https://github.com/045019/DEV-1-Project_045019.git

PROJECT REPORT

COVID-19 Time Series Analysis

Prepared By :

Himanshi Saxena
045019

PGDM, BDA-04

Section - H

Objective

The main goal of this project report is to examine and present the worldwide data on COVID-19, including vital measurements like global confirmed cases, deaths and recovered cases from January 2020 to March 2023.

By rigorously collecting and analysing time-series data and presenting it visually, this report aims to offer a current understanding of how the COVID-19 pandemic has affected the world.

It intends to assist policymakers, healthcare professionals and the public, in general, in decision-making and gaining valuable insights.

General Description of Data

Global Confirmed Cases

```
dataframes_confirmed
```

	b'Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3
0	NaN	Afghanistan	33.939110	67.709953	0.0	0.0	0.0	0.0	0.0	0.0	...	209322.0	209340.0	209358.0	209
1	NaN	Albania	41.153300	20.168300	0.0	0.0	0.0	0.0	0.0	0.0	...	334391.0	334408.0	334408.0	334
2	NaN	Algeria	28.033900	1.659600	0.0	0.0	0.0	0.0	0.0	0.0	...	271441.0	271448.0	271463.0	271
3	NaN	Andorra	42.506300	1.521800	0.0	0.0	0.0	0.0	0.0	0.0	...	47866.0	47875.0	47875.0	47
4	NaN	Angola	-11.202700	17.873900	0.0	0.0	0.0	0.0	0.0	0.0	...	105255.0	105277.0	105277.0	105
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
285	NaN	Winter Olympics 2022	39.904200	116.407400	0.0	0.0	0.0	0.0	0.0	0.0	...	535.0	535.0	535.0	
286	NaN	Yemen	15.552727	48.516388	0.0	0.0	0.0	0.0	0.0	0.0	...	11945.0	11945.0	11945.0	11
287	NaN	Zambia	-13.133897	27.849332	0.0	0.0	0.0	0.0	0.0	0.0	...	343012.0	343012.0	343079.0	343
288	NaN	Zimbabwe	-19.015438	29.154857	0.0	0.0	0.0	0.0	0.0	0.0	...	263921.0	264127.0	264127.0	264
289	'	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	

290 rows × 1147 columns

This data frame presents a comprehensive overview of confirmed COVID-19 cases worldwide between January 2020 and March 2023, with raw data collected from the Human Data Exchange website.

It includes detailed information on confirmed cases date-wise in different countries worldwide.

Description of Columns:

- Province/State: The name of the province or state.
- Country/region: The country with confirmed COVID-19 cases.
- Geographical Location: The Latitude and Longitude values suggest the geographical distribution of these cases.
- Individual Dates: The total number of confirmed cases on particular dates.

Global Recoveries

dataframes_recovered

	b'Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3/3/23
0	NaN	Afghanistan	33.939110	67.709953	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
1	NaN	Albania	41.153300	20.168300	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
2	NaN	Algeria	28.033900	1.659600	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
3	NaN	Andorra	42.506300	1.521800	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
4	NaN	Angola	-11.202700	17.873900	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
270	NaN	Winter Olympics 2022	39.904200	116.407400	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
271	NaN	Yemen	15.552727	48.516388	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
272	NaN	Zambia	-13.133897	27.849332	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
273	NaN	Zimbabwe	-19.015438	29.154857	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
274	'	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

275 rows × 1147 columns

This data frame presents a comprehensive overview of recovered COVID-19 cases worldwide between January 2020 and March 2023, with raw data collected from the Human Data Exchange website.

It includes detailed information on recovered COVID-19 cases, date-wise, in different countries worldwide.

Description of Columns:

- Province/State: The name of the province or state.
- Country/region: The country with the recovered COVID-19 cases.
- Geographical Location: The Latitude and Longitude values suggest the geographical distribution of these cases.
- Individual Dates: The total number of recovered cases on a particular date.

Global Deaths

dataframes_deaths

	b'Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/28/23	3/1/23	3/2/23	3/3/23
0	NaN	Afghanistan	33.939110	67.709953	0.0	0.0	0.0	0.0	0.0	0.0	...	7896.0	7896.0	7896.0	7896.0
1	NaN	Albania	41.153300	20.168300	0.0	0.0	0.0	0.0	0.0	0.0	...	3598.0	3598.0	3598.0	3598.0
2	NaN	Algeria	28.033900	1.659600	0.0	0.0	0.0	0.0	0.0	0.0	...	6881.0	6881.0	6881.0	6881.0
3	NaN	Andorra	42.506300	1.521800	0.0	0.0	0.0	0.0	0.0	0.0	...	165.0	165.0	165.0	165.0
4	NaN	Angola	-11.202700	17.873900	0.0	0.0	0.0	0.0	0.0	0.0	...	1933.0	1933.0	1933.0	1933.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
285	NaN	Winter Olympics 2022	39.904200	116.407400	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
286	NaN	Yemen	15.552727	48.516388	0.0	0.0	0.0	0.0	0.0	0.0	...	2159.0	2159.0	2159.0	2159.0
287	NaN	Zambia	-13.133897	27.849332	0.0	0.0	0.0	0.0	0.0	0.0	...	4057.0	4057.0	4057.0	4057.0
288	NaN	Zimbabwe	-19.015438	29.154857	0.0	0.0	0.0	0.0	0.0	0.0	...	5663.0	5668.0	5668.0	5668.0
289	'	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

290 rows × 1147 columns

This data frame presents a comprehensive overview of COVID-19 deaths worldwide between January 2020 and March 2023, with raw data collected from the Human Data Exchange website. It includes detailed information on fatalities date-wise in different countries worldwide.

Description of columns:

- Province/State: The name of the province or state.
- Country/region: The country with COVID-19 deaths.
- Geographical Location: The Latitude and Longitude values suggest the geographical distribution of these cases.
- Individual Dates: The total number of deaths on a particular date.

NOTE:

All of the above data frames serve as a valuable resource for global analysis of the trends and transmissions during the different waves of COVID-19.

It is also instrumental in understanding the efficacy of vaccination drives and how they have helped boost recoveries.

