import pandas as pd

# 读取四个电站的小时级数据文件

station1\_data = pd.read\_excel('/mnt/data/station1\_hourly\_data.xlsx')

station2\_data = pd.read\_excel('/mnt/data/station2\_hourly\_data.xlsx')

station3\_data = pd.read\_excel('/mnt/data/station3\_hourly\_data.xlsx')

station4\_data = pd.read\_excel('/mnt/data/station4\_hourly\_data.xlsx')

# 查看数据的结构，检查前几行

station1\_data.head(), station2\_data.head(), station3\_data.head(), station4\_data.head()

结果

( Hour ActualEnergy\_kWh Irradiance\_w\_m2 CurrentTemperature \

0 2024-05-01 02:00:00 0.000000 0.000000 6.0

1 2024-05-01 03:00:00 0.000000 0.000000 NaN

2 2024-05-01 04:00:00 0.000000 0.000000 4.0

3 2024-05-01 05:00:00 1.416586 1.157895 NaN

4 2024-05-01 06:00:00 556.718442 46.950000 4.0

MaxTemperature MinTemperature WindSpeed Humidity 电站编号 装机容量

0 21.0 7.0 0.50 52.0 1 4998.3

1 NaN NaN NaN NaN 1 4998.3

2 21.0 7.0 0.81 64.0 1 4998.3

3 NaN NaN NaN NaN 1 4998.3

4 21.0 7.0 0.00 71.0 1 4998.3 ,

Hour ActualEnergy\_kWh Irradiance\_w\_m2 CurrentTemperature \

0 2024-05-01 02:00:00 0.000000 0.000000 19.0

1 2024-05-01 03:00:00 0.000000 0.000000 NaN

2 2024-05-01 04:00:00 0.000000 0.000000 19.0

3 2024-05-01 05:00:00 0.000000 0.000000 NaN

4 2024-05-01 06:00:00 478.179983 44.578947 19.0

MaxTemperature MinTemperature WindSpeed Humidity 电站编号 装机容量

0 20.0 18.0 1.61 97.0 2 5581

1 NaN NaN NaN NaN 2 5581

2 20.0 18.0 0.00 98.0 2 5581

3 NaN NaN NaN NaN 2 5581

4 20.0 18.0 4.00 95.0 2 5581 ,

Hour ActualEnergy\_kWh Irradiance\_w\_m2 CurrentTemperature \

0 2024-05-01 02:00:00 0.00000 0.0 19.0

1 2024-05-01 03:00:00 0.00000 0.0 NaN

2 2024-05-01 04:00:00 0.00000 0.0 19.0

3 2024-05-01 05:00:00 0.00000 0.0 NaN

4 2024-05-01 06:00:00 194.25375 15.0 19.0

MaxTemperature MinTemperature WindSpeed Humidity Unnamed: 10 \

0 22.0 17.0 1.11 100.0 1.0

1 NaN NaN NaN NaN NaN

2 22.0 17.0 1.19 100.0 1.0

3 NaN NaN NaN NaN NaN

4 22.0 17.0 0.00 100.0 1.0

Unnamed: 11 Unnamed: 12 Unnamed: 14 Unnamed: 15 Unnamed: 16 \

0 21.0 19.0 NaN 1.0 6100.0

1 NaN NaN NaN NaN NaN

2 21.0 19.0 NaN 1.0 4300.0

3 NaN NaN NaN NaN NaN

4 22.0 19.0 NaN 0.0 5200.0

Unnamed: 17 Unnamed: 18 Unnamed: 19 电站编号 装机容量

0 180.0 NaN NaN 3 4456

1 NaN NaN NaN 3 4456

2 0.0 NaN NaN 3 4456

3 NaN NaN NaN 3 4456

4 0.0 NaN NaN 3 4456 ,

Hour ActualEnergy\_kWh Irradiance\_w\_m2 CurrentTemperature \

0 2024-05-01 02:00:00 0.000000 0.000000 15.0

1 2024-05-01 03:00:00 0.000000 0.000000 NaN

2 2024-05-01 04:00:00 0.000000 0.052632 15.0

3 2024-05-01 05:00:00 85.822202 41.842105 NaN

4 2024-05-01 06:00:00 1227.709186 266.888889 15.0

MaxTemperature MinTemperature WindSpeed Humidity Unnamed: 10 \

0 27.0 15.0 3.19 41.0 1.0

1 NaN NaN NaN NaN NaN

2 27.0 15.0 5.11 41.0 1.0

3 NaN NaN NaN NaN NaN

4 27.0 15.0 3.81 43.0 4.0

Unnamed: 11 Unnamed: 12 Unnamed: 14 Unnamed: 15 Unnamed: 16 \

0 13.0 15.0 NaN 2.0 23500.0

1 NaN NaN NaN NaN NaN

2 11.0 15.0 NaN 3.0 26200.0

3 NaN NaN NaN NaN NaN

4 12.0 15.0 NaN 3.0 22700.0

Unnamed: 17 Unnamed: 18 Unnamed: 19 电站编号 装机容量

0 270.0 NaN NaN 4 1794.61

1 NaN NaN NaN 4 1794.61

2 270.0 NaN NaN 4 1794.61

3 NaN NaN NaN 4 1794.61

4 270.0 NaN NaN 4 1794.61 )

