DEVELOPMENT PART 2:

DATA ANALYSIS:

- Import and prepare your water quality data in IBM Cognos for analysis.
- Explore and clean the dataset to handle missing values and outliers.
- Calculate relevant statistical measures to gain insights into the data.

MODEL BUILDING:

- Depending on your goals, you can build various models, such as regression models to predict water quality parameters or classification models to detect anomalies.
- Use Cognos for data modeling and to create and train machine learning models.

EVALUATION:

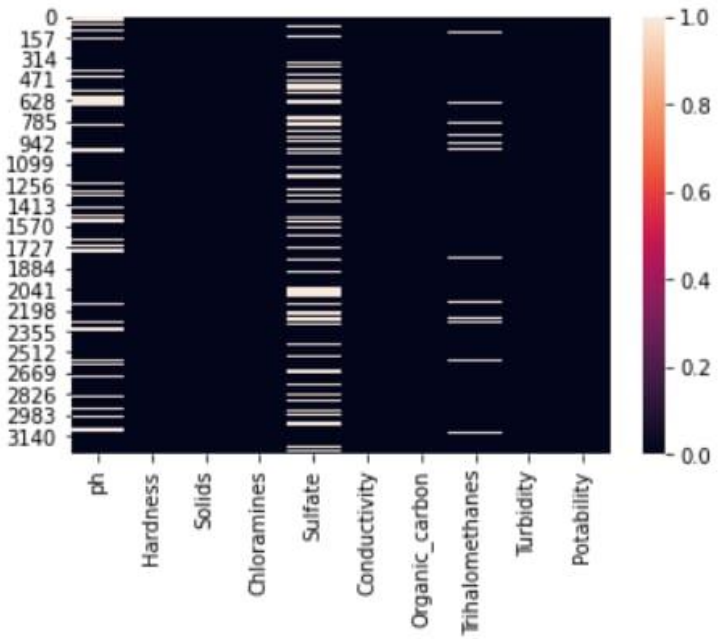
- Assess the performance of your models by using metrics like RMSE (Root Mean Square Error), R-squared, or classification accuracy, depending on the model type.
- Evaluate the models' robustness and generalization to ensure their reliability.

VISUALIZATION:

- Create visualizations in IBM Cognos to represent your data and model results effectively.
- Use charts, graphs, and dashboards to present water quality trends, correlations, and model predictions.

sns.heatmap(df.isnull())

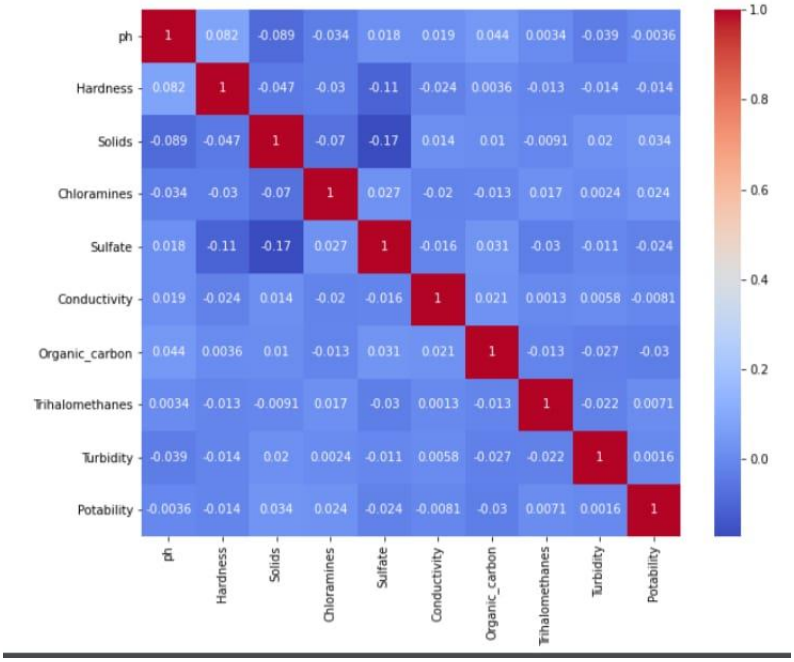
OUTPUT:



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

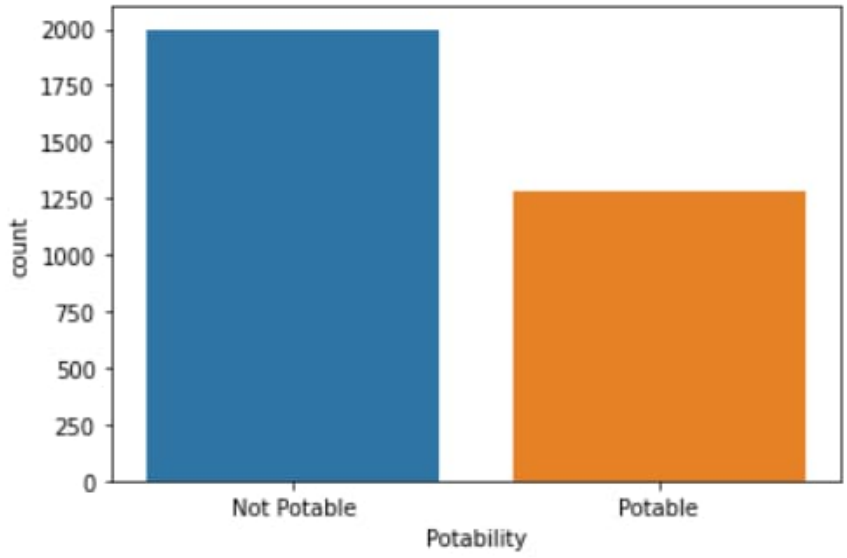
sns.heatmap(df.corr(), annot= True, cmap='coolwarm')

OUTPUT:



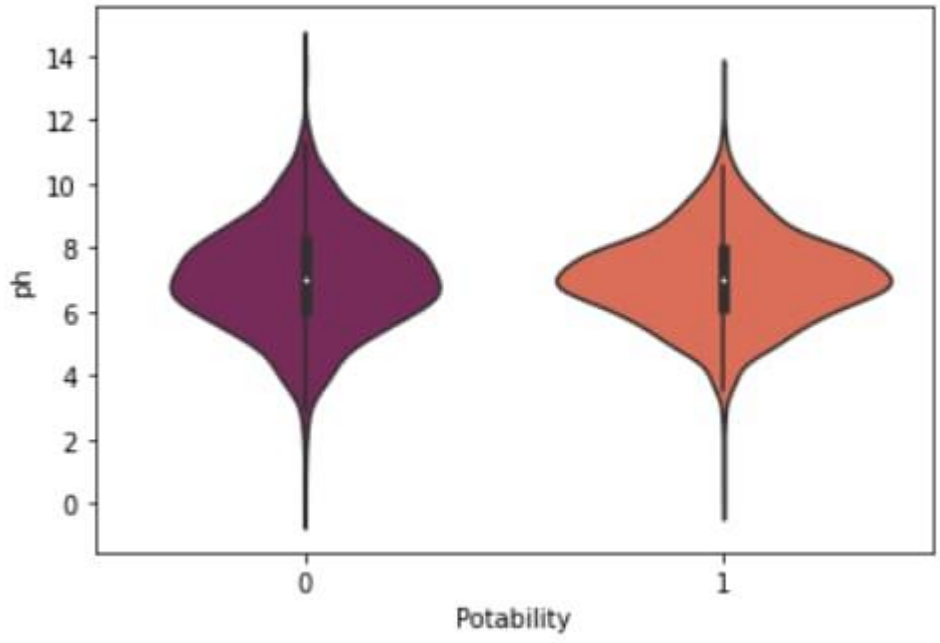
```
ax = sns.countplot(x = "Potability",data= df, saturation=0.8)
plt.xticks(ticks=[0, 1], labels = ["Not Potable", "Potable"])
plt.show()
```

OUTPUT:



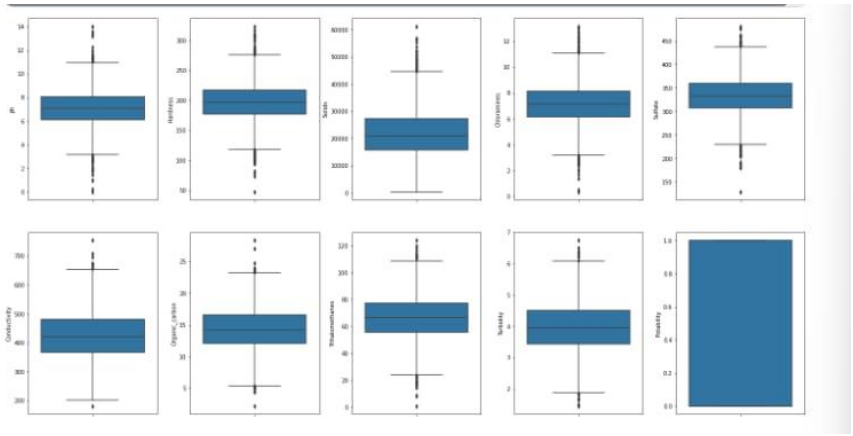

```
sns.violinplot(x='Potability', y='ph', data=df, palette='rocket')
```

OUTPUT:



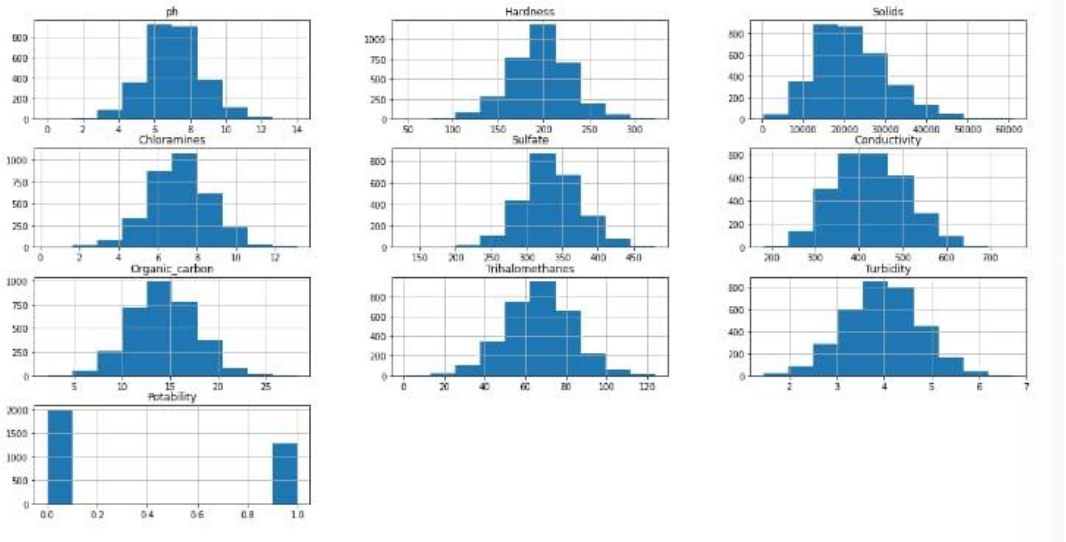
```
fig, ax = plt.subplots(ncols = 5, nrows = 2, figsize = (20, 10))
index = 0
ax = ax.flatten()
for col, value in df.items():
    sns.boxplot(y=col, data=df, ax=ax[index])
    index += 1
plt.tight_layout(pad = 0.5, w_pad=0.7, h_pad=5.0)
```

OUTPUT:



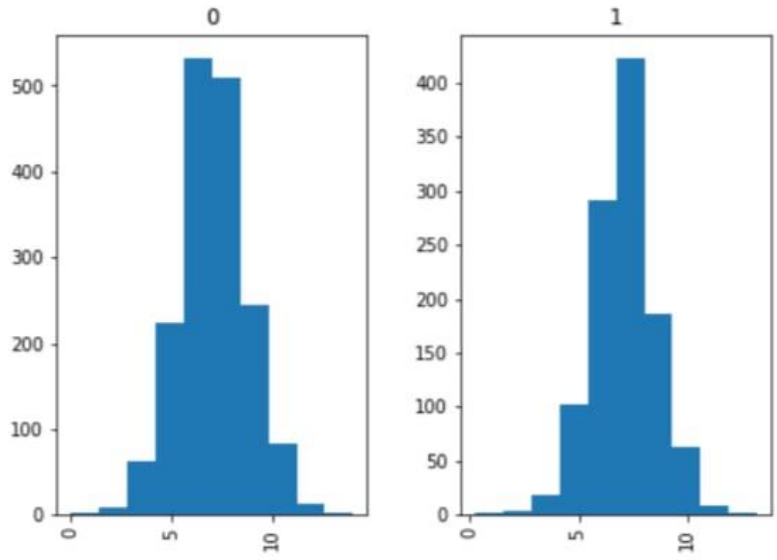
```
plt.rcParams['figure.figsize'] = [20,10]
df.hist()
plt.show()
```

