. We have collected water samples from different locations and measured parameters such as pH, Hardness, dissolved solids, chloramines, conductivity, sulfates, organic carbon, trihalomethanes, turbidity and potability.

Companies can develop and market products such as water filters, purifiers and treatment systems that can improve the quality of water for households, institutions and communities. Businesses can design and install systems that capture, treat and reuse wastewater, reducing the amount of water that is withdrawn from natural sources and helping to conserve water resources.

Access to clean and safe water is a basic human right and has significant social impacts. Poor water quality can lead to waterborne diseases, which can spread quickly and cause illness. Lack of access to clean water can also limit economic growth and development. Addressing water quality and ensuring access to safe water is crucial for promoting public health, reducing poverty and supporting sustainable development.

We will be using classification algorithms such as Logistic Regression, Decision tree, Random forest, KNN, and Xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pkl format. We will be doing flask integration and ibm development

**TECHNICAL ARCHITECTURE**



**4.3 USER STORIES**

**Drinking-water services**

Sustainable Development Goal target 6.1 calls for universal and equitable access to safe and affordable drinking water. The target is tracked with the indicator of “safely managed drinking water services” – drinking water from an improved water source that is located on premises, available when needed, and free from faecal and priority chemical contamination.

In 2020, 5.8 billion people used safely managed drinking-water services – that is, they used improved water sources located on premises, available when needed, and free from contamination. The remaining 2 billion people without safely managed services in 2020 included:

1.2 billion people with basic services, meaning an improved water source located within a round trip of 30 minutes;

282 million people with limited services, or an improved water source requiring more than 30 minutes to collect water;

368 million people taking water from unprotected wells and springs; and

122 million people collecting untreated surface water from lakes, ponds, rivers and streams.

**Water and health**

Contaminated water and poor sanitation are linked to transmission of diseases such as cholera, diarrhoea, dysentery, hepatitis A, typhoid and polio. Absent, inadequate, or inappropriately managed water and sanitation services expose individuals to preventable health risks. This is particularly the case in health care facilities where both patients and staff are placed at additional risk of infection and disease when water, sanitation and hygiene services are lacking. Globally, 15% of patients develop an infection during a hospital stay, with the proportion much greater in low-income countries.

Inadequate management of urban, industrial and agricultural wastewater means the drinking-water of hundreds of millions of people is dangerously contaminated or chemically polluted. Natural presence of chemicals, particularly in groundwater, can also be of health significance, including arsenic and fluoride, while other chemicals, such as lead, may be elevated in drinking-water as a result of leaching from water supply components in contact with drinking-water.

Some 829 000 people are estimated to die each year from diarrhoea as a result of unsafe drinking-water, sanitation and hand hygiene. Yet diarrhoea is largely preventable, and the deaths of 297 000 children aged under 5 years could be avoided each year if these risk factors were addressed. Where water is not readily available, people may decide handwashing is not a priority, thereby adding to the likelihood of diarrhoea and other diseases.

Diarrhoea is the most widely known disease linked to contaminated food and water but there are other hazards. In 2017, over 220 million people required preventative treatment for schistosomiasis – an acute and chronic disease caused by parasitic worms contracted through exposure to infested water.

In many parts of the world, insects that live or breed in water carry and transmit diseases such as dengue fever. Some of these insects, known as vectors, breed in clean, rather than dirty water, and household drinking water containers can serve as breeding grounds. The simple intervention of covering water storage containers can reduce vector breeding and may also reduce faecal contamination of water at the household level.

Economic and social effects

When water comes from improved and more accessible sources, people spend less time and effort physically collecting it, meaning they can be productive in other ways. This can also result in greater personal safety and reducing musculoskeletal disorders by reducing the need to make long or risky journeys to collect and carry water. Better water sources also mean less expenditure on health, as people are less likely to fall ill and incur medical costs and are better able to remain economically productive.

With children particularly at risk from water-related diseases, access to improved sources of water can result in better health, and therefore better school attendance, with positive longer-term consequences for their lives.

Challenges

Historical rates of progress would need to double for the world to achieve universal coverage with basic drinking water services by 2030. To achieve universal safely managed services, rates would need to quadruple. Climate change, increasing water scarcity, population growth, demographic changes and urbanization already pose challenges for water supply systems. Over 2 billion people live in water-stressed countries, which is expected to be exacerbated in some regions as result of climate change and population growth. Re-use of wastewater to recover water, nutrients or energy is becoming an important strategy. Increasingly countries are using wastewater for irrigation; in developing countries this represents 7% of irrigated land. While this practice if done inappropriately poses health risks, safe management of wastewater can yield multiple benefits, including increased food production.

Options for water sources used for drinking-water and irrigation will continue to evolve, with an increasing reliance on groundwater and alternative sources, including wastewater. Climate change will lead to greater fluctuations in harvested rainwater. Management of all water resources will need to be improved to ensure provision and quality.

WHO's response

As the international authority on public health and water quality, WHO leads global efforts to prevent water-related disease, advising governments on the development of health-based targets and regulations.

WHO produces a series of water quality guidelines, including on drinking-water, safe use of wastewater, and recreational water quality. The water quality guidelines are based on managing risks, and since 2004 the Guidelines for drinking-water quality promote the Framework for safe drinking-water. The Framework recommends establishment of health-based targets, the development and implementation of water safety plans by water suppliers to most effectively identify and manage risks from catchment to consumer, and independent surveillance to ensure that water safety plans are effective and health-based targets are being met. The drinking-water guidelines are supported by background publications that provide the technical basis for the Guidelines recommendations. WHO also supports countries to implement the drinking-water quality guidelines through the development of practical guidance materials and provision of direct country support. This includes the development of locally relevant drinking-water quality regulations aligned to the principles in the Guidelines, the development, implementation and auditing of water safety plans and strengthening of surveillance practices.

**CHAPTER 5**

**CODING & SOLUTIONING**

**ABOUT DATASET**

Content

The water\_potability.csv file contains water quality metrics for 3276 different water bodies.

1. pH value: PH is an important parameter in evaluating the acid–base balance of water. It is also the indicator of acidic or alkaline condition of water status. WHO has recommended maximum permissible limit of pH from 6.5 to 8.5. The current investigation ranges were 6.52–6.83 which are in the range of WHO standards.

2. Hardness: Hardness is mainly caused by calcium and magnesium salts. These salts are dissolved from geologic deposits through which water travels. The length of time water is in contact with hardness producing material helps determine how much hardness there is in raw water. Hardness was originally defined as the capacity of water to precipitate soap caused by Calcium and Magnesium.

