

A photograph of a globe with a world map, wearing a blue surgical mask. Two hands, one in a blue glove and one in a red glove, are holding the globe. The globe is oriented with Asia at the top.

Presented By Arpita Deb

# Epidemic Analytics

EVALUATING COVID-19'S  
GLOBAL FOOTPRINT

June 2024

# Contents

01. Introduction

02. Problem Statement

03. Data Description

04. Tools

05. Key Metrics

06. Analysis

07. Conclusion

08. Limitations

09. Future Scope

10. Appendix

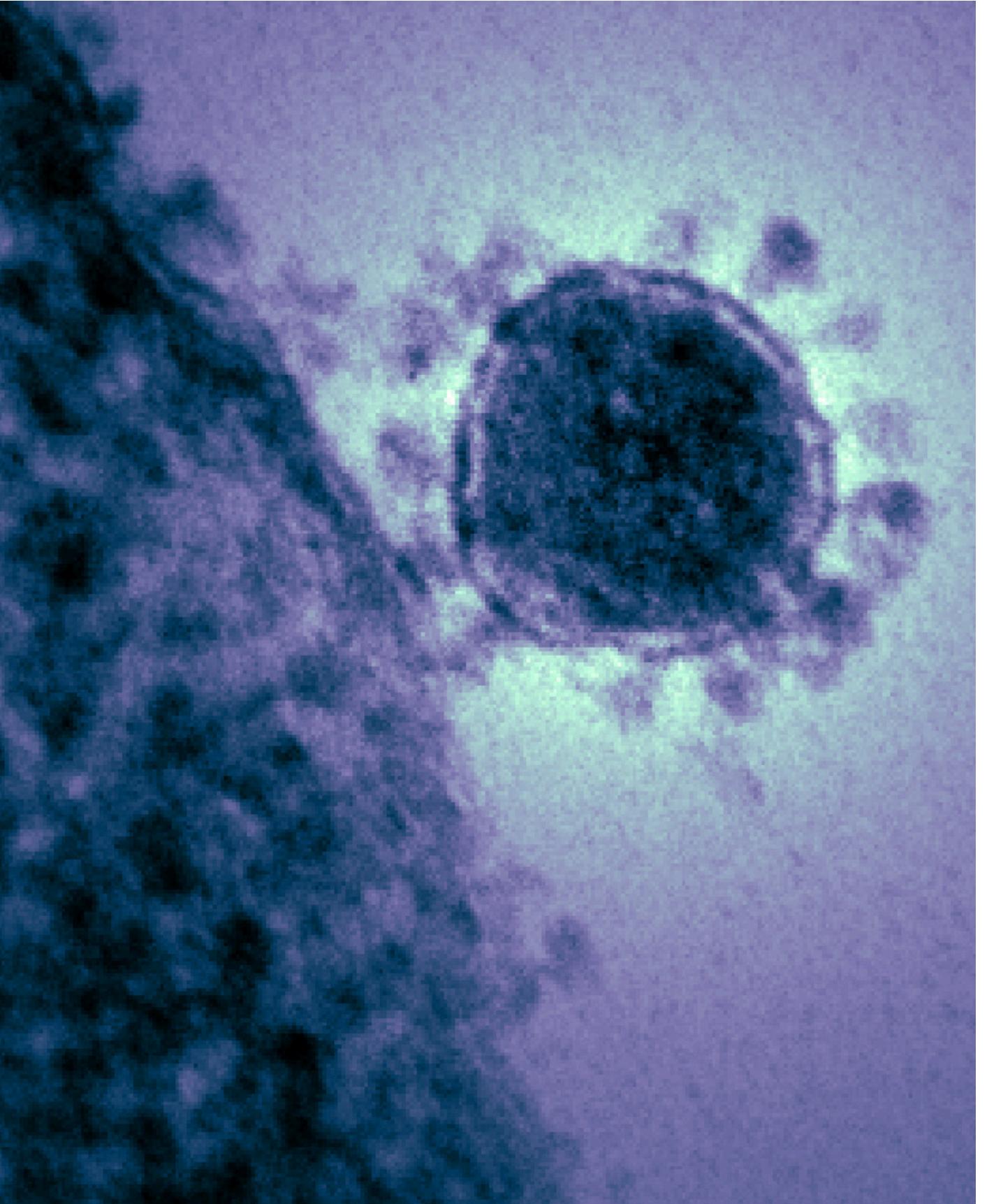
# Introduction

The COVID-19 pandemic, which swept across the globe from early 2020 onwards, led to widespread disruptions in daily life, severe strains on healthcare infrastructure, and significant socio-economic impacts.

This analysis report examines the trajectory of confirmed cases, fatalities, and recoveries from COVID-19 across different countries during the critical period of 2020 to 2021, shedding light on the varying impacts and responses observed globally.

# Problem Statement

“ The Corona Virus pandemic has had a significant impact on public health and has created an urgent need for data-driven insights to understand the spread of the virus. As a data analyst, you have been tasked with analyzing a Corona Virus dataset to derive meaningful insights and present your findings. ”



# Data Description

- **Province:** Geographic subdivision within a country/region.
- **Country/Region:** Geographic entity where data is recorded.
- **Latitude:** North–south position on Earth’s surface.
- **Longitude:** East–west position on Earth’s surface.
- **Date:** Recorded date of Corona Virus data.
- **Confirmed:** Number of diagnosed Corona Virus cases.
- **Deaths:** Number of Corona Virus related deaths.
- **Recovered:** Number of recovered Corona Virus cases.

The Dataset contains 78,387 rows and 8 columns that provides information about the number of confirmed, dead and recovered cases of Corona Virus patients in various countries between January 2020 and June 2021.

# Tools

- |     |                             |                                               |
|-----|-----------------------------|-----------------------------------------------|
| 01. | <b>Microsoft Excel</b>      | For Data Cleaning                             |
| 02. | <b>Microsoft SQL Server</b> | For Data Analysis                             |
| 03. | <b>Power BI</b>             | For Data Visualization and Dashboard creation |
| 04. | <b>Canva</b>                | For creating Project Report                   |
| 05. | <b>GitHub</b>               | For Documentation                             |

# Key Metrics

01. Confirmed Cases per year/month/country

Total and Average Number of Confirmed cases per year, month or country.

02. Death Cases per year/month/country

Total and Average Number of Death cases per year, month or country.

03. Recovered Cases per year/month/country

Total and Average Number of Recovered cases per year, month or country.



