Data Preparation (Data Cleaning and Processing)

1. ***Load the Data***: Read the Excel file to understand the data structure.
2. ***Data Treatment***: Check for missing values and handle them (either by imputation or removal depending on their frequency).
3. identify outliers in the temperature and draft data using statistical techniques (like z-scores) or IQR for filtering.
4. Normalize or scale the data if necessary to prepare it for further analysis.
5. Resample or aggregate data if required to simplify and focus on trends, rather than every 5-minute data point.

Analysis Strategy :-

Time Series Analysis: Plot trends over time for each variable to observe long-term patterns and seasonal variations.

Correlation Analysis: Calculate correlations between the variables to see how different parts of the cyclone system are interrelated.

Anomaly Detection: Use statistical methods like control charts, z-scores, or machine learning models (such as Isolation Forest or Autoencoders) to identify abnormal readings over time. Define thresholds for each variable based on historical data to flag unusual values.

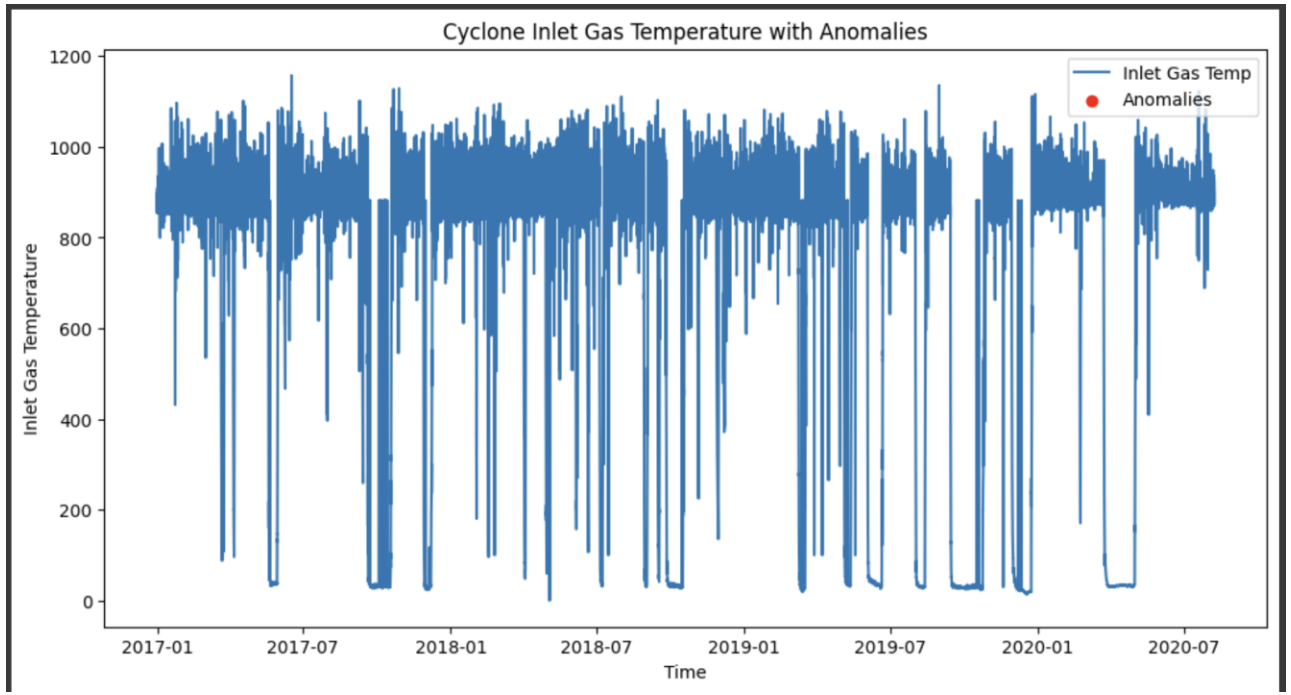
Anomaly Detection

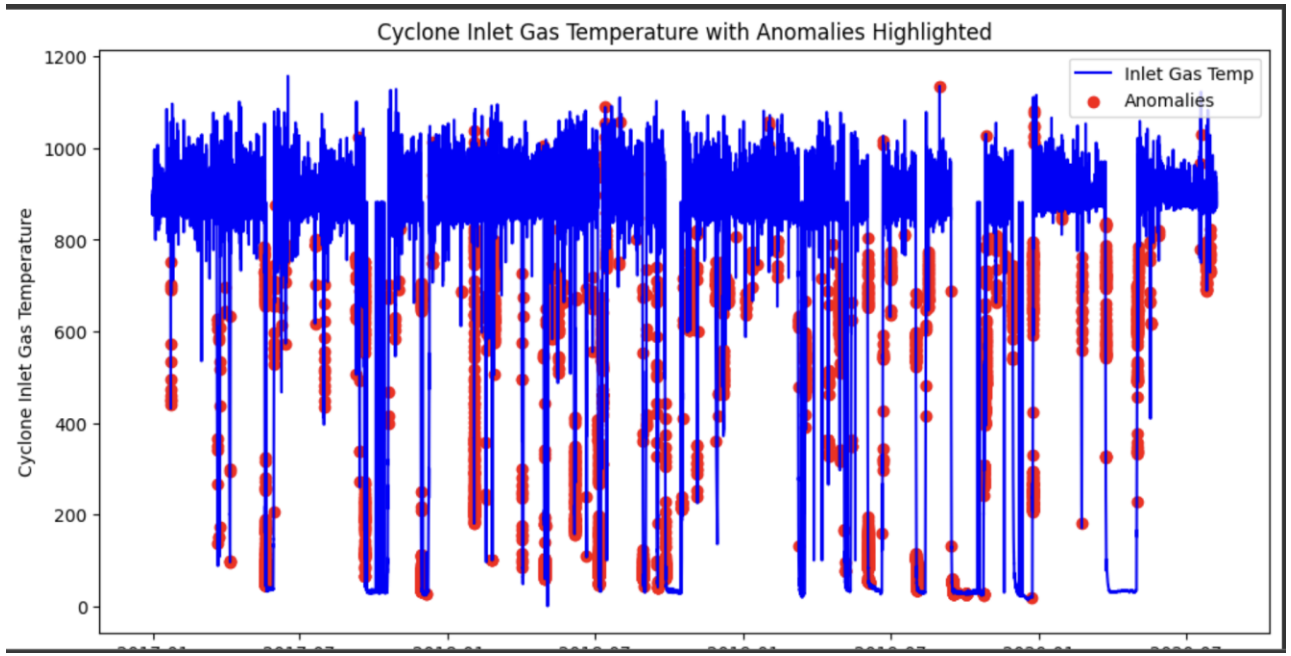
***Statistical Approach*:** Using Z-scores calculated earlier, you can set a threshold (e.g., |Z| > 3) to flag significant anomalies in the data. Highlight these points in the time series plots to make it easier to spot when and where anomalies occur.

Isolation Forest For more complex anomaly detection, try models like *Autoencoders*, which are effective for identifying outliers in multivariate data.

Insight Extraction

1. identify time periods when variables deviated significantly from normal patterns Note down possible causes if identifiable from the data.
2. Report insights like correlations between temperatures and drafts or any systematic changes over time.


Cyclone Inlet Gas Temperature with Anomalies
Inlet Gas Temp
Anomalies
Inlet Gas Temperature
1200
1000
800
600
400
200
0
2017-01
2017-07
2018-01
2018-07
2019-01
2019-07
2020-01
2020-07
Time


Cyclone Inlet Gas Temperature with Anomalies Highlighted
Inlet Gas Temp
Anomalies
Cyclone Inlet Gas Temperature
1200
1000
800
600
400
200
0

Here's how you can implement this approach in Python:

python

```
import pandas as pd
from sklearn.ensemble import IsolationForest
import matplotlib.pyplot as plt

# Assuming 'data_filled' is the preprocessed DataFrame from previous steps

# 1. Feature Selection (Exclude the 'time' column for Isolation Forest)
features = data_filled[['Cyclone_Inlet_Gas_Temp', 'Cyclone_Material_Temp', 'Cyclone_Outlet_Gas_draft',
                        'Cyclone_cone_draft', 'Cyclone_Gas_Outlet_Temp', 'Cyclone_Inlet_Draft']]

# 2. Initialize Isolation Forest
isolation_forest = IsolationForest(contamination=0.01, random_state=42)
data_filled['anomaly_score'] = isolation_forest.fit_predict(features)

# 3. Identify anomalies (anomaly_score of -1 indicates an anomaly)
anomalies = data_filled[data_filled['anomaly_score'] == -1]

# 4. Plot time series with anomalies highlighted
plt.figure(figsize=(12, 6))
plt.plot(data_filled['time'], data_filled['Cyclone_Inlet_Gas_Temp'], label='Inlet Gas Temp', color='blue')

# Plot anomalies
plt.scatter(anomalies['time'], anomalies['Cyclone_Inlet_Gas_Temp'], color='red', label='Anomalies')

plt.xlabel('Time')
plt.ylabel('Cyclone Inlet Gas Temperature')
plt.title('Cyclone Inlet Gas Temperature with Anomalies Highlighted')
plt.legend()
plt.show()
```