

Surakshith Shetty -53026240013

PRACTICAL 3: To analyze crop production trends over time using time series analysis techniques, identifying patterns, seasonality, and forecasting future yields

```
import pandas as pd

data = pd.read_csv('/content/CROP DATA.csv')

data.tail()

{"summary": "{\n    \"name\": \"data\", \n    \"rows\": 5, \n    \"fields\": [\n        {\n            \"column\": \"Year\", \n            \"properties\": {\n                \"dtype\": \"number\", \n                \"std\": 0, \n                \"min\": 2022, \n                \"max\": 2023, \n                \"num_unique_values\": 2, \n                \"samples\": [2023, 2022], \n                \"semantic_type\": \"\", \n                \"description\": \"\"\\n\"\"\n            }\n        }, \n        {\n            \"column\": \"Quarter\", \n            \"properties\": {\n                \"dtype\": \"number\", \n                \"std\": 1, \n                \"min\": 1, \n                \"max\": 4, \n                \"num_unique_values\": 4, \n                \"samples\": [1, 3], \n                \"semantic_type\": \"\", \n                \"description\": \"\"\\n\"\"\n            }\n        }, \n        {\n            \"column\": \"CROPYIELD\", \n            \"properties\": {\n                \"dtype\": \"number\", \n                \"std\": 278.11276890552864, \n                \"min\": 21989.981, \n                \"max\": 22668.986, \n                \"num_unique_values\": 5, \n                \"samples\": [22112.329, 22668.986], \n                \"semantic_type\": \"\", \n                \"description\": \"\"\\n\"\"\n            }\n        }\n    ]\n}, \"type\": \"dataframe\"}

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 308 entries, 0 to 307
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Year        308 non-null    int64  
 1   Quarter     308 non-null    int64  
 2   CROPYIELD  308 non-null    float64 
dtypes: float64(1), int64(2)
memory usage: 7.3 KB

rng = pd.date_range(start='1947', end='2024', freq='Y')

<ipython-input-25-741d5901ddbf>:1: FutureWarning: 'Y' is deprecated
and will be removed in a future version, please use 'YE' instead.
  rng = pd.date_range(start='1947', end='2024', freq='Y')

rng
```

```
DatetimeIndex(['1947-12-31', '1948-12-31', '1949-12-31', '1950-12-31',
                '1951-12-31', '1952-12-31', '1953-12-31', '1954-12-31',
                '1955-12-31', '1956-12-31', '1957-12-31', '1958-12-31',
                '1959-12-31', '1960-12-31', '1961-12-31', '1962-12-31',
                '1963-12-31', '1964-12-31', '1965-12-31', '1966-12-31',
                '1967-12-31', '1968-12-31', '1969-12-31', '1970-12-31',
                '1971-12-31', '1972-12-31', '1973-12-31', '1974-12-31',
                '1975-12-31', '1976-12-31', '1977-12-31', '1978-12-31',
                '1979-12-31', '1980-12-31', '1981-12-31', '1982-12-31',
                '1983-12-31', '1984-12-31', '1985-12-31', '1986-12-31',
                '1987-12-31', '1988-12-31', '1989-12-31', '1990-12-31',
                '1991-12-31', '1992-12-31', '1993-12-31', '1994-12-31',
                '1995-12-31', '1996-12-31', '1997-12-31', '1998-12-31',
                '1999-12-31', '2000-12-31', '2001-12-31', '2002-12-31',
                '2003-12-31', '2004-12-31', '2005-12-31', '2006-12-31',
                '2007-12-31', '2008-12-31', '2009-12-31', '2010-12-31',
                '2011-12-31', '2012-12-31', '2013-12-31', '2014-12-31',
                '2015-12-31', '2016-12-31', '2017-12-31', '2018-12-31',
                '2019-12-31', '2020-12-31', '2021-12-31', '2022-12-31',
                '2023-12-31'],
               dtype='datetime64[ns]', freq='YE-DEC')
```

```
yearly_data = data.groupby('Year')[['CROPYIELD']].sum()
```

```
yearly_data
```

Year	CROPYIELD
1947	8738.457
1948	9098.506
1949	9047.711
1950	9834.127
1951	10625.280
..	
