Surakshith Shetty -53026240013

Practical 2 : To analyze and forecast trends in time series data using statistical and machine learning techniques, identifying patterns, seasonality, and anomalies to derive actionable insights

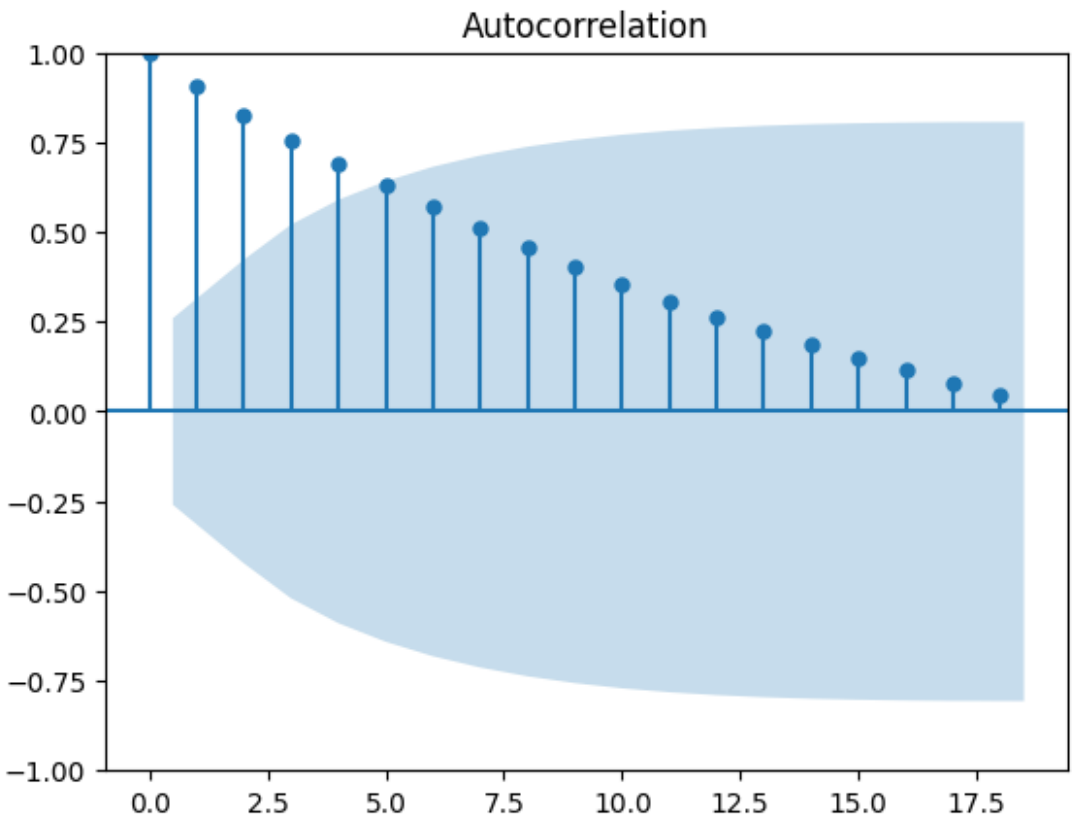
```
import pandas as pd

data = pd.read_csv('/content/GDP.csv')

data
```

```
{"summary":"{\n  \"name\": \"data\",\n  \"rows\": 57,\n  \"fields\":
[\n    {\n      \"column\": \"Year\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 57,\n
\"samples\": [\n          \"1950-51\",\n          \"1955-56\",\n
\"1980-81\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"GDP\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 670848,\n        \"min\": 224786,\n        \"max\": 2848157,\
n        \"num_unique_values\": 57,\n        \"samples\": [\n
224786,\n          268316,\n          641921\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"data"}
```

```
rng = pd.date_range('1950','2007',freq='Y')
s = data.GDP.values
gdpseries = pd.Series(s,rng)
```

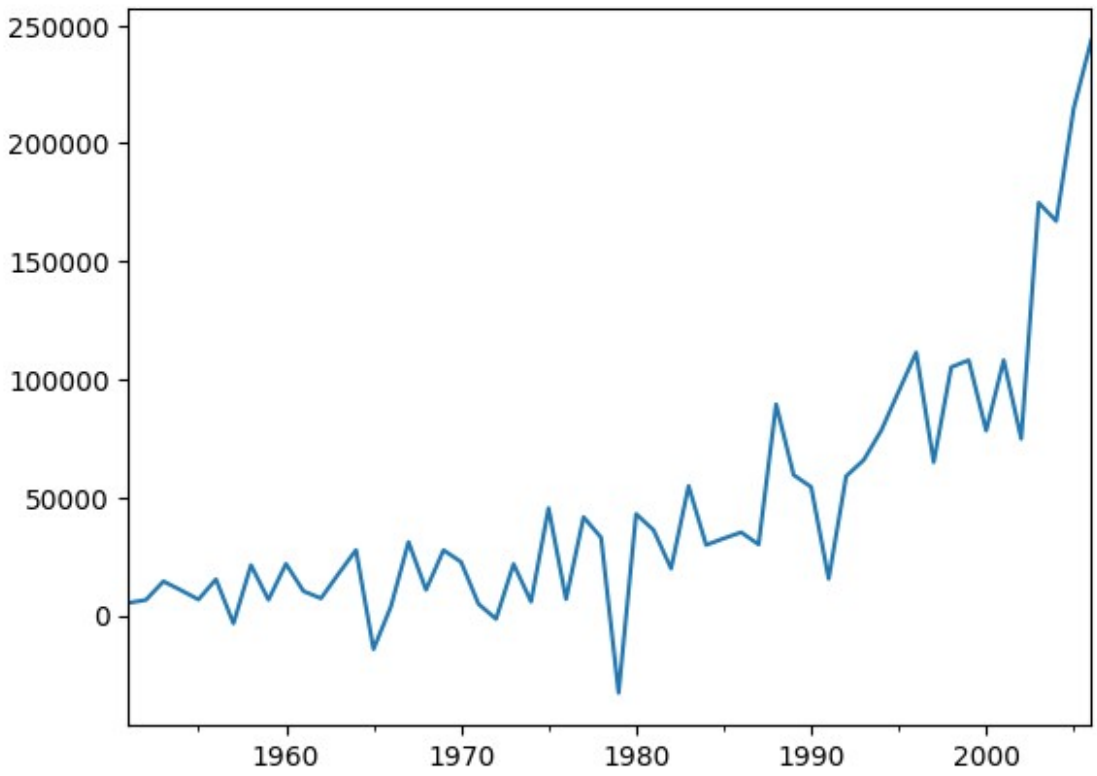
```
<ipython-input-5-4146c9cfbf33>:1: FutureWarning: 'Y' is deprecated and
will be removed in a future version, please use 'YE' instead.
  rng = pd.date_range('1950','2007',freq='Y')
```

```
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
plot_acf(gdpseries)
plt.show()
```

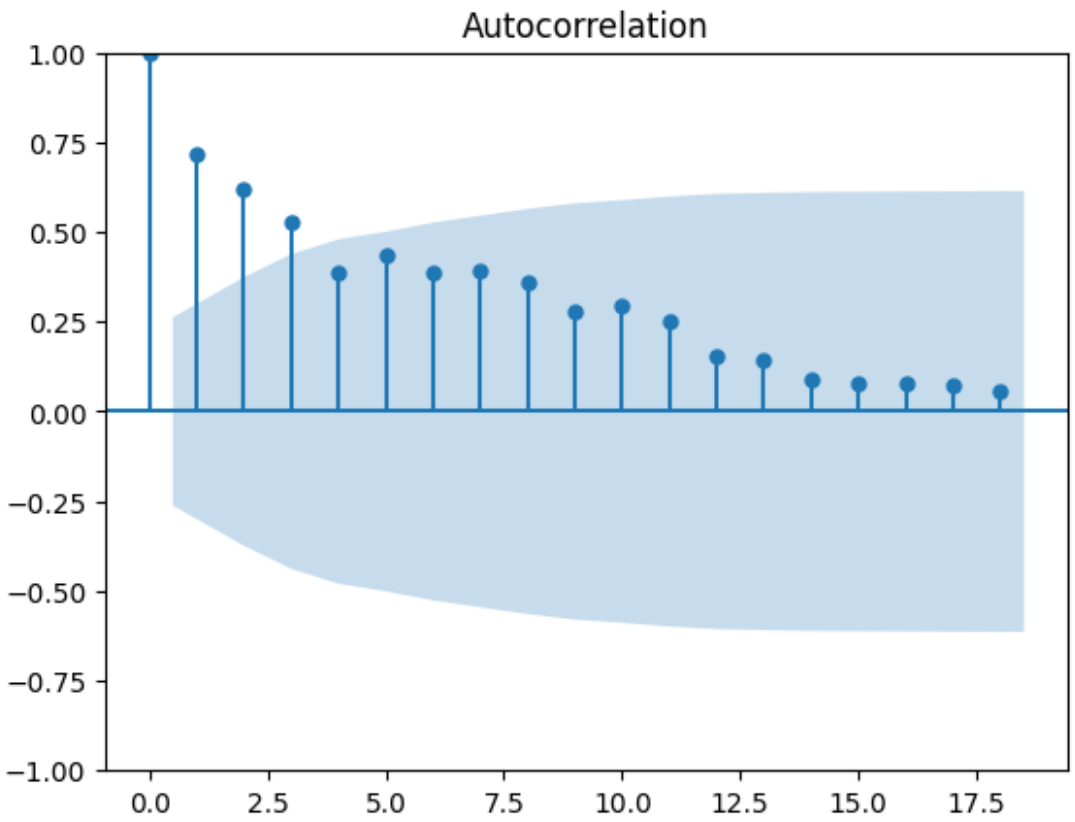



```
from statsmodels.tsa.statespace.tools import diff
gdpdiff = diff(gdpseries)
gdpdiff.plot()
gdpdiff.head()
```

```
1951-12-31     5248.0
1952-12-31     6528.0
1953-12-31    14398.0
1954-12-31    10655.0
