Time Series Analysis

AA 651, April 2021

Submitted by: Kishlay Singh

Roll Number: 2003121005

- **Stationarity in Time Series data:**

  Time series data or time-stamped data is a sequence of data points indexed in time order. These data points consist of successive measurements of a particular quantity over a period of time.

  Stationarity:

  Time series can be stationary or non-stationary.

  Stationary Time Series: The observations in a stationary time series are not dependent on time. Hence, in a stationary time series, we do not observe any clear trends or seasonality and the plot of the time series is almost horizontal.

  When a time series is stationary, it can be easier to model.

  Non-stationary Time Series: Observations from a non-stationary time series show seasonal effects, trends, and other structures that depend on the index.

  Summary statistics like mean and variance change over time in such data sets. So, if we calculate the mean and variance of different sections of the data, we expect to see vast differences in them.

- **Checking for stationarity in Time Series data:**

  A preliminary analysis of stationarity can be done by just looking at the plots and visually checking for any trends or seasonality.

  One 'quick and dirty' method is to splice the data into two or more parts and calculate their mean and variance and compare them.

  A more statistically sound method is the Dickey-Fuller test. In python, it is implemented as an Augmented Dickey-Fuller in Python through the *adfuller()* function found in the *statsmodels* module.

- **Augmented Dickey-Fuller Test:**

  Statistical tests make strong assumptions about your data. They can only be used to inform the degree to which a null hypothesis can be rejected or fail to be reject. The result must be interpreted for a given problem to be meaningful.

Nevertheless, they can provide a quick check and confirmatory evidence that your time series is stationary or non-stationary.

The Augmented Dickey-Fuller Test is a type of statistical test called a unit root test. The intuition behind a unit root test is that it determines how strongly a time series is defined by a trend.

There are a number of unit root tests and the Augmented Dickey-Fuller may be one of the more widely used. It uses an autoregressive model and optimizes an information criterion across multiple different lag values.

The null hypothesis of the test is that the time series can be represented by a unit root, that it is not stationary (has some time-dependent structure). The alternate hypothesis (rejecting the null hypothesis) is that the time series is stationary.

- **Null Hypothesis (H0)**: If failed to be rejected, it suggests the time series has a unit root, meaning it is non-stationary. It has some time dependent structure.
- **Alternate Hypothesis (H1)**: The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary. It does not have time-dependent structure.

We interpret this result using the p-value from the test. A p-value below a threshold (such as 5% or 1%) suggests we reject the null hypothesis (stationary), otherwise a p-value above the threshold suggests we fail to reject the null hypothesis (non-stationary).

- **p-value > 0.05**: Fail to reject the null hypothesis (H0), the data has a unit root and is non-stationary.
- **p-value <= 0.05**: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

The adfuller() function used in the Python system to perform an augmented Dickey-Fuller test has the following syntax:

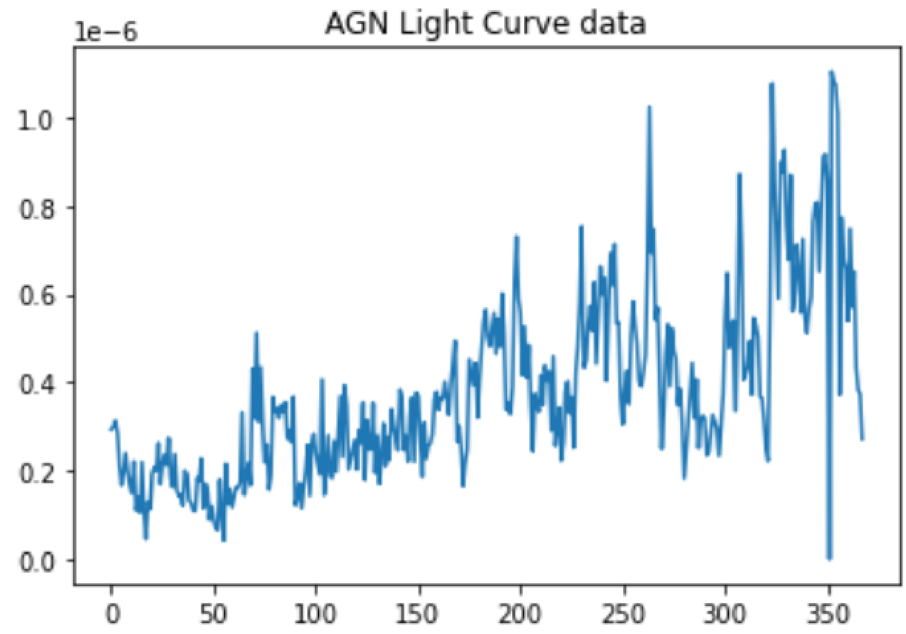
statsmodels.tsa.stattools.adfuller(x, maxlag=None, regression='c', autolag='AIC', store=False, regresults=False)

