**Student’s Name:** Sachin Mahawar

**Roll Number:** B20129

**Mobile No:** 9166843951

**Branch:** CSE

# a.

Chart, line chart

Description automatically generated

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

# Inferences:

1. Changes in one day lagged and original data is not much significant.
2. Number of cases are almost same as a day before.
3. Duration of first wave is July 2020 – December 2020 and duration of second wave is March 2021 – June 2021.

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

# Inferences:

1. Both are highly correlated as pearson’s coefficient is 0.999.
2. As value of correlation is generally high for 1 day lagged data and original data so it hold to good extent.

**c.**

Chart, scatter chart

Description automatically generated

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

# Inferences:

1. By observing plot, we can conclude that correlation is very high as plot is align towards y=x line.
2. Yes, Scatter plot seem to obey the nature reflected by Pearson’s correlation coefficient.
3. Because pearson’s correlation coefficient was 0.999 and we can see from plot that correlation is also very high.

**d.**

**Chart

Description automatically generated**

Figure 3 Correlation coefficient vs. lags in given sequence

**Chart

Description automatically generated**

Figure 4 Correlation coefficient vs. lags in given sequence generated using 'plot\_acf' function

**Inferences**

1. Auto correlation decreases with increase in lag value.
2. As lag increases, corresponding elements are less related to each other as both have large time difference than lesser p values. So that’s why correlation decreases.

# 2.

1. The coefficients obtained from the AR model are w0=59.9548, w1=1.0367, w2=0.2617, w3=0.0275, w4= -0.1753, w5=-0.1524.

**b. i.**

Chart, line chart, scatter chart

Description automatically generated

Figure 5 Scatter plot actual vs. predicted values

# Inferences:

1. Both sequence are highly correlated.
2. Yes, Scatter plot seem to obey the nature reflected by Pearson’s correlation coefficient.
3. Because pearson’s correlation coefficient was high and we can see from plot that correlation is also very high.

**ii.**

**Chart, line chart

Description automatically generated**

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

# Inferences:

1. By observing plot, we can see that plot for predictions and actual plot are almost overlapping each other thus this model can be used to make future predictions accurately.

**iii.**

1. RMSE: 1.824 %
2. MAPE: 1.575

**Inferences:**

1. We can say that our model is accurate by observing RMSE and MAPE values.
2. Because both values are very less.

# 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 | 1.520 |
| 15 | 1.612 | 1.497 |
| 20 | 1.703 | 1.535 |

Chart, bar chart, histogram

Description automatically generated

Figure 7 RMSE(%) vs. time lag

**Inferences:**

1. MAPE value decreases from p=1 to p=15 as we are fitting data more optimally but increases after that (p=25) because from p=25 we are overfitting the data.

Chart, bar chart, histogram

Description automatically generated

Figure 8 MAPE vs. time lag

**Inferences:**

1. MAPE value decreases from p=1 to p=15 as we are fitting data more optimally but increases after that (p=25) because from p=25 we are overfitting the data.

# 4.

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

RMSE: 1.7593 %

MAPE: 2.026

**Inferences**:

1. Based upon the RMSE (%) and MAPE value, Heuristics for calculating the optimal number of lags did not improve the prediction accuracy of the model because data get overfit and thus decreases the accuracy.
2. The prediction accuracies obtained with and without the heuristic for calculating optimal lag with respect to RMSE (%) and MAPE values are (1.7593% , 2.026) and (1.824 %,1.575) respectively.