1955-12-31     6701.0
Freq: YE-DEC, dtype: float64
```

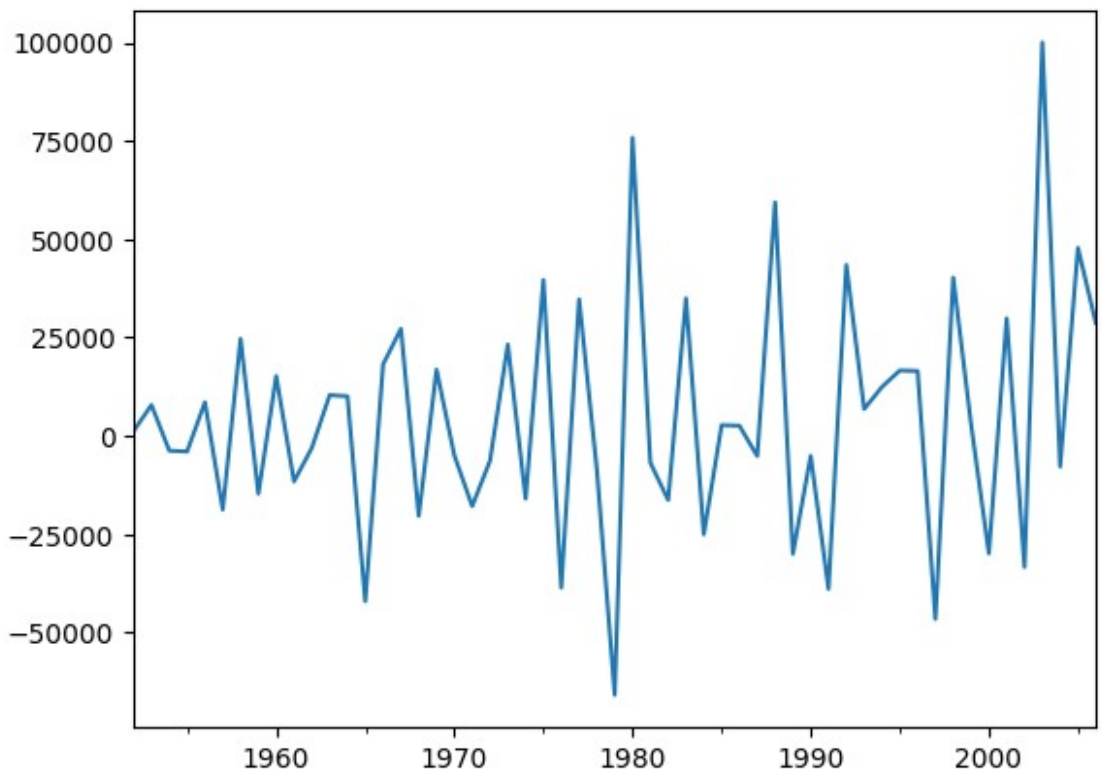



```
plot_acf(gdpdiff)
plt.show()
```

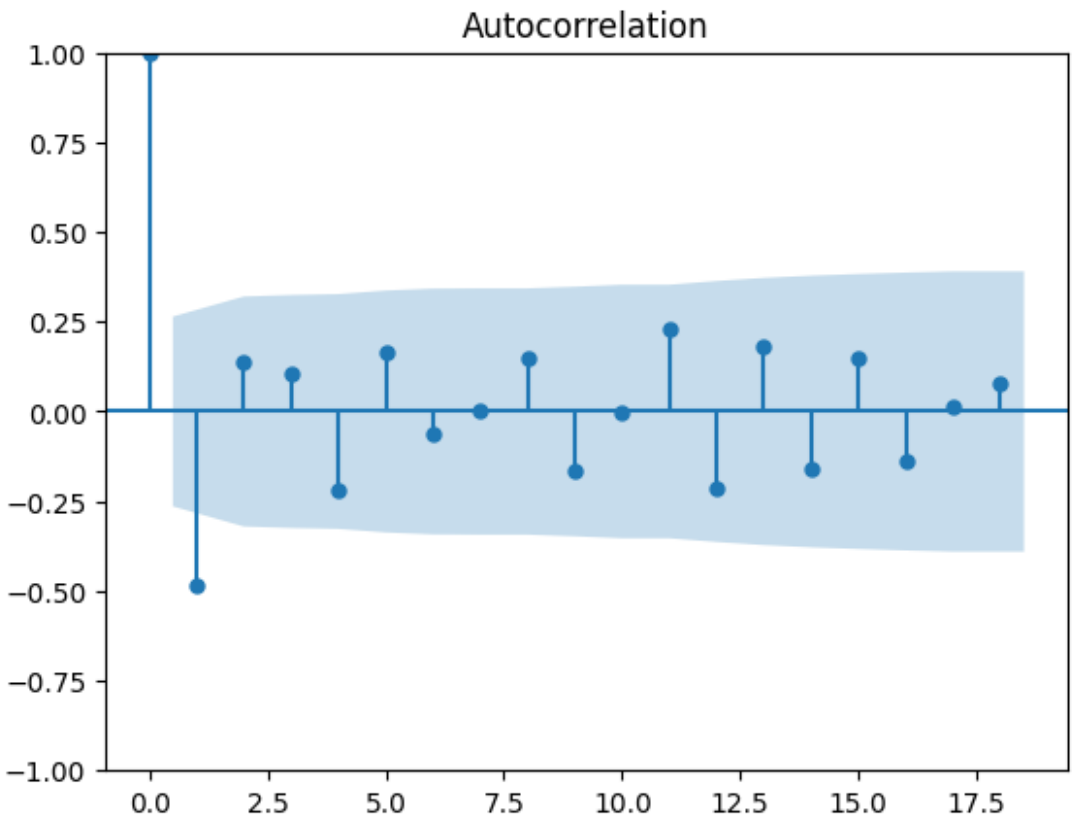



```
from statsmodels.tsa.statespace.tools import diff
gdpdiff1 = diff(gdpdiff)
gdpdiff1.plot()
gdpdiff1.head()
```

```
1952-12-31    1280.0
1953-12-31    7870.0
1954-12-31   -3743.0
1955-12-31   -3954.0
1956-12-31    8572.0
Freq: YE-DEC, dtype: float64
```




```
plot_acf(gdpdiff1)
plt.show()
```




```
!pip install arch

Collecting arch
  Downloading arch-7.2.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (13 kB)
Requirement already satisfied: numpy>=1.22.3 in
/usr/local/lib/python3.10/dist-packages (from arch) (1.26.4)
Requirement already satisfied: scipy>=1.8 in
/usr/local/lib/python3.10/dist-packages (from arch) (1.13.1)
Requirement already satisfied: pandas>=1.4 in
/usr/local/lib/python3.10/dist-packages (from arch) (2.2.2)
Requirement already satisfied: statsmodels>=0.12 in
/usr/local/lib/python3.10/dist-packages (from arch) (0.14.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=1.4->arch)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas>=1.4->arch)
(2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas>=1.4->arch)
(2024.2)
Requirement already satisfied: patsy>=0.5.6 in
/usr/local/lib/python3.10/dist-packages (from statsmodels>=0.12->arch)
```

```
(1.0.1)
Requirement already satisfied: packaging>=21.3 in
/usr/local/lib/python3.10/dist-packages (from statsmodels>=0.12->arch)
(24.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2-
>pandas>=1.4->arch) (1.17.0)
