

| **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:**

Students have to list.

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

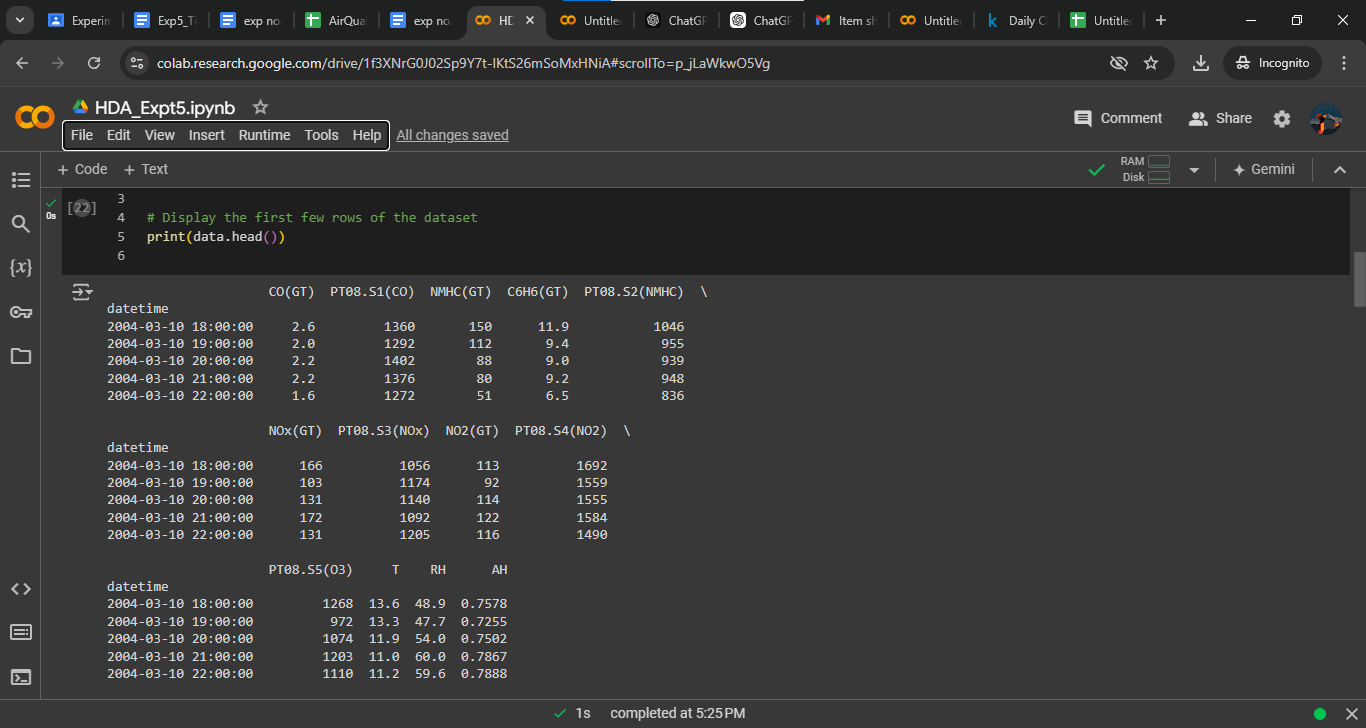
**Pre Lab/ Prior Concepts:**

