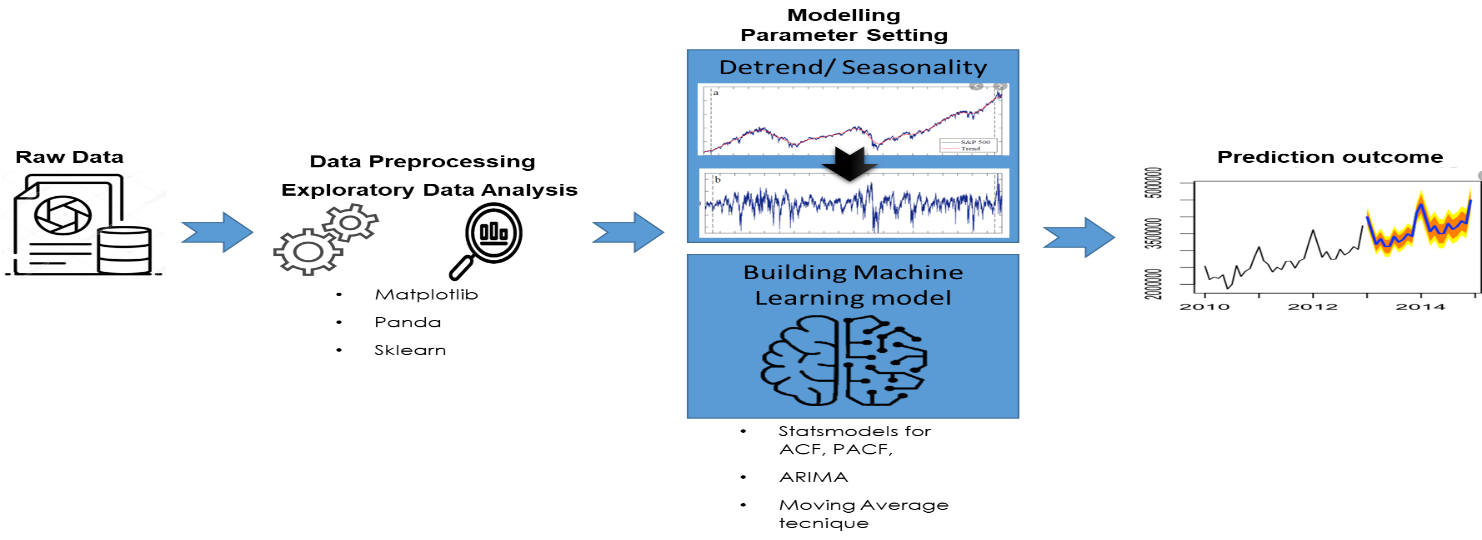
**Submission for Assignment 2**

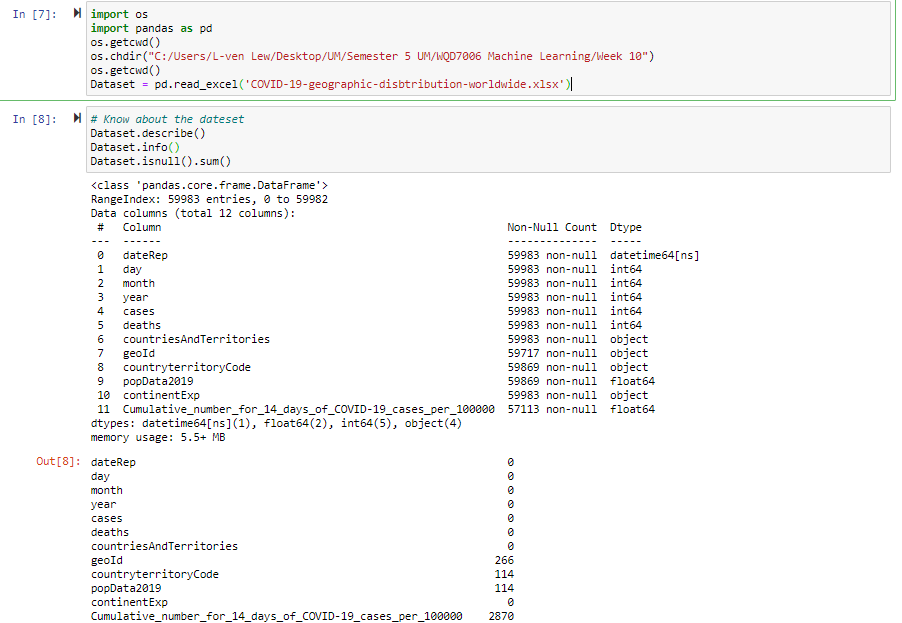
**Name: Lew Teck Wei**

**Matric Number: 170216821**

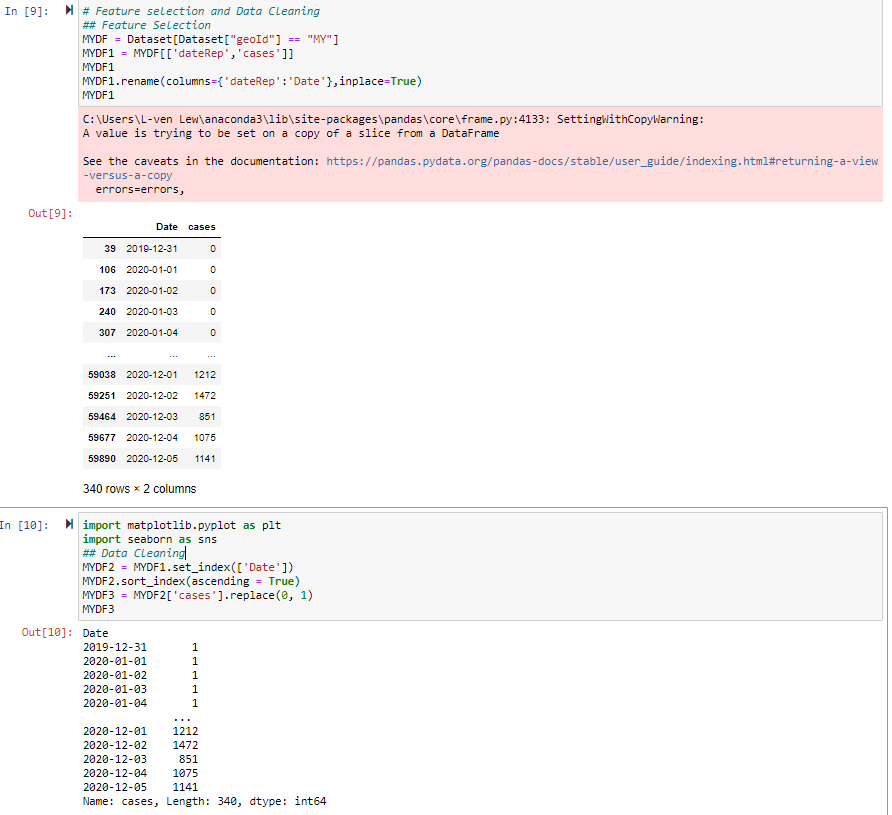
1. **Overview of Methodology**



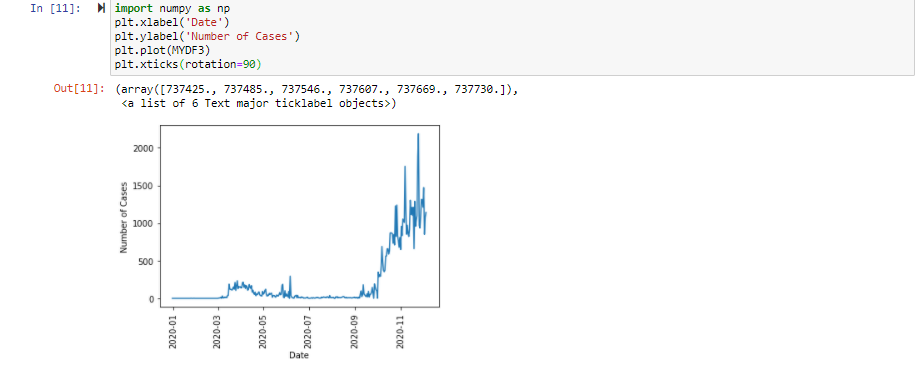
1. **Experiment Process and Output**
2. Import the dataset and get familiar with the dataset