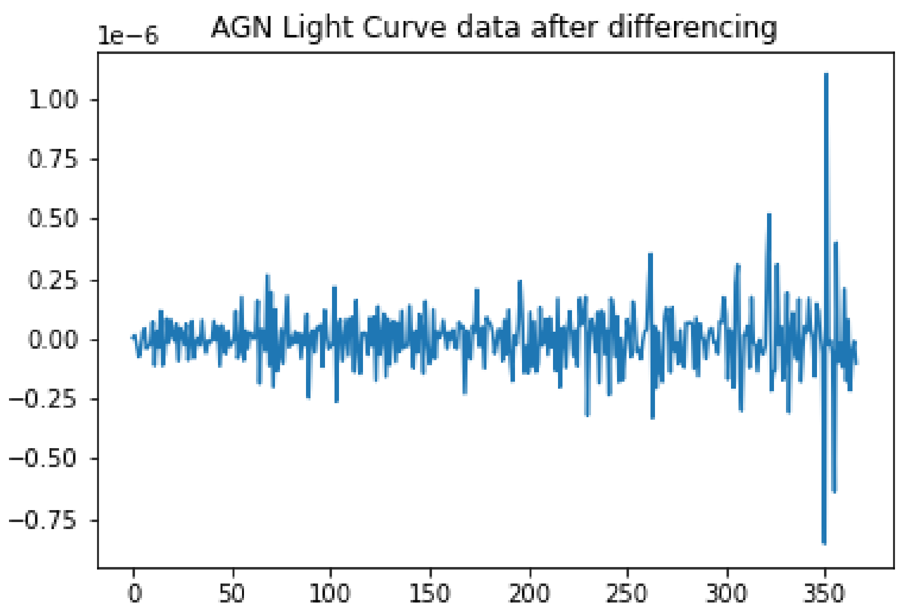
- **Making the non-stationary data into a stationary data set:**

  We use Differencing in this program to attempt making the non-stationary data into stationary data. In this process, each data point is subtracted from its consecutive data point. This helps stabilize the variance in the data. Other such transformative measures can also be used.