Statistical Summary of all the Dataframes

```
Statistical Summary for Confirmed Cases:

             Lat        Long     1/22/20     1/23/20     1/24/20     1/25/20  \
count  89.000000   89.000000   89.000000   89.000000   89.000000   89.000000
mean   22.442517   24.950424    6.157303    7.247191   10.370787   15.831461
std    28.980069   99.535149   47.070281   47.139368   58.537164   81.296348
min   -51.796300 -178.116500    0.000000    0.000000    0.000000    0.000000
25%    12.178400  -63.416800    0.000000    0.000000    0.000000    0.000000
50%    29.183200   45.166244    0.000000    0.000000    0.000000    0.000000
75%    43.666100  115.722100    1.000000    2.000000    3.000000    7.000000
max    71.706900  165.618042  444.000000  444.000000  549.000000  761.000000

            1/26/20      1/27/20      1/28/20      1/29/20  ...       2/27/23  \
count     89.000000    89.000000    89.000000    89.000000  ...  8.900000e+01
mean      23.393258    32.415730    62.000000    68.505618  ...  2.359379e+05
std      113.441476   152.717193   376.977724   378.574363  ...  6.564315e+05
min        0.000000     0.000000     0.000000     0.000000  ...  0.000000e+00
25%        0.000000     0.000000     0.000000     0.000000  ...  3.904000e+03
50%        0.000000     0.000000     0.000000     0.000000  ...  1.225700e+04
75%        9.000000    14.000000    24.000000    27.000000  ...  8.000700e+04
max     1058.000000  1423.000000  3554.000000  3554.000000  ...  3.900969e+06
```

```
             2/28/23       3/1/23       3/2/23       3/3/23       3/4/23  \
count  8.900000e+01  8.900000e+01  8.900000e+01  8.900000e+01  8.900000e+01
mean   2.359467e+05  2.359590e+05  2.362197e+05  2.362336e+05  2.362383e+05
std    6.564460e+05  6.564597e+05  6.573092e+05  6.573213e+05  6.573291e+05
min    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
25%    3.904000e+03  3.904000e+03  3.904000e+03  3.904000e+03  3.904000e+03
50%    1.225700e+04  1.225700e+04  1.225700e+04  1.225700e+04  1.225700e+04
75%    8.000700e+04  8.000700e+04  8.000700e+04  8.000700e+04  8.000700e+04
max    3.900969e+06  3.900969e+06  3.908129e+06  3.908129e+06  3.908129e+06

             3/5/23       3/6/23       3/7/23       3/8/23
count  8.900000e+01  8.900000e+01  8.900000e+01  8.900000e+01
mean   2.362416e+05  2.362604e+05  2.362694e+05  2.362830e+05
std    6.573346e+05  6.573378e+05  6.573520e+05  6.573632e+05
min    0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
25%    3.904000e+03  3.904000e+03  3.904000e+03  3.904000e+03
50%    1.225700e+04  1.225700e+04  1.225700e+04  1.227100e+04
75%    8.000700e+04  8.000700e+04  8.000700e+04  8.001700e+04
max    3.908129e+06  3.908129e+06  3.908129e+06  3.908129e+06

[8 rows x 1144 columns]
```

Statistical Summary for Deaths:

	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	\
count	89.000000	89.000000	89.000000	89.000000	89.000000	89.000000	
mean	22.442517	24.950424	0.191011	0.202247	0.292135	0.471910	
std	28.980069	99.535149	1.801996	1.803908	2.545952	4.240202	
min	-51.796300	-178.116500	0.000000	0.000000	0.000000	0.000000	
25%	12.178400	-63.416800	0.000000	0.000000	0.000000	0.000000	
50%	29.183200	45.166244	0.000000	0.000000	0.000000	0.000000	
75%	43.666100	115.722100	0.000000	0.000000	0.000000	0.000000	
max	71.706900	165.618042	17.000000	17.000000	24.000000	40.000000	

	1/26/20	1/27/20	1/28/20	1/29/20	...	2/27/23	\
count	89.000000	89.000000	89.000000	89.000000	...	89.000000	
mean	0.629213	0.921348	1.471910	1.494382	...	1076.629213	
std	5.511109	8.052702	13.245145	13.244316	...	3135.154032	
min	0.000000	0.000000	0.000000	0.000000	...	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	...	3.000000	
50%	0.000000	0.000000	0.000000	0.000000	...	23.000000	
75%	0.000000	0.000000	0.000000	0.000000	...	311.000000	
max	52.000000	76.000000	125.000000	125.000000	...	18097.000000	

	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	\
count	89.000000	89.000000	89.000000	89.000000	89.000000	
mean	1076.786517	1077.382022	1079.247191	1079.404494	1079.494382	
std	3135.888295	3137.019936	3141.641317	3142.408644	3142.884937	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	3.000000	3.000000	3.000000	3.000000	3.000000	
50%	23.000000	23.000000	23.000000	23.000000	23.000000	
75%	311.000000	311.000000	311.000000	311.000000	311.000000	
max	18101.000000	18105.000000	18114.000000	18124.000000	18131.000000	

	3/5/23	3/6/23	3/7/23	3/8/23
count	89.000000	89.000000	89.000000	89.000000
mean	1079.505618	1079.932584	1080.078652	1080.404494
std	3142.929717	3143.043573	3143.779993	3144.662743
min	0.000000	0.000000	0.000000	0.000000
25%	3.000000	3.000000	3.000000	3.000000
50%	23.000000	23.000000	23.000000	23.000000
75%	311.000000	314.000000	314.000000	314.000000
max	18131.000000	18132.000000	18144.000000	18153.000000

[8 rows x 1144 columns]

Statistical Summary for Recovered Cases:

	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	\
count	74.000000	74.000000	74.000000	74.000000	74.000000	74.000000	
mean	17.335965	46.225257	0.378378	0.405405	0.486486	0.527027	
std	27.877175	94.240150	3.254934	3.260048	3.608374	3.727804	
min	-51.796300	-178.116500	0.000000	0.000000	0.000000	0.000000	
25%	5.992825	-58.721675	0.000000	0.000000	0.000000	0.000000	
50%	24.401900	105.225800	0.000000	0.000000	0.000000	0.000000	
75%	36.043550	117.023350	0.000000	0.000000	0.000000	0.000000	
max	71.706900	165.618042	28.000000	28.000000	31.000000	32.000000	

	1/26/20	1/27/20	1/28/20	1/29/20	...	2/27/23	2/28/23	\
count	74.000000	74.000000	74.000000	74.000000	...	74.0	74.0	
mean	0.662162	0.783784	1.364865	1.621622	...	0.0	0.0	
std	4.885755	5.253232	9.314261	10.236522	...	0.0	0.0	
min	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	
25%	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	
50%	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	
75%	0.000000	0.000000	0.000000	0.000000	...	0.0	0.0	
max	42.000000	45.000000	80.000000	88.000000	...	0.0	0.0	

	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23	3/6/23	3/7/23	3/8/23
count	74.0	74.0	74.0	74.0	74.0	74.0	74.0	74.0
mean	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
75%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
max	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[8 rows x 1144 columns]

Interpretation of individual rows of the Output

- **count:** shows the non-null values in each numerical column. After data cleaning, it provides a count of available data points for each column.

