Time series analysis of the number of cases infected by the new SARS-COV-2 was undertaken by us, the primary aim being to achieve predictions of the cases over the next few days. This was done in the following manner:

1. John Hopkins data was utilized, as it is an independent source and hence would update accurate case numbers to their database. The data acquired was then cleaned and set in a manner fir for model usage.
2. Various tests such as error deviation, seasonality were used to determine if a linear regression model was appropriate for the time series. Upon analysis the ARIMA model was decided to be the best fit for the series.
3. ARIMA model uses a moving average and the residual errors in prediction of the next term in time series. The model uses differenced values to make the values stationary (i.e. the error values do not deviate from mean). This model depends on the number of lagged observations and the differencing order.
4. These parameters (p, q, d) were determined, theoretically by an autocorrelation function for the approximate p values and the q, d values were iterated over to find the best fit. The model error mean and their deviation were taken as the target to minimize.
5. Upon successful model build the next 4 days of prediction were plotted along the with the 95% accuracy range was plotted along with the actual data. The error residuals, residuals and the trends of the data are also plotted.