Trend Analysis and Forecasting COVID-19

# **INTRODUCTION**

This Covid 19 Data Analysis project aims to determine the reason and justify the causes by using different statistical techniques. This study provides a broad overview of the time series data of covid confirmed cases and fatality rates globally and compares reported cases in India with other Countries computes the actual growth rate and daily changes for specific locations by estimating the global trends per location. Also, the SARS Cov-2 trend and forecasting for the next 28 days up from Dec 23. 2021 to Jan 19, 2022, is also predicted by applying different statistical tools and techniques

# **PROBLEM STATEMENT**

The novel coronavirus reported at the end of 2019 has impacted almost every aspect of life. This project mainly focuses on analyzing, forecasting, visualizing, and investigating the current coronavirus spread in different countries.

# **OBJECTIVE**

1. To analyze the time-series data of covid confirmed cases and fatality rates globally
2. To compute the actual growth rate and daily changes for specific locations
3. To estimate the global trends per location
4. To compare reported cases in India with other Countries
5. To Predict the Pandemics Trend and forecast its data
6. To visualize the data by using Single Trend (for single location),itrend (interactive trends),mtrend (multiple locations)
7. To perform analysis for Vaccination & Testing and Genomics data

# **DATASET USED**

A Live data set package from R studio named "covid19.analytics," which fetches current covid data from the JHU Github repository, is used for the analysis.

**JHU's CCSE repository**

<https://github.com/mponce0/covid19.analytics>

**Focused Attributes -** (Province\_State,Country\_Region,Confirmed, Deaths,Recovered and Active cases)

**Total Observation –** 4006

**Total Variables –** 40

# **TOOL DESCRIPTION(LIBRARIES, PACKAGES AND SOFTWARE USED)**

Statistical Software used **–** R Studio

## Libraries Used :

* library(covid19.analytics)
* library(plotly)
* library(lubridate)
* library(ggplot2)
* library(prophet)
* library(dplyr)

## Techniques implemented:

* Summary Statistics
* ONE -WAY ANOVA
* Regression Analysis
* **SIR** MODEL (Susceptible-Infected-Recovered)
* Pandemics Trend Prediction and forecasting(time in day wise)
* Interactive visulaizations

# **CODE:**

library(covid19.analytics)

library(plotly)

library(lubridate)

library(ggplot2)

library(prophet)

library(dplyr)

options(scipen = 6)

covidData <- covid19.data()

## obtain time series data for "confirmed" cases ##

confirmed\_cases <- covid19.data(case = "ts-confirmed")

# India Confirmed Cases

confirmed\_cases\_in  <-

  confirmed\_cases %>% filter(Country.Region == "India")

print(confirmed\_cases\_in)

## Death Count / reads time series data for casualties ##

death\_cases <- covid19.data(case = "ts-deaths")

# India Death Cases

death\_cases\_in <- death\_cases %>% filter(Country.Region == "India")

print(death\_cases\_in)

## obtain time series data for "recovered" cases ##.

recovered\_cases <- covid19.data(case = "ts-recovered")

# India Recovered Cases

recovered\_cases\_in <-

  recovered\_cases %>% filter(Country.Region == "India")

print(recovered\_cases)

#Summary Statistics

#GLOBAL

report.summary(Nentries = 5,

               graphical.output = T)

report.summary(Nentries = 5,

               graphical.output = T,

               geo.loc = "India")

#one way anova

one.way = aov(Confirmed ~ Deaths, data = covidData)

summary(one.way)

# Graphs and Visualization

total\_ts <- covid19.data(case = "ts-ALL")

totals.plt(total\_ts)

#totals per location  //Regression Analysis

tots.per.location(confirmed\_cases, geo.loc = "India")

# growth rates

growth.rate(confirmed\_cases, geo.loc = c('US', 'India'))

growth.rate(confirmed\_cases, geo.loc = c('Brazil', 'India'))

#SIR modelling

SIR = generate.SIR.model(confirmed\_cases, 'India', tot.population = 34478517)

#livemap

live.map(confirmed\_cases)

