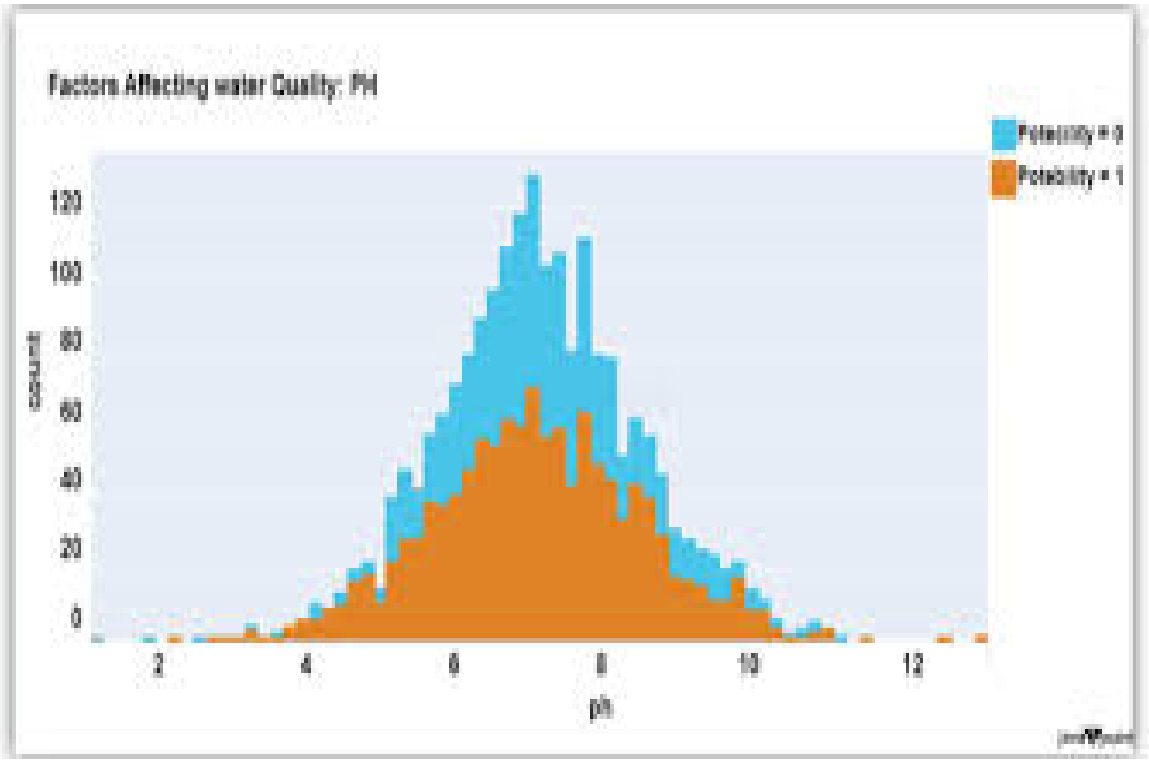
Water quality analysis with python programming

BATCH MEMBER
812121104020:Jaasim J

Project title:Water quality analysis
Phase3:Development part 1
Topic:Start building the water quality analysis model by loading and pre processing data set.



Introduction: Water Quality Analysis

- ➢ Analysing water quality is one of the key topics of machine learning research.
- ➢ In order to train a machine learning model that can determine if a certain water sample is safe or unsafe for eating, we must first understand all the parameters that impact water potability.
- ➢ This process is also known as water potability analysis.
- ➢ We'll be utilising a Kaggle dataset that includes information on all of the key elements that have an impact on the potability of water for the water quality analysis challenge.
- ➢ Before building a model using machine learning to predict whether the water specimen is acceptable or unsafe for eating.
- ➢ we must first quickly examine each characteristic of this dataset because all of the elements that determine water quality are crucial

About dataset

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.

