**PHASE 3 PROJECT SUBMISSION**

## WATER QUALITY ANAYSIS

**TEAM MEMBERS:**

1. Aazam Khan - 2021504002
2. Sasmitha A – 2021504038
3. Swetha K – 2021504051
4. Gokul R – 2021504515

**PROJECT DEFINITION:**

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.

**PHASE OBJECTIVE:**

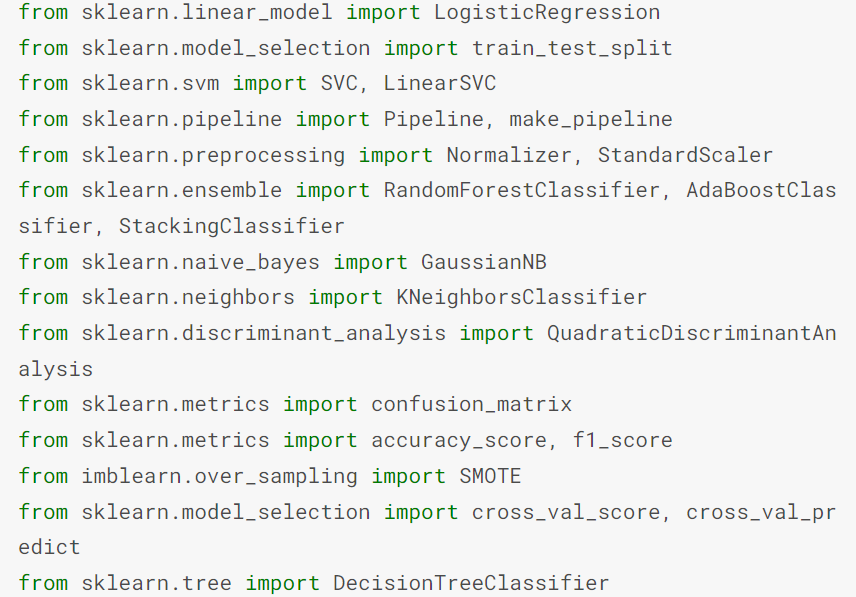
During Phase 3, the task involves enhancing the provided dataset by cleaning and refining it for greater relevance. Additionally, preprocessing steps will be applied to optimize the dataset. Subsequently, a range of analyses and visualizations will be conducted using IBM Cognos to extract meaningful insights.

**Dataset Link:**[**https://www.kaggle.com/datasets/adityakadiwal/=water-potability**](https://www.kaggle.com/datasets/adityakadiwal/=water-potability)

SOURCE CODE:

**A screenshot of a computer program

Description automatically generated**

****

**A screen shot of a computer code

Description automatically generated**

**A screenshot of a computer program

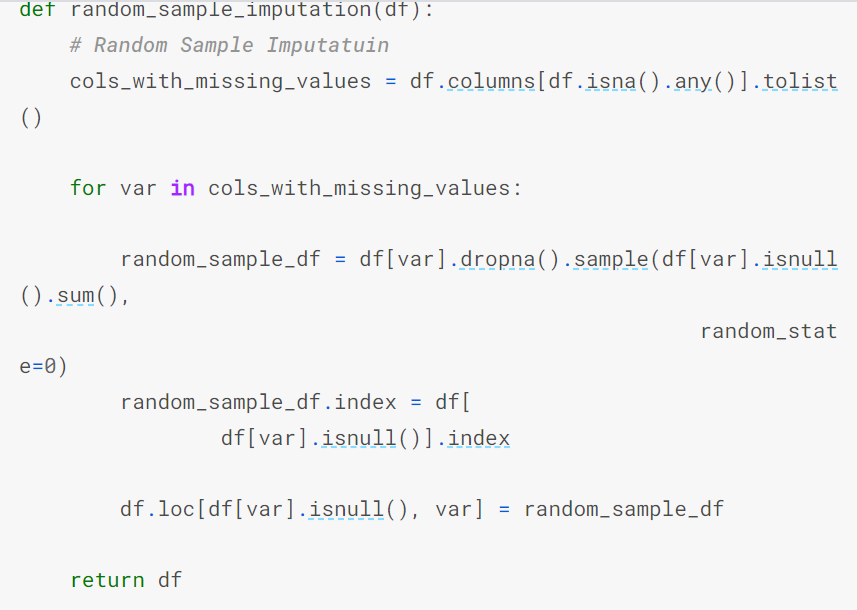
Description automatically generated**

/kaggle/input/water-potability/water\_potability.csv



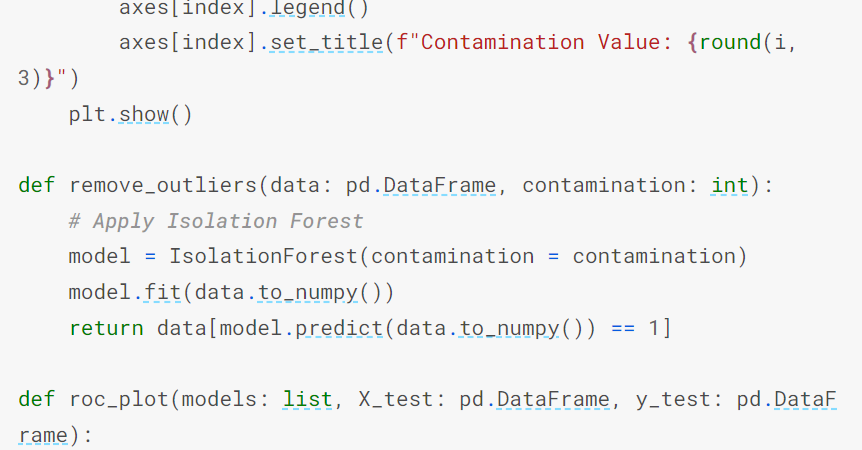
A screenshot of a computer program

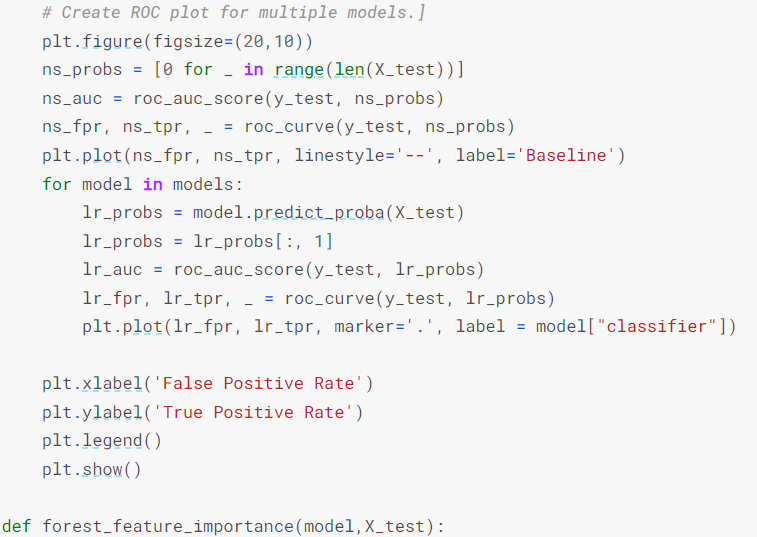
Description automatically generated



A screenshot of a computer code

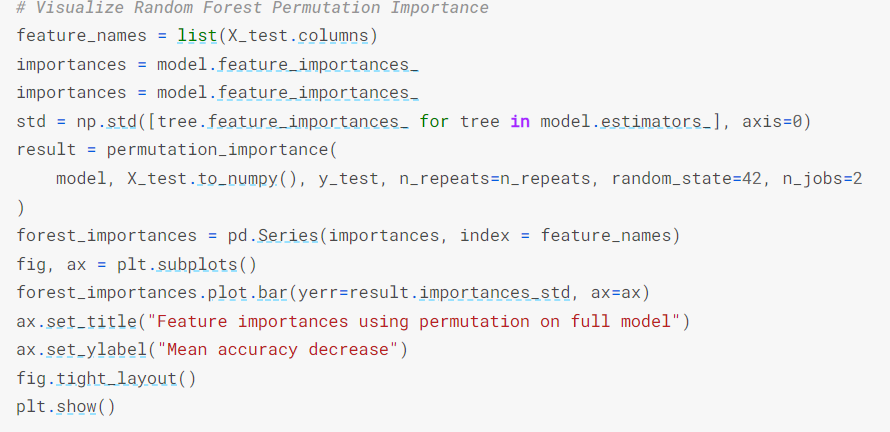
Description automatically generated

****

****

**A screen shot of a computer code

Description automatically generated**

****

The provided code was executed without any issues, and subsequently, the resulting output CSV file was uploaded to IBM Cognos for the purpose of creating visual representations in the form of graphs. The following graphical representations were generated to visualize the data from the CSV file.