OUTPUT:



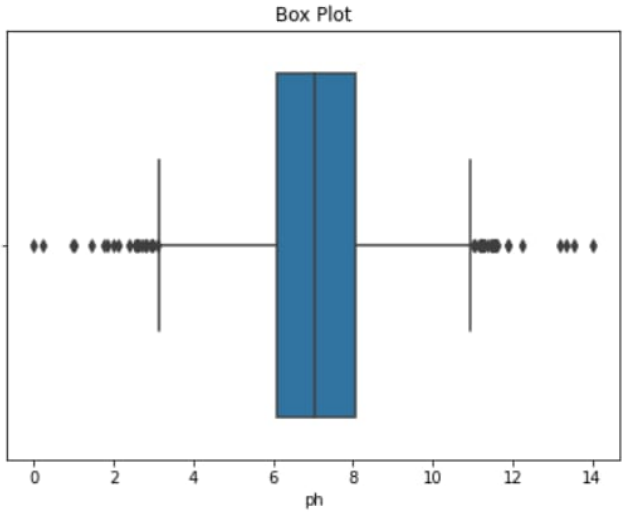
```
plt.rcParams['figure.figsize'] = [7,5]
sns.distplot(df['Potability'])
```

OUTPUT:



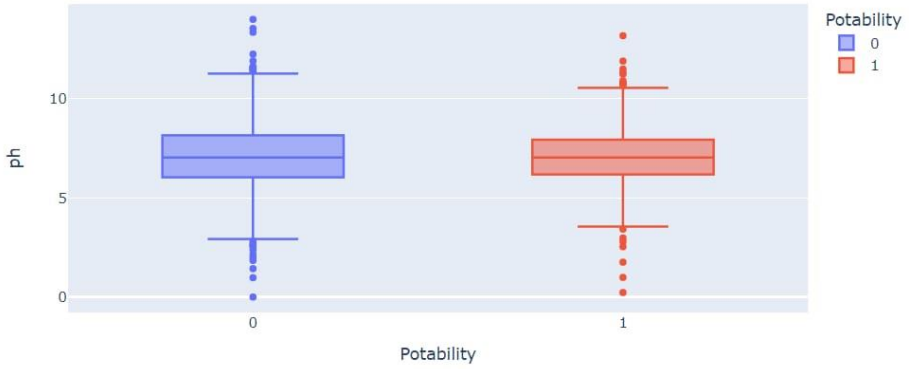

```
# Individual box plot for each feature
def Box(df):
    plt.title("Box Plot")
    sns.boxplot(df)
    plt.show()
Box(df['ph'])
```

OUTPUT:



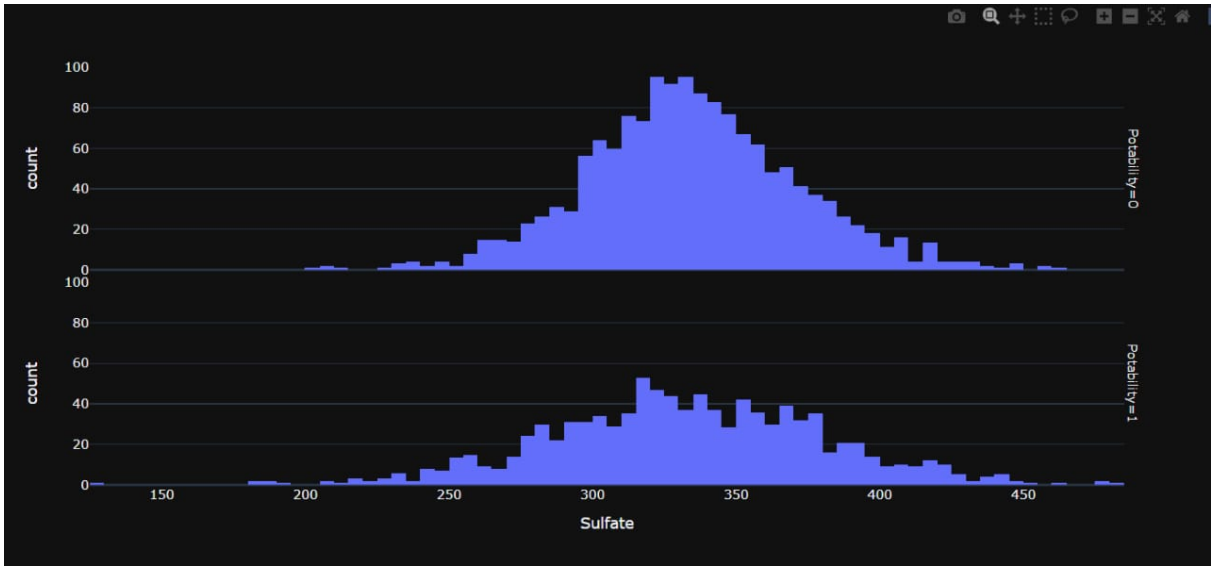
```
fig = px.box(df, x="Potability", y="ph", color="Potability", width=800, height=400)
fig.show()
```

OUTPUT:



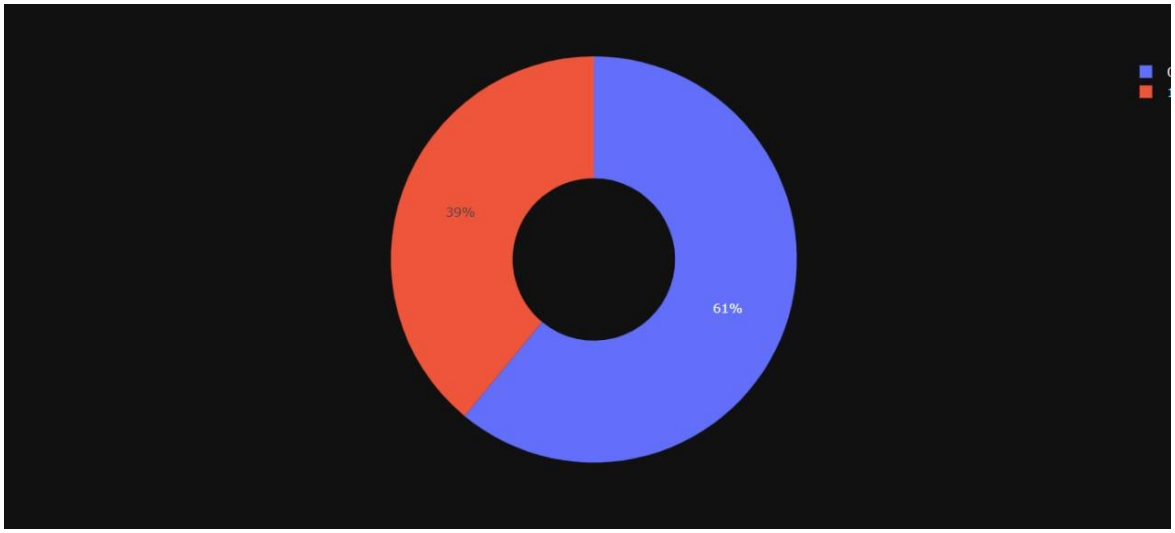

```
fig = px.histogram (df, x = "Sulfate",  facet_row = "Potability",  template = 'plotly_dark')
fig.show ()
```

OUTPUT:



```
fig =  px.pie (df, names = "Potability", hole = 0.4, template = "plotly_dark")
fig.show ()
```

OUTPUT:



Project Conclusion:

In conclusion, data analytics is an indispensable tool in the field of water quality analysis, enabling us to gain valuable insights into the health and safety of water sources. By harnessing the power of data analytics, we can make informed decisions that impact public health, environmental sustainability, and resource management