Assignment 1

Instructor:Dr.Amit Shukla

COURSE CODE:DAASE

Tiasha Biswas, Roll No.-2203121012

Question

(a) Whether these two-time series (TS) are stationarity or not? (b) If these TS are not stationarity, then discuss the method to convert them to stationarity TS and convert them to stationarity TS.
Please use COVID 19 data and NGC 1275 data.

Theory

Time series Analysis:
It is the study of a sequence of data points indexed in time order.These data points typically consist of successive measurements made from the same source over a fixed time interval and are used to track change over time.Time series can be stationary or non stationary.
Stationary time series: A stationary time series has statistical properties (mean ,variance(square of standard deviation) and auto covariance) invariant with time. It does not have trend and seasonality(including time varying seasonality).
Non Stationary time series: Conversely, nonstationarity is the status of a time series whose statistical properties are changing through time.
Application:
1.Time series data is used to predict future data values with the help of previous data.
2.It helps to forecast a certain quantity in the future by analyzing it's previous data, observing the previous trend, analyzing the past behavior, etc.
3.It helps to evaluate the current situation.
Stationary series is easier for statistical models to predict effectively and precisely.Thus it is necessary to convert non stationary time series into stationary.
Tests for checking Stationarity of a Time Series:
1.**Rolling Statistics Test:**The Rolling statistics test gives the visual representation of the dataset.A Moving variance or moving average graph is plotted and then it is observed whether it varies with time or not. In this method, a moving window of time is taken (based on our needs, for eg-10, 12, etc.) and then the mean of that time period is calculated as the current value.
2.**Augmented Dickey Fuller Test:**In this method, we take a null hypothesis that the data is non-stationary. After executing this test, it will give some results comprised of test statistics and some other critical values that help to define the stationarity. If the test statistic is less than the critical value and p value is(≤ 0.5) then we can reject the null hypothesis and say that the series is stationary.

Covid -19 data

Working
At first, data corresponding to no. of patients affected per day is stored in "ydata" and "xdata" contains the dates.Then an array named x which contains the no. of days is initialised.Next,ydata and x is plotted .Using rolling statistics, rolling mean and rolling standard deviation of ydata are evaluated and plotted.Next,for converting this series into stationary one ,at first "differencing method" is employed on ydata .In this method ,datas are transformed by this formula $y(t) = y(t) - y(t-1)$.Then this new values of ydata are stored in numpy array.Then it is converted into pandas series as y1.Then y1 and x are plotted.To check whether the new series is stationary or not both rolling statistics and augmented dickey fuller test is used.Next,original series is modified by first taking cube root and then applying differencing technique on

ydata.Then the modified values are plotted and the series is tested via both above metioned methods.
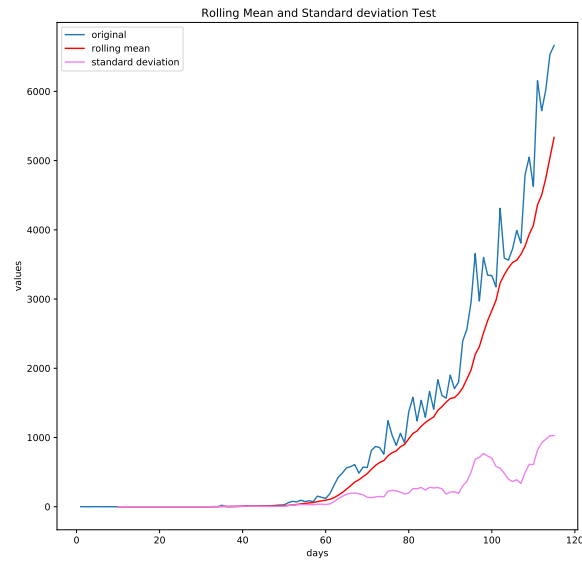
Results,Plots and Discussion

Figure 1: Original Series

Discussion: From the plot,it is observed that series has increasing trend and rolling mean is not stationary therefore the series is not stationary.
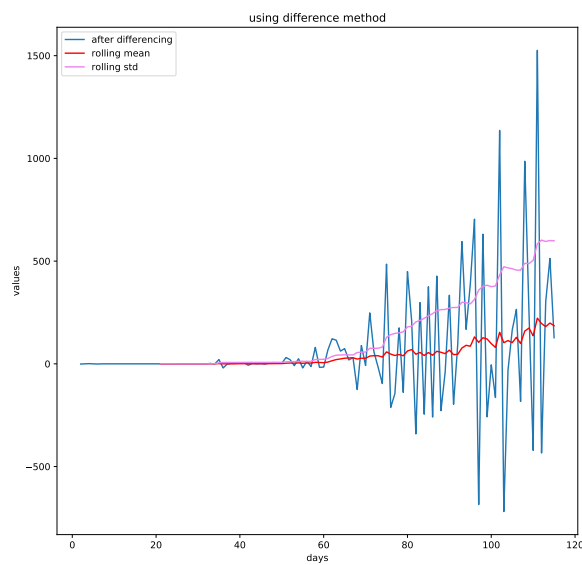**Result and discussion for Differencing method:**

Figure 2: Modified Series 1

When analysis of state of series becomes difficult through visualisation , analysis is then done via p value and critical and statistical values.

p value is : 0.8984569947893324

Test Statistic value: -0.4661368228976262

Critical Value (1%) : -3.4968181663902103

Critical Value (5%) : -2.8906107514600103

Critical Value (10%) : -2.5822770483285953

The test statistic value (-0.46) is not less than the critical values (-3.49 , -2.89 , -2.58) at different percentage and value of p is (>0.5) . Thus, modified series is not stationary.
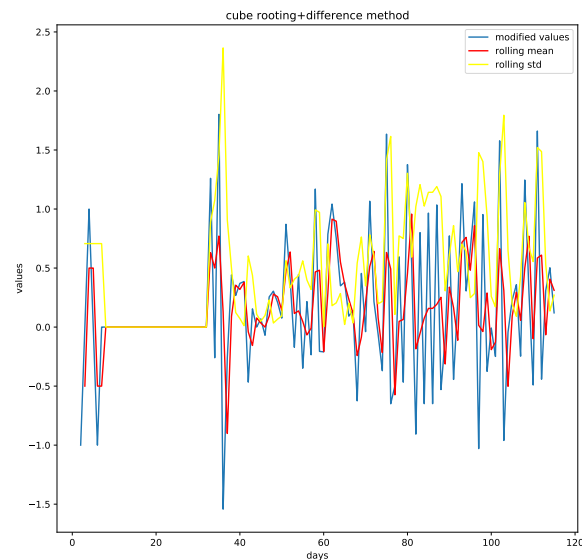
Result and discussion for next method:

Figure 3: Modified Series 2

p value is : $7.704595241392836 X 10^{-29}$

Test Statistic value: -15.932317377812089

Critical Value (1%) : -3.489589552580676

Critical Value (5%) : -2.887477210140433

Critical Value (10%) : -2.580604145195395

The test statistic value (-15.9) is less than the critical values (-3.489 , -2.88 , -2.58) at different percentage and value of p is (<0.5) . Thus, new modified series is stationary.

NGC 1275 data

Working

Similar working like that of previous data but in this case the NAN values are dropped and differencing method is used to convert the series into a stationary one.

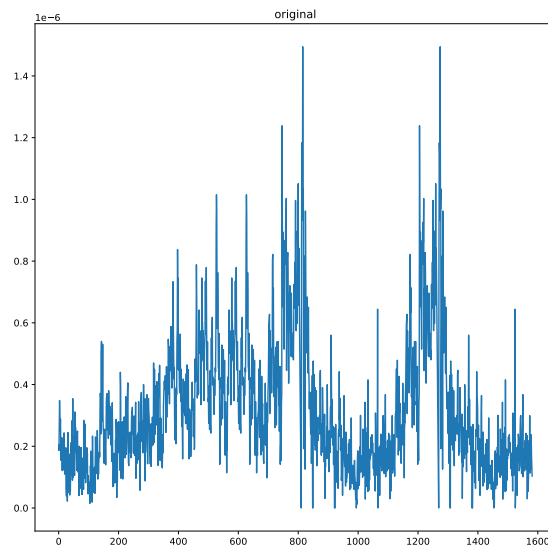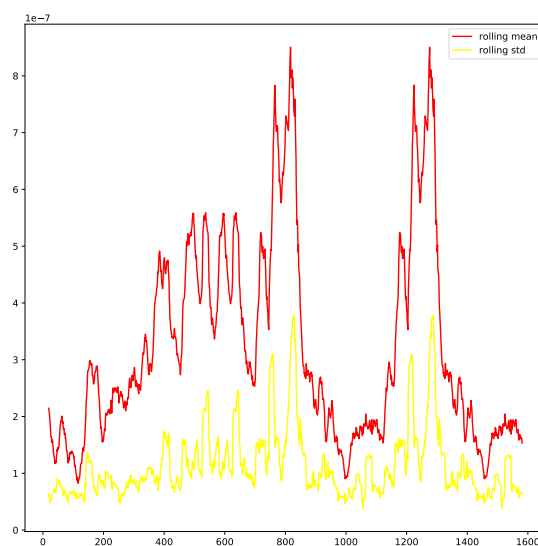**Results,Plots and Discussion Discussion**

Figure 4: Original

Figure 5: Rolling Tests on Original Series

From the graph, it is non stationary due to non uniform moving mean and standard deviation.
Discussion after modification:
It is difficult to conclude correctly about the stationarity of series through plots so augmented dickey fuller test is also performed.
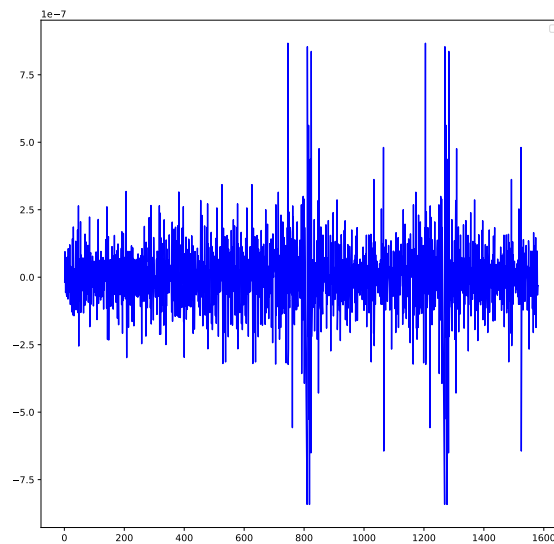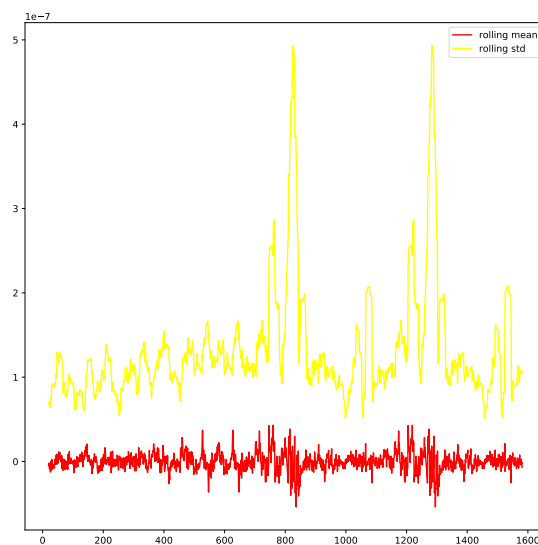

Figure 6: Modified data

Figure 7: Rolling tests on modified data

p value: 5.881295482197483x10-28

Test Statistic value: -15.198123948544005
Critical Value (1%) : -3.4345923608042512
Critical Value (5%) : -2.8634137428910185
Critical Value (10%) : -2.567767551360439
The test statistic value (-15.2) is less than the critical values (-3.43, -2.86 , -2.567) at different percentage but value of p is (<0.5) . Thus, new modified series is stationary.