Downloading arch-7.2.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (985 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 985.1/985.1 kB 11.5 MB/s eta
0:00:00
```

```
from arch.unitroot import ADF
adf = ADF(gdpseries,lags=0)
adf.summary()
```

```
<class 'statsmodels.iolib.summary.Summary'>
"""
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                 15.780
P-value                         1.000
Lags                                0
-------------------------------------

Trend: Constant
Critical Values: -3.55 (1%), -2.91 (5%), -2.60 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
"""
```

```
from arch.unitroot import ADF
adf = ADF(gdpdiff,lags=0)
adf.summary()
```

```
<class 'statsmodels.iolib.summary.Summary'>
"""
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                 -0.599
P-value                         0.871
Lags                                0
-------------------------------------

Trend: Constant
Critical Values: -3.56 (1%), -2.92 (5%), -2.60 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
"""
```

```
from arch.unitroot import ADF
adf = ADF(gdpdiff1,lags=0)
adf.summary()
```

```
<class 'statsmodels.iolib.summary.Summary'>
"""
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                -12.227
P-value                         0.000
Lags                                0
-------------------------------------

Trend: Constant
Critical Values: -3.56 (1%), -2.92 (5%), -2.60 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
"""
```

```
!pip install pmdarima
```

```
Collecting pmdarima
  Downloading pmdarima-2.0.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl.m
etadata (7.8 kB)
Requirement already satisfied: joblib>=0.11 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (1.4.2)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (3.0.11)
Requirement already satisfied: numpy>=1.21.2 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (1.26.4)
Requirement already satisfied: pandas>=0.19 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (2.2.2)
Requirement already satisfied: scikit-learn>=0.22 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (1.6.0)
Requirement already satisfied: scipy>=1.3.2 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (1.13.1)
Requirement already satisfied: statsmodels>=0.13.2 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (0.14.4)
Requirement already satisfied: urllib3 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (2.2.3)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (75.1.0)
Requirement already satisfied: packaging>=17.1 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (24.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.19->pmdarima)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.19->pmdarima)
```

```
(2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.19->pmdarima)
(2024.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn>=0.22-
>pmdarima) (3.5.0)
Requirement already satisfied: patsy>=0.5.6 in
/usr/local/lib/python3.10/dist-packages (from statsmodels>=0.13.2-
>pmdarima) (1.0.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2-
>pandas>=0.19->pmdarima) (1.17.0)
Downloading pmdarima-2.0.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl
(2.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.1/2.1 MB 22.0 MB/s eta
0:00:00
darima
Successfully installed pmdarima-2.0.4
```

```python
import pmdarima
pmdarima.arima.ndiffs(gdpseries)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/
deprecation.py:151: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
```

```
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
```

```
2
```

```
!pip install statsmodels.tsa.stattools
```

```
ERROR: Could not find a version that satisfies the requirement
statsmodels.tsa.stattools (from versions: none)
ERROR: No matching distribution found for statsmodels.tsa.stattools
```

```python
import pmdarima as pm
model = pm.auto_arima(gdpseries,max_p=2, max_q=2, d=2,
seasonal=False,trace=True)
model
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/
deprecation.py:151: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
```

```
Performing stepwise search to minimize aic
 ARIMA(2,2,2)(0,0,0)[0] intercept   : AIC=1294.586, Time=0.10 sec
```

```
 ARIMA(0,2,0)(0,0,0)[0] intercept   : AIC=1295.387, Time=0.03 sec
 ARIMA(1,2,0)(0,0,0)[0] intercept   : AIC=1291.568, Time=0.05 sec
 ARIMA(0,2,1)(0,0,0)[0] intercept   : AIC=1292.152, Time=0.04 sec
 ARIMA(0,2,0)(0,0,0)[0]             : AIC=1294.532, Time=0.02 sec
 ARIMA(2,2,0)(0,0,0)[0] intercept   : AIC=1291.358, Time=0.04 sec
 ARIMA(2,2,1)(0,0,0)[0] intercept   : AIC=1292.717, Time=0.08 sec

/usr/local/lib/python3.10/dist-packages/sklearn/utils/
deprecation.py:151: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(