- **mean:** shows the average value for each column. For example, the 'confirmed cases summary; date (1/24/20)' indicates the average of COVID-19 cases worldwide on 24 January 2020.

- **std:** The standard deviation measures the variability or dispersion of the data. It shows how much the values in each column deviate from the mean. A higher standard deviation indicates more variation.

- **min:** displays the minimum value in each column. For example, the 'confirmed cases summary; date (1/24/20)' indicates the Lowest number of cases confirmed in different countries on 24 January 2020.

- **25%:** represents the 25th percentile value. It shows the value below which 25% of the data falls. It's often referred to as the first quartile (Q1).

- **50%:** corresponds to the median value i.e. 50th percentile in each column. The median is the middlemost value of sorted data.

- **75%:** shows the 75th percentile value, often called the third quartile (Q3). It represents the value below which 75% of the data falls.

- **max:** displays the maximum value in each column. The 'Confirmed cases summary; date (1/24/20)' indicates the Highest number of cases confirmed in different countries on 24 January 2020.

Analysis

The data shows the **statistical summary** of confirmed cases, deaths, and recovered cases for a particular location between January 22, 2020, and March 8, 2023. The key findings from the data are as follows:

- **For Latitude and Longitude**
  Count: 89 data points indicate the latitude and longitude for 89 locations.

- **For Confirmed Cases**

1. Count: 89 data points for each date indicate the number of confirmed COVID-19 cases for each of the 89 locations.

2. Mean: The average number of confirmed cases varies for each date, ranging from 6.16 on 1/22/20 to approximately 236,283 on 3/8/23.

3. Standard Deviation: The standard deviation for confirmed cases varies for each date, indicating the variability in case counts over time.

4. Minimum: The minimum number of confirmed cases for each date is 0, indicating some locations with no reported cases.

5. 25th Percentile (Q1): The 25th Percentile (lower quartile) represents the value below which 25% of the data falls. For most dates, the 25th Percentile is 0, suggesting that many locations reported no cases on those dates.

6. Median (50th Percentile): The median represents the middle value of the data. For most dates, the median is 0, indicating that half of the locations reported no cases.

7. 75th Percentile (Q3): The 75th Percentile (upper quartile) represents the value below which 75% of the data falls. The 75th Percentile is relatively low for most dates, suggesting that most locations reported a small number of cases.

8. Maximum: The maximum number of confirmed cases varies for each date and ranges from 444 on 1/22/20 to approximately 3,908,129 on 3/4/23. These values represent the highest number of cases reported daily among the 89 locations.

- **For Deaths**

1. Count: 89 data points for each date represent the number of deaths reported for each of the 89 locations.

2. Mean: The average number of deaths varies for each date, ranging from approximately 0.19 on 1/22/20 to approximately 1,080.40 on 3/8/23.

3. Standard Deviation: The standard deviation for deaths also varies for each date, indicating variability in the number of deaths over time.

4. Minimum: The minimum number of deaths for each date is 0, indicating some locations with no reported casualties.

5. 25th Percentile (Q1): The 25th Percentile (lower quartile) is 0 for most dates, suggesting that a significant portion of locations reported no deaths on those dates.

6. Median (50th Percentile): The median is 0 for most dates, indicating that half of the locations reported no deaths.

7. 75th Percentile (Q3): The 75th Percentile (upper quartile) is relatively low for most dates, suggesting that most locations reported a few deaths.

8. Maximum Deaths: The maximum number of deaths varies for each date and ranges from 17 on 1/22/20 to approximately 18,153 on 3/4/23. These values represent the highest fatalities reported daily among the 89 locations.

- **For Recoveries**

1. Count: There are 74 data points for each date, representing the total number of recovered cases reported for each of the 74 locations.

2. Mean: The mean (Average) number of recovered cases varies for each date, ranging from approximately 0.38 on 1/22/20 to 0 on the most recent dates.

3. Standard Deviation: The standard deviation for the recovered cases varies for each date and is generally relatively low, indicating limited variability in the number of recovered cases over time/

4. Minimum: The minimum number of recovered cases for each date is 0, indicating that some locations may not have reported any recoveries.

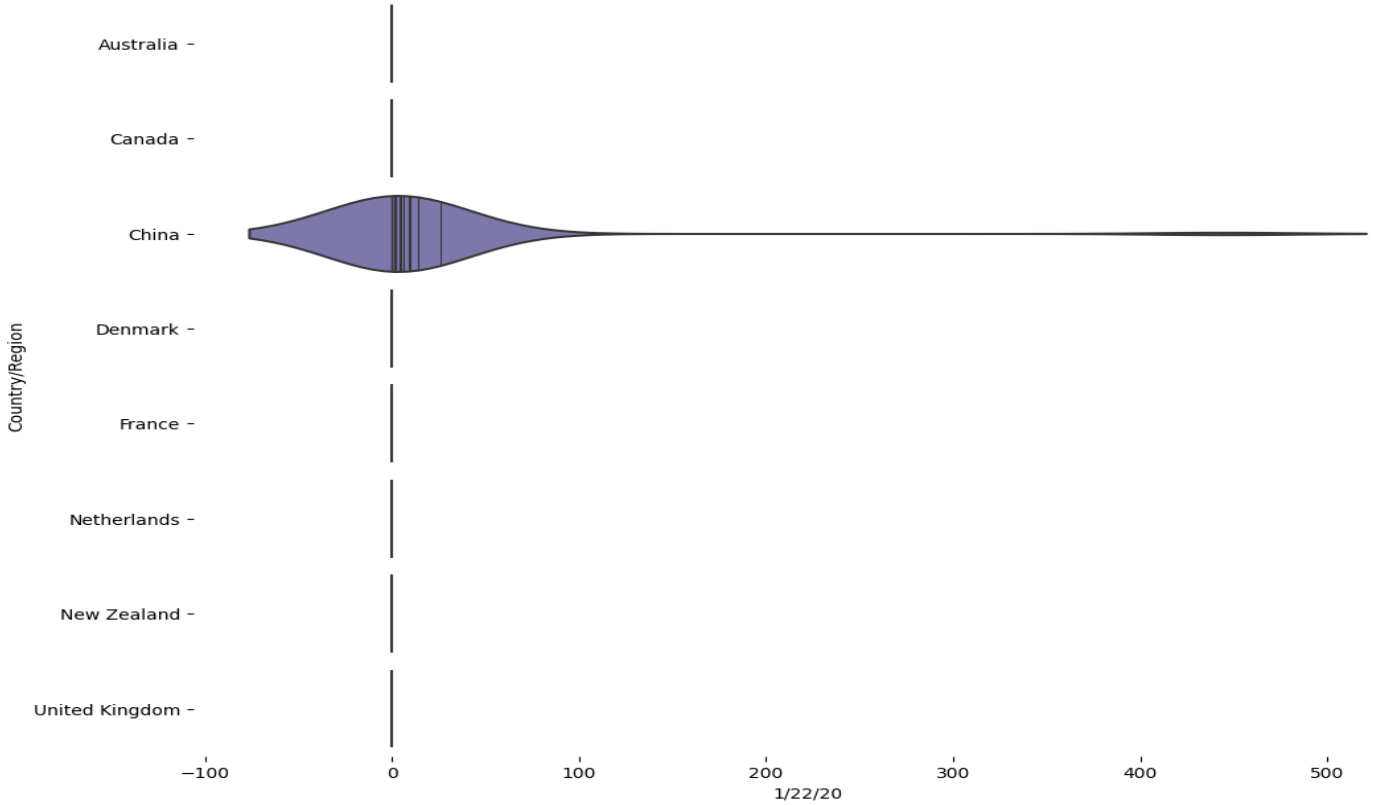
5. 25th Percentile (Q1): The 25th Percentile (lower quartile) is 0 for most dates, suggesting that many locations reported no recoveries on those dates.

