Visualizing Time Series Dataset Of COVID-19

Souhardya Mukherjee

Btech Biotechnology

Amity University Kolkata

Period of Internship: 25th August 2025 - 19th September 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

I. Abstract

This project analyzes the global impact of the COVID-19 pandemic through comprehensive data visualization. Using datasets of confirmed cases and deaths, We applied Python libraries such as Pandas, Matplotlib, Seaborn, and Plotly to uncover key trends and patterns. Static visualizations included line plots, pie charts, and heat maps to highlight daily case surges, country-level death distributions, and regional temporal dynamics. An interactive dashboard built with Plotly provided flexibility to explore new cases, deaths, and cumulative trends across different countries and WHO regions.These visualizations not only enhance understanding of the scale and progression of COVID-19 but also demonstrate the value of data-driven approaches in monitoring global health crises.

II. Introduction

The COVID-19 pandemic has been one of the most significant global health crises in recent history, affecting millions of lives and disrupting societies worldwide. Understanding its progression, impact, and patterns requires not only reliable data but also effective visualization methods. This project was undertaken to analyze global COVID-19 datasets and transform raw data into meaningful insights through visualizations.

The project involved the use of Python as the primary programming language due to its extensive ecosystem of data analysis and visualization libraries. Key technologies included Pandas for data processing, Matplotlib and Seaborn for static visualizations, and Plotly for interactive dashboards. Choropleth maps and heat maps were also used to capture the spatial and temporal intensity of cases and deaths.

Before starting the main tasks, background material surveys were conducted to study existing COVID-19 dashboards developed by organizations such as the WHO, Johns Hopkins University, and Our World in Data. These references helped in framing the scope of the project and guided decisions on the types of plots and dashboards that would be most informative.

The procedure followed included:

1. Cleaning and preparing the dataset.

2. Aggregating data by different time intervals (daily, monthly, quarterly).

3. Creating static visualizations such as line plots, bar charts, and pie charts.

4. Developing comparative and distribution-based charts like stacked bars and heat maps.

5. Building an interactive COVID-19 dashboard with Plotly to allow flexible exploration of trends.

The purpose of this project is to demonstrate how data visualization can make large-scale health data easier to interpret, compare, and communicate. By using multiple visualization approaches, this work highlights the scale, intensity, and geographic spread of the pandemic while providing a platform for further exploration.

Training Topics During First Two Weeks of Internship

During the initial phase of the internship, we received training on a range of foundational and technical topics that equipped us to carry out this project effectively. These included:

1. Basic Python operations – covering data types, functions, and control structures.

2. Object-Oriented Programming (OOP) in Python – classes, objects, inheritance, and modular coding.

3. Machine Learning fundamentals – introduction to supervised and unsupervised learning.

4. Communication skills – effective technical writing, presentation skills, and teamwork strategies.

This training provided both the technical and professional foundation needed to execute the project successfully.

III. Project Objective

The main objectives of this project are:

- To analyze and visualize the global impact of COVID-19 by using reliable datasets of new cases and deaths, highlighting temporal and regional variations.
- To illustrate key patterns and trends such as daily surges, quarterly intensity, regional differences, and the distribution of deaths across the top affected countries.
- To develop interactive tools (using Plotly dashboards) that allow users to dynamically filter, compare, and explore the data from multiple perspectives.
- To evaluate the effectiveness of visualization in simplifying large-scale health data for better interpretation, communication, and decision-making.

No hypothesis testing or sample surveys were performed in this project, as the analysis was entirely based on secondary datasets.

IV. Methodology

The methodology of this project was designed to systematically analyze COVID-19 data and visualize patterns at global, regional, and country levels. The work involved multiple phases, from data collection and cleaning to visualization and dashboard creation. The following subsections describe each step in detail:

1. Data Collection

The dataset was sourced from publicly available repositories (e.g., WHO COVID-19 dataset).

The data included attributes such as Date_reported, Country/Region, WHO_region, New_cases, Cumulative_cases, New_deaths, and Cumulative_deaths.Since no primary survey was conducted, no questionnaire or sampling methodology was applied. The project relied entirely on secondary data.

2. Data Preprocessing and Cleaning

- Inspected the dataset for missing values and inconsistencies.
- Handled null or invalid values (e.g., replacing with 0 where appropriate).
- Converted date fields (Date_reported) into proper datetime format for time-series analysis.
- Added additional derived columns such as month and quarter for aggregation.
- Grouped data by countries, regions, and time intervals (daily, monthly, quarterly).

3. Tools and Technologies Used

- Python: The primary programming language used for analysis.
- Libraries:
  - Pandas → Data manipulation and aggregation.
  - Matplotlib & Seaborn → Static plots (line charts, bar charts, pie charts, heatmaps).
  - Plotly → Interactive visualizations and dashboard creation.
  - Google Colab: Coding environment for development and testing.

4. Data Analysis and Visualization Steps

- Line plots → To show daily new cases globally and for least/most affected countries.
- Bar charts → To compare quarterly cases and deaths across regions.
- Double bar charts → Side-by-side comparison of cases vs deaths.
- Pie charts → To illustrate top 10 countries by cumulative deaths.
- Heat maps → For quarterly deaths by region and monthly cases by top 10 countries
- Interactive dashboard with Plotly → To allow users to filter, compare, and explore cases and deaths dynamically.

V. Data Analysis and Results

COVID-19 data analysis, results and insights:

COVID-19 - Descriptive Analysis

The dataset covers COVID-19 cases from 2020-03-01 to 2023-08-31, across 240 countries.

A total of 770,122,340 new cases and 6,968,439 deaths are recorded.

COVID-19 - Analysis and Insights

The country most affected by COVID-19 cases is United States of America with 103,436,760 cases.

The country with the highest deaths is United States of America with 1,140,868 deaths.

The highest daily surge in cases was 8,401,963 on 2022-01-30.

The peak quarterly cases occurred in 2022Q1 with 199,602,568 cases.

COVID-19 - Data Analysis and Results

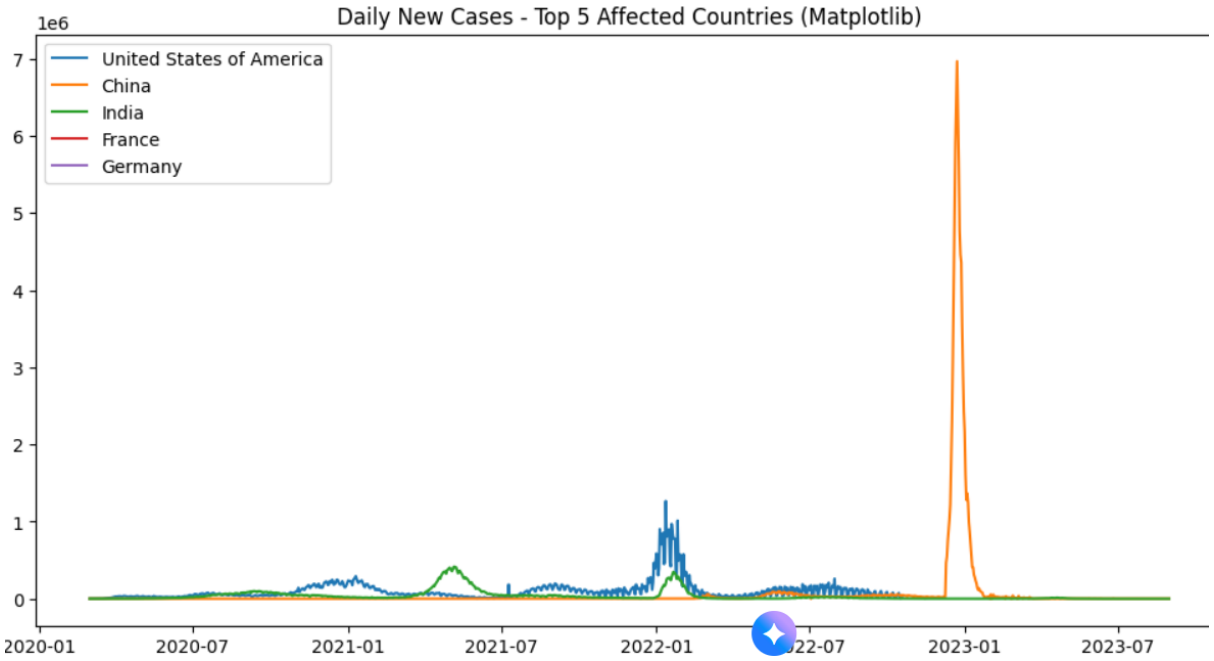
The temporal analysis shows waves of outbreak with clear peaks and declines.