**The following insights were obtained from IBM Cognos based upon the data:**

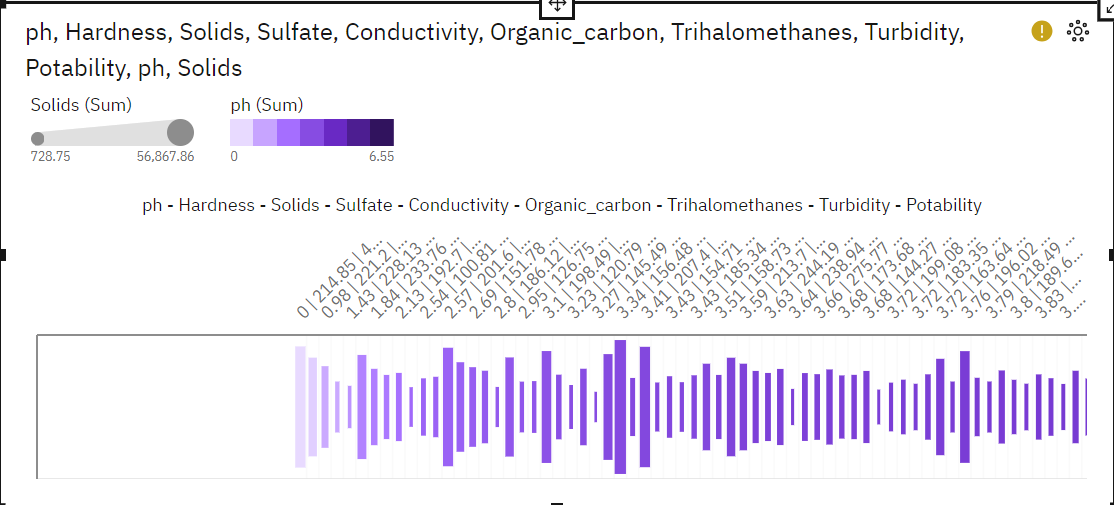
A Bar graph was plotted for the given data:

A graph with colorful bars

Description automatically generated with medium confidence

* The overall number of results for Chloramines, Conductivity, Hardness, Organic carbon, Potability and Turbidity is almost 3500.
* The overall number of results for Sulphate is almost 2500.
* The overall number of results for Trihalomethanes and ph is over three thousand.

An heat map was plotted for the given data:



Line graph :

A graph with a line and numbers

Description automatically generated with medium confidence

* Conductivity ranges from 1788, when ph (Group) is less than 1.4000000000000000, to 1823, when ph (Group) is 12.5999999999999980 and above.
* Hardness ranges from 623.5, when ph (Group) is 12.5999999999999980 and above, to 721.8, when ph (Group) is less than 1.4000000000000000.
* Organic\_carbon ranges from 52.48, when ph (Group) is less than 1.4000000000000000, to 57.87, when ph (Group) is 12.5999999999999980 and above.
* Potability ranges from 1, when ph (Group) is 12.5999999999999980 and above, to 2, when ph (Group) is less than 1.4000000000000000.
* Solids ranges from over 75 thousand, when ph (Group) is 12.5999999999999980 and above, to almost 137 thousand, when ph (Group) is less than 1.4000000000000000.
* Sulfate ranges from 1062, when ph (Group) is less than 1.4000000000000000, to 1091, when ph (Group) is 12.5999999999999980 and above.

Pie chart:

A pie chart with numbers and a few letters

Description automatically generated with medium confidence

* Over all values of Sulfate - Conductivity - Turbidity - Potability - ph, the sum of ph is 6.
* ph ranges from 0, when Sulfate - Conductivity - Turbidity - Potability - ph is (no value)|583.4488493097667|4.928839979650702|0|0, to 3, when Sulfate - Conductivity - Turbidity - Potability - ph is 444.3757306927798|322.29119123589123|5.1606037881606035|1|0.9899122128791387.
* For ph, the most significant values of Sulfate - Conductivity - Turbidity - Potability - ph are 444.3757306927798|322.29119123589123|5.1606037881606035|1|0.9899122128791387 and 333.6778429254695|439.1127646925047|2.533996269917253|0|0.975577989772022, whose respective ph values add up to 5, or 83.3 % of the total.

**CONCLUSION**:

Thus the given dataset was successfully cleaned and visualised using IBM Cognos .