

Spread Visualization and Prediction of the COVID-19 Using Machine Learning and Deep learning.



By
Mohammad Faisal Danish


Objective

The current destructive pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first reported in Wuhan, China, in December 2019. The outbreak has affected millions of people around the world and the number of infections and mortalities has been growing at an alarming rate. In such a situation, forecasting and proper study of the pattern of disease spread can inspire design better strategies to make more efficient decisions. Moreover, such studies play an important role in achieving accurate predictions.

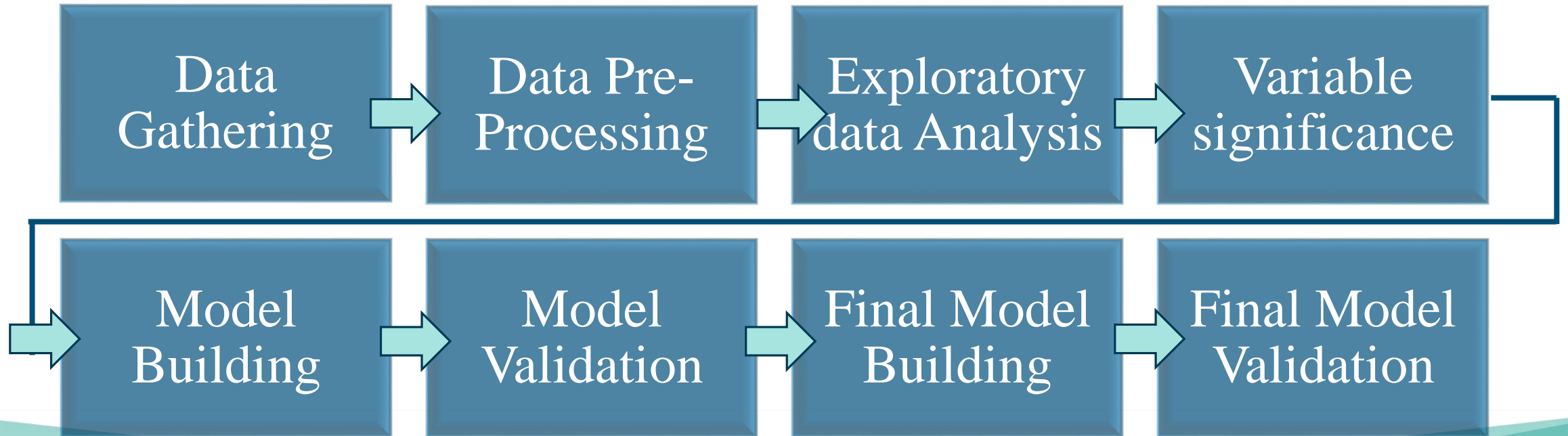
Problem Statement

- Lets use Problem Solving Formula that is 6W+2H.
- What , Why ,When , Who , Where , Which , How , How Much.
- **What :-** Have to do Spread Analysis and Visualization of Covid-19 virus and Forecasting the future spread of virus. On Global data and India data
- **Why :-** For applying better strategies to handle the spread of virus. And taking measures to stop the spread of virus.
- **Who :-** This Project Handle by Mohammad Faisal Danish (Jr. Data Scientist) from Data Analytics Department.
- **When :-** From 15/07/2021 to 15/08/2021

Problem Statement

- What , Why ,When , Who , Where , Which , How , How Much.
 - **Where :-** Imarticus Learning , Pune.
 - **Which :-** Agile Methodology and “Ask , Prepare , Process , Analyze , Share , Act” Approach.
 - **How :-** Please Refer the figure on next slide.
 - **How Much :-** it will cost , One Month of time and One Manpower for fully deployed for this project.
- 

HOW (To Do This.)



Ask phase (I am interested to finding below question's Answered.)

- How Covid-19 virus spread across the World?
- What is the recovery rate and Mortality of the virus?
- Which country is most affected by Covid-19. and what's the confirmed cases and recovery rate and mortality rate?
- Deaths across the world due to Covid-19?
- how much Confirmed, Active, Recovered and Death cases in 10 Most-affected Countries?
- which model is best for forecasting the covid-19 deaths and confirmed cases?
- what's the spread rate of covid-19 is top 5 countries?
- how many active cases in India and what's the recovery rate , and mortality rate?
- What's the Forecasting of India covid-19 confirmed cases and deaths?

Prepare Phase

- Where is the Data Located and how data is Organized?
 - The Data is collected from multiple sources.
 - - COVID-19 datasets from Johns Hopkins (data Updated on daily Basis)
 - - World-happiness dataset from Kaggle.
 - - Population by country - 2020 from Kaggle.
 - - World Health Organization (WHO)
 - The data Gathered by this sources are not organized in required format.
 - It is safe to assume that the data is unbiased and credible. And Privacy is protected by country/region and Date for confirmed , recovered and deaths cases.
- Our data is sufficient for our objective. And sufficient to give the answer to our questions.
- Data is not in a standard format it required close watch for organizing the data.

Process Phase:-

- For Processing of our Data I used Python , Excel and Tableau.
- After gathering the Data (in 3 CSV files) for confirmed , recovered and Deaths. We observed that we have some common columns in 3 tables.
- So, we used common columns for merging the data and then create a one full uncleaned csv fil from 3 csv files. And converting wide data into long Data.
- All details about the columns names are explain in a data Dictionary please refer it.
- After Checking NULL values we observed that there are NA in state(96%) columns and also NA in Latitude and in Longitude up to 0.00722%.(mostly missing values from 2 countries like CHINA and CANADA and state/ province are Unknown and Repatriated Travelers respectively.)
- We decide to impute NULL value by blank in state and decide to remove the records which have NULL values in Latitude and Longitude.

Process Phase:-

- We observed that we have some ship names in country columns because there are 2 ship names Diamond prince and MS zandan. (we know this because we found it in news) # remove rows in which 'Country/Region' is a ship name.
- We arrange the data from 22 January 2020 to 7 July 2021.
- We observed that the date is not in a standard format. That's why we have to fix it.
- We scrap the WHO side for region. African Region (AFRO) , Region of the Americas (PAHO), South-East Asia Region (SEARO) , European Region (EURO) , Eastern Mediterranean Region (EMRO) , Western Pacific Region (WPRO). We added region into uncleaned table with respect to country name. and save the file as a cleaned data into CSV file.
- After doing all this we observed that the data have redundancy issue that's why we have to split the data into multiple tables.
- We split the data into 8 CSV files for better understanding please refer the data dictionary for that.
- we create a new columns like active cases and new confirmed cases as well as new recovered and new deaths columns. And many more.

Process Phase:- Details About Data.

	Date	Country/Region	Confirmed	Deaths	Recovered	Active	New confirmed	New deaths	New recovered	WHO region
34403	2020-07-23	West Bank and Gaza	9744	67	2720	6957	346	1	770	EMRO
34404	2020-07-23	Western Sahara	10	1	8	1	0	0	0	AFRO
34405	2020-07-23	Yemen	1654	461	762	431	14	3	11	EMRO
34406	2020-07-23	Zambia	3789	134	1677	1978	206	6	0	AFRO
34407	2020-07-23	Zimbabwe	2124	28	510	1586	90	2	0	AFRO

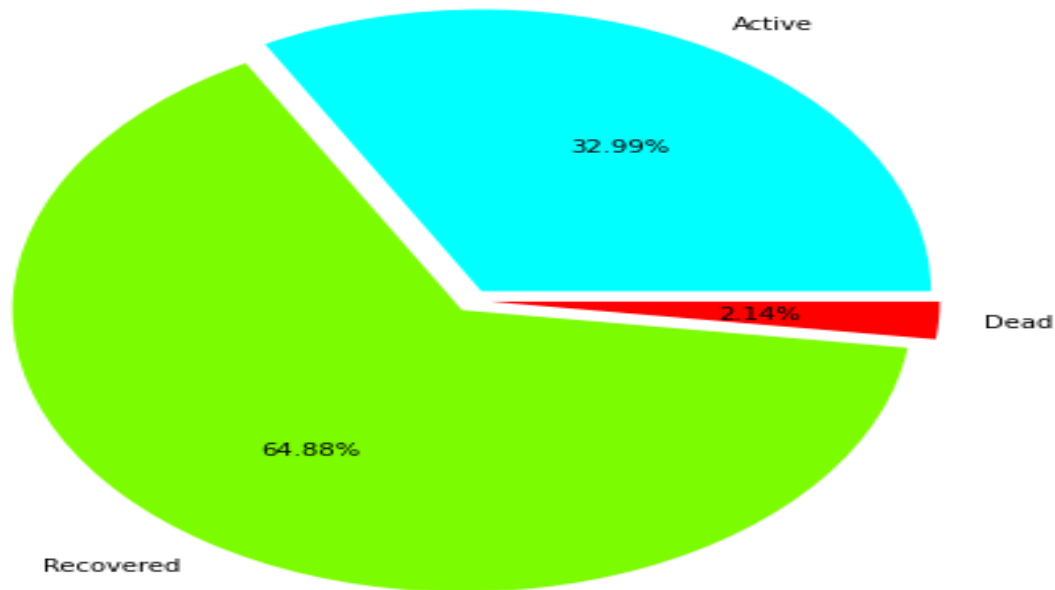
	Country/Region	Confirmed	Deaths	Recovered	Active	New confirmed	New deaths	New recovered	Recovery rate(per 100)	Mortality rate(per 100)	WHO region
0	Afghanistan	35928	1211	24550	10167	201	21	626	68.33	3.37	EMRO
1	Albania	4466	123	2523	1820	108	3	60	56.49	2.75	EURO
2	Algeria	25484	1124	17369	6991	612	13	386	68.16	4.41	AFRO
3	Andorra	889	52	803	34	0	0	0	90.33	5.85	EURO
4	Angola	851	33	236	582	39	0	15	27.73	3.88	AFRO