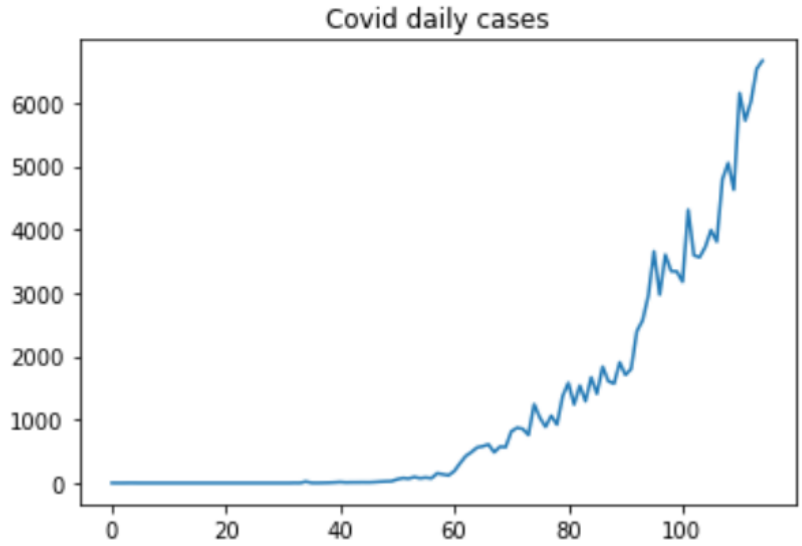
- **Results:**

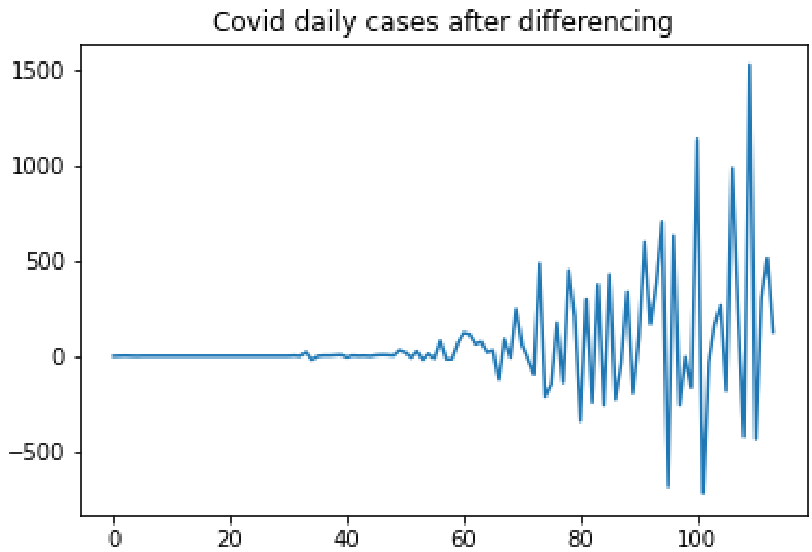
  Null hypothesis could not be rejected. AGN Light Curve data is non-stationary.
  p-Value is: 0.43344426663682534
  AGN Light Curve data has been made stationary through differencing
  p=Value is: 4.305082496591425e-10



