

Implement programs for time series data cleaning, loading and handling time series data and pre-processing techniques

```
import pandas as pd
import matplotlib.pyplot as plt

# Step 1: Load the dataset with the correct delimiter (semicolon)
file_path = r"C:\Users\AI_LAB\Downloads\archive (4)\AirQuality.csv"
df = pd.read_csv(file_path, delimiter=";")

# Step 2: Inspect the column names
print(df.columns)

# Step 3: Clean the 'Time' column by replacing periods with colons
df['Time'] = df['Time'].str.replace('.', ':', regex=False)

# Step 4: Combine 'Date' and 'Time' columns into a single 'datetime' column
df['datetime'] = pd.to_datetime(df['Date'] + ' ' + df['Time'], format='%d/%m/%Y %H:%M:%S')

# Step 5: Drop the original 'Date' and 'Time' columns (optional)
df.drop(['Date', 'Time'], axis=1, inplace=True)

# Step 6: Set 'datetime' as the index
df.set_index('datetime', inplace=True)

# Step 7: Handle missing data by forward filling
df.fillna(method='ffill', inplace=True)

# Step 8: Resample the data to daily frequency (mean of each day)
df_daily = df.resample('D').mean()
```

```
# Step 9: Apply a 7-day moving average to 'NO2(GT)' column
```

```
df_daily['NO2_SMA_7'] = df_daily['NO2(GT)'].rolling(window=7).mean()
```

```
# Step 10: Visualize the data
```

```
plt.figure(figsize=(12, 6))
```

```
plt.plot(df_daily.index, df_daily['NO2(GT)'], label='NO2(GT)', color='blue', alpha=0.7)
```

```
plt.plot(df_daily.index, df_daily['NO2_SMA_7'], label='7-day Moving Average', color='red',  
linestyle='--')
```

```
plt.title('NO2 Levels with 7-day Moving Average')
```

```
plt.xlabel('Date')
```

```
plt.ylabel('NO2 Levels')
```

```
plt.legend()
```

```
plt.grid(True)
```

```
plt.show()
```

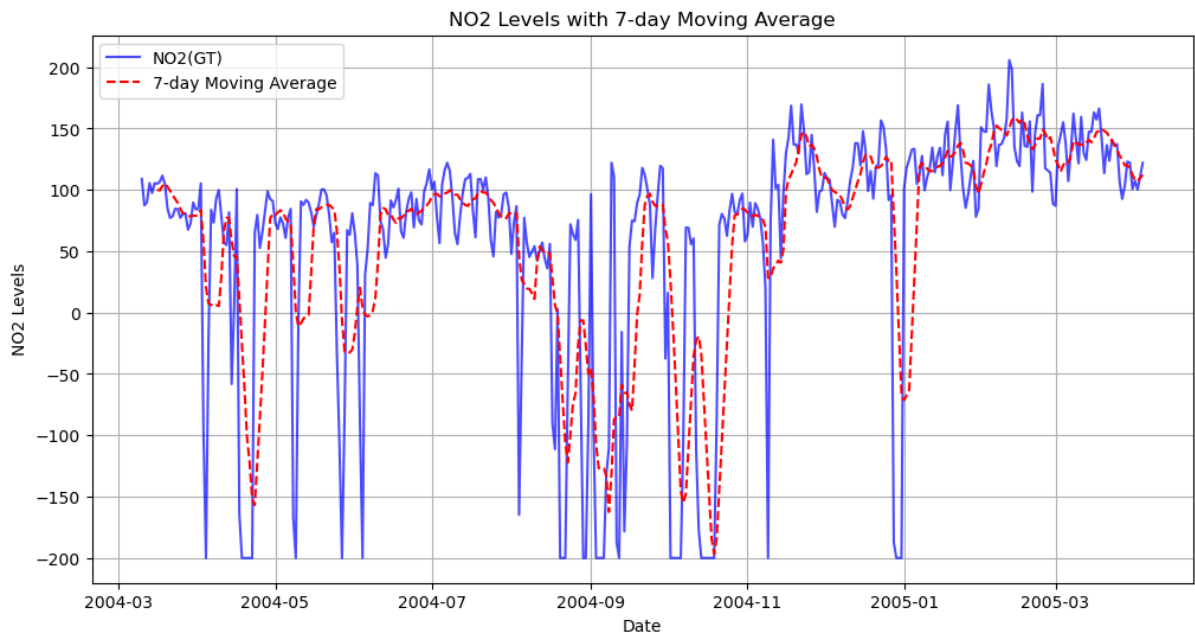
```
# Step 11: Save the cleaned data to a CSV file
```

```
df_daily.to_csv(r"C:\Users\AI_LAB\Downloads\cleaned_air_quality_data.csv")
```

```
# Check the cleaned data
```

```
print(df_daily.head())
```

OUTPUT:



	PT08.S1(CO)	NMHC(GT)	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)	\
datetime						
2004-03-10	1316.500000	86.500000	912.333333	132.000000	1167.333333	
2004-03-11	1244.166667	104.500000	851.958333	130.041667	1277.250000	
2004-03-12	1281.666667	141.500000	1008.291667	142.583333	1101.875000	
2004-03-13	1330.666667	139.250000	992.833333	168.416667	993.208333	
2004-03-14	1361.125000	116.958333	943.916667	132.166667	1001.291667	

	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	Unnamed: 15	Unnamed: 16	\
datetime						
2004-03-10	108.833333	1545.500000	1096.000000	NaN	NaN	
2004-03-11	87.375000	1522.833333	885.250000	NaN	NaN	
2004-03-12	89.916667	1627.291667	1084.375000	NaN	NaN	
2004-03-13	105.583333	1595.791667	1245.916667	NaN	NaN	
2004-03-14	97.458333	1602.375000	1234.208333	NaN	NaN	

	NO2_SMA_7
datetime	
2004-03-10	NaN
2004-03-11	NaN
2004-03-12	NaN
2004-03-13	NaN
2004-03-14	NaN