3. Solids (Total dissolved solids - TDS): Water has the ability to dissolve a wide range of inorganic and some organic minerals or salts such as potassium, calcium, sodium, bicarbonates, chlorides, magnesium, sulfates etc. These minerals produced un-wanted taste and diluted color in appearance of water. This is the important parameter for the use of water. The water with high TDS value indicates that water is highly mineralized. Desirable limit for TDS is 500 mg/l and maximum limit is 1000 mg/l which prescribed for drinking purpose.

4. Chloramines: Chlorine and chloramine are the major disinfectants used in public water systems. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water. Chlorine levels up to 4 milligrams per liter (mg/L or 4 parts per million (ppm)) are considered safe in drinking water.

5. Sulfate: Sulfates are naturally occurring substances that are found in minerals, soil, and rocks. They are present in ambient air, groundwater, plants, and food. The principal commercial use of sulfate is in the chemical industry. Sulfate concentration in seawater is about 2,700 milligrams per liter (mg/L). It ranges from 3 to 30 mg/L in most freshwater supplies, although much higher concentrations (1000 mg/L) are found in some geographic locations.

6. Conductivity: Pure water is not a good conductor of electric current rather’s a good insulator. Increase in ions concentration enhances the electrical conductivity of water. Generally, the amount of dissolved solids in water determines the electrical conductivity. Electrical conductivity (EC) actually measures the ionic process of a solution that enables it to transmit current. According to WHO standards, EC value should not exceeded 400 μS/cm.

7. Organic\_carbon: Total Organic Carbon (TOC) in source waters comes from decaying natural organic matter (NOM) as well as synthetic sources. TOC is a measure of the total amount of carbon in organic compounds in pure water. According to US EPA < 2 mg/L as TOC in treated / drinking water, and < 4 mg/Lit in source water which is use for treatment.

8. Trihalomethanes: THMs are chemicals which may be found in water treated with chlorine. The concentration of THMs in drinking water varies according to the level of organic material in the water, the amount of chlorine required to treat the water, and the temperature of the water that is being treated. THM levels up to 80 ppm is considered safe in drinking water.

9. Turbidity: The turbidity of water depends on the quantity of solid matter present in the suspended state. It is a measure of light emitting properties of water and the test is used to indicate the quality of waste discharge with respect to colloidal matter. The mean turbidity value obtained for Wondo Genet Campus (0.98 NTU) is lower than the WHO recommended value of 5.00 NTU.

10. Potability: Indicates if water is safe for human consumption where 1 means Potable and 0 means Not potable.

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

import os

for dirname, \_, filenames in os.walk('/kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))

# It is always consider as a good practice to make a copy of original dataset.

main\_df = pd.read\_csv("/kaggle/input/water-potability/water\_potability.csv")

df = main\_df.copy()

# Getting top 5 row of the dataset

df.head()

Output

|  | Ph | Hardness | Solids | Chloramines | Sulfate | Conductivity | Organic\_carbon | Trihalomethanes | Turbidity | Potability |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | NaN | 204.890455 | 20791.318981 | 7.300212 | 368.516441 | 564.308654 | 10.379783 | 86.990970 | 2.963135 | 0 |
| 1 | 3.716080 | 129.422921 | 18630.057858 | 6.635246 | NaN | 592.885359 | 15.180013 | 56.329076 | 4.500656 | 0 |
| 2 | 8.099124 | 224.236259 | 19909.541732 | 9.275884 | NaN | 418.606213 | 16.868637 | 66.420093 | 3.055934 | 0 |
| 3 | 8.316766 | 214.373394 | 22018.417441 | 8.059332 | 356.886136 | 363.266516 | 18.436524 | 100.341674 | 4.628771 | 0 |
| 4 | 9.092223 | 181.101509 | 17978.986339 | 6.546600 | 310.135738 | 398.410813 | 11.558279 | 31.997993 | 4.075075 | 0 |

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

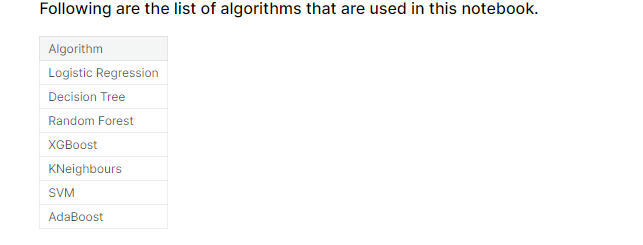
%matplotlib inline

import plotly.express as px

import warnings

warnings.filterwarnings('ignore')

OUTPUT



print(df.shape)

OUTPUT

(3276, 10)

print(df.columns)

Index(['ph', 'Hardness', 'Solids', 'Chloramines', 'Sulfate', 'Conductivity',

'Organic\_carbon', 'Trihalomethanes', 'Turbidity', 'Potability'],

dtype='object')

df.describe()

OUTPUT

|  | ph | Hardness | Solids | Chloramines | Sulfate | Conductivity | Organic\_carbon | Trihalomethanes | Turbidity | Potability |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| count | 2785.000000 | 3276.000000 | 3276.000000 | 3276.000000 | 2495.000000 | 3276.000000 | 3276.000000 | 3114.000000 | 3276.000000 | 3276.000000 |
| Mean | 7.080795 | 196.369496 | 22014.092526 | 7.122277 | 333.775777 | 426.205111 | 14.284970 | 66.396293 | 3.966786 | 0.390110 |
| Std | 1.594320 | 32.879761 | 8768.570828 | 1.583085 | 41.416840 | 80.824064 | 3.308162 | 16.175008 | 0.780382 | 0.487849 |
| min | 0.000000 | 47.432000 | 320.942611 | 0.352000 | 129.000000 | 181.483754 | 2.200000 | 0.738000 | 1.450000 | 0.000000 |
| 25% | 6.093092 | 176.850538 | 15666.690297 | 6.127421 | 307.699498 | 365.734414 | 12.065801 | 55.844536 | 3.439711 | 0.000000 |
| 50% | 7.036752 | 196.967627 | 20927.833607 | 7.130299 | 333.073546 | 421.884968 | 14.218338 | 66.622485 | 3.955028 | 0.000000 |
| 75% | 8.062066 | 216.667456 | 27332.762127 | 8.114887 | 359.950170 | 481.792304 | 16.557652 | 77.337473 | 4.500320 | 1.000000 |
| max | 14.000000 | 323.124000 | 61227.196008 | 13.127000 | 481.030642 | 753.342620 | 28.300000 | 124.000000 | 6.739000 | 1.000000 |