# 01. Overview of Confirmed Cases

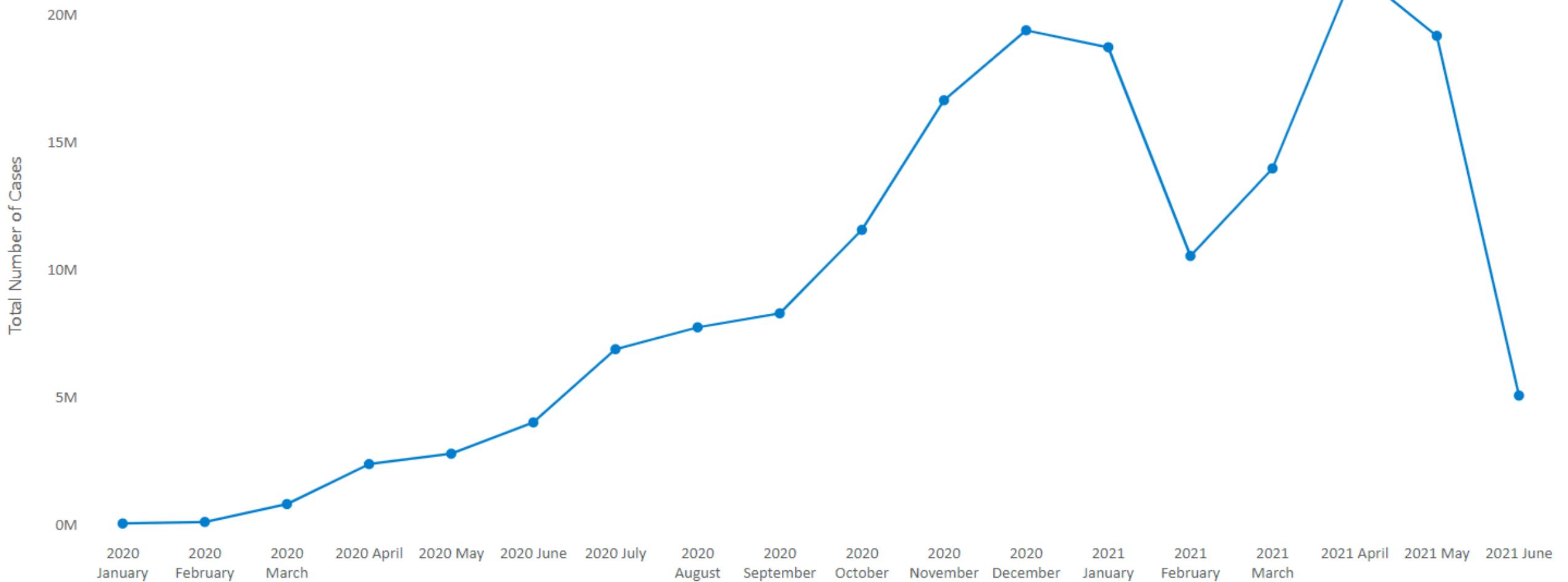
Between January 2020 and June 2021, a staggering **169 million individuals** were identified as COVID-19 positive globally.

# Confirmed Case Dynamics

An initial gradual increase in cases was observed from January to July 2020, followed by a significant surge from October to December. The year 2021 saw a dip in cases from January to March, with a subsequent rise in April.

**Monthly Confirmed Cases from Jan 2020 to June 2021**

Line chart showing the total number of confirmed cases per month.



# Confirmed Cases Statistical Breakdown

The statistics tell us that April saw the most cases overall, December was the worst on average each day, and there was a lot of variation (in terms of variance and standard deviation) in the number of cases from day to day in most months. The situation was quite volatile, with some days having very few cases and others having a very high number of cases.

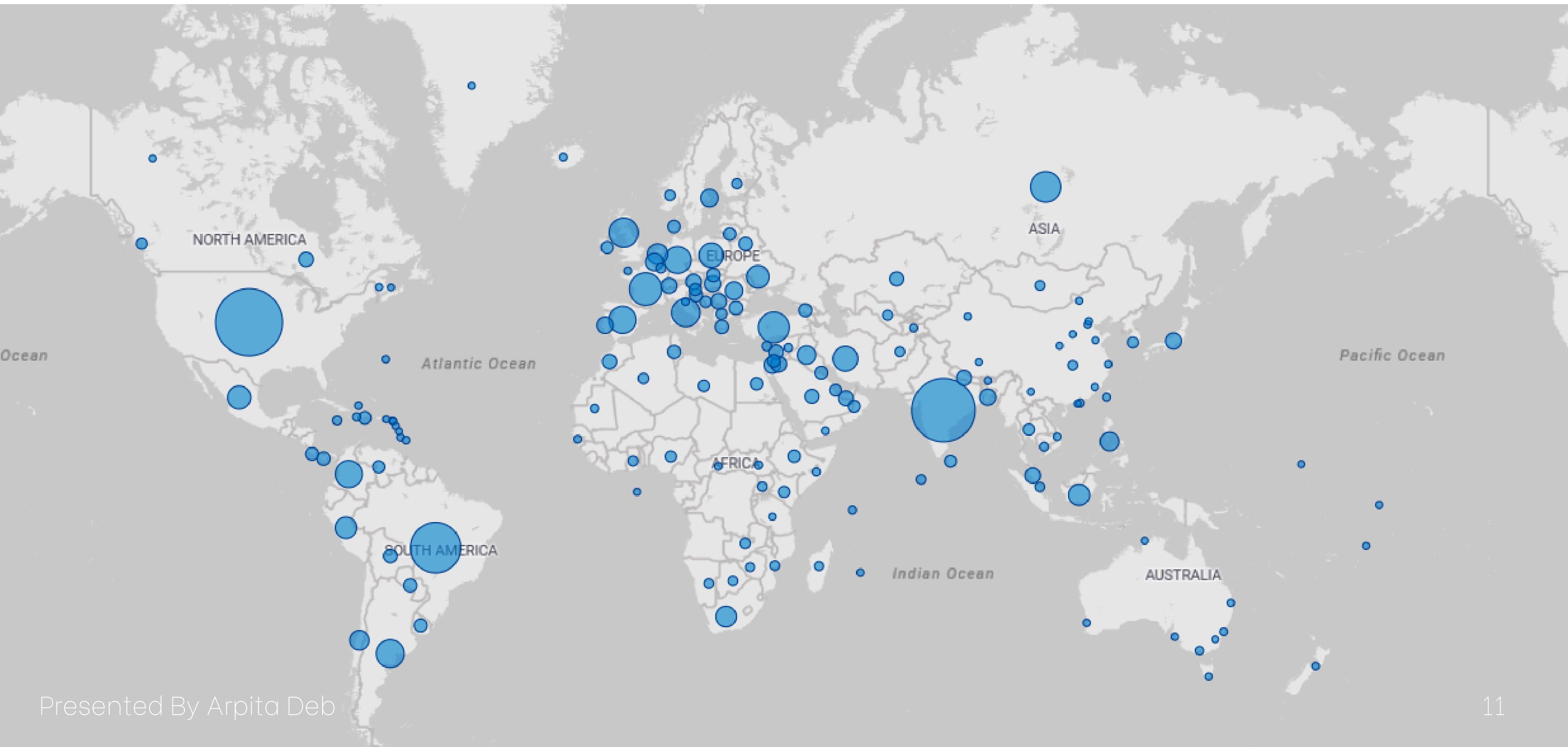
```
WITH OrderedCases AS (
    SELECT [Confirmed],
        ROW_NUMBER() OVER (ORDER BY [Confirmed]) AS RowNum,
        COUNT(*) OVER () AS TotalCount
    FROM [dbo].[CoronaVirusData]
)
SELECT
    MONTH([Date]) AS Month_Num,
    DATENAME(month, [Date]) AS Month,
    SUM([Confirmed]) AS Total_Confirmed_Cases,
    ROUND(AVG([Confirmed]),0) AS Avg_Confirmed_Cases,
    MIN([Confirmed]) AS Min_Confirmed_Cases,
    (SELECT [Confirmed] FROM OrderedCases WHERE RowNum = (TotalCount + 1) / 2) AS Median_Confirmed_Cases,
    MAX([Confirmed]) AS Max_Confirmed_Cases,
    ROUND(VAR([Confirmed]),2) AS Variance,
    ROUND(STDEV([Confirmed]),2) AS Standard_Deviation
FROM
    [dbo].[CoronaVirusData]
GROUP BY
    MONTH([Date]),
    DATENAME(month, [Date])
ORDER BY
    MONTH([Date]),
    DATENAME(month, [Date]);
```