################################################################################

# Pandemics Trend Prediction (time in day wise)

tsc <- covid19.data(case = 'ts-confirmed')

tsc <- tsc %>% filter(Country.Region == 'India')

tsc  <- data.frame(t(tsc))

tsc <- cbind(rownames(tsc), data.frame(tsc, row.names = NULL))

colnames(tsc) <- c('Date', 'Confirmed')

tsc <- tsc[-c(1:4), ]

tsc$Date <- ymd(tsc$Date)

str(tsc)

tsc$Confirmed <- as.numeric(tsc$Confirmed)

#Plot

qplot(Date, Confirmed, data = tsc, main = 'Covid19 confirmed cases in India')

ds <- tsc$Date

y <- tsc$Confirmed

df <- data.frame(ds, y)

# Forecasting

m <- prophet(df)

# Prediction

future <- make\_future\_dataframe(m, periods = 28)

forecast <- predict(m, future)

#Plot forecast

plot(m, forecast)

dyplot.prophet(m, forecast)

#Forecast components

prophet\_plot\_components(m, forecast)

#Model Performance

pred <- forecast$yhat[1:121]

actual <- m$history$y

plot(actual, pred)

#Single Trend for India

indiaData <-

  confirmed\_cases[confirmed\_cases$Country.Region == "India" , ]

single.trend(indiaData)

#Multiple Location

mtrends (confirmed\_cases, geo.loc = c ("US", "India"))

#Interactive Locations trend daily cases

itrends (covid19.data("ts-confirmed") , geo.loc = "India")

itrends (covid19.data("ts-confirmed") , geo.loc = "ALL")

#TESTING

covidTest <- covid19.testing.data(tgt = "testing", disclaimer = TRUE)

#covidTest<-covidTest %>% filter(Country.Region == "India")

print(covidTest)

#VACCINATION

covidvaccine <-

  covid19.vaccination(tgt = "global",

                      data.fmt = "orig",

                      disclaimer = TRUE)

print(covidvaccine)

#GENOMIC

covidgenomic <- covid19.genomic.data(

  type = "genome",

  src = "livedata",

  graphics.ON = TRUE,

  accOnly = TRUE

)

print(covidgenomic)

# **OUTPUT**

**Graphical user interface

Description automatically generated with low confidence**

Graphical user interface

Description automatically generated with low confidenceChart, line chart

Description automatically generated

**Chart, line chart

Description automatically generated**

**A picture containing application

Description automatically generated**

**Chart

Description automatically generated with medium confidenceA picture containing text

Description automatically generated**

**Chart, line chart

Description automatically generatedChart

Description automatically generated with medium confidence**

**Diagram

Description automatically generatedChart, line chart

Description automatically generated**

**Diagram

Description automatically generatedChart, line chart

Description automatically generated**

**Chart

Description automatically generated**

**Table

Description automatically generated with low confidence** **Table

Description automatically generated with medium confidenceChart, shape, square

Description automatically generated**

# **ANALYSIS REPORTS AND INFERENCE:**

* Comparing Confirmed and Death cases globally shows that the two entities are significantly different as the F value is positive. We reject the Null Hypothesis. Infers that confirmed and instances of death vary, indicating that the confirmed cases are more than the death cases.
* Compared to India, brazil has more Covid cases
* Thus with the SIR model, the covid infected cases are getting increased in India as the number of persons who are prone covid is more
* Forecasting for the next 28 days predicts that the covid cases will increase concerning the actual data (up to Jan 19, 2022).
* It infers that the SARS COV-2 gets evolved in different variants by mutating its genome in a way by changing its characteristics and traits to spread even more vigorously, thus resulting in the rise of COVID cases globally

**PROOF for TRUTHFULNESS OF THE FORECASTING DATA**

Graphical user interface

Description automatically generated

# **CONCLUSION:**

## JUSTIFICATION:

Though Herd immunity is developing among the people to fight against the novel coronavirus, it mutates itself by changing its characteristics and traits, resulting in a widespread virus.

## SOLUTION TO THIS PROBLEM:

* The sequencing of the Genomes and testing can be done to prevent the spread of the mutated virus
* The initiation of vaccination can increase among the people
* People can maintain the Covid restrictions put forth by the Government to bring the covid in control