2019	82768.347
2020	80936.296
2021	85630.770
2022	87288.146
2023	89497.357

```
Name: CROPYIELD, Length: 77, dtype: float64
```

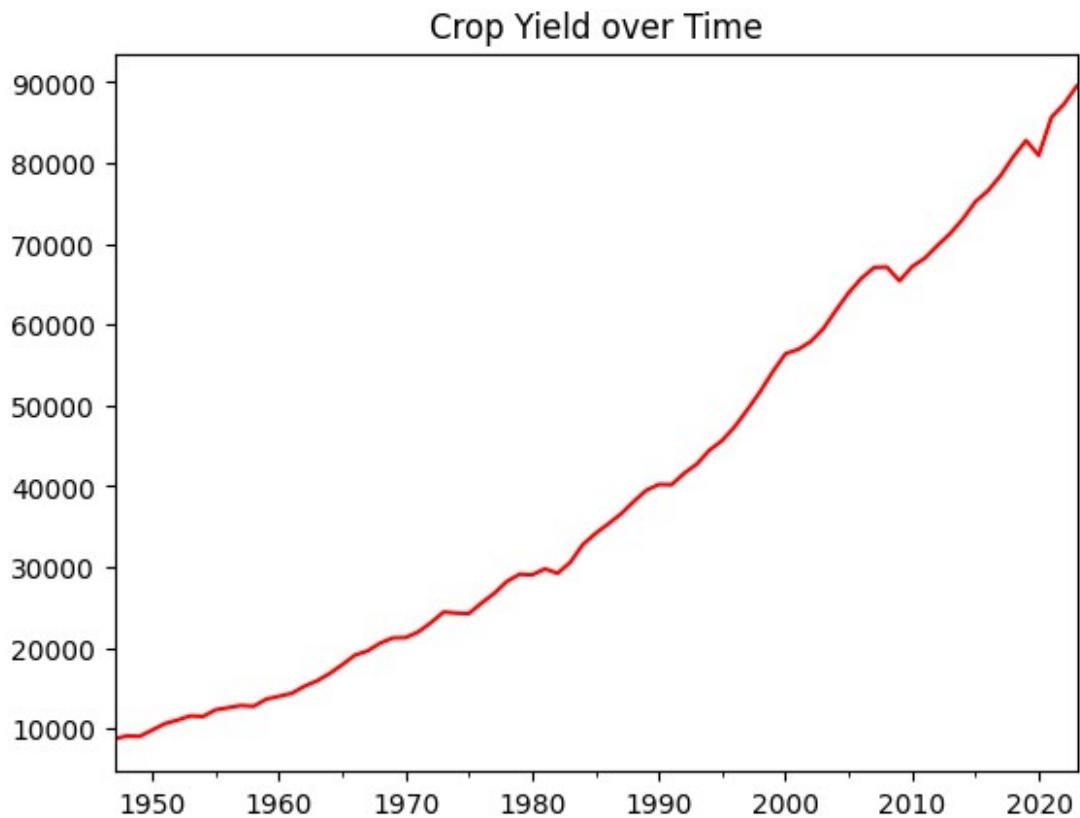
```
rng = pd.date_range(start='1947', end='2024', freq='Y')
s = yearly_data.values
Crop = pd.Series(s,rng)
```

```
<ipython-input-29-a7621b9330fe>:1: FutureWarning: 'Y' is deprecated
and will be removed in a future version, please use 'YE' instead.