SQL Code for creating the Summary Stats table

Month_Num	Month	Total_Confirmed_Cases	Avg_Confirmed_Cases	Min_Confirmed_Cases	Median_Confirmed_Cases	Max_Confirmed_Cases	Variance	Standard_Deviation
1	January	18678589	2958	0	23	300462	242011851.22	15556.73
2	February	10560976	1203	0	23	134975	40601692.52	6371.95
3	March	14694026	1539	0	23	100158	44278902.42	6654.24
4	April	24047819	2603	0	23	401993	258437079.6	16075.98
5	May	21865416	2290	0	23	414188	320331062.75	17897.79
6	June	8991916	1358	0	23	134154	43731651.84	6612.99
7	July	6838092	1432	0	23	75866	46923851.93	6850.1
8	August	7694938	1612	0	23	85687	54419982.4	7376.99
9	September	8244794	1785	0	23	97894	69329705.03	8326.45
10	October	11515841	2412	0	23	99264	69002612.88	8306.78
11	November	16595938	3592	0	23	207933	195858271.38	13994.94
12	December	19336799	4050	0	23	823225	459981798.11	21447.19

# Geographical Distribution

Visualization of confirmed cases is represented by blue bubbles, with larger sizes indicating higher case counts.

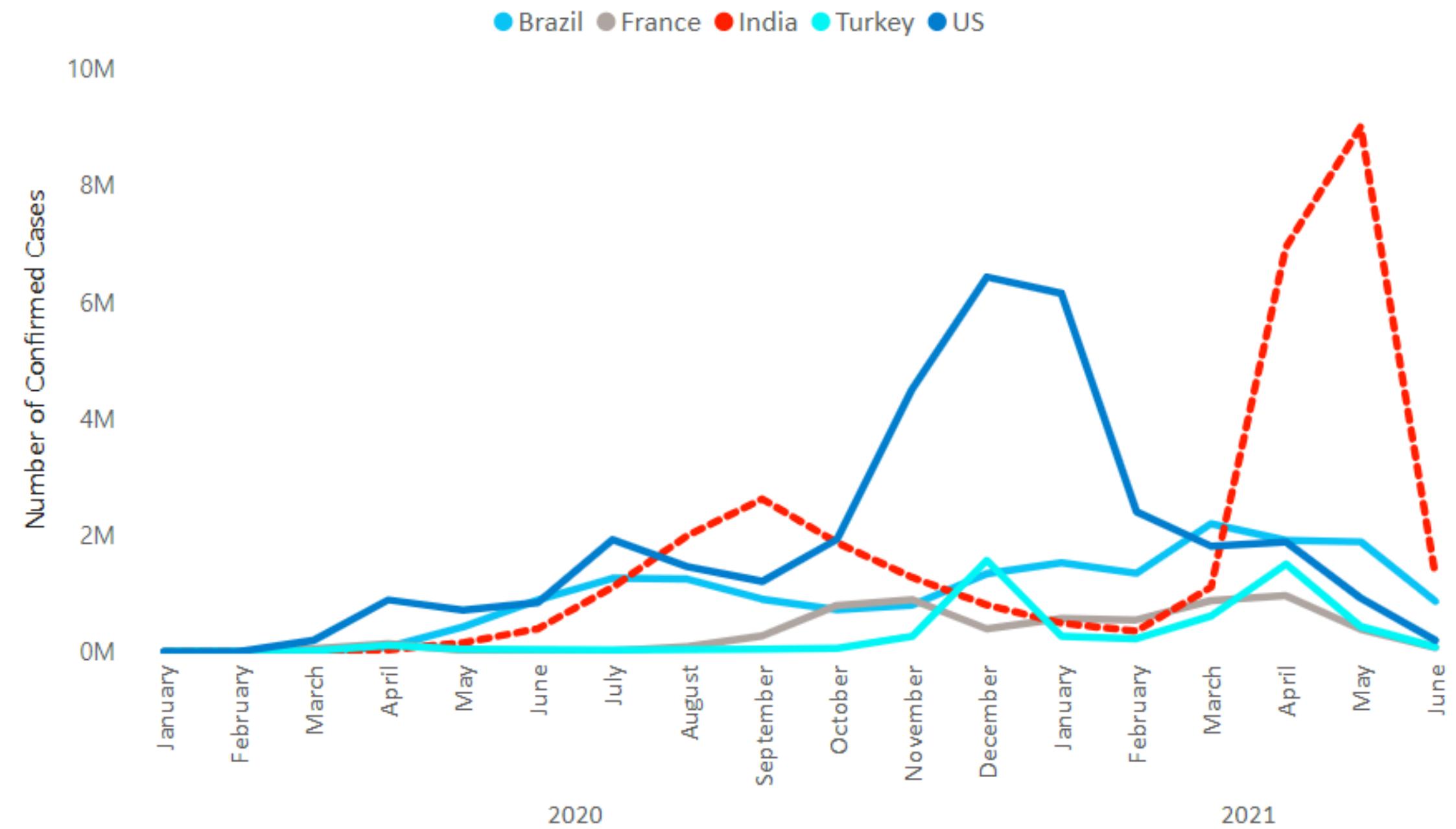


# Top 5 Impacted Nations Trajectory

The United States, India, Brazil, France, and Turkey experienced the highest infection rates.

Monthly trends are depicted through line charts, with India's situation highlighted in red, showing it as the second most affected country with 29 million cases, following the United States' 33 million cases.

**Confirmed Cases in Top 5 Countries from January 2020 to June 2021**



## 02. Mortality Analysis

During the 18-month period, COVID-19 claimed **3.65 million** lives.

# Mortality Trends

A significant spike in deaths was observed in April 2020. Following this, death cases decreased, only to rise steadily from October 2020 to January 2021, reaching a peak. The numbers then fell until March, rose in May, and declined again in June.

**Monthly Death Cases from Jan 2020 to June 2021**

Line chart showing the total number of death cases per month.



# Death Cases Statistical Breakdown

The statistics tell us that April saw the most death cases overall, December was the worst on average each day. There was comparatively lower variation in the number of cases from day to day in most months. The situation was less volatile, as the values tend to be closer to the mean.

