UNIVERSITY OF TORONTO

Content Table:

1. Data Cleaning
2. Exploratory Analysis
3. Model Selection & Implementation
4. Second Dataset Visualization
5. Proposed Policy & Guidance

Name- TUHIN RANJAN
Program-MEng MIE

UNIVERSITY OF TORONTO

Data Cleaning

- *Used Dataset* : 'time_series_covid19_confirmed_global.csv' to analyze the causes for COVID-19 cases and its projection.
- *Selected region* : **USA** as a whole country.

Selecting USA as Country for analysis purposes and modelling

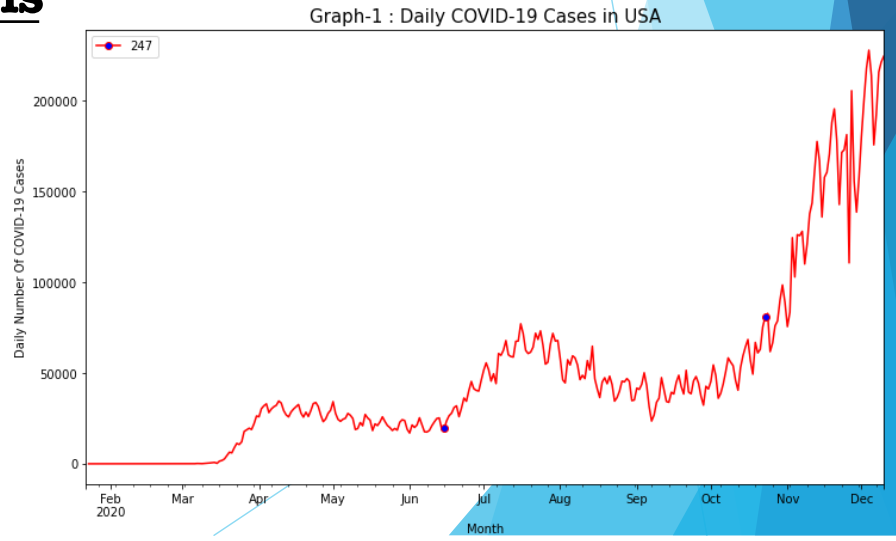
```
df_USA = df_raw[df_raw['Country/Region'] == 'US']
df_USA.head()
```

Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	12/1/20	12/2/20	12/3/20	12/4/20	12/5/20	1
NaN	US	40.0	-100.0	1	1	2	2	5	5	...	13721822	13921929	14139609	14367437	14581318	14

