Trends for Covid-19

Submitted in partial fulfilment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

By

Aayush Rokade (21102064)

Aditya Raorane (21102173)

Ajayraj Rathod (21102085)

Unnati Pohankar (21102070)

Under the Guidance of

Prof. Kadambari Deherkar

For BDA Project

Department of Computer Engineering

A.P. SHAH INSTITUTE OF TECHNOLOGY, THANE

(2023-2024)

Problem Statement:

The COVID-19 pandemic has significantly impacted global health, economy, and daily life, leading to complex and rapidly changing trends across various regions. Traditional methods of analyzing pandemic data have proven inadequate due to the sheer volume, velocity, and variety of information generated worldwide. This project aims to leverage big data analytics to uncover hidden patterns and trends in COVID-19 spread, deaths, and financial effects across different regions particularly India and USA for the duration 2019-2020.

Objective:

Trend Analysis of COVID-19 Spread and Mortality: Analyse historical and real-time COVID-19 case and death data for India and the USA to identify significant trends, peak periods, and mortality rates across various states and regions.

Regional Comparison of Pandemic Impact: Compare the pandemic's trajectory and intensity between India and the USA by examining infection rates, death ratios, and recovery patterns to understand region-specific factors influencing these differences.

Financial Impact Assessment: Investigate the financial effects of the pandemic on key economic indicators (such as GDP growth, unemployment rates, and healthcare expenditures) in India and the USA to highlight sectors most affected.

Predictive analysis for Outbreak Forecasting: Analyse graphs and data to forecast future outbreaks and potential waves, enabling proactive measures and preparedness in each country.

Policy Evaluation and Recommendation: Assess the impact of various containment and vaccination strategies employed in India and the USA, offering data-driven recommendations to optimize pandemic management and minimize future disruptions.

Technical Approach (Implementation):

1. **Data Acquisition and Preprocessing:**
   Kaggle datasets: https://www.kaggle.com/datasets/shashwatwork/impact-of-covid19-pandemic-on-the-global-economy

2. **Load Required Libraries:**
   Make sure you load any necessary libraries at the beginning of your R script. Common libraries for data analytics include:

```
install.packages("naniar")
install.packages("visdat")
install.packages("dplyr")
```

3. **Read Your CSV Files**
   Use the appropriate function to read your CSV files into R. Depending on the size of your data, you can choose different functions:

```
data <- read.csv("your_file.csv")
```

4. **Clean the Data**:

The **naniar** package is specifically aimed at exploring and visualizing missing data. It provides tools to help identify, visualize, and deal with missing values in your datasets.

```
library(naniar)
vis_miss(covid19_data) #Visualizing the Data which is missing

colSums(is.na(covid19_data))
library(dplyr)
glimpse(covid19_data) #getting type of data in the dataset with respect to columns

#Replacing NA values with means due to Numeric Values
covid19_data$total_cases[is.na(covid19_data$total_cases)] <-
median(covid19_data$total_cases, na.rm = TRUE)
covid19_data$total_deaths[is.na(covid19_data$total_deaths)] <-
median(covid19_data$total_deaths, na.rm = TRUE)
covid19_data$stringency_index[is.na(covid19_data$stringency_index)] <-
median(covid19_data$stringency_index, na.rm = TRUE)
```

```
covid19_data$gdp_per_capita[is.na(covid19_data$gdp_per_capita)] <-
median(covid19_data$gdp_per_capita, na.rm = TRUE)
covid19_data$human_development_index[is.na(covid19_data$human_development_i
ndex)] <- median(covid19_data$human_development_index, na.rm = TRUE)

sum(is.na(covid19_data))
```

Further the **visdat** package is focused on visualizing the structure of datasets, including data types and missingness. It helps you understand the composition of your data through visual tools.