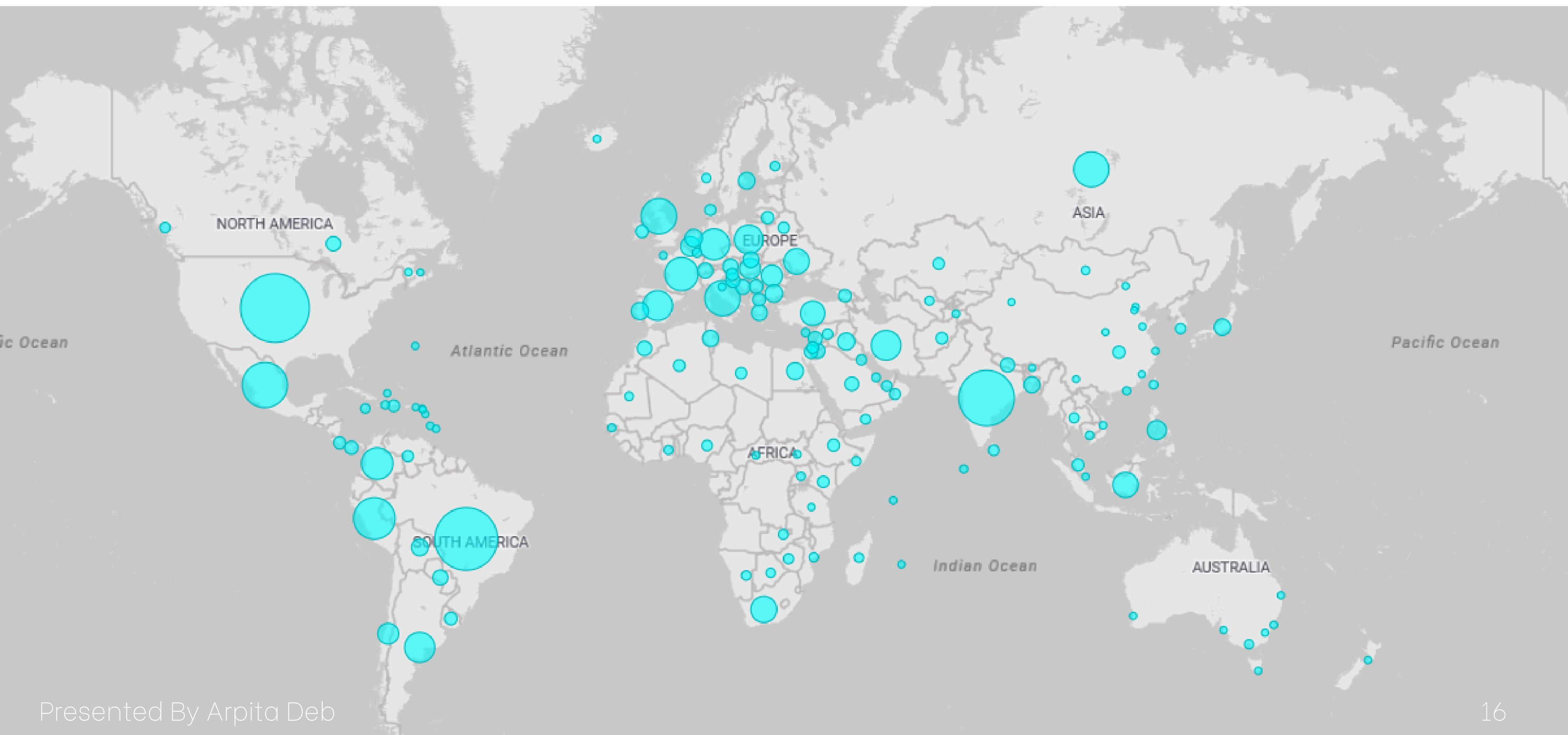
```
WITH OrderedCases AS (
    SELECT [Deaths],
        ROW_NUMBER() OVER (ORDER BY [Deaths]) AS RowNum,
        COUNT(*) OVER () AS TotalCount
    FROM [dbo].[CoronaVirusData]
)
SELECT
    MONTH([Date]) AS Month_Num,
    DATENAME(month, [Date]) AS Month,
    SUM([Deaths]) AS Total_Death_Cases,
    ROUND(AVG([Deaths]),0) AS Avg_Death_Cases,
    MIN([Deaths]) AS Min_Death_Cases,
    (SELECT [Deaths] FROM OrderedCases WHERE RowNum = (TotalCount + 1) / 2) AS Median_Death_Cases,
    MAX([Deaths]) AS Max_Death_Cases,
    ROUND(VAR([Deaths]),2) AS Variance,
    ROUND(STDEV([Deaths]),2) AS Standard_Deviation
FROM
    [dbo].[CoronaVirusData]
GROUP BY
    MONTH([Date]),
    DATENAME(month, [Date])
ORDER BY
    MONTH([Date]),
    DATENAME(month, [Date]);
```

SQL Code for creating the Summary Stats table

Month_Num	Month	Total_Death_Cases	Avg_Death_Cases	Min_Death_Cases	Median_Death_Cases	Max_Death_Cases	Variance	Standard_Deviation
1	January	402083	64	0	0	4475	79012.04	281.09
2	February	300890	34	0	0	3907	34852.62	186.69
3	March	323966	34	0	0	3869	29785.05	172.58
4	April	554220	60	0	0	4249	67905.92	260.59
5	May	511110	54	0	0	4529	76775.78	277.08
6	June	270414	41	0	0	7374	46250.19	215.06
7	July	167613	35	0	0	1595	21144.58	145.41
8	August	179200	38	0	0	1505	23277.87	152.57
9	September	160671	35	0	0	1703	20107.12	141.8
10	October	175484	37	0	0	3351	17583.75	132.6
11	November	262247	57	0	0	2259	27779.81	166.67
12	December	339996	71	0	0	3752	65359.06	255.65

# Geographical Distribution

Visualization of death cases is represented by teal bubbles, with larger sizes indicating higher case counts.