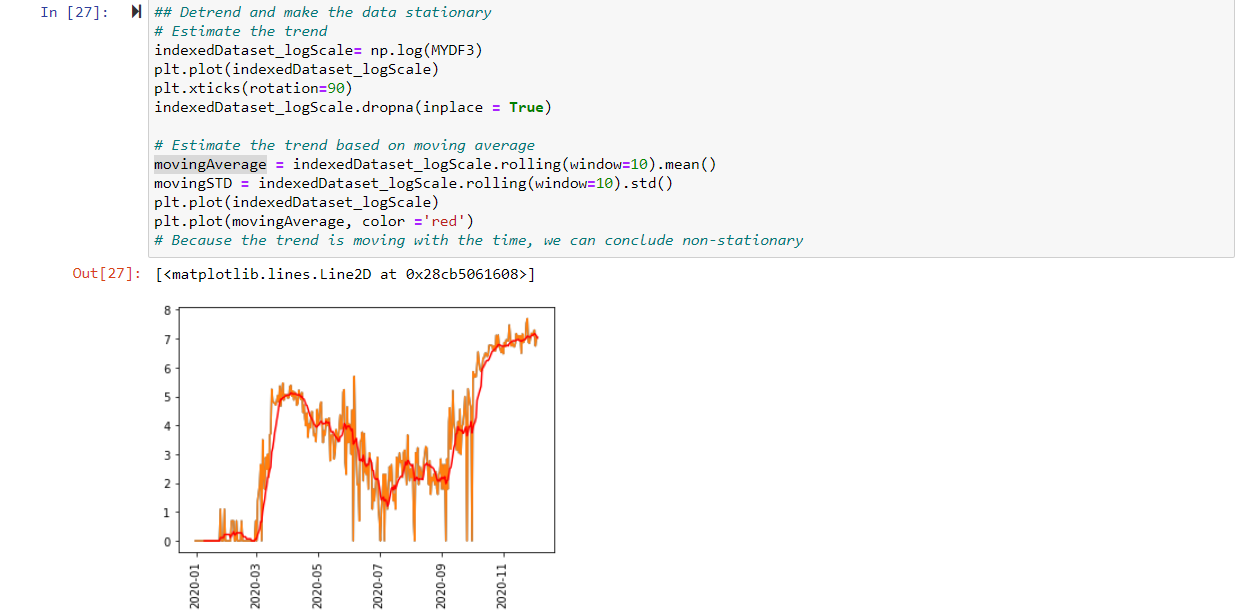
1. Feature selection and data cleaning



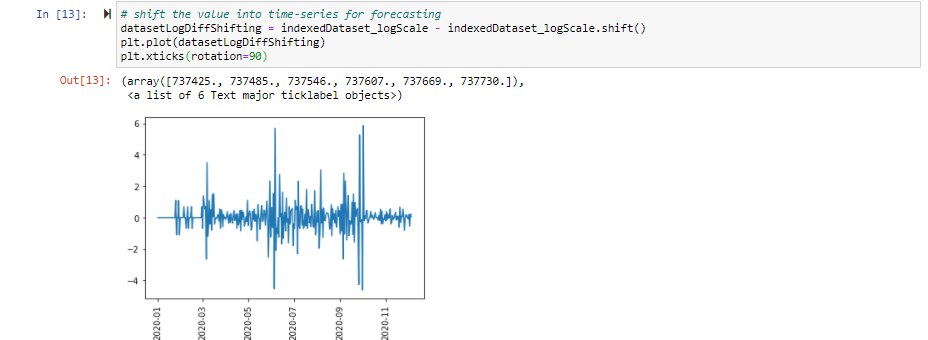
1. **Exploratory Data Analysis**



1. **Data Normalization & calculating the moving average for De-trending purpose**