 ARIMA(1,2,1)(0,0,0)[0] intercept   : AIC=1293.092, Time=0.06 sec
 ARIMA(2,2,0)(0,0,0)[0]             : AIC=1290.274, Time=0.06 sec
 ARIMA(1,2,0)(0,0,0)[0]             : AIC=1290.280, Time=0.03 sec

/usr/local/lib/python3.10/dist-packages/sklearn/utils/
deprecation.py:151: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/utils/deprecation.py:1
51: FutureWarning: 'force_all_finite' was renamed to
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(

 ARIMA(2,2,1)(0,0,0)[0]             : AIC=1292.003, Time=0.11 sec
 ARIMA(1,2,1)(0,0,0)[0]             : AIC=1290.670, Time=0.05 sec

Best model:  ARIMA(2,2,0)(0,0,0)[0]
Total fit time: 0.709 seconds

/usr/local/lib/python3.10/dist-packages/sklearn/utils/
deprecation.py:151: FutureWarning: 'force_all_finite' was renamed to
```

```
'ensure_all_finite' in 1.6 and will be removed in 1.8.
  warnings.warn(

ARIMA(order=(2, 2, 0), scoring_args={}, suppress_warnings=True,
      with_intercept=False)
```

```
from statsmodels.tsa.arima.model  import ARIMA
model = ARIMA(gdpseries, order=(2,2,0)).fit()
model.params
print(model.params)
model.aic
print(model.aic)
```

```
ar.L1    -6.732785e-01
ar.L2    -2.369390e-01
ma.L1     3.545751e-01
ma.L2     3.514057e-02
sigma2    7.970162e+08
dtype: float64
1293.8912617376225
```

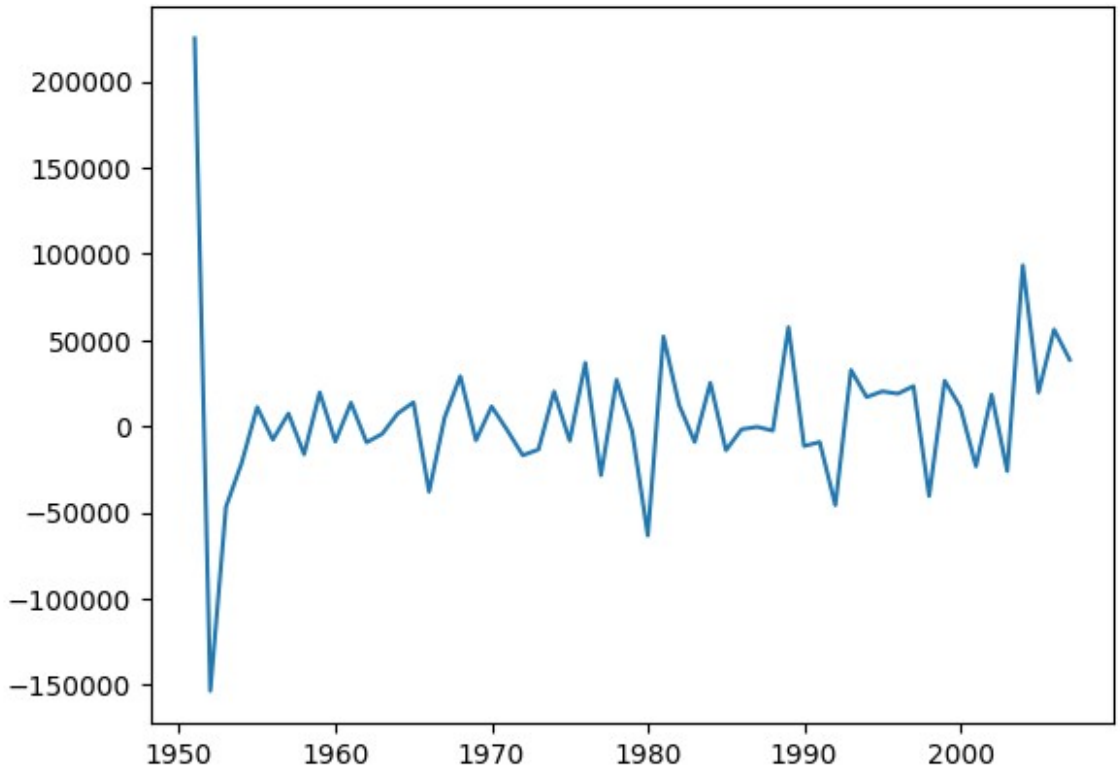
```
resi = model.resid
from statsmodels.stats.diagnostic import acorr_ljungbox
acorr_ljungbox(resi,lags=8,boxpierce=True)
```


```
{"summary":"{\n  \"name\":
\"acorr_ljungbox(resi,lags=10,boxpierce=True)\",\n  \"rows\": 10,\n
