Course Name:	Data Analysis Laboratory (216H03L501)	Semester:	V
Date of Performance:	_13__ / _10__ / __25____	DIV/ Batch No:	D-2
Student Name:	Shreyans Tatiya	Roll No:	16010123325

TITLE: Perform forecasting/predict using time series analysis (AR/MA/ARIMASARIMA)

AIM: To perform forecasting using time series analysis

Expected OUTCOME of Experiment:

CO4: Perform Time series Analytics and forecasting

Books/ Journals/ Websites referred:

https://www.datacamp.com/tutorial/arima

Pre Lab/ Prior Concepts:

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

Procedure:

Data set Used: AirQuality dataset

Note: google colab is shared please refer

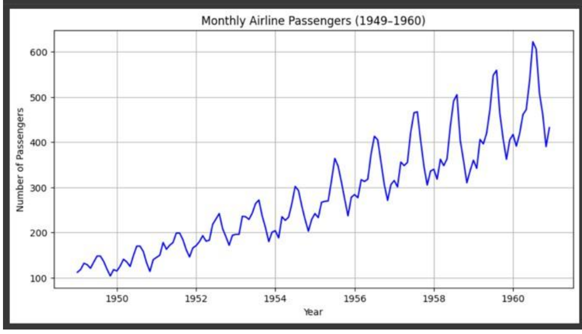
Step1: Select and Load the dataset

```
# Step 1: Select and Load the Dataset

import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm

# Load dataset from statsmodels
data = sm.datasets.get_rdataset("AirPassengers").data

# Display first few rows
print(data.head())
```

```
          time  value
0  1949.000000    112
1  1949.083333    118
2  1949.166667    132
3  1949.250000    129
4  1949.333333    121
```

Step2: Visualize the data

```
# Step 2: Visualize the data

plt.figure(figsize=(10,5))
plt.plot(data['time'], data['value'], color='blue')
plt.title('Monthly Airline Passengers (1949–1960)')
plt.xlabel('Year')
plt.ylabel('Number of Passengers')
plt.grid(True)
plt.show()
```



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

```python
# Step 3: Fit ARIMA model

from statsmodels.tsa.arima.model import ARIMA

# Convert to time series (index as datetime)
data.index = pd.date_range(start='1949-01', periods=len(data), freq='M')
ts = data['value']

# Fit ARIMA model (p,d,q) = (1,1,1)
model = ARIMA(ts, order=(1,1,1))
model_fit = model.fit()

# Summary of the model
print(model_fit.summary())
```

```
/tmp/ipython-input-1352181933.py:6: FutureWarning: 'M' is deprecated and will be removed in a future version, please use 'ME' instead.
  data.index = pd.date_range(start='1949-01', periods=len(data), freq='M')
                               SARIMAX Results
==============================================================================
Dep. Variable:                  value   No. Observations:                  144
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -694.341
Date:                Mon, 13 Oct 2025   AIC                           1394.683
Time:                        15:34:30   BIC                           1403.571
Sample:                    01-31-1949   HQIC                          1398.294
                         - 12-31-1960
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4742      0.123     -3.847      0.000      -0.716      -0.233
ma.L1          0.8635      0.078     11.051      0.000       0.710       1.017
sigma2       961.9270    107.433      8.954      0.000     751.362    1172.492
===================================================================================
Ljung-Box (L1) (Q):                   0.21   Jarque-Bera (JB):                 2.14
Prob(Q):                              0.65   Prob(JB):                         0.34
Heteroskedasticity (H):               7.00   Skew:                            -0.21
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.43
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Step4: Forecast future values

