TSA Data Conversion

In [1]:

```
#importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
df = pd.read_csv("AirPassengers.csv")
df
```

Out[2]:

	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

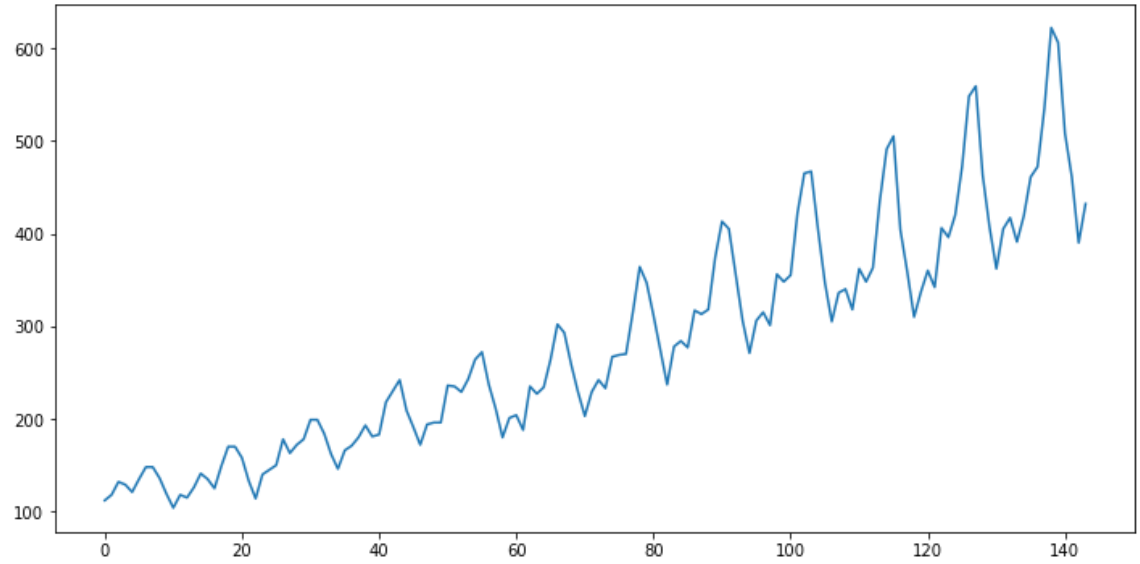
144 rows × 2 columns

In [3]:

```
plt.rcParams.update({'figure.figsize':(12,6)})
df['#Passengers'].plot()
```

Out[3]:

```
<AxesSubplot:>
```



Method 1 - Differencing and Seasonal differencing

In [4]:

```
#differencing meaning y(t) = y(t)-y(t-1)
```

In [5]:

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

Out[5]:

```
0        NaN
1      112.0
2      118.0
3      132.0
4      129.0
       ...
139    622.0
140    606.0
141    508.0
142    461.0
143    390.0
Name: #Passengers, Length: 144, dtype: float64
```

In [6]:

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

In [7]:

```
df.head()
```

Out[7]:

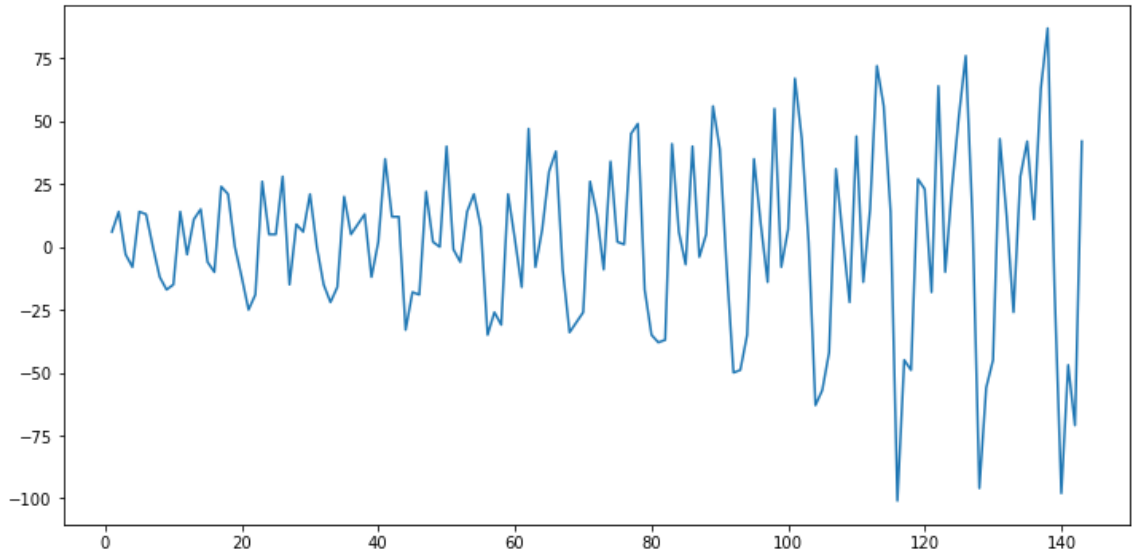
	Month	#Passengers	#Passengers_diff
0	1949-01	112	NaN
1	1949-02	118	6.0
2	1949-03	132	14.0
3	1949-04	129	-3.0
4	1949-05	121	-8.0

In [8]:

```
df['#Passengers_diff'].dropna().plot()
```

Out[8]:

```
<AxesSubplot:>
```



In [9]:

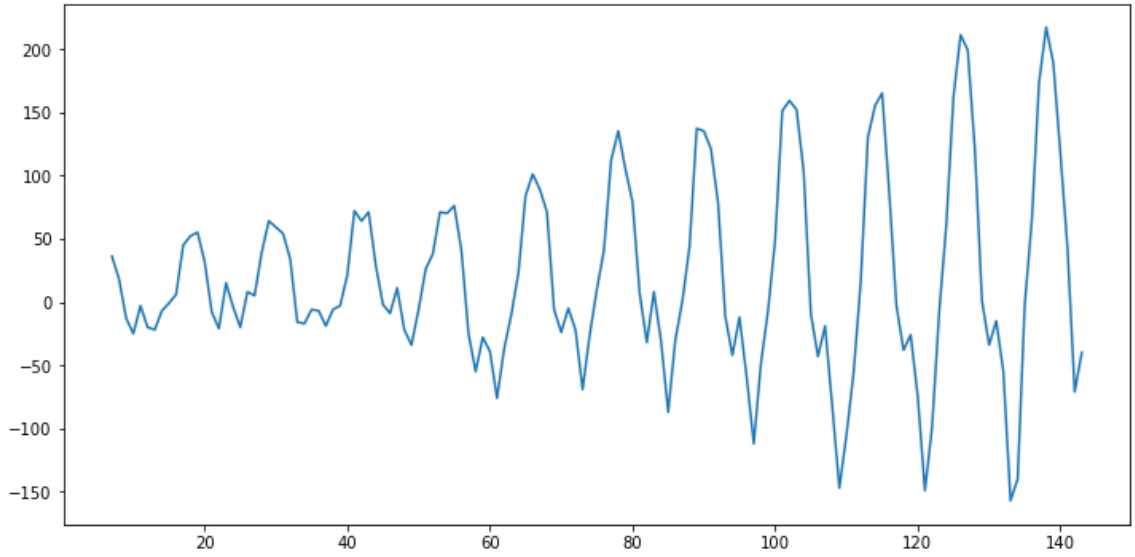
```
# Seasonal Differencing meaning - y(t) = y(t) - y(t-n)
```

In [10]:

```
df['#Passengers_diff_7'] = df['#Passengers'] - df['#Passengers'].shift(7)
df['#Passengers_diff_7'].dropna().plot()
```

Out[10]:

```
<AxesSubplot:>
```



In [11]:

```
df.head(10)
```

Out[11]:

	Month	#Passengers	#Passengers_diff	#Passengers_diff_7
0	1949-01	112	NaN	NaN
1	1949-02	118	6.0	NaN
2	1949-03	132	14.0	NaN
3	1949-04	129	-3.0	NaN
4	1949-05	121	-8.0	NaN
5	1949-06	135	14.0	NaN
6	1949-07	148	13.0	NaN
7	1949-08	148	0.0	36.0
8	1949-09	136	-12.0	18.0
9	1949-10	119	-17.0	-13.0

Method 2 - Transformation

In [12]:

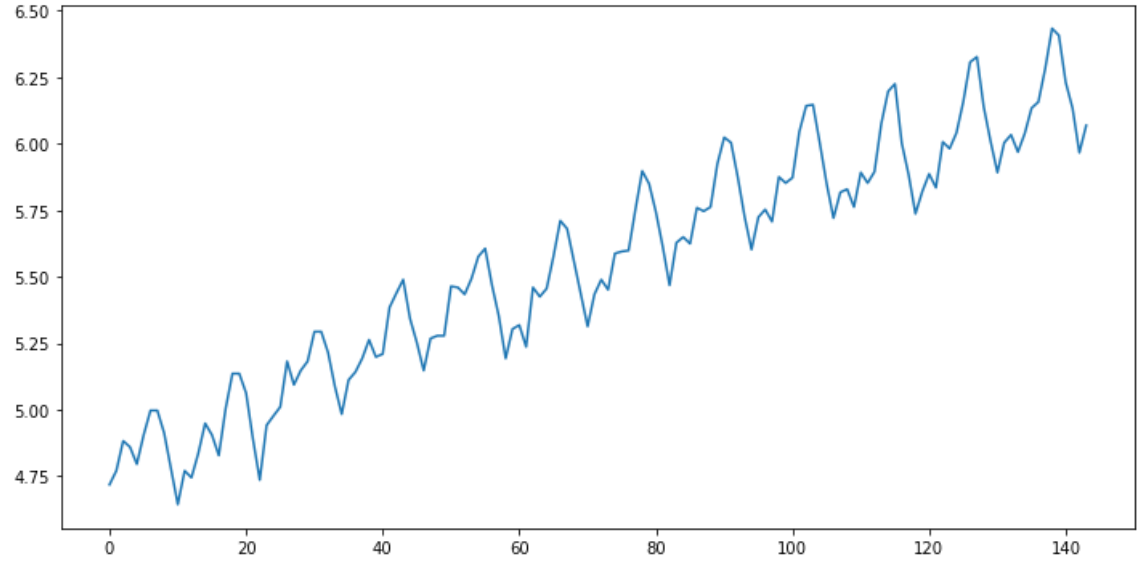
```
#Create transformation columns
import numpy as np

#calculate the log
df['adj_log'] = np.log(df['#Passengers'])

#calculate the square root
df['adj_sqrt'] = np.sqrt(df['#Passengers'])

#calculate the cubed root
df['adj_cbrt'] = np.cbrt(df['#Passengers'])
```

In [13]:

```
df['adj_log'].dropna().plot()
```

Out[13]:

<AxesSubplot:>

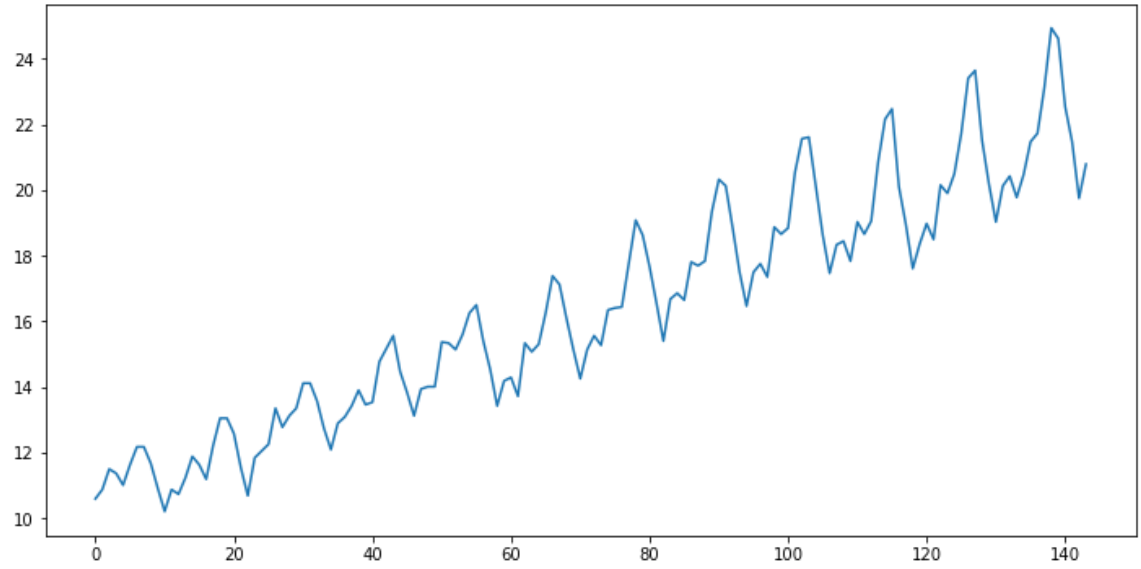


In [14]:

```
df['adj_sqrt'].dropna().plot()
```

Out[14]:

<AxesSubplot:>

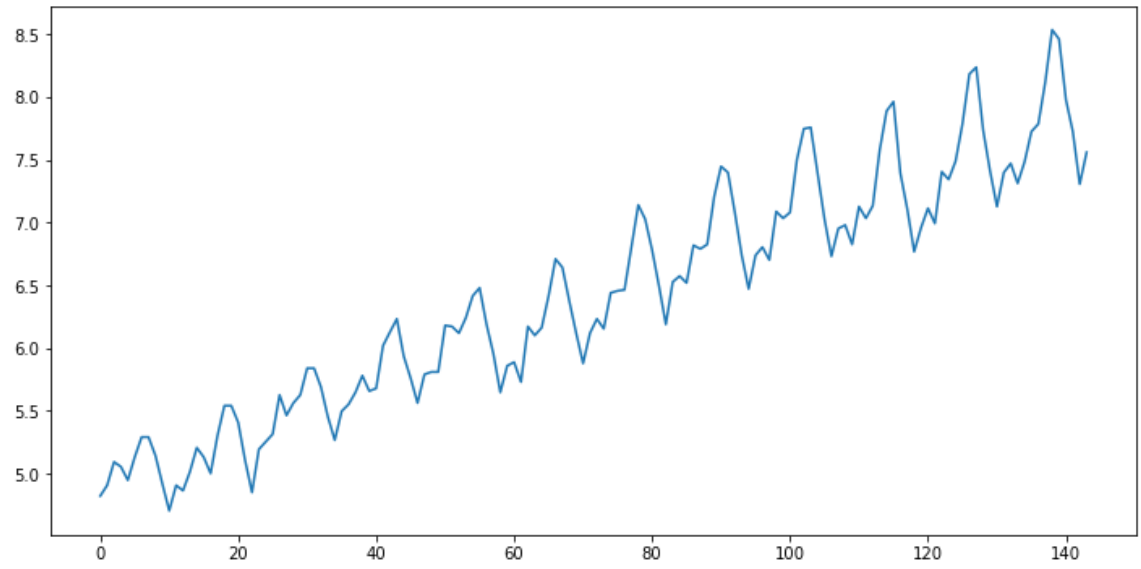


In [15]:

```
df['adj_cbrt'].dropna().plot()
```

Out[15]:

<AxesSubplot:>

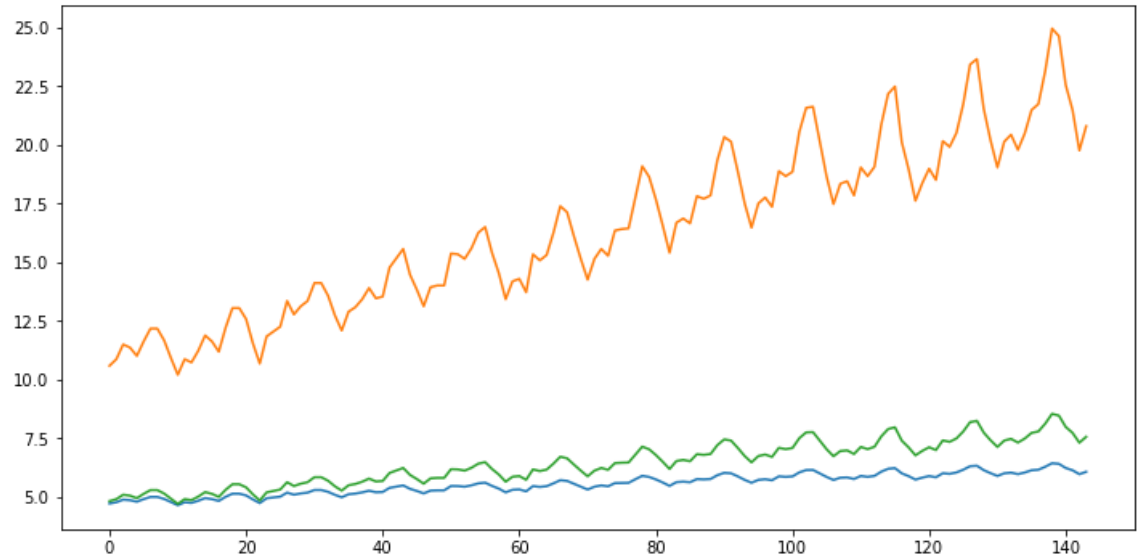


In [16]:

```
df['adj_log'].dropna().plot()
df['adj_sqrt'].dropna().plot()
df['adj_cbrt'].dropna().plot()
```

Out[16]:

<AxesSubplot:>

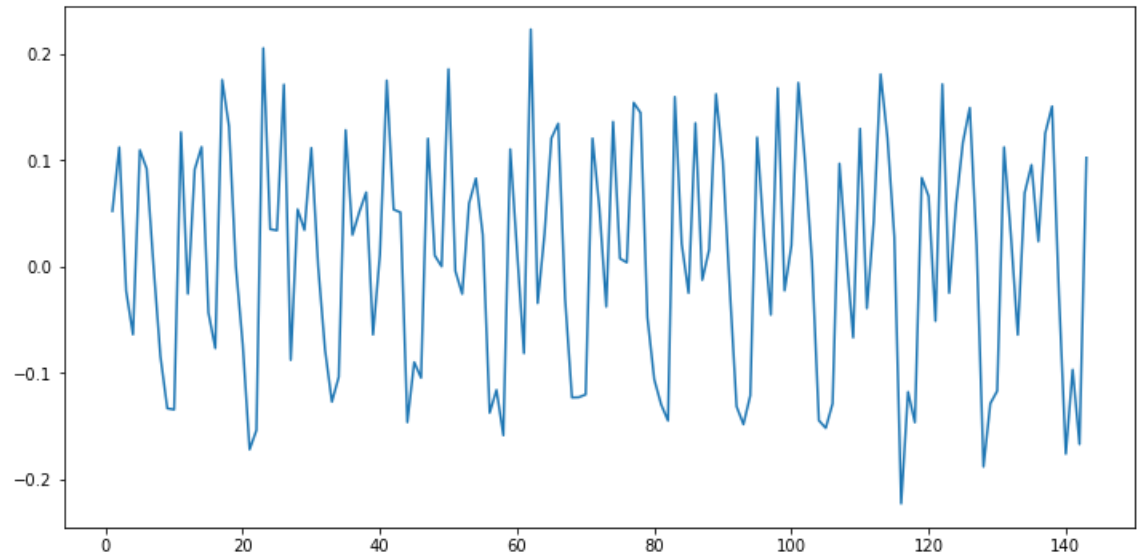


In [17]:

```
df['training_diff'] = df['adj_log'] - df['adj_log'].shift(1)
df['training_diff'].dropna().plot()
```