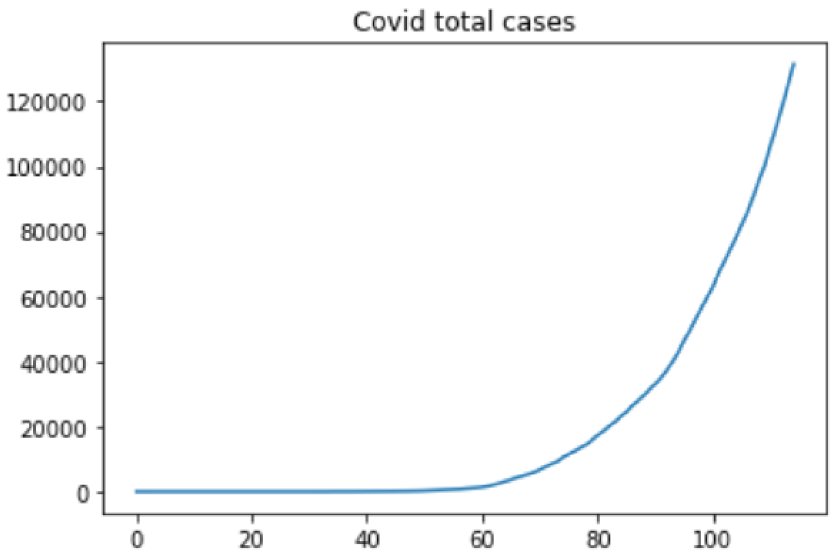


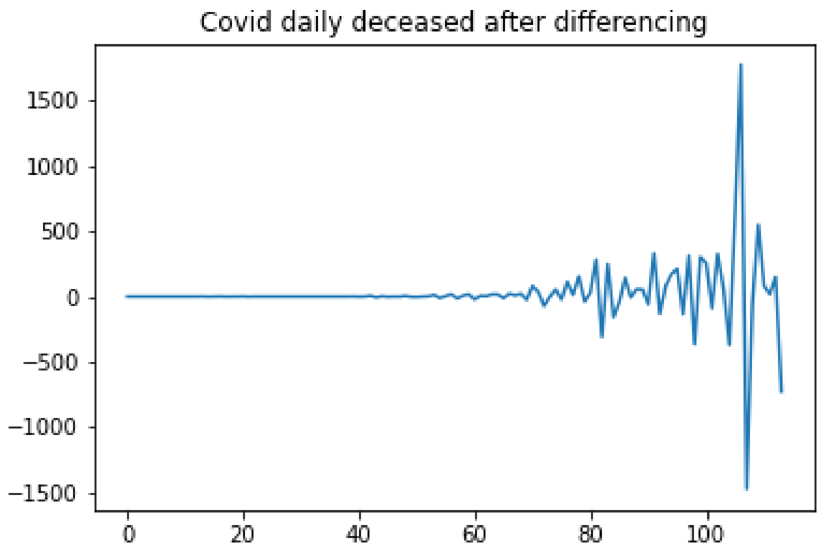
Null hypothesis could not be rejected. Covid daily cases is non-stationary.
p-Value is: 1.0
Differencing didn't work on Covid daily cases
p-Value is: 0.8984569947893335



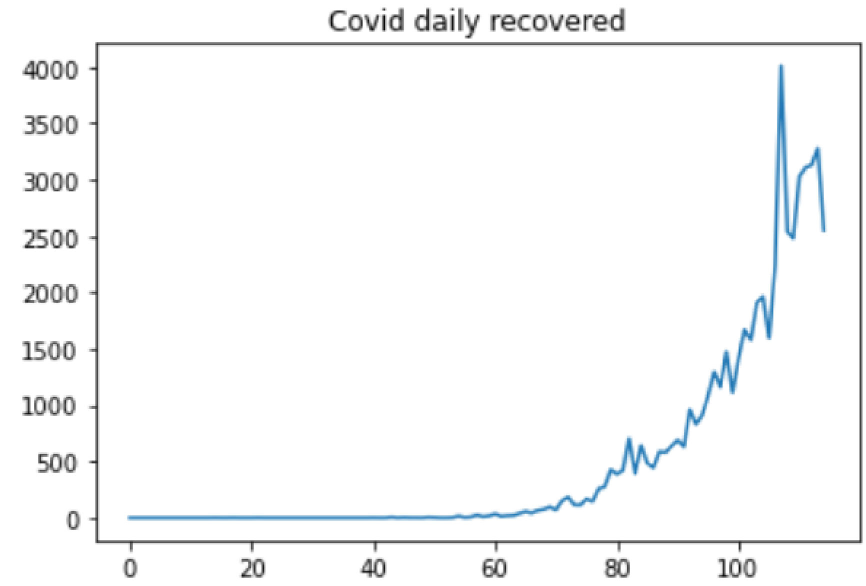


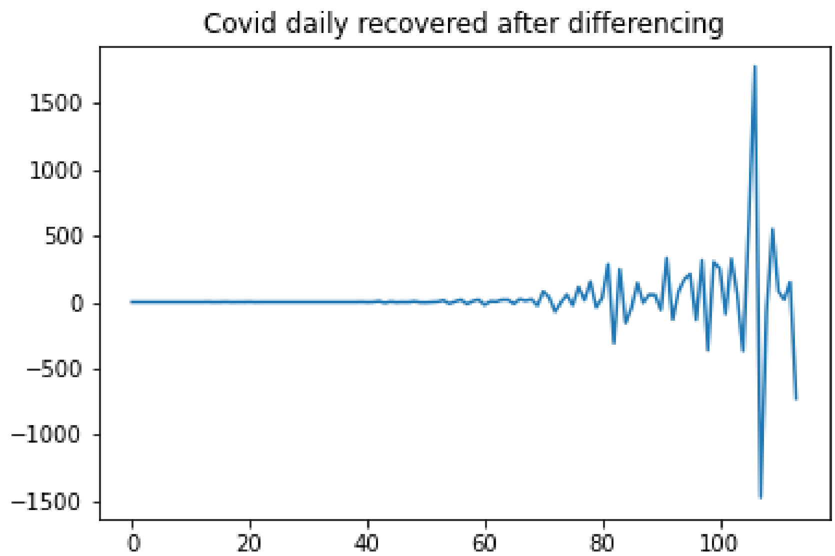
Null hypothesis could not be rejected. Covid total cases is non-stationary.
p-Value is: 0.5896105084111981
Differencing didn't work on Covid total cases
p-Value is: 1.0