	Date	Confirmed	Deaths	Recovered	Active	New confirmed	New deaths	New recovered	Recovery rate(per 100)	Mortality rate(per 100)	Number of countries
0	2020-01-22	555	17	28	510	0	0	0	5.05	3.06	6
1	2020-01-23	654	18	30	606	99	1	2	4.59	2.75	8
2	2020-01-24	941	26	36	879	287	8	6	3.83	2.76	9
3	2020-01-25	1434	42	39	1353	493	16	3	2.72	2.93	11
4	2020-01-26	2118	56	52	2010	684	14	13	2.46	2.64	13

Analyze & Share Phase :-

- Total status of the COVID-19 Cases of the world upto 22 july 2021

Total COVID-19 Cases of the world

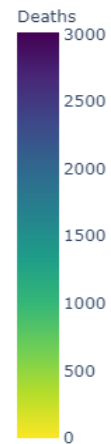
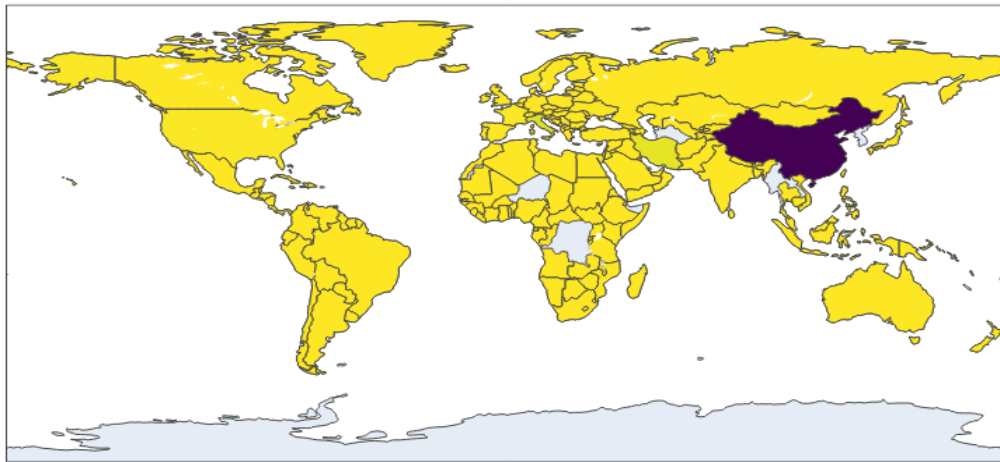


Confirmed	Recovered	Deaths	Active	Recovery rate(per 100)	Mortality rate(per 100)	Number of countries	
554	196625810.00	127565256.00	4198924.00	64861630.00	-0.65	2.14	193.00

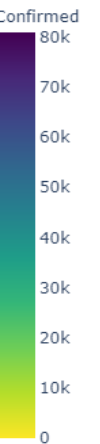
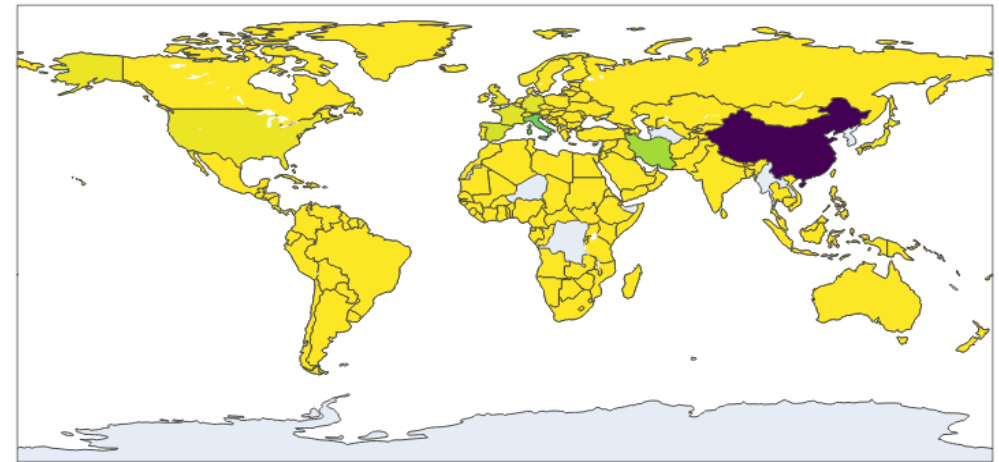
Analyze & Share Phase :-

For Details understanding of spread of covid-19 and deaths country wise please run the cell 104 and 105 cells for better understanding. It's dynamic Dashboard.