6. Median (50th Percentile): The median is 0 for most dates, indicating that half of the locations reported no recoveries.

7. 75th Percentile (Q3): The 75th Percentile (upper quartile) is 0 for most dates, suggesting that most locations reported no or very few recoveries.

8. Maximum Recovered Cases: The maximum number of recovered cases varies for each date and is generally low, with the highest values reaching up to 88 on some earlier dates. However, for the most recent dates (from 2/27/23 onwards), the maximum recovered cases are reported as 0 for all locations as fewer cases exist.

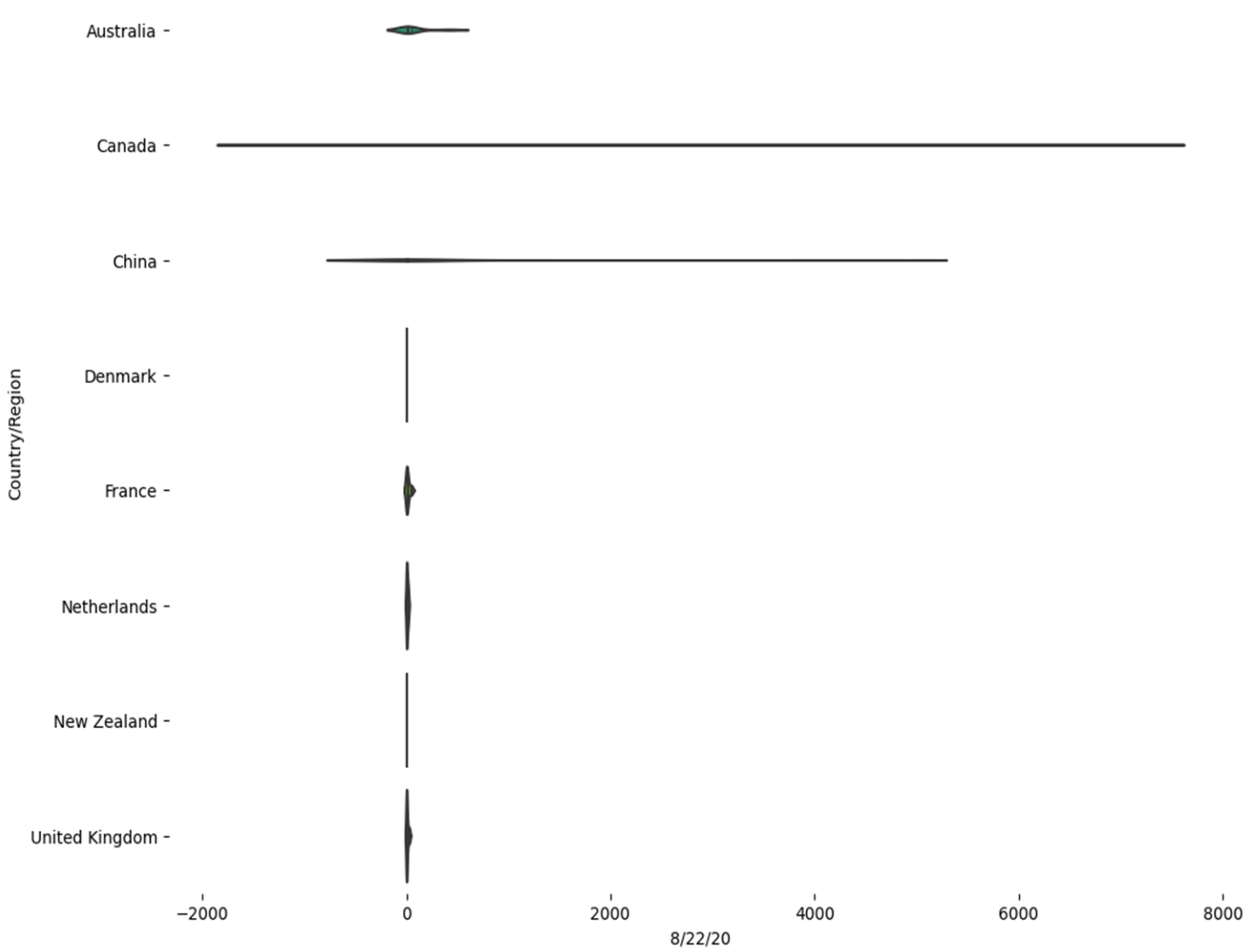
Violin Plot for Confirmed cases in the First wave of COVID-19



According to the violin plot, the number of cases in China increased sporadically in the days leading up to January 22. There was also a steady increase in confirmed cases in the weeks following January 22, 2020.