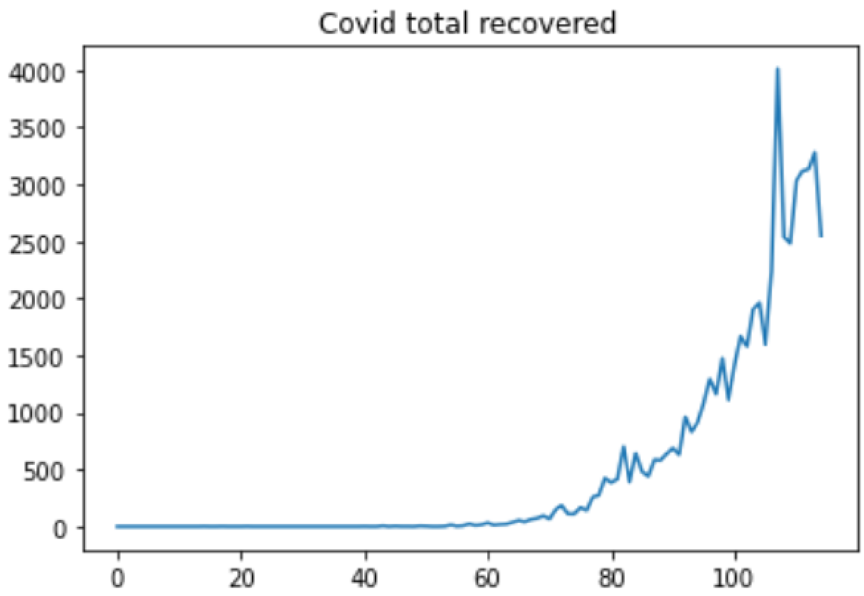


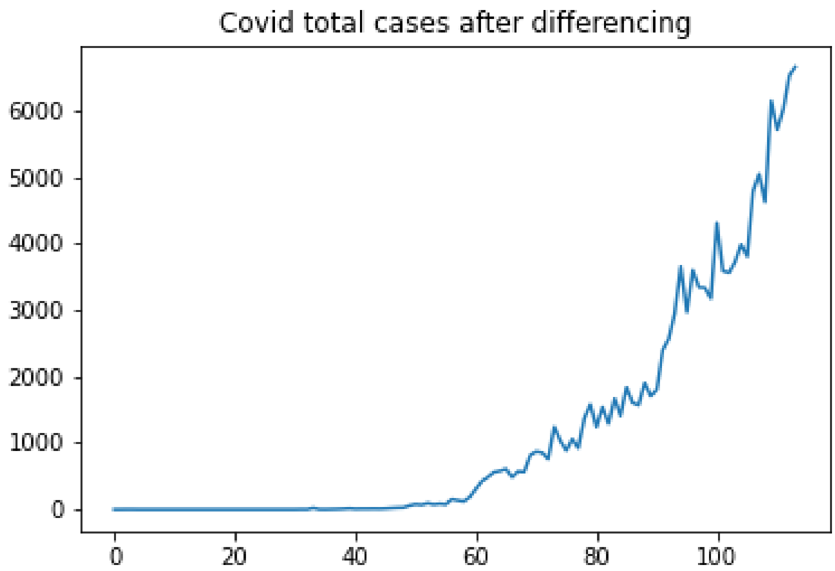
Null hypothesis could not be rejected. Covid daily recovered is non-stationary.
p-Value is: 1.0
Differencing didn't work on Covid daily recovered
p-Value is: 0.9946503715930034