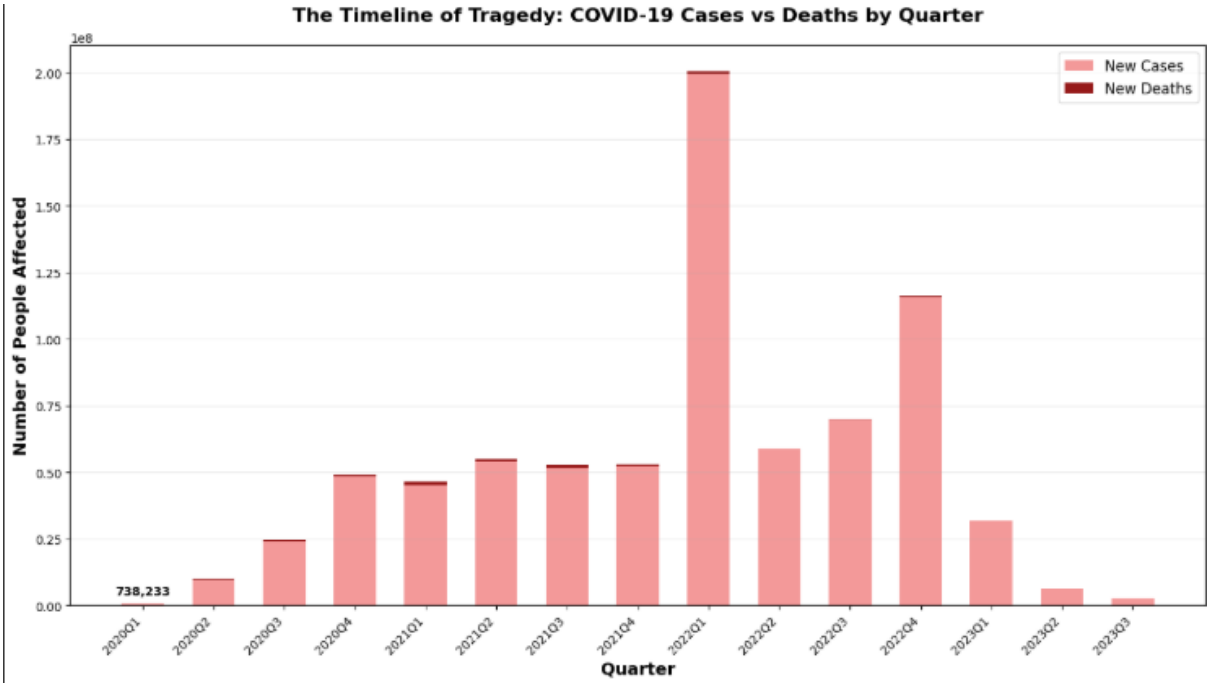
Stacked bar charts indicate that deaths follow similar trends as cases but at lower magnitude.

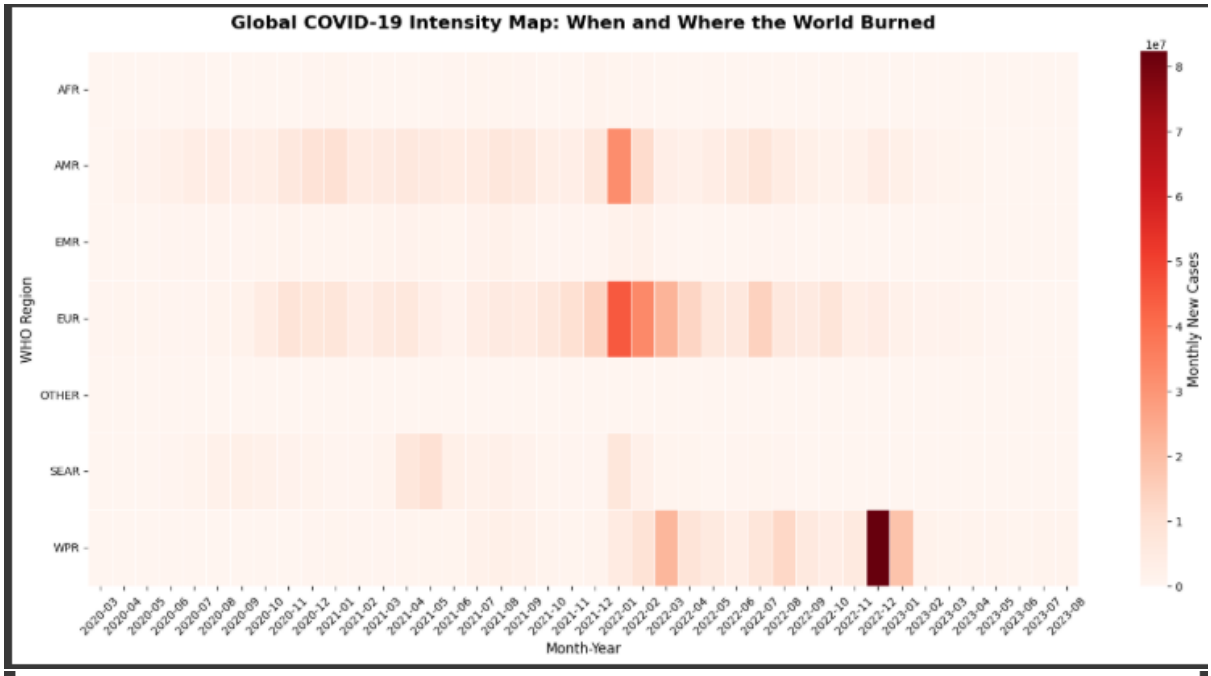
Heatmaps highlight the most affected countries and months, showing concentrated surges in specific regions.

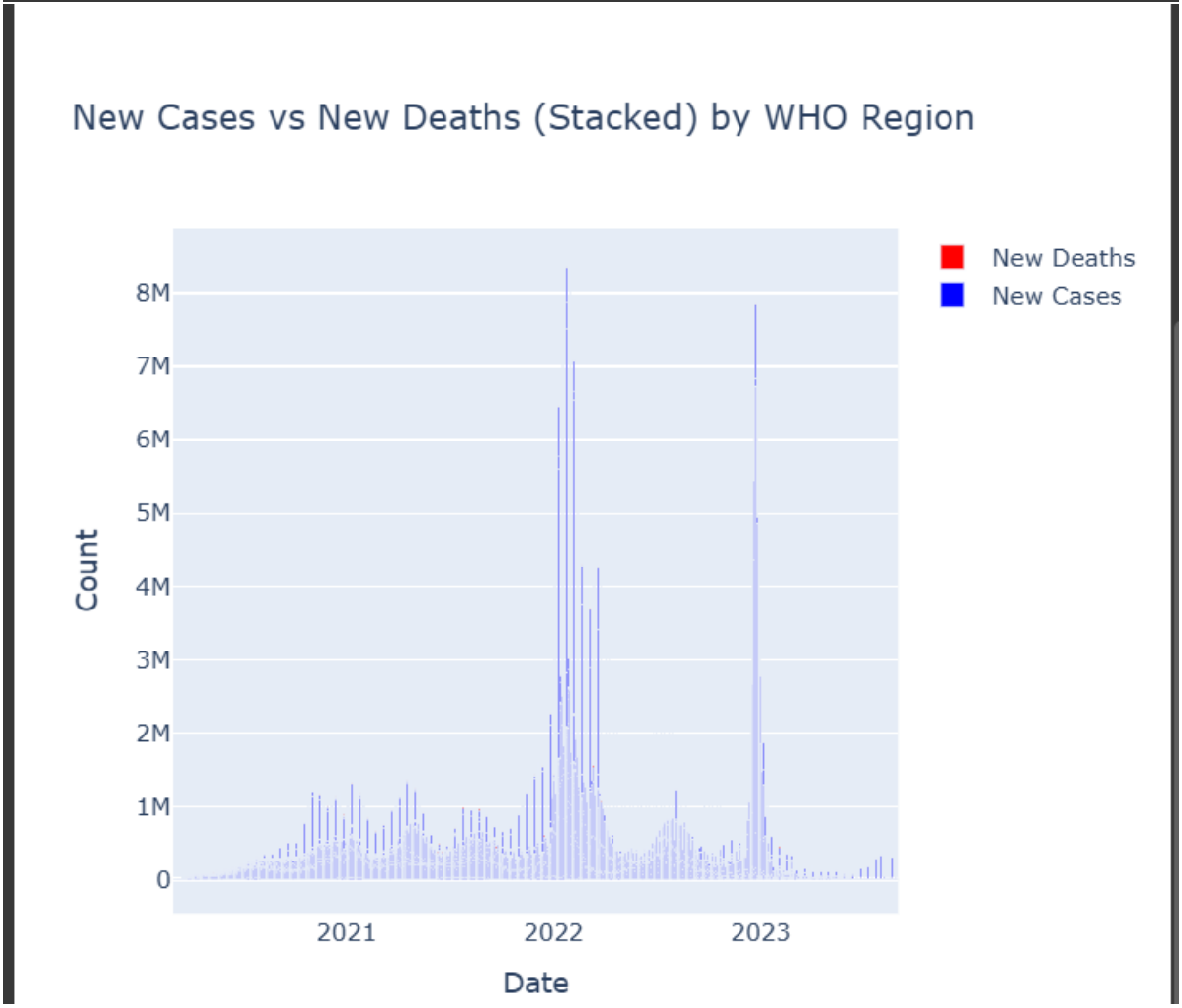
The interactive dashboard provides a comprehensive tool to explore cases, deaths, and distributions dynamically.