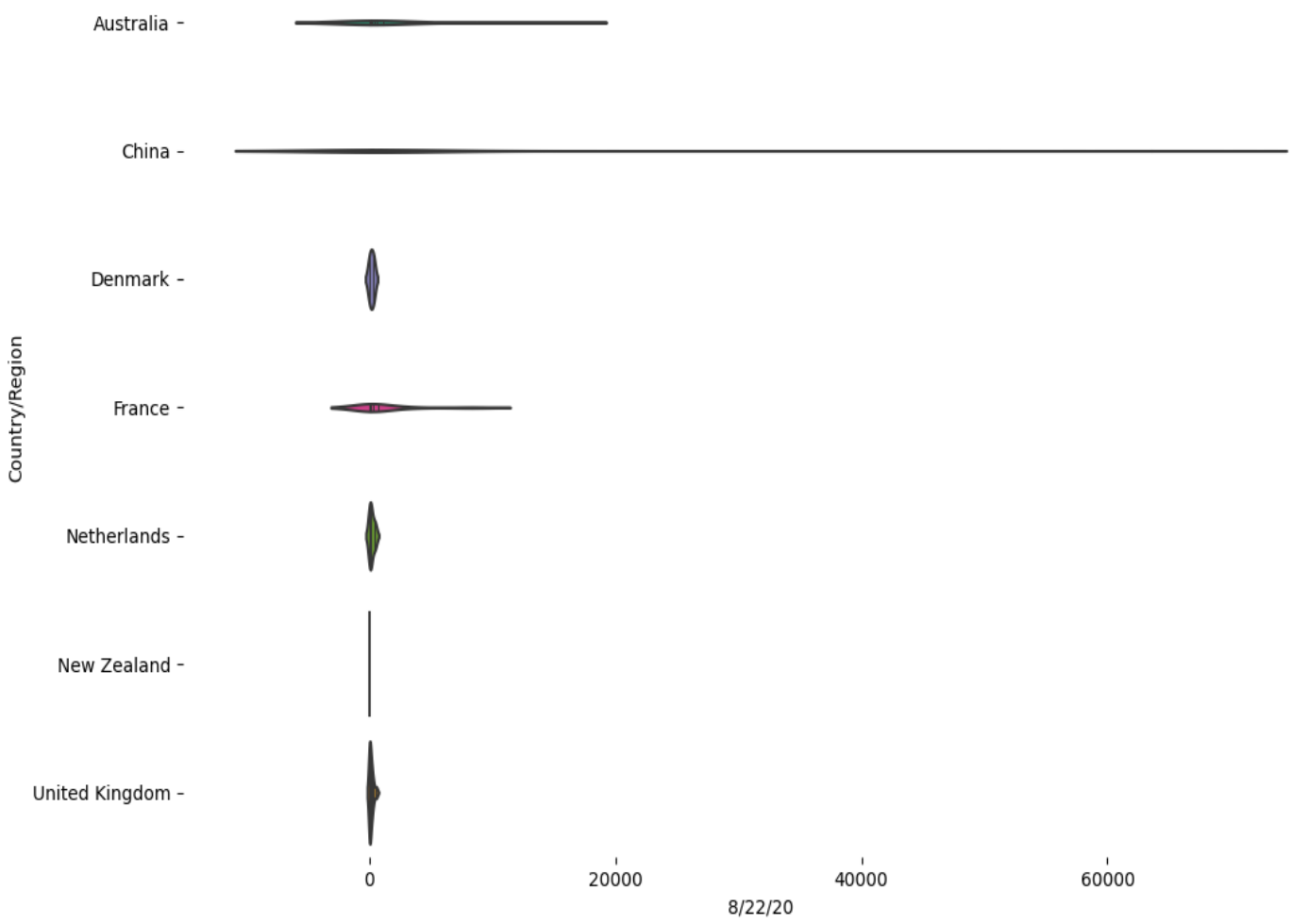
According to the World Health Organization (WHO), there were 295 confirmed cases of COVID-19 in China on January 22, 2020. All of the cases were reported in Wuhan, Hubei Province.

Violin Plot for Deaths in the First wave of COVID-19



As per the plot, the number of deaths has been increasing worldwide since the pandemic began in early 2020. Meanwhile, in many countries, here, China and Canada, we see the number of deaths increasing steadily.

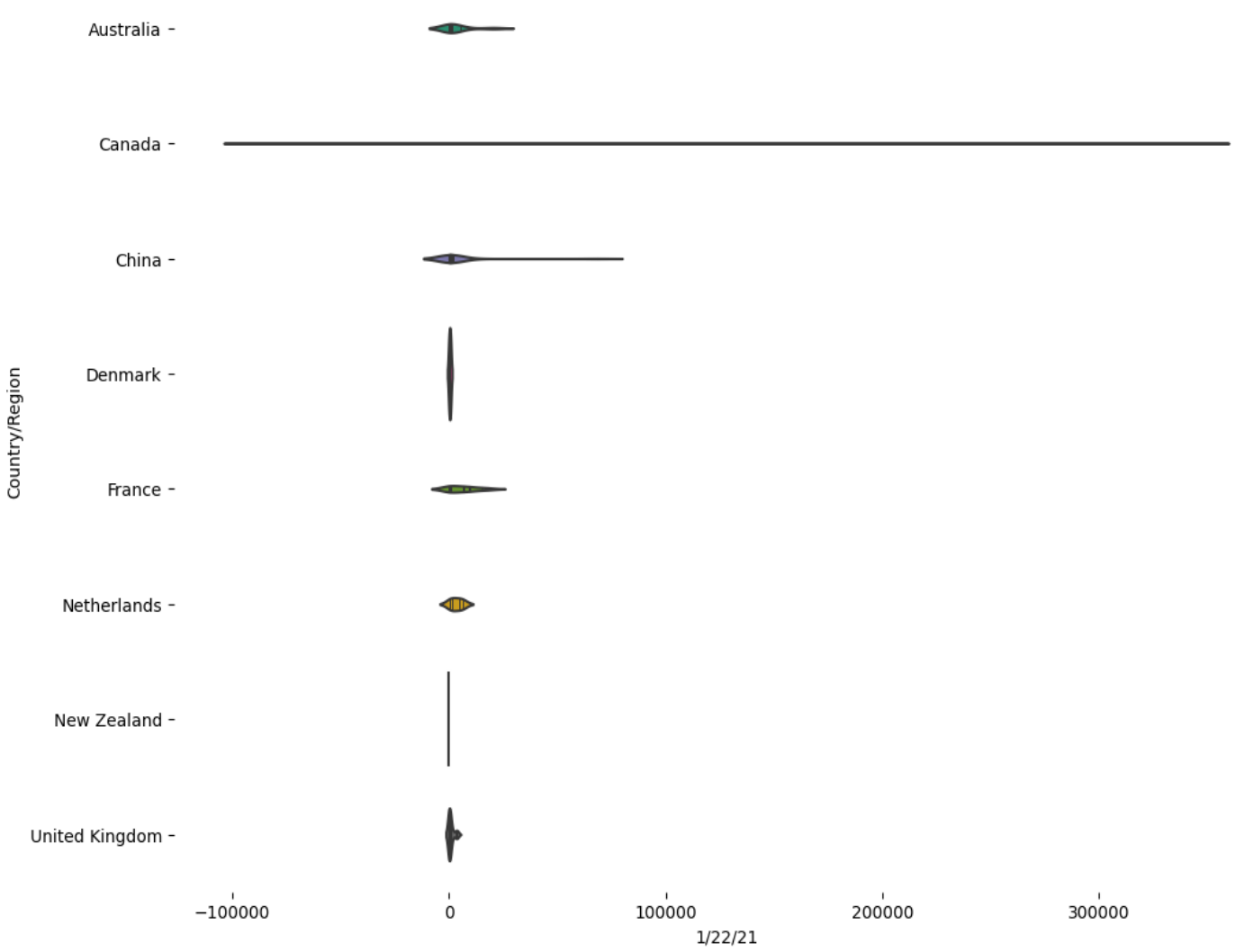
Violin Plot for Recoveries in the First wave of COVID-19



