#Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn import metrics

from sklearn.neighbors import KNeighborsClassifier

from sklearn.linear\_model import LogisticRegression

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

#Read CSV data

df=pd.read\_csv("https://github.com/YBI-Foundation/Dataset/raw/main/Air%20Quality%20Missing%20Data.csv")

#Display top 5 rows of dataset

print("Displaying first 5 rows of DataFrame : ")

print(df.head())

#Display last 5 rows of dataset

print("\n\nDisplaying last 5 rows of DataFrame : ")

print(df.tail())

#Display shape of DataFrame

print("\n\nShape of DataFrame : ")

print(df.shape)

#Display column names of DataFrame

print("\n\nThe Column name of DataFrame : ")

print(df.columns)

#Display data types of each column

print("\n\nData types of each column : ")

print(df.dtypes)

#Display information about dataset like total no. of rows,total no. of columns,datatype of each column and memory requirement

print("\n\nInformation about Dataset : ")

print(df.info())

#Display summary statistics of DataFrame

print("\n\nSummary Statistics of DataFrame : ")

print(df.describe(include='all'))

#Display the count the no. of unique values in each column

print("\n\nCount of unique values in each column : ")

print(df.nunique())

#Display the number of missing values in each column

print("\n\nThe number of missing values in each column : ")

print(df.isnull().sum())

#Fill missing values with mean(for numerical columns)

df.fillna(df.mean(numeric\_only=True),inplace=True)

#Drop duplicates

df.drop\_duplicates(inplace=True)

print("\n\nDataFrame after dropping duplicates : ")

print(df.head())

#Return the number of dimensions of an array

print("Return the no. of dimensions of an array : ")

print(df.ndim)

#Return the size of frame along the axis

print("Return the size of frame along the axis : ")

print(df.size)

#Assuming df is your DataFraame and 'Temp' is the column

Temp\_dummies=pd.get\_dummies(df['Temp'],prefix='Temp')

print(Temp\_dummies)

#Assuming df is your original DataFrame and Temp\_dummies is the DataFrame with dummy variables

df\_with\_dummies=pd.concat([df['Wind'],Temp\_dummies],axis=1)

print(df\_with\_dummies)

#Merging data with itself

merged=pd.merge(df,df,on='Temp')

print(merged.info())