**PHASE 3**

**DEVELOPMENT PART 1**

**Steps for preprocessing:**

* Previously we selected Support Vector Machine algorithm.
* In this part we will begin building our project by loading and preprocessing the dataset.
* Preprocessing the data and performing exploratory data analysis for the given water quality analysis test.
* Handling missing values and outliers by isnull methods in python.

**# IMPORTING PACKAGES**

import pandas as pd

import seaborn as sns

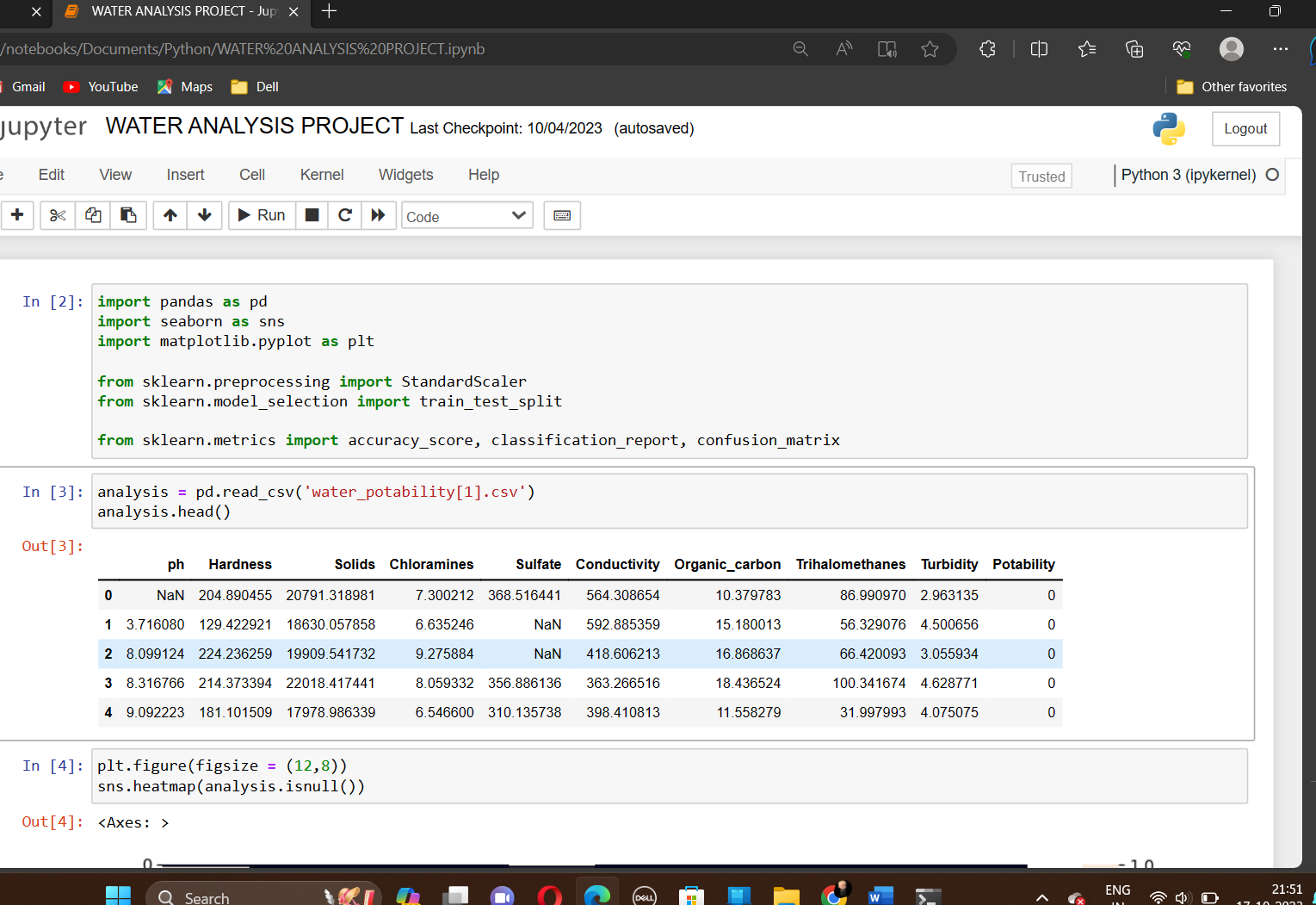
import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