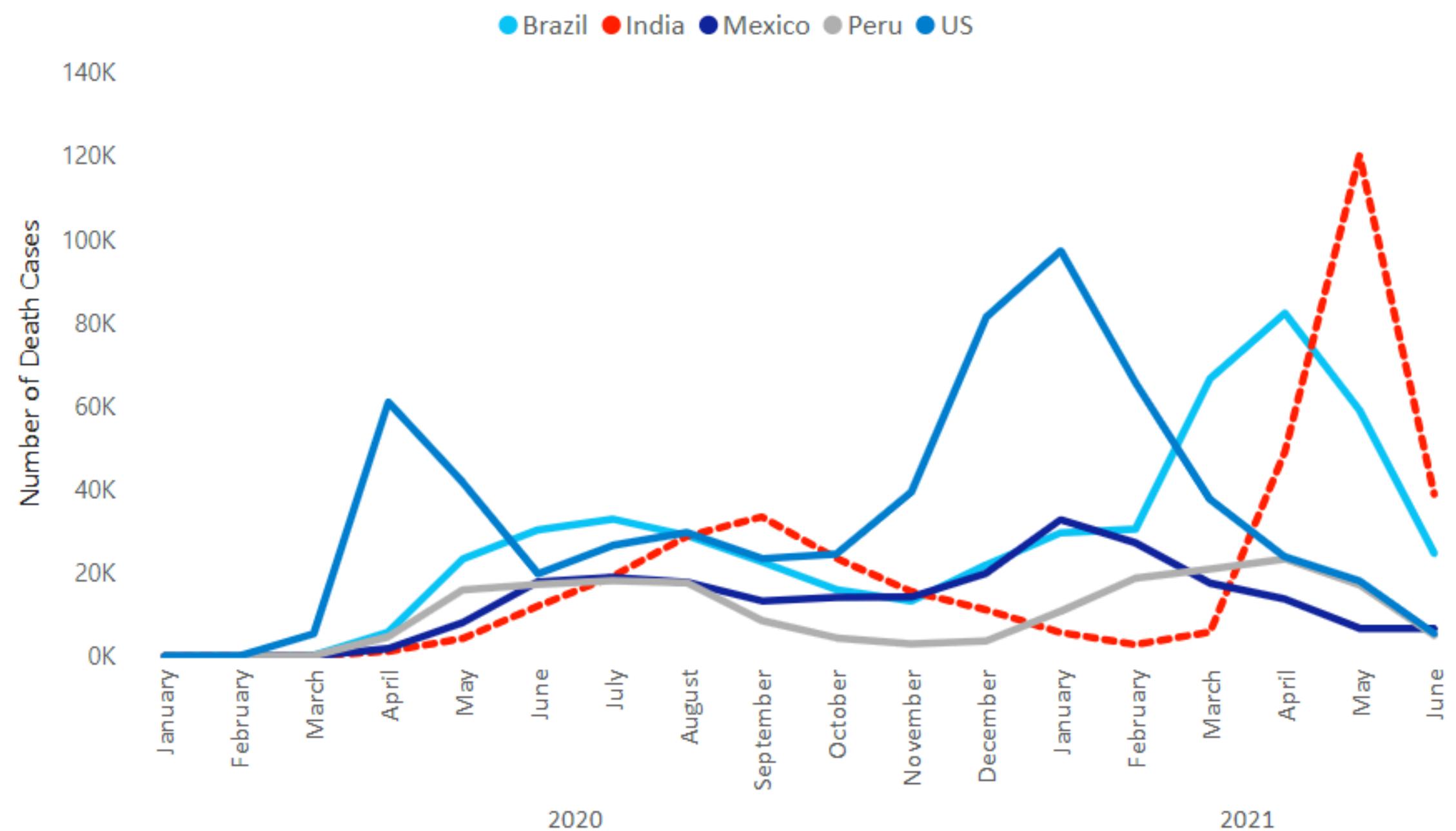
# Global Mortality Impact

The highest death tolls were recorded in the United States, Brazil, India, Mexico, and Peru.

The fewest deaths occurred in the Marshall Islands, Samoa, Kiribati, Dominica, and Bhutan.

The death-to-confirmed case ratio was approximately 2.16%, indicating a relatively small impact on the global population.

**Death Cases in Top 5 Countries from January 2020 to June 2021**



# 03. Recovery Insights



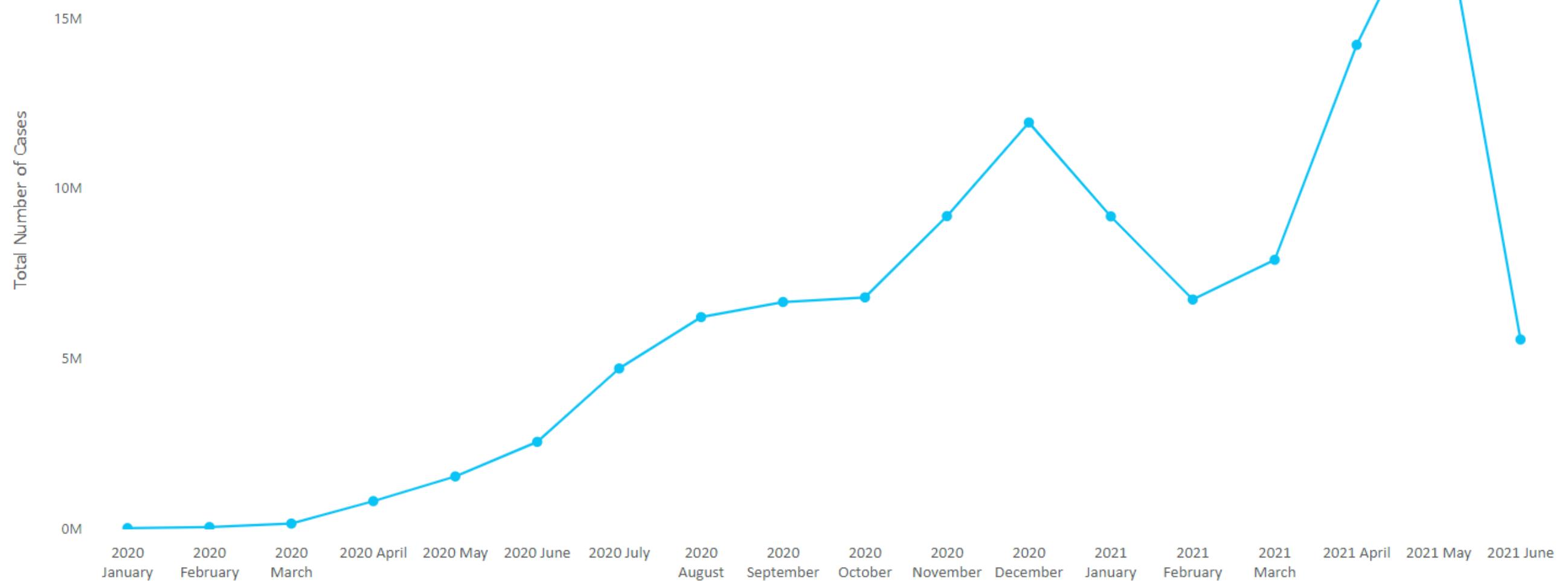
Approximately  
**113 million** people  
recovered during this  
period due to effective  
mitigating efforts and  
timely interventions.

# Recovery Patterns

The trend in recovered cases mirrored that of confirmed cases, indicating that as more individuals were diagnosed, a corresponding number of patients recovered.

**Monthly Recovered Cases from Jan 2020 to June 2021**

Line chart showing the total number of recovered cases per month.