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3276 entries, 0 to 3275

Data columns (total 10 columns):

OUTPUT

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 ph 2785 non-null float64

1 Hardness 3276 non-null float64

2 Solids 3276 non-null float64

3 Chloramines 3276 non-null float64

4 Sulfate 2495 non-null float64

5 Conductivity 3276 non-null float64

6 Organic\_carbon 3276 non-null float64

7 Trihalomethanes 3114 non-null float64

8 Turbidity 3276 non-null float64

9 Potability 3276 non-null int64

dtypes: float64(9), int64(1)

memory usage: 256.1 KB

print(df.nunique())

ph 2785

Hardness 3276

Solids 3276

Chloramines 3276

Sulfate 2495

Conductivity 3276

Organic\_carbon 3276

Trihalomethanes 3114

Turbidity 3276

Potability 2

dtype: int64

print(df.isnull().sum())

ph 491

Hardness 0

Solids 0

Chloramines 0

Sulfate 781

Conductivity 0

Organic\_carbon 0

Trihalomethanes 162

Turbidity 0

Potability 0

dtype: int64

df.dtypes

OUTPUT

ph float64

Hardness float64

Solids float64

Chloramines float64

Sulfate float64

Conductivity float64

Organic\_carbon float64

Trihalomethanes float64

Turbidity float64

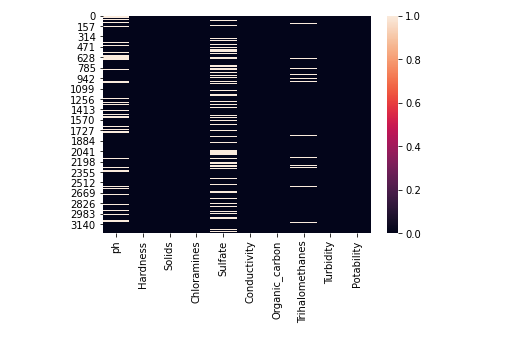
Potability int64

dtype: object

sns.heatmap(df.isnull())

OUTPUT

<AxesSubplot:>

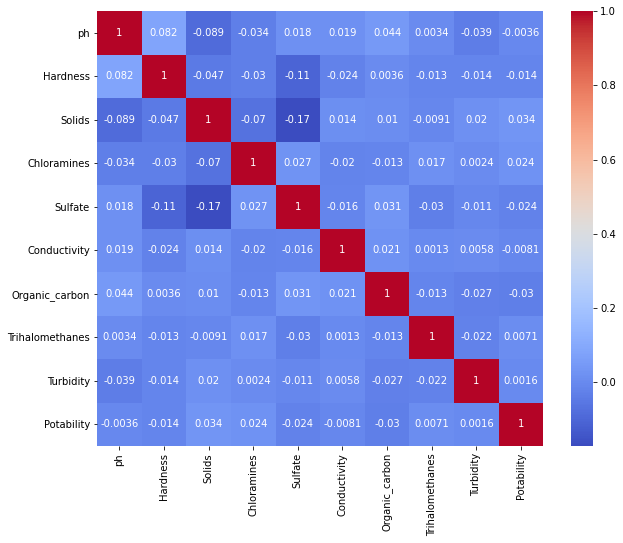


plt.figure(figsize=(10, 8))

sns.heatmap(df.corr(), annot= True, cmap='coolwarm')

OUTPUT

<AxesSubplot:>



# Unstacking the correlation matrix to see the values more clearly.

corr = df.corr()

c1 = corr.abs().unstack()

c1.sort\_values(ascending = False)[12:24:2]

OUTPUT

Hardness Sulfate 0.106923

ph Solids 0.089288

Hardness ph 0.082096

Solids Chloramines 0.070148

Hardness Solids 0.046899

ph Organic\_carbon 0.043503

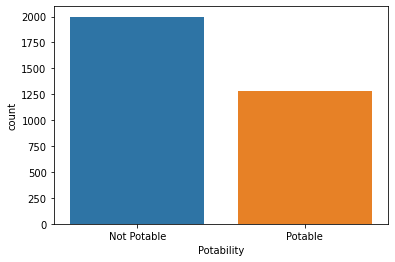
dtype: float64

ax = sns.countplot(x = "Potability",data= df, saturation=0.8)

plt.xticks(ticks=[0, 1], labels = ["Not Potable", "Potable"])

plt.show()

OUTPUT



x = df.Potability.value\_counts()

labels = [0,1]

print(x)

0 1998

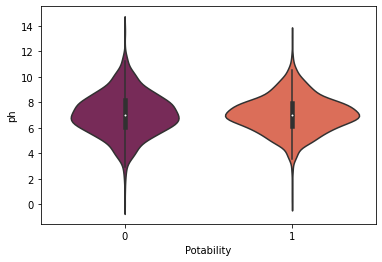
1 1278

Name: Potability, dtype: int64

sns.violinplot(x='Potability', y='ph', data=df, palette='rocket')

OUTPUT

<AxesSubplot:xlabel='Potability', ylabel='ph'>



# Visualizing dataset and also checking for outliers

fig, ax = plt.subplots(ncols = 5, nrows = 2, figsize = (20, 10))

index = 0

ax = ax.flatten()