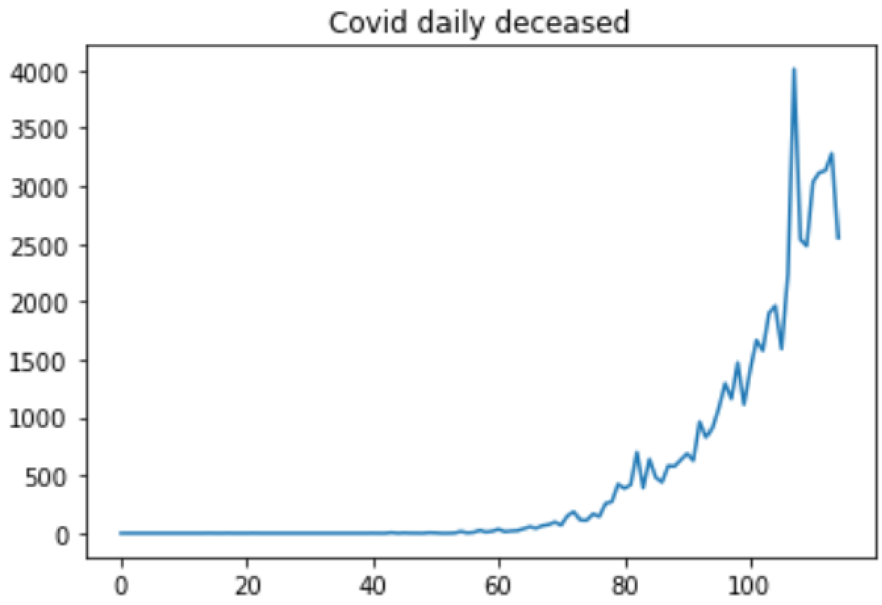


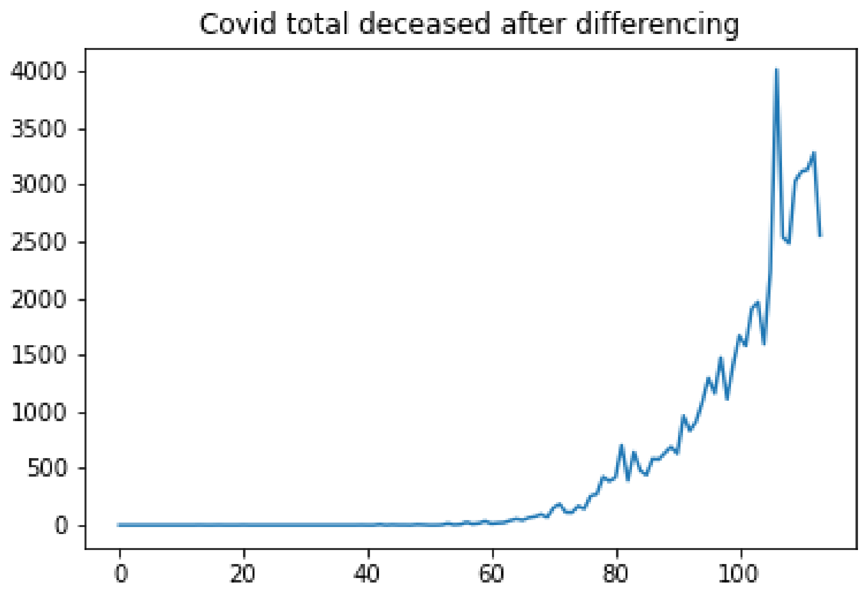
Null hypothesis could not be rejected. Covid total recovered is non-stationary.
p-Value is: 1.0
Differencing didn't work on Covid total recovered
p-Value is: 0.9946503715930034



