**Batch: D- 1 Roll No.: 16010122096**

**Experiment No. 5**

|  |
| --- |
| **TITLE : To perform forecasting using time series analysis** |

**AIM:** To perform forecasting using time series analysis

**Expected OUTCOME of Experiment:**

CO4: Perform Time series Analytics and forecasting

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Books/ Journals/ Websites referred:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Pre Lab/ Prior Concepts:**

Students should have a basic understanding of: Time series Analytics and forecasting

**Procedure:**

**Data set Used: Temperature data**

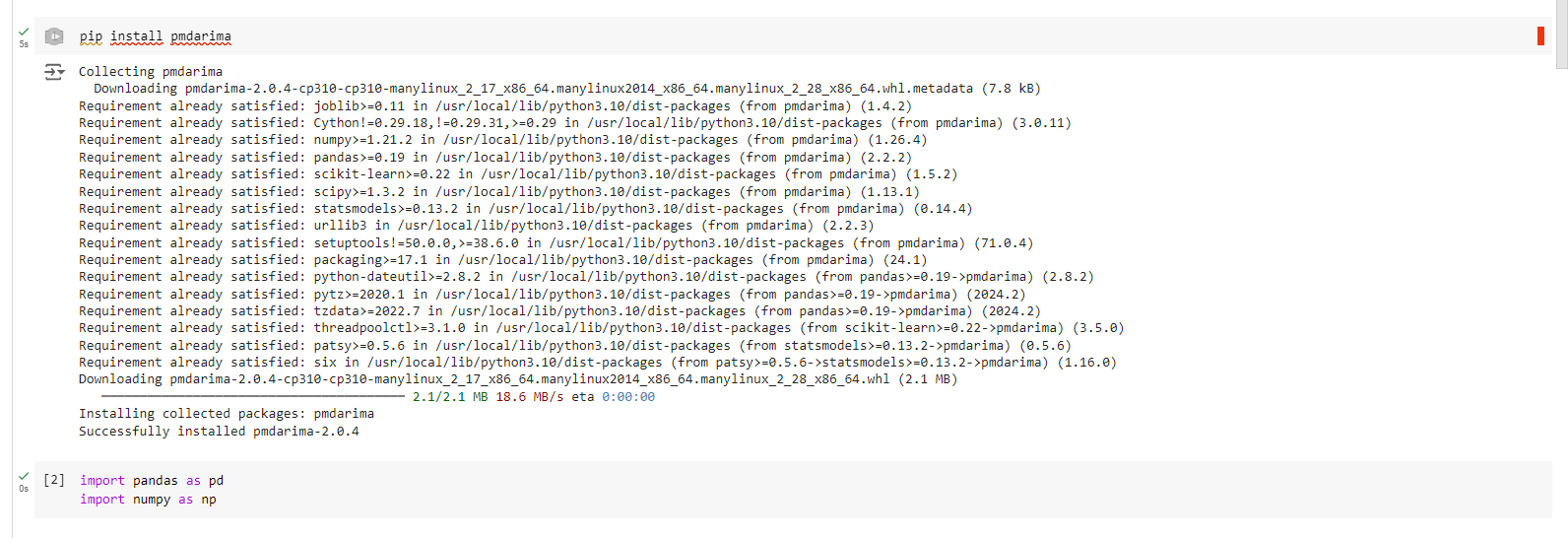
**Step1: Select and Load the dataset**

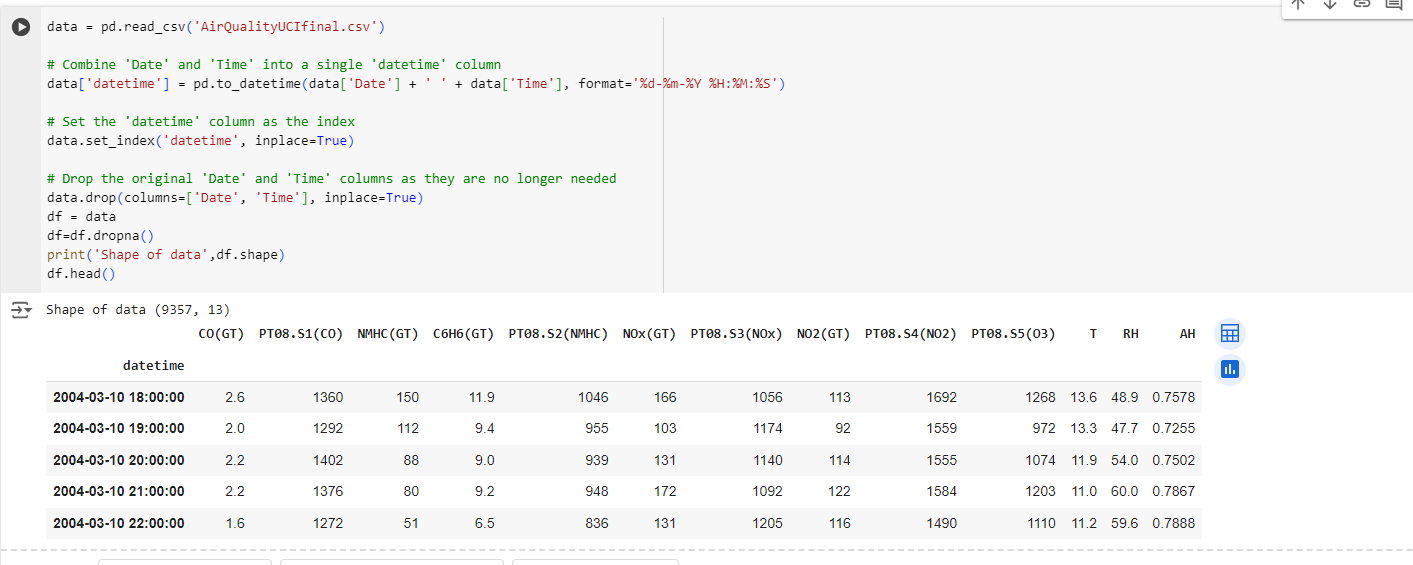
df=pd.read\_csv('AirQualityUCIfinal.csv',index\_col='Date',parse\_dates=True)

df=df.dropna()

print('Shape of data',df.shape)

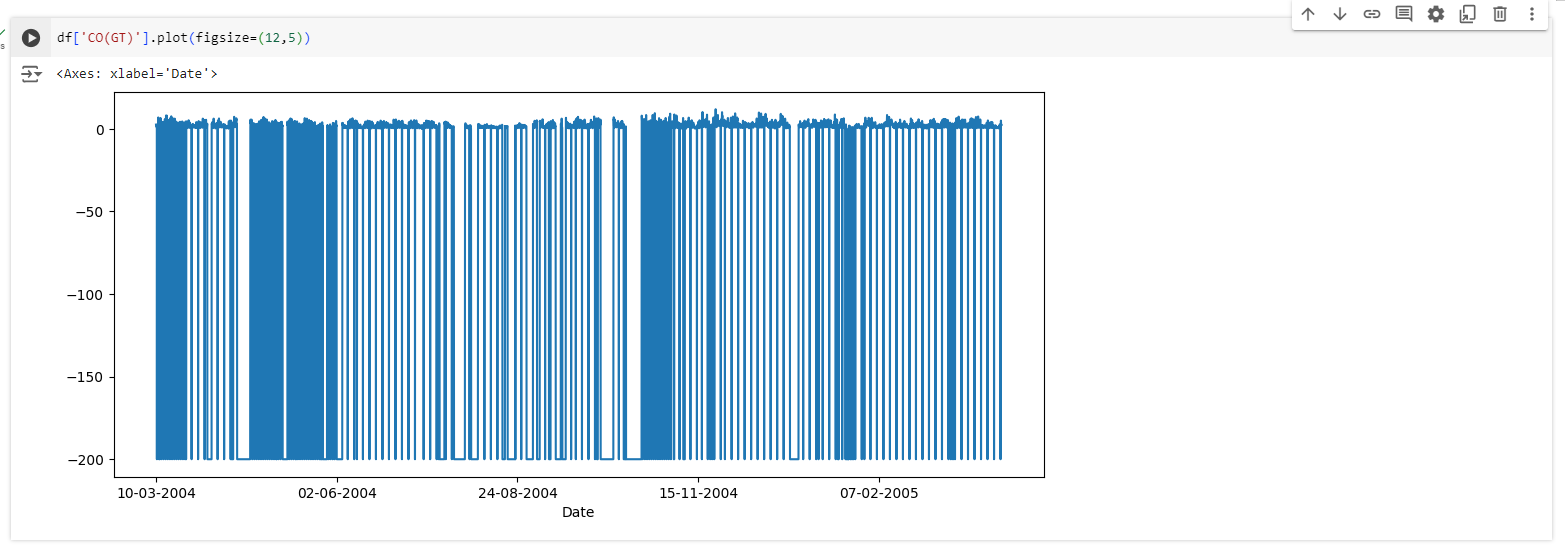
df.head()