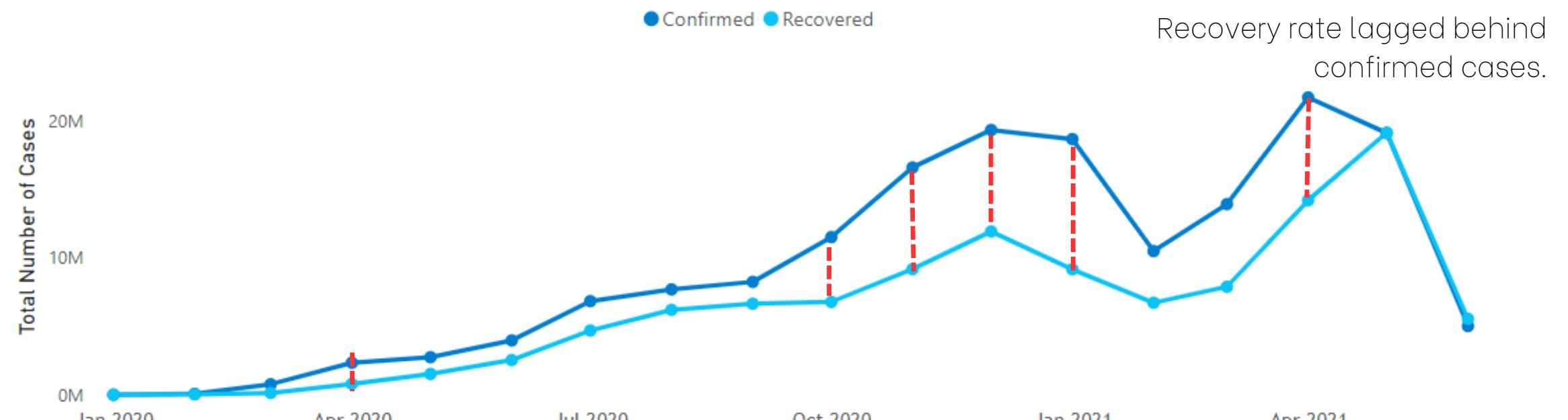


# Effectiveness of Recovery Strategies

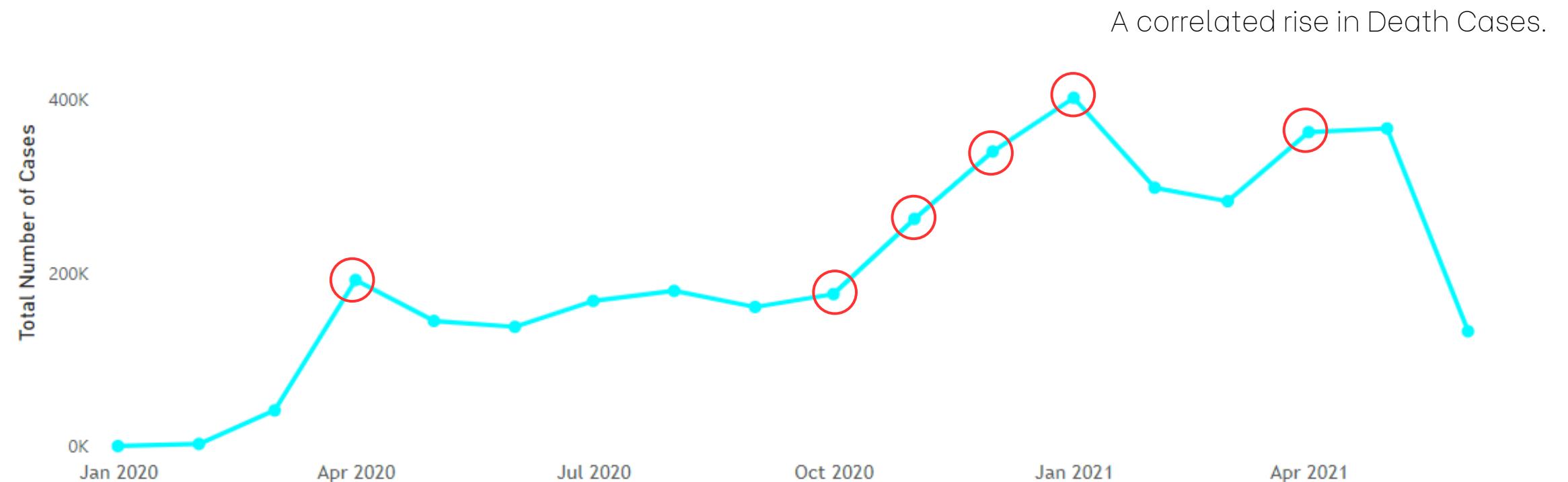
An analysis of the charts suggests that effective recovery measures led to a decrease in death rates, as evidenced from January to March, May to September in 2020, and February to March in 2021.

However, deaths increased when these measures could not keep pace with the rise in cases, especially in April 2020 and from October 2020 to January 2021.

Monthly Confirmed and Recovered Cases from Jan 2020 to June 2021



Monthly Death Cases from Jan 2020 to June 2021



# Recovered Cases Summary Statistics

May saw the most recovered cases, December had on average higher recovered cases and there was a lot of variation (in terms of variance and standard deviation) in the number of cases from day to day in most months. The situation was quite volatile, with some days having very few recovered cases and others having very high recovered cases.

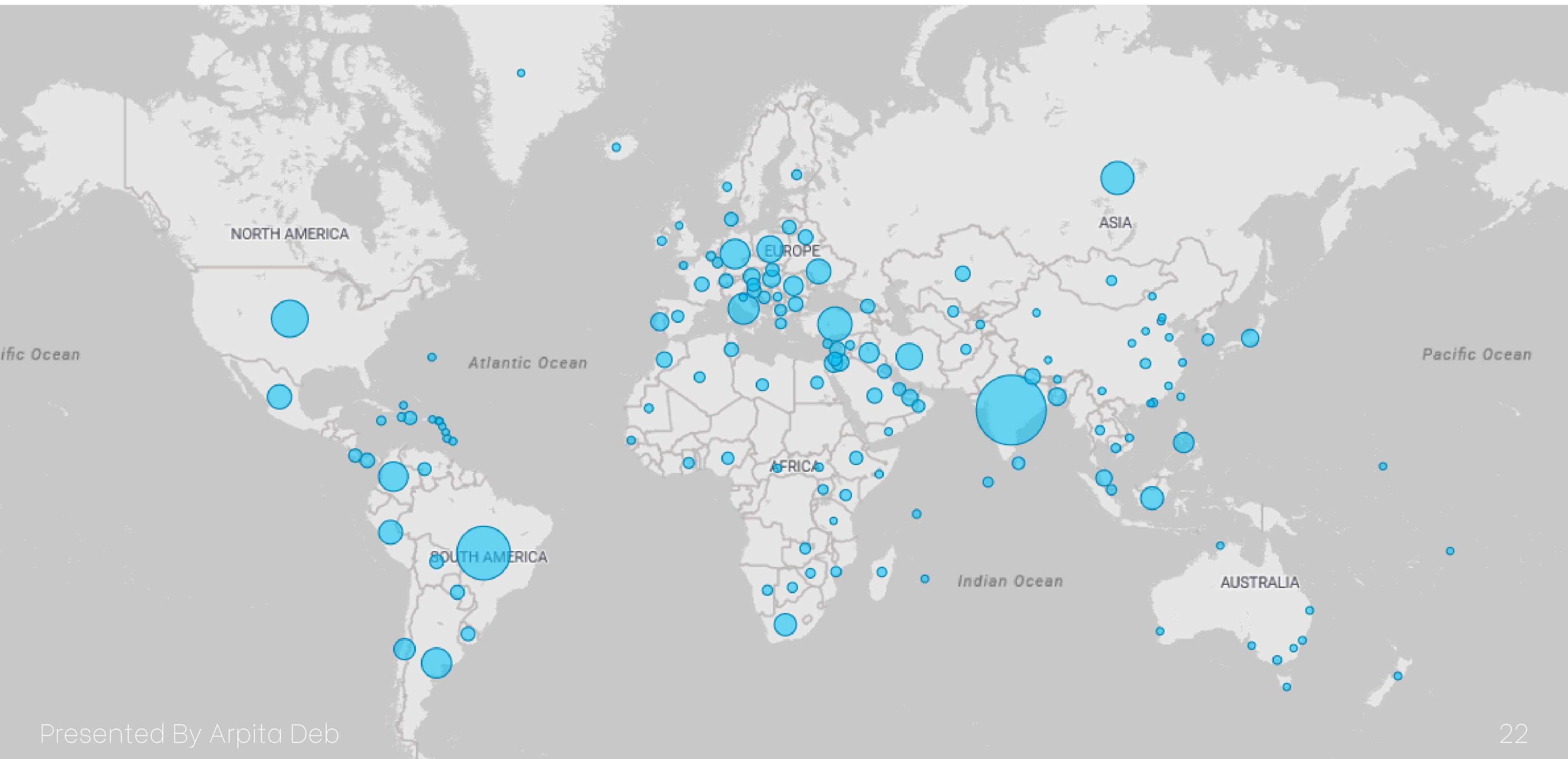
```
WITH OrderedCases AS (
    SELECT [Recovered],
        ROW_NUMBER() OVER (ORDER BY [Recovered]) AS RowNum,
        COUNT(*) OVER () AS TotalCount
    FROM [dbo].[CoronaVirusData]
)
SELECT
    MONTH([Date]) AS Month_Num,
    DATENAME(month, [Date]) AS Month,
    SUM([Recovered]) AS Total_Recovered_Cases,
    ROUND(AVG([Recovered]),0) AS Avg_Recovered_Cases,
    MIN([Recovered]) AS Min_Recovered_Cases,
    (SELECT [Recovered] FROM OrderedCases WHERE RowNum = (TotalCount + 1) / 2) AS Median_Recovered_Cases,
    MAX([Recovered]) AS Max_Recovered_Cases,
    ROUND(VAR([Recovered]),2) AS Variance,
    ROUND(STDEV([Recovered]),2) AS Standard_Deviation
FROM
    [dbo].[CoronaVirusData]
GROUP BY
    MONTH([Date]),
    DATENAME(month, [Date])
ORDER BY
    MONTH([Date]),
    DATENAME(month, [Date]);
```