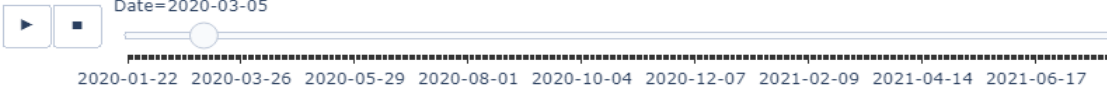
Global Spread of Coronavirus-Deaths



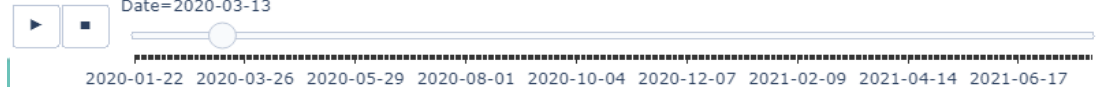
Global Spread of Coronavirus-Confirmed



Date=2020-03-05

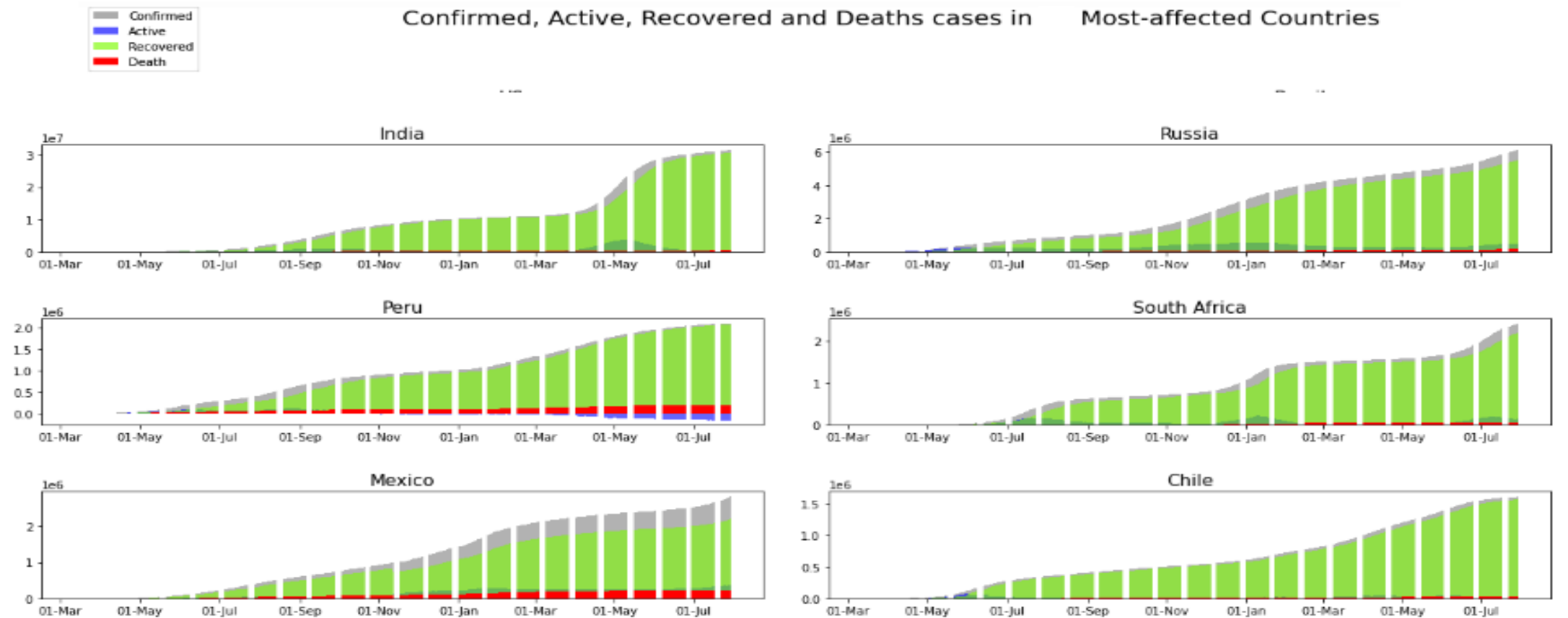


Date=2020-03-13

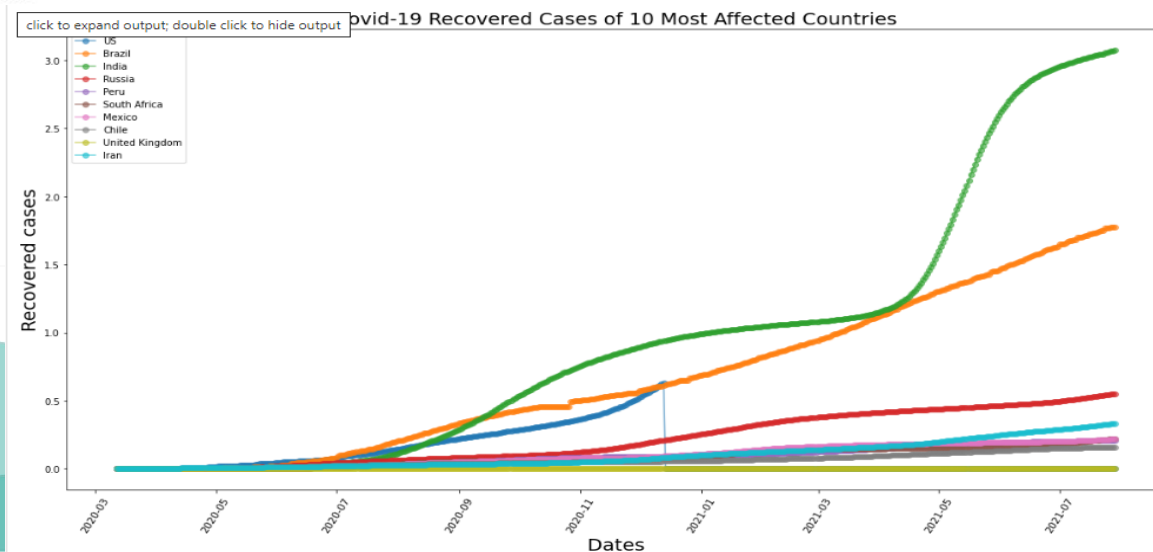
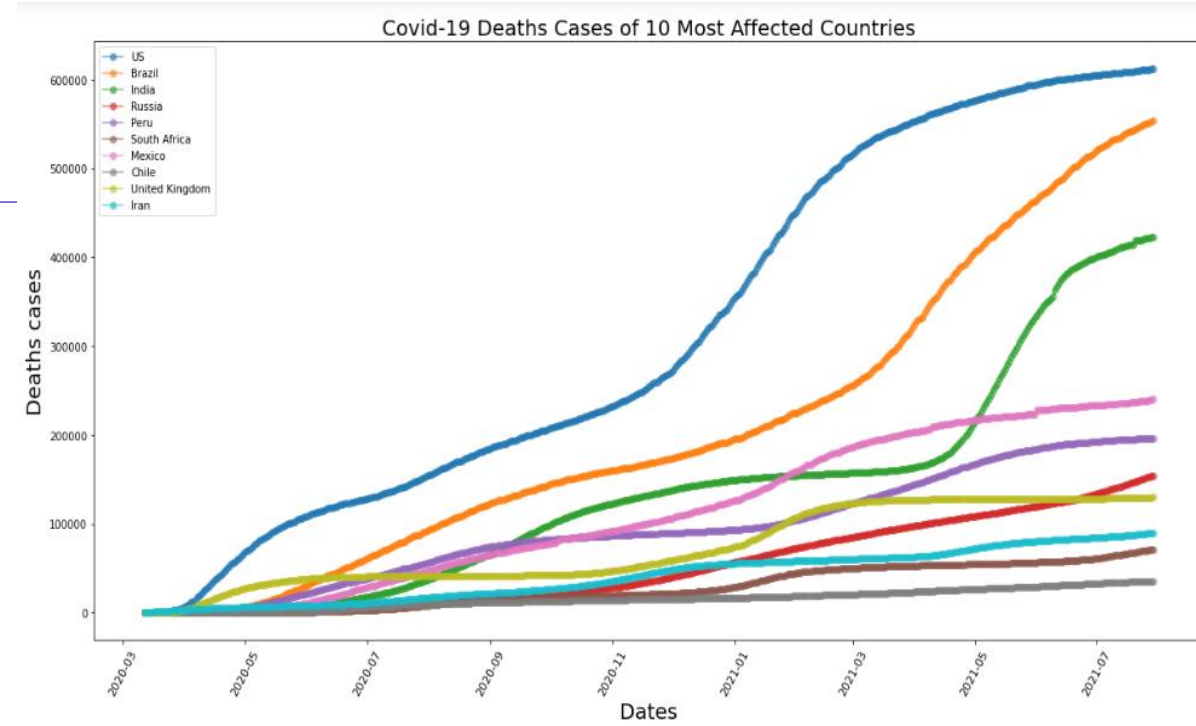
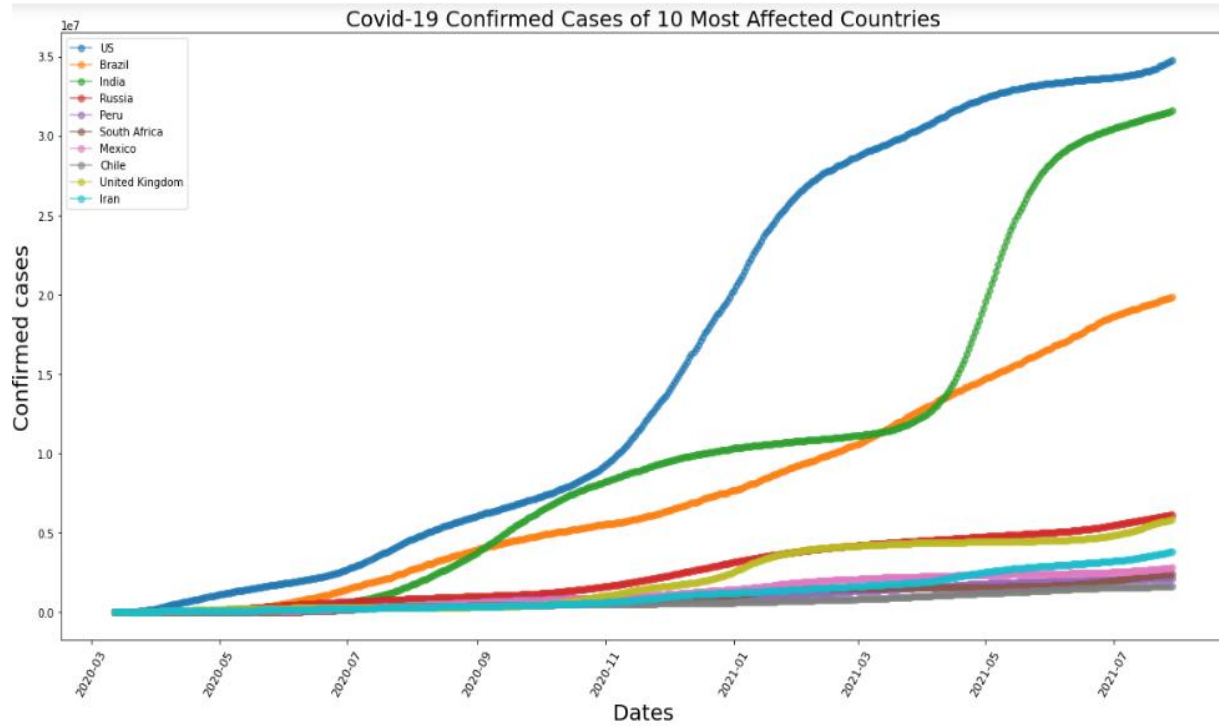


Analyze & Share Phase :-

Most Affected countries visualization of active cases(Blue color) recovered cases(Green color) and deaths(red color).

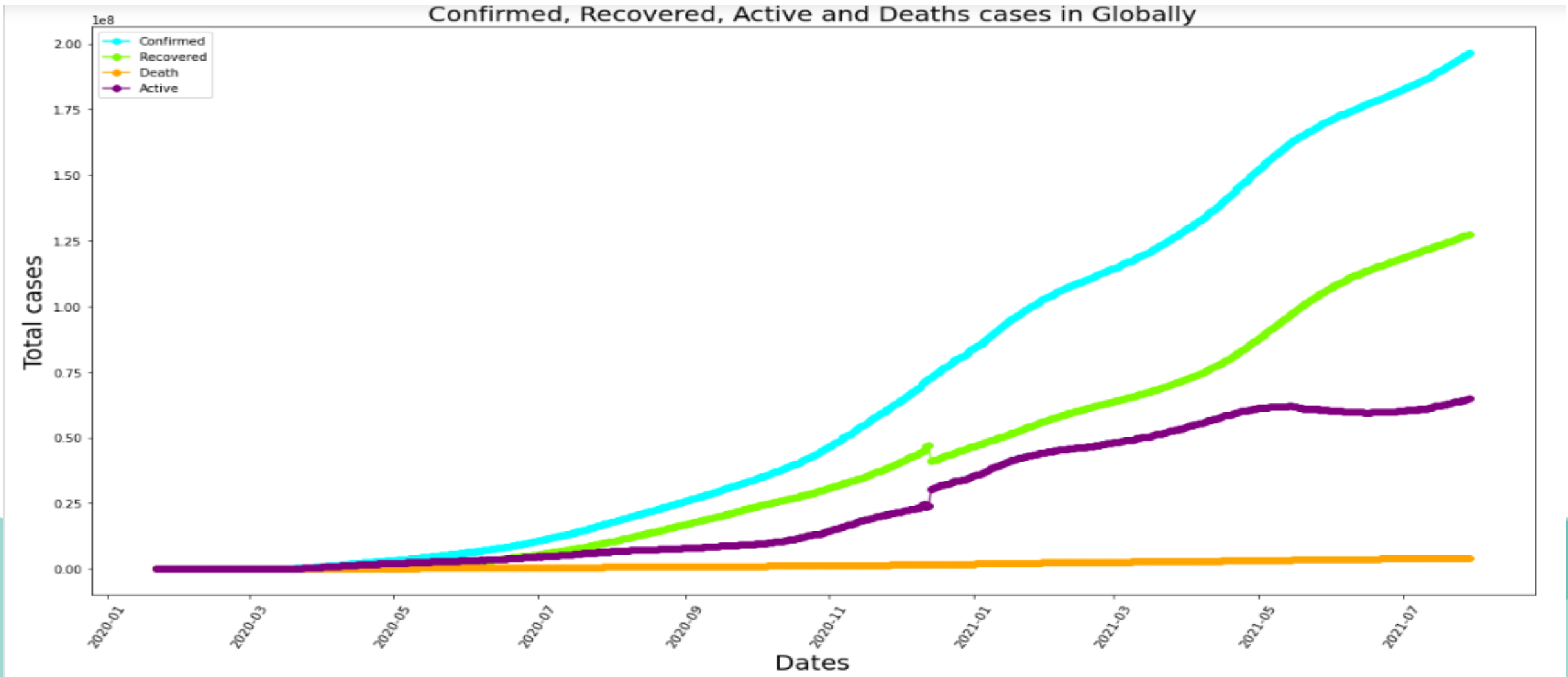


Analyze & Share Phase :-

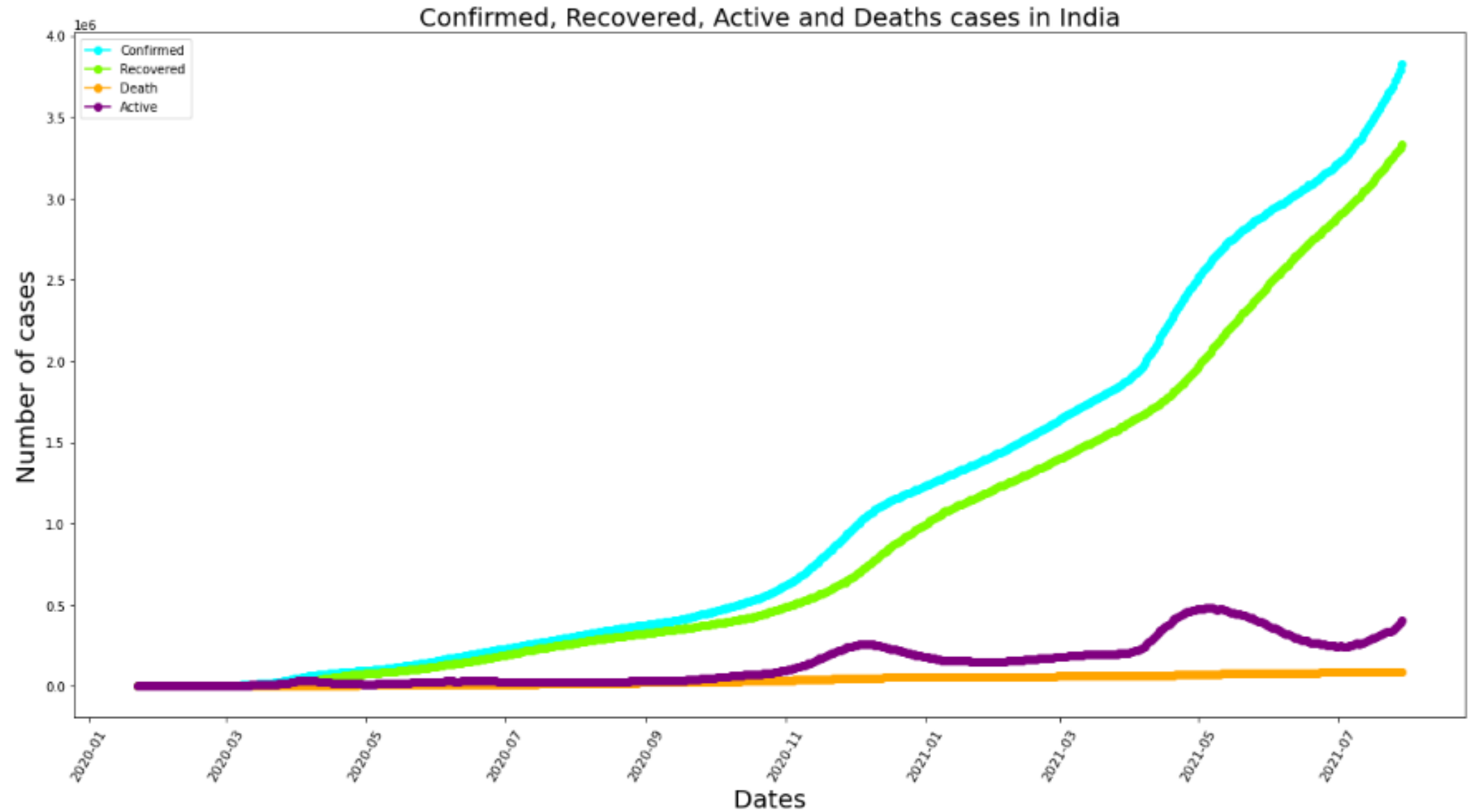


10 countries with
Confirmed cases
Recover cases
deaths

Analyze & Share Phase :-



Analyze & Share Phase :-



Analyze & Share Phase :-

Top 30 countries all Details.

click to expand output; double click to hide output										
Country/Region	Confirmed	Deaths	Recovered	Active	New confirmed	New deaths	New recovered	Recovery rate(per 100)	Mortality rate(per 100)	
0	US	34750860	612122	0	34138738	78170	321	0	0.000000	1.760000
1	India	31572344	423217	30743972	406155	44230	555	42360	-38.680000	1.340000
2	Brazil	19839369	554497	17771228	1513644	42283	1318	28696	89.580000	2.790000
3	France	6142282	111951	412373	5617958	25429	28	576	6.710000	1.820000
4	Russia	6138969	154404	5500409	484156	22720	784	20259	89.600000	2.520000
5	United Kingdom	5828316	129809	20638	5677869	30871	91	914	0.350000	2.230000
6	Turkey	5682630	51184	5443501	187945	22161	60	5483	95.790000	0.900000
7	Argentina	4905925	105113	4542904	257908	14115	291	13645	92.600000	2.140000
8	Colombia	4766829	120126	4557829	88874	9890	325	12483	95.620000	2.520000
9	Spain	4422291	81442	150376	4190473	26689	46	0	3.400000	1.840000
10	Italy	4336906	128029	4130393	78484	6167	19	1825	95.240000	2.950000
11	Iran	3826447	90074	3329065	407308	34433	292	18902	87.000000	2.350000
12	Germany	3772326	91622	3649670	31034	0	-87	-4320	96.750000	2.430000
13	Indonesia	3331206	90552	2686170	554484	43479	1893	45494	80.640000	2.720000
14	Poland	2882630	75257	2653529	153844	165	5	83	92.050000	2.610000
15	Mexico	2810097	239997	2192477	377623	19223	381	11186	78.020000	8.540000
16	South Africa	2422151	71431	2194762	155958	30928	1093	14268	90.610000	2.950000
17	Ukraine	2330440	55489	2253803	21148	1396	40	1085	96.710000	2.380000
18	Peru	2108595	196214	2075381	-162980	722	76	1109	96.420000	9.310000
19	Netherlands	1888741	18102	28149	1842490	4543	4	75	1.490000	0.960000
20	Czechia	1673219	30363	1639849	3007	202	1	151	96.010000	1.810000
21	Chile	1613288	35295	1568294	9699	1371	119	1058	97.210000	2.190000
22	Iraq	1603787	18533	1447160	138094	13259	49	9401	90.230000	1.160000
23	Philippines	1572287	27577	1488437	56273	5620	176	3723	94.670000	1.750000
24	Canada	1437111	26544	0	1410567	912	14	0	0.000000	1.850000
25	Bangladesh	1226253	20255	1050220	155778	15271	239	14336	85.640000	1.650000
26	Belgium	1122951	25235	0	1097716	1862	4	0	0.000000	2.250000
27	Sweden	1099414	14656	0	1084758	619	1	0	0.000000	1.330000
28	Romania	1082880	34275	1047528	1077	170	1	84	96.740000	3.170000
29	Malaysia	1078646	8725	890742	179179	17170	174	12930	82.580000	0.810000

Analyze & Share Phase :-

After visualization, we investigated data modeling and prediction based on univariate time series, using Linear regression, Support vector machine, Random forests, XGBoost, Multilayer perceptron (MLP), and a recurrent neural network, Long Short-Term Memory network (LSTM-RNN) to forecast the number of confirmed cases and deaths in the world and some other countries such as India. Some of our results are summarized in the following tables:

Table 1: Prediction errors of total confirmed cases of the world

Regressor	RMSE
Support Vector Machine	68123619.22
Random Forest Regressor	34229279.33
XGBoost	68123619.22

Table 2: Prediction errors of total Deaths of the world

Regressor	RMSE
Support Vector Machine	1289165.65
Linear Regression	1446097.27

Table 3: Accuracy of predicting the total cases of India using MLP and LSTM-RNN

Neural Network	MAPE%	ACCURECY(Percent)
MLP	0.07492	99.99925
LSTM-RNN	0.123195	99.998768

Table 3: Accuracy of predicting the Death cases of India using MLP and LSTM-RNN

Neural Network	MAPE	ACCURECY(Percent)
MLP	0.68879	99.99311
LSTM-RNN	0.780975	99.99219

Analyze & Share Phase :-

Neural Network results.

Table 3: Accuracy of predicting the Confirmed cases of World using MLP and LSTM-RNN



Neural Network	MAPE	ACCURECY(Percent)
MLP	0.29073	99.9970
LSTM-RNN	0.15863	99.9984



Table 3: Accuracy of predicting the Death cases of World using MLP and LSTM-RNN

Neural Network	MAPE	ACCURECY(Percent)
MLP	0.08641	99.99913
LSTM-RNN	0.20931	99.99790

Analyze & Share Phase :-

Predicted confirmed cases by using MLP and LSTM(RNN).

CONFIRMED CASES BY MLP of INDIA

	Confirmed	Confirmed_predicted
2021-07-20	31216337	3.120539e+07
2021-07-21	31257720	3.124217e+07
2021-07-22	31293062	3.127798e+07
2021-07-23	31293062	3.131446e+07
2021-07-24	31371901	3.135093e+07
2021-07-25	31411262	3.138640e+07
2021-07-26	31440951	3.142150e+07
2021-07-27	31484605	3.145737e+07
2021-07-28	31528114	3.149296e+07
2021-07-29	31572344	3.152751e+07
2021-07-30	NaN	3.156227e+07
2021-07-31	NaN	3.159765e+07
2021-08-01	NaN	3.163274e+07
2021-08-02	NaN	3.166792e+07
2021-08-03	NaN	3.170302e+07
2021-08-04	NaN	3.173786e+07
2021-08-05	NaN	3.177266e+07

CONFIRMED CASES BY LSTM of INDIA

	Confirmed	Confirmed_predicted
2021-07-20	31216337	3.123666e+07
2021-07-21	31257720	3.127766e+07
2021-07-22	31293062	3.131917e+07
2021-07-23	31293062	3.136131e+07
2021-07-24	31371901	3.140355e+07
2021-07-25	31411262	3.144584e+07
2021-07-26	31440951	3.148833e+07
2021-07-27	31484605	3.153130e+07
2021-07-28	31528114	3.157446e+07
2021-07-29	31572344	3.161798e+07
2021-07-30	NaN	3.166242e+07
2021-07-31	NaN	3.170573e+07
2021-08-01	NaN	3.174923e+07
2021-08-02	NaN	3.179289e+07
2021-08-03	NaN	3.183672e+07
2021-08-04	NaN	3.188070e+07
2021-08-05	NaN	3.192485e+07

Analyze & Share Phase :-

India Death Cases by LSTM

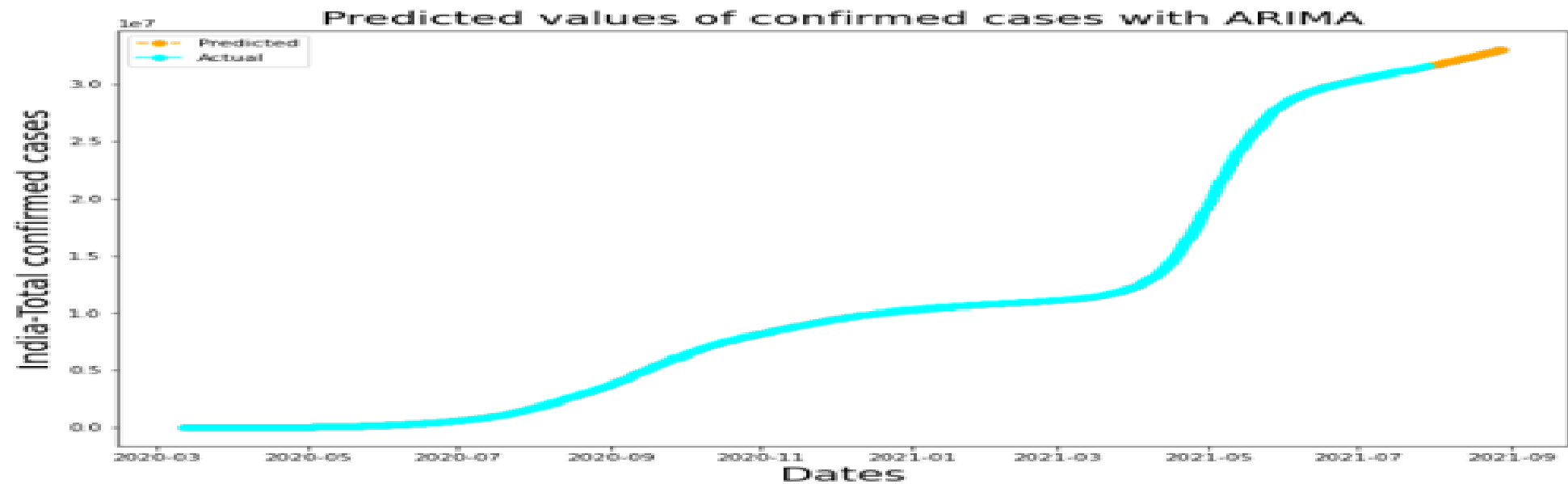
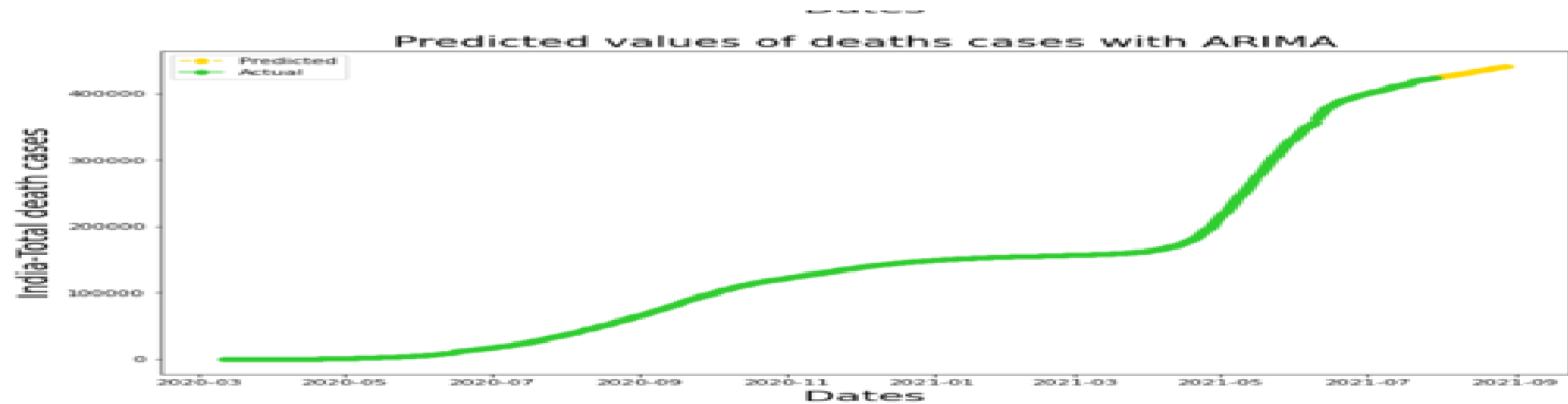
	Deaths	Deaths_predicted
2021-07-20	418480	415310.854507
2021-07-21	418987	415855.996193
2021-07-22	419470	416407.808243
2021-07-23	419470	416968.168241
2021-07-24	420551	417529.714081
2021-07-25	420967	418092.050483
2021-07-26	421382	418656.956211
2021-07-27	422022	419228.285252
2021-07-28	422662	419802.035389
2021-07-29	423217	420380.677125
2021-07-30	NaN	420971.621978
2021-07-31	NaN	421547.447339
2021-08-01	NaN	422125.693795
2021-08-02	NaN	422706.262526
2021-08-03	NaN	423288.955891
2021-08-04	NaN	423873.724481
2021-08-05	NaN	424460.716525
2021-08-06	NaN	423607.749887
2021-08-07	NaN	424024.968812
2021-08-08	NaN	424532.311767
2021-08-09	NaN	425023.744669
2021-08-10	NaN	425479.256425
2021-08-11	NaN	425846.768784
2021-08-12	NaN	426171.887273
2021-08-13	NaN	426534.409212
2021-08-14	NaN	426923.316146
2021-08-15	NaN	427188.500178

India's Death Cases by MLP

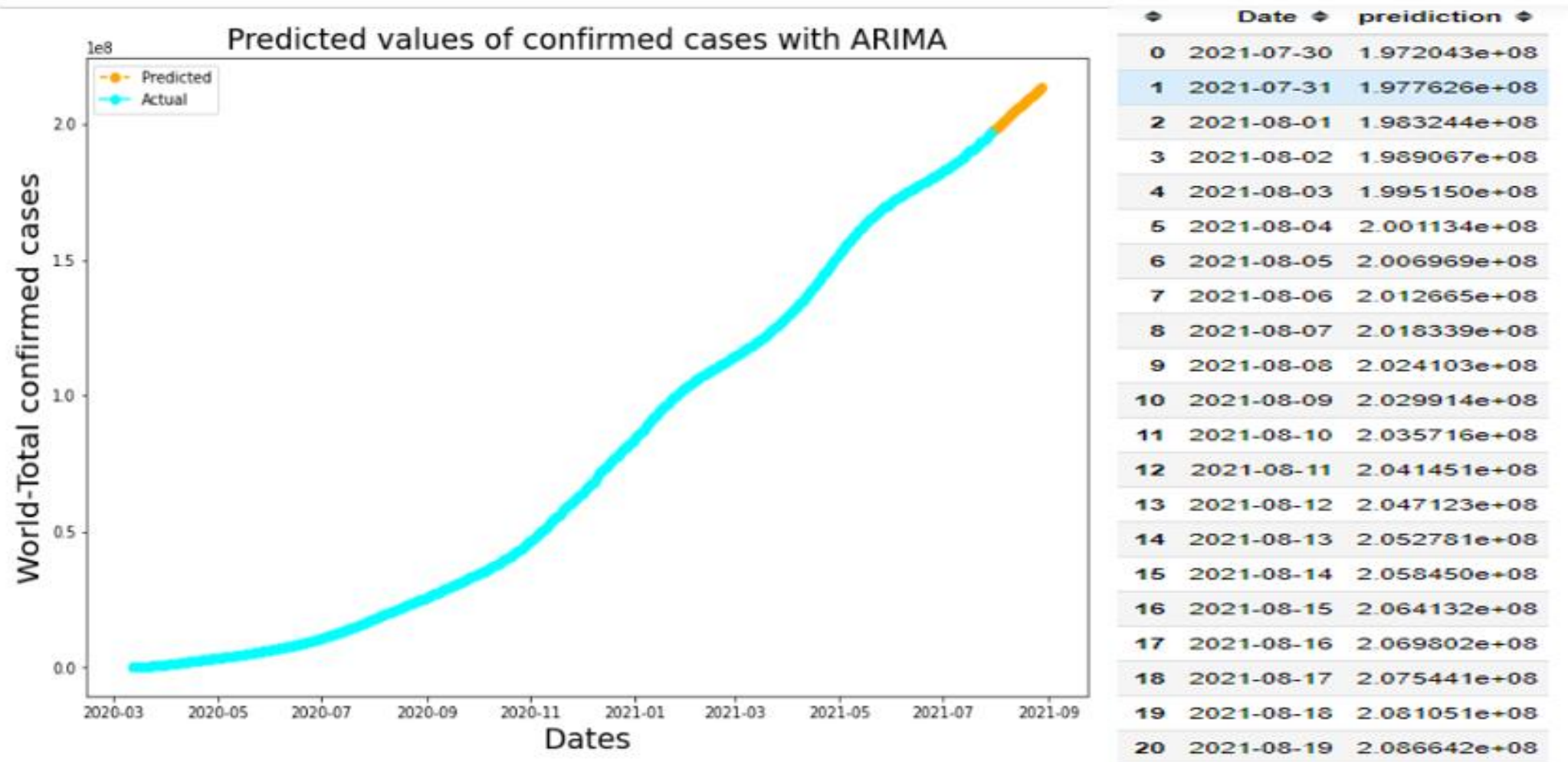
	Deaths	Deaths_predicted
2021-07-20	418480	414877.774939
2021-07-21	418987	415483.789874
2021-07-22	419470	416087.235483
2021-07-23	419470	416630.400765
2021-07-24	420551	417172.923715
2021-07-25	420967	417717.818350
2021-07-26	421382	418266.418743
2021-07-27	422022	418816.847310
2021-07-28	422662	419370.536944
2021-07-29	423217	419929.513459
2021-07-30	NaN	420500.051939
2021-07-31	NaN	421081.411231
2021-08-01	NaN	421664.401057
2021-08-02	NaN	422249.564928
2021-08-03	NaN	422839.867449
2021-08-04	NaN	423435.555672
2021-08-05	NaN	424036.876647

Analyze & Share Phase :-

ARIMA & PROPHET FORECASTING:-

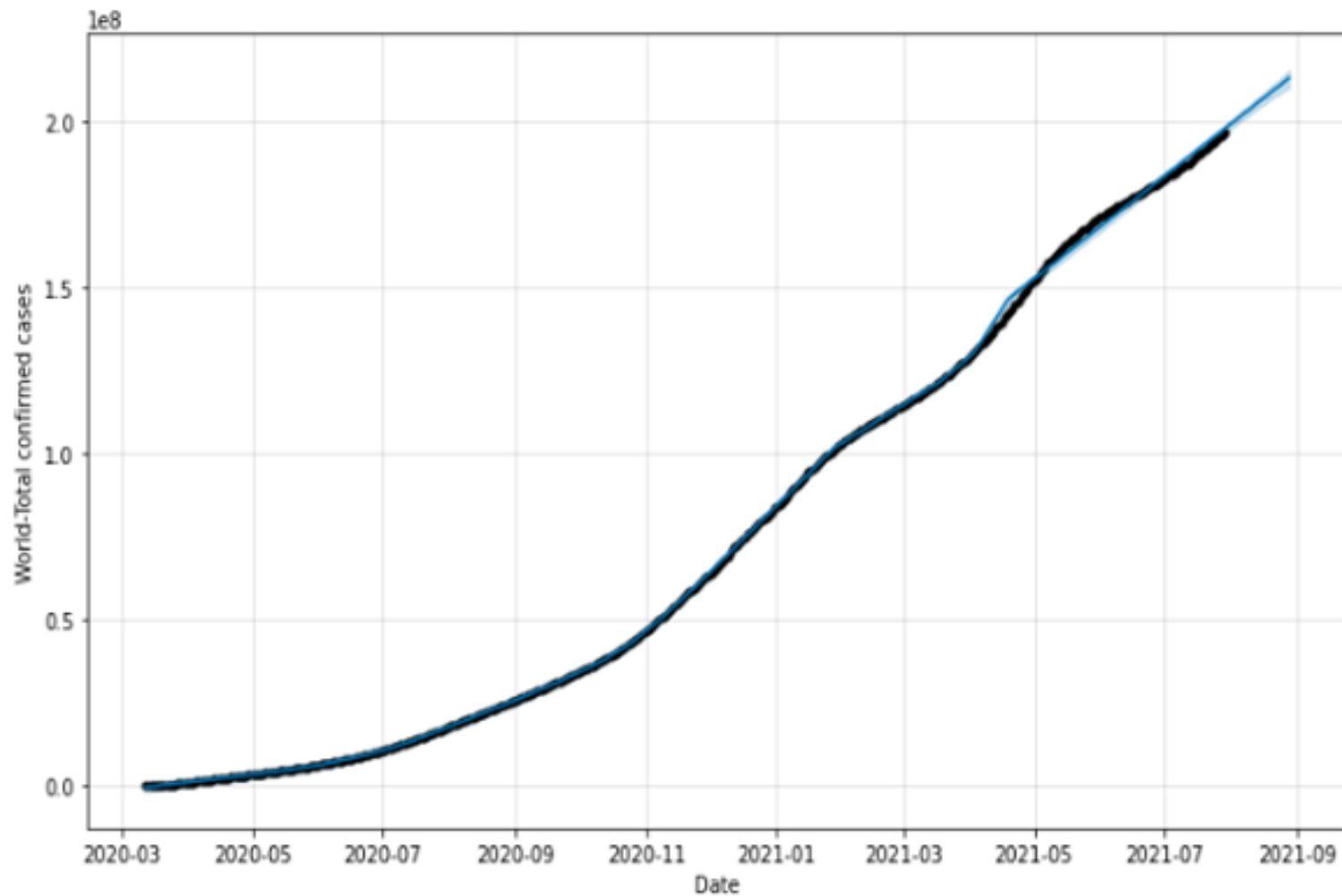


Analyze & Share Phase :-



World Confirmed cases forecasted by ARIMA

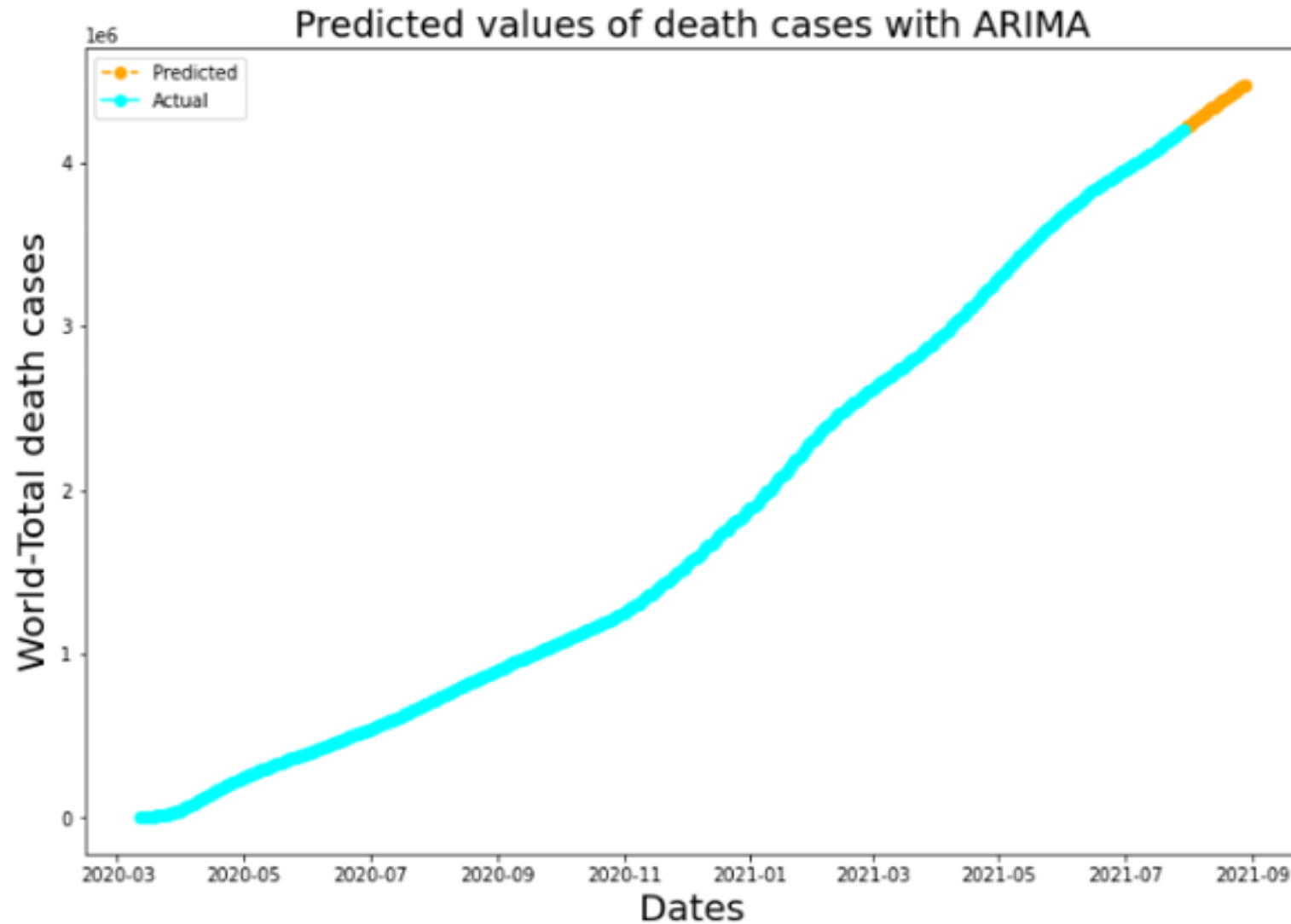
Analyze & Share Phase :-



	ds	yhat
505	2021-07-30	1.983193e+08
506	2021-07-31	1.988187e+08
507	2021-08-01	1.992549e+08
508	2021-08-02	1.996984e+08
509	2021-08-03	2.002052e+08
510	2021-08-04	2.007467e+08
511	2021-08-05	2.013106e+08
512	2021-08-06	2.018702e+08
513	2021-08-07	2.023695e+08
514	2021-08-08	2.028058e+08
515	2021-08-09	2.032493e+08
516	2021-08-10	2.037561e+08
517	2021-08-11	2.042976e+08
518	2021-08-12	2.048615e+08
519	2021-08-13	2.054211e+08
520	2021-08-14	2.059204e+08
521	2021-08-15	2.063567e+08
522	2021-08-16	2.068002e+08
523	2021-08-17	2.073070e+08
524	2021-08-18	2.078485e+08
525	2021-08-19	2.084124e+08

World Confirmed Cases Forecasted by Prophet

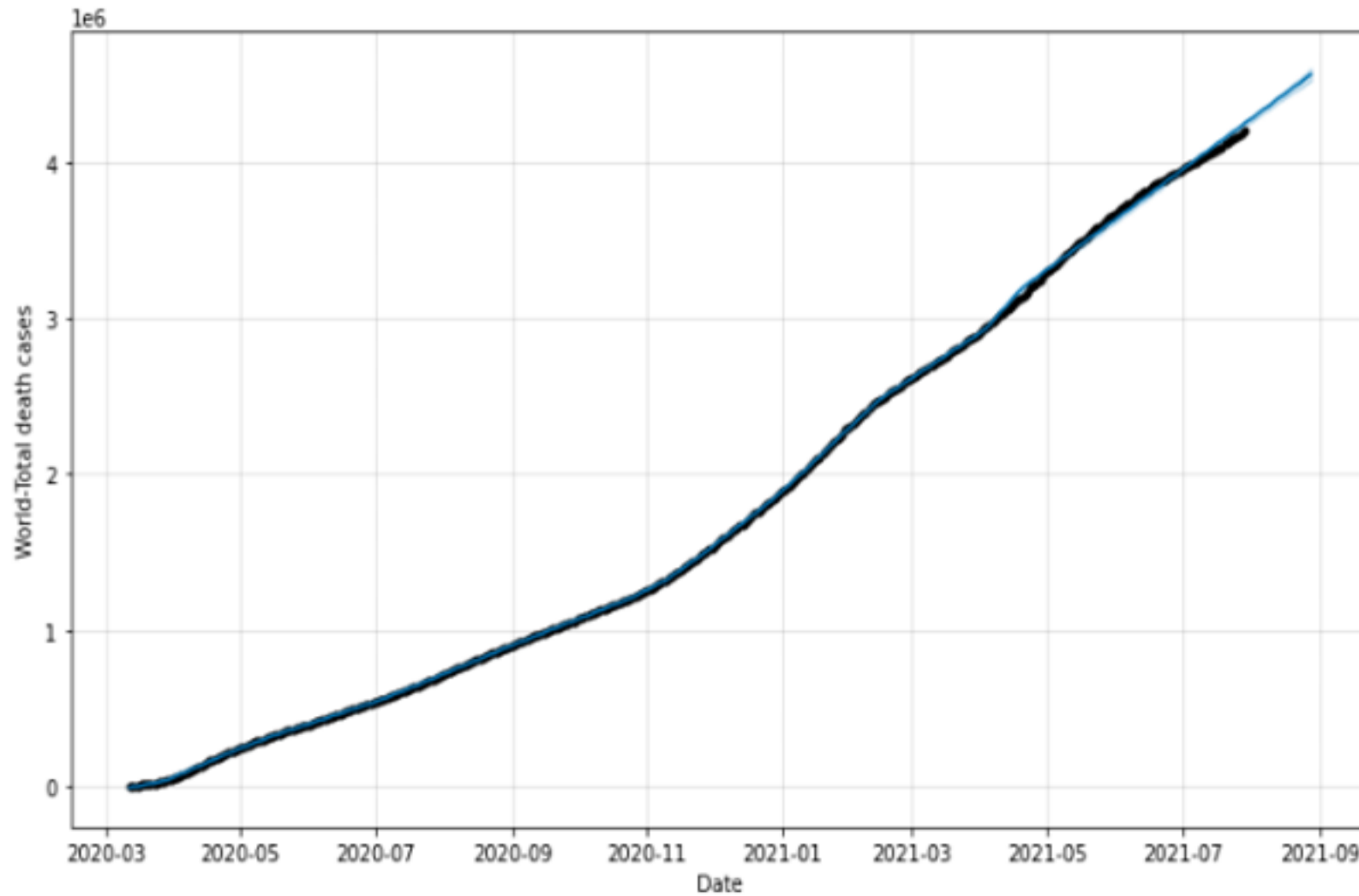
Analyze & Share Phase :-



	Date	preidiction
0	2021-07-30	4.209391e+06
1	2021-07-31	4.218767e+06
2	2021-08-01	4.228138e+06
3	2021-08-02	4.237550e+06
4	2021-08-03	4.247410e+06
5	2021-08-04	4.257362e+06
6	2021-08-05	4.267028e+06
7	2021-08-06	4.276465e+06
8	2021-08-07	4.285786e+06
9	2021-08-08	4.295150e+06
10	2021-08-09	4.304585e+06
11	2021-08-10	4.313991e+06
12	2021-08-11	4.323304e+06
13	2021-08-12	4.332522e+06
14	2021-08-13	4.341689e+06
15	2021-08-14	4.350845e+06
16	2021-08-15	4.359989e+06
17	2021-08-16	4.369100e+06
18	2021-08-17	4.378162e+06
19	2021-08-18	4.387176e+06
20	2021-08-19	4.396157e+06