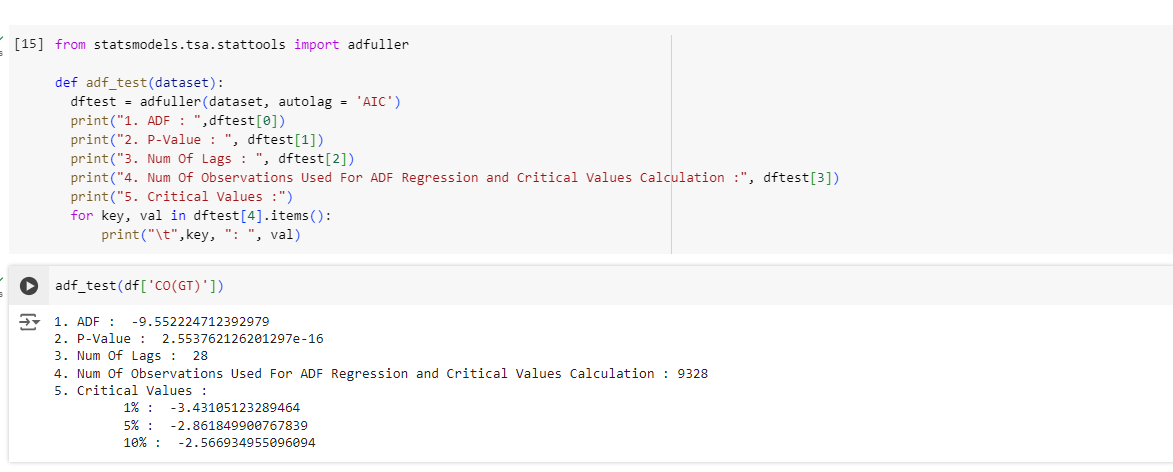


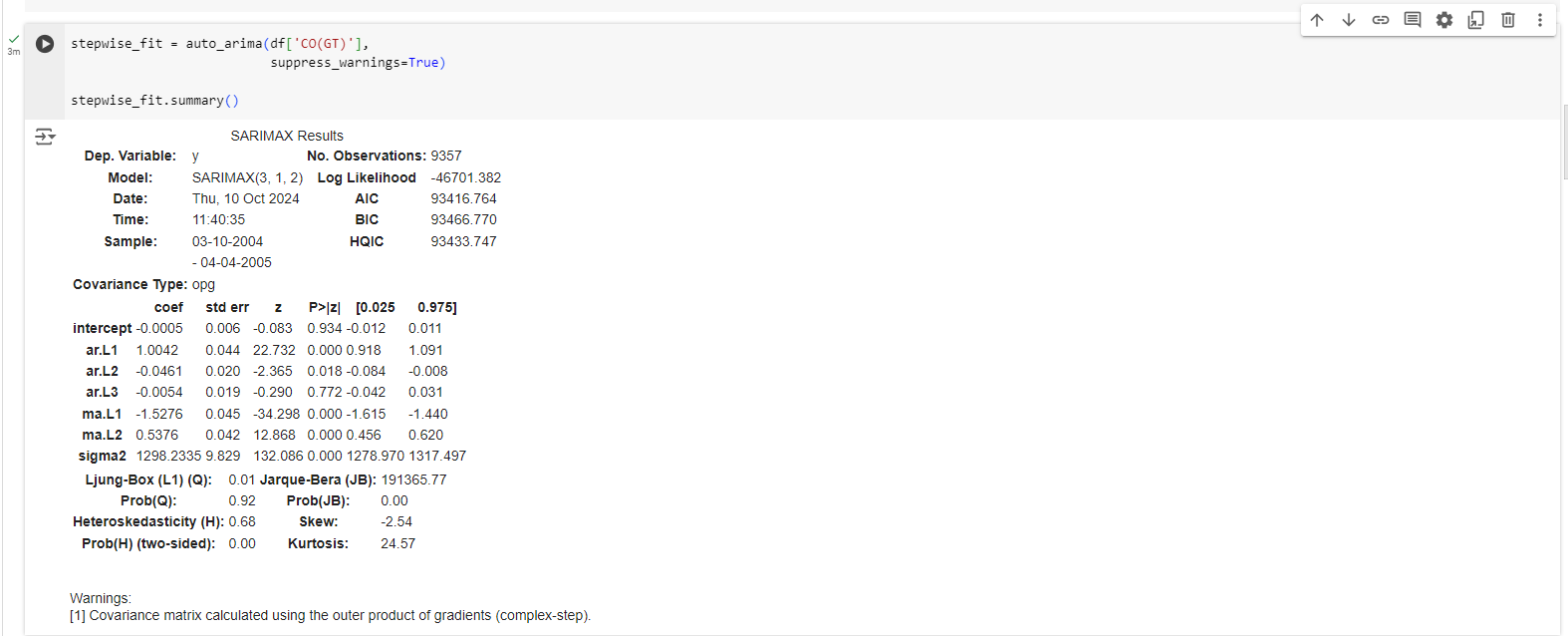
