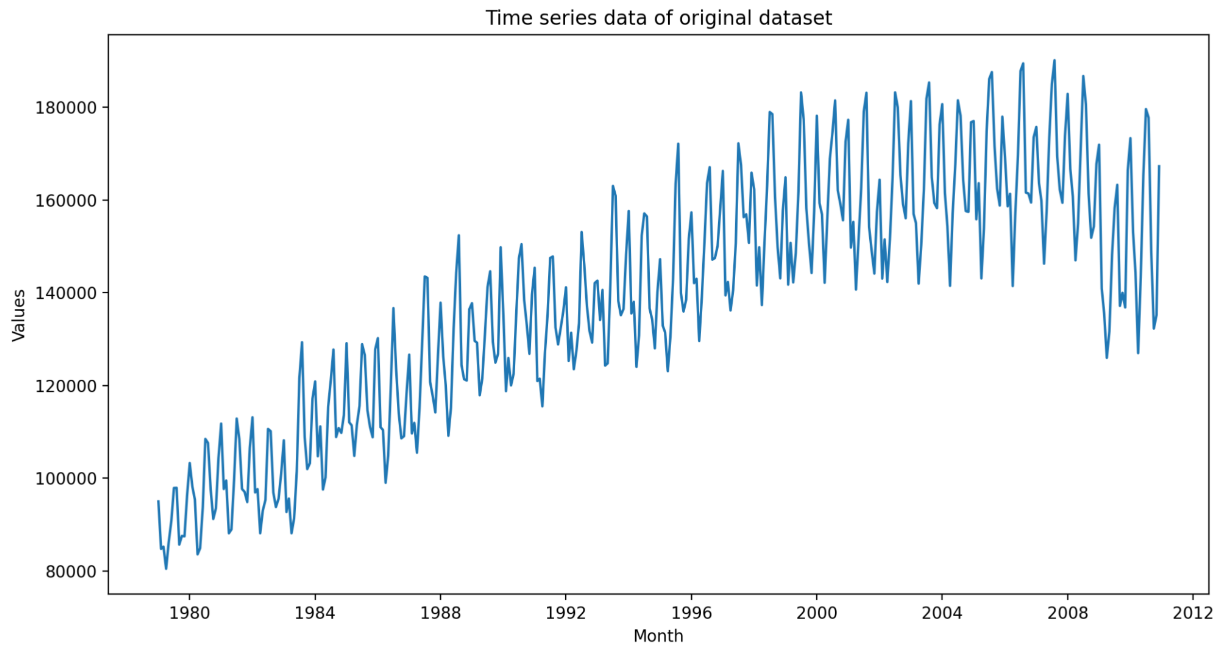
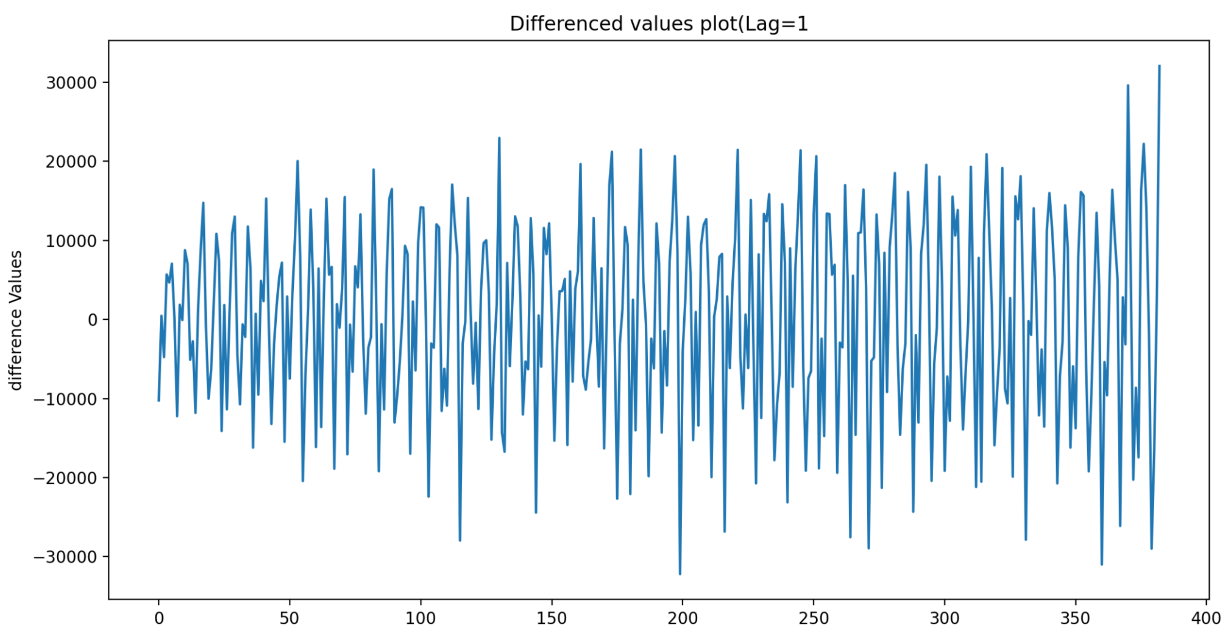
Q1. For electricity.csv dataset:

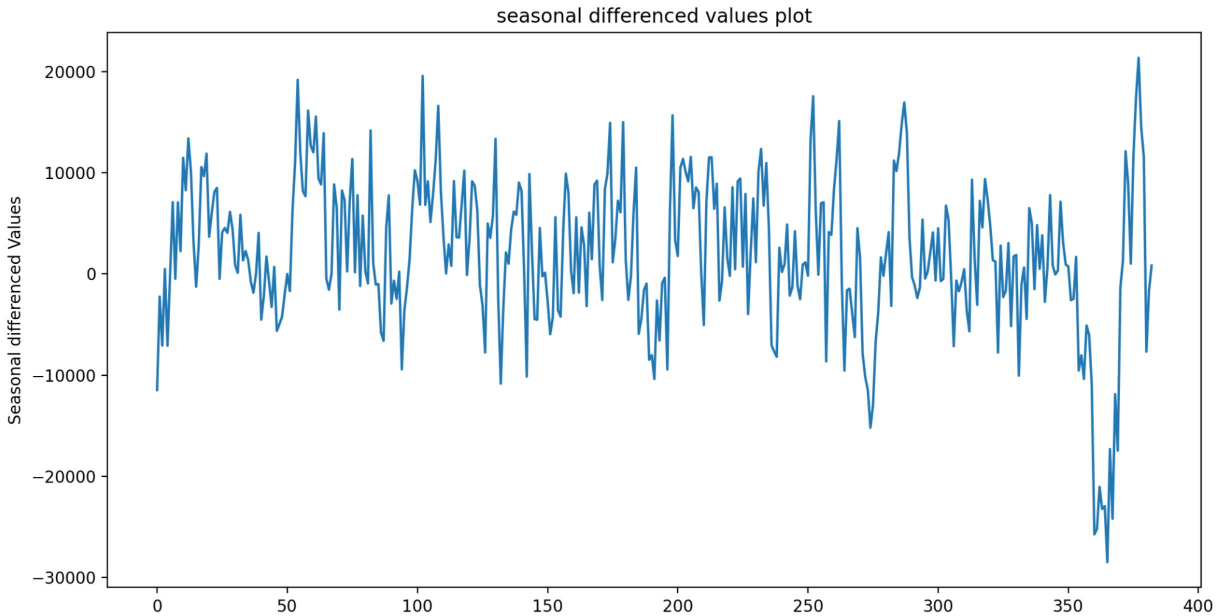
(i) the original time series of data



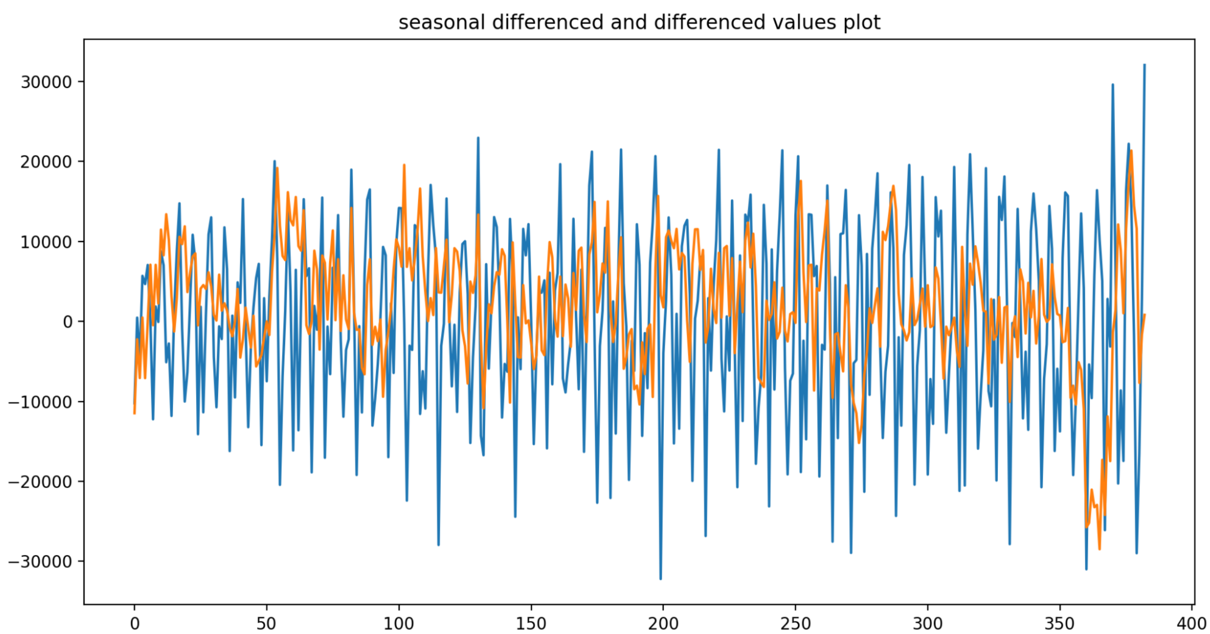
(ii) the differenced series of data



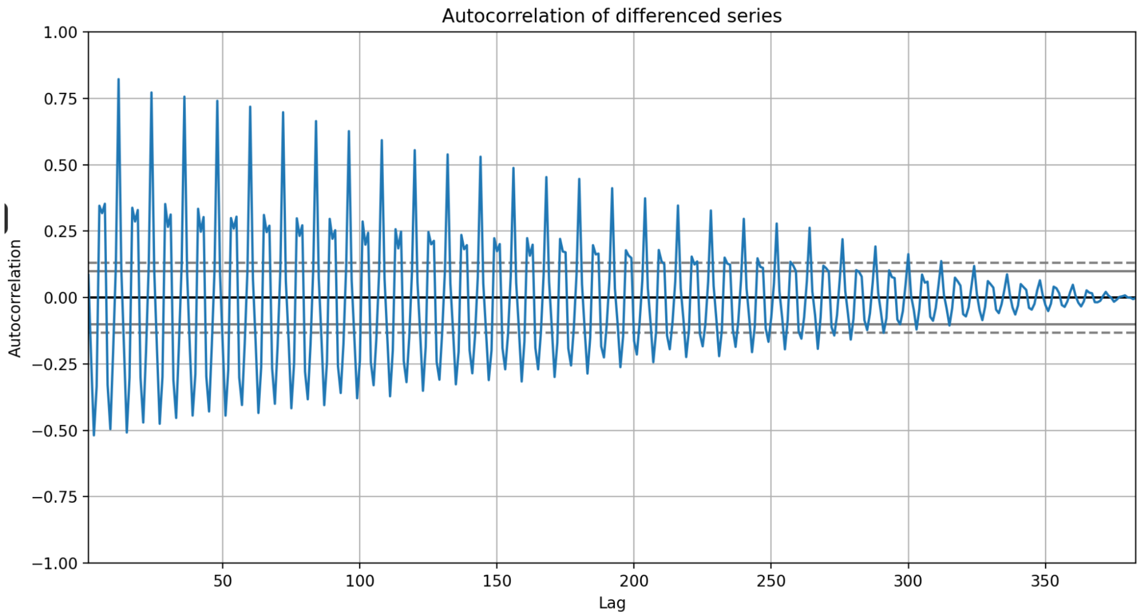
(iii) the seasonally differenced series of data

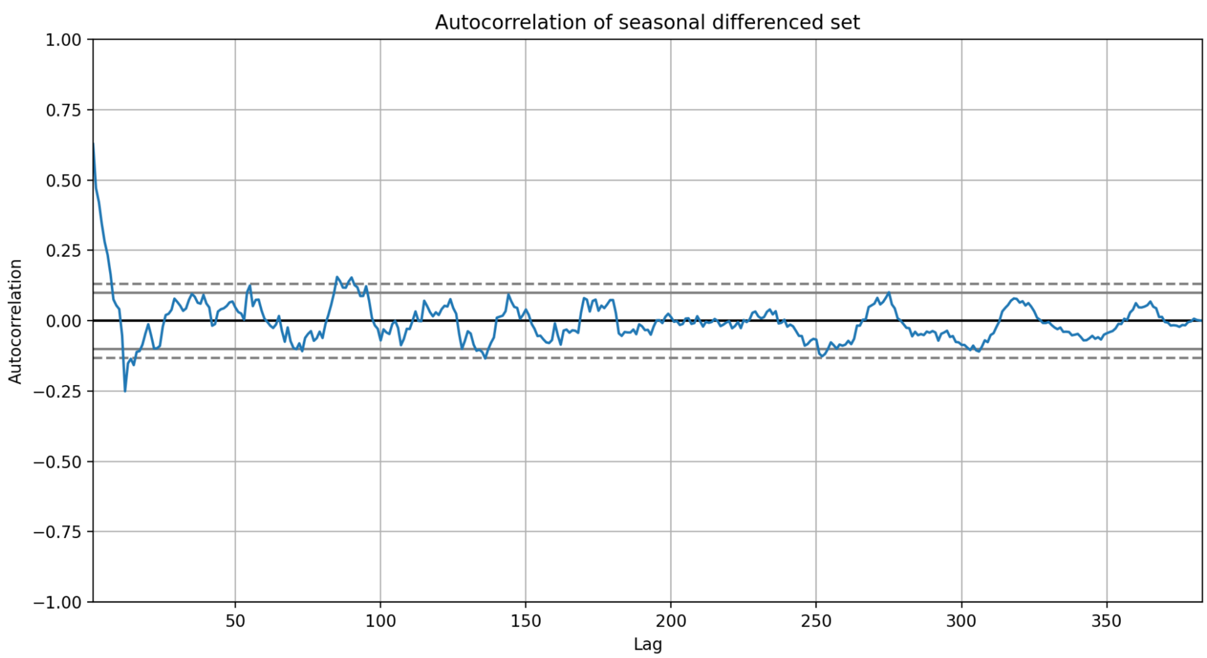


(iv) the differenced and seasonally differenced series of data



What kind of differencing is needed to make the time series stationary?





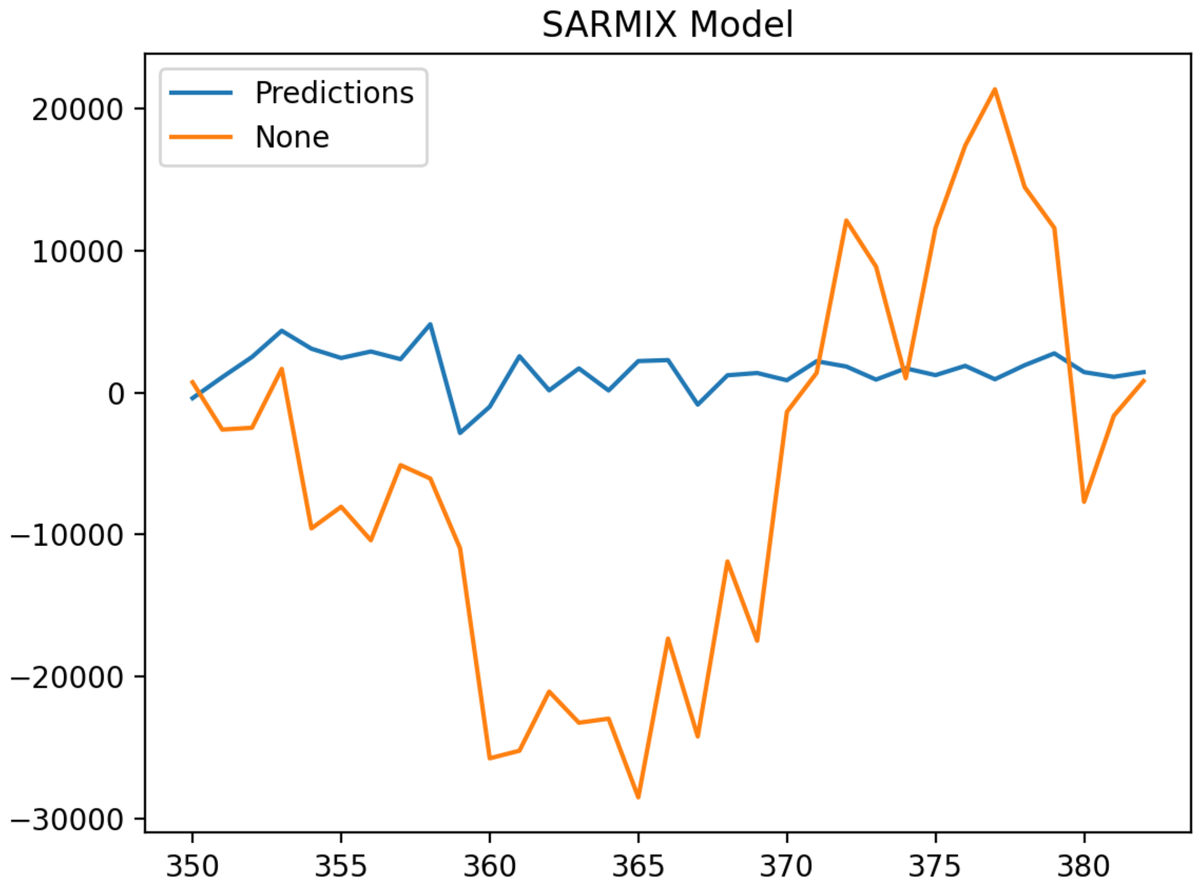
As we can see from the autocorrelation plot of differenced time series (lag=1), we can see that it is non stationary data. But in the autocorrelation plot of seasonal differenced time series data we can see that the data is stationary. So, seasonal differencing is needed for making time series stationary.

Examine plots of sample autocorrelation functions and partial autocorrelation functions for a seasonally differenced (and possibly further differenced) and fit the following types of ARIMA-and Seasonal ARIMA models to the time series of electricity consumption:

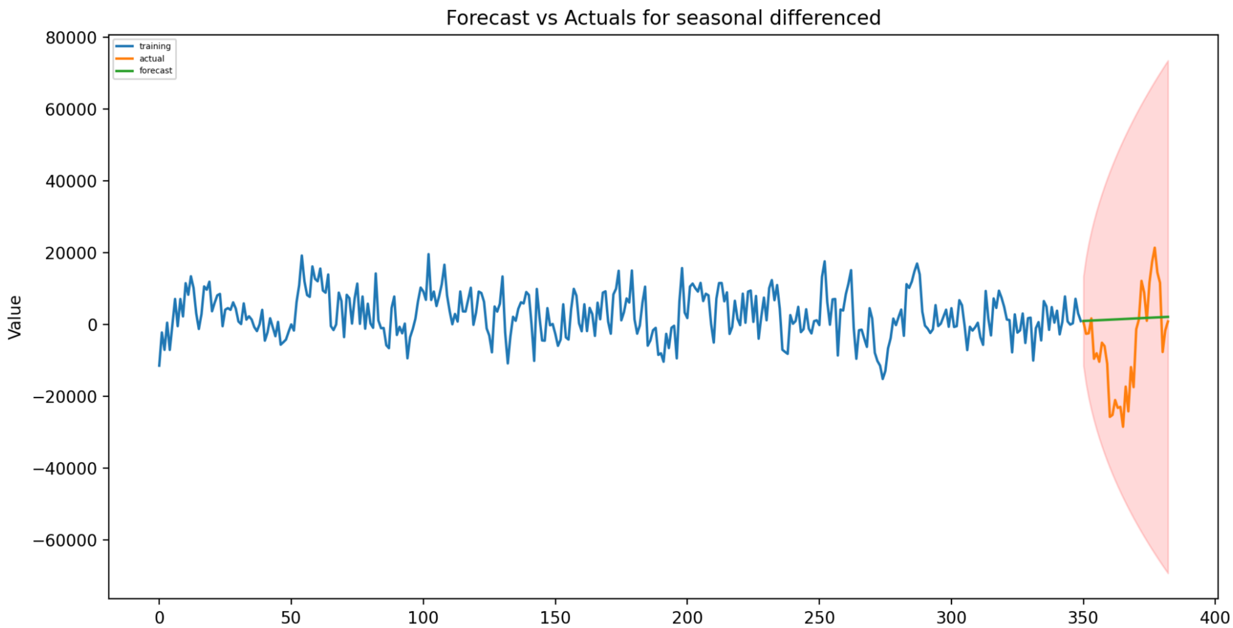
(i) an ordinary ARMA-model of a seasonally differenced (and possibly further differenced) series.

(ii) a seasonal ARMA-model of a seasonally differenced (and possibly further differenced) Series.

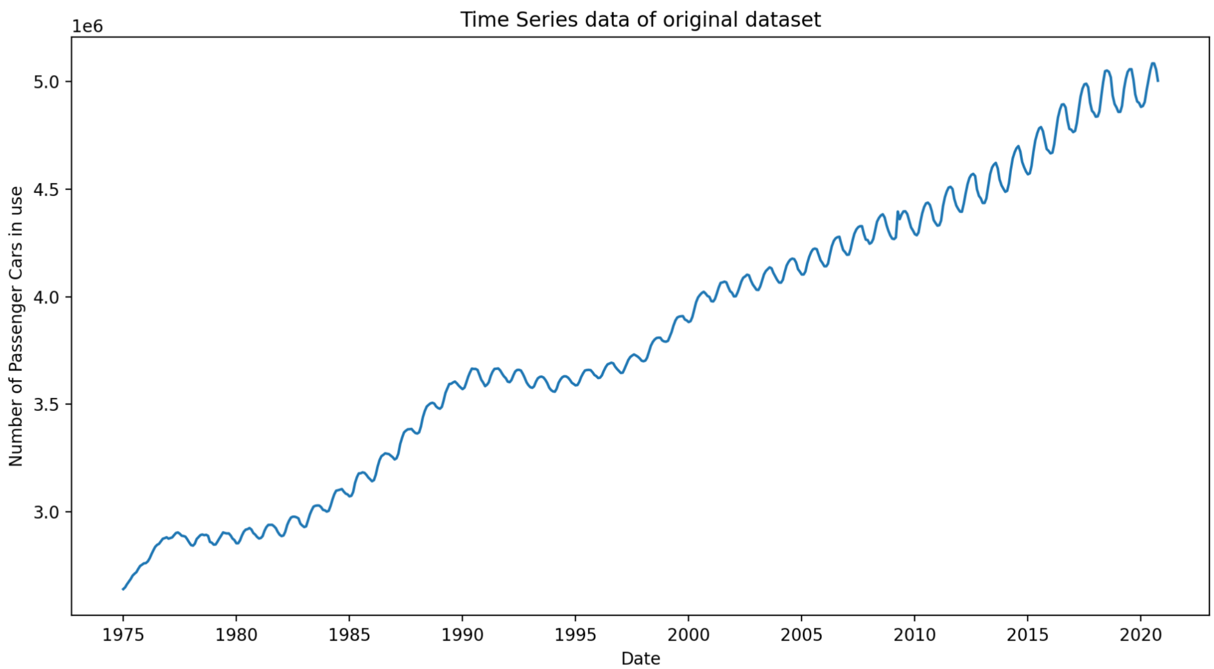
```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                  382
Model:                 ARIMA(0, 1, 0)   Log Likelihood               -4246.333
Method:                           css   S.D. of innovations          16270.602
Date:                Mon, 16 May 2022   AIC                           8496.665
Time:                        23:21:55   BIC                           8504.556
Sample:                             1   HQIC                          8499.796

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const        110.7616    832.476      0.133      0.894   -1520.862    1742.385
==============================================================================
                             ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                  382
Model:                 ARIMA(0, 1, 0)   Log Likelihood               -3893.705
Method:                           css   S.D. of innovations           6464.013
Date:                Mon, 16 May 2022   AIC                           7791.409
Time:                        23:21:55   BIC                           7799.300
Sample:                             1   HQIC                          7794.540

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         32.2154    330.728      0.097      0.922    -615.999     680.430
==============================================================================
```

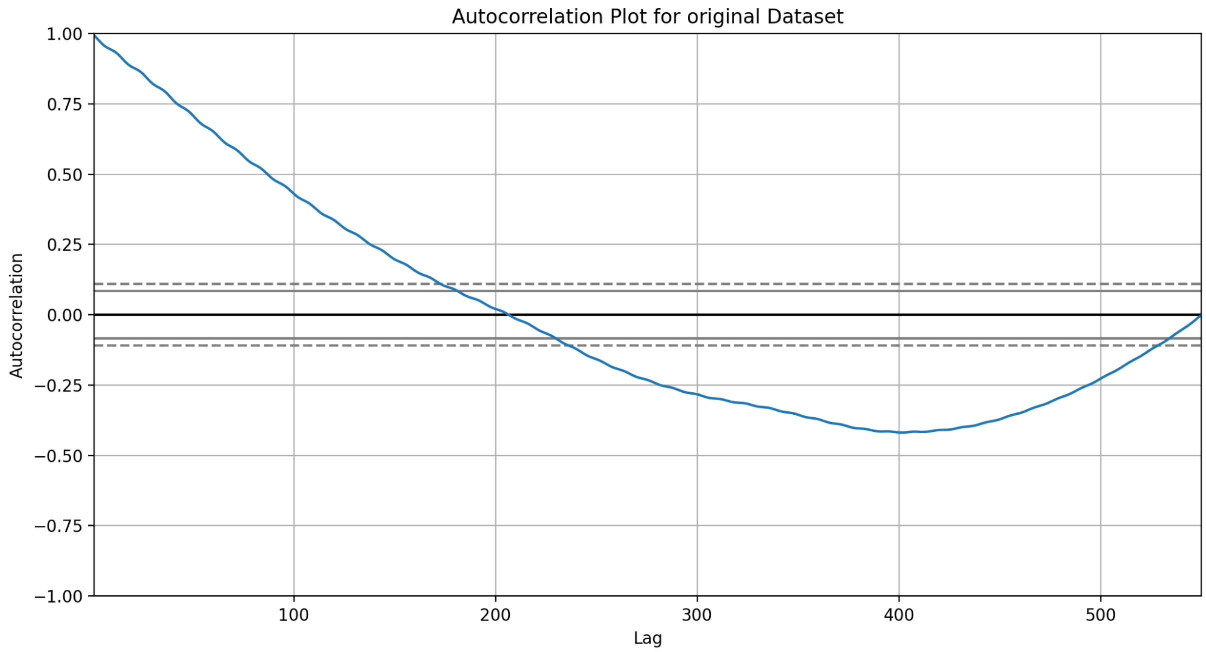


Compute forecasts of the electricity consumption for a period of 12 months and plot observed and forecasted values. Examine, by visual inspection, in what respect the two model classes above produce different forecasts.

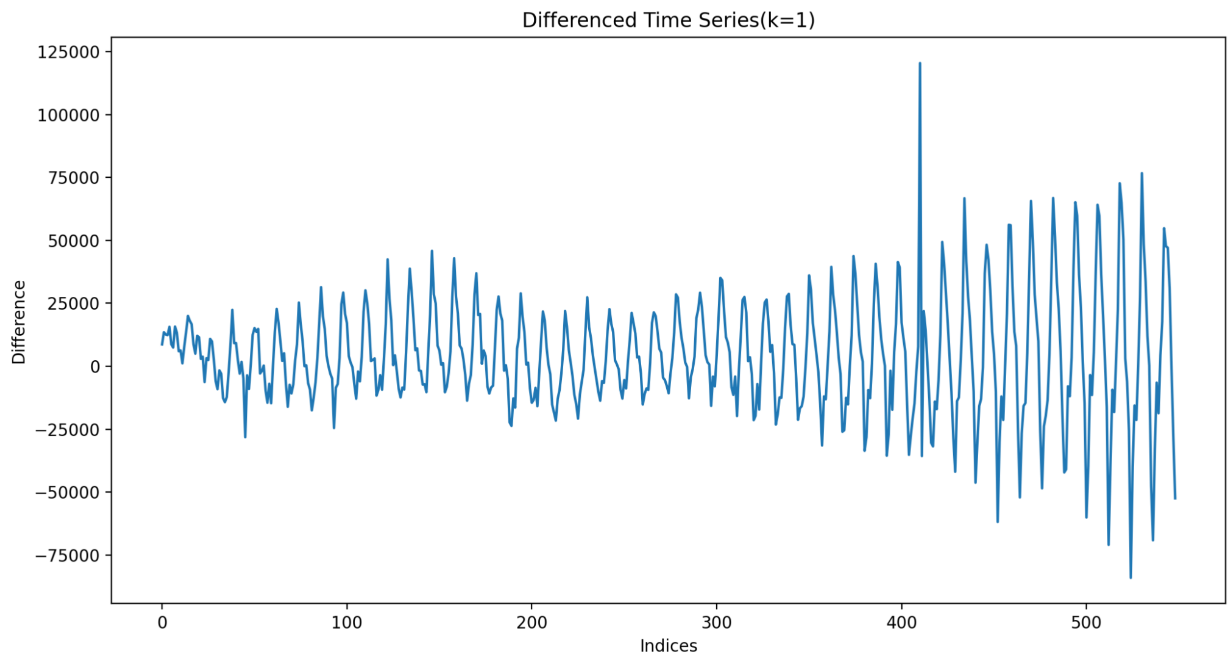

Forecast vs Actuals for seasonal differenced
training
actual
forecast
Value
80000
60000
40000
20000
0
−20000
−40000
−60000
0
50
100
150
200
250
300
350
400

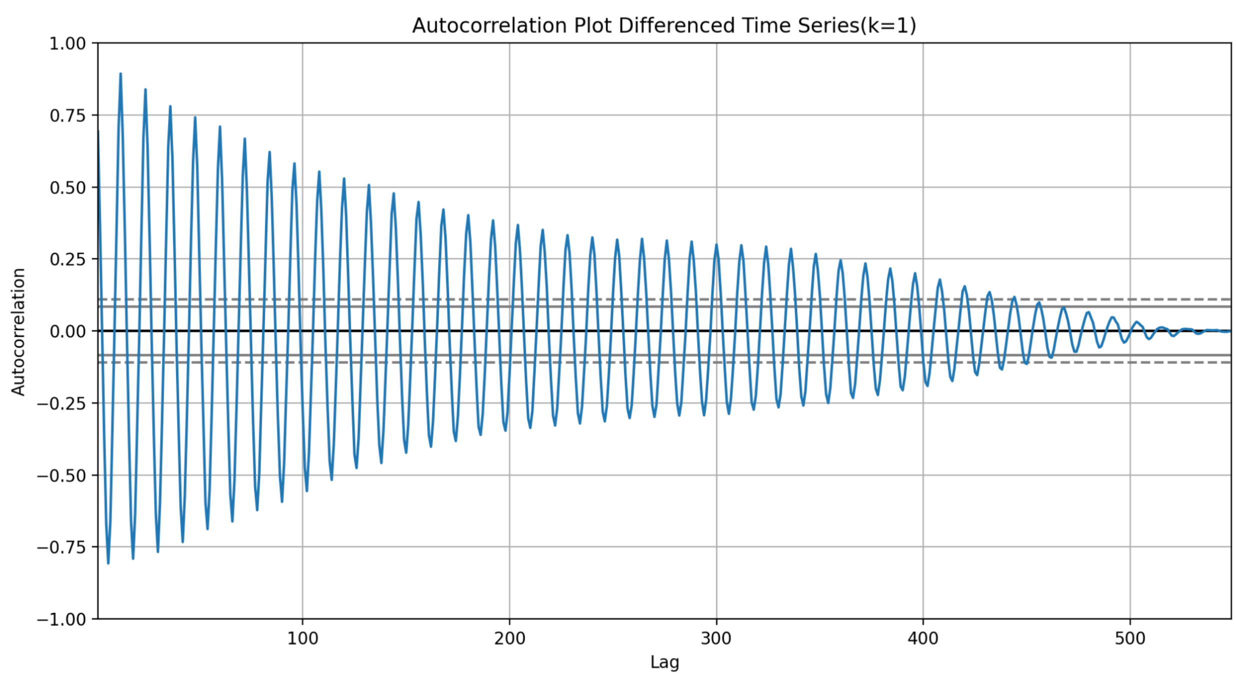
Q2. Make a time series plot of the number of cars in use by using Passenger cars dataset.
Carry out additional differencing and make time series plots of the differenced series.
What differencing operations are needed to make the series stationary?Compute sample autocorrelation functions for the original series as well as each of the
differenced series. Do the estimated autocorrelations indicate that you have obtained a stationary series after the last differencing? If not, try another type of differencing.



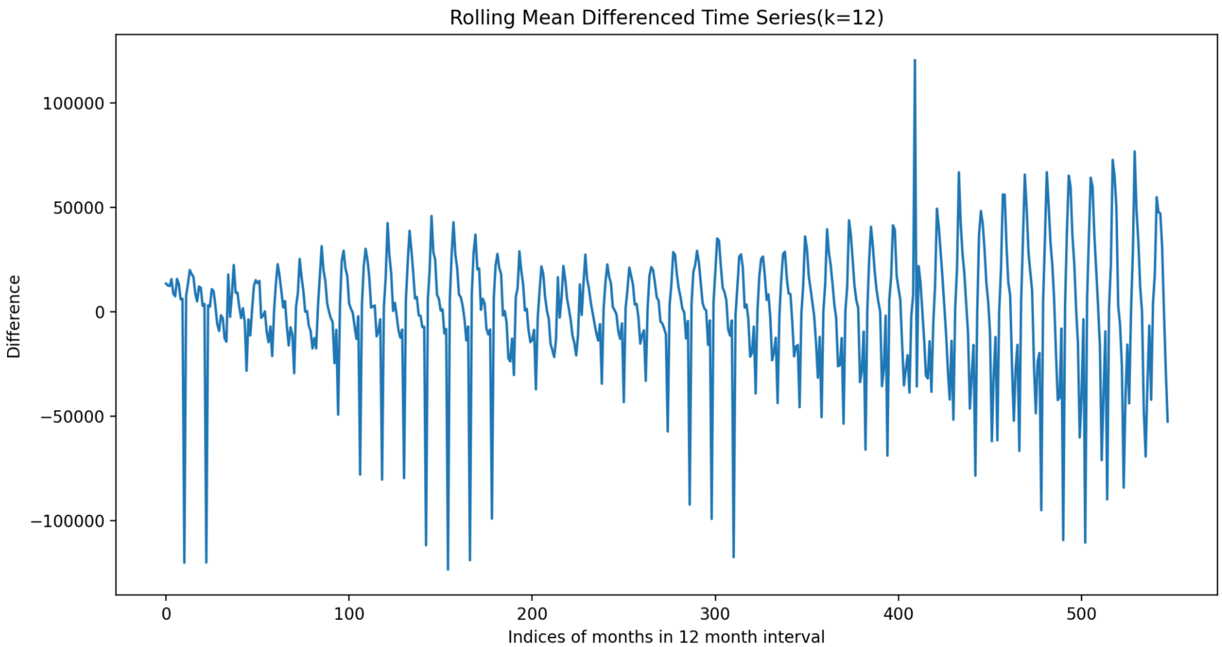

Autocorrelation Plot for original Dataset
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Autocorrelation
100
200
300
400
500
Lag

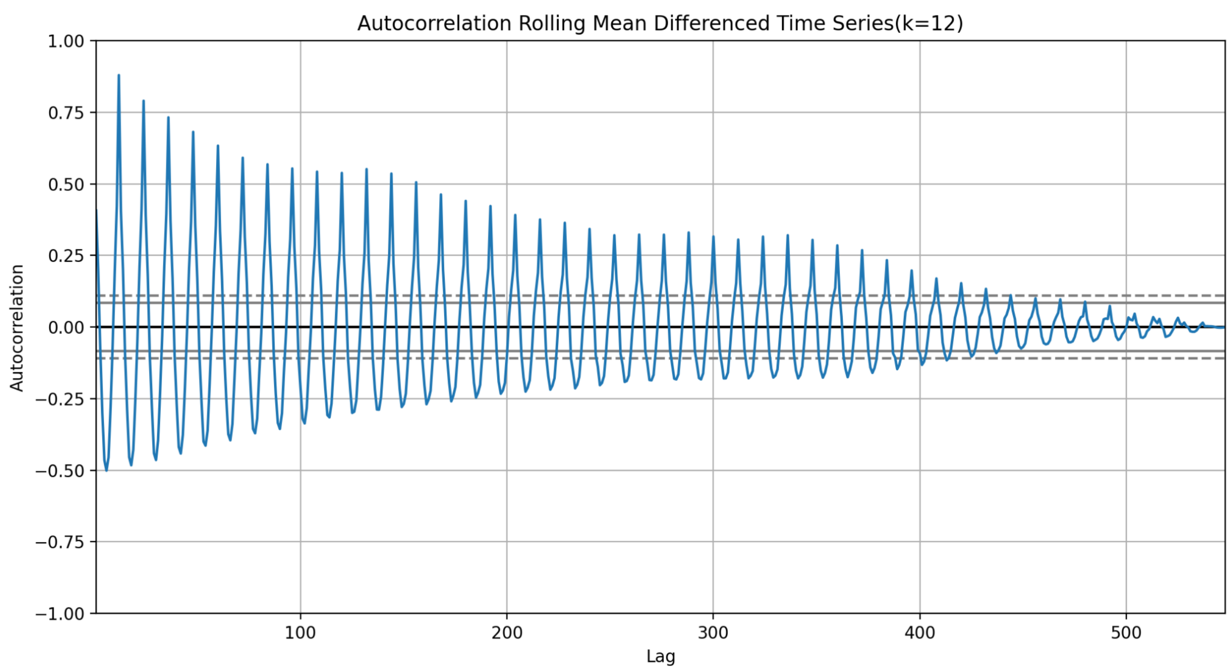
Additional differencing operations on given time series data and their corresponding autocorrelation plot:



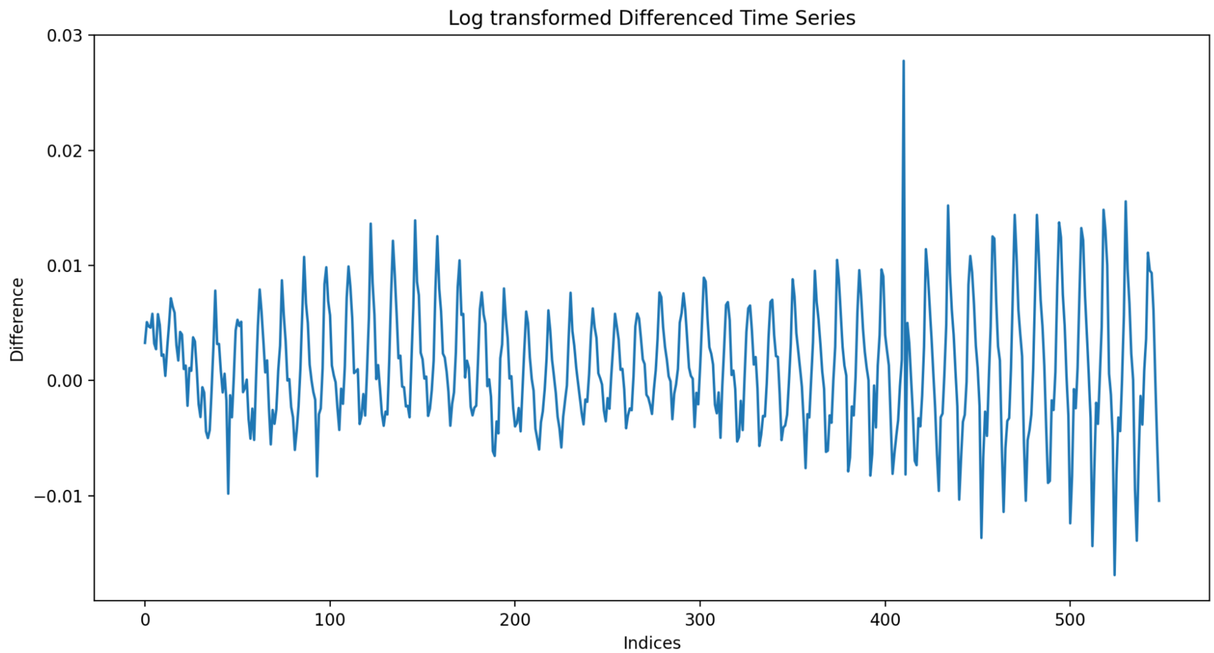


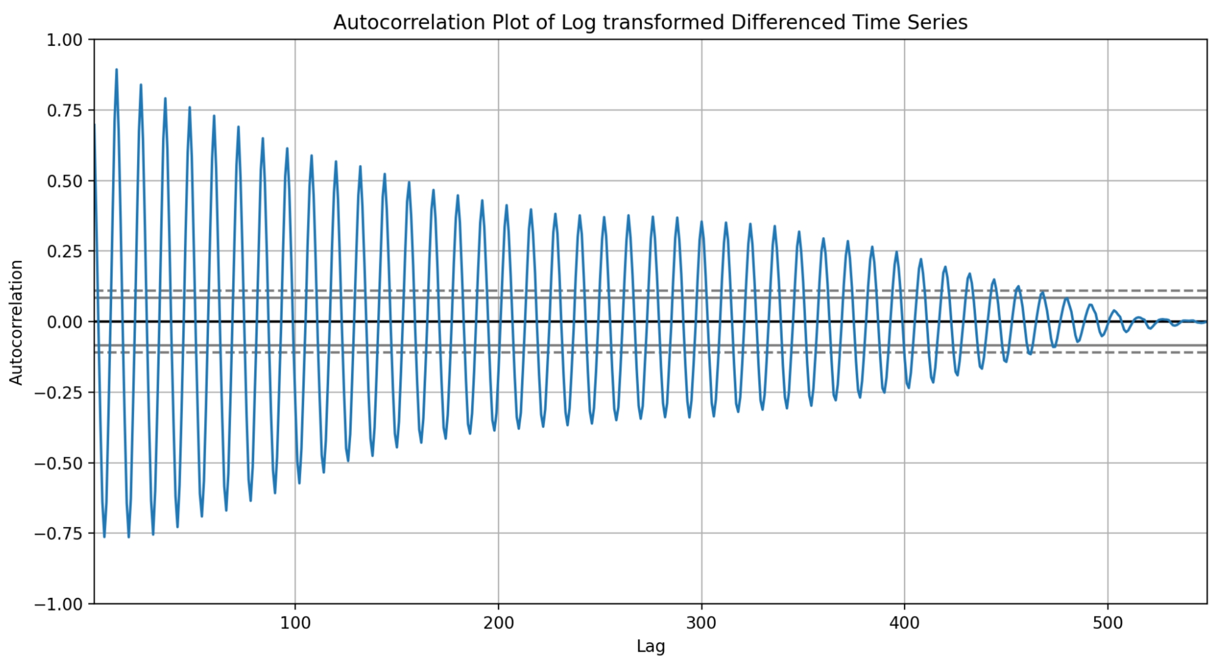
Additional differencing operations : Rolling mean (taking 12 as interval) on given time series data and their corresponding autocorrelation plot:

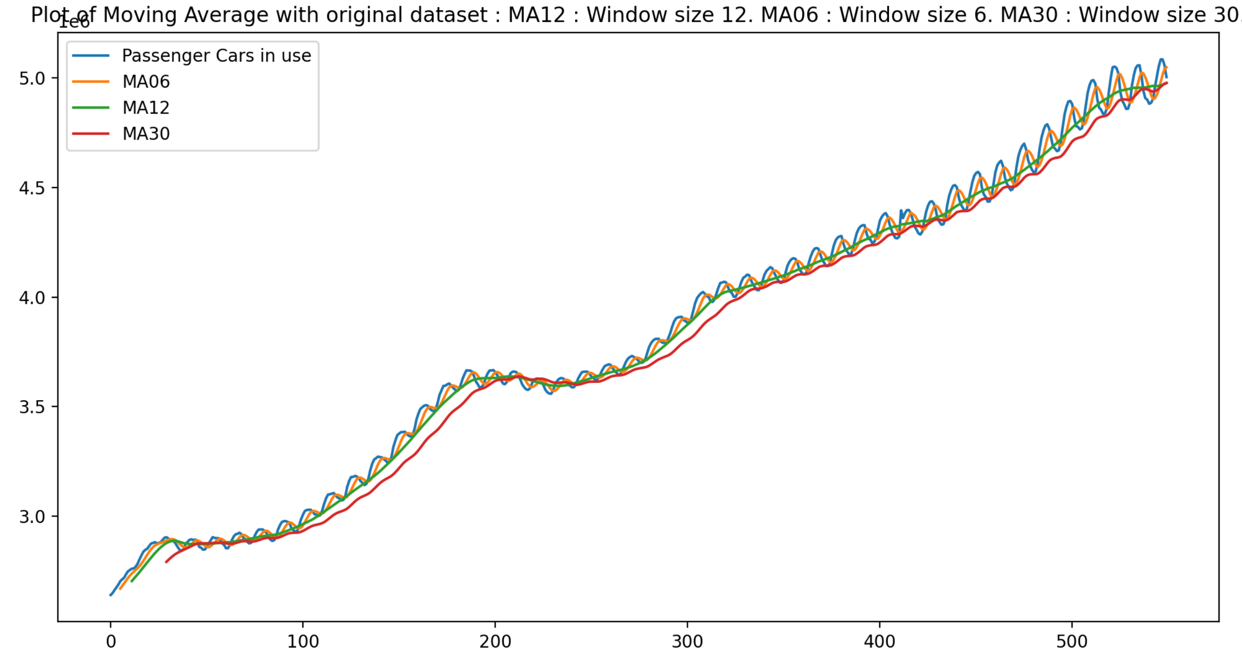


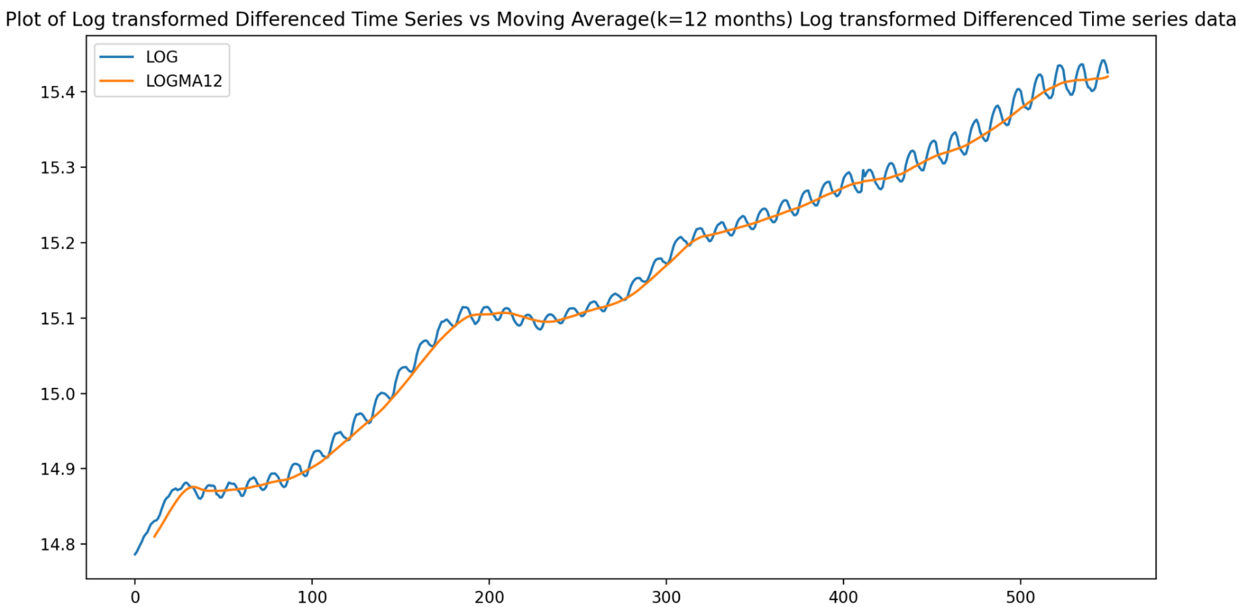


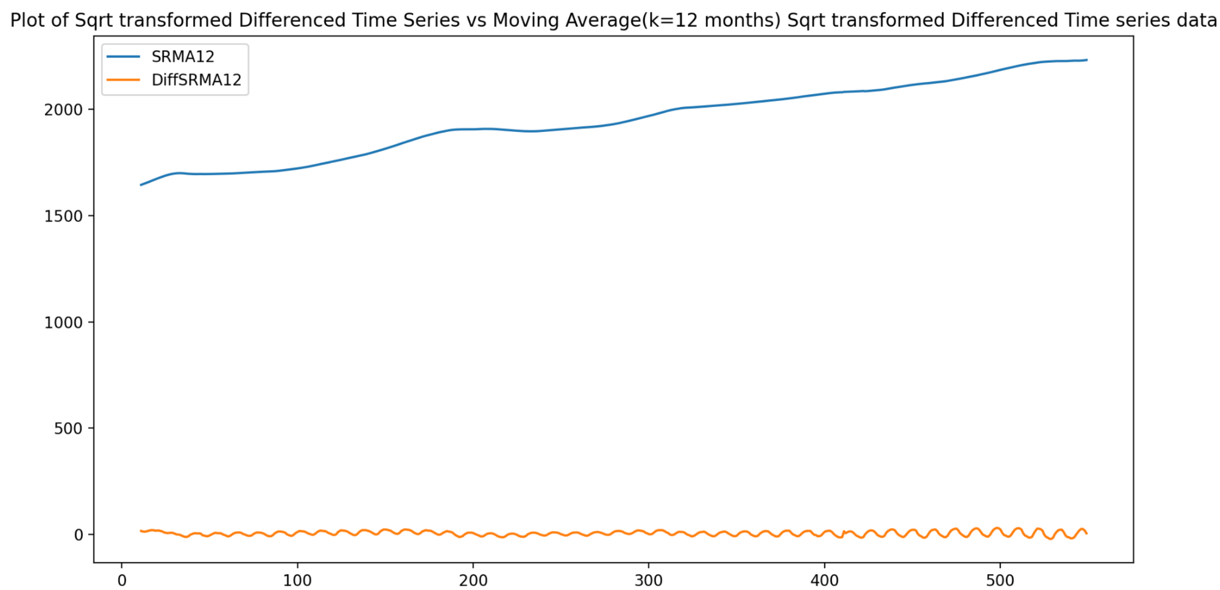
Additional differencing operations : Logarithmic difference on given time series data and their corresponding autocorrelation plot:



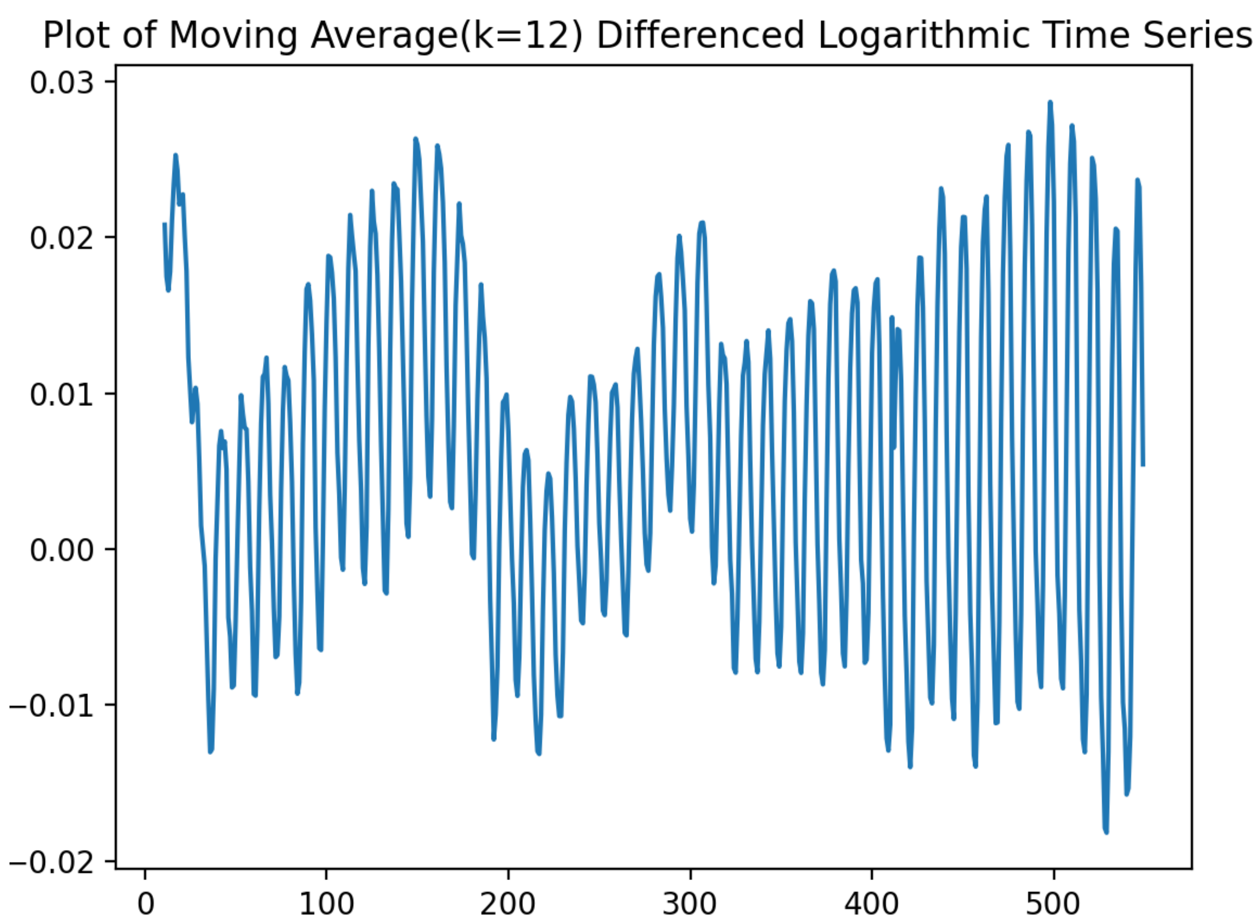


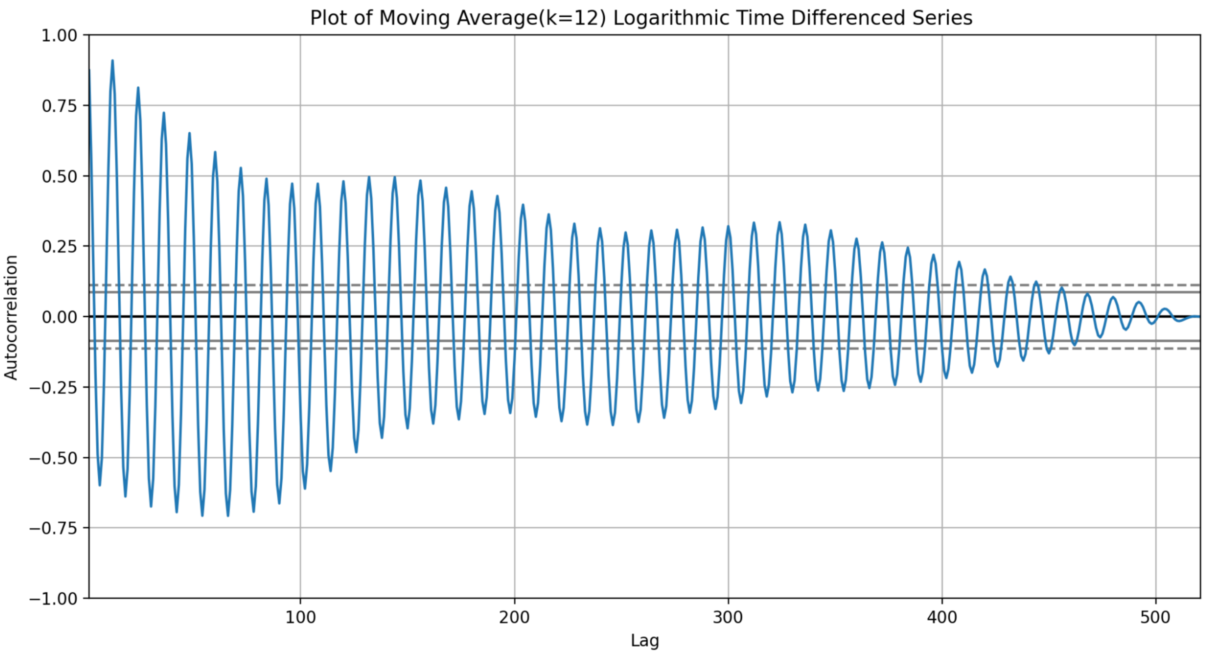

Plot of Moving Average with original dataset : MA12 : Window size 12. MA06 : Window size 6. MA30 : Window size 30.
1e6
Passenger Cars in use
MA06
MA12
MA30
5.0
4.5
4.0
3.5
3.0
0
100
200
300
400
500


Plot of Log transformed Differenced Time Series vs Moving Average(k=12 months) Log transformed Differenced Time series data
LOG
LOGMA12
15.4
15.3
15.2
15.1
15.0
14.9
14.8
0
100
200
300
400
500



Additional differencing operations : Logarithmic difference and applying moving average(k=12) on given time series data and their corresponding autocorrelation plot:






```
Fuller test hypothesis

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
```

```
Fuller test on original sataset:
ADF Statistic: 0.756482
p-value: 0.990900
Critical Values:
	1%: -3.443
	5%: -2.867
	10%: -2.570
Fuller test on Moving Average(k=12) dataset:
ADF Statistic: -3.806840
p-value: 0.002838
Critical Values:
	1%: -3.443
	5%: -2.867
	10%: -2.570
Fuller test on Moving Average of Logarithmic time series dataset:
ADF Statistic: -3.654957
p-value: 0.004793
Critical Values:
	1%: -3.443
	5%: -2.867
	10%: -2.570
Fuller test on Moving Average of Square root time series dataset: dataset:
ADF Statistic: -3.766447
p-value: 0.003270
Critical Values:
	1%: -3.443
	5%: -2.867
	10%: -2.570
```

Fig : 1.0

We have seen in the auto-correlation plot that in original dataset the correlation is depended on lag in a definitive order. And when we perform the Dickey-fuller test on the original dataset it can be clearly seen that p-value is 0.990900. Which implies that it is non-stationary data.

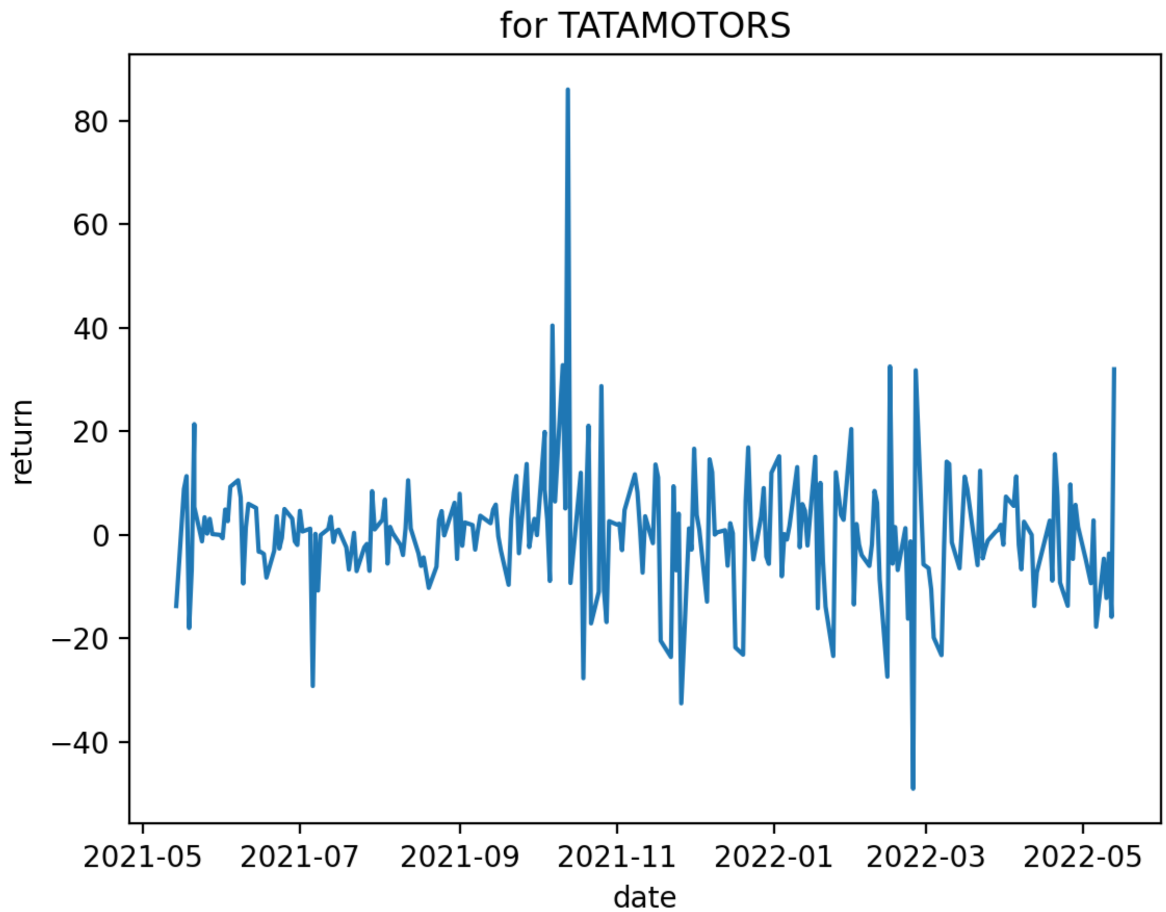
Now we have used following differenced data series models to remove stationarity:

1. Simple Time Difference model : where we find difference between data at time=t and time=t-1
2. Rolling Mean Difference model : where we take mean of 12 datasets and replace the original data with the difference between original data and the average of this 12 datasets.
3. Logarithmic Difference Time series model: where we convert the original data into logarithmic series and apply simple time difference, that is find difference between data at instance t and data with lag=1

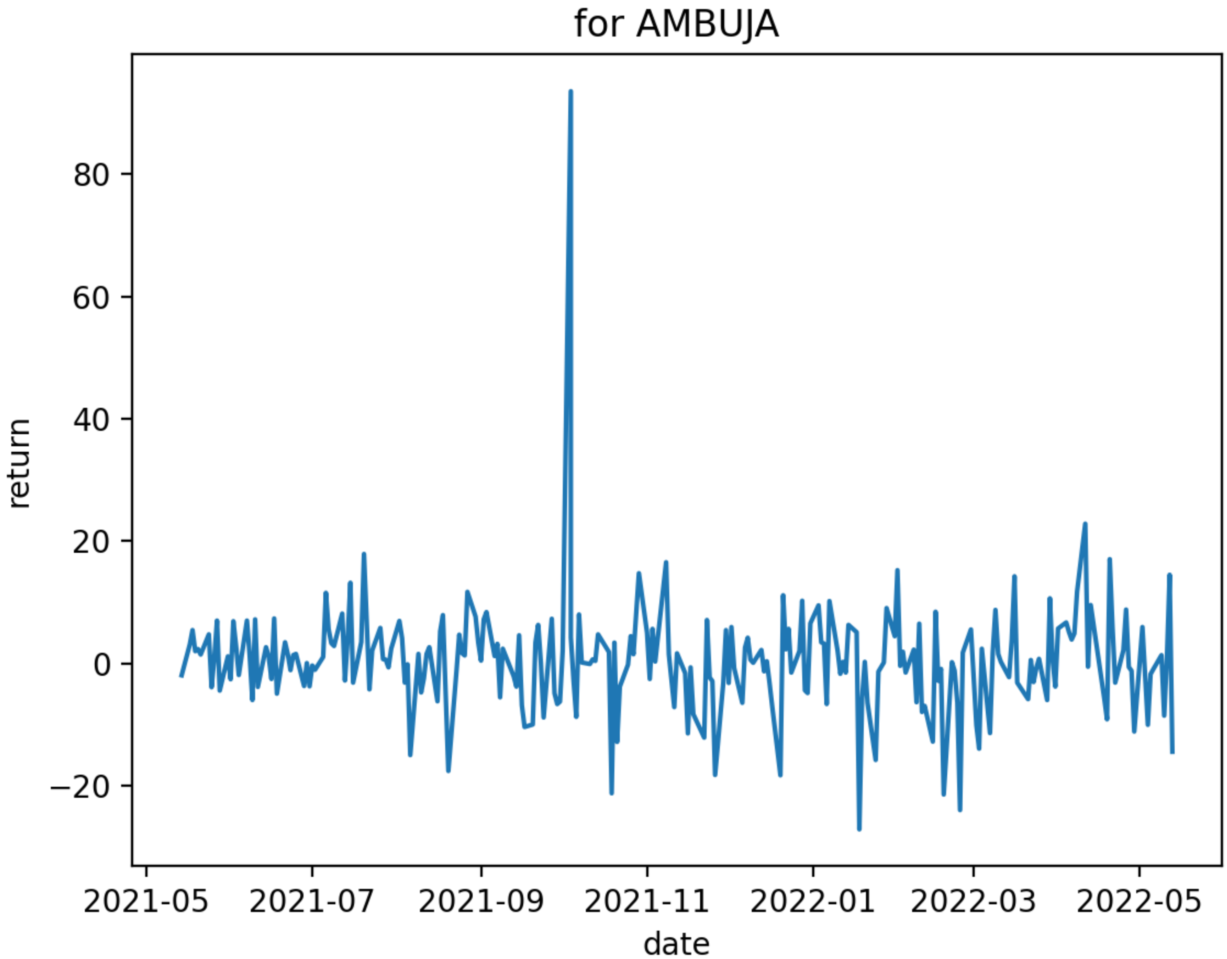
4. Logarithmic difference and Moving Average: We can apply more than one transformation as well. We'll first apply log transformation to time-series, then take a rolling mean over a period of 12 months and then subtract rolled time-series from log-transformed time-series to get final time-series.
5. Square Root Differenced and Moving Average: We'll first apply square root transformation to time-series, then take a rolling mean over a period of 12 months and then subtract rolled time-series from power-transformed time-series to get final time-series.

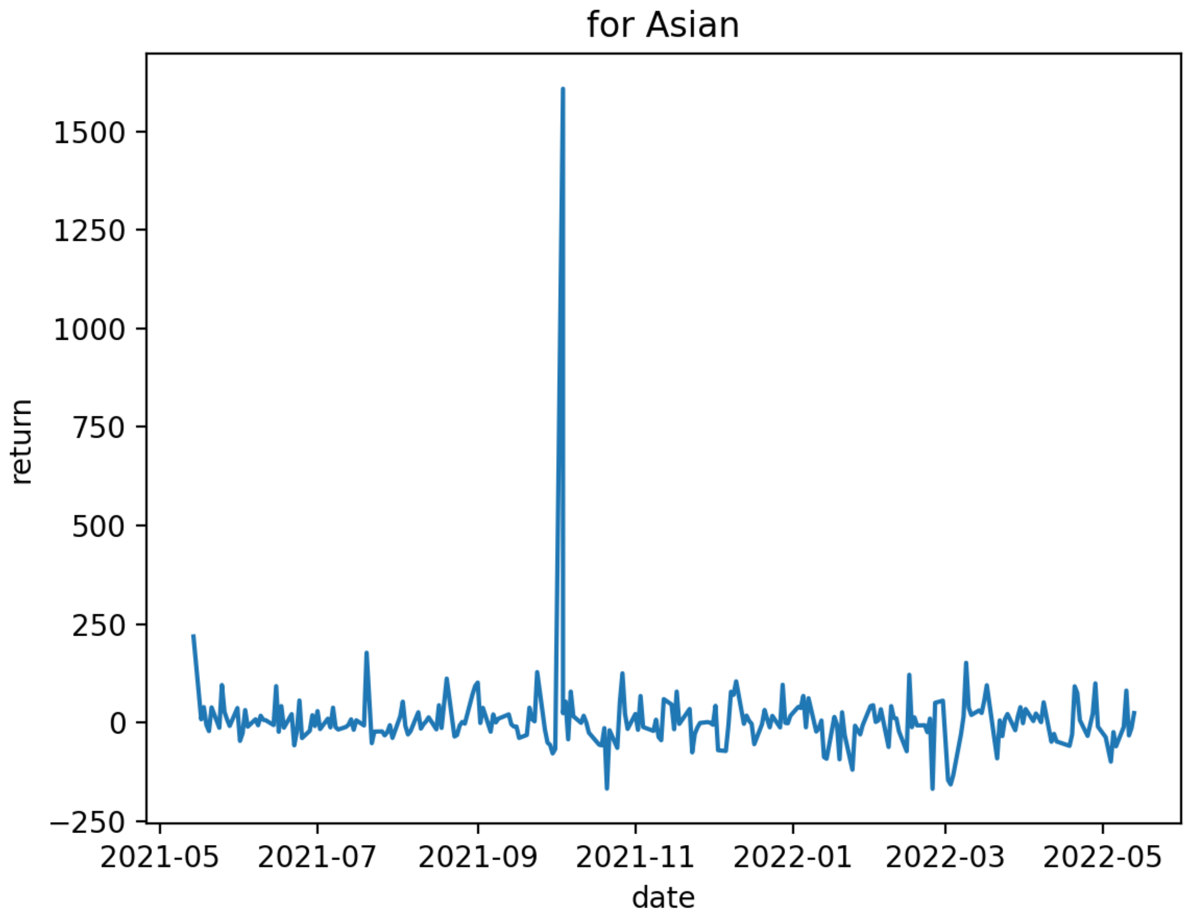
We can observe that in the time series plot for Simple Time difference, Rolling mean Difference are non stationary. But forMoving Average, Logarithmic and Moving Average, Square Root and Moving Average we can achieve stationarity.
This can be confirmed by p-values obtained from Dickey-fuller test where p values for these are lesser than 0.05. Thus stationarity can be achieved.

Q3. Write a program for the following:
a) Create a excel file containing tickers of three random stocks listed in the NSE. Read tickers from excel file and for each ticker fetch the daily historical data for last one year.
b) For each stock do the following:
1) Calculate the return and generate plot for each stock.

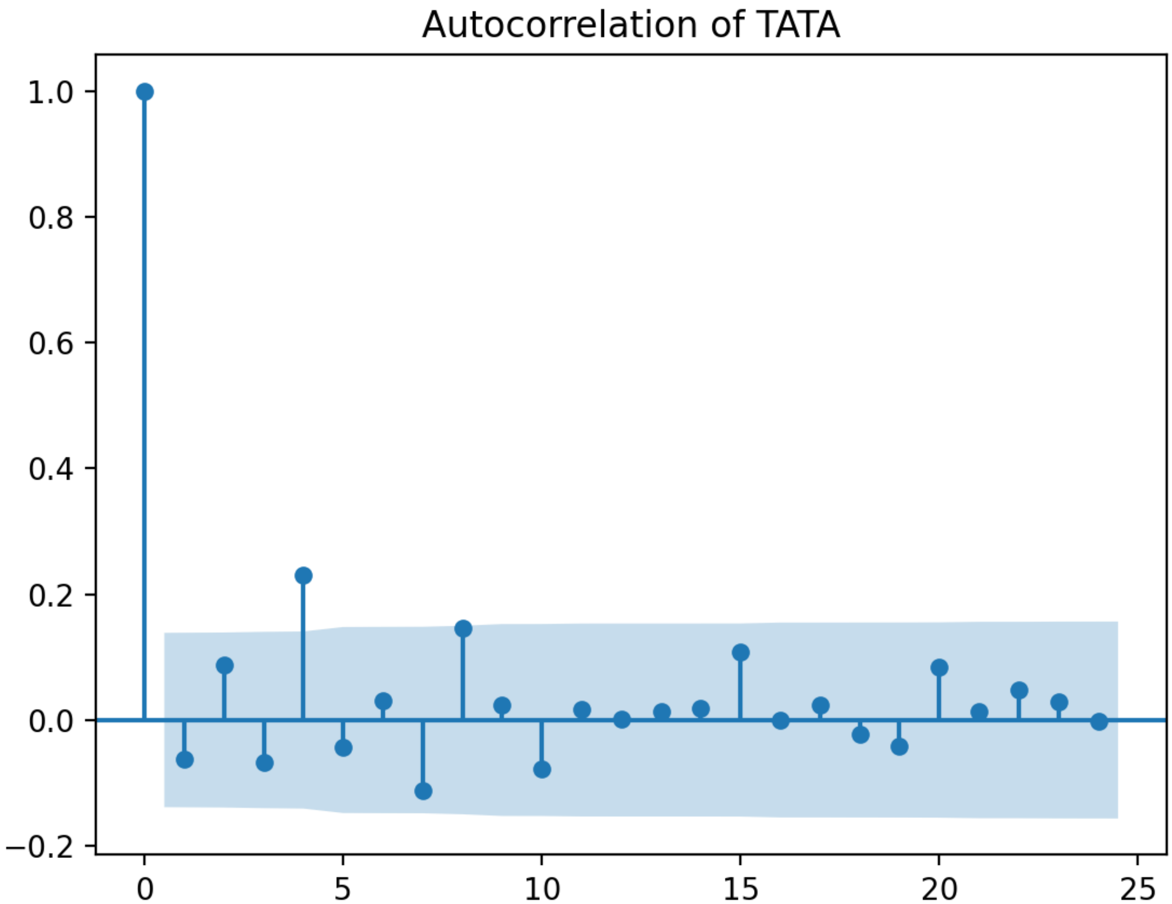


for TATAMOTORS
80
60
40
20
0
−20
−40
return
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
date

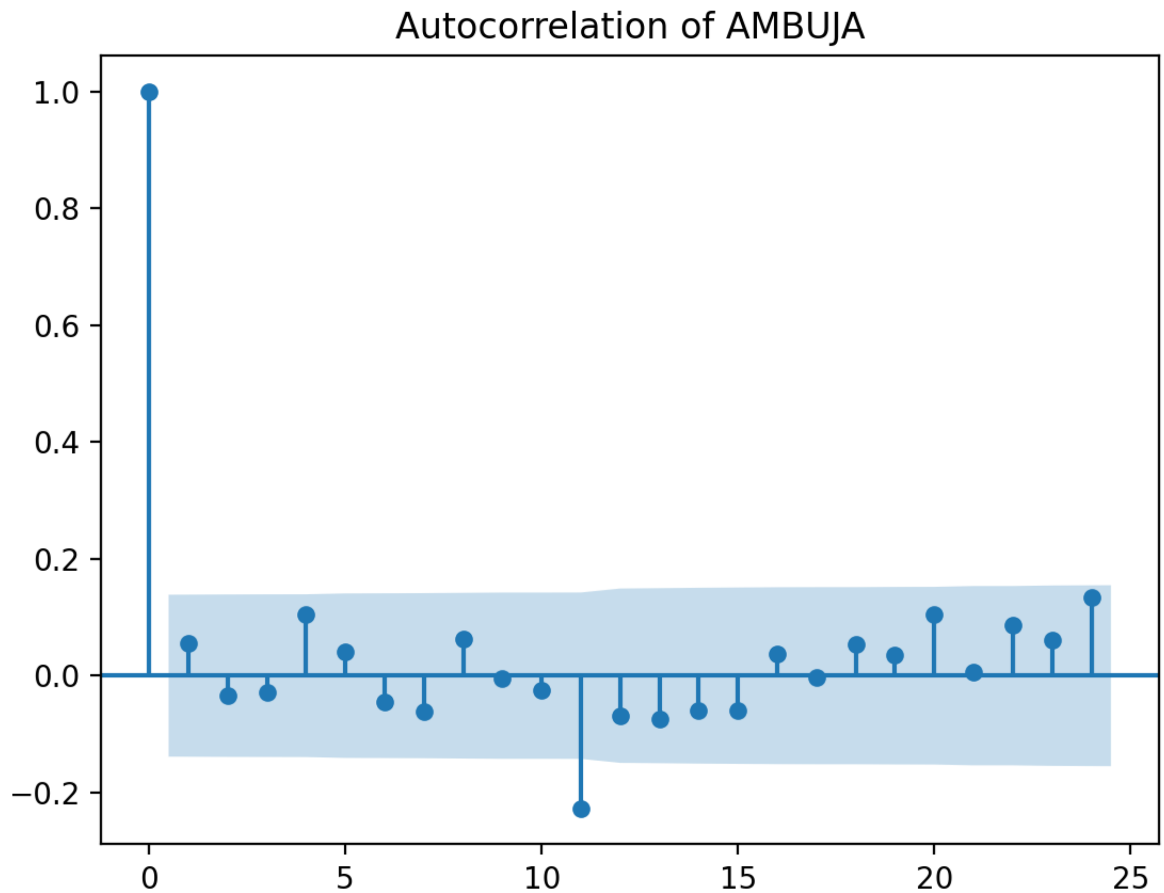

for AMBUJA
return
80
60
40
20
0
−20
2021-05
2021-07
2021-09
2021-11
2022-01
2022-03
2022-05
date

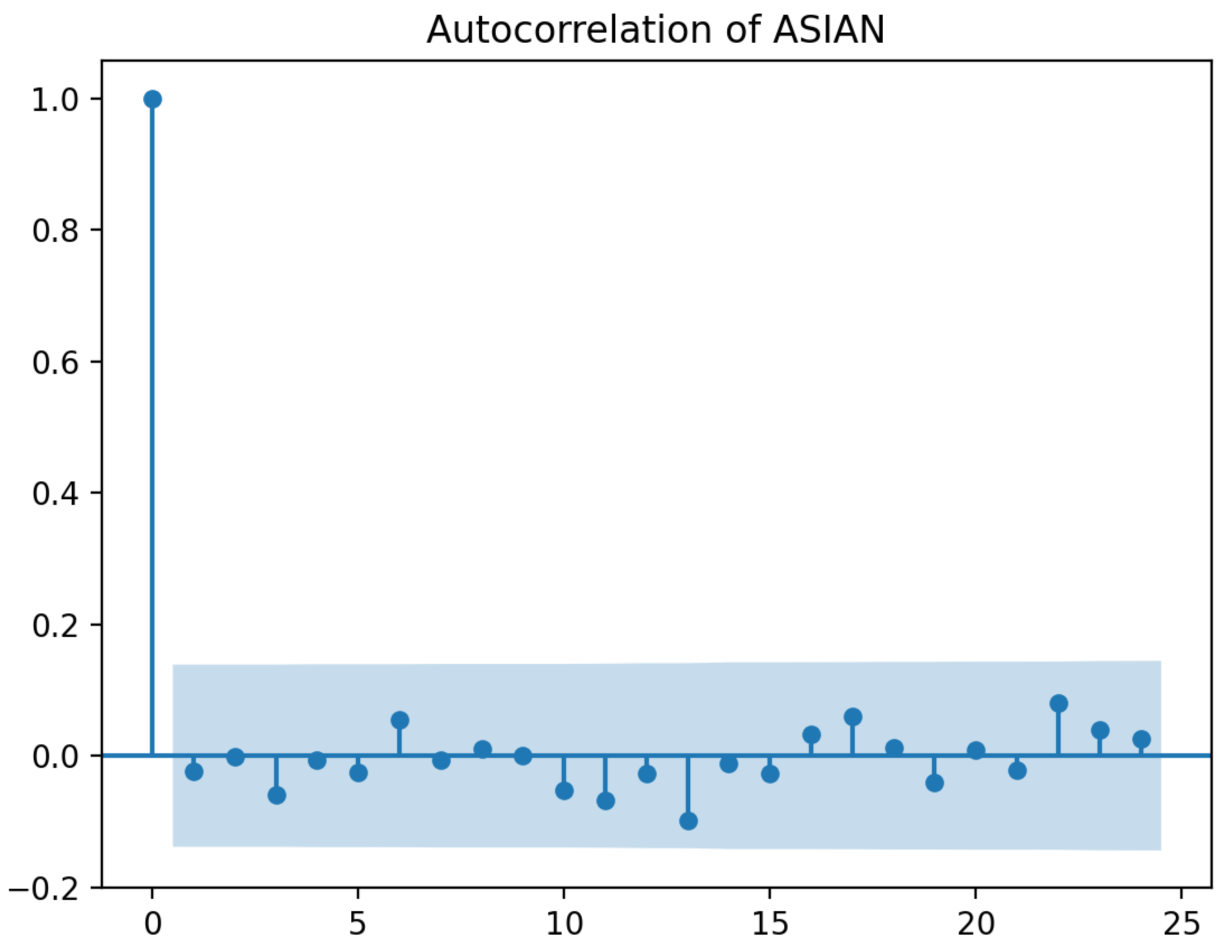


2) Check the stationarity of the return data time series for each stock. If the series is not stationary, then first make it stationary



Autocorrelation of TATA
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
25


Autocorrelation of AMBUJA
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20
25



Autocorrelation clearly shows that the data of return is stationary in nature

3) Divide return into training and test set of ratios 80:20 and plot the following for training set

I.

Autocorrelation correlogram : it is shown above

II.

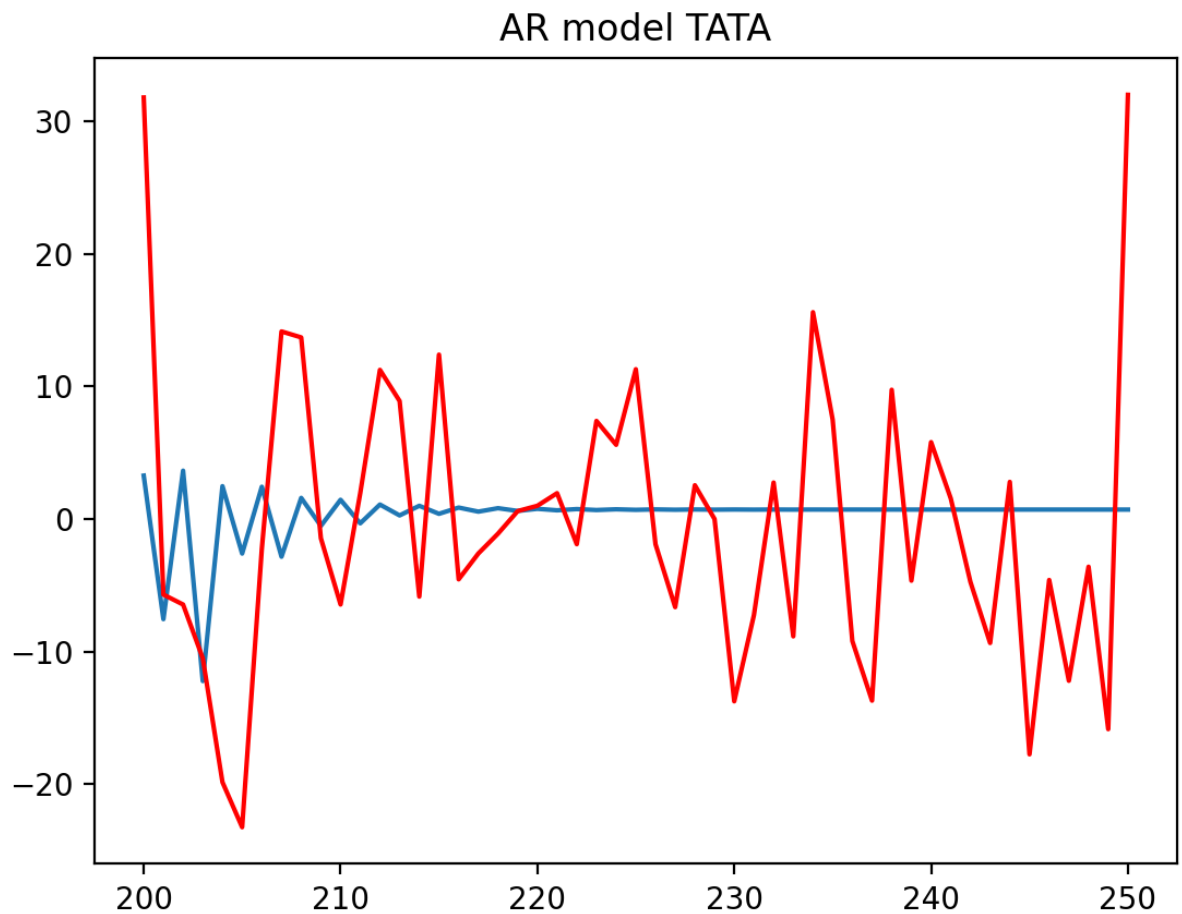
Partial autocorrelation

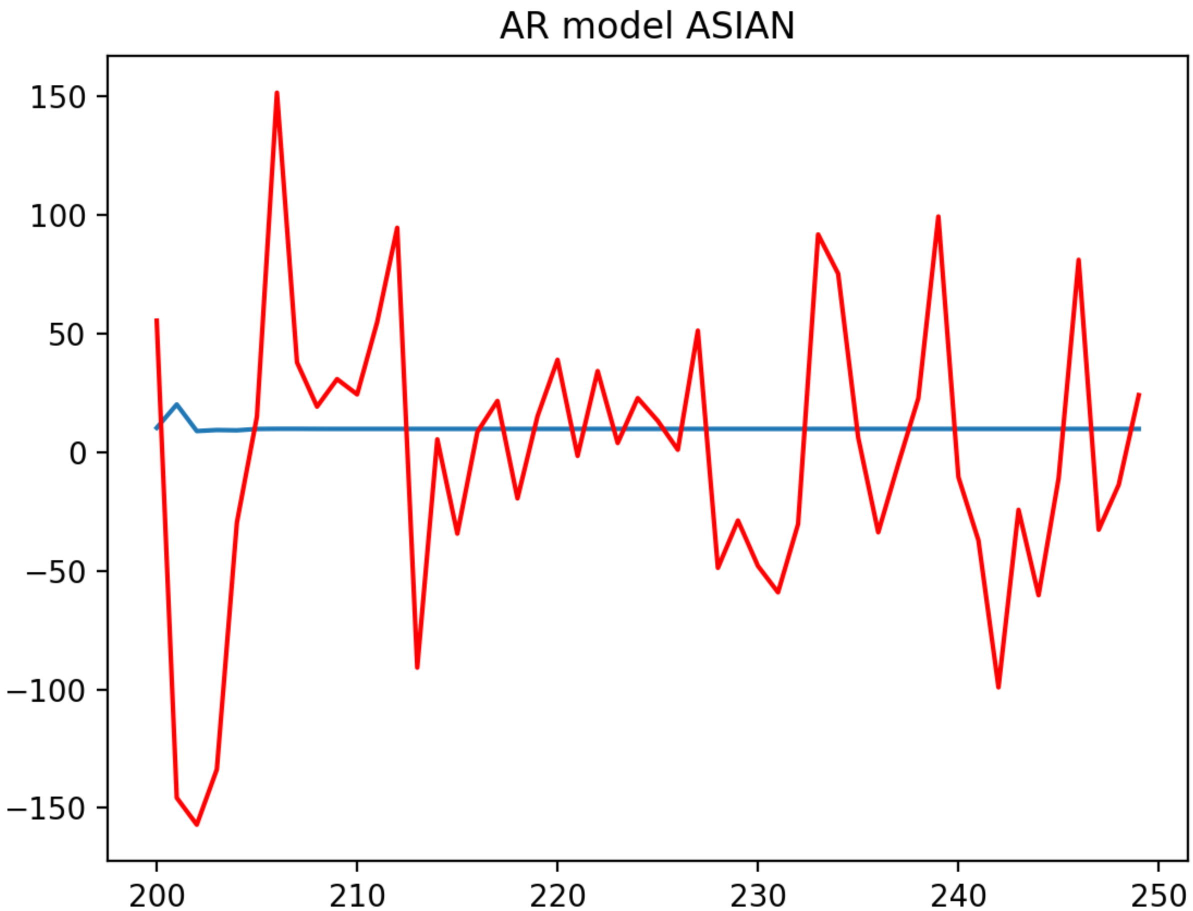
III.

4) Train the following Model on training data set for:

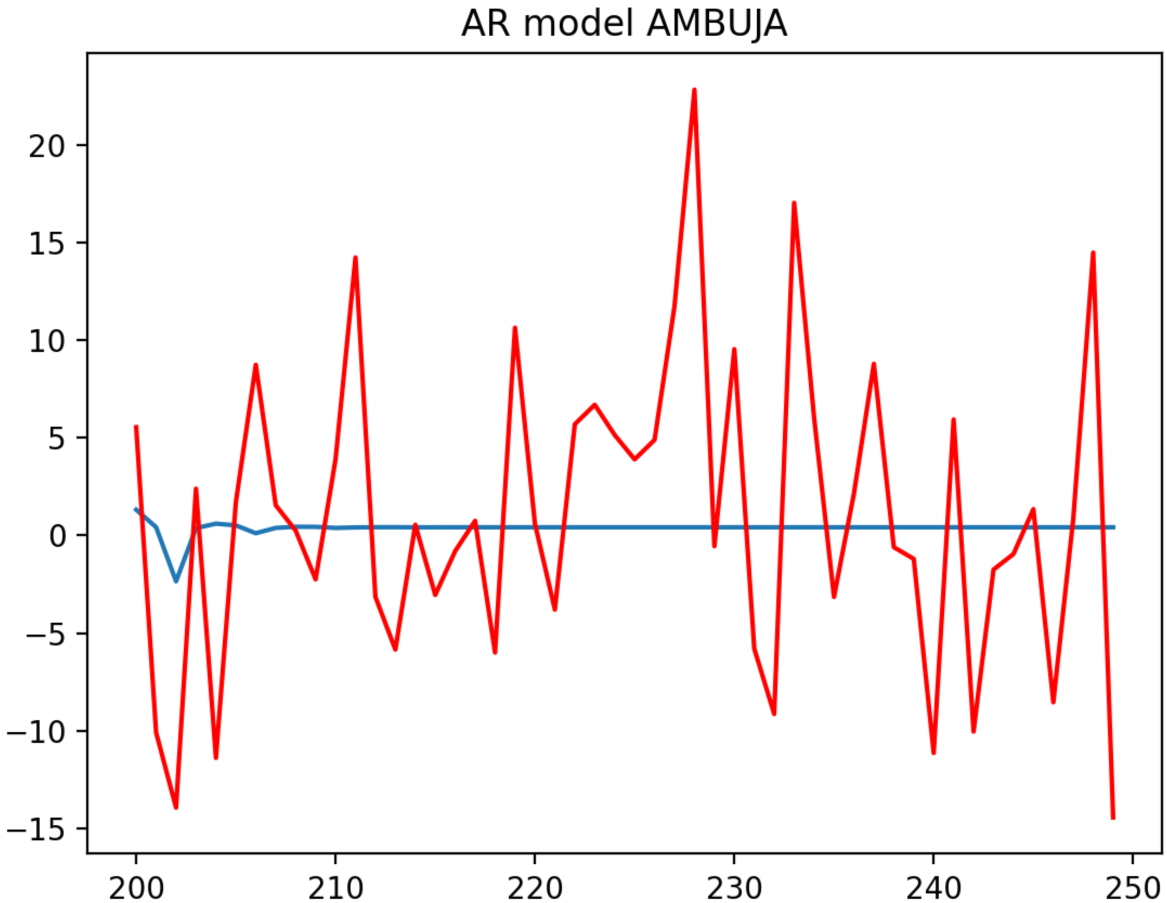
I.

AR Model


AR model TATA
30
20
10
0
−10
−20
200
210
220
230
240
250

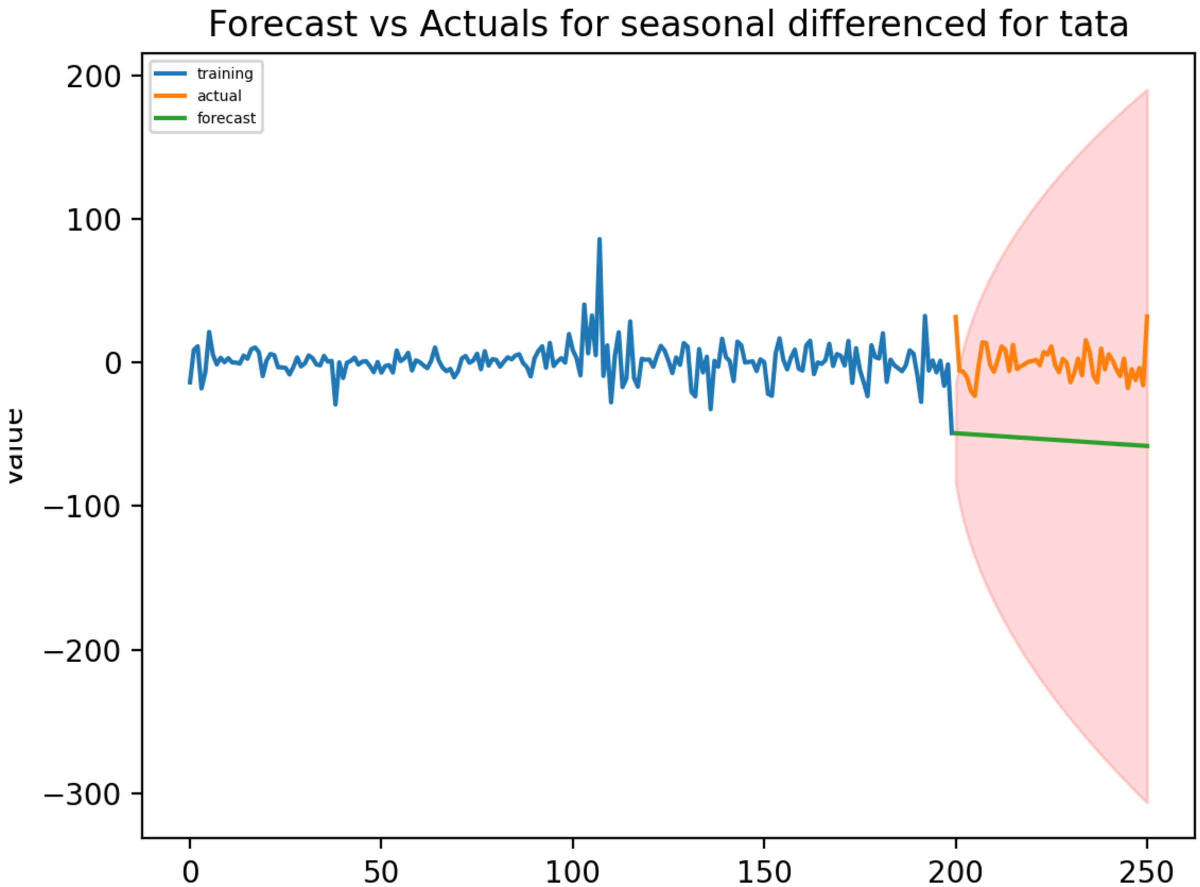


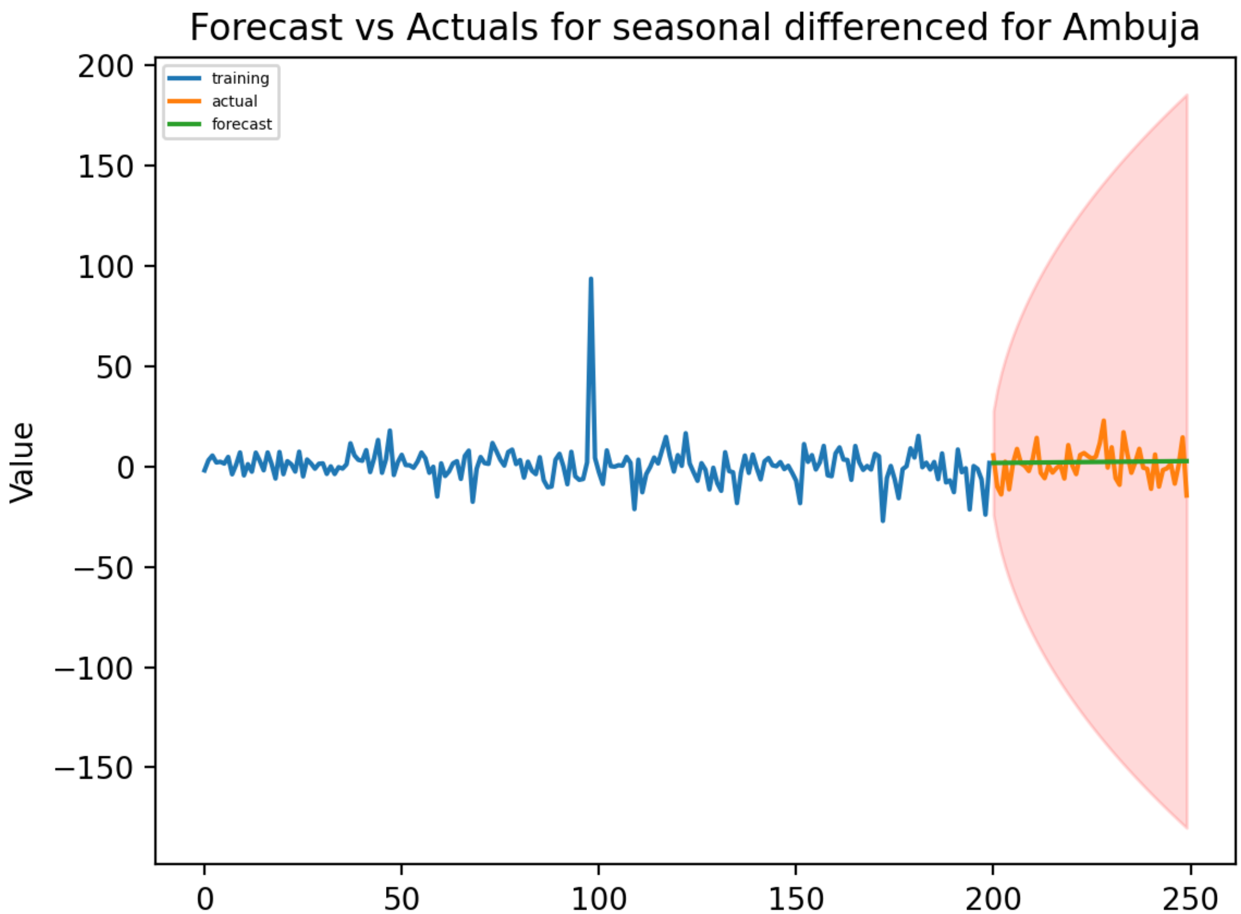
AR model ASIAN
150
100
50
0
−50
−100
−150
200
210
220
230
240
250

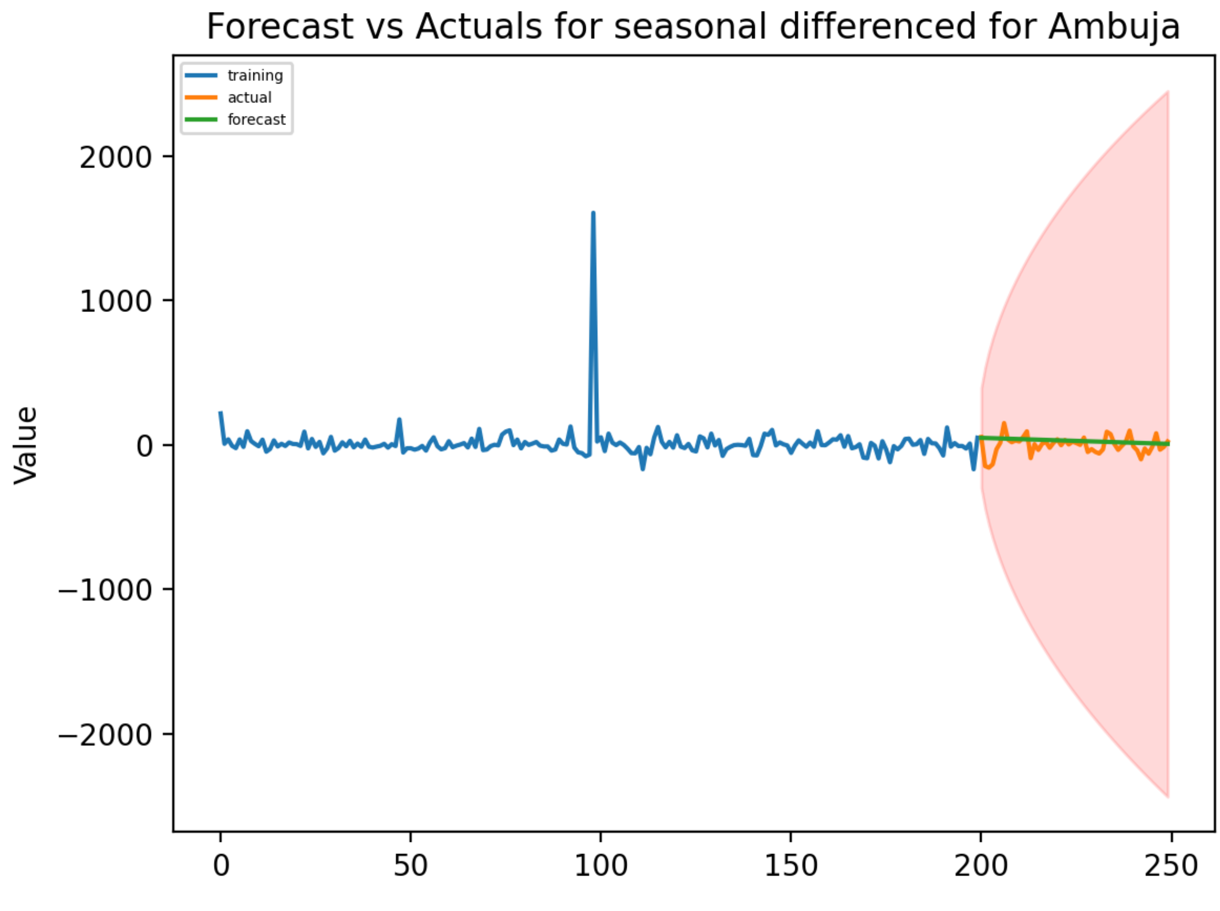


II.

ARIMA Model


Forecast vs Actuals for seasonal differenced for tata
training
actual
forecast
value
200
100
0
−100
−200
−300
0
50
100
150
200
250


Forecast vs Actuals for seasonal differenced for Ambuja
training
actual
forecast
Value
200
150
100
50
0
−50
−100
−150
0
50
100
150
200
250



For each model plot the actual return and predicted return for each stock in the original return scale.