Dataset

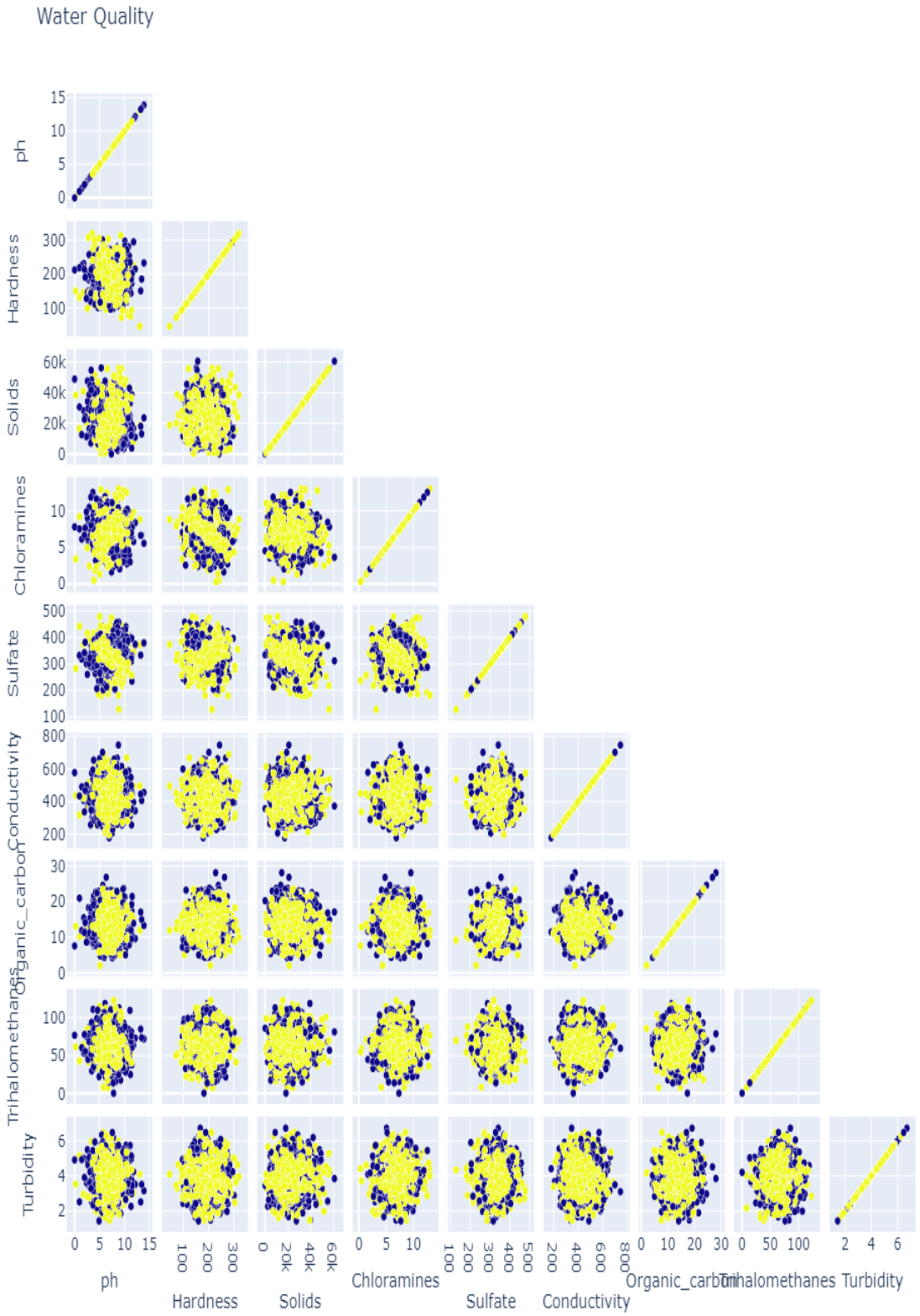
# ph	# Hardness	# Solids	# Chloramines	# Sulfate	# Conductivi...	# Organic_c...
8.757257397440991	200.19140044205727	21536.224687445414	4.915101054543186	317.88290049783706	404.71779915644453	13.768323306421337
	168.38843077429533	27492.307306587 81	7.0462247976964 13	299.82047791082 16	383.79501999032584	16.18206649367 37
7.8096318980194 1	100.45761509158 31	12013.550628764531	5.2123146020653 52	247.20082604764 31	605.22012435000 56	9.611348740011 38
6.6524880908557 07	145.01017191908 34	19871.788448305862	4.9610663801915 02	288.05219173685 15	545.97499376240 3	10.942024256012132
9.1471970553369 21	211.71414177764134	11920.610835646206	7.2307947693374	339.75191880079 38	527.70891001304 07	18.275312109277 65
10.560744636218196	181.89336556155354	21783.651033363374	6.9912599962381 34	340.39037835517297	456.55640212234 38	16.482835366963 91
7.4842548377274 66	260.09217257337195	30616.615148853696	9.3791335405076 24		404.67077732816267	15.934483871207078
8.5208065729703 03	238.33511245550966	28779.650011033 6	8.2828084645096 09	381.64932287189 46	481.31880268861 31	6.0163366092713 07
4.9994138107969 19	190.28705014936028	24323.865903045946	7.2301641268620 36	324.89303784225 57	405.33048202759 03	8.236557503105303

Program:

```
import plotly.graph_objs as go
index_vals                                                    =
data['Potability'].astype('category').cat.codes

fig = go.Figure(data=go.Splom(
        dimensions=[dict(label='ph',
        values=data['ph']),
        dict(label='Hardness',
        values=data['Hardness']),
       dict(label='Solids',
       values=data['Solids']),
       dict(label='Chloramines',
       values=data['Chloramines'])
```

```
dict(label='Sulfate',
     values=data['Sulfate']),
    dict(label='Conductivity',
    values=data['Conductivity']),
   dict(label='Organic_carbon',
   values=data['Organic_carbon']),
   dict(label='Trihalomethanes',
  values=data['Trihalomethanes']),

dict(label='Turbidity',values=data['Turbidity'])],
showupperhalf=False,
text=data['Potability'],
marker=dict(color=index_vals,
showscale=False,
line_color='white', line_width=0.5)
))
 fig.update_layout(
 title='Water Quality',
 width=1000,
  height=1000,
)

fig.show().
```

Output:


Water Quality
ph
Hardness
Solids
Chloramines
Sulfate
Conductivity
Organic_carbon
Trihalomethanes
Turbidity