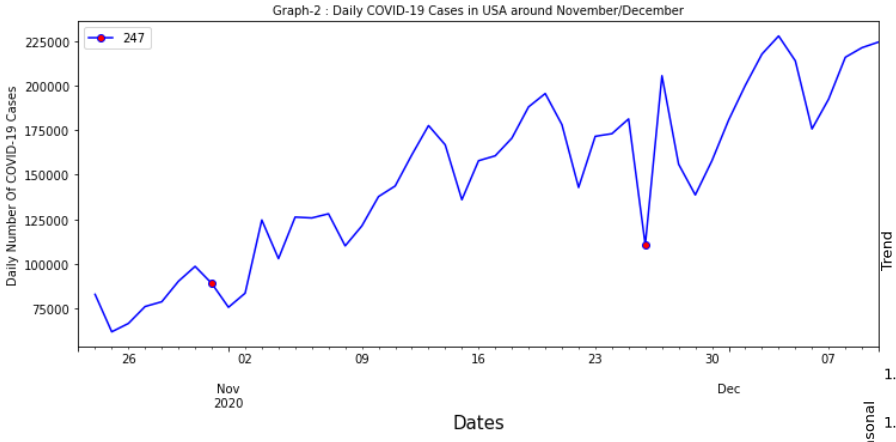
ws × 328 columns

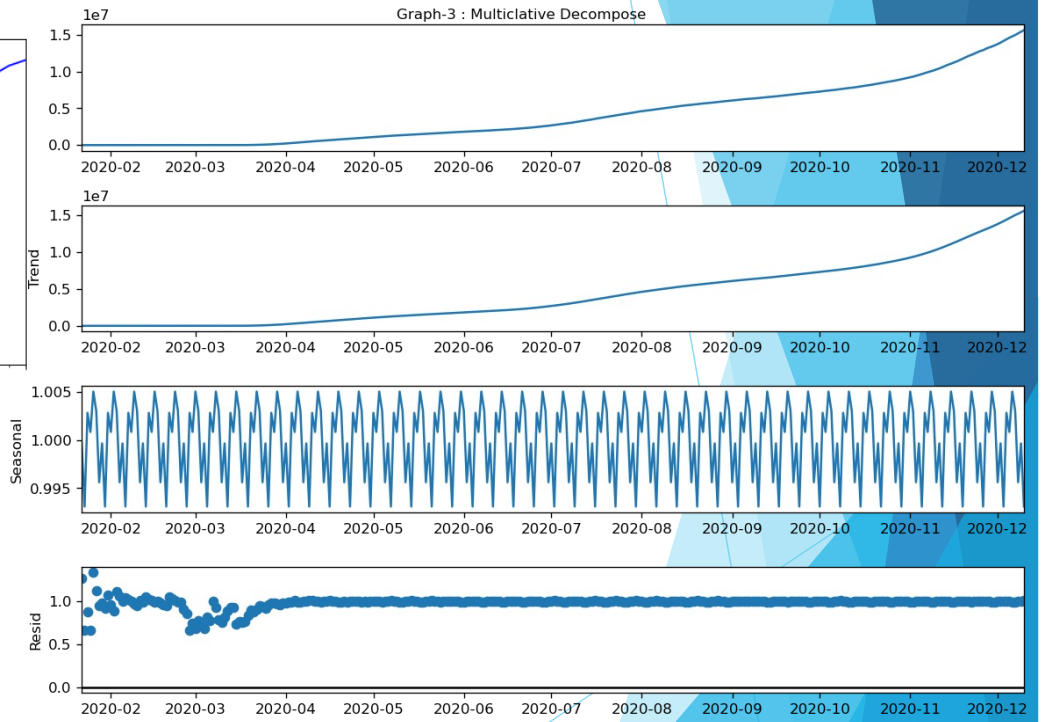
Exploratory Analysis

- Graph-1 shows how the COVID-19 cases in USA is getting into the second wave of infection.
- Graph-2 shows the detailed expansion of daily cases in the month of November and December.
  - The two red points (1) 31st October,2020 : Halloween (2) 26th November,2020 : Thanksgiving shows one of the reason for high increase of daily cases specifically after Thanks giving holiday due to more people gathering. It will be discussed in later part.



UNIVERSITY OF TORONTO



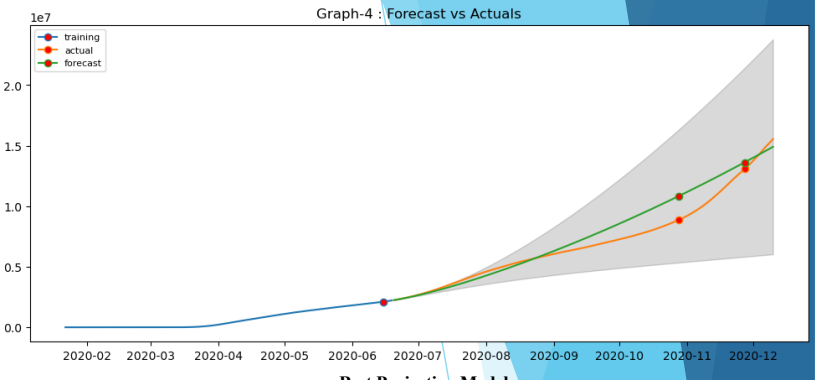


- Graph-3 shows the plots of the data being decomposed into components, nature selected- Multiclative due to non-linearty of the original curve:
  - Trend is increasing with time.
  - Seasonal data is cyclic.
  - Residual is pretty much constant.
  - Thus, the decomposition of the data shows that model-ARIMA can be used for projecting the COVID-19 case.

