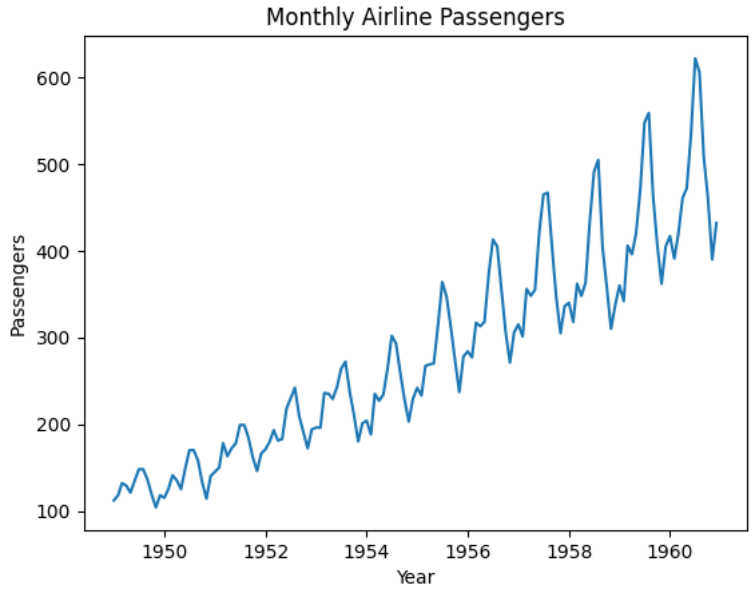
AIM: Implementation of ARIMA Model.

Tool: Google Collab

Roll No. 412039

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA


df = pd.read_csv('/content/AirPassengers.csv', index_col='Month', parse_dates=True)


train_size = int(len(df) * 0.8)
train_data, test_data = df[:train_size], df[train_size:]


plt.plot(df)
plt.xlabel('Year')
plt.ylabel('Passengers')
plt.title('Monthly Airline Passengers')
plt.show()
```




```
train_size = int(len(df) * 0.8)
train_data, test_data = df[:train_size], df[train_size:]


model = ARIMA(train_data, order=(2, 1, 2))
model_fit = model.fit()
print(model_fit.summary())
```

```
/usr/local/lib/python3.9/dist-packages/statsmodels/tsa/base/tsa_model.py:471: ValueWarning: No frequency information was provided,
  self._init_dates(dates, freq)
/usr/local/lib/python3.9/dist-packages/statsmodels/tsa/base/tsa_model.py:471: ValueWarning: No frequency information was provided,
  self._init_dates(dates, freq)
/usr/local/lib/python3.9/dist-packages/statsmodels/tsa/base/tsa_model.py:471: ValueWarning: No frequency information was provided,
  self._init_dates(dates, freq)
                               SARIMAX Results
==============================================================================
Dep. Variable:            #Passengers   No. Observations:                  115
Model:                 ARIMA(2, 1, 2)   Log Likelihood                -523.758
Date:                Sun, 02 Apr 2023   AIC                           1057.516
Time:                        17:14:03   BIC                           1071.197
Sample:                    01-01-1949   HQIC                          1063.069
                         - 07-01-1958
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.3280      0.145      2.268      0.023       0.045       0.611
ar.L2          0.2521      0.165      1.528      0.126      -0.071       0.575
ma.L1         -0.0125      0.109     -0.114      0.909      -0.227       0.202
ma.L2         -0.7544      0.130     -5.812      0.000      -1.009      -0.500
sigma2       568.4920    103.877      5.473      0.000     364.897     772.087
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 3.39
```

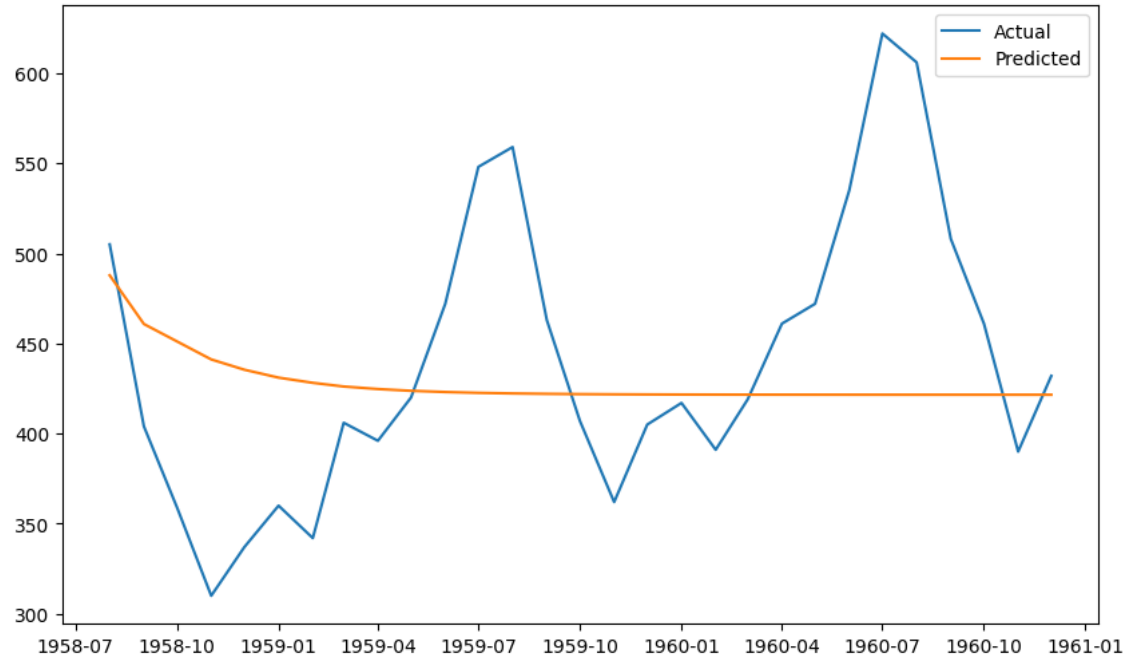
```
Prob(Q):                              0.90   Prob(JB):                         0.18
Heteroskedasticity (H):               5.24   Skew:                             0.11
Prob(H) (two-sided):                  0.00   Kurtosis:                         2.19
================================================================================

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

we are using an ARIMA model with p=2, d=1, and q=2. The order parameter is used to set the values of p, d, and q. The fit() method is used to fit the ARIMA model to the training data. The summary() method is used to print a summary of the model.

Use the ARIMA model to make predictions on the testing data:

```
predictions = model_fit.forecast(len(test_data))
```

```
fig, ax = plt.subplots(figsize=(10, 6))
plt.plot(test_data, label='Actual')
plt.plot(predictions, label='Predicted')
plt.legend()
plt.show()
```




```
from sklearn.metrics import mean_squared_error
# Calculate MSE
mse = mean_squared_error(test_data, predictions)

# Calculate RMSE
rmse = np.sqrt(mse)

print('MSE:', mse)
print('RMSE:', rmse)
```

```
MSE: 6808.3970474928465
RMSE: 82.51301138301066
```

Conclusion: Hence we have implemented the Arima Model for the AirPassengers Dataset.

✓ 1s completed at 10:51 PM