**Step2: Visualize the data**

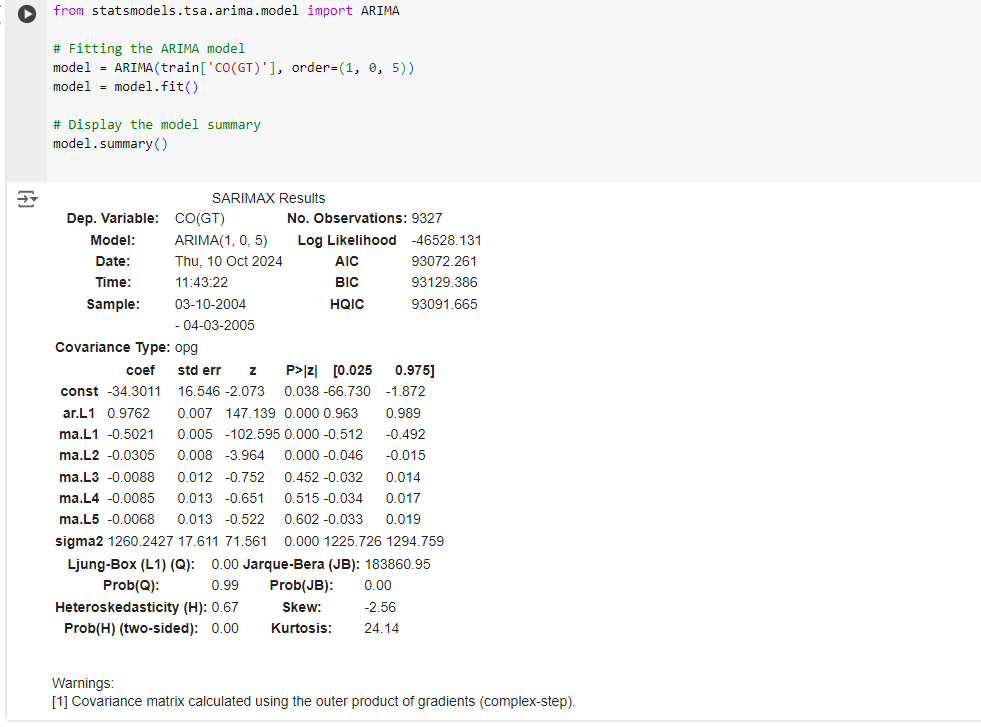
df['CO(GT)'].plot(figsize=(12,5))

****

**Step 3: Fit the model (ARIMA Model is Used)**

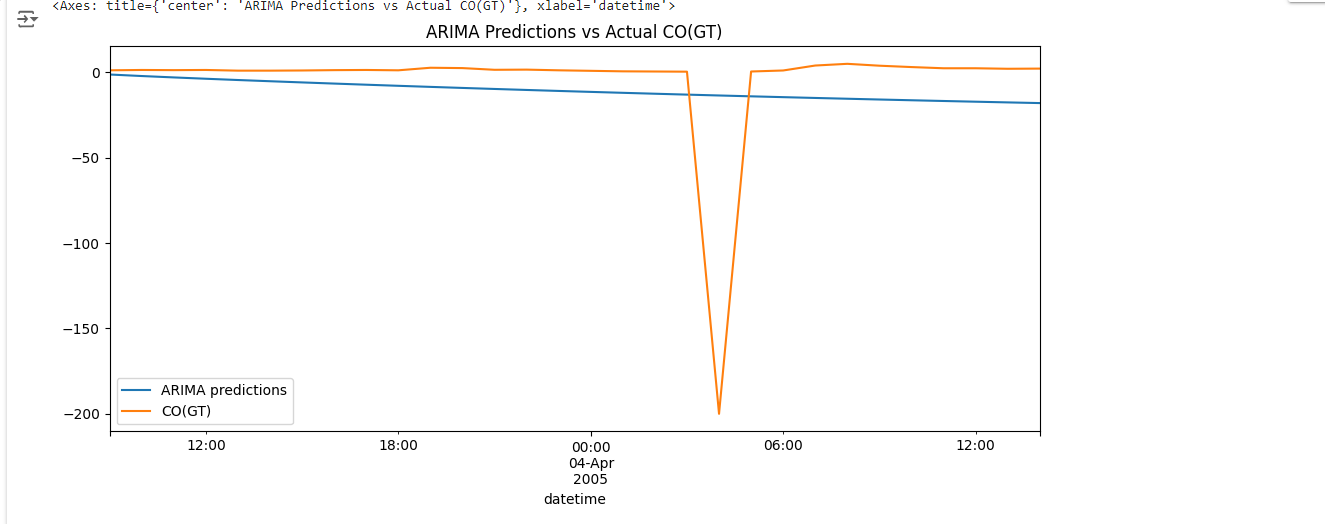
****

****

****

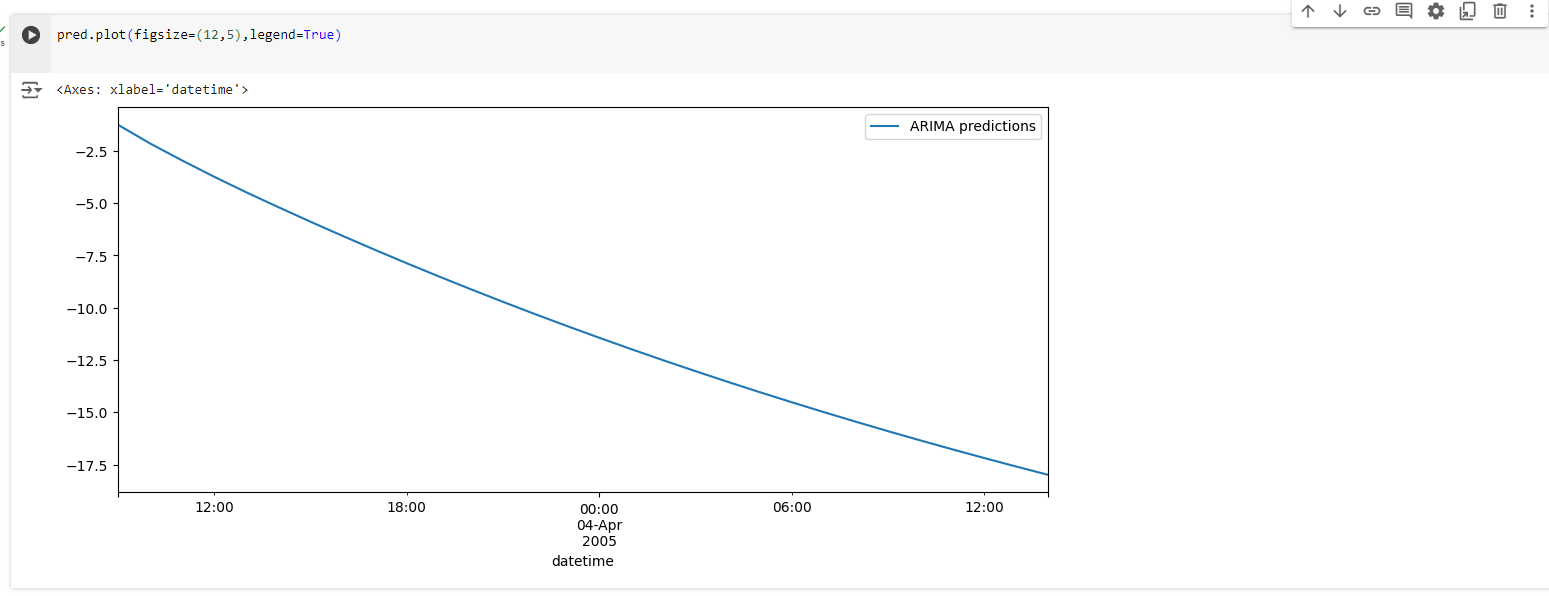
**Step4: Forecast future values**



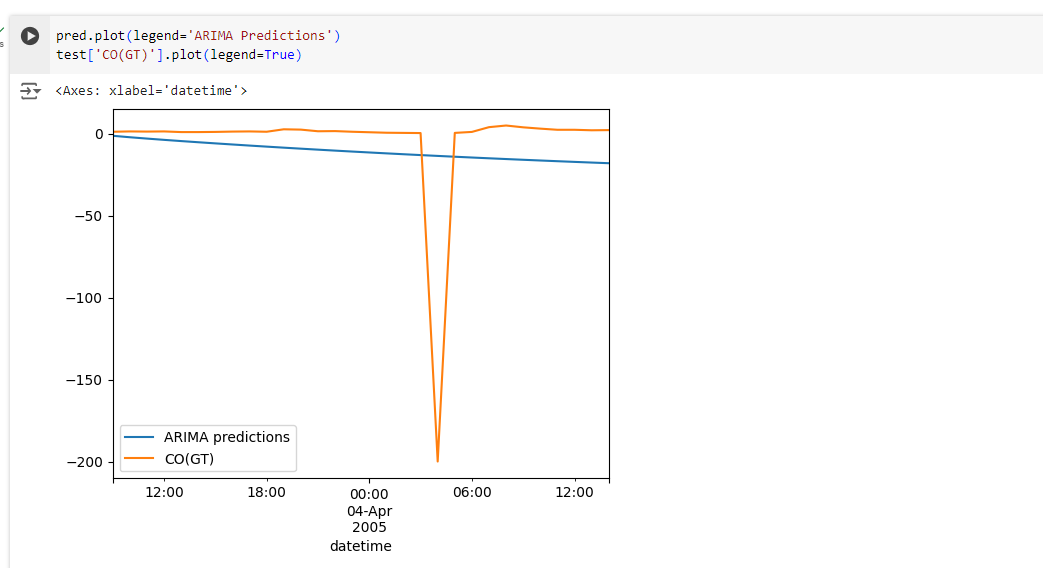


****

**Step 5: Create a DataFrame for the forecast**

****

**Step 6: Plot the results**



**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_ Signature of faculty in-charge**

**Post Lab Descriptive Questions:**

1. What are the key components of a time series, and how do they affect the analysis?

A time series typically consists of four key components:

* **Trend**: The long-term movement in the data. It shows the overall direction (upward, downward, or stable) over time.
* **Seasonality**: Regular patterns or fluctuations that occur at specific intervals, such as monthly, quarterly, or yearly. This component reflects periodic changes.
* **Cyclic**: Fluctuations that occur at irregular intervals due to economic or other cycles. Unlike seasonality, cyclic patterns are not fixed in frequency or duration.
* **Residual (or Irregular) Component**: The random noise or irregular variations in the data that cannot be attributed to trend, seasonality, or cyclic patterns.

These components affect analysis by helping to understand the underlying patterns and guiding forecasting models. By identifying and separating these components, analysts can make more accurate predictions and better understand the data's behavior.

1. What is the purpose of decomposing a time series into trend, seasonal, and residual components?

Decomposing a time series into its components serves several purposes:

* **Understanding Data Structure**: It helps in identifying the underlying patterns, making it easier to analyze and interpret the data.
* **Improving Forecasting**: By isolating the trend and seasonal effects, analysts can apply forecasting techniques more effectively, leading to improved accuracy.
* **Modeling**: Decomposing the series can assist in selecting appropriate models for analysis. For example, seasonal decomposition helps when choosing seasonal ARIMA models.
* **Anomaly Detection**: Identifying the residual component allows for the detection of outliers or unusual events that may require further investigation.

1. Explain how the ARIMA model works and what the terms (p, d, q) represent.

ARIMA stands for AutoRegressive Integrated Moving Average, and it is used for forecasting time series data.

* **p** (AutoRegressive order): This parameter indicates the number of lagged observations included in the model. It captures the influence of past values on the current value.
* **d** (Integrated order): This parameter represents the number of times the raw observations are differenced to make the time series stationary. A stationary series has constant mean and variance over time, which is crucial for effective modeling.
* **q** (Moving Average order): This parameter indicates the number of lagged forecast errors in the model. It captures the relationship between an observation and a residual error from a moving average model applied to lagged observations.