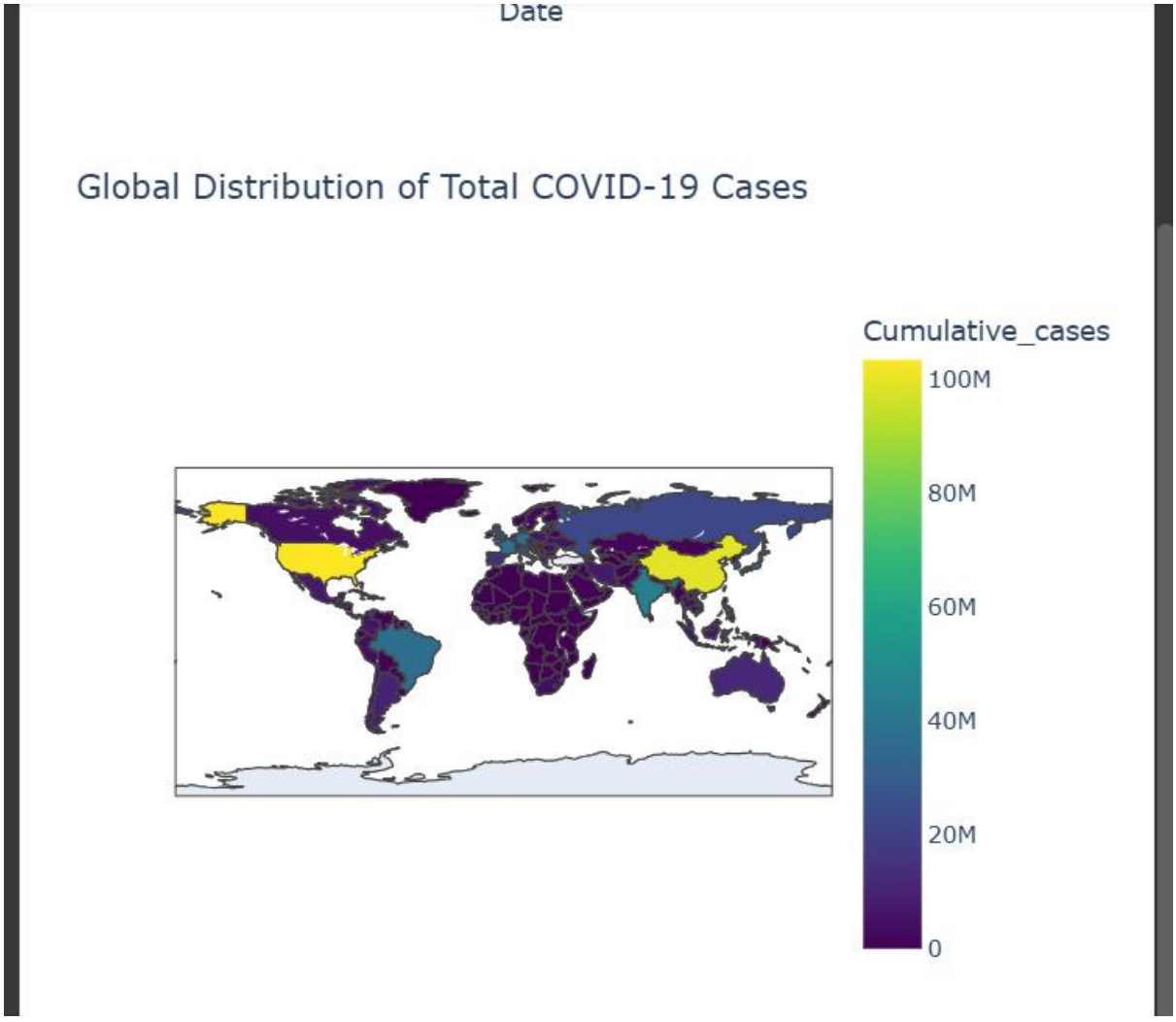
GRAPHS AND CHARTS GENERATED:

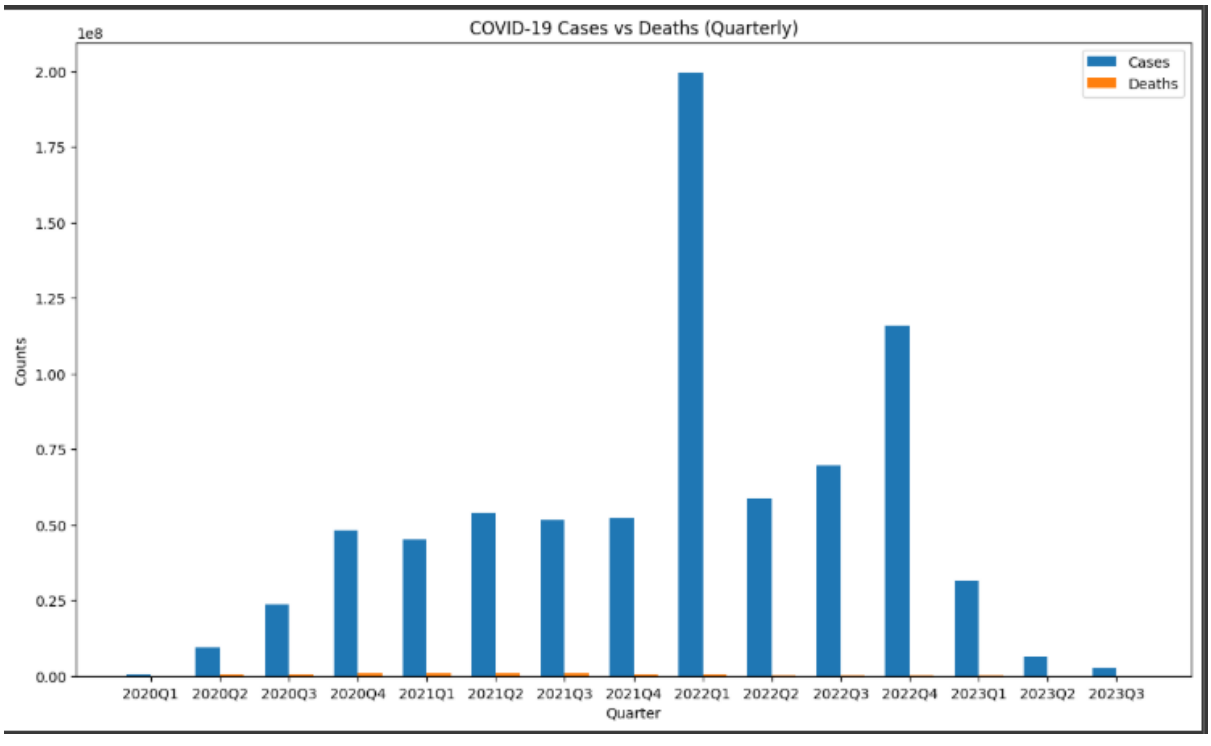


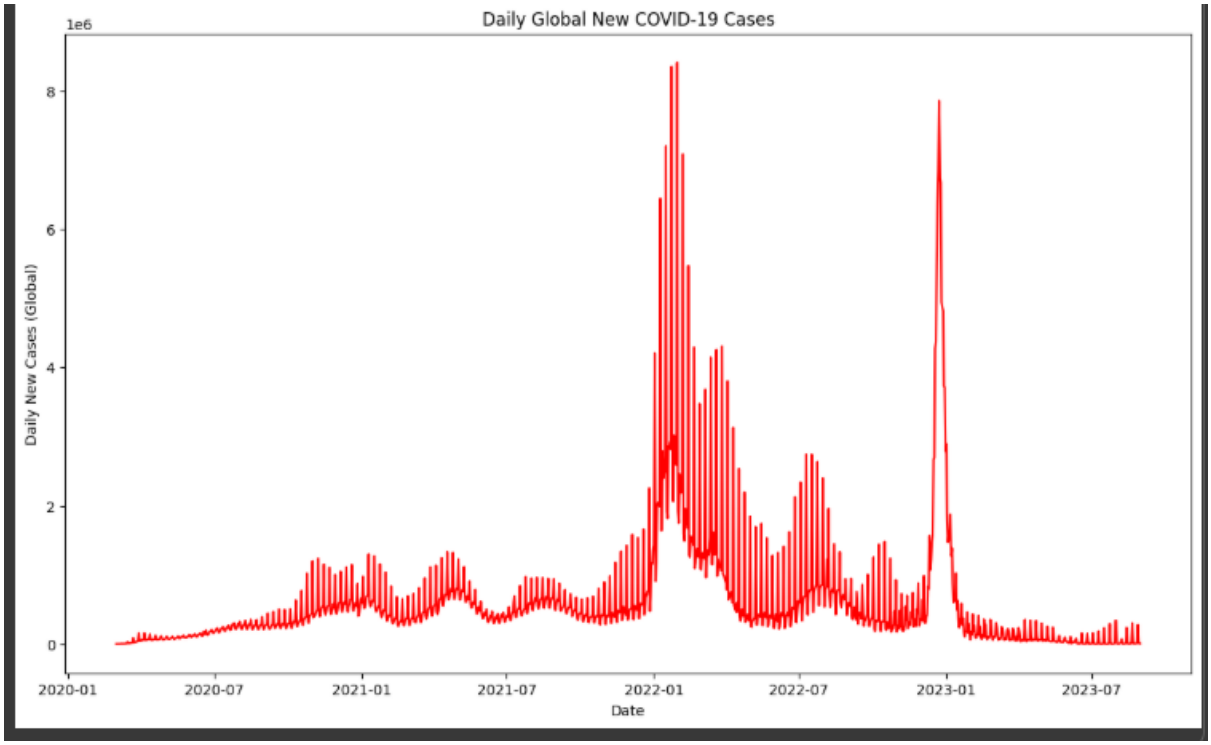


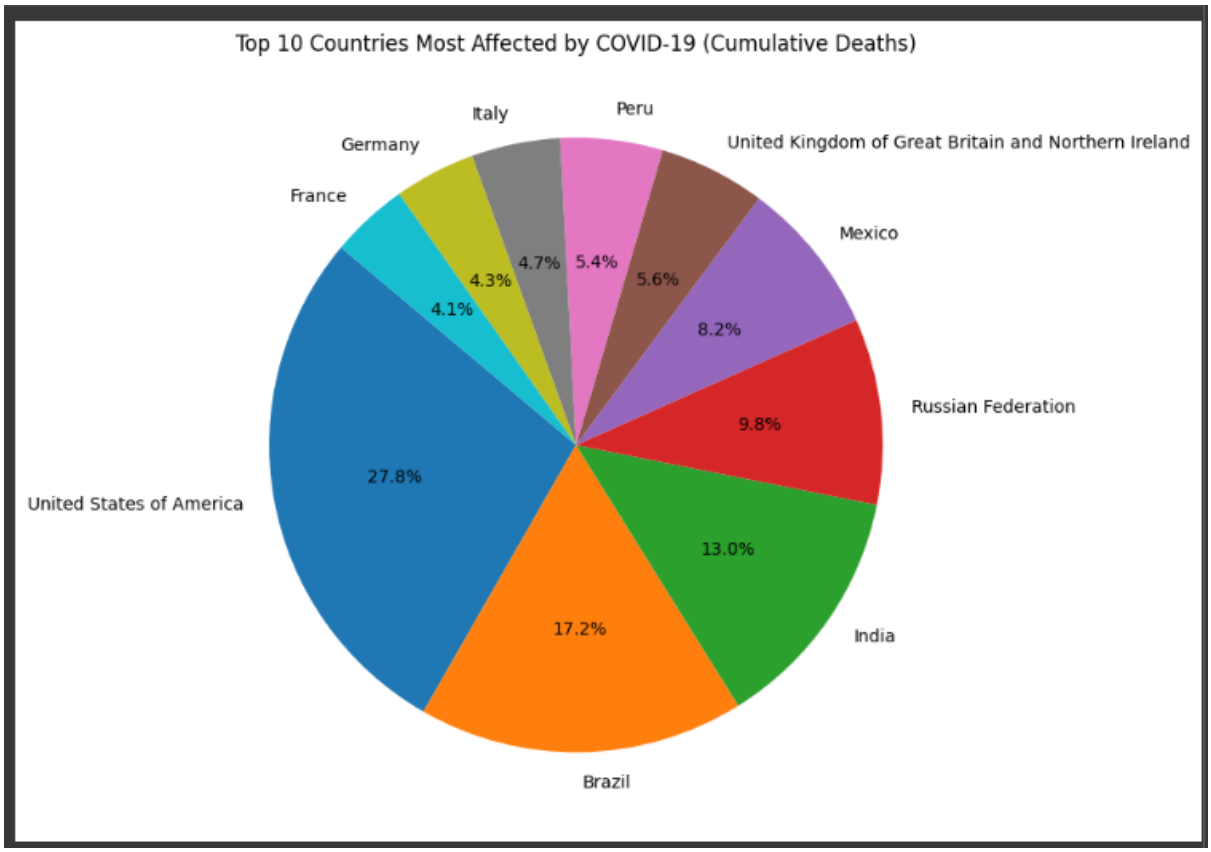

Global COVID-19 Intensity Map: When and Where the World Burned
WHO Region
AFR
AMR
EMR
EUR
OTHER
SEAR
WPR
Month-Year
Monthly New Cases
1e7

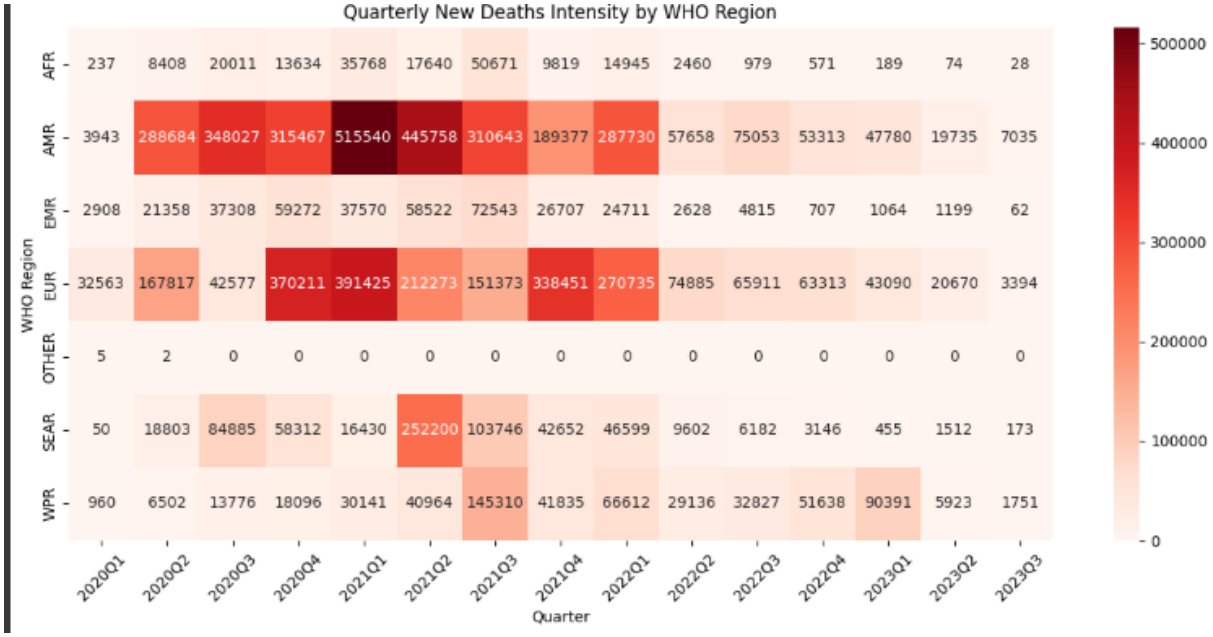

New Cases vs New Deaths (Stacked) by WHO Region
New Deaths
New Cases
Count
8M
7M
6M
5M
4M
3M
2M
1M
0
2021
2022
2023
Date

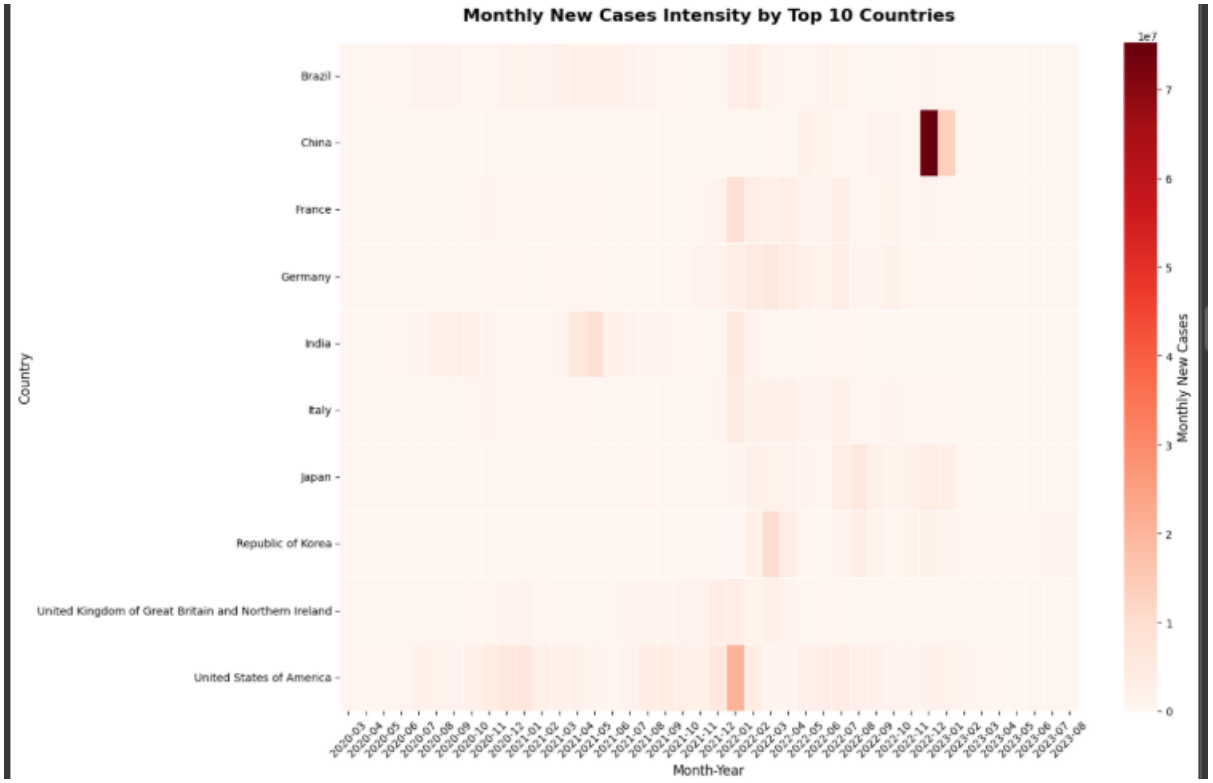

Date
Global Distribution of Total COVID-19 Cases
Cumulative_cases
100M
80M
60M
40M
20M
0

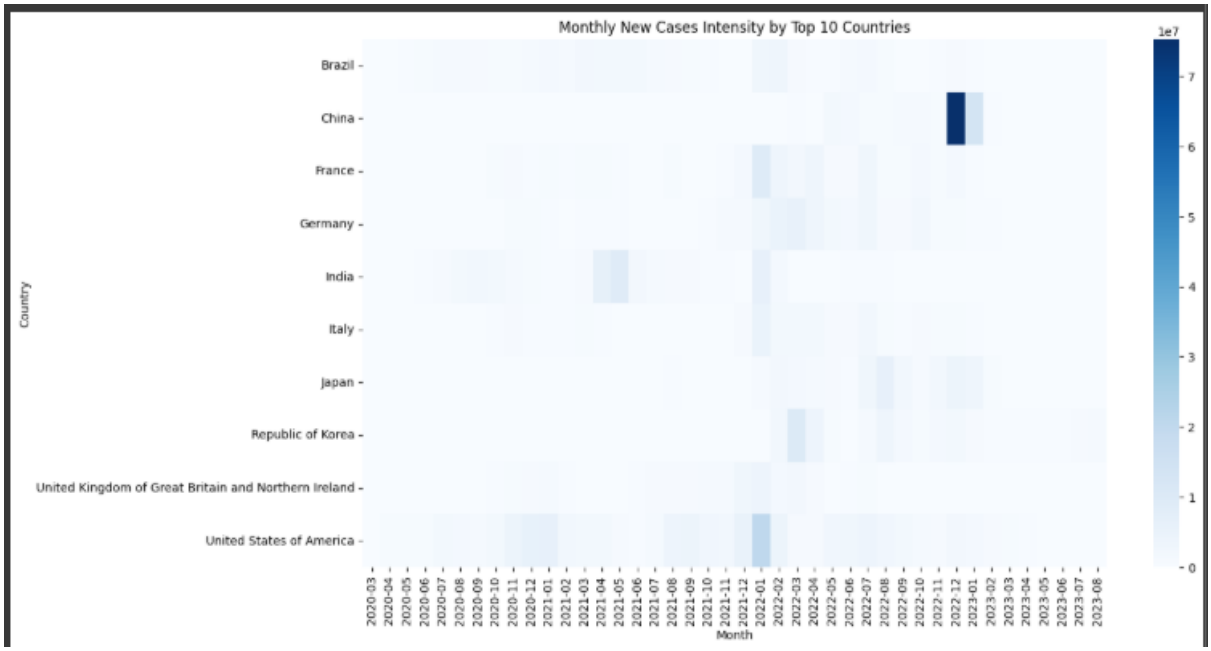

COVID-19 Cases vs Deaths (Quarterly)
1e8
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Counts
Cases
Deaths
2020Q1
2020Q2
2020Q3
2020Q4
2021Q1
2021Q2
2021Q3
2021Q4
2022Q1
2022Q2
2022Q3
2022Q4
2023Q1
2023Q2
2023Q3
Quarter

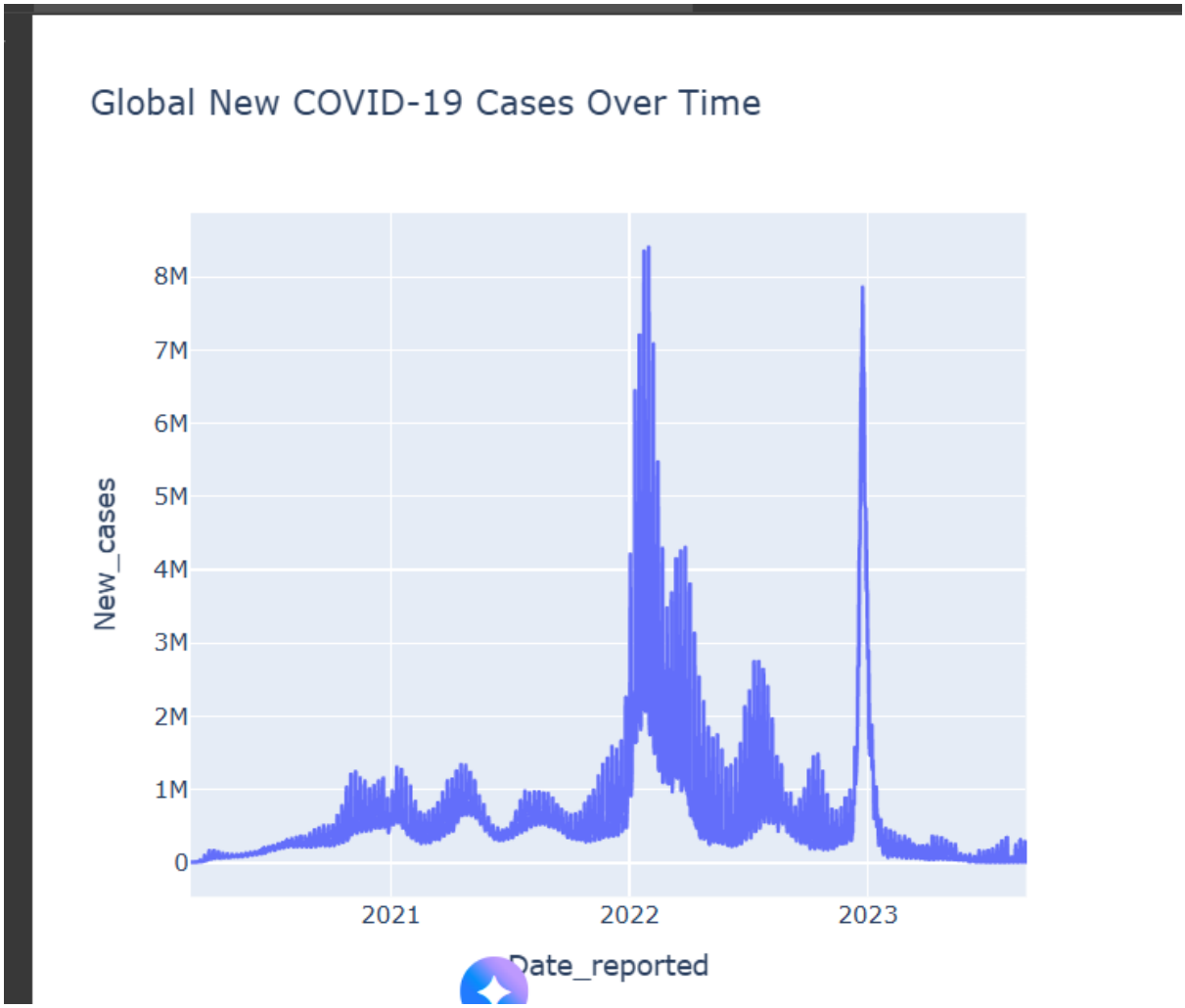

Daily Global New COVID-19 Cases
1e6
Daily New Cases (Global)
Date
0
2
4
6
8
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07

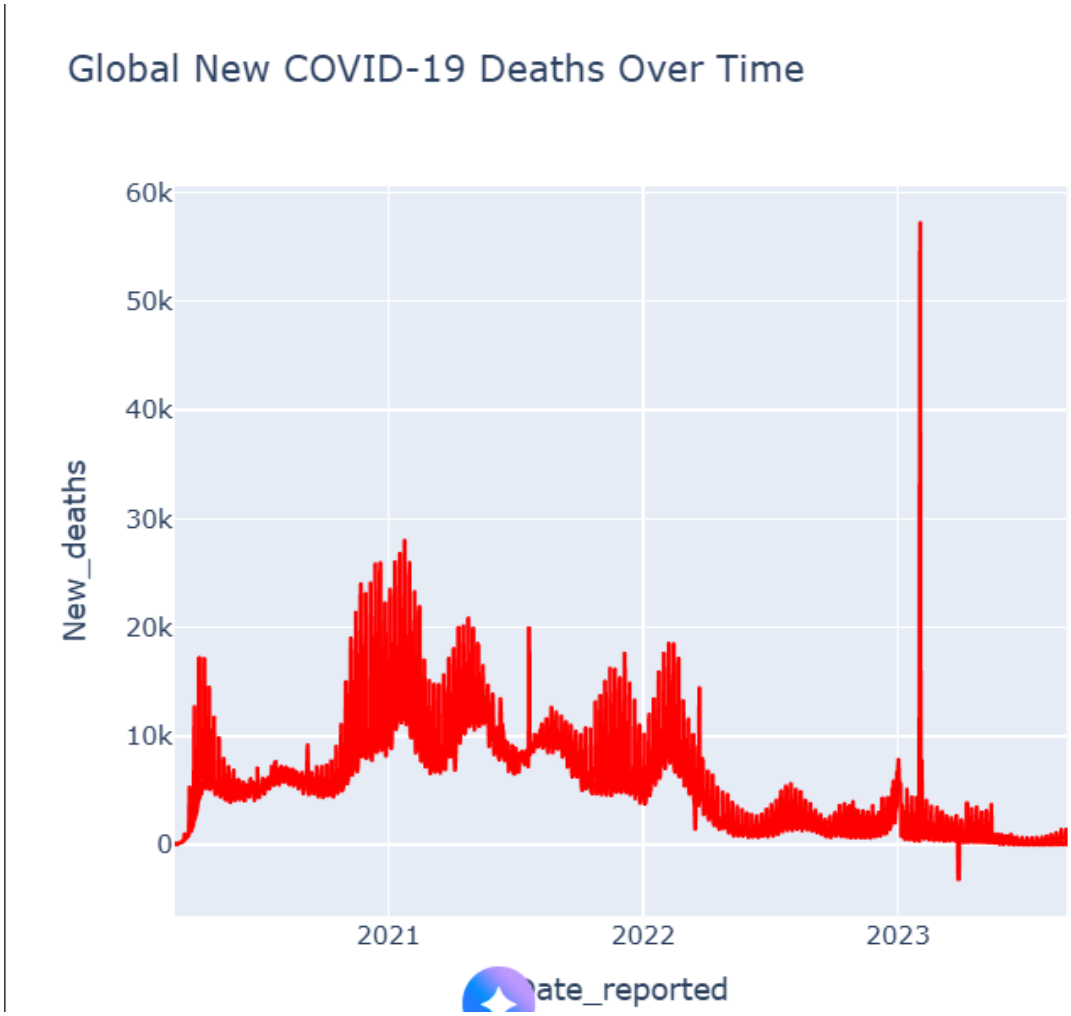

Top 10 Countries Most Affected by COVID-19 (Cumulative Deaths)
Italy
Peru
Germany
United Kingdom of Great Britain and Northern Ireland
France
Mexico
4.7%
5.4%
4.3%
5.6%
4.1%
8.2%
Russian Federation
9.8%
27.8%
United States of America
13.0%
India
17.2%
Brazil


Quarterly New Deaths Intensity by WHO Region
WHO Region
AFR 237 8408 20011 13634 35768 17640 50671 9819 14945 2460 979 571 189 74 28
AMR 3943 288684 348027 315467 515540 445758 310643 189377 287730 57658 75053 53313 47780 19735 7035
EMR 2908 21358 37308 59272 37570 58522 72543 26707 24711 2628 4815 707 1064 1199 62
EUR 32563 167817 42577 370211 391425 212273 151373 338451 270735 74885 65911 63313 43090 20670 3394
OTHER 5 2 0 0 0 0 0 0 0 0 0 0 0 0 0
SEAR 50 18803 84885 58312 16430 252200 103746 42652 46599 9602 6182 3146 455 1512 173
WPR 960 6502 13776 18096 30141 40964 145310 41835 66612 29136 32827 51638 90391 5923 1751
2020Q1 2020Q2 2020Q3 2020Q4 2021Q1 2021Q2 2021Q3 2021Q4 2022Q1 2022Q2 2022Q3 2022Q4 2023Q1 2023Q2 2023Q3
Quarter
500000
400000
300000
200000
100000
0


Monthly New Cases Intensity by Top 10 Countries
Country
Brazil
China
France
Germany
India
Italy
Japan
Republic of Korea
United Kingdom of Great Britain and Northern Ireland
United States of America
Month-Year
1e7
7
6
5
4
3
2
1
0
Monthly New Cases


Monthly New Cases Intensity by Top 10 Countries
Brazil
China
France
Germany
India
Italy
Japan
Republic of Korea
United Kingdom of Great Britain and Northern Ireland
United States of America
Country
Month
1e7


Global New COVID-19 Cases Over Time
New_cases
8M
7M
6M
5M
4M
3M
2M
1M
0
2021
2022
2023
Date_reported



Ebola data analysis, results and insights:

Ebola - Descriptive Analysis

The dataset covers Ebola cases from 2014-08-29 to 2016-03-23, across 10 countries.

A total of 836,618 new cases and 375,702 deaths are recorded.

Ebola - Analysis and Insights

The country most affected by Ebola cases is Liberia with 819,411 cases.

The country with the highest deaths is Liberia with 369,614 deaths.

The highest daily surge in cases was 10,749 on 2015-07-30.

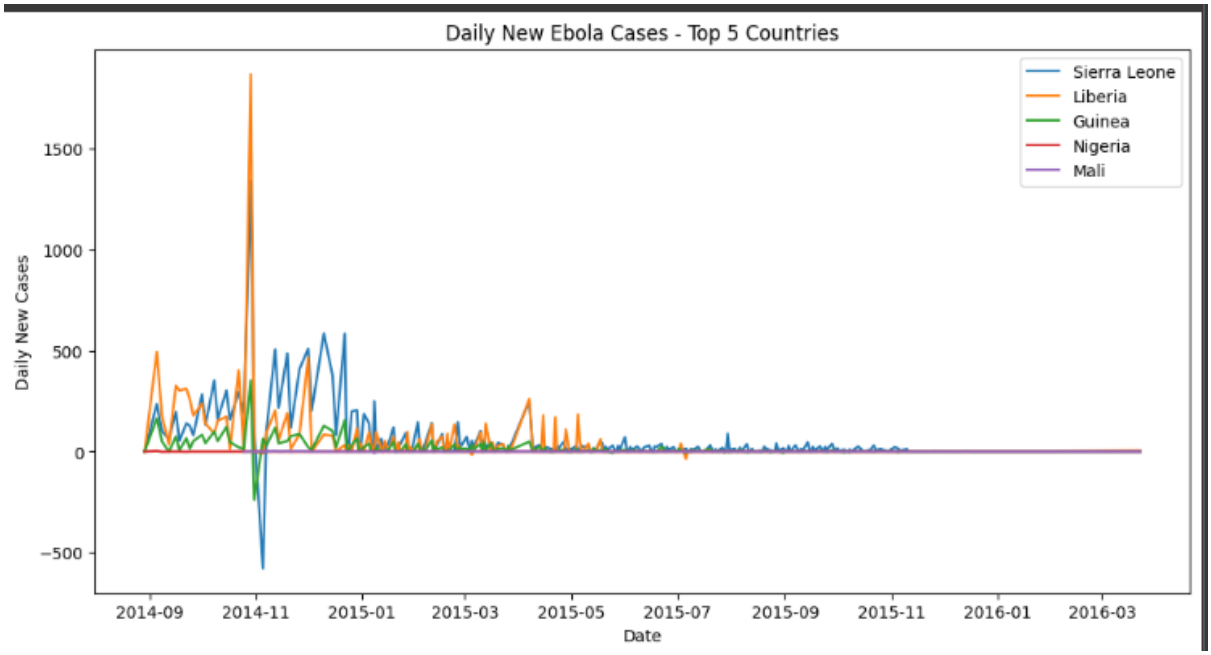
The peak quarterly cases occurred in 2015Q3 with 491,253 cases.

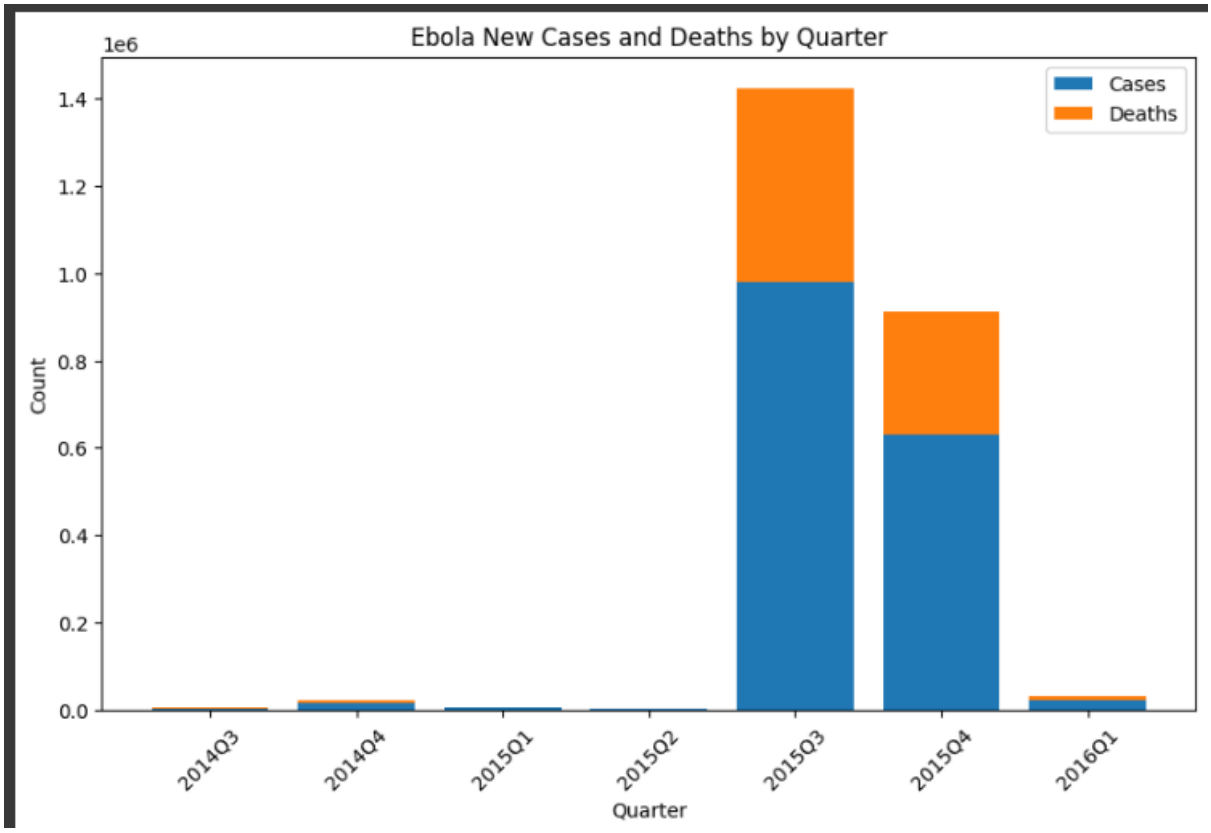
Ebola - Data Analysis and Results

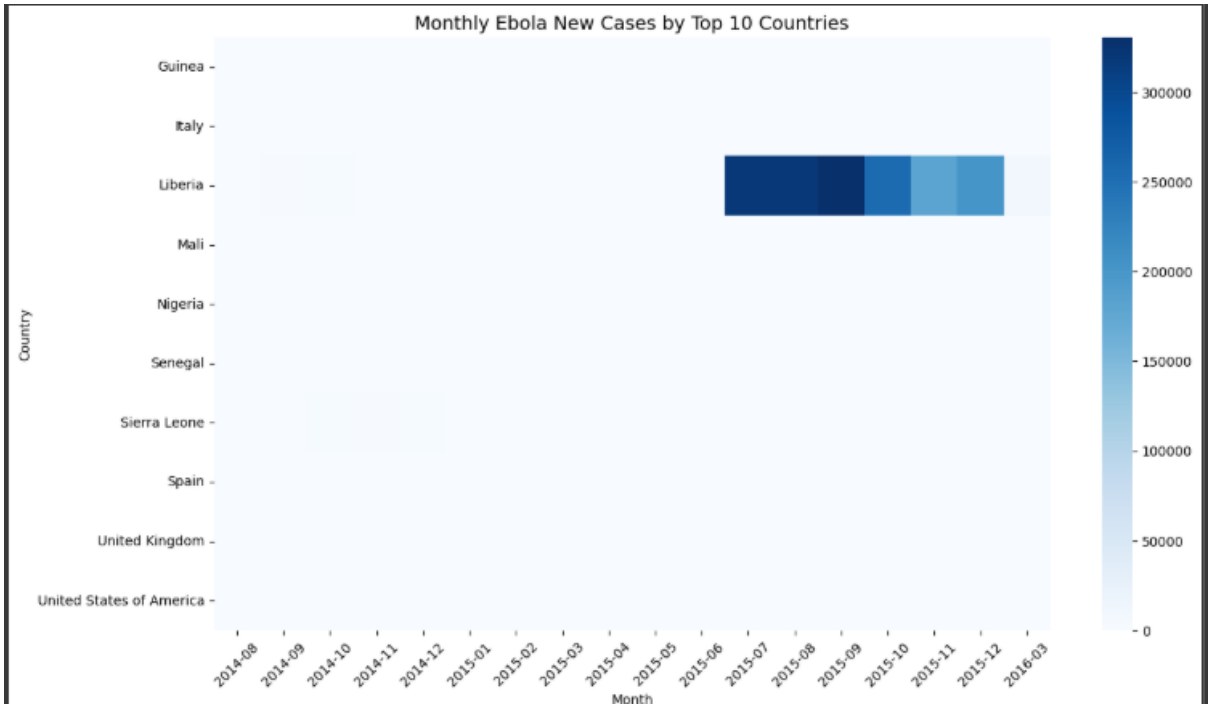
The temporal analysis shows waves of outbreak with clear peaks and declines.