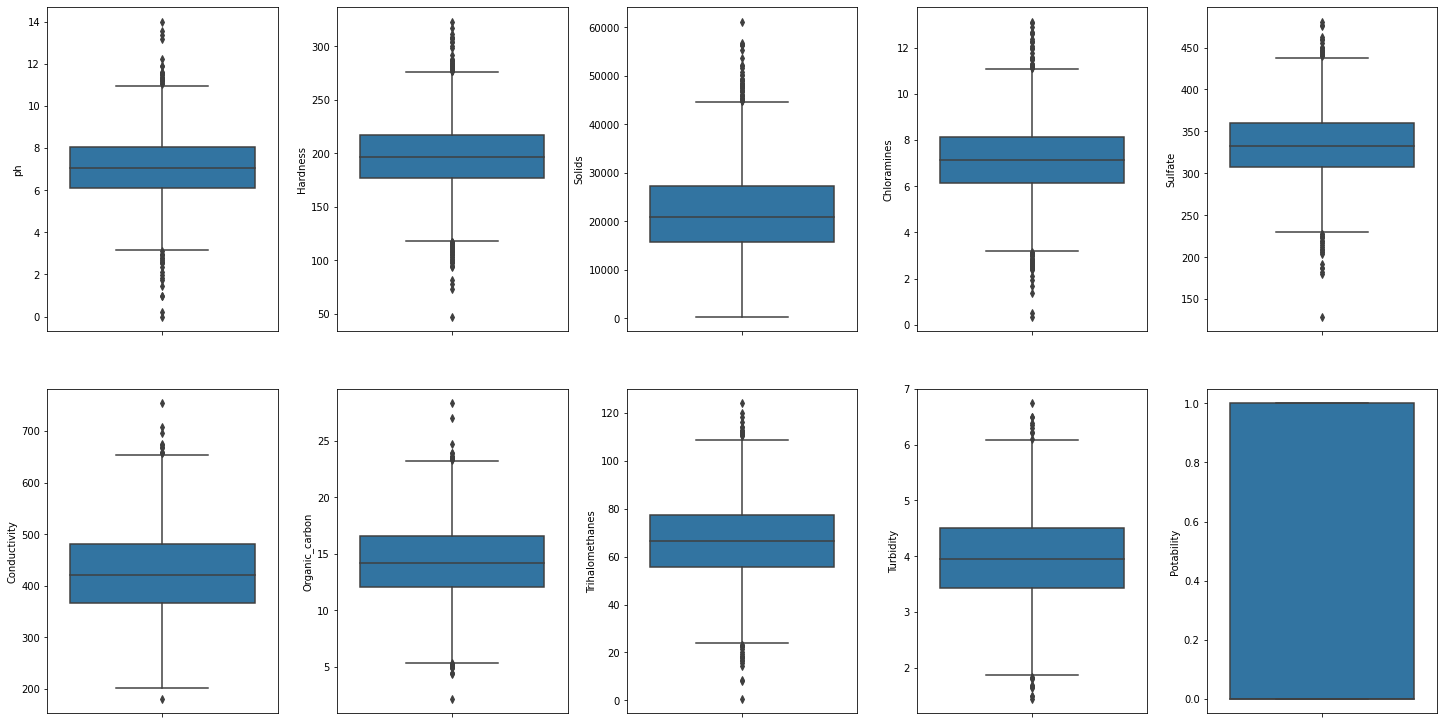
for col, value in df.items():

sns.boxplot(y=col, data=df, ax=ax[index])

index += 1

plt.tight\_layout(pad = 0.5, w\_pad=0.7, h\_pad=5.0)

OUTPUT

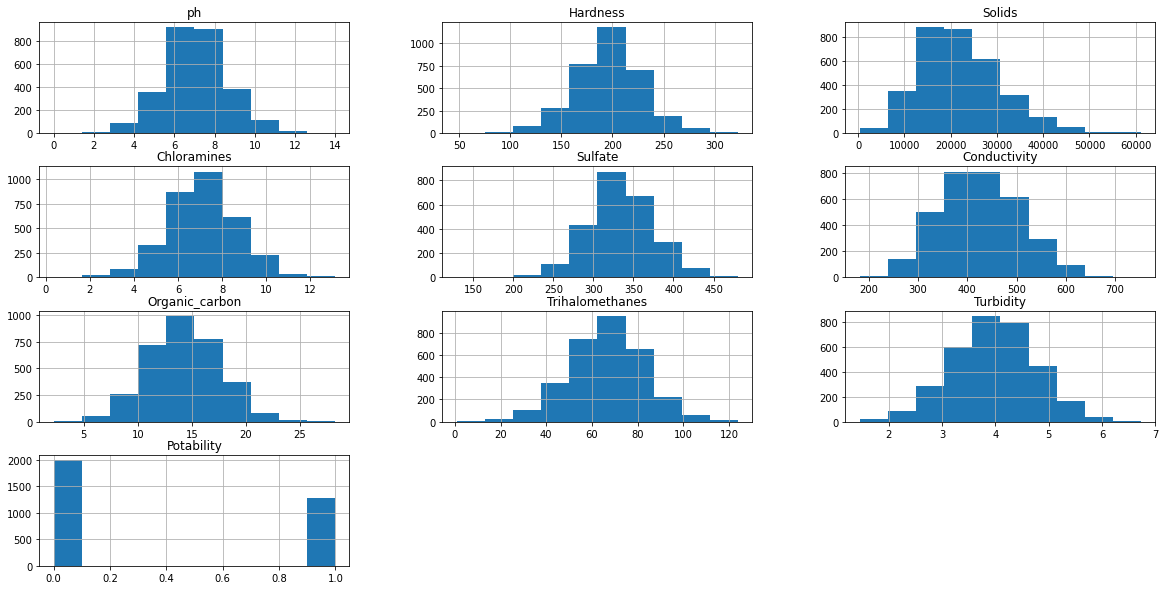


plt.rcParams['figure.figsize'] = [20,10]

df.hist()

plt.show()