```python
# Step 4: Forecast next 12 months

forecast = model_fit.forecast(steps=12)
print(forecast)
```

```
1961-01-31    475.735059
1961-02-28    454.996073
1961-03-31    464.830415
1961-04-30    460.167010
1961-05-31    462.378378
1961-06-30    461.329756
1961-07-31    461.827008
1961-08-31    461.591213
1961-09-30    461.703026
1961-10-31    461.650005
1961-11-30    461.675148
1961-12-31    461.663225
Freq: ME, Name: predicted_mean, dtype: float64
```

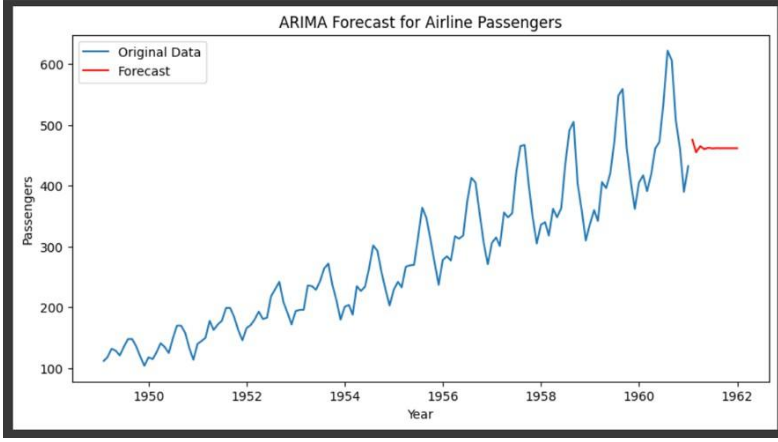
Step 5: Create a DataFrame for the forecast

```python
# Step 5: Create a DataFrame for the forecast

forecast_df = pd.DataFrame({
    'Date': pd.date_range(start=ts.index[-1] + pd.DateOffset(months=1), periods=12, freq='M'),
    'Forecasted_Passengers': forecast
})

print(forecast_df)
```

```
                 Date  Forecasted_Passengers
1961-01-31 1961-01-31             475.735059
1961-02-28 1961-02-28             454.996073
1961-03-31 1961-03-31             464.830415
1961-04-30 1961-04-30             460.167010
1961-05-31 1961-05-31             462.378378
1961-06-30 1961-06-30             461.329756
1961-07-31 1961-07-31             461.827008
1961-08-31 1961-08-31             461.591213
1961-09-30 1961-09-30             461.703026
1961-10-31 1961-10-31             461.650005
1961-11-30 1961-11-30             461.675148
1961-12-31 1961-12-31             461.663225
/tmp/ipython-input-277842258.py:4: FutureWarning: 'M' is deprecated and will be removed in a future version, please use 'ME' instead.
  'Date': pd.date_range(start=ts.index[-1] + pd.DateOffset(months=1), periods=12, freq='M'),
```

Step 6: Plot the results

```
# Step 6: Plot the results

plt.figure(figsize=(10,5))
plt.plot(ts, label='Original Data')
plt.plot(forecast_df['Date'], forecast_df['Forecasted_Passengers'], label='Forecast', color='red')
plt.title('ARIMA Forecast for Airline Passengers')
plt.xlabel('Year')
plt.ylabel('Passengers')
plt.legend()
plt.show()
```



Implementation details:
https://colab.research.google.com/drive/1PhnXRZsQS6KEYjLRnpaOgOdOJeE_uVho?usp=sharing

Date: _____________ **Signature of faculty in-charge**

<u>Post Lab Descriptive Questions:</u>

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

   A time series has **three main components**:

   - **Trend:** The long-term movement or direction in the data (e.g., upward sales growth).
   - **Seasonality:** Regular, repeating patterns over fixed periods (e.g., higher sales in December).

- **Residual (Noise):** Random fluctuations that cannot be explained by trend or seasonality.
  Understanding these helps choose the right forecasting model and improve accuracy.

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

   Decomposition helps **separate the data into trend, seasonal, and residual parts** to better understand underlying patterns.
   It allows analysts to **study each component individually**, identify seasonality strength, and make **more accurate forecasts** by modeling trend and seasonality separately.

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

   The **ARIMA (AutoRegressive Integrated Moving Average)** model combines autoregression, differencing, and moving averages to make forecasts.

   - **p:** Number of lag observations (AutoRegressive part).
   - **d:** Number of times the data is differenced to make it stationary.
   - **q:** Size of the moving average window (how past errors affect current prediction).

   Together, these parameters help model both trend and temporal dependencies in time series data.