**DATA ANALYTICS WITH COGNOS -**

**GROUP 5**

**PROJECT: WATER QUALITY ANALYSIS**

**PHASE 4: DEVELOPMENT PART 2**

**SUBMITTED BY**

**Joslin Pinu .P: 963321106050.**

# WATER QUALITY ANALYSIS

**CONTINUE BUILDING THE ANALYSIS BY CREATING VISUALIZATIONS AND BUILDING A PREDICTIVE MODEL**

**Introduction:**

Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level.

In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions.

Some of the water quality parameters are,

* pH value
* Hardness
* Total Dissolved Solids
* Chloramines
* Sulfate
* Conductivity
* Organic carbon
* Trihalomethanes
* Turbidity  Potability

**1. Data Preparation:**

* Import necessary libraries (e.g., pandas, numpy, matplotlib, scikitlearn).
* Load your dataset.
* Explore and preprocess your data. This includes handling missing values, encoding categorical variables, and scaling numerical features.

**2.Exploratory Data Analysis (EDA):**

* Create visualizations to better understand your data. Common libraries for this are Matplotlib and Seaborn.
* Examples of visualizations: histograms, scatter plots, box plots, etc., depending on your data type.

**3.Feature Engineering:**

* If needed, create new features or transform existing ones to improve the performance of your predictive model.

**4.Splitting Data:**

* Split your data into training and testing sets to evaluate your model.

**5.Building a Predictive Model:**

* Select an appropriate algorithm for your problem (e.g., linear regression, decision tree, random forest, or neural network).
* Train your model on the training data.
* Evaluate its performance using metrics like mean squared error (MSE), R-squared, etc.

**6.Predictions:**

* Make predictions on your test data.

**7.Visualize Predictions:**

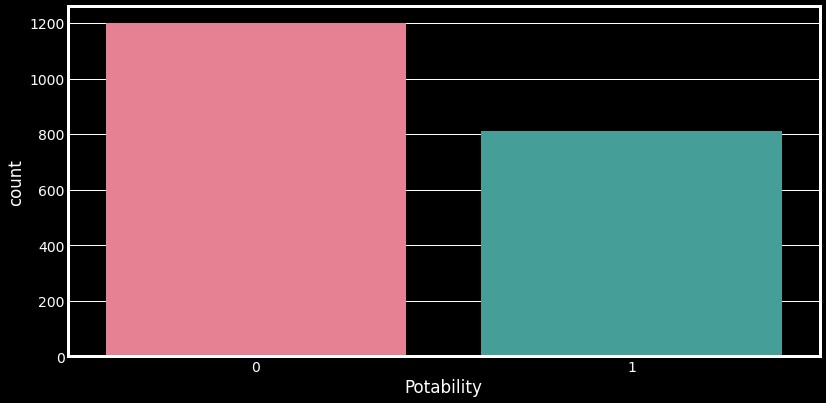
* Create a bar chart or any other suitable visualization to display the predicted values alongside the actual values for comparison.

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

**Program:**

|  |
| --- |
| import matplotlib.pyplot as plt import matplotlib.pyplot as plt plt.style.use('fivethirtyeight') plt.style.use('dark\_background') import numpy as np import pandas as pd import seaborn as sns  from matplotlib.colors import ListedColormap from scipy.stats import norm, boxcox |
| from sklearn.metrics import confusion\_matrix, classification\_report, accuracy\_ score  from collections import Counter from scipy import stats from tqdm import tqdm\_notebook    *## Importing LuciferML*  from luciferml.supervised.classification import Classification from luciferml.preprocessing import Preprocess as prep    import warnings  warnings.simplefilter(action='ignore', category=**Warning**) plt.figure(figsize=(12, 6))  sns.countplot(x="Potability", data=dataset, palette='husl'); |

**Output:**



**Program:**

**import** pandas as pd **import** pandas as pd

**import** matplotlib.pyplot as plt

# reading the database data **=** pd.read\_csv("tips.csv")

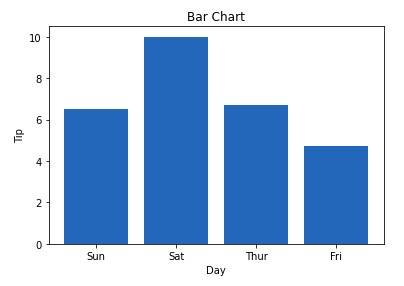
# Bar chart with day against tip plt.bar(data['day'], data['tip'])

plt.title("Bar Chart")

# Setting the X and Y labels plt.xlabel('Day') plt.ylabel('Tip')

# Adding the legends plt.show()

**Output:**



**Conclusion:**

Good data visualization should communicate a data set clearly and effectively by using graphics. The best visualizations make it easy to comprehend data at a glance.