outlier HW

```
In [74]: # create time series data
         dates = pd.date_range(start='2023-01-01', periods=100, freq='D')
         values = np.random.randn(100)*10 #정규분포를 따르는 값
         # Add outlier
         values[15] = 70
         df = pd.DataFrame({'Date':dates, 'Value':values})
  Out [74]:
                    Date
                           Value
           0 2023-01-01 13.993554
             1 2023-01-02 9.246337
            2 2023-01-03 0.596304
             3 2023-01-04 -6.469368
             4 2023-01-05 6.982233
             5 2023-01-06 3.934854
             6 2023-01-07 8.951932
             7 2023-01-08 6.351718
             8 2023-01-09 10.495527
             9 2023-01-10 -5.352352
            10 2023-01-11 13.173941
```

```
In [76]: # Detect outlier
z_scores = np.abs(stats.zscore(df['Value']))

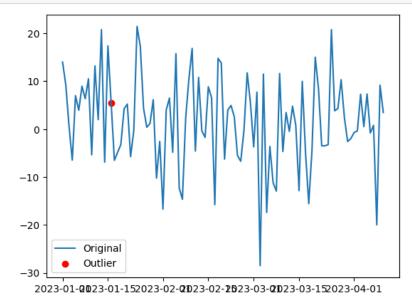
#Threshold
threshold = 3

#outlier
outliers = z_scores > threshold
df.loc[outliers,'Value'] = np.nan

df['Value'] = df['Value'].interpolate()
```

outlier HW 1

```
In [77]: #visualize
    plt.plot(df['Date'], df['Value'], label='Original')
    plt.scatter(df[outliers]['Date'], df[outliers]['Value'], color='r', label='Outlier')
    plt.legend()
    plt.show()
```



outlier HW 2