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

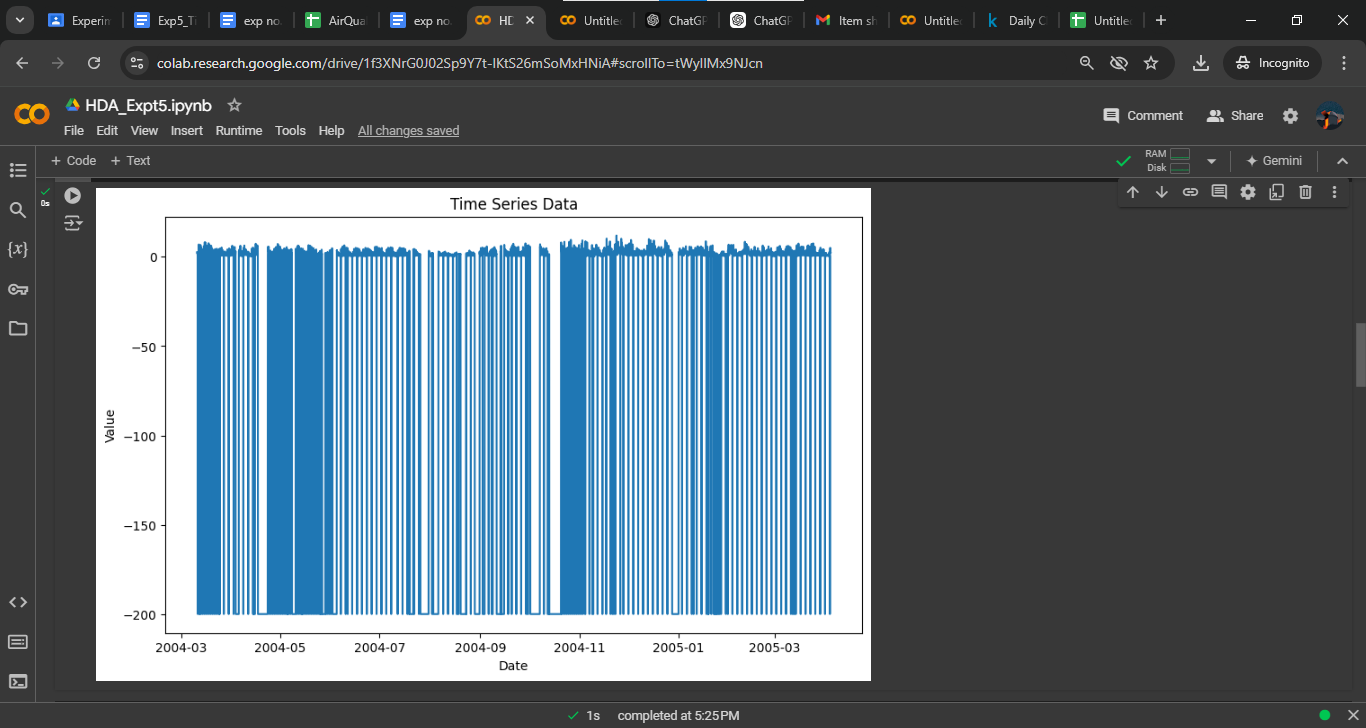
**Procedure:**

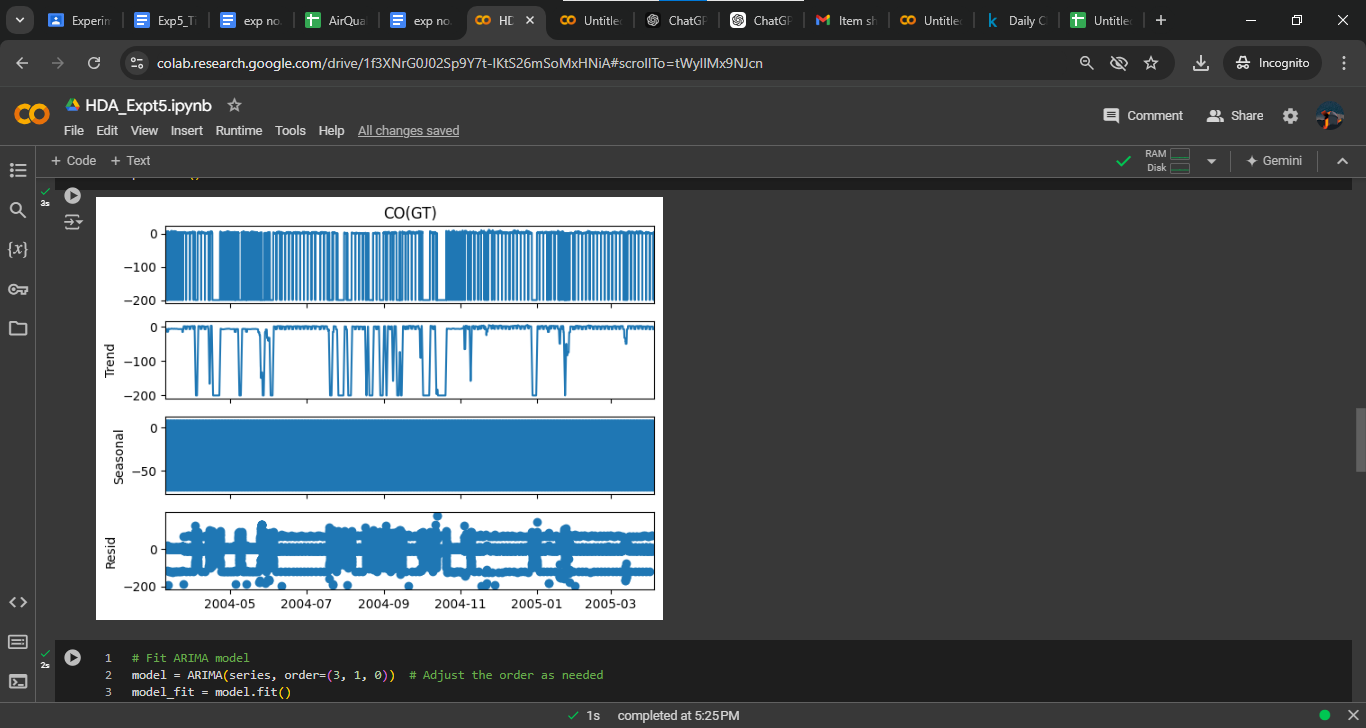
**Data set Used: Air Quality**

**Step1: Select and Load the dataset**

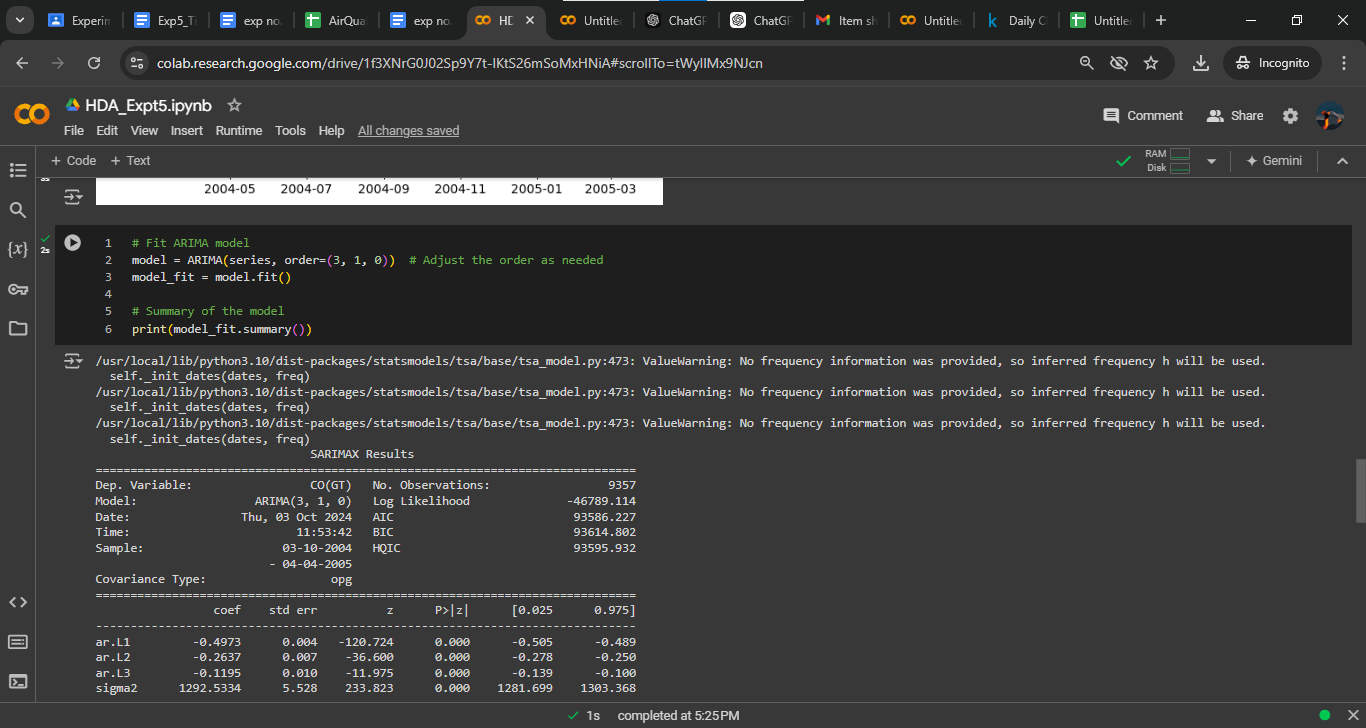


**Step2: Visualize the data**

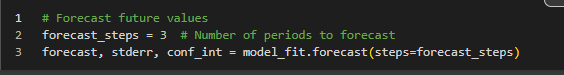




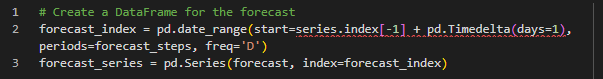
**Step 3: Fit the model (ARIMA Model is Used)**



