In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

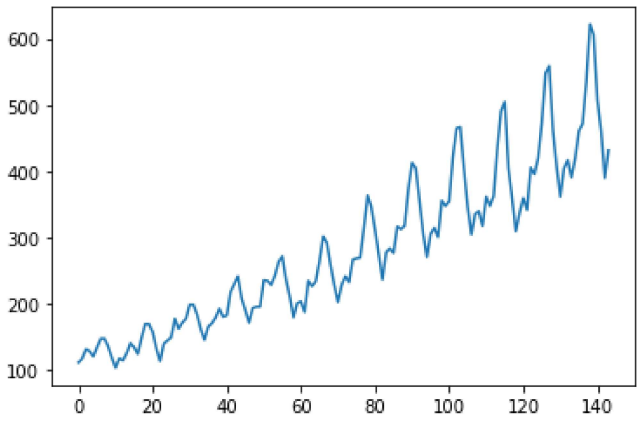
In [2]:
```
df = pd.read_csv('datasets/AirPassengers.csv')
df.head()
```

Out[2]:

	Month	#Passengers
0	1949-01	112
1	1949-02	118
2	1949-03	132
3	1949-04	129
4	1949-05	121

In [3]:
```
df['#Passengers'].plot()
```

Out[3]: <AxesSubplot:>



In [4]:
```
df['diff_shift_1'] = df['#Passengers']-df['#Passengers'].shift(1)
```

In [5]:
```
df
```

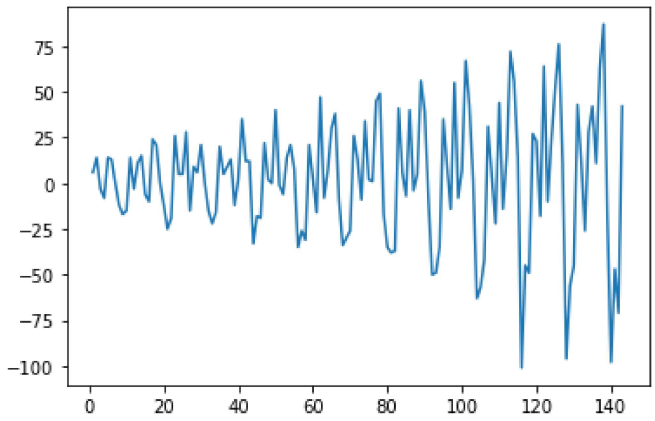
Out[5]:

	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

In [6]:
```
df['diff_shift_1'].plot()
```

Out[6]: <AxesSubplot:>



In [7]:
```
from statsmodels.tsa.stattools import adfuller
```

In [8]:
```
adfuller(df['diff_shift_1'].dropna())
```

Out[8]:
```
(-2.8292668241699888,
 0.054213290283826945,
 12,
 130,
 {'1%': -3.4816817173418295,
  '5%': -2.8840418343195267,
  '10%': -2.578770059171598},
 988.5069317854084)
```

In [9]:
```
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 th
    else:
        print('weak evidence against null hypothesis ,indicating that the data is non-stationary
```

In [10]:
```
adf_test(df['diff_shift_1'].dropna())
```

```
p_values : 0.054213290283826945
weak evidence against null hypothesis ,indicating that the data is non-stationary
```

In [11]:
```
df['diff_shift_2'] = df['#Passengers']-df['#Passengers'].shift(2)
```

In [12]:
```
df
```

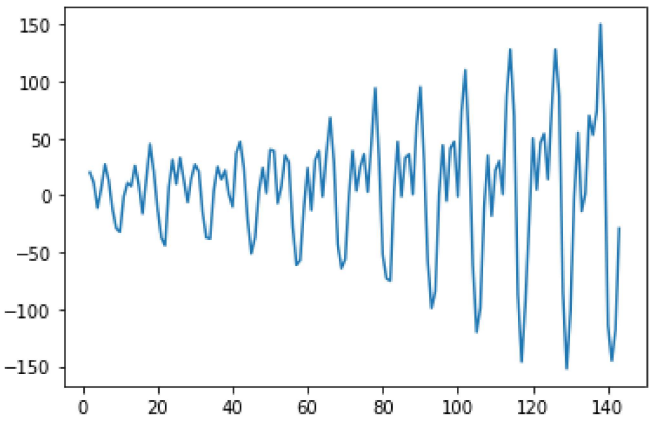
Out[12]:

	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

In [13]:
```
df['diff_shift_2'].plot()
```

Out[13]: <AxesSubplot:>



In [14]:
```
adf_test(df['diff_shift_2'].dropna())
```

```
p_values : 0.03862975767698791
Strong evidence againsts the null hypothesis,reject null hypothesis,indicating that data is st
ationary
```

In [15]:
```
from statsmodels.tsa.ar_model import AutoReg
```

In [16]:
```
dff= df['diff_shift_2'].dropna()
dff.shape
```

Out[16]: (142,)

In [17]:
```
train = dff[:len(dff)-7]
```

In [18]:
```
train.shape
```

Out[18]: (135,)

In [19]:
```
test = dff[len(dff)-7:]
```

In [20]:
```
test.shape
```

Out[20]:
```
(7,)
```

In [21]:
```
test
```

Out[21]:
```
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
```

In [60]:
```
model = AutoReg(df['diff_shift_2'].dropna(),lags=1).fit()
```

```
C:\Users\User15\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarnin
g: An unsupported index was provided and will be ignored when e.g. forecasting.
  self._init_dates(dates, freq)
```

In [61]:
```
pred = model.predict(start =136,end=142)
```

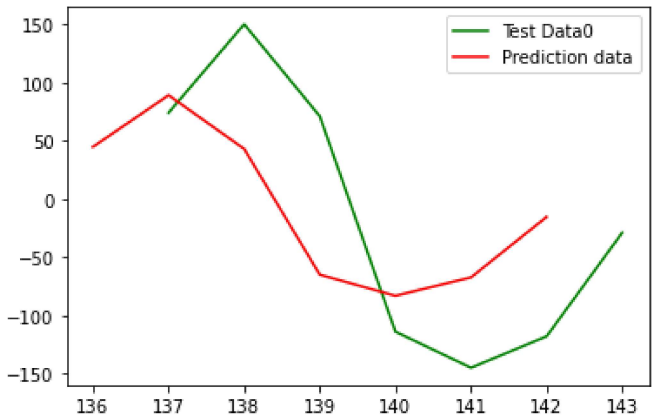
```
C:\Users\User15\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:834: ValueWarnin
g: No supported index is available. Prediction results will be given with an integer index beg
inning at `start`.
  return get_prediction_index(
```

In [62]:
```
len(dff)-1
```

Out[62]:
```
141
```

In [63]:
```
plt.plot(test,label='Test Data0',color='g')
plt.plot(pred,label='Prediction data',color='r')
plt.legend()
```

Out[63]:
```
<matplotlib.legend.Legend at 0x1827d3e1640>
```



In [64]:
```
from sklearn.metrics import mean_squared_error
```

In [65]:
```
pred.shape
```

Out[65]:
```
(7,)
```

In [66]:
```
rmse = np.sqrt(mean_squared_error(test,pred))
```

In [67]:
```
rmse
```

Out[67]:
```
45.23437643176114
```

In [ ]:
```

```