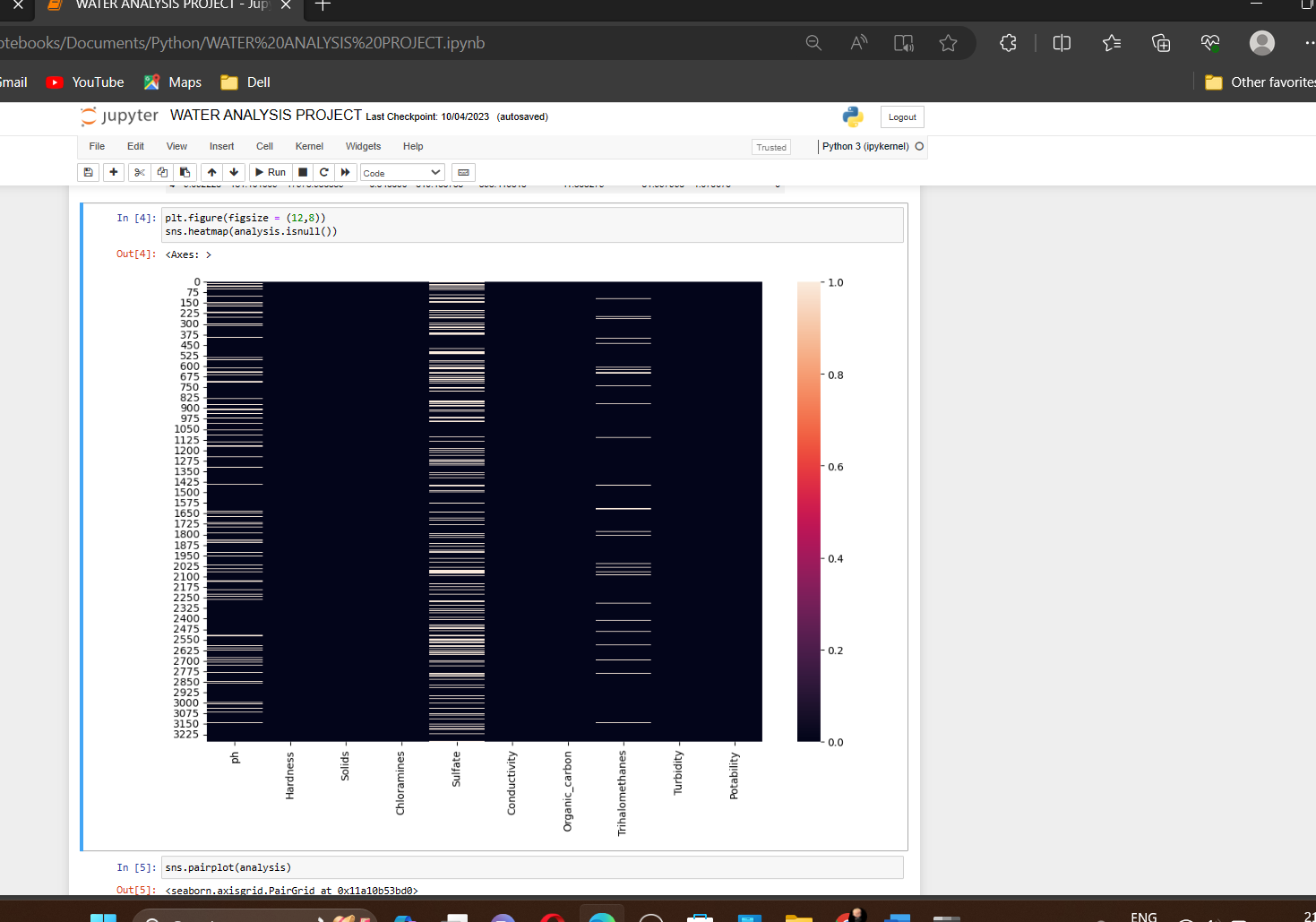
from sklearn.model\_selection import train\_test\_split