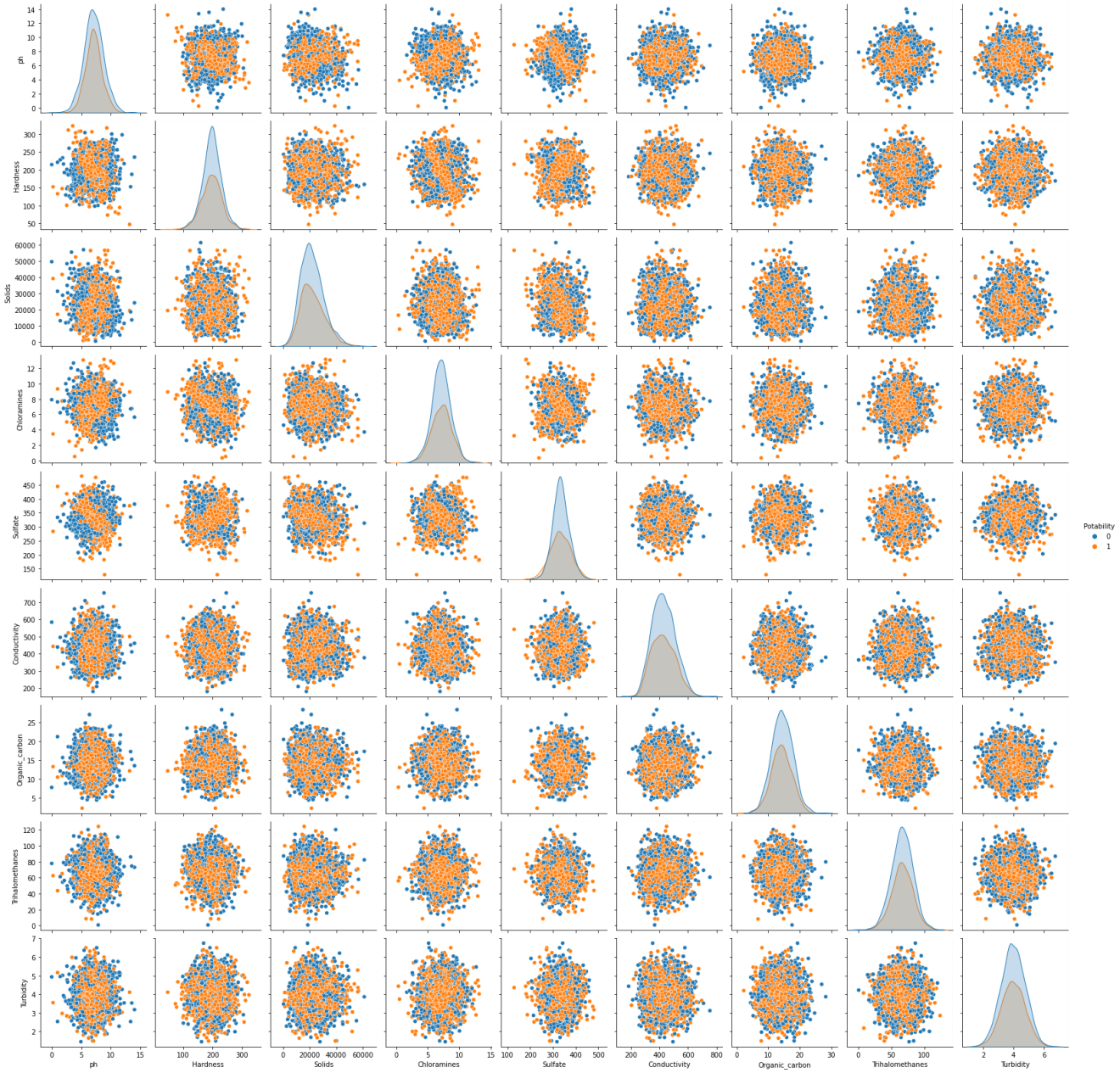
OUTPUT



sns.pairplot(df, hue="Potability")

OUTPUT

<seaborn.axisgrid.PairGrid at 0x7f73daa726d0>

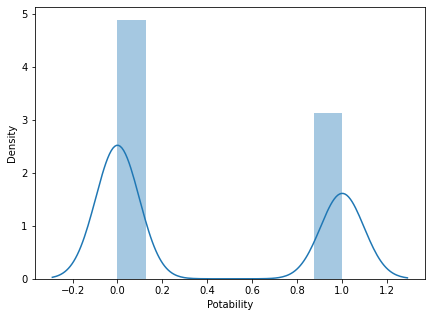


plt.rcParams['figure.figsize'] = [7,5]

sns.distplot(df['Potability'])

OUTPUT

<AxesSubplot:xlabel='Potability', ylabel='Density'>



df.hist(column='ph', by='Potability')

OUTPUT

array([<AxesSubplot:title={'center':'0'}>,

<AxesSubplot:title={'center':'1'}>], dtype=object)

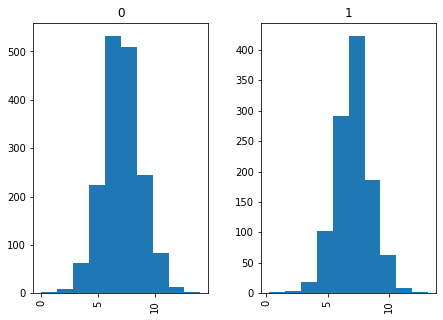


fig = px.histogram (df, x = "Sulfate", facet\_row = "Potability", template = 'plotly\_dark')

fig.show ()

OUTPUT

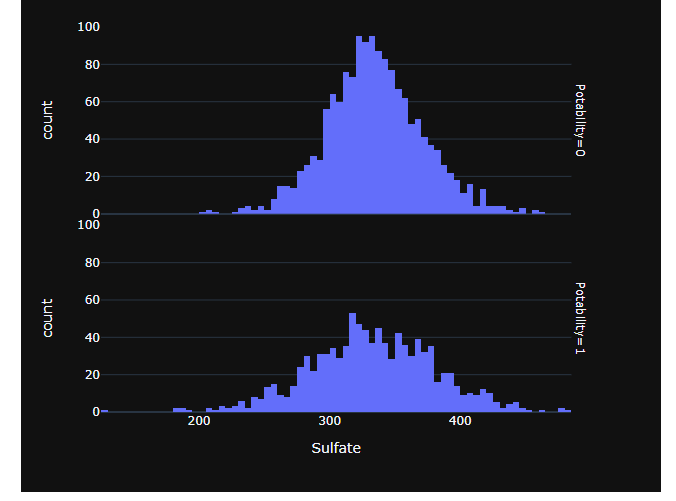


fig = px.scatter (df, x = "ph", y = "Sulfate", color = "Potability", template = "plotly\_dark", trendline="ols")

fig.show ()

