

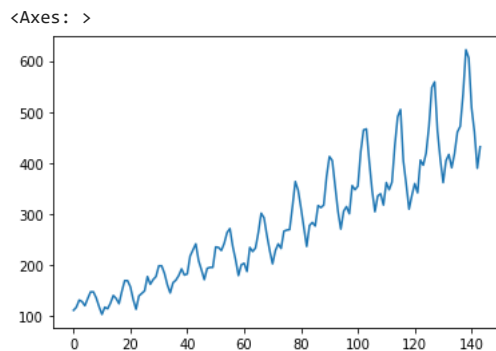
```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv("/content/drive/MyDrive/Datasets-20220805T064950Z-001.zip (Unzipped Files)/Datasets/AirPassengers - AirPassengers.csv")
df
```

	Month	#Passengers
0	1949-01	112
1	1949-02	118
2	1949-03	132
3	1949-04	129
4	1949-05	121
...	...	...
139	1960-08	606
140	1960-09	508
141	1960-10	461
142	1960-11	390
143	1960-12	432

144 rows × 2 columns

```
df['#Passengers'].plot()
```



```
df['diff_shift_1'] = df['#Passengers'] - df['#Passengers'].shift(1)
```

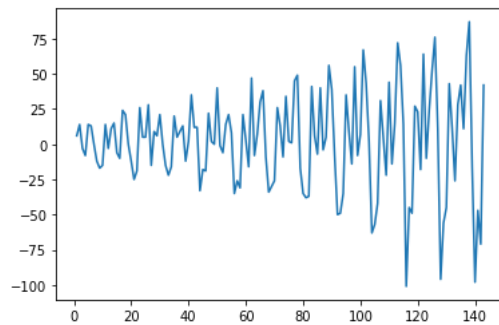
```
df
```

	Month	#Passengers	diff_shift_1
0	1949-01	112	NaN
1	1949-02	118	6.0
2	1949-03	132	14.0
3	1949-04	129	-3.0
4	1949-05	121	-8.0
...	...	...	...
139	1960-08	606	-16.0
140	1960-09	508	-98.0
141	1960-10	461	-47.0
142	1960-11	390	-71.0
143	1960-12	432	42.0

144 rows × 3 columns

```
df['diff_shift_1'].plot()
```

&lt;Axes: &gt;



```
from statsmodels.tsa.stattools import adfuller
```

```
adfuller(df['diff_shift_1'].dropna())
```

```
(-2.8292668241700047,
 0.05421329028382478,
 12,
 130,
 {'1%': -3.4816817173418295,
  '5%': -2.8840418343195267,
  '10%': -2.578770059171598},
 988.5069317854085)
```

```
def adf_test(series):
    result = adfuller(series)
    print('p - values : {}'.format(result[1]))
    if result[1]<=0.05:
        print('Strong evidence againsts the null hypothesis,reject null hypothesis,indicating that data is Stationary')
    else:
        print('weak evidence against null hypothesis, indicating that the data is non-stationary')
```

```
adf_test(df['diff_shift_1'].dropna())
```

```
p - values : 0.05421329028382478
weak evidence against null hypothesis, indicating that the data is non-stationary
```

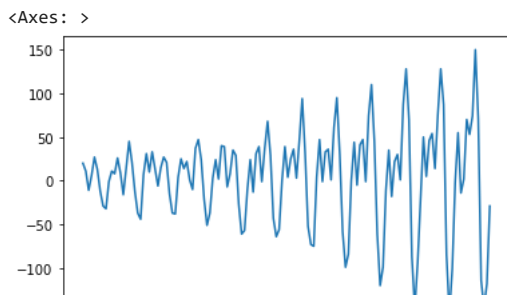
```
df['diff_shift_2'] = df['#Passengers'] - df['#Passengers'].shift(2)
```

```
df
```

	Month	#Passengers	diff_shift_1	diff_shift_2	
0	1949-01	112	NaN	NaN	
1	1949-02	118	6.0	NaN	
2	1949-03	132	14.0	20.0	
3	1949-04	129	-3.0	11.0	
4	1949-05	121	-8.0	-11.0	
...	...	...	...	...	
139	1960-08	606	-16.0	71.0	
140	1960-09	508	-98.0	-114.0	
141	1960-10	461	-47.0	-145.0	
142	1960-11	390	-71.0	-118.0	
143	1960-12	432	42.0	-29.0	

```
144 rows × 4 columns
```

```
df['diff_shift_2'].plot()
```



```
adf_test(df['diff_shift_2'].dropna())
```

```
p - values : 0.03862975767698902
```

```
Strong evidence againts the null hypothesis,reject null hypothesis,indicating that data is Stationary
```

```
from statsmodels.tsa.ar_model import AutoReg
```

```
dff = df['diff_shift_2'].dropna()
```

```
dff.shape
```

```
(142,)
```

```
train = dff[:len(dff)-7]
```

```
train.shape
```

```
(135,)
```

```
test = dff[len(dff)-7:]
```

```
test.shape
```

```
(7,)
```

```
test
```

```
137      74.0
```

```
138     150.0
```

```
139      71.0
```

```
140    -114.0
```

```
141    -145.0
```

```
142    -118.0
```

```
143     -29.0
```

```
Name: diff_shift_2, dtype: float64
```

```
model = AutoReg(df['diff_shift_2'].dropna(),lags=1).fit()
```

```
/usr/local/lib/python3.9/dist-packages/statsmodels/tsa/base/tsa_model.py:471: ValueWarning: An unsupported index was provided and will self._init_dates(dates, freq)
```

```
pred = model.predict(start = 136, end = 142)
```

```
/usr/local/lib/python3.9/dist-packages/statsmodels/tsa/base/tsa_model.py:834: ValueWarning: No supported index is available. Prediction return get_prediction_index()
```

```
len(dff)-1
```

```
141
```

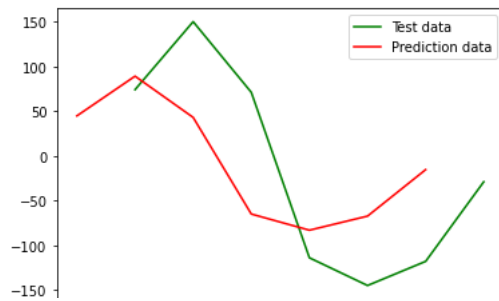
```
import matplotlib.pyplot as plt
```

```
plt.plot(test,label="Test data",color='g')
```

```
plt.plot(pred,label="Prediction data",color='r')
```

```
plt.legend()
```

<matplotlib.legend.Legend at 0x7f0346b69520>



```
from sklearn.metrics import mean_squared_error
```

```
pred.shape
```

```
(7,)
```

```
test.shape
```

```
(7,)
```

```
rmse = np.sqrt(mean_squared_error(test,pred))
```

```
rmse
```

```
45.234376431761135
```