**BIGDATA ANALYTICS-DA1**

**CODE**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**# Define the file path**

**file\_path = r"C:\Users\HP\Desktop\emc java\AirQuality.csv" # Use raw string for Windows paths**

**# Load the data**

**try:**

**df = pd.read\_csv(file\_path, delimiter=';')**

**print("Data loaded successfully:")**

**print(df.head()) # Print the first few rows of the DataFrame**

**except FileNotFoundError:**

**print(f"File not found at {file\_path}. Please check the path.")**

**exit(1)**

**# Check the shape of the DataFrame**

**print(f"DataFrame shape: {df.shape}")**

**# Inspect the DataFrame columns**

**print("Columns in the DataFrame:")**

**print(df.columns.tolist())**

**# Replace commas with dots and convert numerical columns to proper format**

**df['CO(GT)'] = df['CO(GT)'].str.replace(',', '.').astype(float)**

**# Combine 'Date' and 'Time' into a single datetime column**

**df['Datetime'] = pd.to\_datetime(df['Date'] + ' ' + df['Time'], format='%d/%m/%Y %H.%M.%S', errors='coerce')**

**# Check for NaN values in relevant columns**

**print("Count of NaN values in relevant columns:")**

**print(df[['Datetime', 'CO(GT)']].isnull().sum())**

**# Drop rows with NaN values in the relevant columns**

**df = df.dropna(subset=['Datetime', 'CO(GT)'])**

**# Verify that we have data to plot**

**print(f"DataFrame shape after dropping NaNs: {df.shape}")**

**# Plot the data**

**plt.figure(figsize=(10, 5))**

**plt.plot(df['Datetime'], df['CO(GT)'], label='CO(GT)', color='blue')**

**plt.xlabel('Datetime')**

**plt.ylabel('CO(GT)')**

**plt.title('CO(GT) Levels Over Time')**

**plt.legend()**

**plt.xticks(rotation=45) # Rotate x-axis labels for better visibility**

**plt.tight\_layout() # Adjust layout for better spacing**

**plt.show()**