SQL Code for creating the Summary Stats table

Month_Num	Month	Total_Recovered_Cases	Avg_Recovered_Cases	Min_Recovered_Cases	Median_Recovered_Cases	Max_Recovered_Cases	Variance	Standard_Deviation
1	January	9164490	1451	0	2	87090	24495691.2	4949.31
2	February	6751190	769	0	2	98389	12608693.19	3550.87
3	March	8021083	840	0	2	102138	18130329.88	4257.97
4	April	14998494	1623	0	2	299988	114714213.66	10710.47
5	May	20651389	2163	0	2	422436	382019449.6	19545.32
6	June	8079855	1220	0	2	231456	76059821.13	8721.23
7	July	4693120	983	0	2	140050	24849082.94	4984.89
8	August	6202833	1299	0	2	95881	40178838.38	6338.68
9	September	6647749	1439	0	2	101468	57035911.88	7552.21
10	October	6782150	1421	0	2	388340	73747150.17	8587.62
11	November	9172292	1985	0	2	139292	50738601.25	7123.1
12	December	11924903	2498	0	2	1123456	326763170.52	18076.59

# Geographical Distribution

Visualization of recovered cases is represented by turquoise bubbles, with larger sizes indicating higher case counts.



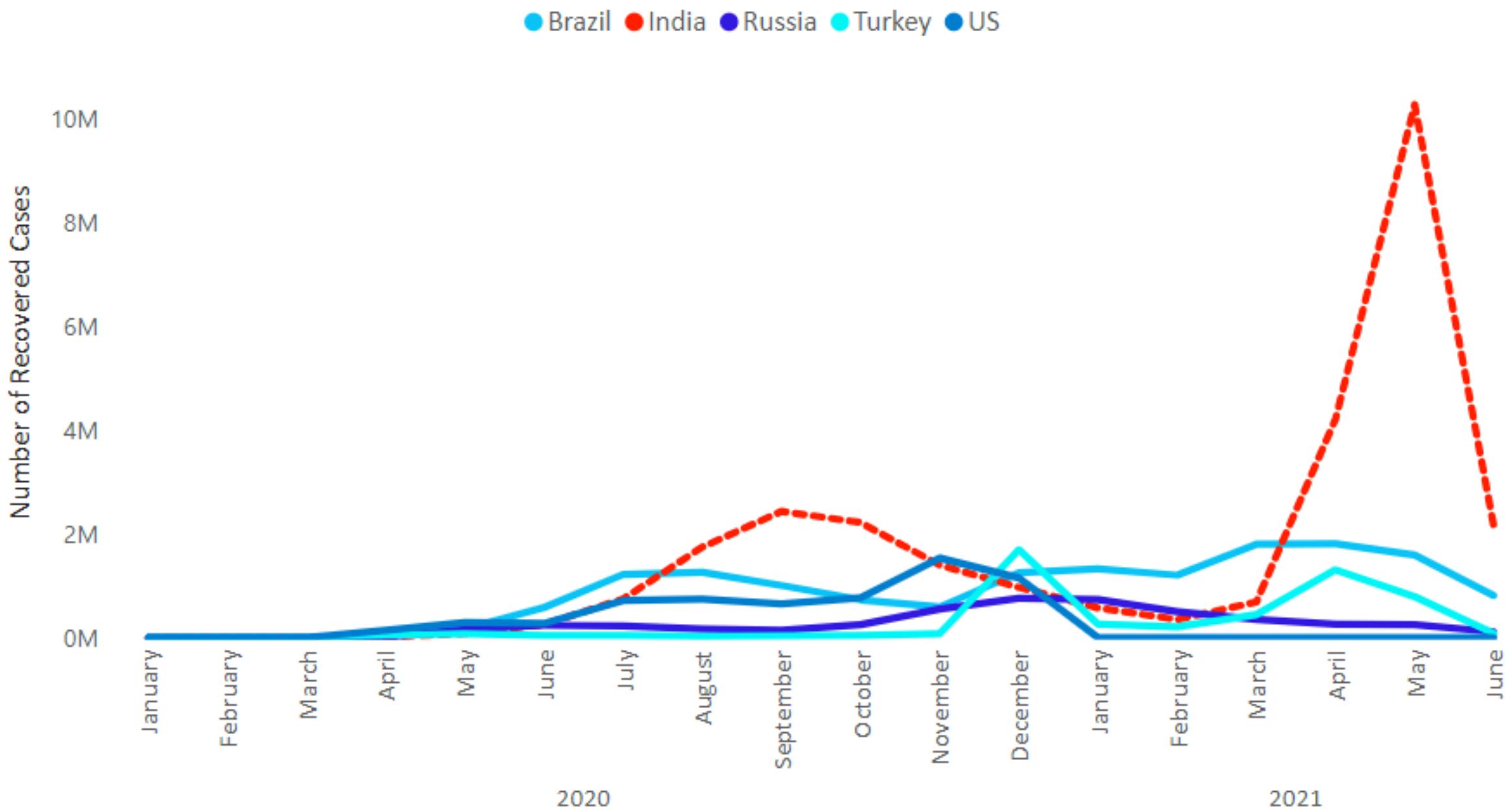
# Recovery Success Stories

India demonstrated a remarkable recovery rate, with 28.1 million citizens recovering over 18 months.