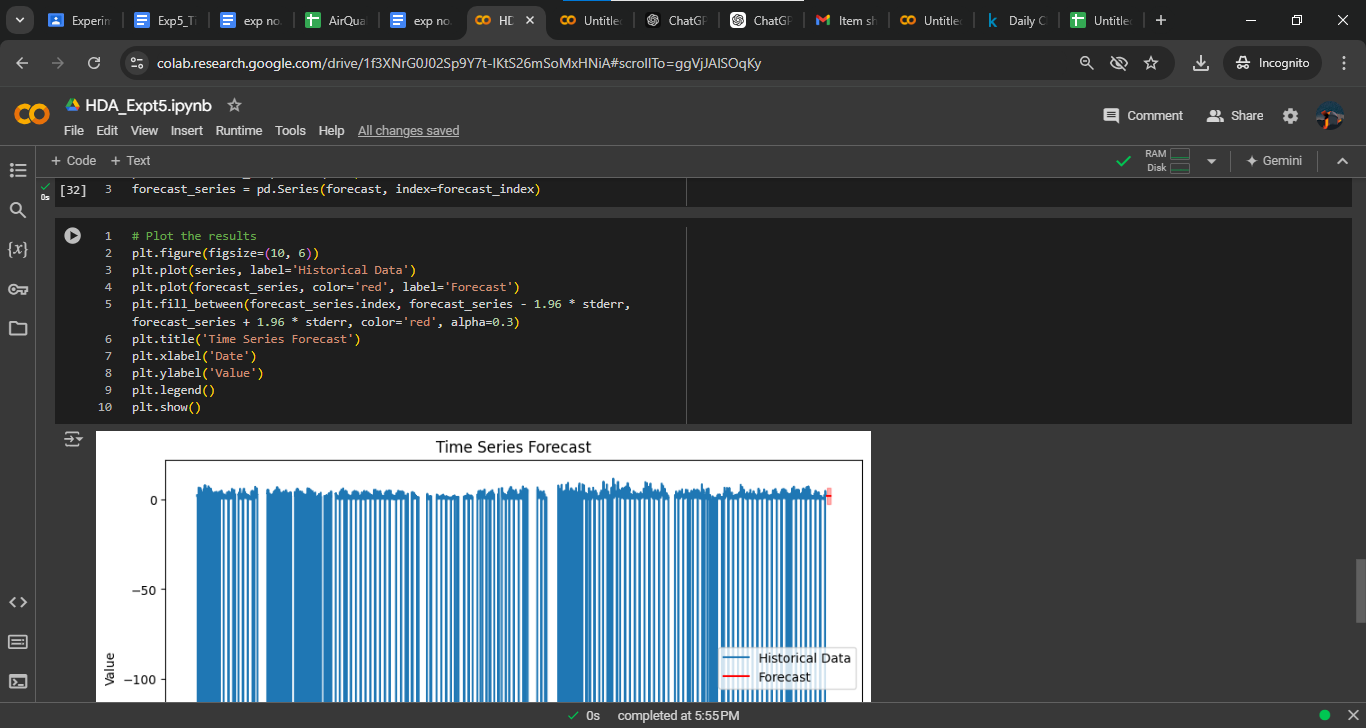
**Step4: Forecast future values**



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



**Step 6: Plot the results**

****

Implementation details:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from statsmodels.tsa.arima.model import ARIMA

from statsmodels.tsa.seasonal import seasonal\_decompose

# Load the data

import pandas as pd

# Load the dataset

data = pd.read\_csv('AirQualityUCIfinal.csv')

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

data['datetime'] = pd.to\_datetime(data['Date'] + ' ' + data['Time'], format='%d-%m-%Y %H:%M:%S')

# Set the 'datetime' column as the index

data.set\_index('datetime', inplace=True)

# Drop the original 'Date' and 'Time' columns as they are no longer needed

data.drop(columns=['Date', 'Time'], inplace=True)

# Optional: Convert columns to numeric if they are not already

data = data.apply(pd.to\_numeric, errors='coerce')

# Display the first few rows of the dataset

print(data.head())

series = data['CO(GT)'] # Adjust column names as needed

# Visualize the data

plt.figure(figsize=(10, 6))

plt.plot(series)

plt.title('Time Series Data')

plt.xlabel('Date')

plt.ylabel('Value')

plt.show()

# Decompose the data

decomposition = seasonal\_decompose(series, model='additive')

fig = decomposition.plot()

plt.show()

# Fit ARIMA model

model = ARIMA(series, order=(3, 1, 0)) # Adjust the order as needed

model\_fit = model.fit()

# Summary of the model

print(model\_fit.summary())

# Forecast future values

forecast\_steps = 3 # Number of periods to forecast

forecast, stderr, conf\_int = model\_fit.forecast(steps=forecast\_steps)

# Create a DataFrame for the forecast

forecast\_index = pd.date\_range(start=series.index[-1] + pd.Timedelta(days=1), periods=forecast\_steps, freq='D')

forecast\_series = pd.Series(forecast, index=forecast\_index)

# Plot the results

plt.figure(figsize=(10, 6))

plt.plot(series, label='Historical Data')

plt.plot(forecast\_series, color='red', label='Forecast')

plt.fill\_between(forecast\_series.index, forecast\_series - 1.96 \* stderr, forecast\_series + 1.96 \* stderr, color='red', alpha=0.3)

plt.title('Time Series Forecast')

plt.xlabel('Date')

plt.ylabel('Value')

plt.legend()

plt.show()

**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?
2. What is the purpose of decomposing a time series into trend, seasonal, and residual components?
3. Explain how the ARIMA model works and what the terms (p, d, q) represent.