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

Student's Name: Ravi Ranjan **Mobile No: 9304228691**

Roll Number: B20313 **Branch:**EE

1 a.

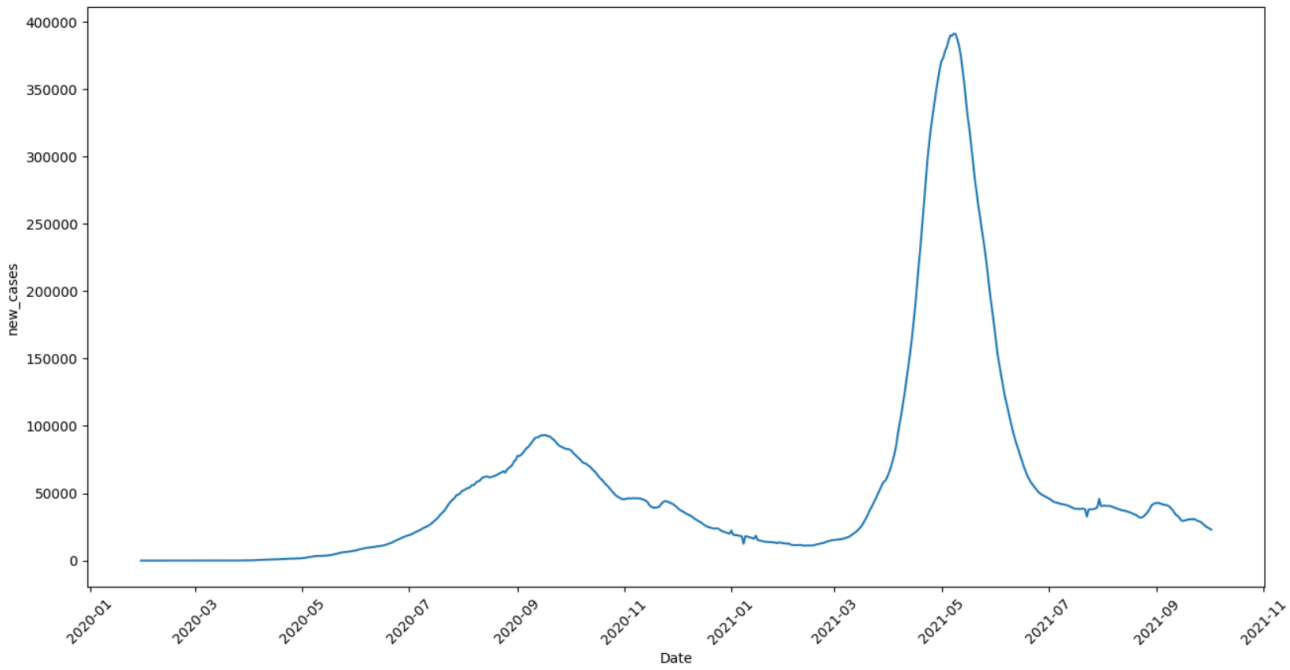

Figure 1 No. of COVID-19 cases vs. days

Inferences:

1. First wave lasts in 7 months and second wave last in 5 months.

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

Inferences:

1. Series is strongly correlated that's mean future value is highly dependent on past value.
2. The Pearson's coefficient is 0.999 that mean variation of covid cases highly depend upon in past value.

c.