5. Write and run R Scripts:

- using Line plot to visualize the trend between cases in both countries:

```
library(ggplot2)
ggplot(combined_long, aes(x = date, y = Total_Cases, color = Country)) +
  geom_line(size = 1) +
  labs(title = "Total COVID-19 Cases: India vs. USA",
       x = "Date",
       y = "Total Cases",
       color = "Country") +
  theme_minimal() +
  scale_color_manual(values = c("total_cases_IND" = "blue", "total_cases_USA" =
"red"))
```

- Average GDP Per Capita by Country:

```
ggplot(average_gdp_per_country, aes(x = average_gdp , y = reorder(location,
average_gdp))) +
  geom_bar(stat = "identity", fill = "skyblue") +  # Bar graph
  coord_flip() +  # Flip coordinates for better readability
  labs(title = "Average GDP Per Capita by Country",

       x = "Country",
       y = "Average GDP Per Capita") +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  xlim(c(0, max(average_gdp_per_country$average_gdp) * 1.0))
```

- Pie chart for cases v/s deaths in India:

```
ggplot(pie_data, aes(x = "", y = Count, fill = Category)) +
  geom_bar(width = 1, stat = "identity") +  # Create a bar chart with one bar per
category
  coord_polar("y") +  # Convert the bar chart into a pie chart
```

```
  labs(title = "Pie Chart: Average Total Cases, Total Deaths, and Remaining
Population in India") +
  theme_void() +  # Remove background and axis details for a clean pie chart
  scale_fill_manual(values = c("Average Total Cases" = "orange",
                               "Average Total Deaths" = "blue"))+
  geom_text(aes(label = paste0(round(Percentage, 1), "%")),
            position = position_stack(vjust = 0.5),
            color = "white")
```

Output:

```
stringency_index   population          gdp_per_capita     human_development_index
Min.   :  0.00   Min.   :       809   Min.   :   661.2   Min.   :0.000
1st Qu.: 37.96   1st Qu.:   1399491   1st Qu.:  5338.4   1st Qu.:0.601
Median : 61.11   Median :   8278737   Median : 13913.8   Median :0.752
Mean   : 56.16   Mean   :  42516010   Mean   : 20818.7   Mean   :0.720
3rd Qu.: 78.70   3rd Qu.:  29136808   3rd Qu.: 31400.8   3rd Qu.:0.847
Max.   :100.00   Max.   :1439323774   Max.   :116935.6   Max.   :0.953
NA's   :7126                          NA's   :5712       NA's   :6202
>
```

Fig1. Summary

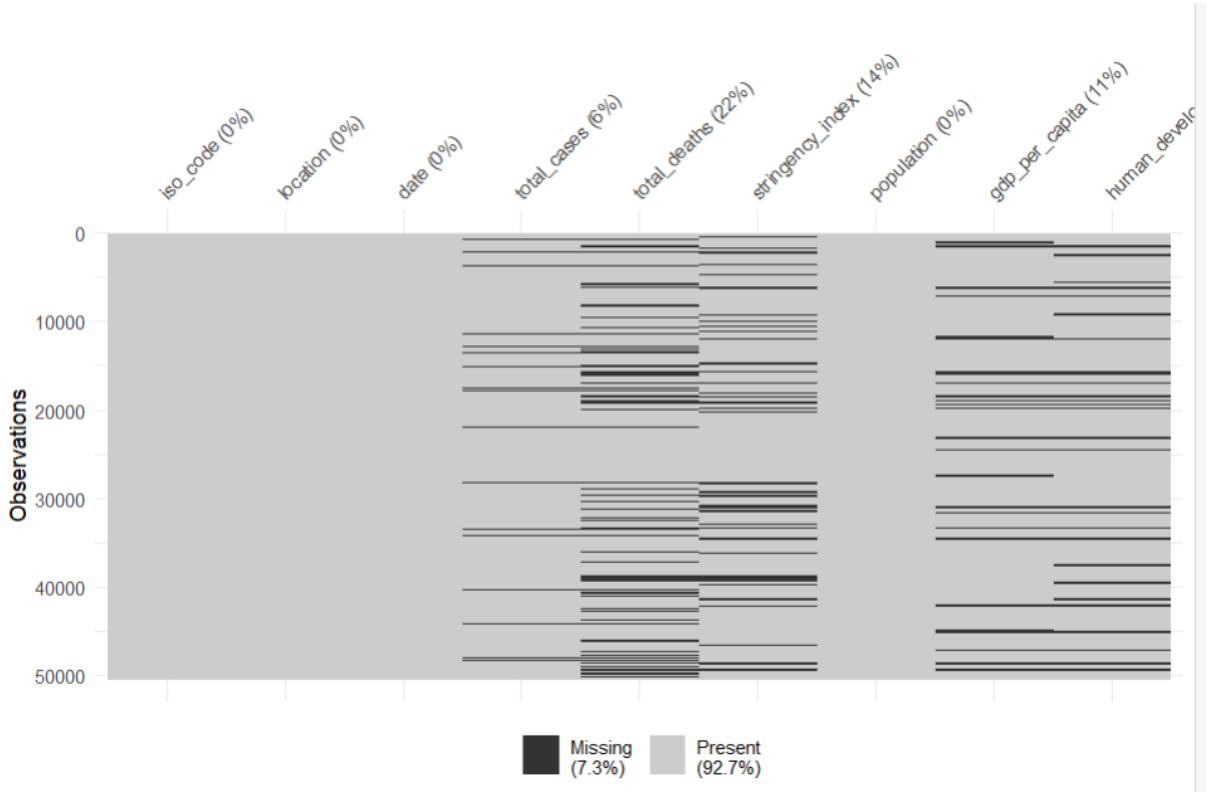

Fig2. Missing v/s present

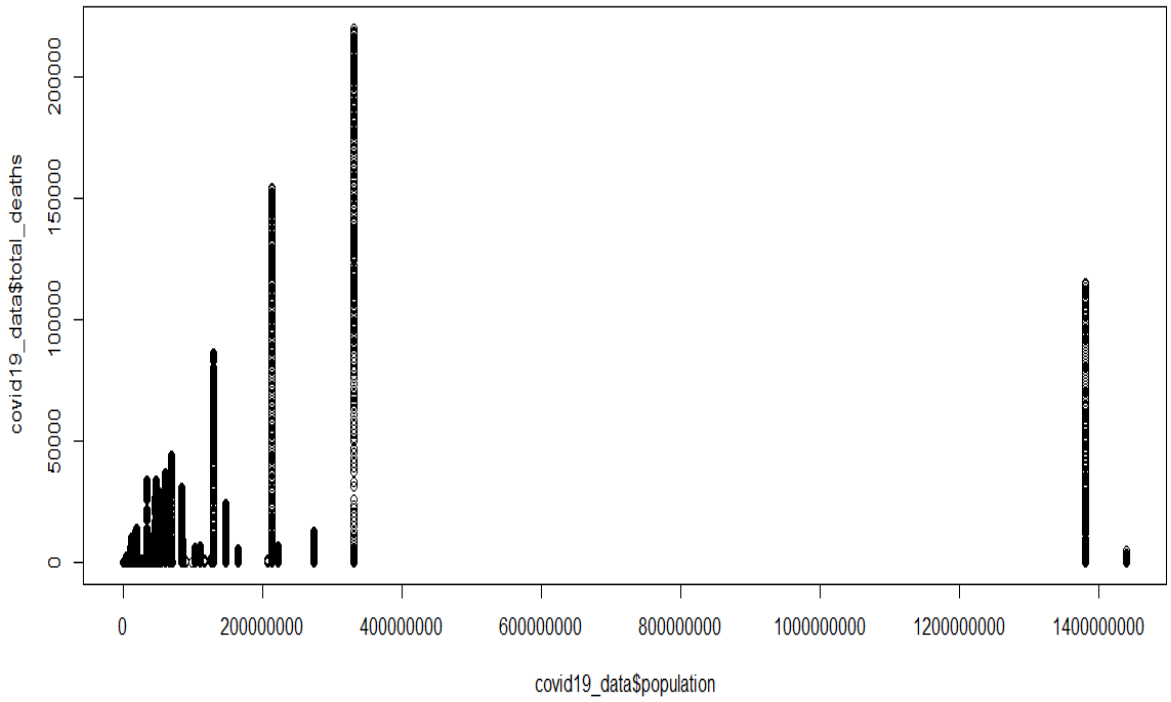

Fig3. Graph

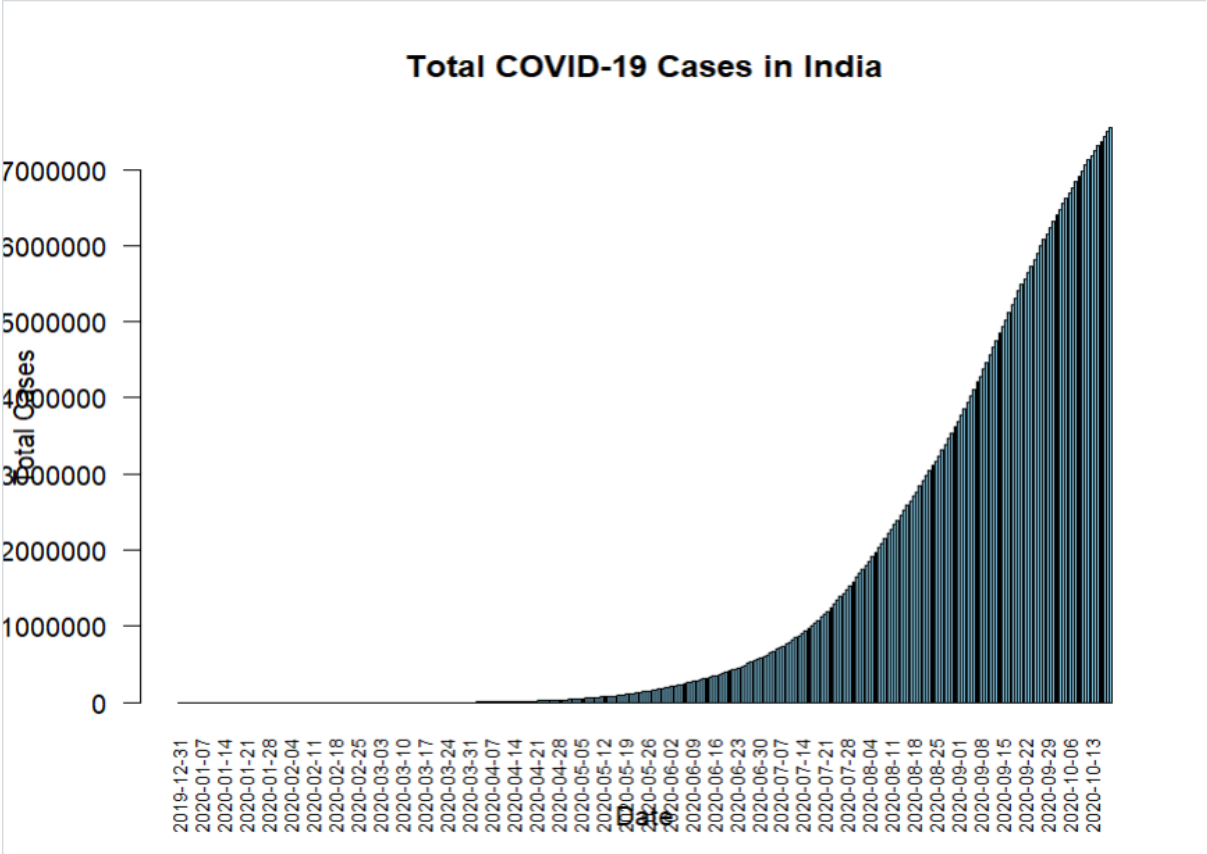

Fig4. Covid India

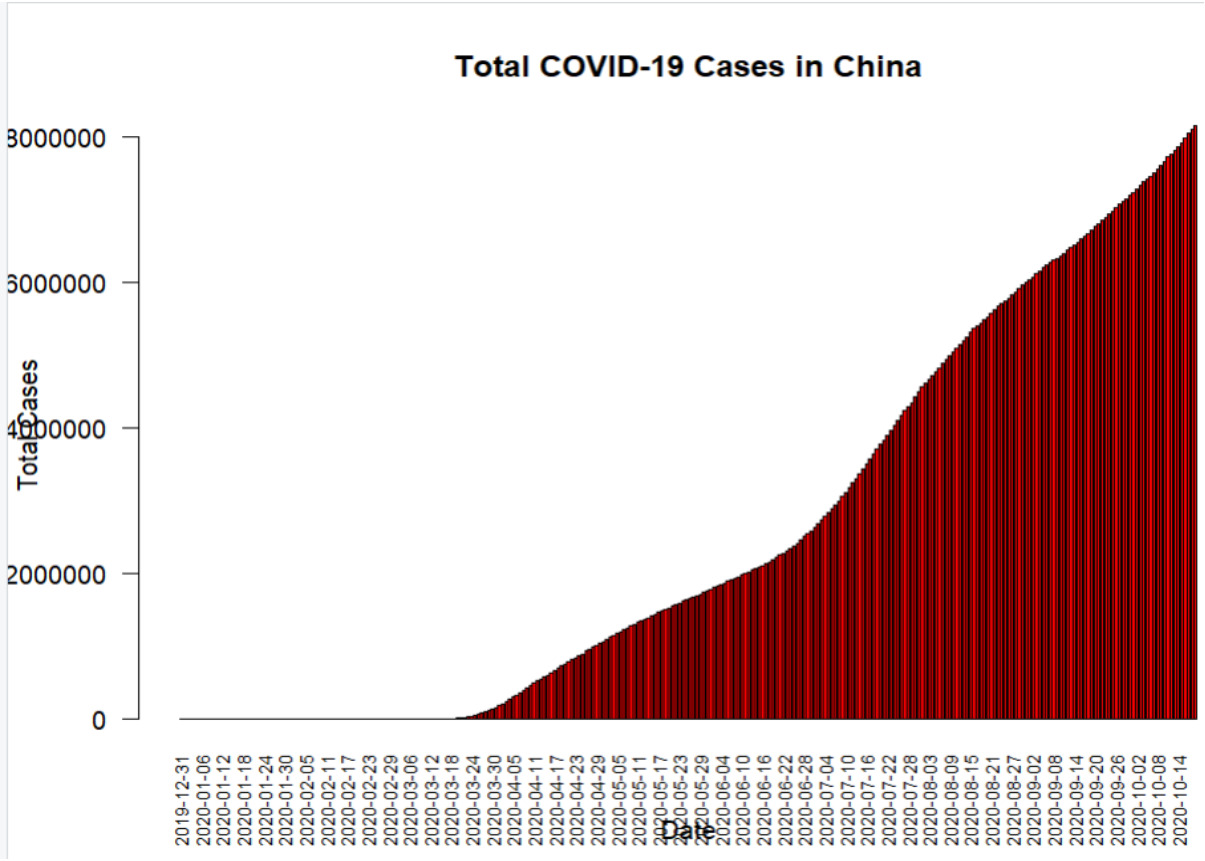

Fig5. USA Covid

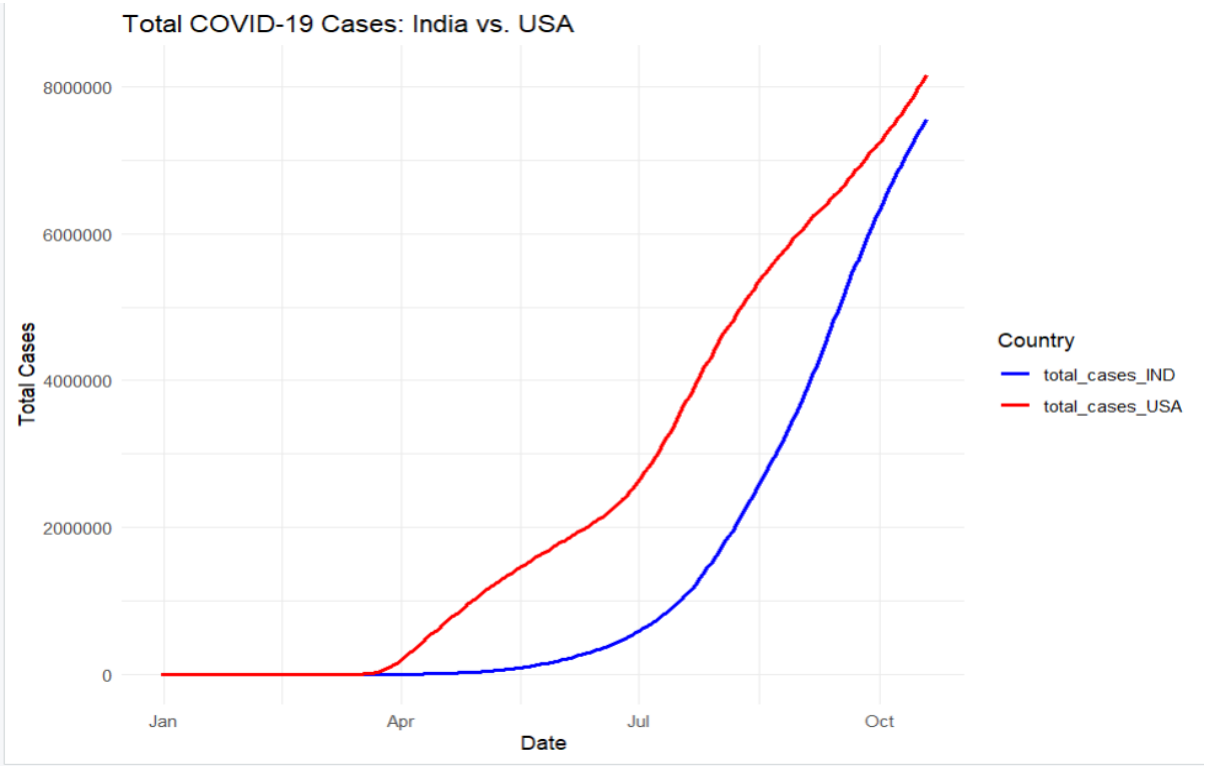

Fig6. COVID Ind v/s USA

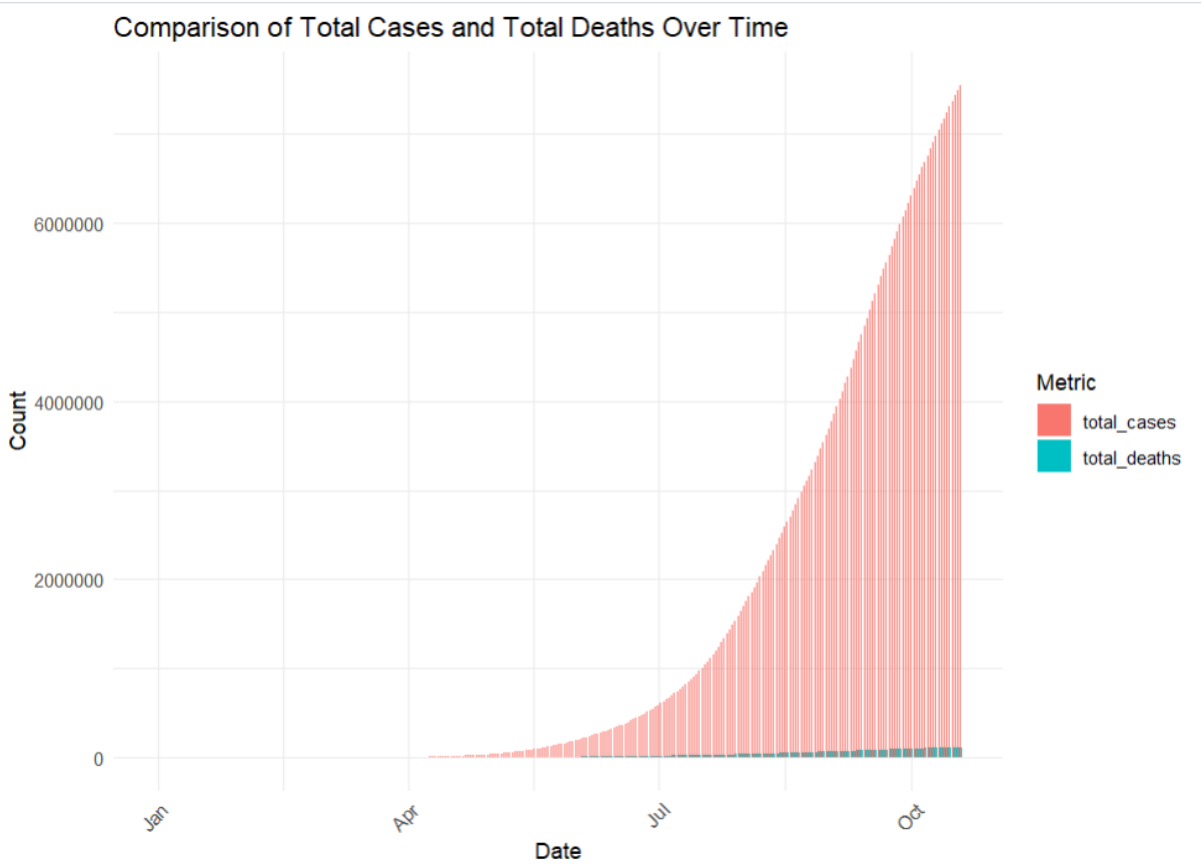

Fig7. COVID Deaths IND v/s USA

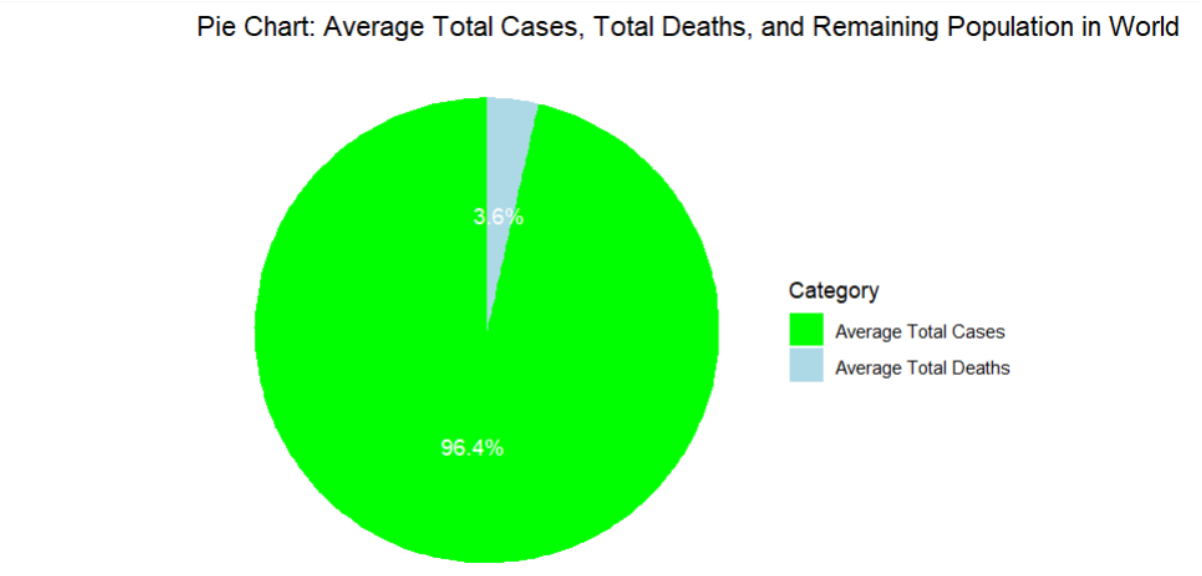

Fig8. Deaths v/s cases around the world

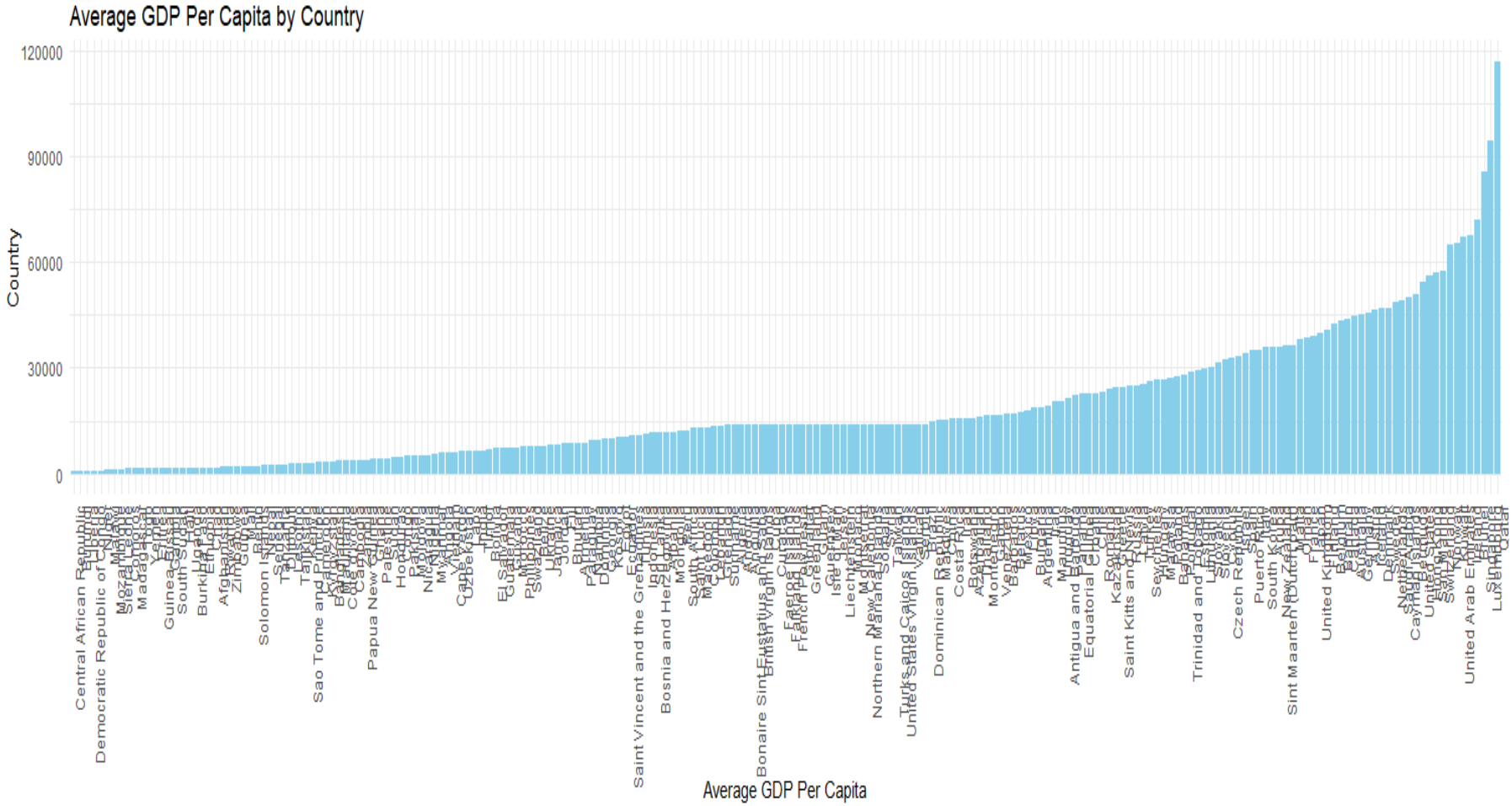
Average GDP Per Capita by Country
120000
90000
60000
30000
0
Country
Average GDP Per Capita

Fig8. GDP

Conclusion:

The project on analysing COVID-19 trends and their economic impacts using big data analytics has provided valuable insights into the complex dynamics of the pandemic across different regions, particularly focusing on India and the USA. Through the application of advanced data manipulation and visualization techniques with R, we have successfully addressed key research questions and identified significant patterns and correlation. “And that we Find Qatar is having the Highest GDP and Central Africa is having the lowest”.

Trends in COVID-19 Spread: Our analysis revealed distinct trends in the spread of COVID-19, highlighting peak periods of infection and regional variations. Visualizations facilitated a clear understanding of how case numbers and mortality rates fluctuated over time, allowing for the identification of critical periods requiring intervention.

Economic Impact Assessment: The investigation into the economic repercussions of the pandemic underscored the substantial effects on key economic indicators, including GDP growth and deprecation.

References:

H. Kang, "Analysis of the Impact of COVID-19 on the U.S. Economy Based on Big Data Processing and Experimental Analyses," 2021 2nd International Conference on Big Data Economy and Information Management (BDEIM), Sanya, China, 2021, pp. 365-370, doi: 10.1109/BDEIM55082.2021.00080. keywords: {COVID-19;Epidemics;Economic indicators;Big Data;Market research;Information management;History;COVID-19;economy;United States;GDP;unemployment rate},[1]

Rajan, Gupta., Saibal, K., Pal. (2020). Trend Analysis and Forecasting of COVID-19 outbreak in India. medRxiv, doi: 10.1101/2020.03.26.20044511[2]

Muhammad, Qasim., Waqas, Ahmad., Minami, Yoshida., Maree, Gould., Muhammad, Yasir. (2020). Analysis of the Worldwide Corona Virus (COVID-19) Pandemic Trend; A Modelling Study to Predict Its Spread. medRxiv, doi: 10.1101/2020.03.30.20048215[3]

Parag, Verma., Ankur, Dumka., Anuj, Bhardwaj., Alaknanda, Ashok., Mukesh, Chandra, Kestwal., Praveen, Kumar. (2021). A Statistical Analysis of Impact of COVID19 on the Global Economy and Stock Index Returns. 2(1):27-27. doi: 10.1007/S42979-020-00410-W[4]

Michael, König., Adalbert, Winkler. (2021). COVID-19: Lockdowns, Fatality Rates and GDP Growth: Evidence for the First Three Quarters of 2020. Intereconomics, 56(1):32-39. doi: 10.1007/S10272-021-0948-Y[5]