Out[17]:

<AxesSubplot:>

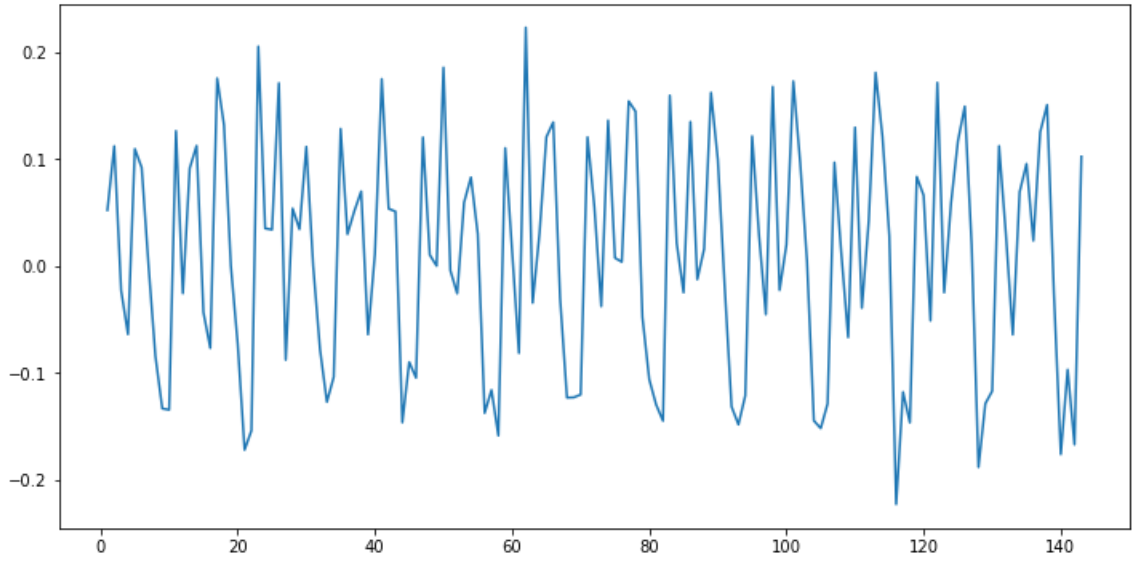


In [18]:

```
df['#Passengers_log_diff'] = df['adj_log'] - df['adj_log'].shift(1)
df['#Passengers_log_diff'].dropna().plot()
```

Out[18]:

<AxesSubplot:>

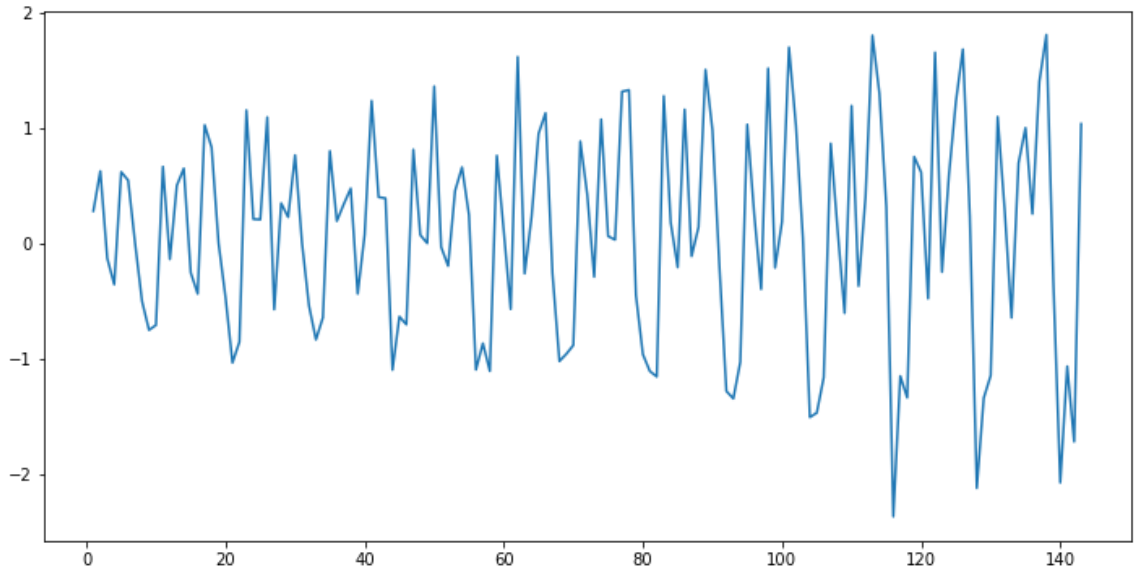


In [19]:

```
df['#Passengers_sqrt_diff'] = df['adj_sqrt'] - df['adj_sqrt'].shift(1)
df['#Passengers_sqrt_diff'].dropna().plot()
```

Out[19]:

<AxesSubplot:>

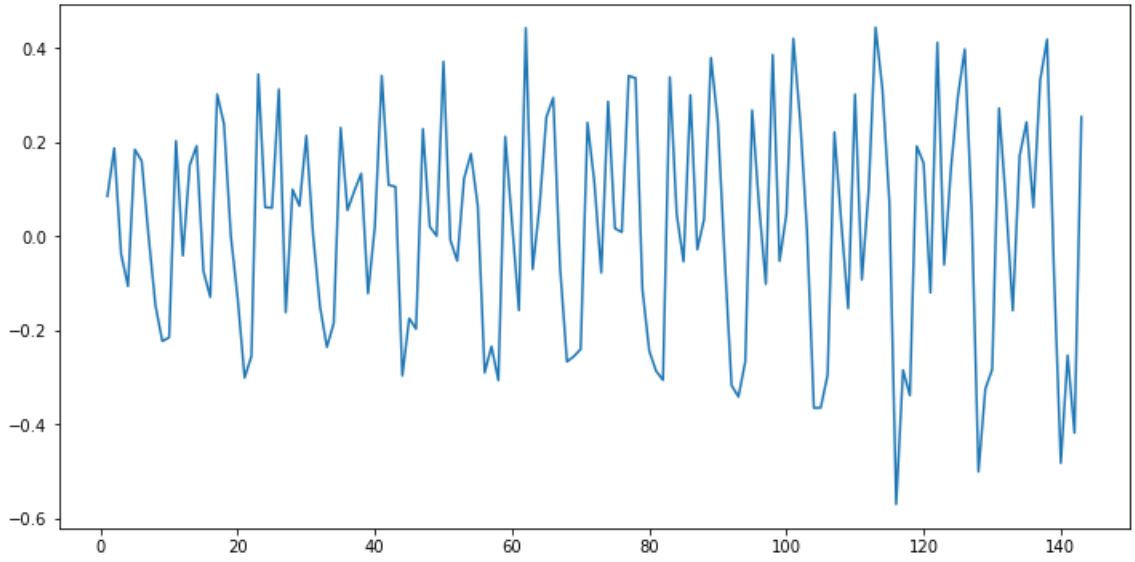


In [20]:

```
df['#Passengers_cbrt_diff'] = df['adj_cbrt'] - df['adj_cbrt'].shift(1)
df['#Passengers_cbrt_diff'].dropna().plot()
```

Out[20]:

```
<AxesSubplot:>
```



In [ ]: