IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – VI
Auto-regression

Student's Name: ISHAAN GUPTA **Mobile No: 9179242114**

Roll Number: B20292 **Branch:M.E.**

1 a.

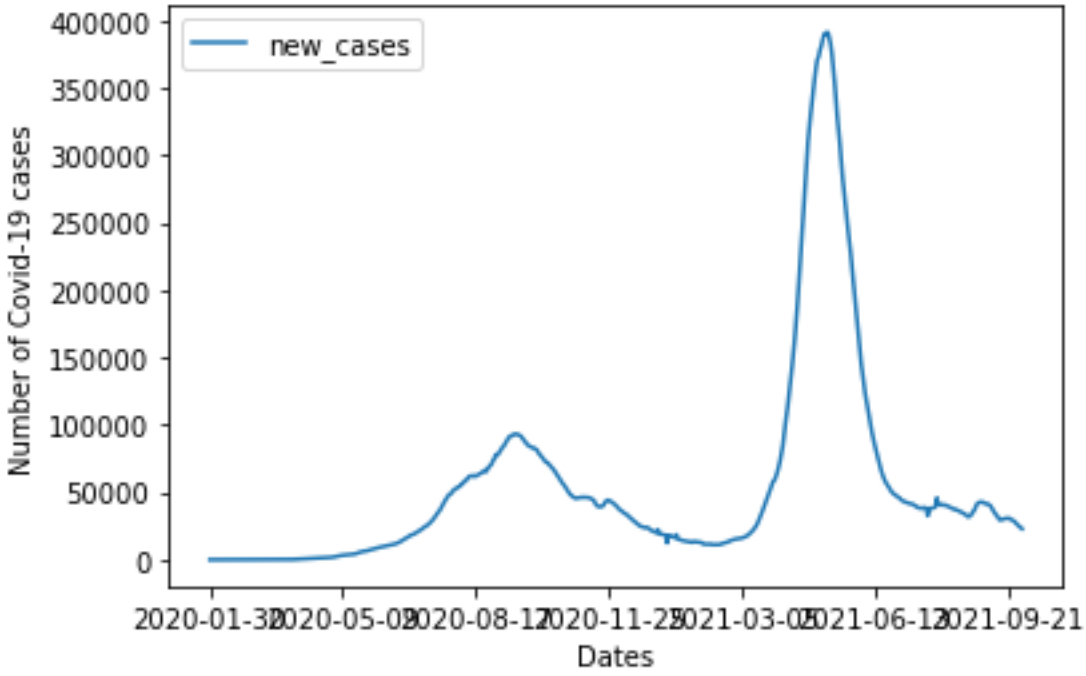

Figure 1 No. of COVID-19 cases vs. days

Inferences:

1. The time series data consists of 2 distinct peaks.
2. August-2020 consisted of first wave.
3. May-2021 consisted of second wave

b. The value of the Pearson's correlation coefficient is 0.999.

Inferences:

1. Both series have high positive correlation.
2. This is because higher the pearson coefficient, higher will be the similarity.

c.

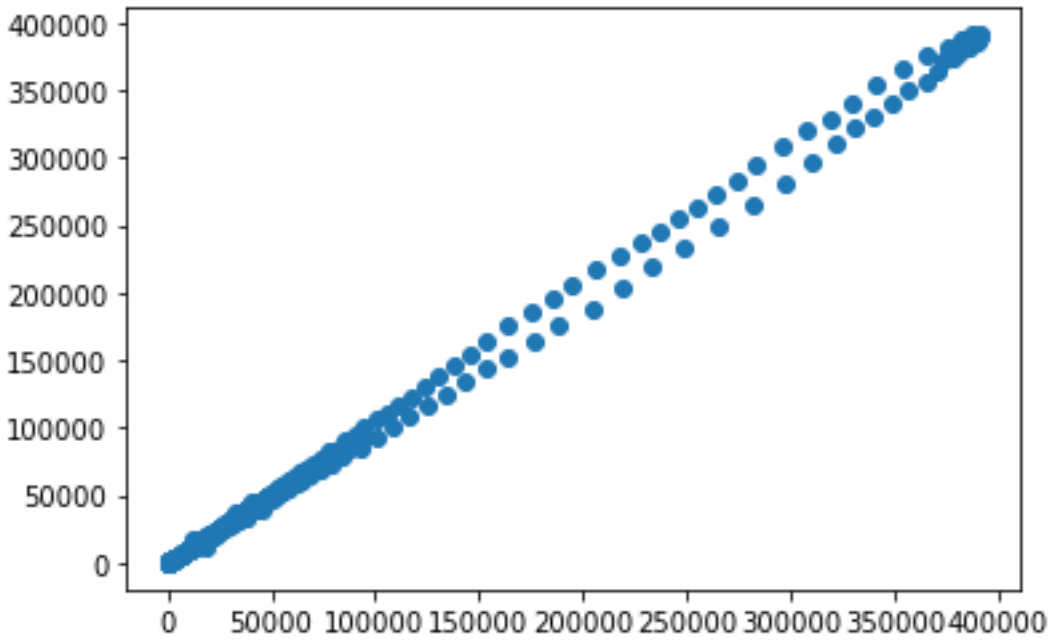

Figure 2 Scatter plot one day lagged sequence vs. given time sequence

Inferences:

1. Both series have high positive correlation.
2. It obeys the pearson coefficient as the scatter plot deviates very little from a straight line with slope 1.

d.

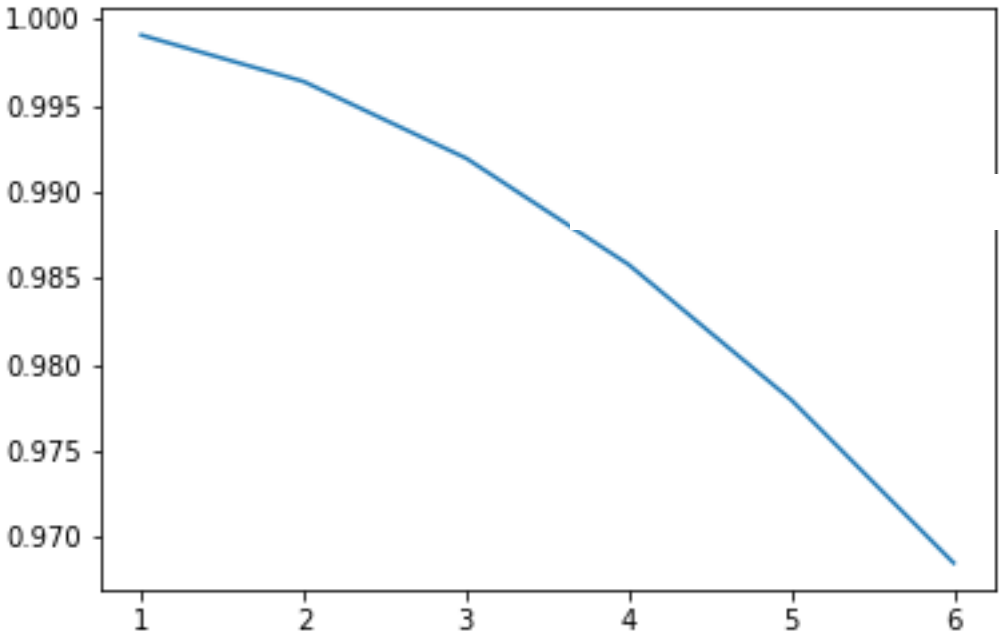

Figure 3 Correlation coefficient vs. lags in given sequence

Inferences:

1. Correlation decreases with increase in lags.
2. The number of new cases in 2 consecutive days have little changes .

e.

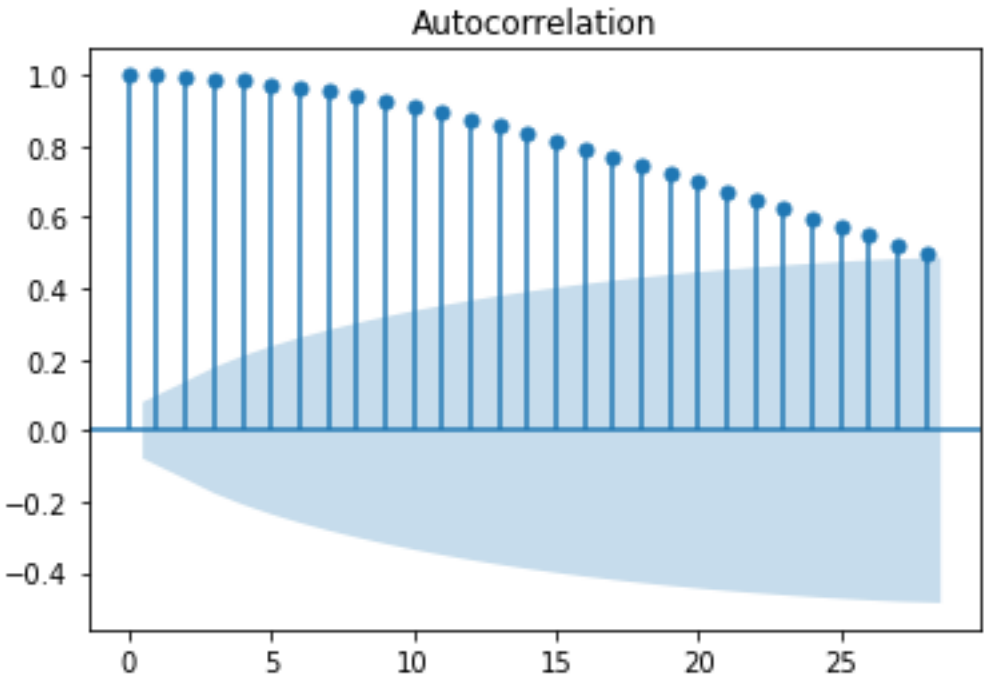

Figure 4 Correlation coefficient vs. lags in given sequence generated using 'plot_acf' function

Inferences:

1. Correlation decreases with increase in lags.
2. As the new cases depend upon existing cases , therefore correlation is more with lesser lagged data.

2

a. Coefficients obtained from the AR model : [5.99548333e+01 1.03675933e+00 2.61712336e-01 2.75612628e-02 -1.75391955e-01 -1.52461366e-01]

b. i.

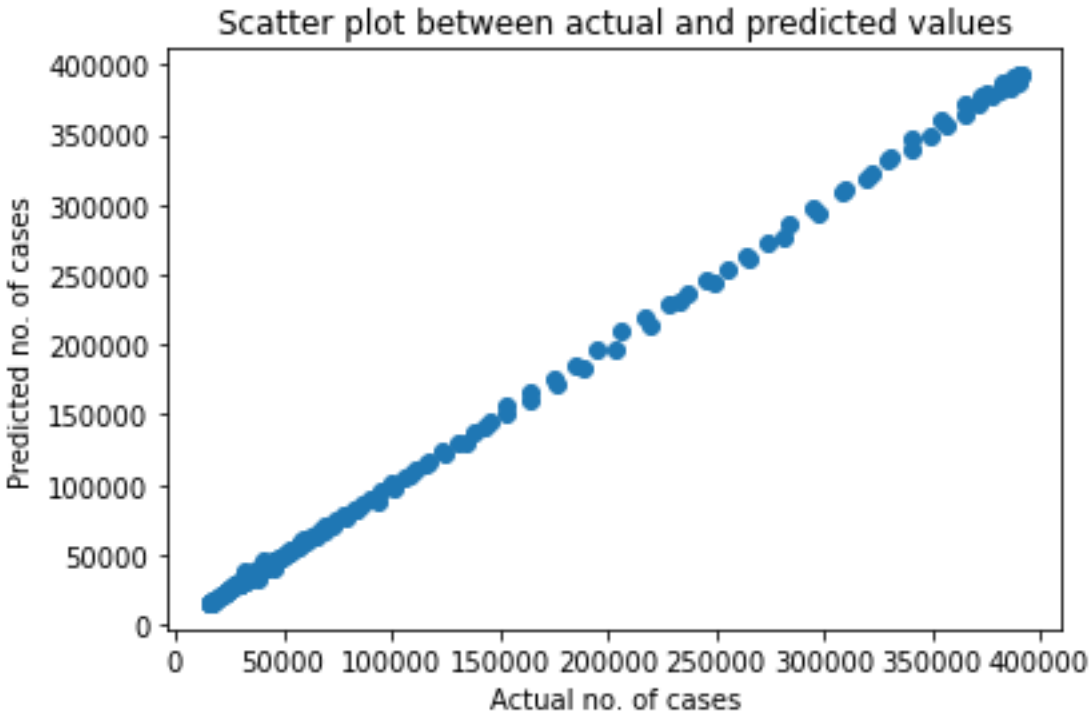

Figure 5 Scatter plot actual vs. predicted values

Inferences:

1. Both series have high positive correlation.
2. It obeys the pearson coefficient as the scatter plot deviates very little .

ii.

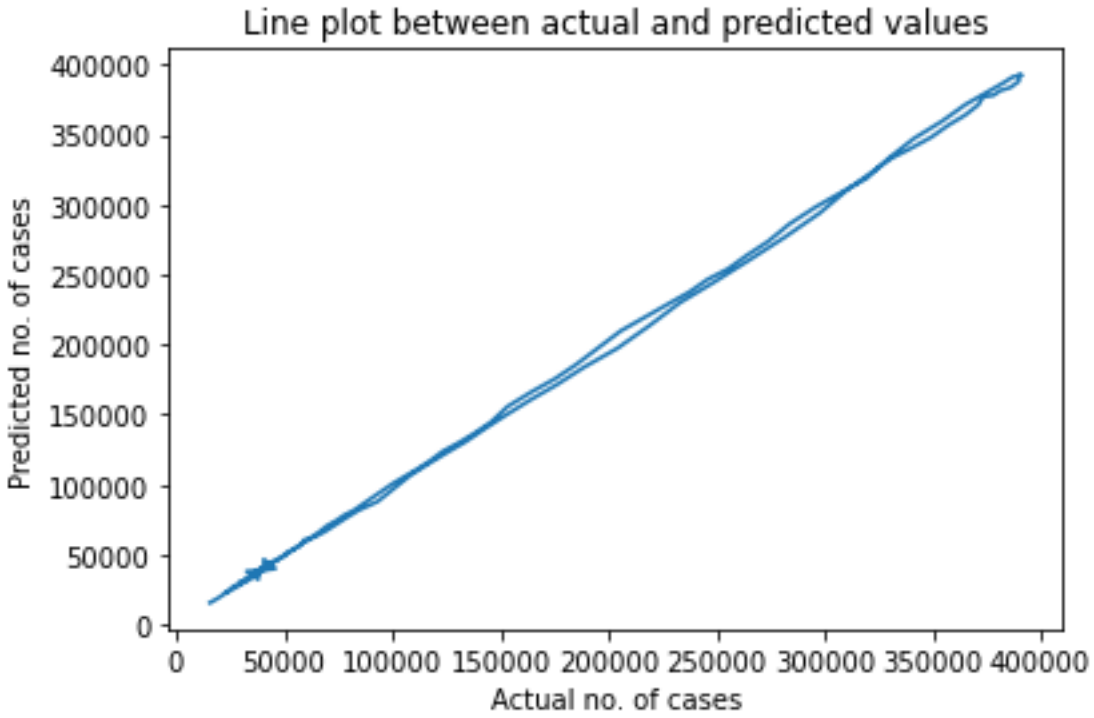

Figure 6 Predicted test data time sequence vs. original test data sequence

Inferences:

1. As they majorly overlap they they are reliable.

iii.

MAPE between actual and predicted test data 1.575 %.

RMSE between actual and predicted test data 1.825 %.

Inferences:

1. Both RMSE and MAPE are small and under 2%.

3

Table 1 RMSE (%) and MAPE between predicted and original data values wrt lags in time sequence

Lag value	RMSE (%)	MAPE(%)
1	5.373	3.447
5	1.825	1.575
10	1.686	1.519
15	1.612	1.496
25	1.703	1.535

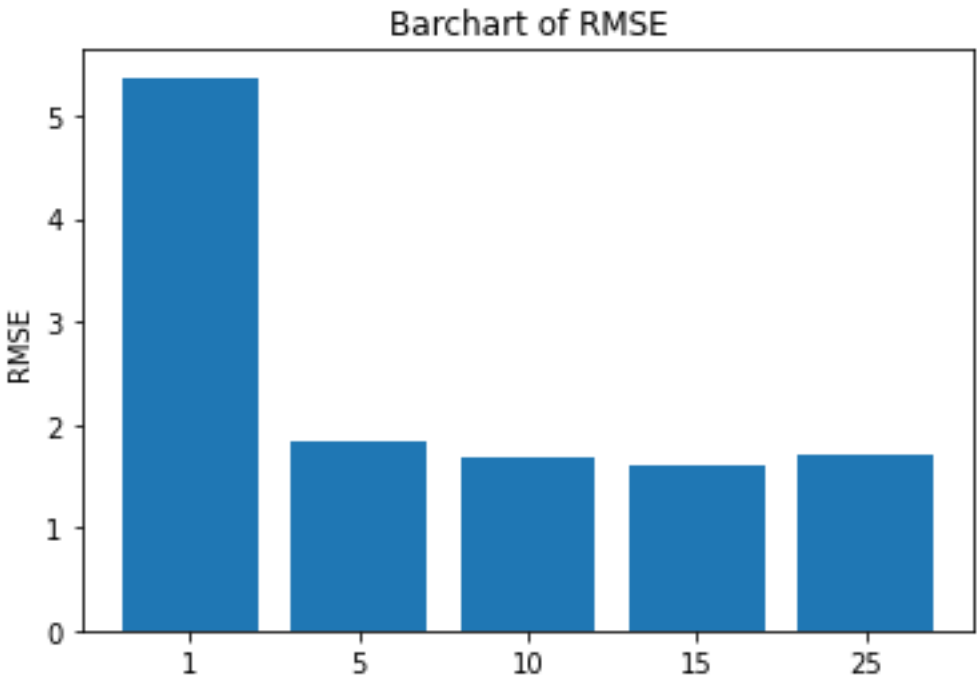

Figure 7 RMSE(%) vs. time lag

Inferences:

RMSE decreases upto lag = 15, after which the RMSE starts increasing.

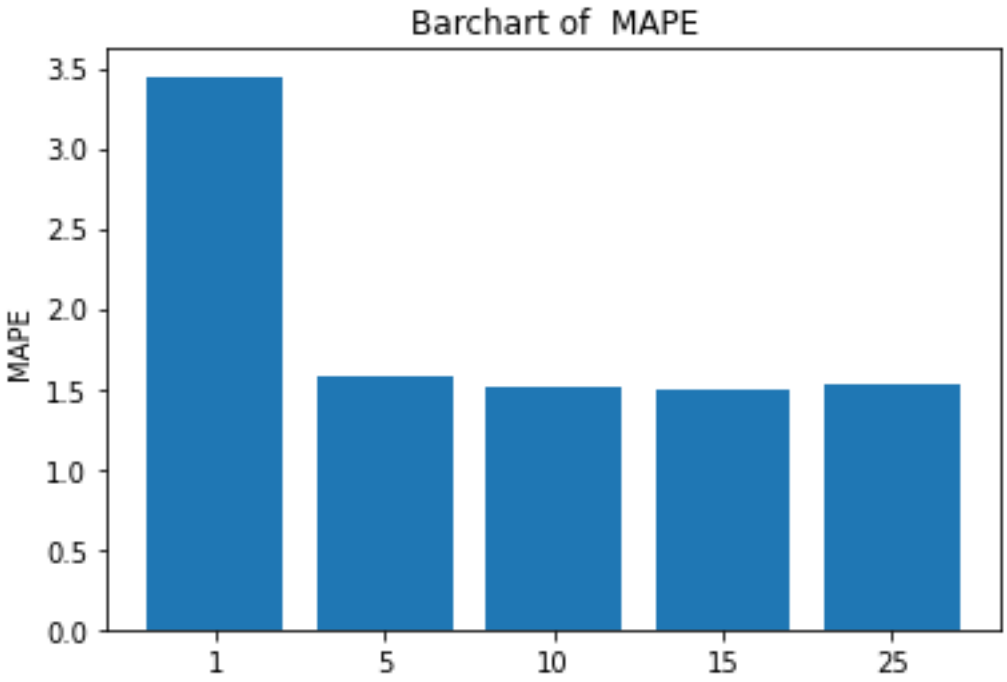

Figure 8 MAPE vs. time lag

Inferences:

MAPE decreases up to lag = 15, after which the RMSE starts increasing.

4

The heuristic value for the optimal number of lags is 77.

RMSE : 1.759 %

MAPE : 2.026 %

Inferences:

Both RMSE and MAPE value for heuristic 77 are more than lag=15, but lesser than lag =1 ,this tells that the optimal solution from heuristic can be used for prediction.