The plot shows a steady recovery rate in a few countries: China, Australia and France. Meanwhile, in other countries, there have been sporadic yet limited recoveries. The recovery rate varies as it depends on factors, including the age and health of the individual, the severity of the illness, and the availability of healthcare.
For instance, younger people and people with mild cases are more likely to recover quickly.

According to the WHO, as of August 31, 2020, there were 11,636,329 recovery cases out of 22,789,768 confirmed cases of COVID-19 worldwide. This means the recovery rate as of August 31, 2020 was about 51.3%.

Violin Plot for Confirmed cases in the Second wave of COVID-19

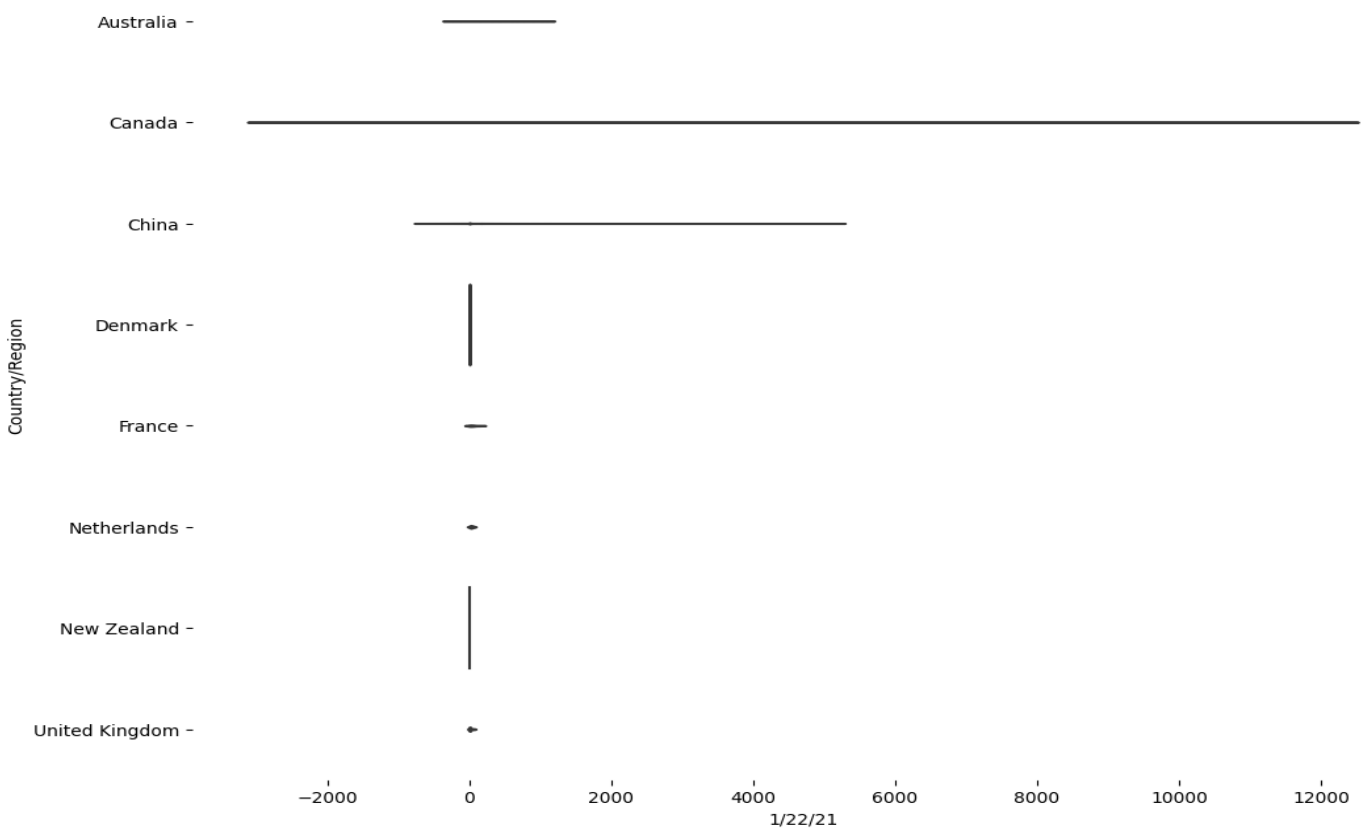


According to the plot, there was a significant increase in the confirmed cases in January 2021.

It is likely due to several factors, including the increased transmissibility of the virus, the relaxation of social distancing measures in some countries, and the winter holidays, which led to increased travel and mixing of people.

According to the WHO, there were 81,963,694 confirmed cases of COVID-19 worldwide in January 2021. The number of cases increased by an average of 4.4 million per week in January 2021.