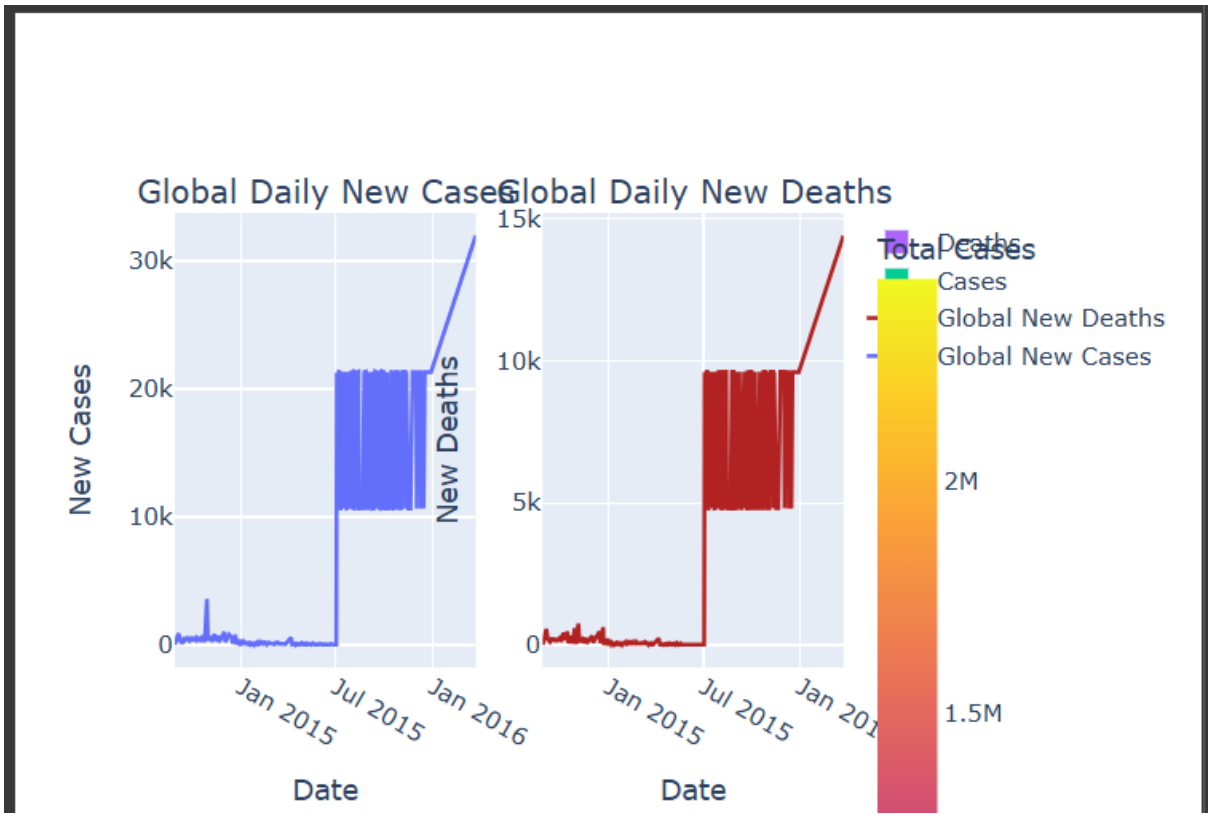
Stacked bar charts indicate that deaths follow similar trends as cases but at lower magnitude.

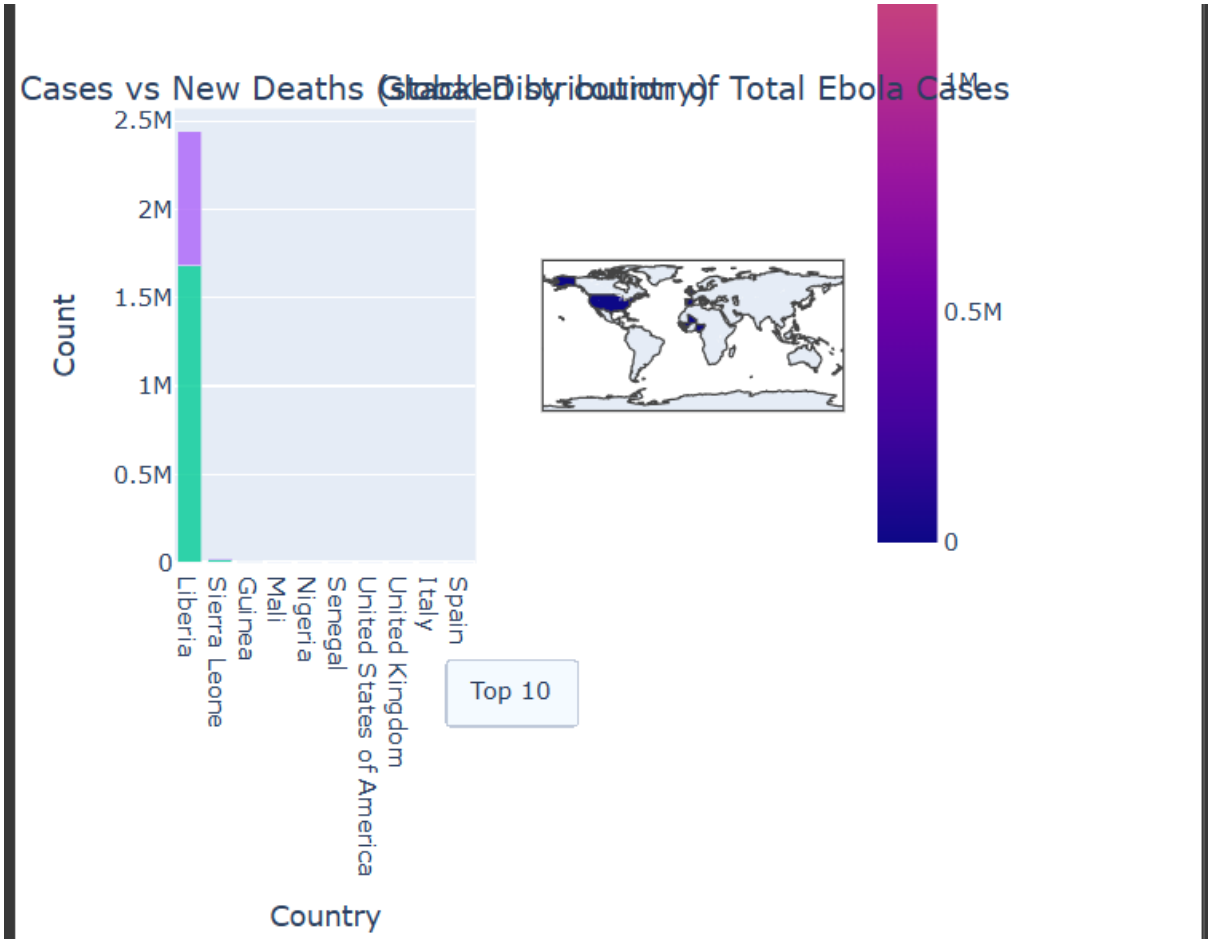
GRAPHS AND CHARTS GENERATED:


Daily New Ebola Cases - Top 5 Countries
Sierra Leone
Liberia
Guinea
Nigeria
Mali
Daily New Cases
1500
1000
500
0
−500
2014-09
2014-11
2015-01
2015-03
2015-05
2015-07
2015-09
2015-11
2016-01
2016-03
Date


Ebola New Cases and Deaths by Quarter
1e6
Cases
Deaths
Count
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2014Q3
2014Q4
2015Q1
2015Q2
2015Q3
2015Q4
2016Q1
Quarter


Monthly Ebola New Cases by Top 10 Countries
Guinea
Italy
Liberia
Mali
Nigeria
Senegal
Sierra Leone
Spain
United Kingdom
United States of America
Country
2014-08
2014-09
2014-10
2014-11
2014-12
2015-01
2015-02
2015-03
2015-04
2015-05
2015-06
2015-07
2015-08
2015-09
2015-10
2015-11
2015-12
2016-03
Month
300000
250000
200000
150000
100000
50000
0


Global Daily New Cases
Global Daily New Deaths
New Cases
New Deaths
30k
20k
10k
0
15k
10k
5k
0
Jan 2015
Jul 2015
Jan 2016
Jan 2015
Jul 2015
Date
Date
Total Cases
Deaths
Cases
Global New Deaths
Global New Cases
2M
1.5M



Heatmaps highlight the most affected countries and months, showing concentrated surges in specific regions.

The interactive dashboard provides a comprehensive tool to explore cases, deaths, and distributions dynamically.

VI. Conclusion

The project successfully demonstrated how systematic data analysis and visualization can uncover meaningful insights into the COVID-19 pandemic. By applying Python tools and visualization libraries, we were able to transform raw datasets into clear, interpretable patterns.

Our findings showed that:

The top 10 most affected countries by cumulative deaths were consistently those with high population density and international connectivity, confirming global observations. The quarterly heat map of deaths clearly highlighted peak mortality periods during major pandemic waves, with certain WHO regions experiencing more severe surges.The monthly heat map of new cases for top countries revealed contrasting patterns of outbreak intensity, reflecting differences in containment measures and vaccination progress. The global daily cases line plot resembled a mountain-like trajectory, emphasizing how the pandemic unfolded in successive

waves.The interactive dashboard proved effective in offering dynamic exploration, enabling policymakers and researchers to compare trends across regions and timelines with greater flexibility.

From these outcomes, it is concluded that visualization and dashboard-based approaches make pandemic data more understandable, actionable, and engaging compared to static reports.

VII. APPENDICES

References

1. World Health Organization (WHO). Coronavirus (COVID-19) Dashboard. Retrieved from: https://covid19.who.int

2. Johns Hopkins University. COVID-19 Global Data Repository.

3. Python Libraries: Pandas, Numpy, Matplotlib, Seaborn, Plotly documentation.

4. Kaggle Ebola Dataset

GitHub Repository

All Python scripts and Jupyter Notebooks developed for this project have been uploaded to GitHub:

🔗 [GitHub Repository Link – https://github.com/Souhardya-Mukherjee/IDEAS-TIH-ISI-Kol-covid-19-dataset-project.git]

Contents include:

Data preprocessing scripts

Visualization codes (line plots, bar charts, heat maps, pie charts)

Plotly interactive dashboard code

PDF copy of the final report