\"fields\": [\n    {\n      \"column\": \"lb_stat\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.4870372045280769,\n        \"min\": 5.251549063626597,\n
\"max\": 6.6122249595046005,\n        \"num_unique_values\": 10,\n
\"samples\": [\n          6.449060354718018,\n
5.284770792897964,\n          5.502523390013077\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"lb_pvalue\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
0.27098826605803555,\n        \"min\": 0.021927242163080854,\n
\"max\": 0.7614756429531917,\n        \"num_unique_values\": 10,\n
\"samples\": [\n          0.6942680067151882,\n
0.07119124750301303,\n          0.48115177116806274\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"bp_stat\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.4052853010924584,\n        \"min\": 4.9845211451371085,\n
\"max\": 6.122318781638787,\n        \"num_unique_values\": 10,\n
\"samples\": [\n          5.99234019816473,\n
5.015490553779909,\n          5.211751073618509\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"bp_pvalue\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
```

```
0.285099069182749,\n        \"min\": 0.025575062345128668,\n
\"max\": 0.8048839423670772,\n        \"num_unique_values\": 10,\n
\"samples\": [\n          0.7406846739557202,\n
0.08145168334813911,\n          0.5169555556335477\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```


```
plt.plot(resi)
```

```
[<matplotlib.lines.Line2D at 0x7ee2e19a2920>]
```




```
model.forecast(steps=3)
```

```
2007-12-31    3.076728e+06
2008-12-31    3.309975e+06
2009-12-31    3.543641e+06
Freq: YE-DEC, Name: predicted_mean, dtype: float64
```