OUTPUT

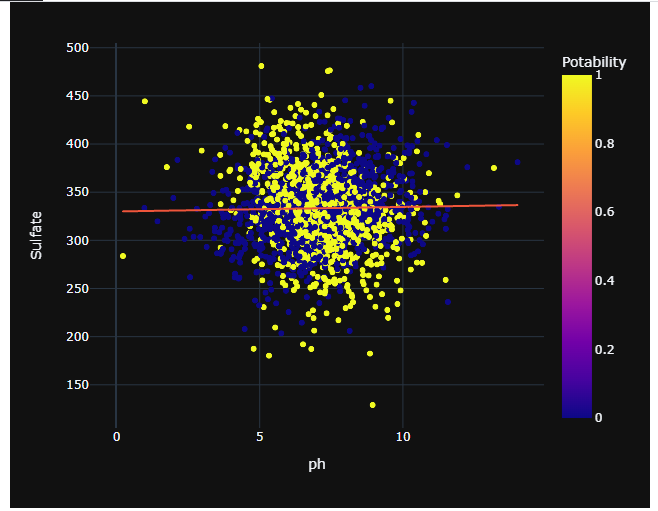
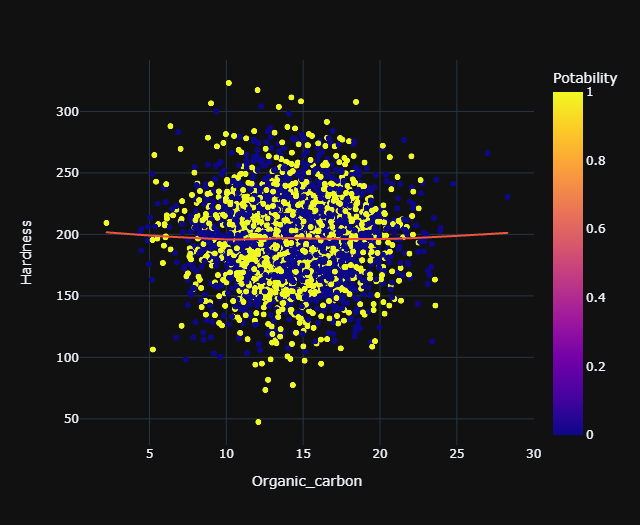


fig = px.scatter (df, x = "Organic\_carbon", y = "Hardness", color = "Potability", template = "plotly\_dark", trendline="lowess")

fig.show ()

OUTPUT



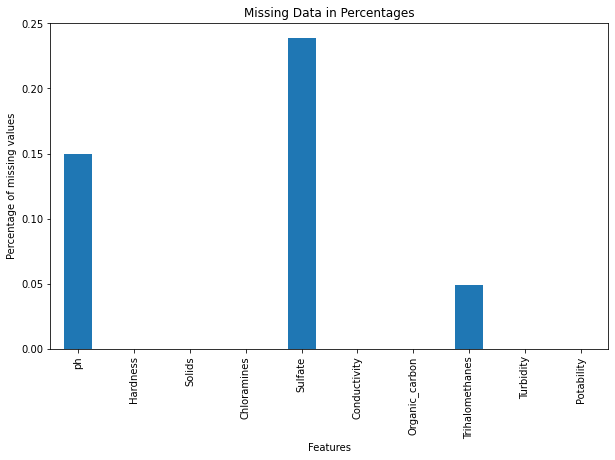
df.isnull().mean().plot.bar(figsize=(10,6))

plt.ylabel('Percentage of missing values')

plt.xlabel('Features')

plt.title('Missing Data in Percentages');

OUTPUT



models = pd.DataFrame({

'Model':['Logistic Regression', 'Decision Tree', 'Random Forest', 'XGBoost', 'KNeighbours', 'SVM', 'AdaBoost'],

'Accuracy\_score' :[lg, dt, rf, xgb, kn, sv, ada]

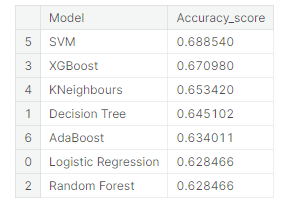
})

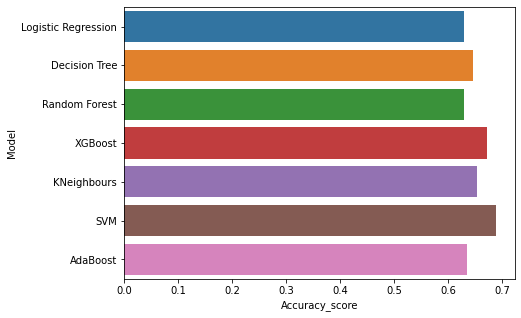
models

sns.barplot(x='Accuracy\_score', y='Model', data=models)

models.sort\_values(by='Accuracy\_score', ascending=False)

OUTPUT





5.3 DATABASE SCHEMA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ph | Hardness | Solids | Chloramines | Sulfate | Conductivity | Organic\_carbon |
|  | 204.890455 | 20791.32 | 7.300211873 | 368.5164 | 564.3086542 | 10.37978308 |
| 3.71608 | 129.422921 | 18630.06 | 6.635245884 |  | 592.8853591 | 15.18001312 |
| 8.099124 | 224.236259 | 19909.54 | 9.275883603 |  | 418.6062131 | 16.86863693 |
| 8.316766 | 214.373394 | 22018.42 | 8.059332377 | 356.8861 | 363.2665162 | 18.4365245 |
| 9.092223 | 181.101509 | 17978.99 | 6.546599974 | 310.1357 | 398.4108134 | 11.55827944 |
| 5.584087 | 188.313324 | 28748.69 | 7.544868789 | 326.6784 | 280.4679159 | 8.39973464 |
| 10.22386 | 248.071735 | 28749.72 | 7.513408466 | 393.6634 | 283.6516335 | 13.78969532 |
| 8.635849 | 203.361523 | 13672.09 | 4.563008686 | 303.3098 | 474.6076449 | 12.3638167 |
|  | 118.988579 | 14285.58 | 7.804173553 | 268.6469 | 389.3755659 | 12.70604897 |
| 11.18028 | 227.231469 | 25484.51 | 9.077200017 | 404.0416 | 563.8854815 | 17.92780641 |
| 7.36064 | 165.520797 | 32452.61 | 7.550700907 | 326.6244 | 425.3834195 | 15.58681044 |
| 7.974522 | 218.6933 | 18767.66 | 8.110384501 |  | 364.0982305 | 14.5257457 |
| 7.119824 | 156.704993 | 18730.81 | 3.606036091 | 282.3441 | 347.7150273 | 15.92953591 |
|  | 150.174923 | 27331.36 | 6.838223471 | 299.4158 | 379.7618348 | 19.37080718 |
| 7.496232 | 205.344982 | 28388 | 5.072557774 |  | 444.6453523 | 13.2283111 |
| 6.347272 | 186.732881 | 41065.23 | 9.629596276 | 364.4877 | 516.7432819 | 11.53978119 |
| 7.051786 | 211.049406 | 30980.6 | 10.09479601 |  | 315.1412672 | 20.39702184 |
| 9.18156 | 273.813807 | 24041.33 | 6.904989726 | 398.3505 | 477.9746419 | 13.38734078 |
| 8.975464 | 279.357167 | 19460.4 | 6.204320859 |  | 431.44399 | 12.88875905 |

