

# PREDICTION & ANALYSIS ON SPREAD OF NOVEL COVID-19

## CAPSTONE GROUP PROJECT PRESENTATION

Group members :

JOSINA P JOY

KRISHNAMRAJU NALIMELA

SUBHALAXMI BABOO

# Contents

1	Project Introduction
2	Project Proposal
3	Pedagogy
3.1	Data Cleaning and Preparation
3.2	Exploratory Data Analysis and Statistics
3.2.1	Power BI
3.3	Time Series Forecasting
3.3.1	Prophet
3.3.2	Arima
3.3.3	Comparison of TS model
4	Summary and Inference
5	References

# 1. PROJECT INTRODUCTION

## INTRODUCTION

- Corona Virus Disease-2019 (COVID-19), an infectious disease caused by a novel coronavirus.
- COVID-19 outbreak is first observed in Wuhan City, China
- Mainly affects the human respiratory system & is highly contagious.
- More than 200 countries are infected .

## OBJECTIVE

- This project aims to analyze, visualize, perform live data comparisons using Johns Hopkins University dataset.
- To predict overall Growth Rate, Recovery Rate, and Mortality Rate
- Using Auto-Regressive Integrated Moving Average (ARIMA) & PROPHET .
- Further, the results to be validated over R accuracy metrics.

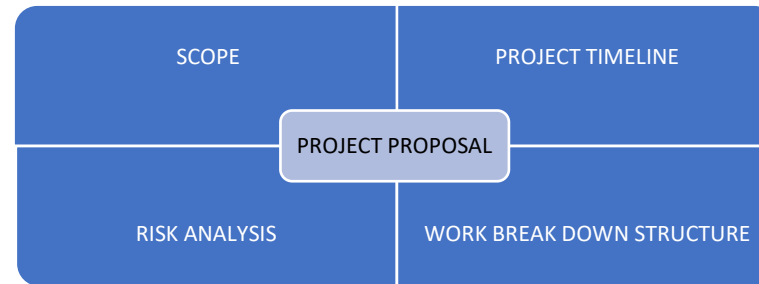
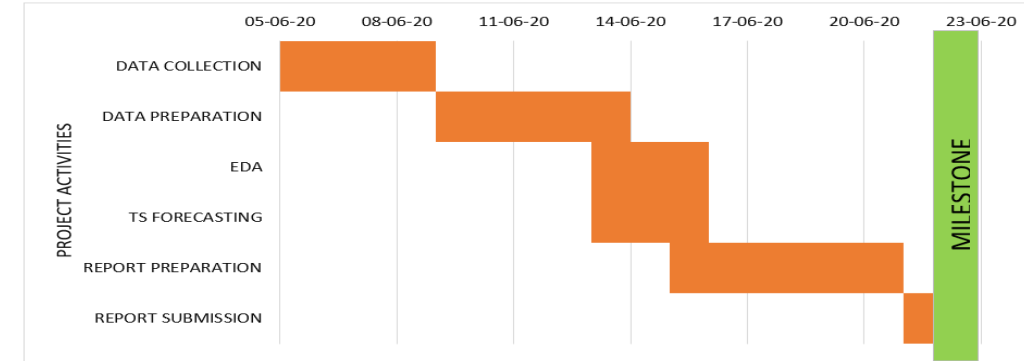
## PROJECT ADVANTAGES

- One can study and analyse the growth and spread of the virus.
- One can alarm the authorities about the approximate number of infected cases in the next 40 days.
- Helps the Government to take adequate health care measures such as arranging necessary equipments and covid specialised hospitals, preparing front line workers.

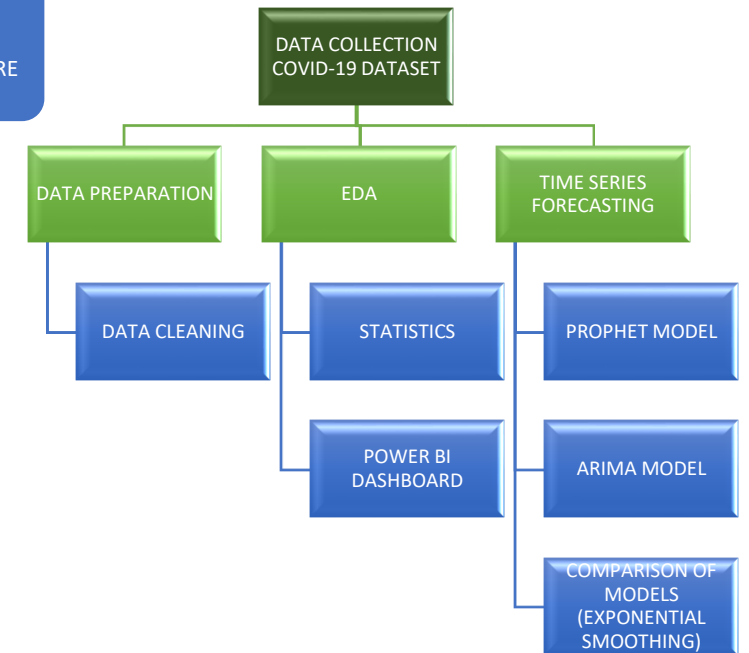
## 2. PROJECT PROPOSAL

Project scope will include :

- Identification of appropriate dataset.
- Data preparation & EDA using R tools.
- Develop a dynamic Power BI dashboard.
- Build TS forecasting models.



Risk no:	Risk	Severity	Occurrence	Criticality	Mitigation
1	Miscommunication between team members	9	8	72	Created a WhatsApp group with team members and project coordinator
2	Lack of coordination	6	8	48	Scheduled a daily zoom meeting to share the project updates
3	Lack of knowledge in setting ts frequency	9	5	45	Referred similar journal, websites.
4	Improper requirement definition	10	3	30	Identified the necessary attributes



# 3.PEDAGOGY

## 3.1 DATA PREPARATION & CLEANING

## 3.2 EXPLORATORY DATA ANALYSIS & STATISTICS

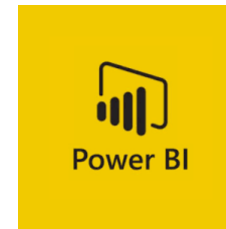
### 3.2.1 POWER BI

## 3.3 TIME SERIES FORECASTING

### 3.3.1 PROPHET

### 3.3.2 ARIMA

### 3.3.3 COMPARISON OF TS MODELS



# 3.1 DATA PREPARATION & CLEANING

## a) Procedure

- Checking not available values from the data set.
- Converting attributes into appropriate class.
- Creating confirmed, recovered and deaths columns

## b) Result

Fig: 1 Converted raw data into tabular form

Province/State, Country/Region, Lat, Long, 1/22/20, 1/23/20, 1/24/20, 2/11/20, 2/12/20, 2/13/20, 2/14/20, 2/15/20, 2/16/20, 2/17/20, 3/5/20, 3/6/20, 3/7/20, 3/8/20, 3/9/20, 3/10/20, 3/11/20, 3/12/20, 3/29/20, 3/30/20, 3/31/20, 4/1/20, 4/2/20, 4/3/20, 4/4/20, 4/5/20, 4/20/20, 4/23/20, 4/24/20, 4/25/20, 4/26/20, 4/27/20, 4/28/20, 4/29/20, 5/17/20, 5/18/20, 5/19/20, 5/20/20, 5/21/20, 5/22/20, 5/23/20, 5/10/20, 6/11/20, 6/12/20, 6/13/20, 6/14/20, 6/15/20, 6/16/20, 6/17/20, Afghanistan, 33.0, 65.0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 237, 273, 281, 299, 349, 367, 423, 444, 484, 521, 555, 607, 665, 714, 784, 4963, 5226, 5639, 6053, 6402, 6664, 7072, 7653, 8145, 8676, 9216, 9998, 24766, 25527, 26310, 26874, 27532, 27878, 28424, 28833, 29157, 29481, Albania, 41.1533, 20.1683, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 223, 243, 259, 277, 304, 333, 361, 377, 383, 400, 409, 416, 433, 446, 4916, 933, 946, 948, 949, 964, 969, 981, 989, 998, 1004, 1029, 1050, 1076, 047, 2114, 2192, 2269, 2330

Raw data transformed

	Country <chr>	date <dtm>	confirmed <dbl>	recovered <int>	deaths <dbl>
1	Afghanistan	2020-01-22 00:00:00	0	0	0
2	Albania	2020-01-22 00:00:00	0	0	0
3	Algeria	2020-01-22 00:00:00	0	0	0
4	Andorra	2020-01-22 00:00:00	0	0	0
5	Angola	2020-01-22 00:00:00	0	0	0
6	Antigua and Barbuda	2020-01-22 00:00:00	0	0	0

## 3.2 EXPLORATORY DATA ANALYSIS & STATISTICS

a) Result

Fig 1: List of top 6 infected countries

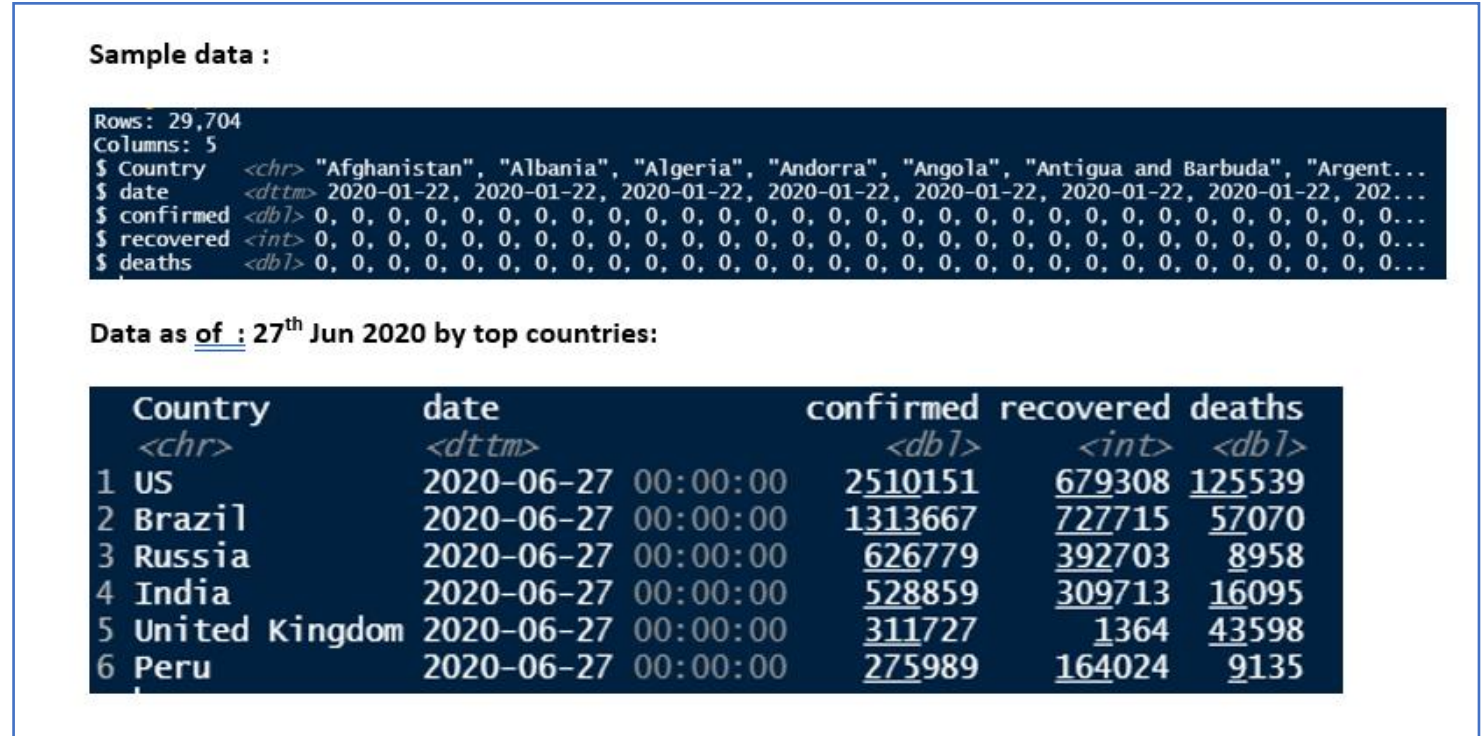


Fig 2: No: of cases in US



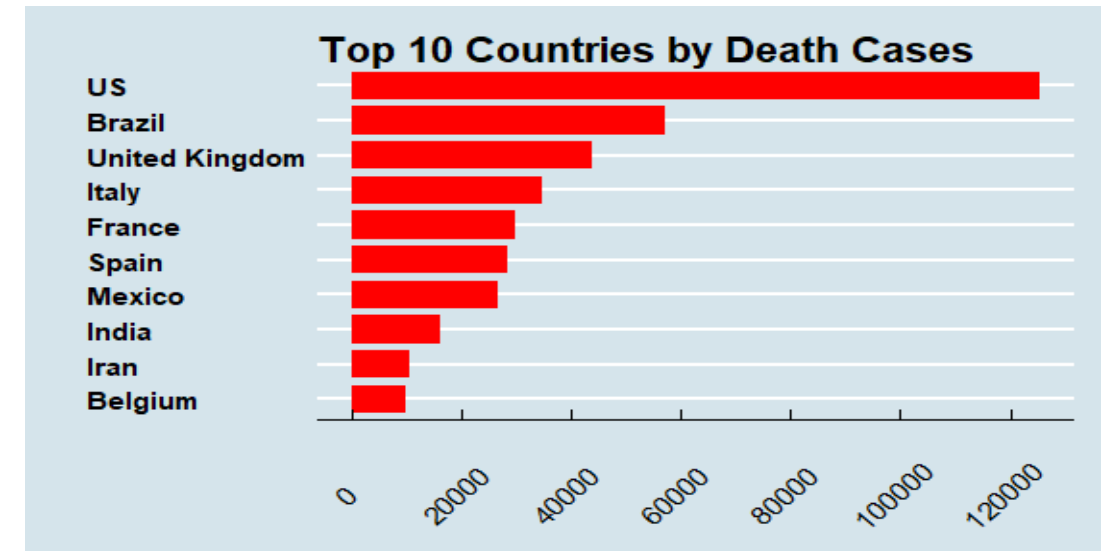
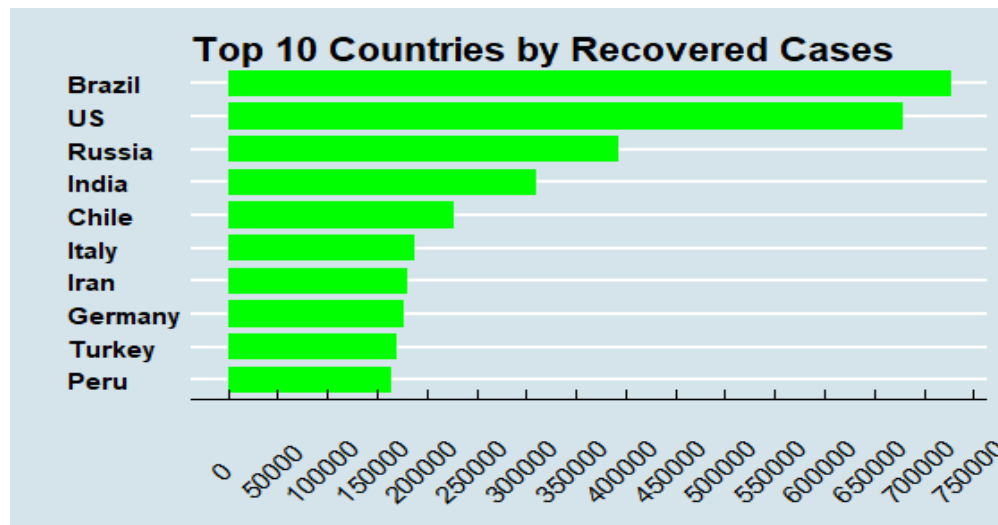
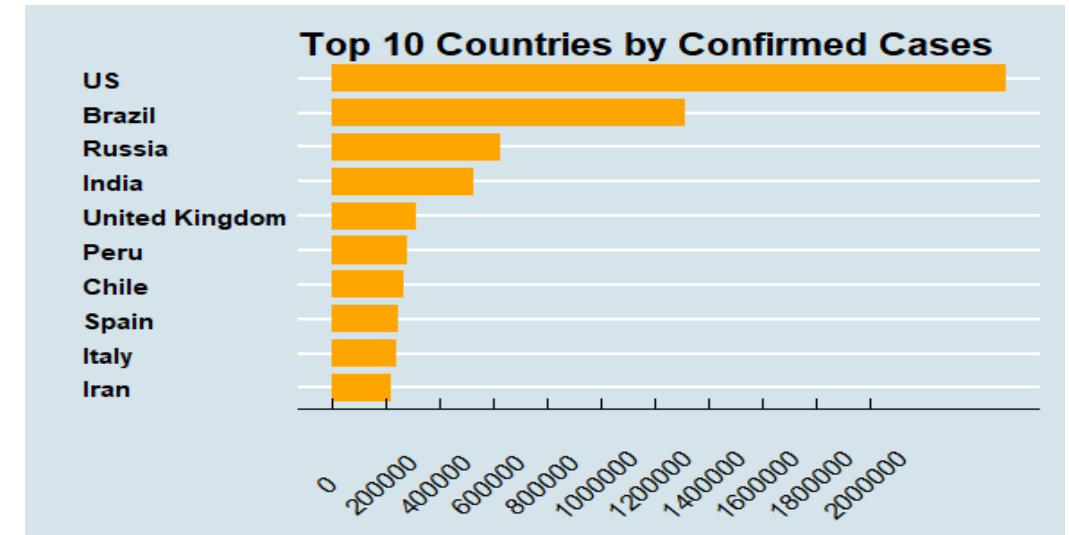
# 3.2 EXPLORATORY DATA ANALYSIS & STATISTICS.....Cont

a) Result

Fig 1: List of top 10 countries by confirmed countries

Fig 2: List of top 10 countries by recovered countries

Fig :3 List of top 10 countries by deceased countries





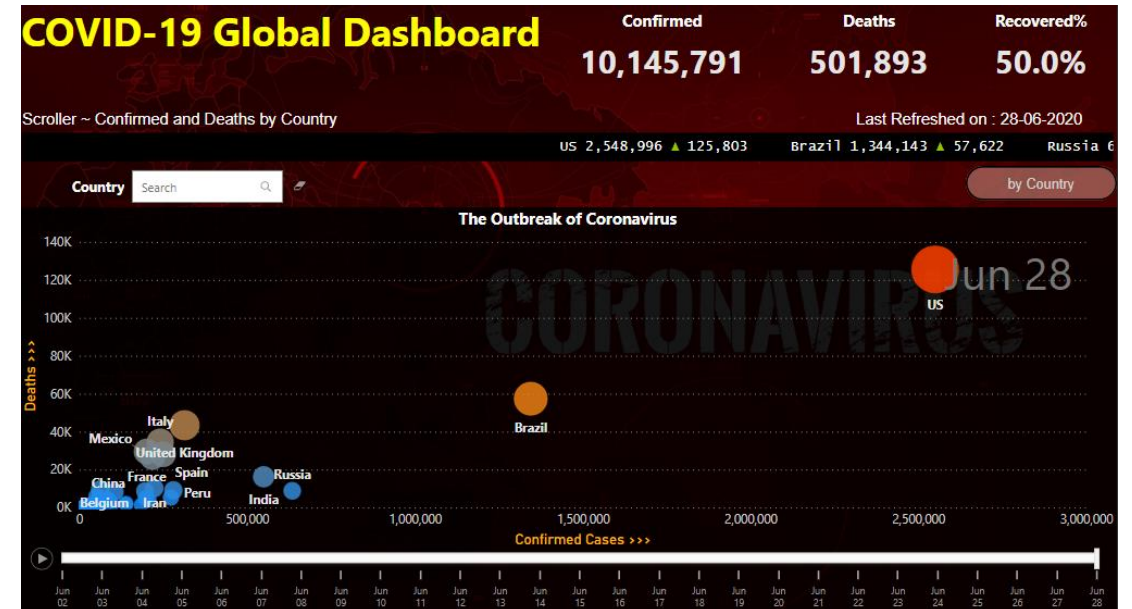
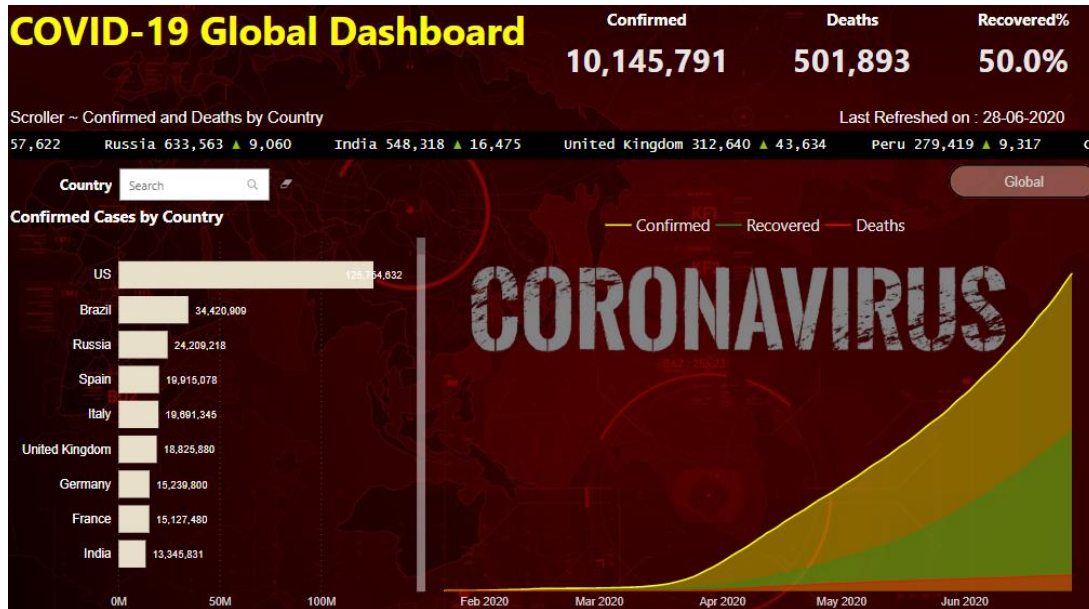
## 3.2.1 POWER BI

Power BI Dashboard link : [Link](#)

a) Result

Fig 1: Homepage of dynamic Power BI dashboard

Fig 2: Animated representation of infection growth



# 3.3 TIME SERIES FORECASTING

- Performing forecasting techniques on a time series data ie the prediction over time.
- Set of observations generated sequentially with time on a single variable or it is indexed by time.



## PROPHET

- To forecast confirmed cases.
- Created historical data frame with date & outcome variables.
- Fitting prophet model
- Automatically disables seasonality.
- Built-in helper function to create future data frame.
- No need to set frequency, it will set by default.

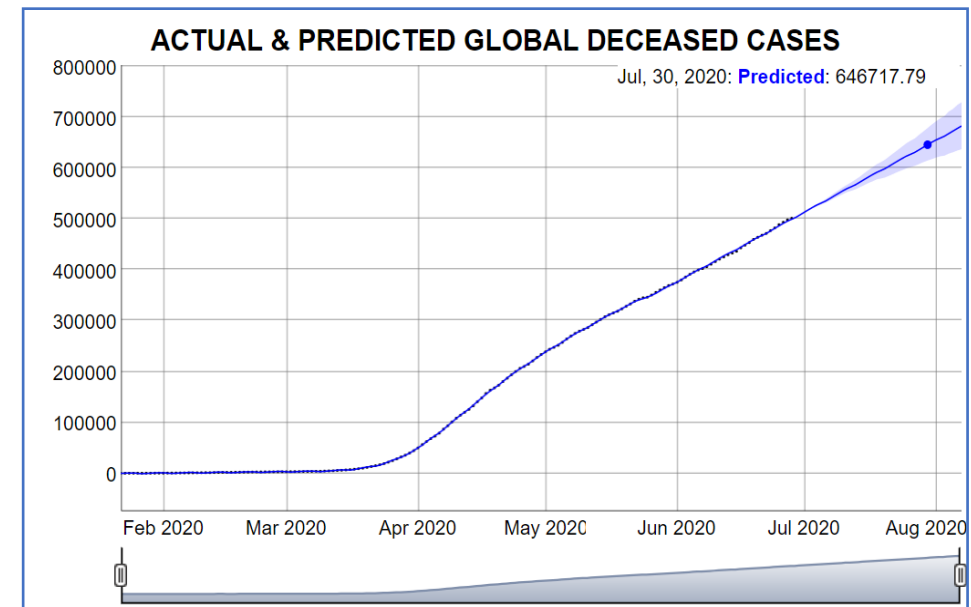
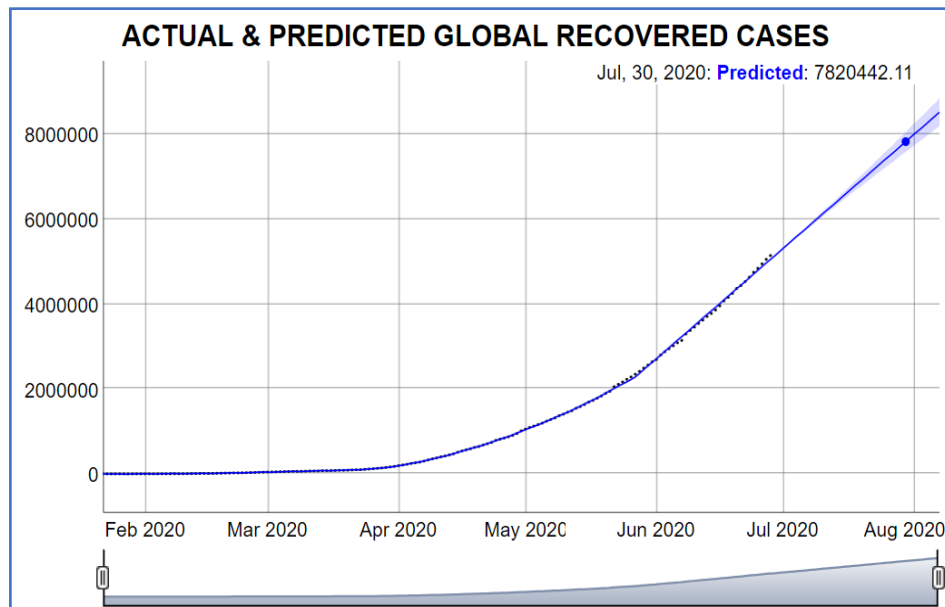
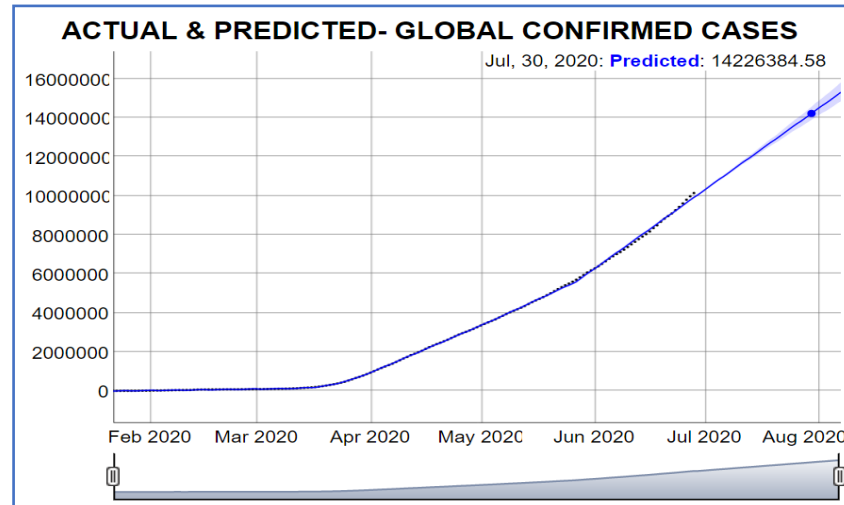
## ARIMA

- Created time series object using xts package,
- Performed acf and pacf to find the p, d, q values.
- Build ARIMA model and forecasted the cases.
- Also build auto arima model and forecasted the values.

## COMPARISON OF MODELS

- Compared the performance of prophet model, arima model, automated arima model and exponential smoothing model.
- The metrics used to compare the accuracy are MAPE, MASE, RMSE, AIC, BIC

# 3.3.1 PROPHET

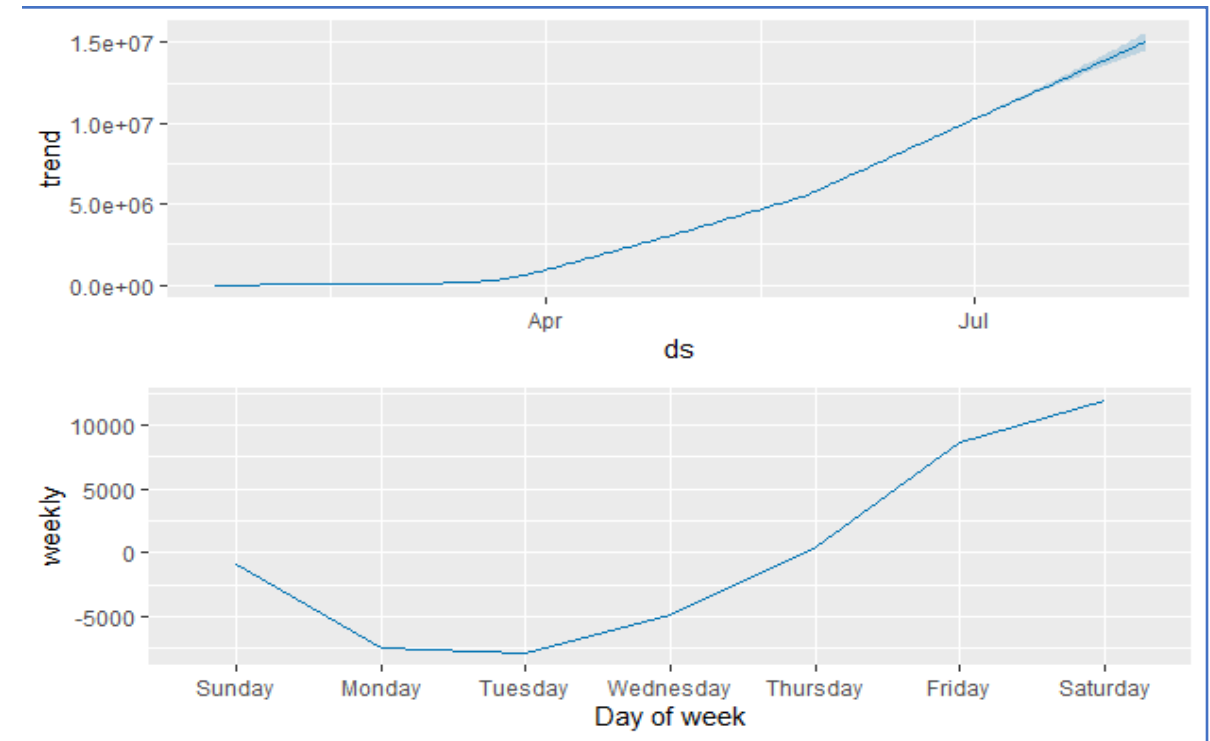
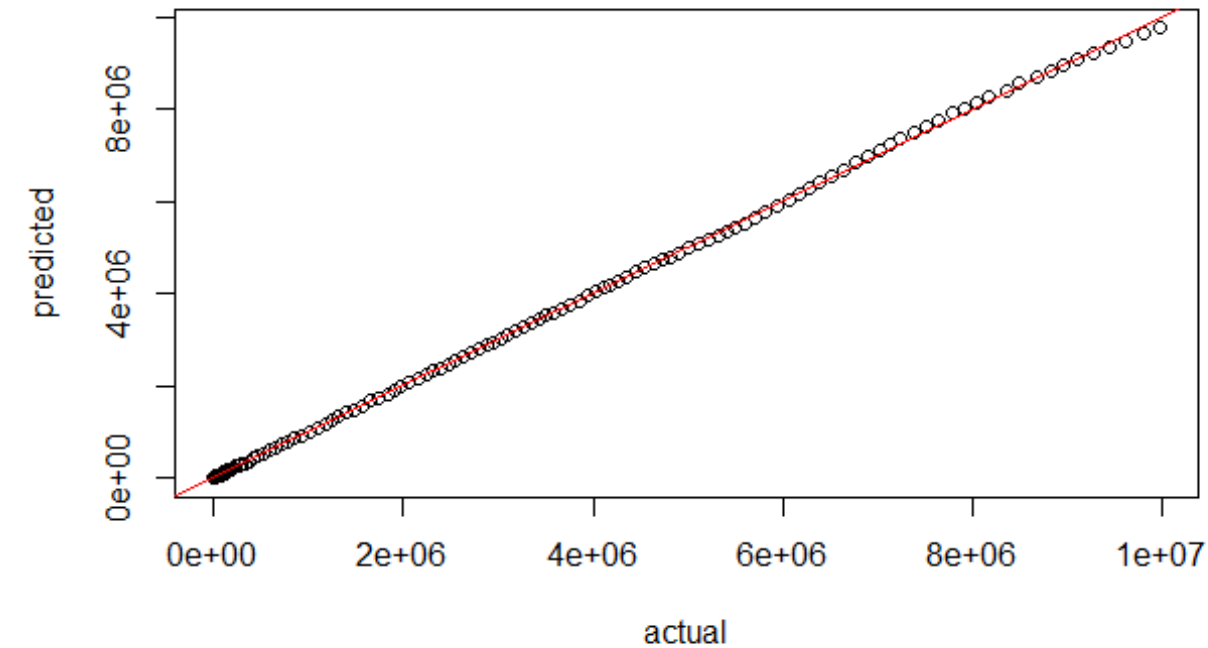


## 3.3.1 PROPHET.....Cont

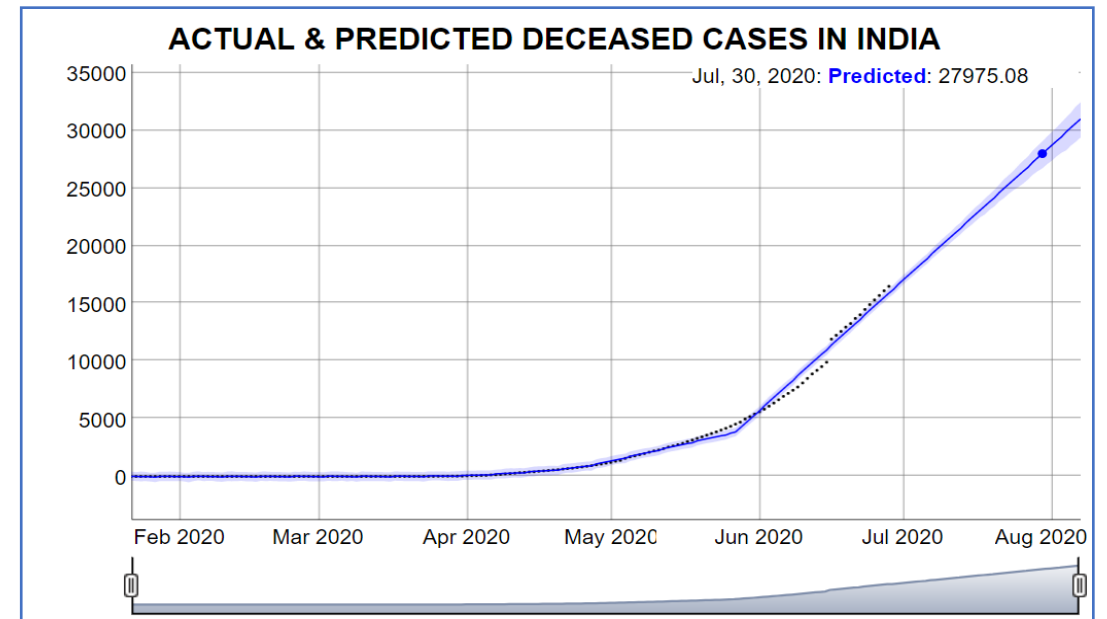
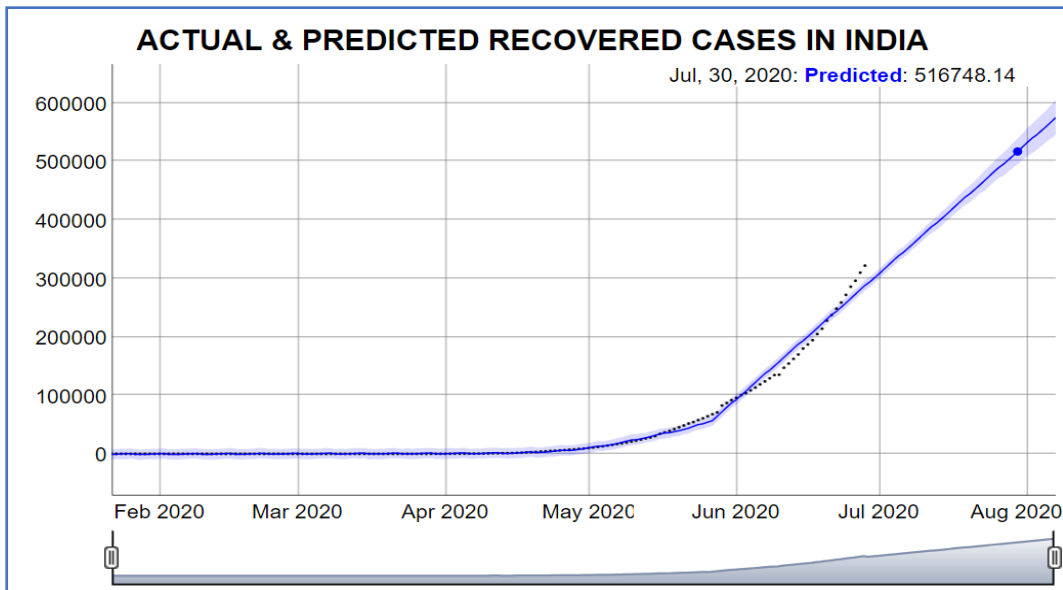
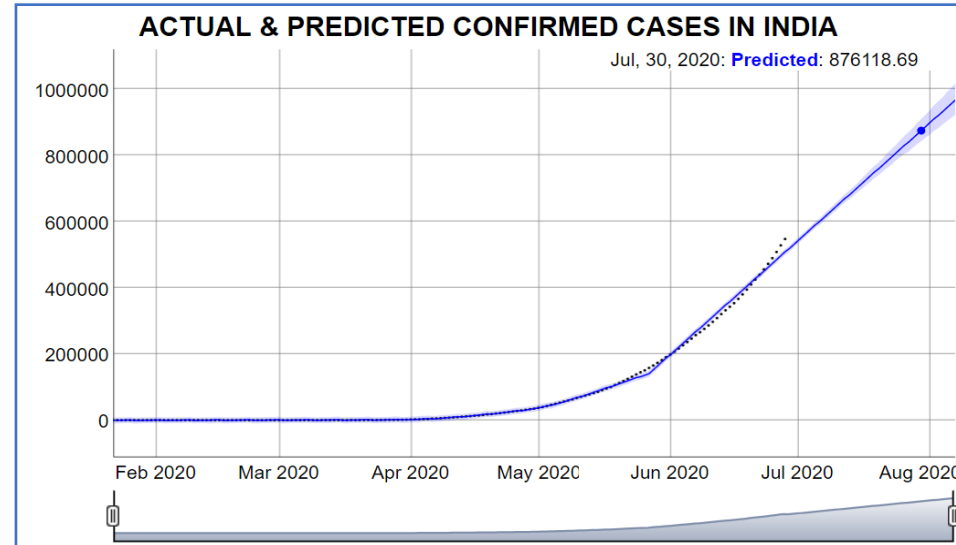
a) Result

Fig 1: Fit line of actual and predicted cases

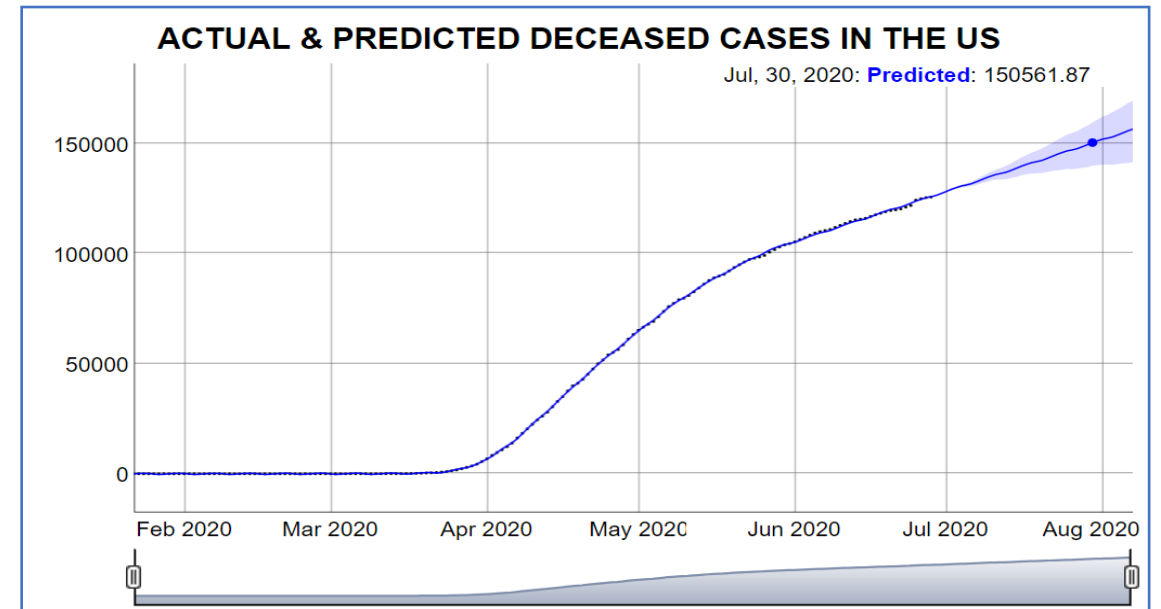
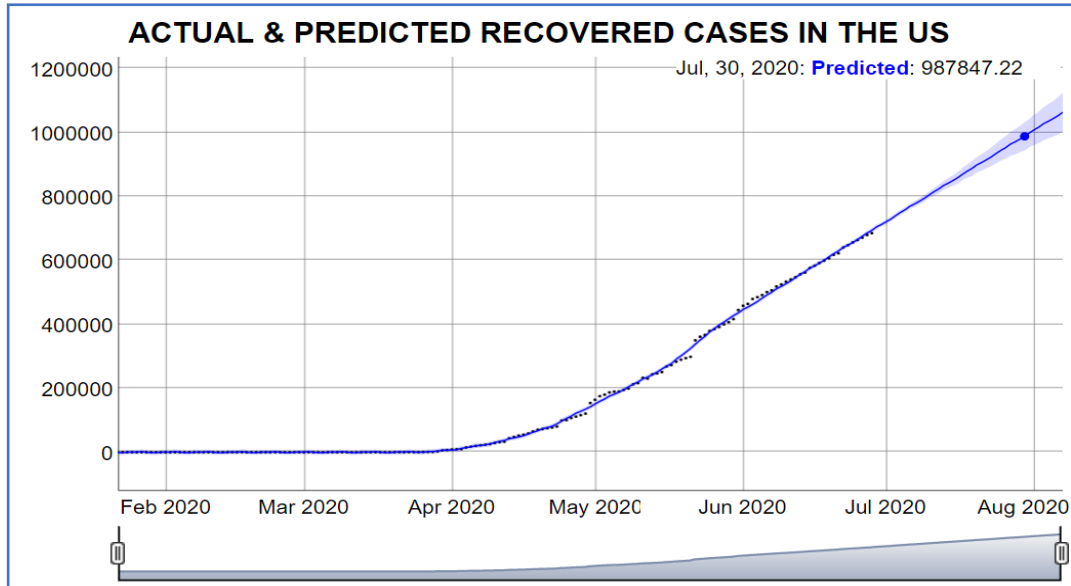
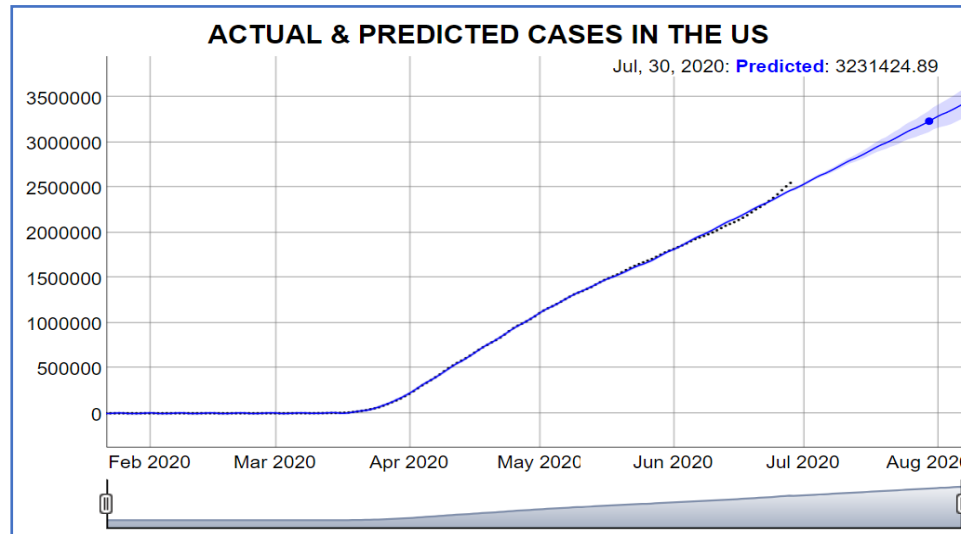
Fig 2: Time series components ( trend & weekly seasonality)



# 3.3.1 PROPHET.....Cont

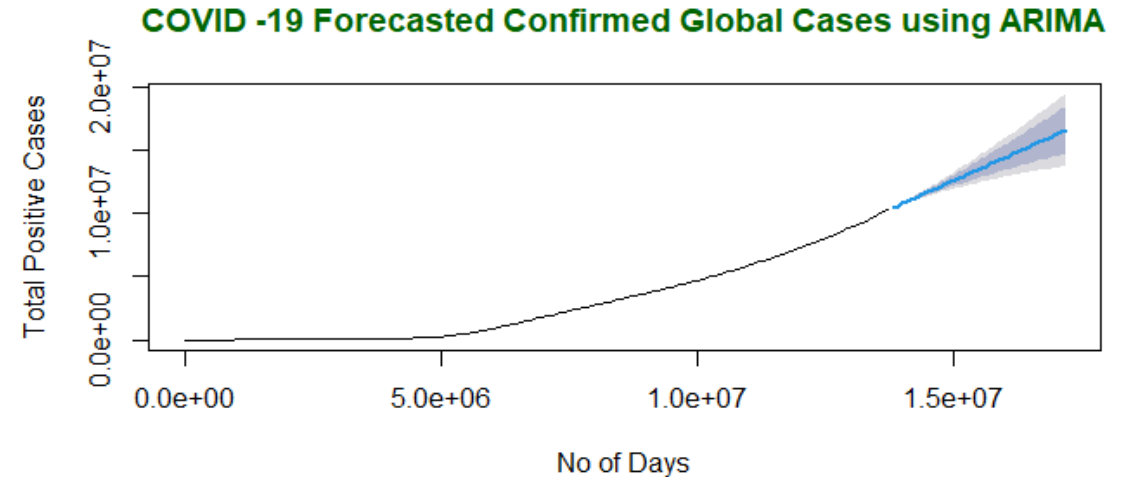


## 3.3.1 PROPHET.....Cont



## 3.3.2 ARIMA

- Confirmed cases will be around 15608413.
- Created time series object and forecasted the future cases.



## 3.3.3 COMPARISON OF TS MODELS

- Compared prophet, arima , auto arima & exponential model.
- Compared model performance using R accuracy metrics such as MAPE, RMSE, MASE, AIC, BIC, ACF
- Accuracy() function from forecast package is used.
- MAPE value is lowest for ARIMA model.
- All the 4 models have MAPE value less than 20.

```
> prophet_model_accuracy
               ME      RMSE      MAE      MPE      MAPE      MASE
Training set -2.789627e-11 44075.66 23860.78 -1.116391 14.8997 0.009271408
> ets_model_accuracy
               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1647.794 9543.191 6388.593 -0.2297444 5.352697 0.1005122 0.02553355
> auto_arima_accuracy
               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1575.132 8573.352 5656.424 0.8776556 1.827148 0.08899292 0.001954441
> arima_model_accuracy
               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1125.72 9831.788 6265.79 0.6221695 1.827937 0.09858011 -0.2175761
```

# 4.SUMMARY & INFERENCE

The results and insights we have got from the project is as follows:

- From the EDA, we could realize that around 49.94% of total corona infected people has recovered and 5% overall death rate.
- We could understand that US is the worst affected country with highest mortality rate.
- Power BI dashboard helps in analyzing and studying the spread of this virus.
- The trend of confirmed cases are still going up but death rates are comparatively low and growth rate is increasing.
- Country wise differences are observed due to health policies, preliminary measures and economic levels.



TIME SERIES FORECASTED VALUES: 30th July 2020			
COUNTRY	CONFIRMED	RECOVERED	DECEASED
GLOBAL	15608413	7820442	646718
INDIA	876119	516748	27975
US	3231425	987847	150562

TOTAL NO: OF CONFIRMED CASES				
Date	Prophet	Arima	Auto Arima	Actual
29-06-2020	10039570	10311325	10316875	1024460
30-07-2020	14226385	15608413	15831730	---



# 5. REFERENCES

1. [Johns Hopkins University](#) for making the data available for educational and academic research purposes
2. <https://www.r-bloggers.com/>
3. <https://rpubs.com/>
4. <https://stackoverflow.com/>
5. <https://facebook.github.io/prophet/>
6. [https://facebook.github.io/prophet/docs/quick\\_start.html#r-api](https://facebook.github.io/prophet/docs/quick_start.html#r-api)
7. World Health Organization (WHO): <https://www.who.int/>