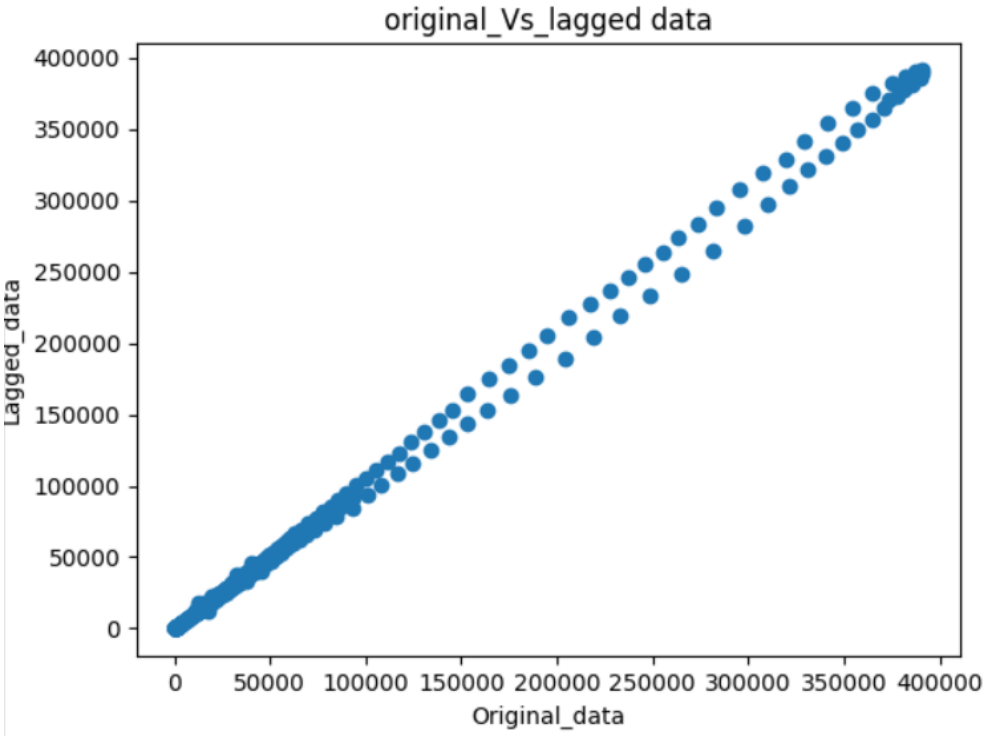

Figure 2 Scatter plot one day lagged sequence vs. given time sequence

Inferences:

1. Pearson's correlation is very high that's mean variable are highly dependent on each other.
2. the scatter plot seems to obey the nature reflected by Pearson's correlation coefficient.

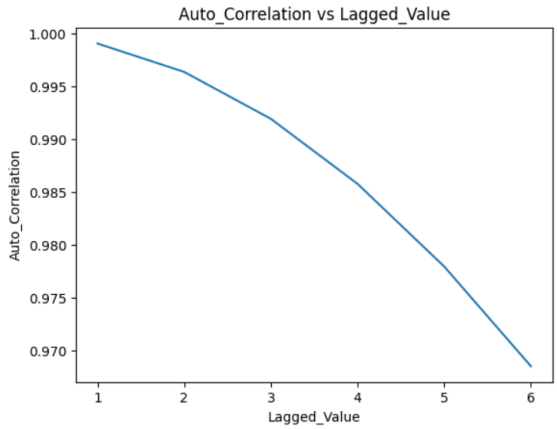

Figure 3 Correlation coefficient vs. lags in given sequence

1. **d.**
    a. correlation coefficient decreases as lag value increases.

e.

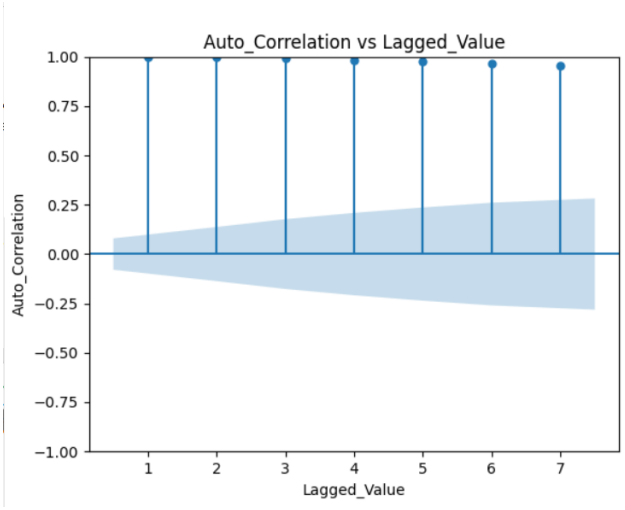

Figure 4 Correlation coefficient vs. lags in given sequence generated using 'plot_acf' function

Inferences:

1. Correlation coefficient decreases as value of lag increases.

2

a. The coefficients obtained from the AR model are; [59.955,1.03,0.26,0.028, -0.175, -0.152]

b. i.

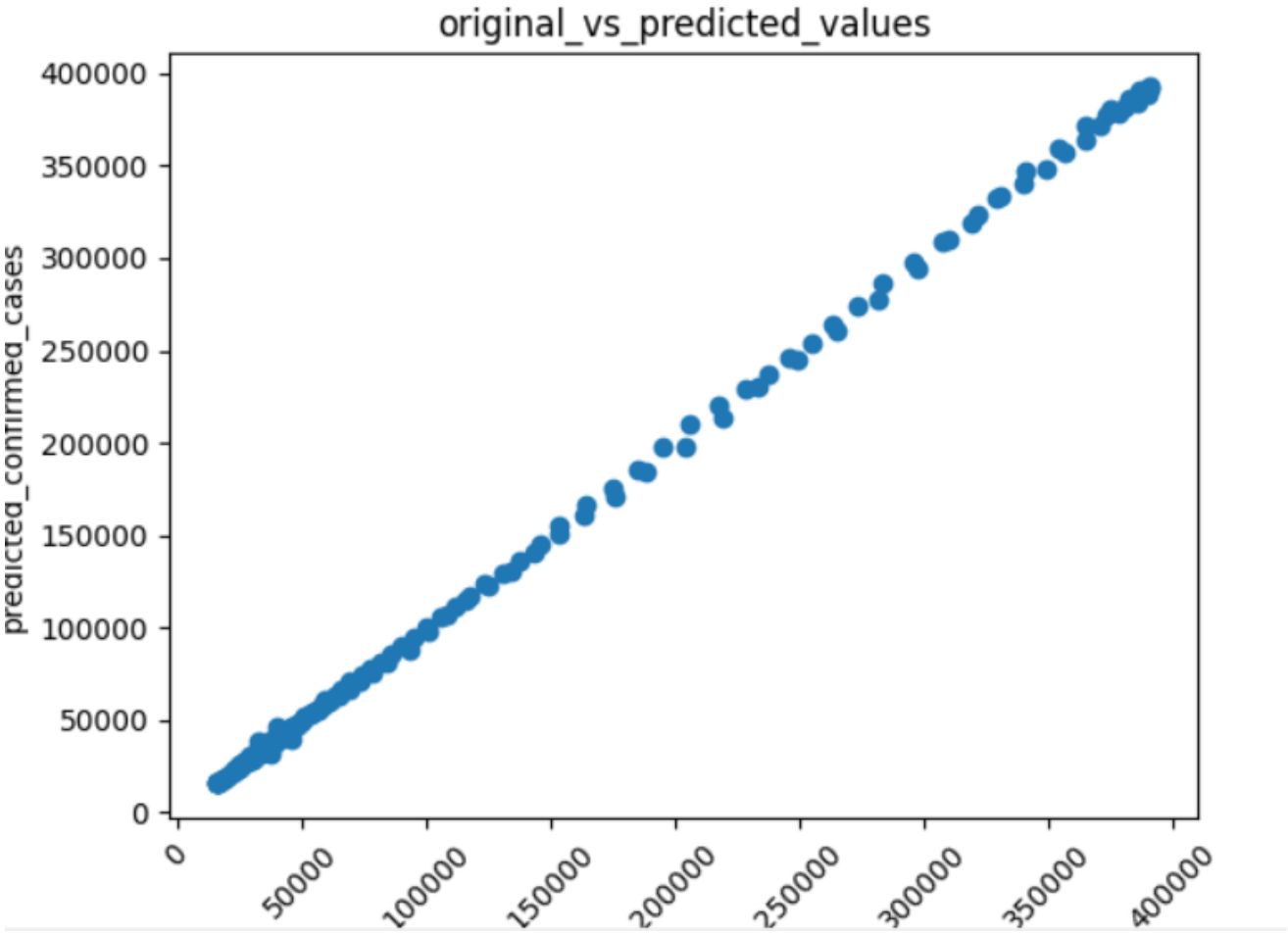

Figure 5 Scatter plot actual vs. predicted values

Inferences:

1. Both variables are strongly positive correlated.
2. Yes, scatter plot seems to obey the nature reflected by Pearson's correlation coefficient calculated.

ii.