    rng = pd.date_range(start='1947', end='2024', freq='Y')
```

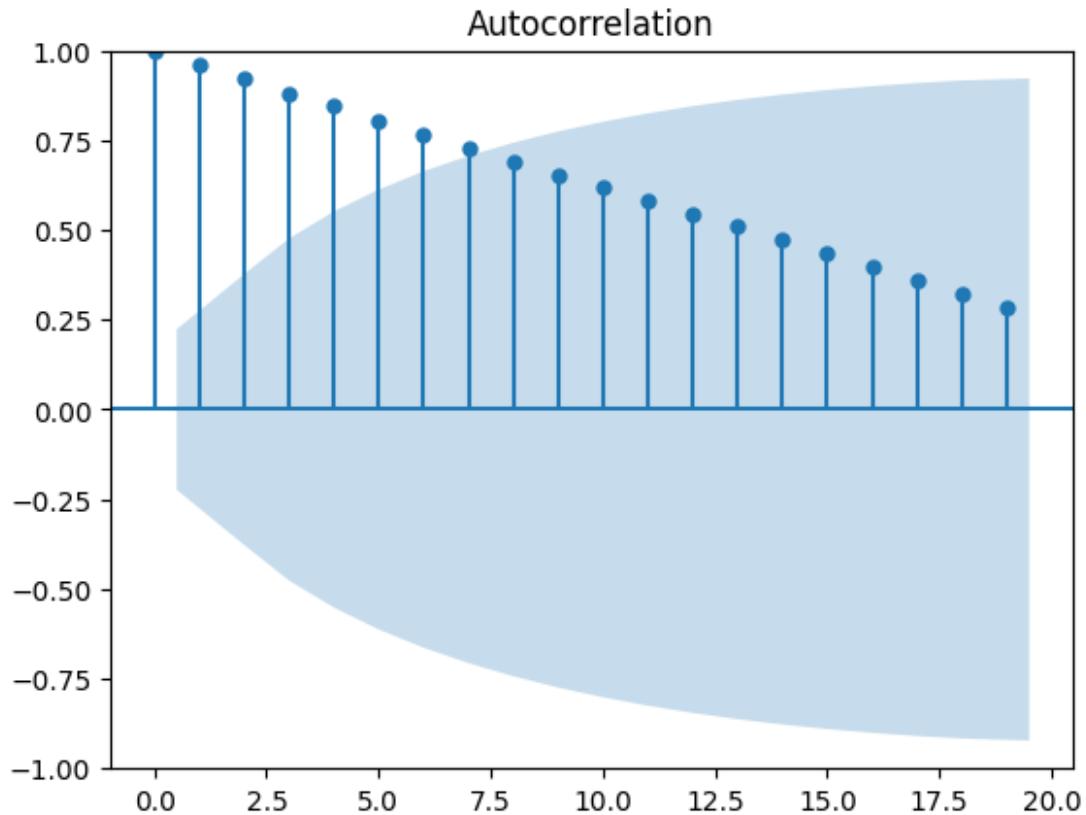
```
Crop
```

```
1947-12-31      8738.457
1948-12-31      9098.506
1949-12-31      9047.711
1950-12-31      9834.127
1951-12-31     10625.280
...
2019-12-31     82768.347
2020-12-31     80936.296
2021-12-31     85630.770
2022-12-31     87288.146
2023-12-31     89497.357
Freq: YE-DEC, Length: 77, dtype: float64

Crop.plot(color='red',title = 'Crop Yield over Time')
<Axes: title={'center': 'Crop Yield over Time'}>
```

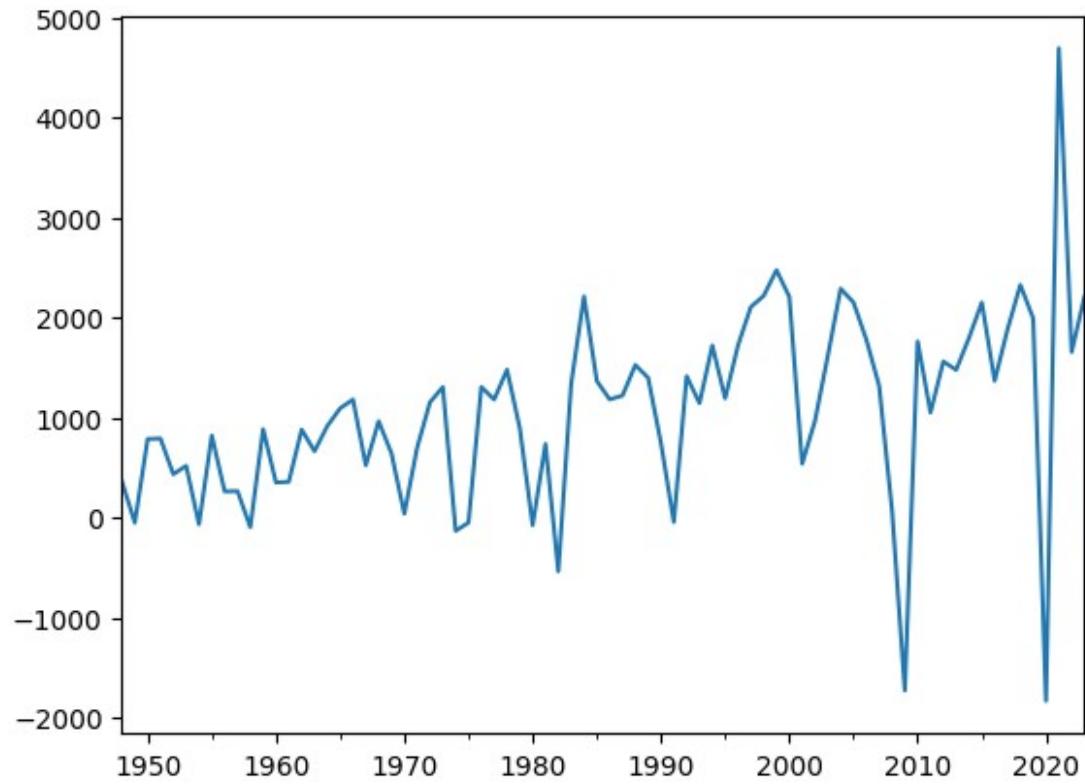


```
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
plot_acf(Crop)
plt.show()
```

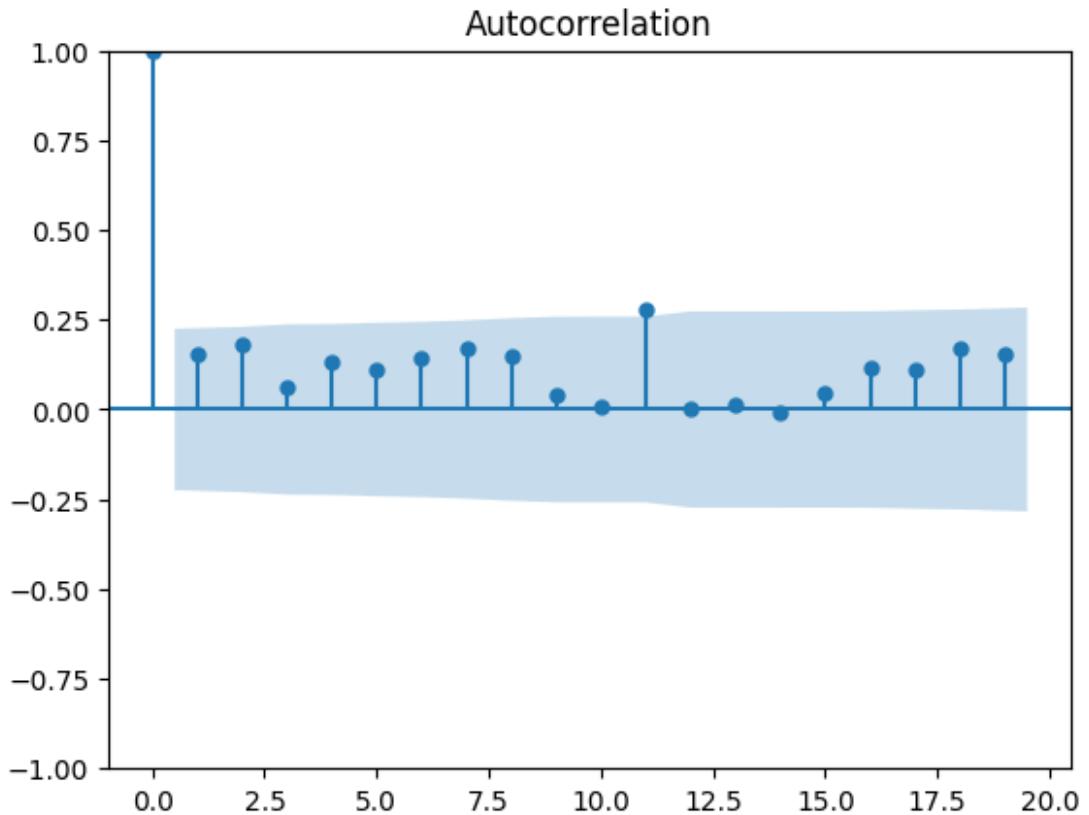


```
from statsmodels.tsa.statespace.tools import diff
Cropdiff = diff(Crop)
Cropdiff.plot()
Cropdiff.head()

1948-12-31    360.049
1949-12-31    -50.795
1950-12-31    786.416
1951-12-31    791.153
1952-12-31    433.932
Freq: YE-DEC, dtype: float64
```

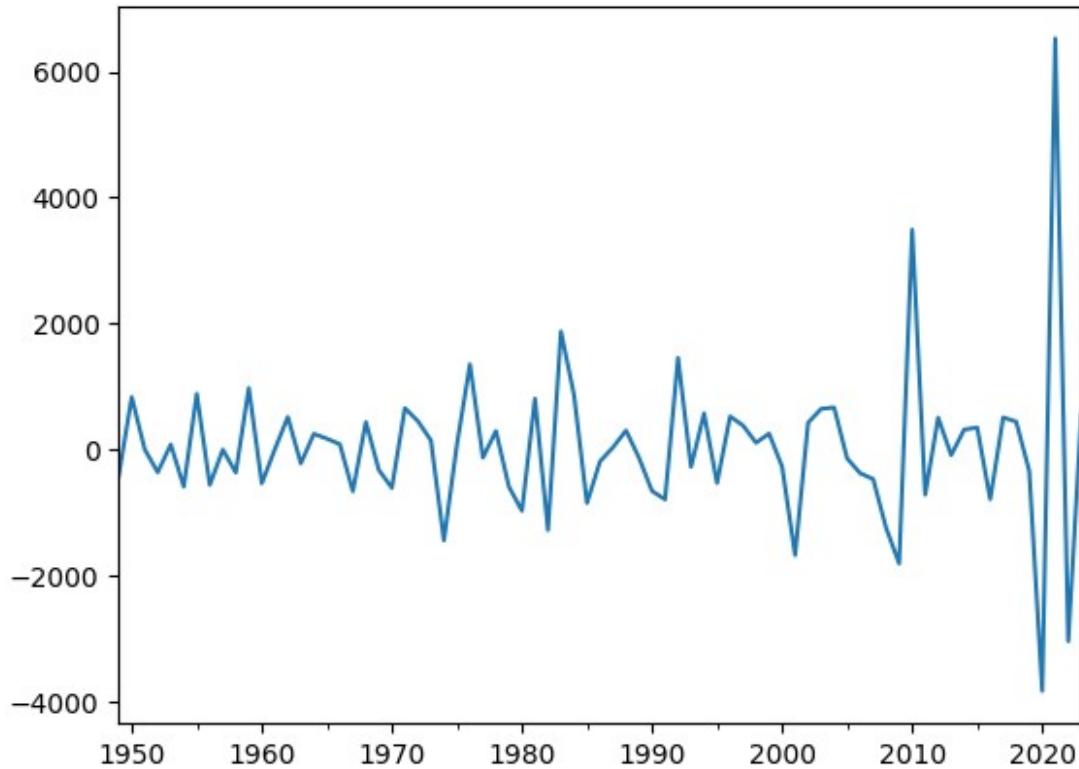


```
plot_acf(Cropdiff)  
plt.show()
```



```
from statsmodels.tsa.statespace.tools import diff
Cropdiff1 = diff(Cropdiff)
Cropdiff1.plot()
Cropdiff1.head()

1949-12-31    -410.844
1950-12-31     837.211
1951-12-31      4.737