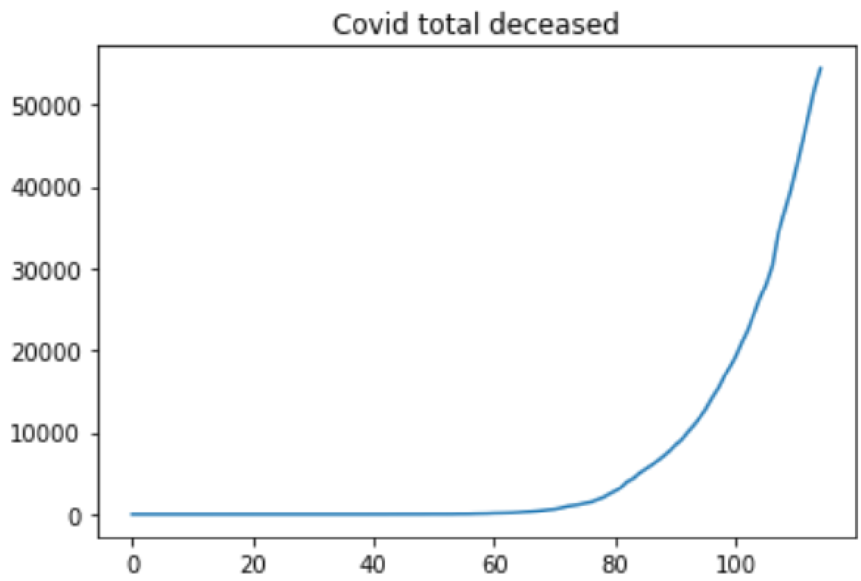


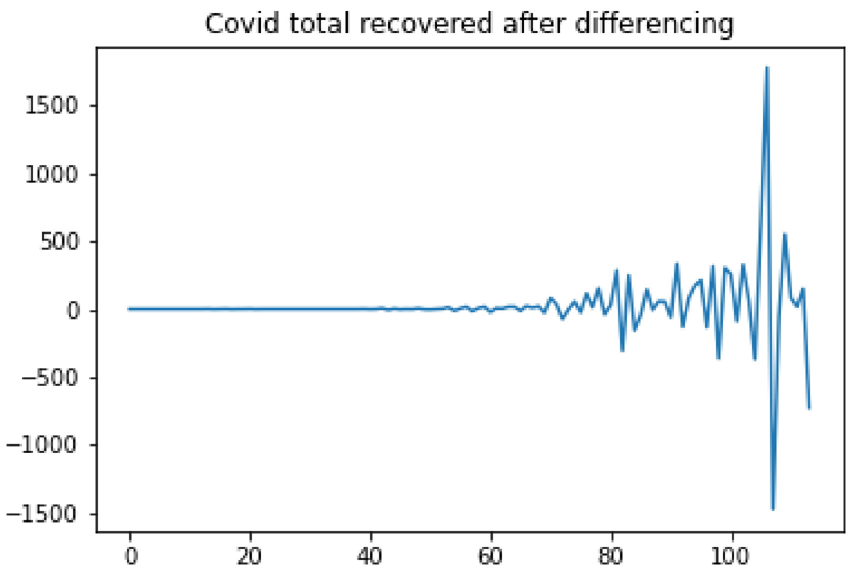
Null hypothesis could not be rejected. Covid daily deceased is non-stationary.
p-Value is: 1.0
Differencing didn't work on Covid daily deceased
p-Value is: 0.9946503715930034





Null hypothesis could not be rejected. Covid total deceased is non-stationary.
p-Value is: 0.570203611362623
Differencing didn't work on Covid total deceased
p-Value is: 1.0





- **Acknowledgements:**
    1. [statsmodels.tsa.stattools.adfuller — statsmodels]
    2. [8.1 Stationarity and differencing | Forecasting: Principles and Practice (2nd ed) (otexts.com)]