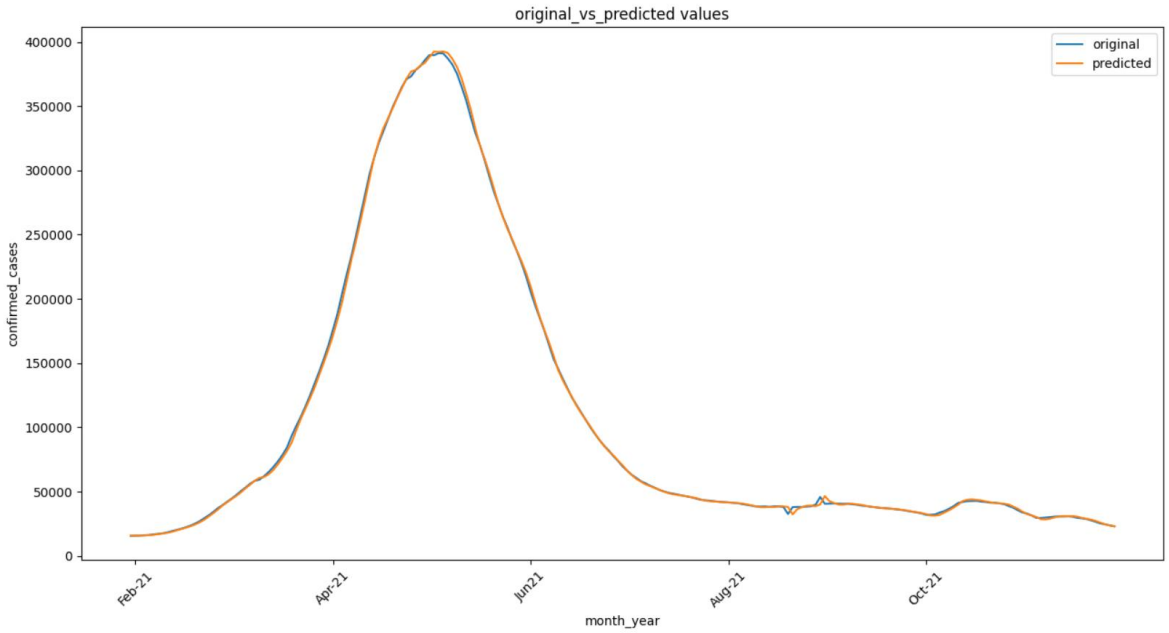

Figure 6 Predicted test data time sequence vs. original test data sequence

Inferences:

1. This model is very highly reliable as both the graph almost coincide each other.

iii.

The RMSE(\%) and MAPE between predicted power consumed for test data and original values for test data are 1.8245,1.5274 respectively.

3

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

Lag value	RMSE (%)	MAPE
1	5.44%	3.48
5	1.83%	1.62
10	1.74%	1.55
15	1.70%	1.56
25	1.79%	1.60

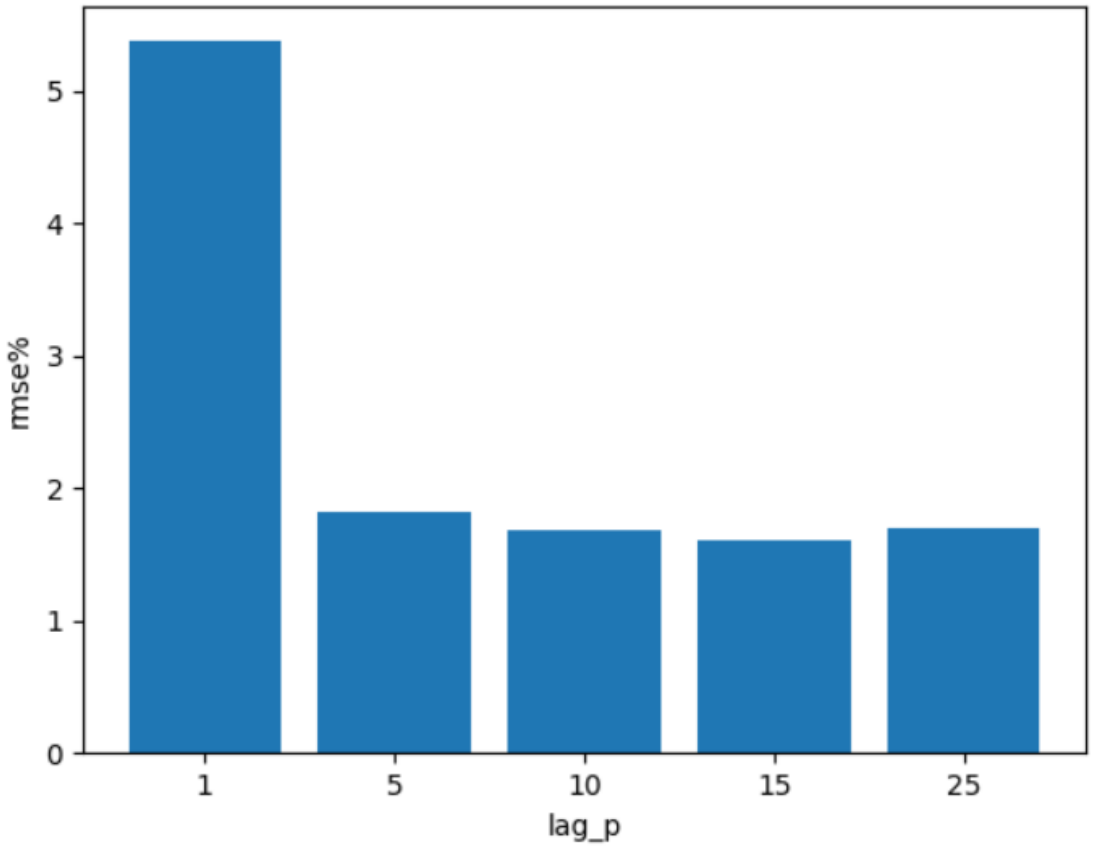

Figure 7 RMSE(%) vs. time lag

Inferences:

1. Lag decreases as rmse value increases.

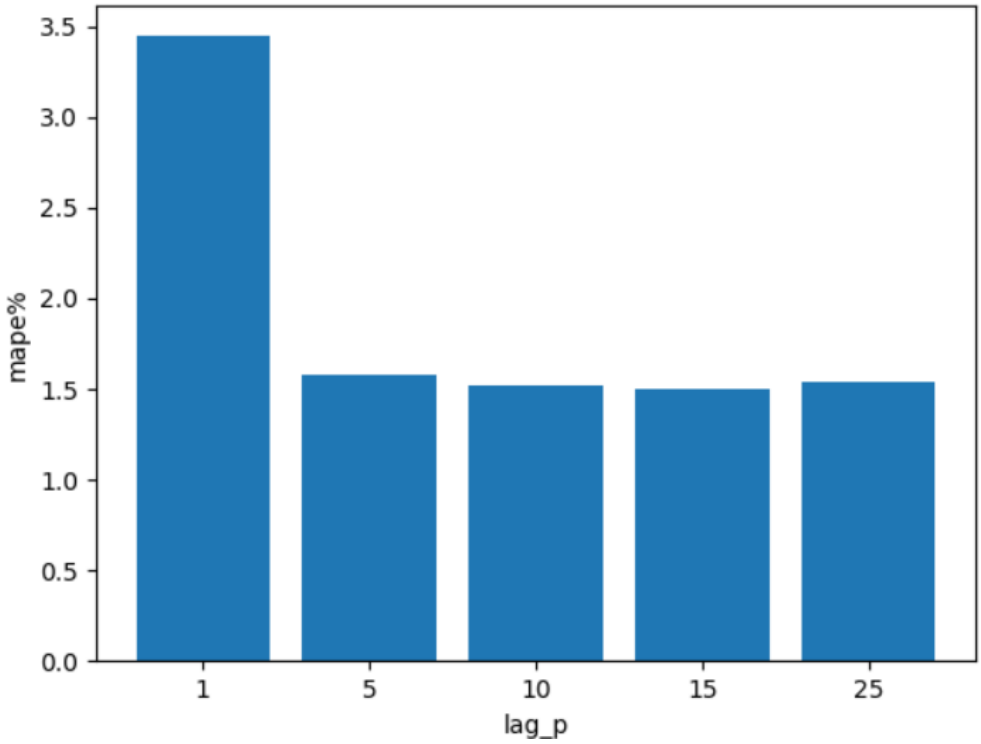

Figure 8 MAPE vs. time lag

Inferences:

1. MAPE increases as lag value decreases.

4

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

The RMSE(%) and MAPE value between test data time sequence and original test data sequence are1.7593,2.0264 respectively.

Inferences:

1. Based upon the RMSE(%) and MAPE value, heuristics for calculating the optimal number of lags does not improve the prediction accuracy of the model.