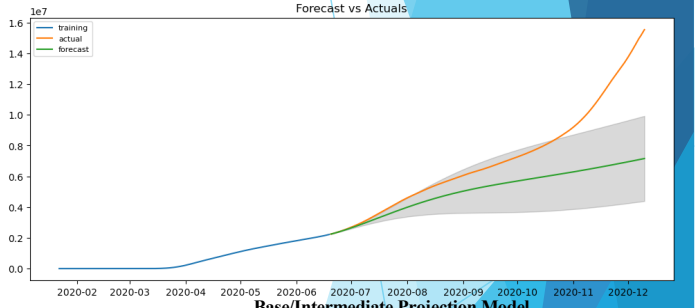
UNIVERSITY OF TORONTO

Model Preparation

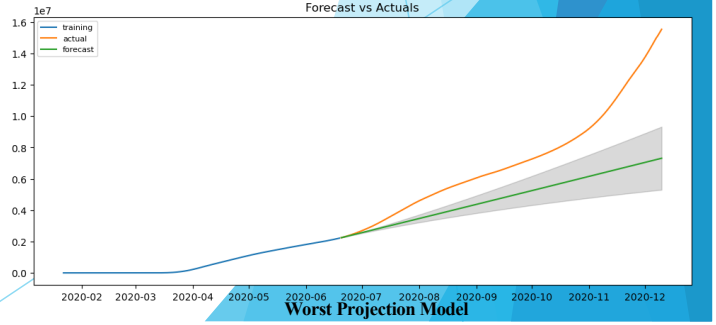
- Selected Model- **ARIMA** (AutoRegressive Integrated Moving Average', is a forecasting algorithm based on the idea that the information in the past values of the time series can alone be used to predict the future values).
- Best Case Projection Model - The projection is quite accurate and all the projection data comes under 95% conf band.
- Base Case Projection Model - The projection till November looks great as in lie around 95% conf band but after that it is way over the band.
- Worst Case Projection Model - The projected data for the COVID-19 cases in this model looks horrible as it does not even fit under 95% conf band.



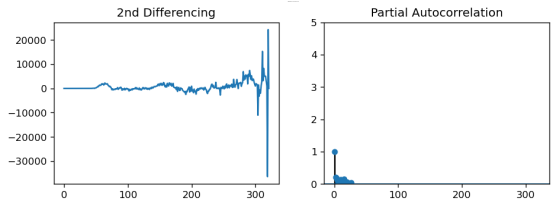
Best Projection Model



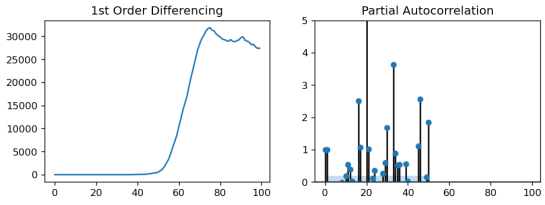
Base/Intermediate Projection Model



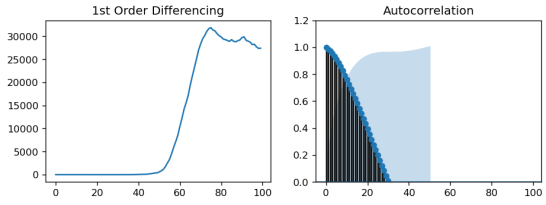
Worst Projection Model



Analysis for selecting parameter, d=2



PACF Analysis for selecting parameter, q=2

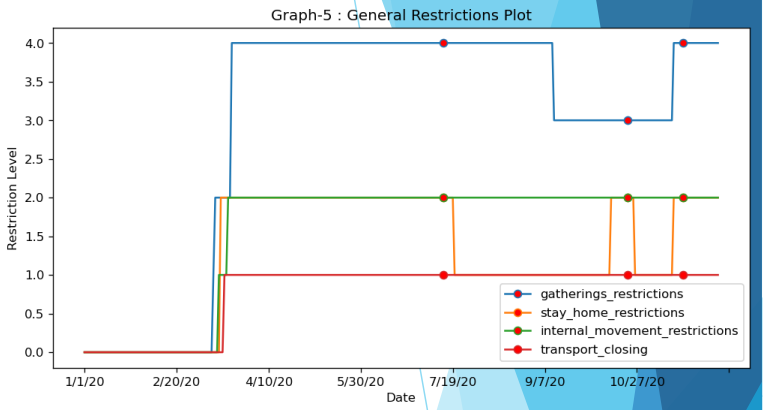


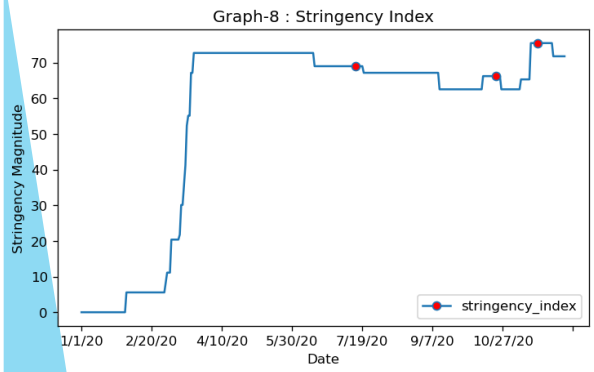
ACF Analysis for selecting parameter, p=1

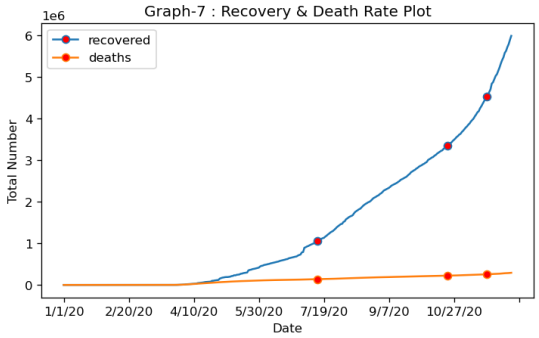
UNIVERSITY OF TORONTO

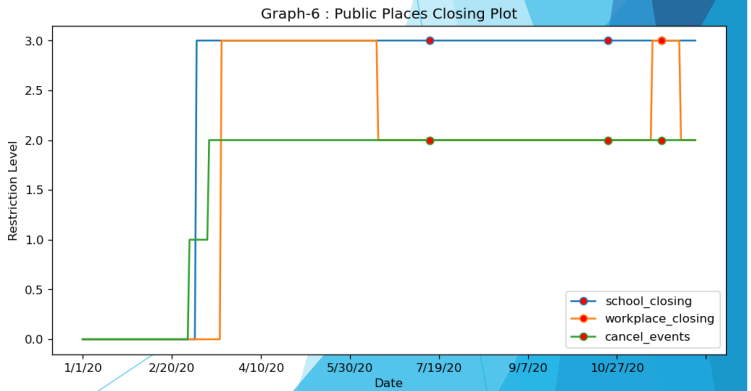
Second Dataset Visualization

1. Selected dataset : “data-1.csv”/Selected location : USA
2. The three red marker on the graphs for various restrictions referred to the starting of major spikes in increase of the COVID-19 cases across USA.
3. Referring to Graph-1 and Graph-4, during late October and November, there is sudden increase in the number of cases and the reason we can see on the Graph-5, as the restriction level for the gathering were decreased during these period.
4. Further the month of July,late October and late November are being considered in account because of the holiday seasons as in July, Independence day celebration, October-Halloween and November-ThanksGiving.
5. Workplace closing restriction level has been decrease to level-2 in between June, and it can be seen during early July, there is sudden increase in cases.
6. The recovery rate is huge as comparison to death rate, so there is misconception across the people that the COVID-19 virus is not deadly at all.
7. Many countries have the stringency index over 95% during first wave itself but USA after being badly hit by the second wave, the index is still approx. 75%









UNIVERSITY OF TORONTO

Proposed Policy & Guidance

Conclusions:

- Referring Graph-4 for the Forecast & Actual cases, the red marker around October end, shows the difference in actual and forecast value. The forecast value is showing more than the actual one. Under such scenario, the policy that government should follow is to decrease the restrictions as well as stringency magnitude which can easily be reflected on Graph-5 & 8.
- Referring Graph-4 for the Forecast & Actual cases, the red marker around November end, shows the approximately same values for the cases. Under such circumstances, the policy that government should follow is to keep the same level of the restriction and observe for another 7-14 days period after that. If there is decrease in number of cases, then decrease the restriction level after some say 7 days time.
- If in case the forecast value is showing less than the actual one, then increase the restriction level from immediate effect by increasing the stringency index.

Potential Guidance to Control the Pandemic Situation:

- For USA or any other country Government :
  - Increase the stringency index as soon as possible to approx. 95% level. This will ensure less pressure on the medical services.
  - Increase the lockdowns/restriction in working areas. Focus more on the work from home culture as new guidelines for people.
  - Hold down the internal mode of transportation like flights or metro for 2-3 days period to break the spread cycle.
  - Impose stay at home orders in the heavily affected areas from COVID-19.

- For General Public
  - Avoid social gathering irrespective of the holiday seasons.
  - COVID-19 is a deadly virus and still we don't know the exact reasons why it is affecting one person and not the other one. So the recovery rate should not be manipulated in wrong direction and people should take precaution like wearing mask and avoid gatherings.