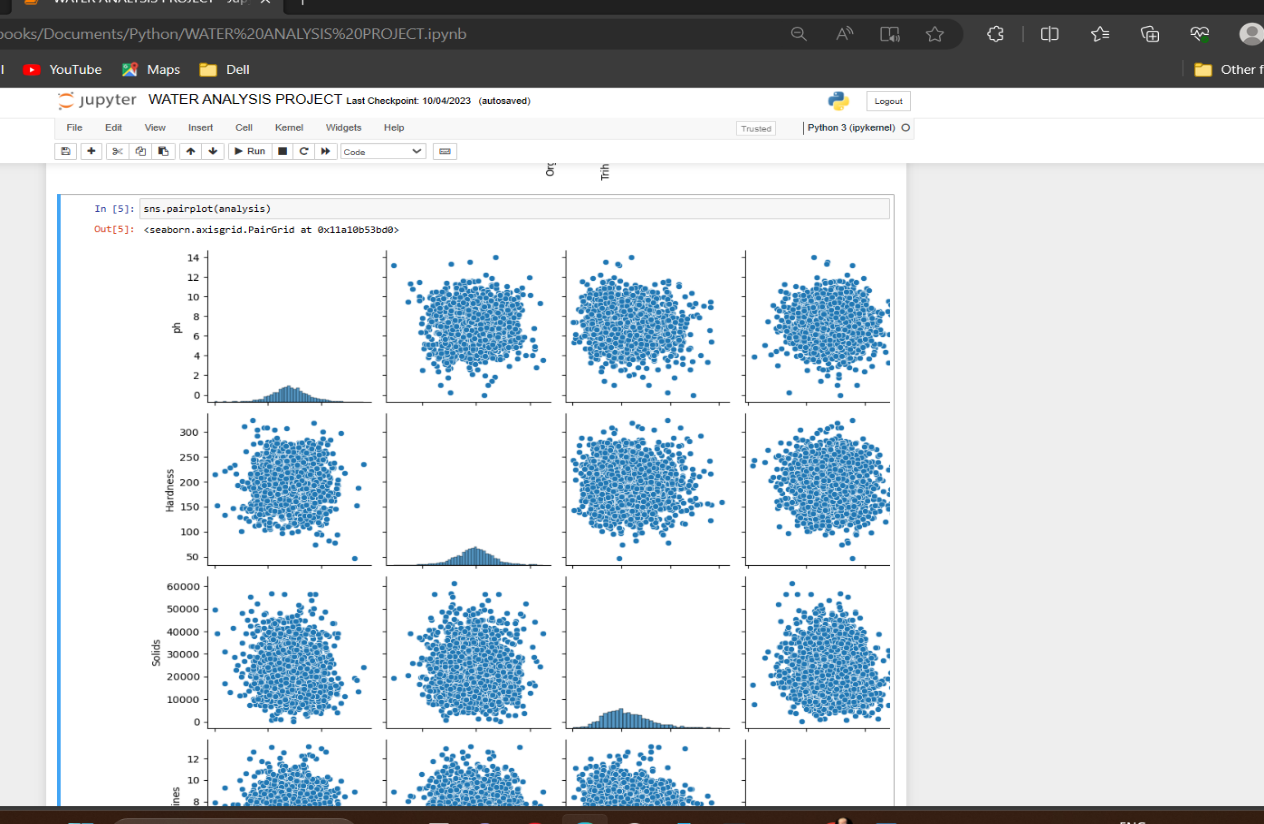
from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix

**#Deriving datasets using pandas**



**#Visualization using heatmap**



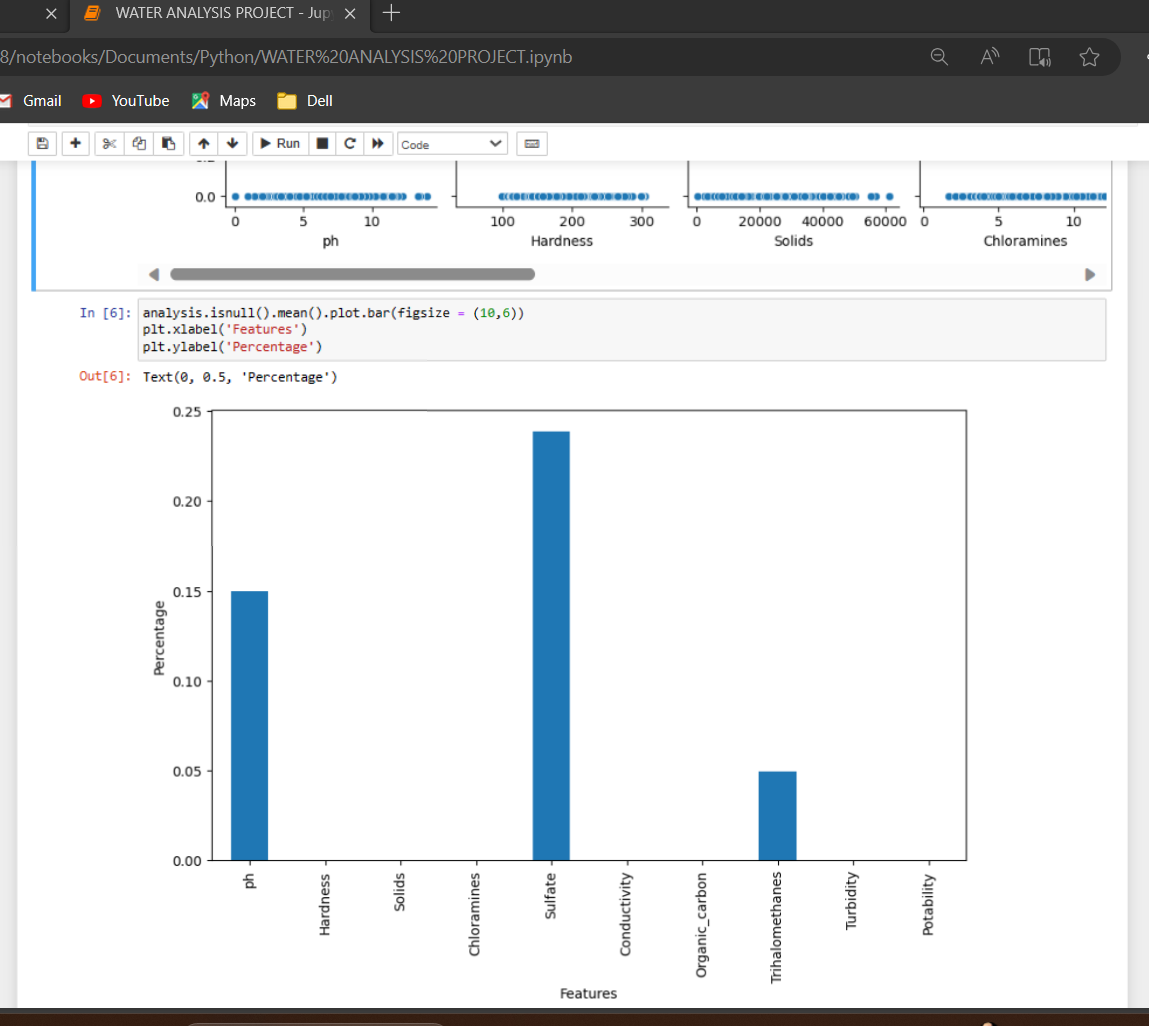
**#Visualization using Pairplots**

**#Visualization using Barplots**

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

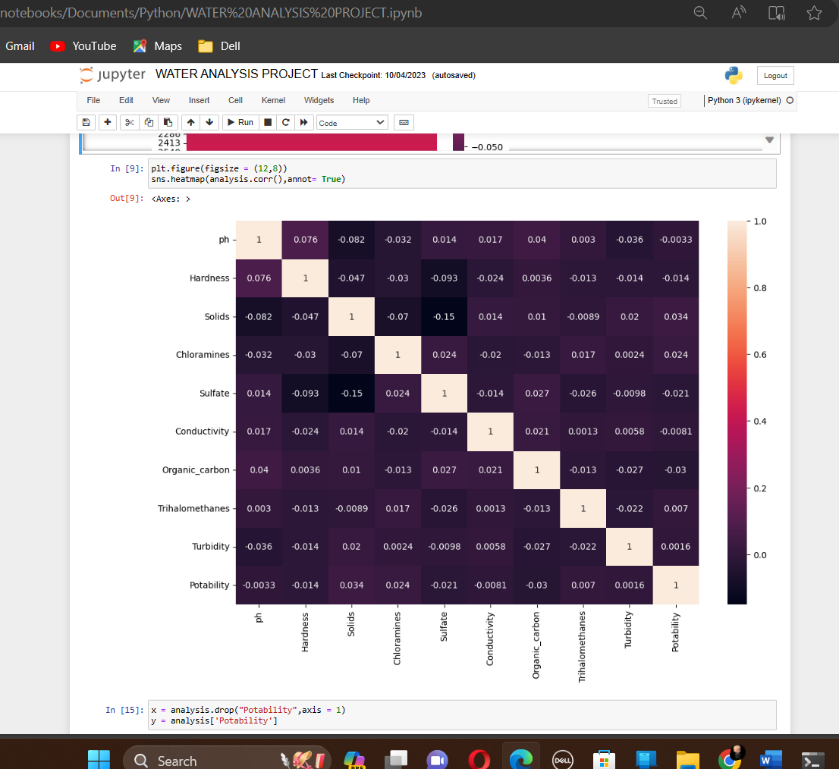
plt.xlabel('Features')

plt.ylabel('Percentage')



**#Final Visualization using Elements**

[{"metadata":{"trusted":true},"id":"922b6b0d","cell\_type":"code","source":"analysis['ph'] = analysis['ph'].fillna(analysis['ph'].mean())\nanalysis['Sulfate'] = analysis['Sulfate'].fillna(analysis['Sulfate'].mean())\nanalysis['Trihalomethanes'] = analysis['Trihalomethanes'].fillna(analysis['Trihalomethanes'].mean())","execution\_count":23,"outputs":[]}]



In this WATER QUALITY ANALYSIS, we represent various visualization methods in order to understand the relation and correlations between the datasets.