1952-12-31    -357.221
1953-12-31     84.503
Freq: YE-DEC, dtype: float64
```



```
!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)
=====
          0.0/985.1 kB ? eta -:----:-
=====
         985.1/985.1 kB 31.5 MB/s eta
0:00:00

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

<class 'statsmodels.iolib.summary.Summary'>
"""
    Augmented Dickey-Fuller Results
=====
Test Statistic                 4.038
P-value                      1.000
Lags                          0
-----
Trend: Constant
Critical Values: -3.52 (1%), -2.90 (5%), -2.59 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
"""

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

<class 'statsmodels.iolib.summary.Summary'>
"""
    Augmented Dickey-Fuller Results
=====
Test Statistic                 -7.241
P-value                      0.000
Lags                          0
-----
Trend: Constant
Critical Values: -3.52 (1%), -2.90 (5%), -2.59 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
"""

!pip install pmdarima
```

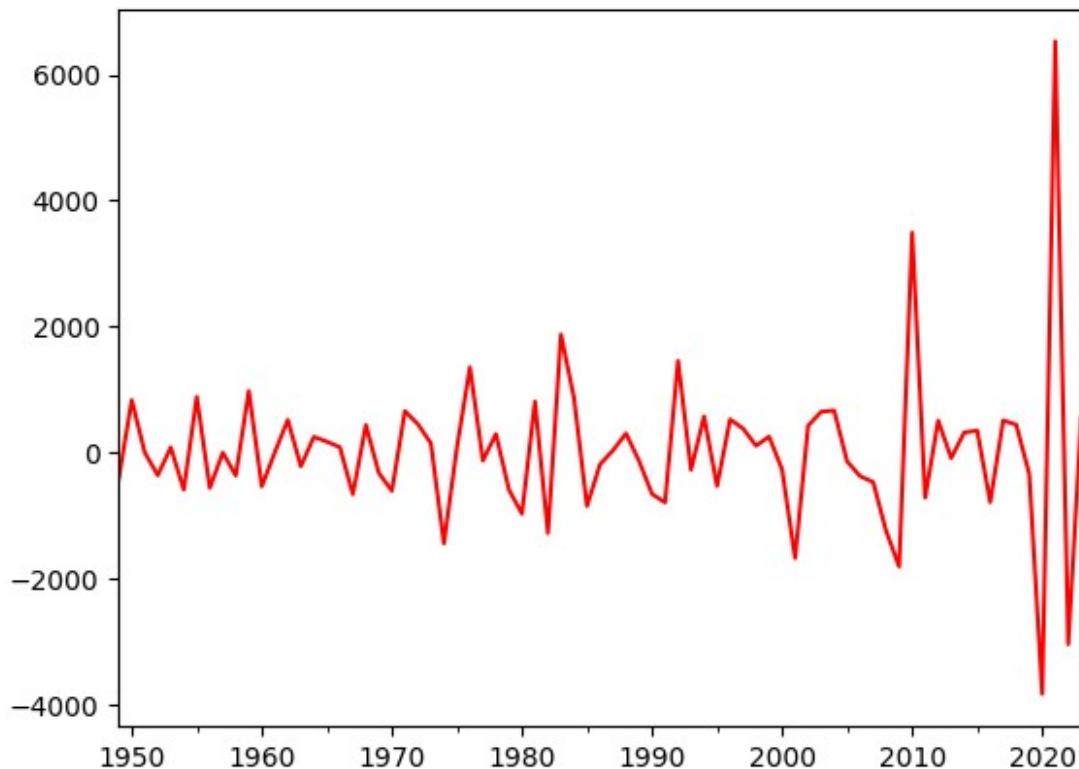
```
Collecting pmdarima
  Downloading pmdarima-2.0.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl.m
etadata (7.8 kB)