**CHAPTR 6**

**RESULTS**

**6.1 PERFORMANCE METRICS**

In this investigation, WSP-QA Tool along with the WSP manual presented by WHO and IWA were used to evaluate the stages of implementation of the WSP in the water supply system of Birjand city. The results shown in Table 2 indicate that from the 440 points for complete application of the plan and 328 points associated with the investigated phases, 190 points were obtained and a 43.18% congruence with WSP implementation was observed. The system description stage (7 out of the total 8 points) showed the maximum percentage of congruent implementation with the WSP (87.5%), whereas the phase related to management procedures had the lowest score (9 out of the total 36 points) and claimed the minimum percentage of implementation (25%). These results reveal that, although management procedures are an important part of the WSP, in the current approach, they are not of great interest to the company. These procedures include standard operating procedures (SOPs) and corrective activities (at the time of incident). Unexpected accidents and deviations in water supply systems may occur when there is no corrective activity at a certain site. Therefore, these procedures should be developed by experienced employees and updated and documented if necessary. In 2014, Aghaei evaluated the safety of the water supply system in Ardebil city and observed 21% congruent implementation with the WSP.In this study, the system description stage also claimed the maximum percentage of congruent implementation with the WSP, whereas the phase related to management procedures claimed the minimum percentage (10). These results are in line with those of the current research. The potential of the software is such that evaluation of some stages of implementation of the WSP (including WSP review, improvement plan, and supporting programs) necessitates complete implementation of the plan in water supply system management; as WSP was not implemented completely in Birjand city, the questions and scores related to phases that could not be evaluated were not included in the final analysis. This is also justifiable considering the feature of the software in independent analysis of the data of each phase. Evaluation of the application of the eight parameters , which is somehow in line with corresponding Environmental Health Engineering and Management Journal 2018, 5(1), 39–47 45 Eslami et al phases in the WSP, for different components of water supply system (catchment, treatment, distribution, and point of use) indicated that the processes associated with the identification of stakeholders across different parts (resources, treatment plant, and distribution system) as well as the process related to operational monitoring at the final point of consumption gained the maximum score. It is notable that, although system description indicated the maximum percentage of congruent implementation with the WSP, the processes associated with risk assessment across all four parts of the water supply system (especially distribution system and the final point of use with a score of zero) were among the processes that gained the lowest scores. This suggests a lack of concentration of the water supplier system on the risk assessment process. The best way to achieve this target, according to the WSP manual, is to draw asimple table with the systematic registration of all hazardous events and relevant risks along with an estimation of their risk level. The risk assessment process can be based on a quantitative, semi quantitative (through a risk assessment matrix), or a simple qualitative solution (based on the judgment of the WSP expert panel). A WSP workshop in Berlin in 2014 also introduced this plan as an effective method for risk management in the 21st century. It further stated that, even in cases of high correspondence with the drinking water directive of the European Union (EU-DWD), the WSP is a systematic method and suitable management instrument to ensure continuous delivery of healthy drinking water to consumers (21). In a study by Gholami et al on the water supply system in Khoy city, identification and risk assessment stage showed 12% congruent implementation with the WSP (14). The results of the study by Aghaei on the water supply system in Ardebil city indicated that, considering the low percentage of general application of different WSP phases of and regarding the lack of attention to some key parameters including risk assessment and catchment management by the water supplier system, the current system does not enjoy the necessary safety, and the governing controlling approach does not have the efficiency required for integrated management of a healthy drinking water provision system (10). In this study, analysis of the results obtained from the diagrams revealed that, considering the percentage of general application of different WSP phases, the operational concentration of the supplier system is mainly on the verification stages and system description, such that the processes associated with risk assessment and management procedures gained a low score across all four main parts of the water supply system. Therefore, the major disadvantage of the current approach governing the qualitative management of drinking water in Birjand city, i.e. dependence on end-product experiments. This is in accordance with the results obtained by Gholami et al in their evaluation of drinking water safety in Khoy city (14).

According to the report of progression of development of the WSP in Korea in 2013, the objectives in developing this plan is to protect and enhance the safety of drinking water

as well as to identify and assess the risks and hazardous events in the water supply system. The report stated that the main characteristic of WSP implementation in this country is the presentation of a water safety index (WSI) to detect the safety level of drinking water and to identify the weak points in the water supply system. It is highly useful to gain a quick understanding of the water safety level in a water supply system and to detect the vulnerable points of the system (29). The most effective method for achieving sustainable confidence about the safety of a drinking

water supply system is the application and management of a comprehensive risk assessment approach, such that it covers all the stages involved in water supply from catchment to the consumer. Such solutions are called WSP(30)

**CHAPTER 7**

**ADVANTAGES & DISADVANTAGES**

**ADVANTAGES**

The population of this city is 486495 people

The most important flowing river in this region is Zanjan Rood . Zanjan city has a cold semiarid climate – according to the Martonne – and its height is 1659 meters . Water supplies for drinking and agriculture of the region are formed by superficial and underground waters. The water needed for Zanjan citizens is now supplied from wellbores situated in residential region (center and west of the city),15 well bore sofeastandal so140 Assessment of Zanjan Drinking Water System

Eslami A et al/ J. Hum. Environ. Health Promot. 2017; 2(3):138-146 18 wellbores of Bonab (65% of underground water) and Taham dam (35% superficial water)

This is a cross-sectional descriptive analytic study which was done on drinking water supply system of Zanjan city of Zanjan province in 2016. The software which is used in this study is based on Excel software provided from WHO and IWA in 2010. Since the software is new, it is not applied in qualitative management of drinking water in different countries of the world, even in countries which have the applied WSP. The major advantage of the application of this software for assessment of WSP by water suppliers is the identification of the fields which needs

development and opportunities which can be enhanced. Another feature of this software is the application of it as a manual in a place which WSP has not implanted totally or is at first stages of implementation. In this study, the latter feature is applied

**DISADVANTAGE**

The study, books, references and various research articles were investigated and the most current risks in water supply system and also the probable specific risks which can occur in the process of water supply were detected. Then, experts of water supply (water resources, treatment and distribution), health and environment specialists from the organization or out of the organization were chosen. Efforts were made to choose the experienced and expert staff who were aware of water resources, treatment, distribution of water, and the risks which can influence water safety of water supply system from the catchment area to the consumption point. The staff should have the ability of detection and management of associated risks.