Violin Plot for Deaths in the Second wave of COVID-19

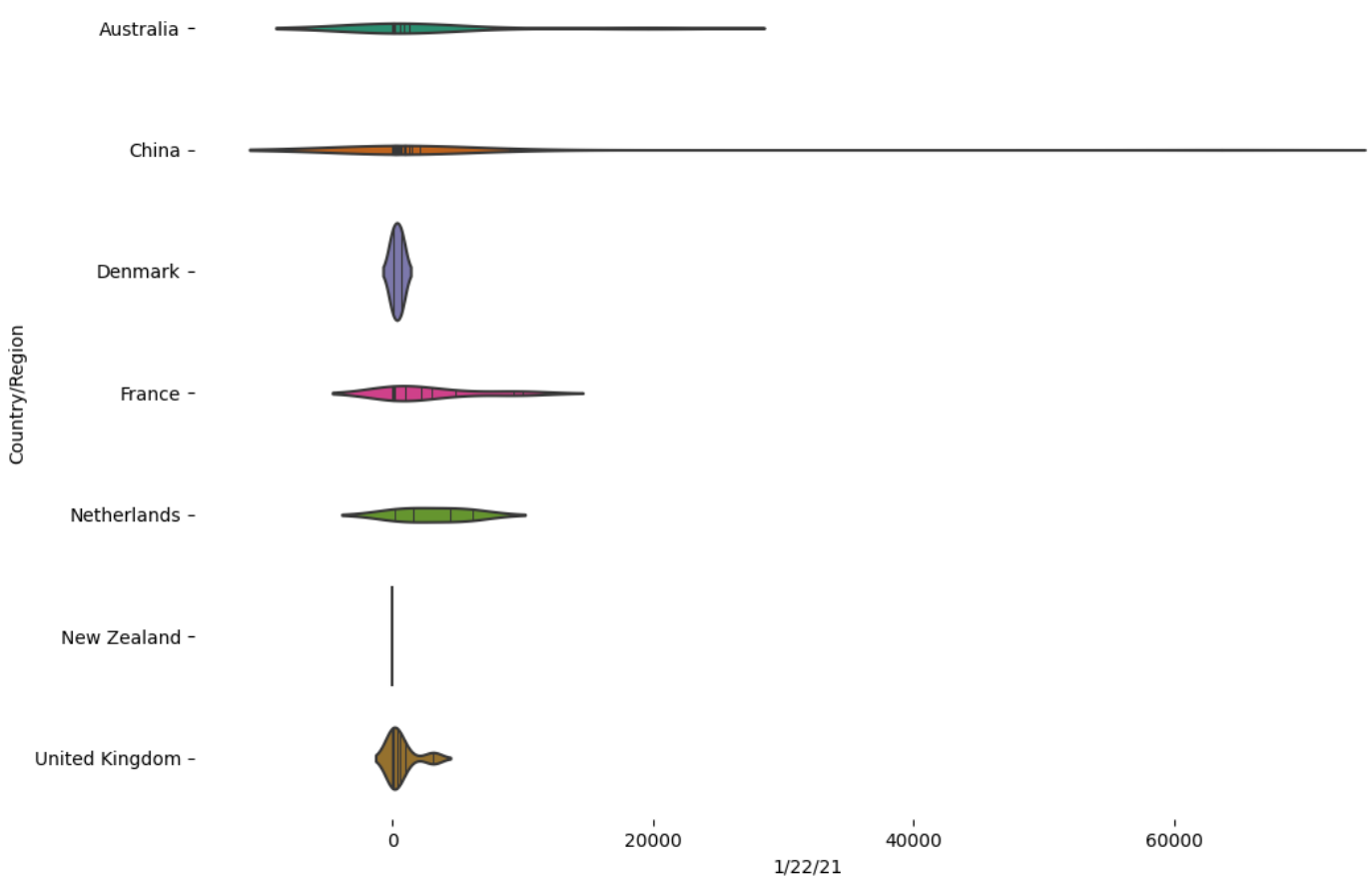


As per the plot, the deaths in January 2021 were significantly higher than in the previous month.

It is likely due to the same factors that led to the increase in the confirmed cases and the fact that the virus is more likely to be fatal in older people and people with underlying health conditions.

According to the WHO, there were 1,813,188 deaths from COVID-19 worldwide in January 2021. Fatalities increased steadily by an average of 150,000 per week in January 2021.

Violin Plot for Recoveries in the Second wave of COVID-19



According to the violin plot, there was a steady increase in the recovered cases worldwide. As seen in the case of many countries, here, China and Australia.

However, many countries, like Denmark and the United Kingdom, have a limited and concentrated amount of recoveries.

Other Statistical Insights

- **Average confirmed cases (Mean during the first wave of COVID):**
  The average confirmed cases in the first Covid wave (roughly January 2020-May 2020) in the dataset is approximately 34361 per day. The number of confirmed Covid cases increased rapidly, from 98 on January 22, 2020, to 111,790 on May 22, 2020. The peak number (Max) of confirmed cases was 128,123 on March 12, 2020.

- **Average confirmed cases (Mean during the second wave of COVID-19):**
  The average number of confirmed COVID-19 cases between December 2020 and August 2021 was 238,381 per day. The number of confirmed Covid cases increased rapidly, from 1,558,549 on December 1, 2020, to 243,896 on August 31, 2021. The peak number (Max) of confirmed cases was 264,775 on July 19, 2021.

- **Average confirmed cases (Mean during the third wave of COVID-19):**
  The average number of confirmed COVID-19 cases between November 2021 and May 2022 was 140,277 per day. The number of confirmed Covid cases increased rapidly, from 4,489,903 on November 1, 2021, to 147,337 on May 22, 2022. The peak number (Max) of confirmed cases was 275,141 on January 19, 2022.

- **Average recovery cases and death cases (Mean during the first wave of COVID):**
  The average number of death cases was around 430,882 in the first COVID wave (roughly Jan 2020-May 2020), while the average recoveries were 60,94,895.