数据分析

# 计算PR值并找出疑似清洗时间点

# 假设装机容量为4998.3kWp（每个电站的装机容量不同，我们可以在每个表格中提取）

installed\_capacity\_station1 = 4998.3 # 电站1装机容量

installed\_capacity\_station2 = 5581 # 电站2装机容量

installed\_capacity\_station3 = 4456 # 电站3装机容量

installed\_capacity\_station4 = 1794.61 # 电站4装机容量

# 计算理论发电量

def calculate\_theoretical\_energy(df, installed\_capacity):

# 理论发电量 = 辐照强度 \* 装机容量 \* 转换效率（假设转换效率为1）

df['TheoreticalEnergy\_kWh'] = df['Irradiance\_w\_m2'] \* installed\_capacity

return df

# 计算PR值

def calculate\_pr(df):

df['PR'] = df['ActualEnergy\_kWh'] / df['TheoreticalEnergy\_kWh']

return df

# 计算每个电站的数据

station1\_data = calculate\_theoretical\_energy(station1\_data, installed\_capacity\_station1)

station2\_data = calculate\_theoretical\_energy(station2\_data, installed\_capacity\_station2)

station3\_data = calculate\_theoretical\_energy(station3\_data, installed\_capacity\_station3)

station4\_data = calculate\_theoretical\_energy(station4\_data, installed\_capacity\_station4)

station1\_data = calculate\_pr(station1\_data)

station2\_data = calculate\_pr(station2\_data)

station3\_data = calculate\_pr(station3\_data)

station4\_data = calculate\_pr(station4\_data)

# 清洗预警规则（PR值低于0.8，实际发电量与理论发电量差异大，湿度高且辐照强度低）

def cleaning\_alert(df, pr\_threshold=0.8, energy\_diff\_threshold=1000, humidity\_threshold=70, irradiance\_threshold=500):

alerts = []

for i in range(len(df)):

# 清洗预警条件

if df['PR'][i] < pr\_threshold:

alerts.append(True)

elif abs(df['ActualEnergy\_kWh'][i] - df['TheoreticalEnergy\_kWh'][i]) > energy\_diff\_threshold:

alerts.append(True)

elif df['Humidity'][i] > humidity\_threshold and df['Irradiance\_w\_m2'][i] < irradiance\_threshold:

alerts.append(True)

else:

alerts.append(False)

return alerts

# 应用清洗预警规则

station1\_data['CleaningAlert'] = cleaning\_alert(station1\_data)

station2\_data['CleaningAlert'] = cleaning\_alert(station2\_data)

station3\_data['CleaningAlert'] = cleaning\_alert(station3\_data)

station4\_data['CleaningAlert'] = cleaning\_alert(station4\_data)

# 获取清洗预警的时间点

alert\_station1 = station1\_data[station1\_data['CleaningAlert'] == True]

alert\_station2 = station2\_data[station2\_data['CleaningAlert'] == True]

alert\_station3 = station3\_data[station3\_data['CleaningAlert'] == True]

alert\_station4 = station4\_data[station4\_data['CleaningAlert'] == True]

# 输出清洗预警的时间点

alert\_station1[['Hour', 'PR', 'ActualEnergy\_kWh', 'TheoreticalEnergy\_kWh', 'Humidity', 'Irradiance\_w\_m2']]

alert\_station2[['Hour', 'PR', 'ActualEnergy\_kWh', 'TheoreticalEnergy\_kWh', 'Humidity', 'Irradiance\_w\_m2']]

alert\_station3[['Hour', 'PR', 'ActualEnergy\_kWh', 'TheoreticalEnergy\_kWh', 'Humidity', 'Irradiance\_w\_m2']]

alert\_station4[['Hour', 'PR', 'ActualEnergy\_kWh', 'TheoreticalEnergy\_kWh', 'Humidity', 'Irradiance\_w\_m2']]

结果

Hour PR ActualEnergy\_kWh TheoreticalEnergy\_kWh \

2 2024-05-01 04:00:00 0.000000 0.000000 9.445316e+01

3 2024-05-01 05:00:00 0.001143 85.822202 7.509026e+04

4 2024-05-01 06:00:00 0.002563 1227.709186 4.789615e+05

5 2024-05-01 07:00:00 0.005892 5152.719844 8.745418e+05

6 2024-05-01 08:00:00 0.011662 13674.412815 1.172542e+06

... ... ... ... ...

8187 2025-03-09 00:00:00 NaN NaN 0.000000e+00

8189 2025-03-09 22:00:00 NaN NaN 0.000000e+00

8193 2025-03-10 22:00:00 NaN NaN 0.000000e+00

8197 2025-03-11 22:00:00 NaN NaN 0.000000e+00

8213 2025-03-15 22:00:00 NaN NaN 0.000000e+00

Humidity Irradiance\_w\_m2

2 41.0 0.052632

3 NaN 41.842105

4 43.0 266.888889

5 NaN 487.315789

6 37.0 653.368421

... ... ...

8187 74.5 0.000000

8189 75.0 0.000000

8193 73.0 0.000000

8197 86.0 0.000000

8213 93.0 0.000000

[6147 rows x 6 columns]