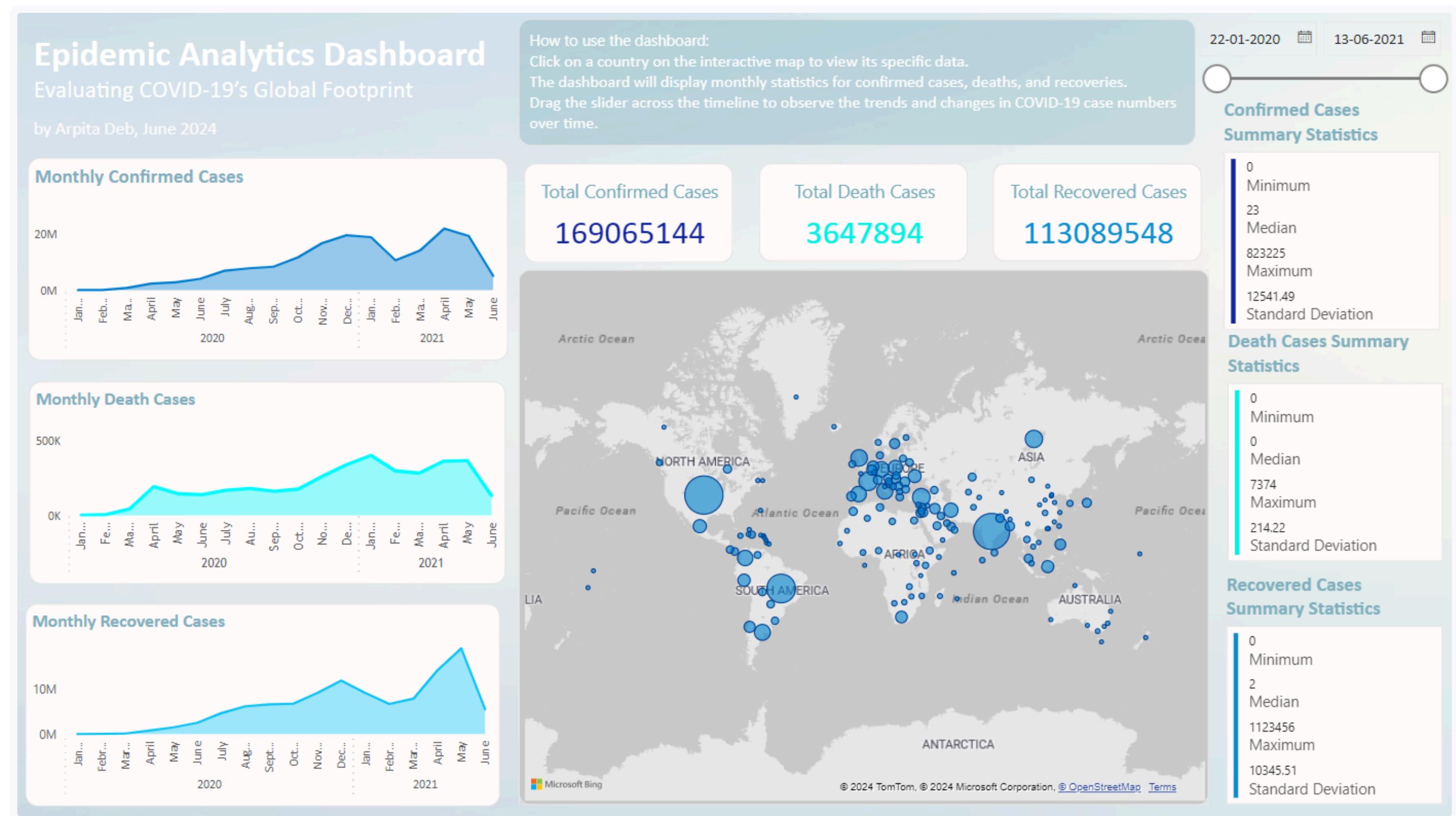
Brazil, the United States, Turkey, and Russia also reported high numbers of recoveries.

The recovery-to-confirmed ratio was approximately 67%, indicating the majority of patients overcame the virus.

**Recovered Cases in Top 5 Countries from January 2020 to June 2021**



# Dashboard



# Conclusion

## Summary

From January 2020 to June 2021, the world saw **169 million** COVID-19 cases, **3.65 million** deaths, and **113 million** recoveries. The global death and recovery rates were **2.16%** and **67%**, respectively.

## Trend Analysis

Case averages rose notably in **January, April, and August-December**, while February and June saw declines. Effective recovery strategies led to reduced mortality.

## Country-Specific Outcomes

The **United States, India, Brazil, France, and Turkey** reported the highest case counts, with the U.S., Brazil, India, Mexico, and Peru experiencing the most deaths. Conversely, **Kiribati, Samoa, Marshall Islands, Dominica, and Tanzania** had the lowest case numbers, with the fewest deaths in the Marshall Islands, Samoa, Kiribati, Dominica, and Bhutan.

## Recovery Highlights

**India's** recovery efforts were notably successful, with **28.1 million recoveries**. Brazil, the U.S., Turkey, and Russia also achieved significant recovery figures.

# Limitations

## **Timeframe:**

It covers COVID-19 cases from January 2020 to June 2021, missing nearly two years of data up to the WHO's declared end of the emergency in May 2023.

## **Data Depth:**

Limited to counts of confirmed cases, deaths, and recoveries, it lacks detailed statistics like hospitalization and testing rates, or case severity.

## **Demographics:**

Absence of patient demographics such as age, gender, and health conditions prevent detailed vulnerability analysis.

## **Mitigation Measures:**

No data on country-specific responses like lockdowns and vaccination campaigns, which impedes analysis of strategy effectiveness.

# Future Scope

01.

## Data Expansion

Extend the dataset to include cases up to the end of the WHO's emergency declaration, providing a complete picture of the pandemic's impact.

02.

## Enhanced Metrics

Incorporate additional data points such as hospitalization rates, testing frequency, and case severity to enable more comprehensive statistical analysis.

03.

## Demographic Details

Collect demographic information and details on mitigation measures to facilitate segmentation analysis and evaluate response strategies' effectiveness.

# Appendix

- Project Details Document
- GitHub Project Documentation
- Data Source
- SQL Code File
- Power BI Dashboard

# Thank You

For your Attention



Presented by  
Arpita Deb

Dated  
May, 2024

LinkedIn  
<https://www.linkedin.com/in/arpitadeb1811/>

GitHub  
<https://github.com/Arpita-deb/Arpita-deb>