Requirement already satisfied: joblib>=0.11 in
/usr/local/lib/python3.10/dist-packages (from pmdarima) (1.4.2)
Requirement already satisfied: Cython!=0.29.18,!>=0.29.31,>=0.29 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 33.8 MB/s eta
0:00:00
darima
Successfully installed pmdarima-2.0.4
```



```
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

from statsmodels.tsa.statespace.tools import diff
Cropdiff = diff(Crop)
Cropdiff = diff(Cropdiff)
Cropdiff1.plot(color='red')

<Axes: >
```



```
import pmдарима as pm
model = pm.auto_arima(Crop,max_p=2, max_q=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
```

Performing stepwise search to minimize aic

```
optimizing stepwise search to maximize AIC
ARIMA(2,2,2)(0,0,0)[0] intercept : AIC=inf, Time=0.31 sec
ARIMA(0,2,0)(0,0,0)[0] intercept : AIC=1283.929, Time=0.02 sec
ARIMA(1,2,0)(0,0,0)[0] intercept : AIC=1267.411, Time=0.03 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=inf, Time=0.09 sec
ARIMA(0,2,0)(0,0,0)[0] : AIC=1281.960, Time=0.02 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(  
/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(2,2,0)(0,0,0)[0] intercept : AIC=1261.666, Time=0.09 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(  
  
ARIMA(2,2,1)(0,0,0)[0] intercept : AIC=inf, Time=0.24 sec  
ARIMA(1,2,1)(0,0,0)[0] intercept : AIC=inf, Time=0.15 sec  
ARIMA(2,2,0)(0,0,0)[0] : AIC=1259.699, 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(1,2,0)(0,0,0)[0] : AIC=1265.363, Time=0.03 sec  
ARIMA(2,2,1)(0,0,0)[0] : AIC=1247.884, Time=0.11 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(
    ARIMA(1,2,1)(0,0,0)[0] : AIC=1245.917, Time=0.12 sec
    ARIMA(0,2,1)(0,0,0)[0] : AIC=1244.431, Time=0.04 sec
    ARIMA(0,2,2)(0,0,0)[0] : AIC=1245.903, Time=0.09 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(1,2,2)(0,0,0)[0] : AIC=1248.282, Time=0.11 sec

Best model: ARIMA(0,2,1)(0,0,0)[0]
Total fit time: 1.510 seconds

ARIMA(order=(0, 2, 1), scoring_args={}, suppress_warnings=True,
      with_intercept=False)

from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(Crop, order=(0,2,1)).fit()
model.params
print(model.params)
model.aic
print(model.aic)

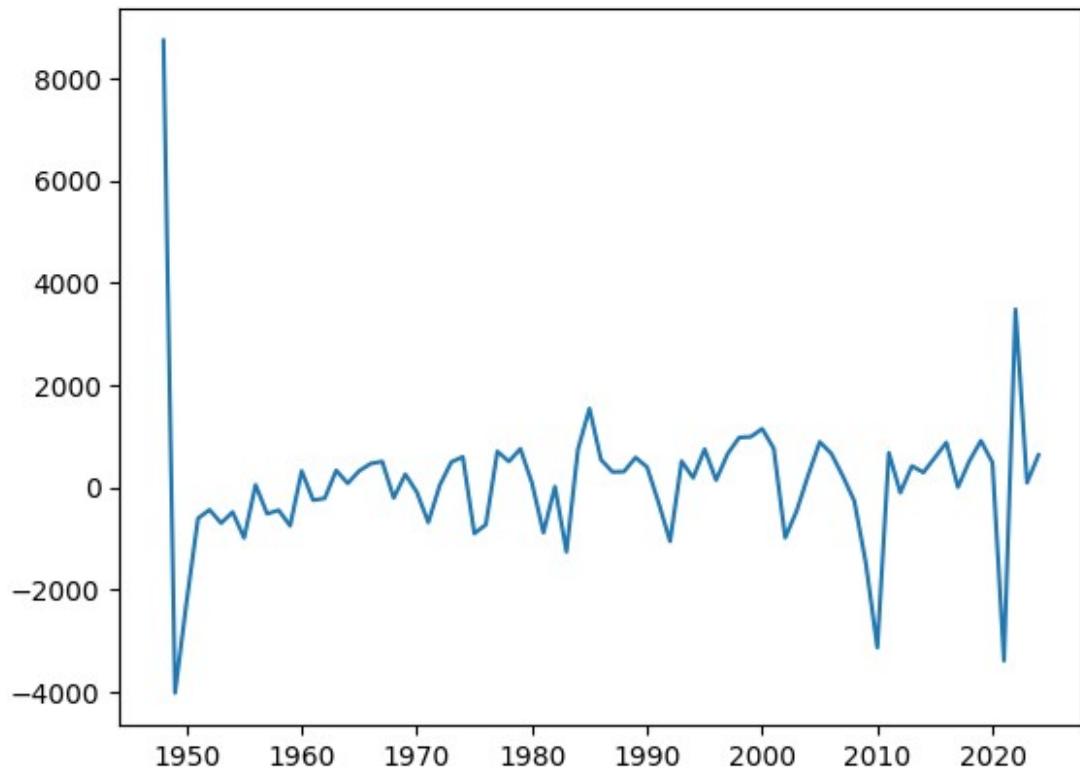
ma.L1          -0.898574
sigma2       822606.170846
dtype: float64
1244.4313001285873

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

{"summary": {
    "name": "acorr_ljungbox(resi, lags=8, boxpierce=True)", "rows": 8,
    "fields": [
        {"column": "lb_stat", "properties": {
            "dtype": "number", "min": 0.29111141090819614, "max": 2.8920102110465598, "samples": [2.5444527289573586]
        }
    }
}]


```

```
2.749660143875047,\n          1.9524470560493912\n      ],\n      {\\"semantic_type\\": \"\",\\n          \\"description\\": \"\\n      }\n    },\\n    {\\"n      \\"column\\": \"lb_pvalue\",\\n\n      \\"properties\\": {\\"n          \\"dtype\\": \"number\",\\n          \\"std\\\":\n0.2928440373240872,\\n          \\"min\\\": 0.16232340100637732,\\n\n          \\"max\\\": 0.9409504809955623,\\n          \\"num_unique_values\\\": 8,\\n\n          \\"samples\\\": [\n            0.28020708371381436,\n            0.8395470837959395,\n            0.16232340100637732\n          ],\n          {\\"semantic_type\\": \"\",\\n              \\"description\\": \"\\n      }\n        },\\n        {\\"n          \\"column\\": \"bp_stat\",\\n          \\"properties\\": {\\"n            \\"dtype\\": \"number\",\\n            \\"std\\\":\n0.2731689137022754,\\n            \\"min\\\": 1.8783034969589079,\n            \\"max\\\": 2.7552874938213296,\\n            \\"num_unique_values\\\": 8,\\n\n            \\"samples\\\": [\n              2.440334199086725,\n              2.630364830956817,\n              1.8783034969589079\n            ],\n            {\\"semantic_type\\": \"\",\\n                \\"description\\": \"\\n      }\n              },\\n              {\\"n                \\"column\\": \"bp_pvalue\",\\n\n                \\"properties\\": {\\"n                  \\"dtype\\": \"number\",\\n                  \\"std\\\":\n0.29232829351084194,\\n                  \\"min\\\": 0.1705270917041129,\n                  \\"max\\\": 0.9487637665010837,\\n                  \\"num_unique_values\\\": 8,\\n\n                  \\"samples\\\": [\n                    0.2951808382194512,\n                    0.8536019686669327,\n                    0.1705270917041129\n                  ],\n                  {\\"semantic_type\\": \"\",\\n                      \\"description\\": \"\\n      }\n                }\n              ]\\n},\n      \"type\": \"dataframe\"\n    }\n  }\n}\nplt.plot(resi)\n[<matplotlib.lines.Line2D at 0x789bd87ce800>]
```



```
model.forecast(steps=3)

2024-12-31    91138.301869
2025-12-31    92779.246738
2026-12-31    94420.191607
Freq: YE-DEC, Name: predicted_mean, dtype: float64
```