The common detected risks associated with the compartments of water supply system involved 4 stages:

• Detected risks of the source

• Detected risks of the treatment

• Detected risks of water distribution network and reservoir sources

• Detected risks of consumption point

For this purpose, 10 members of the staff were chosen and they were asked to detect the risks in each phase and classify and prioritize them according to the importance of the risk in water supply system of Zanjan city. For assessment of each risk, the risk score of 1 allocated to the least important risk (lowest priority) and 10 (or more

depending on number of risks detected in each phase) allocated to the most important risk (highest priority) according to counseling with the experts of drinking water supply of Zanjan city.

Total number of questions asked about the risks from the experts were 56. 17 questions were about the water source, 15 questions were about the refinery, 13 questions were

about the distribution network and storage source and finally 11 questions were about the consumption point. In this study, risk scoring matrix of 5×5 of chapter 4

which exists at the third version of the WHO book was applied.

In this study, for the reliability of the designed questionnaire, completion of the questionnaire was done two times with one- week interval by the experts to check the certainty of the given responses. For this purpose, the reliability test of the asked questions were done in two interviews with the experts and by means of Pearson Correlation Coefficient by SPSS software (version 23). In this study, the resulted correlation coefficient of 5 questions from two tests, were lower than the accepted limit (<0.6) therefore they were omitted from the questionnaire and eventually the reliability of 51 questions were confirmed.

**CONCLUSION**

In a general view, according to the percentage of general implementation of various phases of WSP in water supply system of Zanjan city (52.95%) and low attention of water supply system to some key parameters such as the WSP review, WSP team,

and control and accreditation criteria, especially in distribution network and final consumption point, reforming the present approach for more conformity with WSP in which forming a team is one its phases is necessary .

Data resulted from assessment of important risks in this system show that for achieving a desirable drinking water with high quality, removal or reduction of high concentration of nitrite and nitrate in some water sources, matching the refinement processes, and quality preservation Eslami A et al/ J. Hum. Environ. Health Promot. 2017; 2(3):138-146 of drinking water in distribution system by keeping enough chlorine in the network, and reconstruction of infrastructural installations and old pipelines of the city is necessary. According to water safety plans, it is necessary that organizations in the charge participate in accomplishing water safety plans

**FUTURE SCOPE**

Oceans and seas represent 97% of global water sources, making desalination—the process of removing salt and other particles from seawater—one of the best options for increasing supplies of freshwater.

In 2020, desalination met about 1% of global freshwater demand. Now with more than 150 desalination projects in the pipeline, the desalination market could grow around 9% annually through 2025. Meanwhile, new technologies, such as nano-membranes that desalinate water faster, could further improve capacity.

Desalinated water may still be too expensive for agricultural use, but it could be a viable option for households, businesses and industries to tap new sources of freshwater.

Businesses Rethinking Their Water Use

Agriculture may be one of the biggest users of fresh water, but the growing water crisis will no doubt require virtually all industries to rethink their water use.“From an investment perspective, we think that heavily water intensive companies should be prepared for potentially more-volatile water pricing over time as more market-based and dynamic pricing mechanisms are put into use,” says Alsford.Energy companies and utilities are the most water-intensive by far, followed by mining and cement. Companies in these sub-industries will need to get serious about using desalinated seawater, water recycling and other novel technologies.

Investors in pharmaceuticals, beverages, semiconductors, apparel and data centers should also pay attention to what companies in these sectors are doing to make every drop count.

**APPENDIX**

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

import os

for dirname, \_, filenames in os.walk('/kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))

# It is always consider as a good practice to make a copy of original dataset.

main\_df = pd.read\_csv("/kaggle/input/water-potability/water\_potability.csv")

df = main\_df.copy()

# Getting top 5 row of the dataset

df.head()

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

import plotly.express as px

import warnings

warnings.filterwarnings('ignore')

import warnings

warnings.filterwarnings('ignore')

print(df.columns)

Index(['ph', 'Hardness', 'Solids', 'Chloramines', 'Sulfate', 'Conductivity',

'Organic\_carbon', 'Trihalomethanes', 'Turbidity', 'Potability'],

dtype='object')

df.describe()

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3276 entries, 0 to 3275

Data columns (total 10 columns):

print(df.nunique())

print(df.isnull().sum())

df.dtypes

sns.heatmap(df.isnull())

plt.figure(figsize=(10, 8))

sns.heatmap(df.corr(), annot= True, cmap='coolwarm')

# Unstacking the correlation matrix to see the values more clearly.

corr = df.corr()

c1 = corr.abs().unstack()

c1.sort\_values(ascending = False)[12:24:2]

ax = sns.countplot(x = "Potability",data= df, saturation=0.8)

plt.xticks(ticks=[0, 1], labels = ["Not Potable", "Potable"])

plt.show()

x = df.Potability.value\_counts()

labels = [0,1]

print(x)

sns.violinplot(x='Potability', y='ph', data=df, palette='rocket') # Visualizing dataset and also checking for outliers

fig, ax = plt.subplots(ncols = 5, nrows = 2, figsize = (20, 10))

index = 0

ax = ax.flatten()

for col, value in df.items():

sns.boxplot(y=col, data=df, ax=ax[index])

index += 1

plt.tight\_layout(pad = 0.5, w\_pad=0.7, h\_pad=5.0)

plt.rcParams['figure.figsize'] = [20,10]

df.hist()

plt.show()

sns.pairplot(df, hue="Potability")

plt.rcParams['figure.figsize'] = [7,5]

sns.distplot(df['Potability'])

df.hist(column='ph', by='Potability')

fig = px.histogram (df, x = "Sulfate", facet\_row = "Potability", template = 'plotly\_dark')

fig.show ()

fig = px.scatter (df, x = "ph", y = "Sulfate", color = "Potability", template = "plotly\_dark", trendline="ols")

fig.show ()

fig = px.scatter (df, x = "Organic\_carbon", y = "Hardness", color = "Potability", template = "plotly\_dark", trendline="lowess")

fig.show ()

df.isnull().mean().plot.bar(figsize=(10,6))

plt.ylabel('Percentage of missing values')

plt.xlabel('Features')

plt.title('Missing Data in Percentages');

models = pd.DataFrame({

'Model':['Logistic Regression', 'Decision Tree', 'Random Forest', 'XGBoost', 'KNeighbours', 'SVM', 'AdaBoost'],

'Accuracy\_score' :[lg, dt, rf, xgb, kn, sv, ada]

})

models

sns.barplot(x='Accuracy\_score', y='Model', data=models)

models.sort\_values(by='Accuracy\_score', ascending=False)