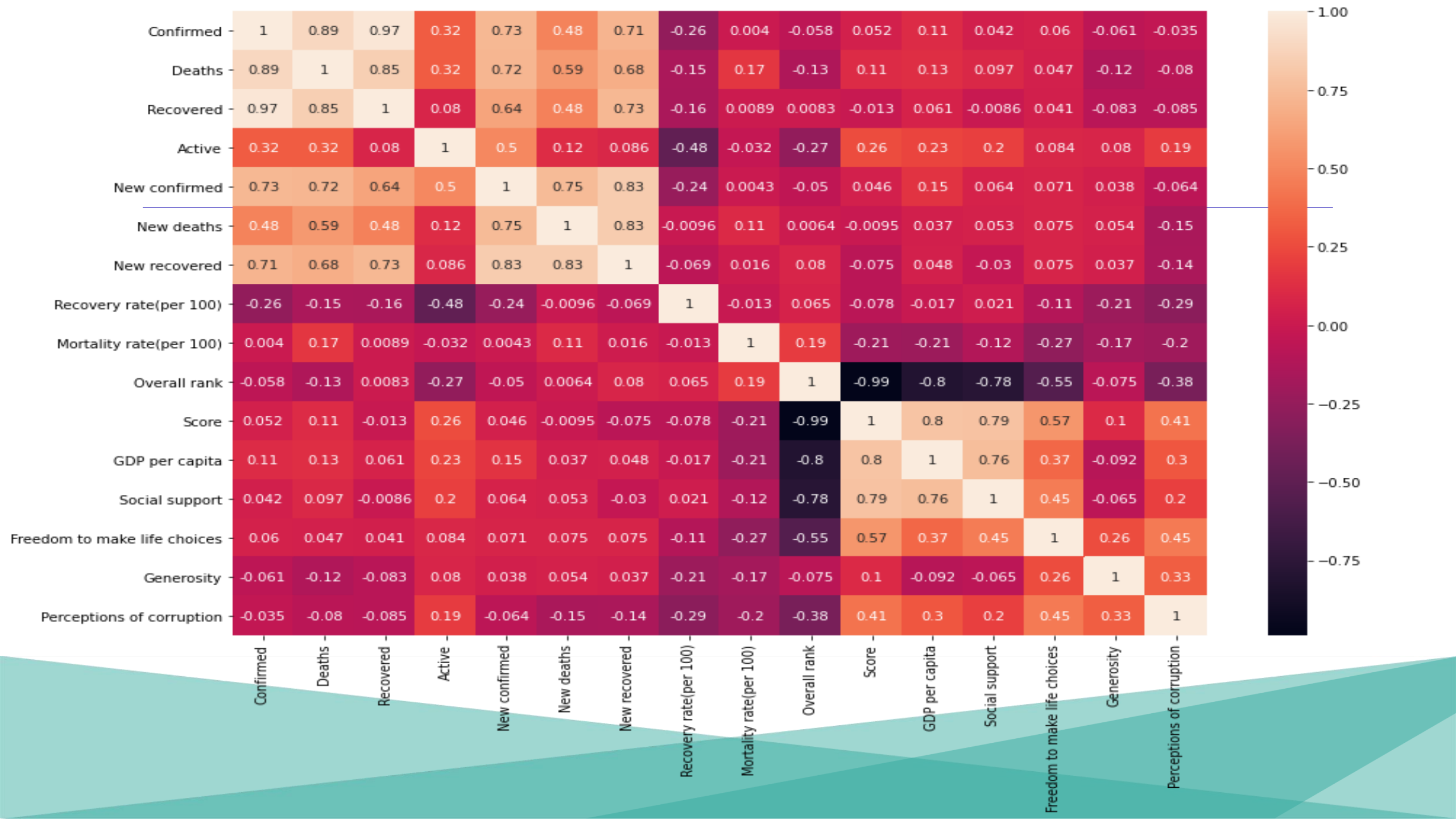
Analyze & Share Phase :-

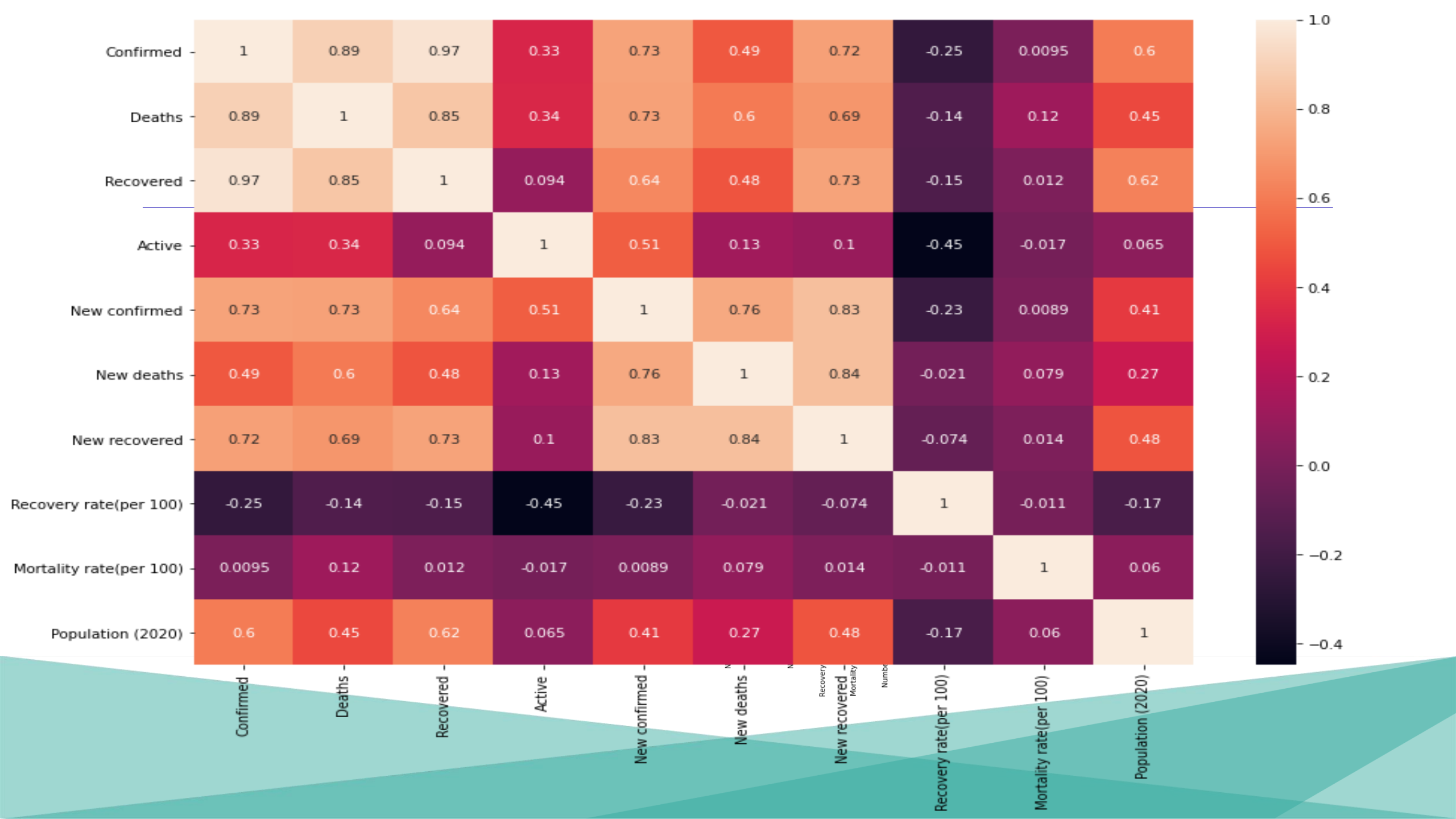
World Deaths Cases by ARIMA



	ds	yhat
505	2021-07-30	4.258170e+06
506	2021-07-31	4.268042e+06
507	2021-08-01	4.275966e+06
508	2021-08-02	4.284816e+06
509	2021-08-03	4.296549e+06
510	2021-08-04	4.308193e+06
511	2021-08-05	4.319730e+06
512	2021-08-06	4.331540e+06
513	2021-08-07	4.341412e+06
514	2021-08-08	4.349336e+06
515	2021-08-09	4.358186e+06
516	2021-08-10	4.369919e+06
517	2021-08-11	4.381563e+06
518	2021-08-12	4.393100e+06
519	2021-08-13	4.404910e+06
520	2021-08-14	4.414782e+06
521	2021-08-15	4.422707e+06
522	2021-08-16	4.431556e+06
523	2021-08-17	4.443289e+06
524	2021-08-18	4.454933e+06
525	2021-08-19	4.466470e+06

World Deaths Cases by Prophet





Analyze & Share , Act Phase :-

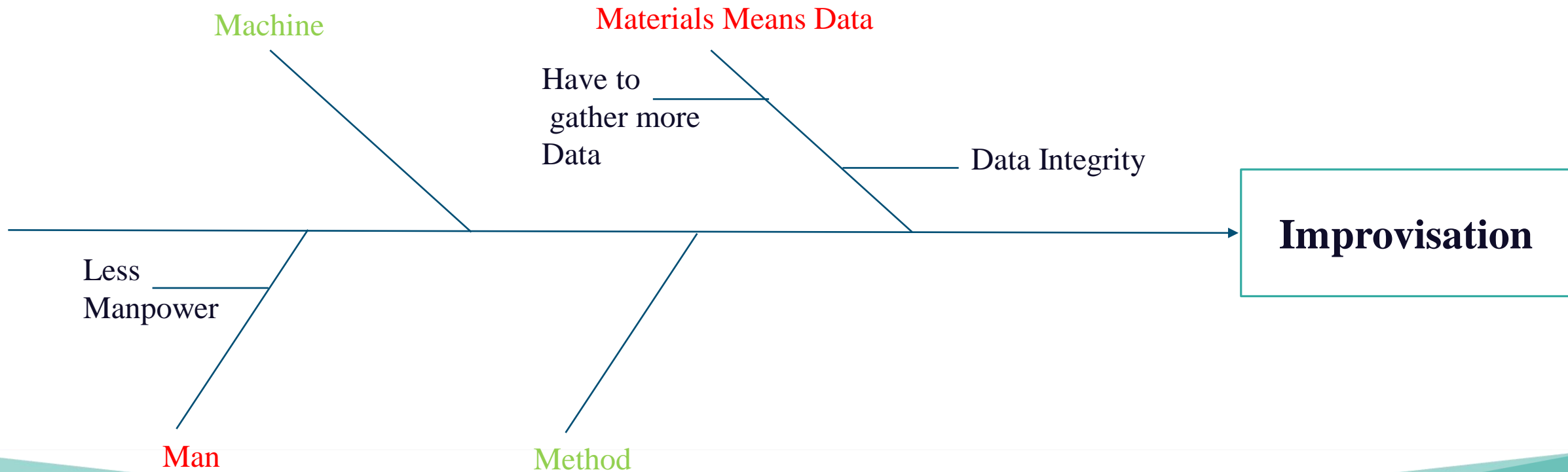
- What is your final conclusion based on your analysis?

- As a conclusion based on the analysis of the observations, it seems that even though the total number of confirmed cases and deaths in the world are monotonically (almost exponentially) increasing, the recovery rate shows some increase whereas the mortality rate shows some decrease. On the other hand, by data modeling and prediction based on univariate time series, using Linear regression, Support vector machine, Random forests and XGBoost we concluded that Support vector machine and Random forests performed the best and the worst accuracy, respectively. Moreover, both of Multilayer perceptron and LSTM-RNN performed high accuracy, more than 99.98 in percent.
- Furthermore, by examining the correlations between the features, it seems that there exist weak correlations between the new parameters, life expectancy, GDP per capita, social support, and freedom to make life choices, generosity, and the primary ones, confirmed, deaths, recovered, active cases, recovery rate and mortality rate. Also, it seems that the correlation between population and confirmed and, population and active cases is moderate (near 0.6). As future works, by considering the population of each country, we may investigate the percentage of total populations that will be affected by COVID-19. Also, the impact of some other parameters in prediction of COVID-19 spread can be considered. Moreover, data modeling and prediction based on multivariate time series using Multilayer perceptron and LSTM-RNN can be considered.

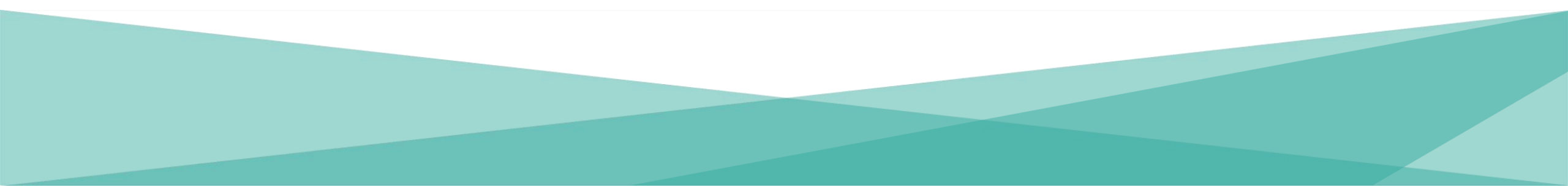
Act Phase

- How could your team and business apply your insights?
 - Well, a government organization can use my insight for taking decision with respect to Lockdown and Vaccination drive.
- What next steps would you or your stakeholders take based on your findings?
 - Well , I have to prepare a live Dynamic dashboard of my finding and have to use reinforcement learning for forecasting.
 - One more thing I want to do, is that I want to apply all this strategies only for India and its state. from there my insight also helpful for state government for taking measures to stop the spread of Covid-19.
 - We can also use covid-19 vaccination Data to perform some analysis and find some insight related to vaccination drive like where we have to supply more vaccine as compared to other state.
 - We can also track the all vaccination Data which we can use to analyze where we have to focus to reduce wastage of Vaccine.

Act Phase



Act Phase

- From Above diagram we Observed that we have to improve our Data gathering process and required more accurate data because we gather the data from public resources where, there is a high chances of losing Data accuracy and integrity.
 - Second we observed that there is a shortage of manpower. If a team of 4 people's work on same project then we can conduct brainstorming session which can gives us idea of visualization and more insight from same data.
 - Due to Manpower shortage it create a time constraint which directly impact on finding insight from data.
 - Machine is ok for this Project because we used google Colab for LSTM and MLP.
 - Method is also OK because we used Most detailed approach to reach out our question which is asked in a asked Phase.
- 

THANK YOU

An abstract decorative pattern at the bottom of the slide, composed of overlapping translucent teal and light blue geometric shapes, creating a layered, mountain-like effect.