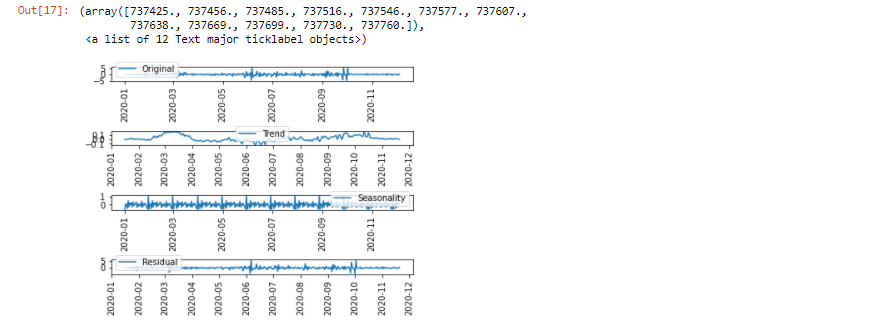
1. **De-trending work and process**
2. **De-trending**



Description: the purpose of de-trending is to identify cyclical and other pattern. Detrending shows a different aspect of time series data by removing deterministic and stochastic trends.

1. **Checking to ensure de-trended data**

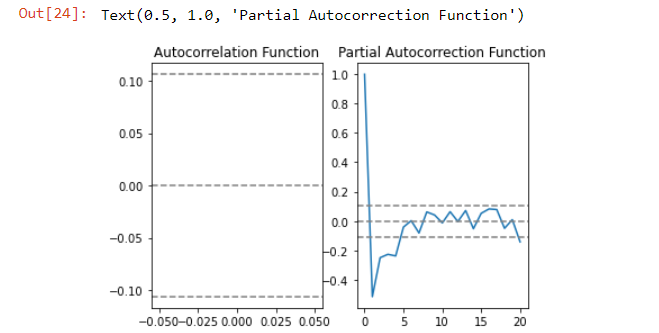




Description: as shown by the Original and Trend graph, the uptrend pattern as presented by item (3). Exploratory Data Analysis have been removed for time-series forecasting

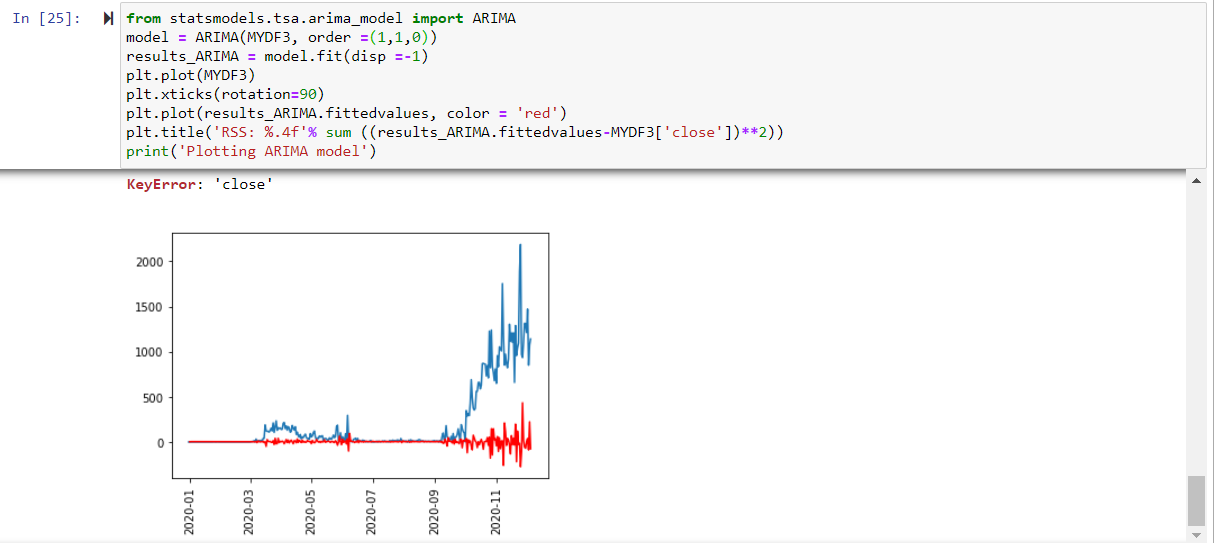
1. **ACF (Auto-correction Function) & PACF (Partial Auto-Correlation Function)**
   1. Using ACF to observe the correlation between the current spot and previous spot
   2. Using PACF to remove the influence of the day before yesterday to observe the correlation between two time spots given that we consider both observation is correlated to other time spots.

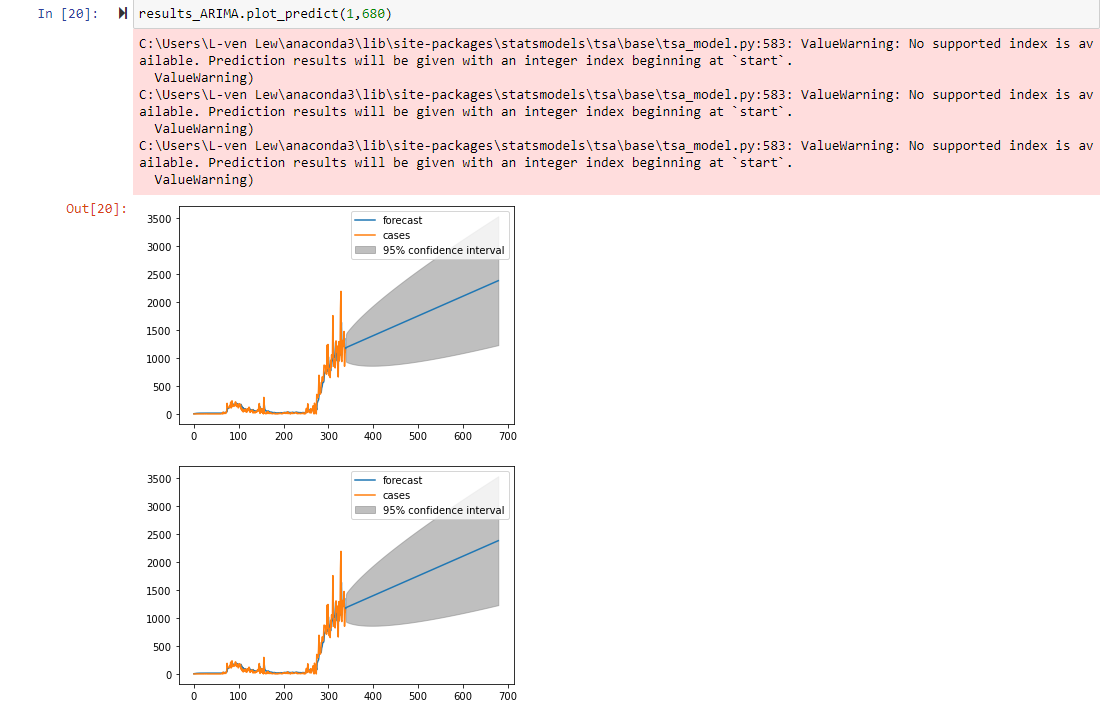




Description:

1. To obtain the p and q value, refer to the value where the graph cut off
2. Partial Auto-correction function - AR (p): Auto-regression. A model that uses the dependent relationship between an observation and some number of lagged observations.
3. Integrated (d). The use of differencing of raw observations (d values e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
4. ACF – Autocorrelation function - Moving Average (q). A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.
5. p = 1, and q = 0
6. Note that the PACF plot has a significant spike only at lag 1,
7. Meaning that higher-order partial autocorrelations are effectively explained by the lag-1 autocorrelation.
8. **Combining Auto Regression (AR) and Moving Average (MA) = ARIMA for the next year**





Result/ Output : The Time-series forecasting shows that the number of cases is expected to be on uptrend with the highest and lowest number demonstrated in the grey region.