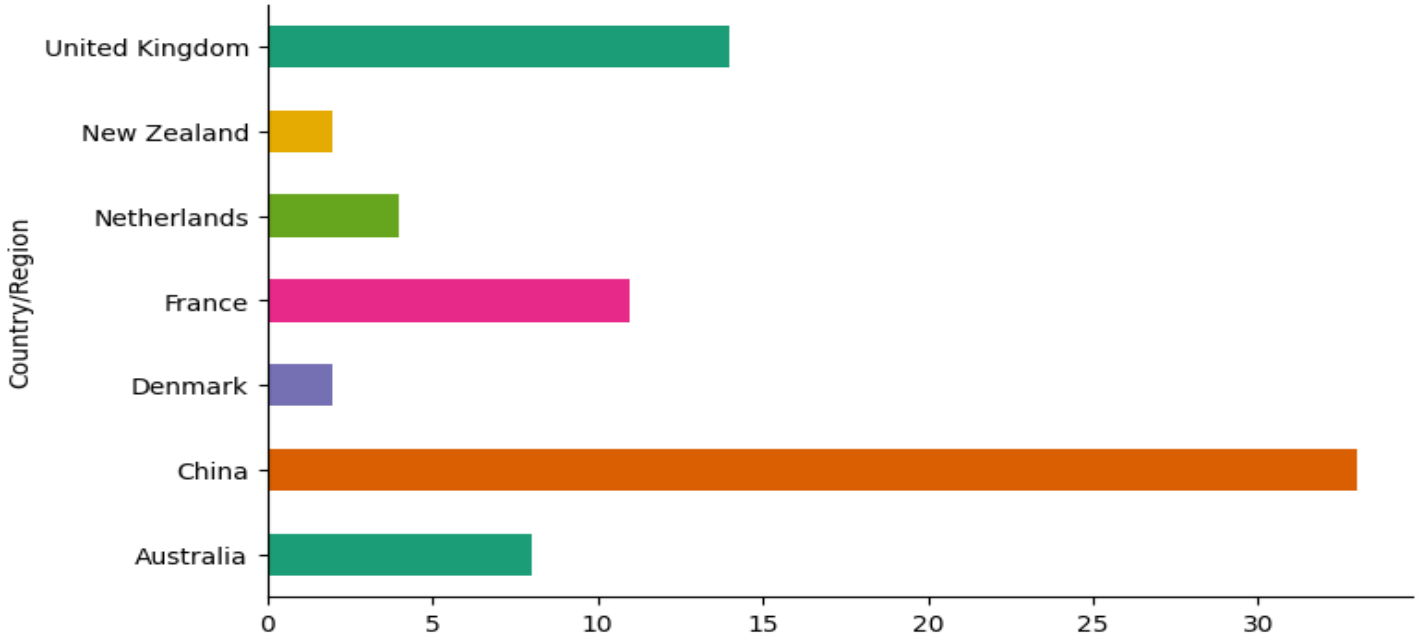
- **Average recovery cases and death cases (Mean during the second wave of COVID):**
  The average death cases was around 29,95,321 between December 2020 and August 2021, while the average recoveries were 13,46,98,418.

- **Average recovery cases and death cases (Mean during the third wave of COVID):**
  The average number of death cases was around 12,66,919 between November 2021 and May 2022, while the average recoveries were 28,47,08,680.

- **Median**:

  The median number of confirmed cases in all three waves is 13,76,93,739. The median is the middle value when all cases are sorted in ascending order. It's a robust measure of central tendency and is not influenced by extreme outliers. The median number of deaths during the three COVID waves was 29,95,321.

Bar Graph depicting Scaled/Normalised Recovered Cases

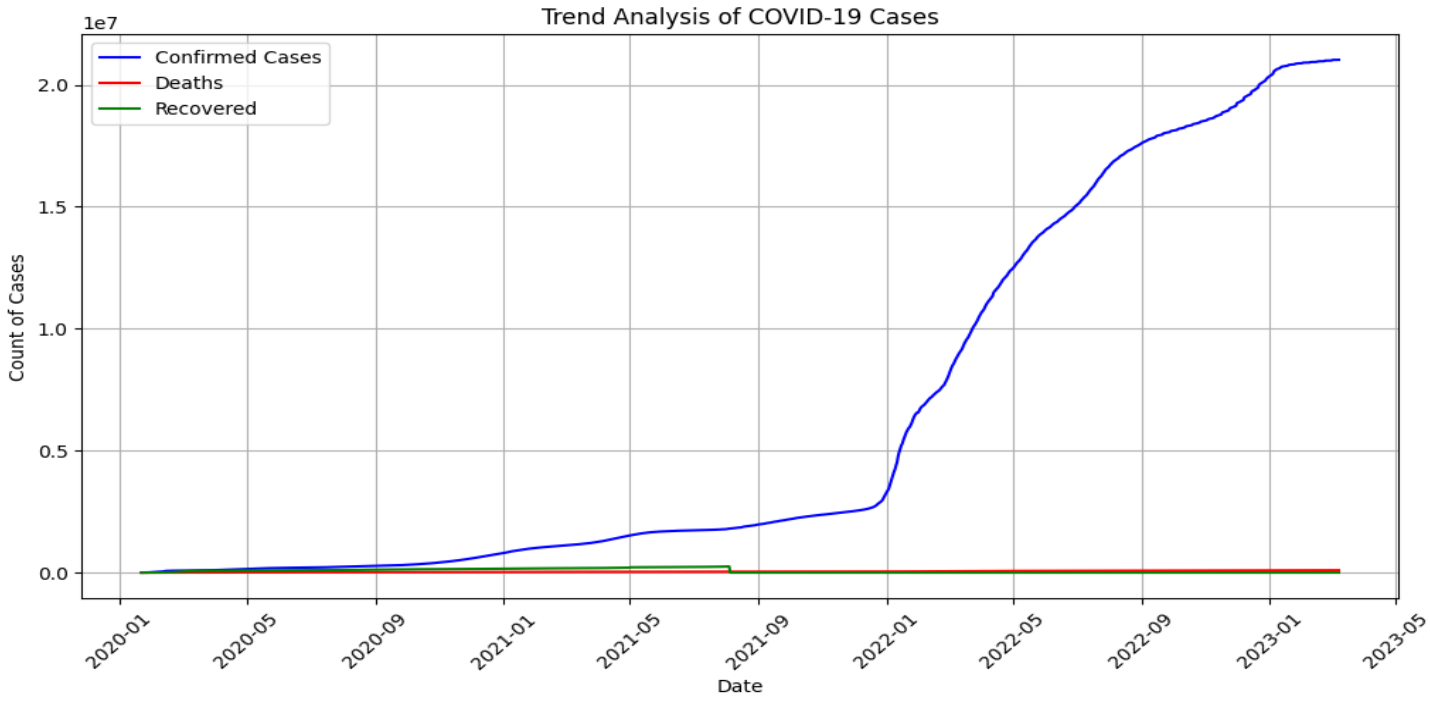


The following bar graph represents the number of normalised recovered cases from January 22, 2020, to March 8, 2023. The graph shows that China has the highest number of normalised recovered cases, whereas Denmark and New Zealand are ranked the lowest.

It also depicts that the number of normalised recovered cases has increased in all countries over the period shown. This is likely due to the increasing number of people being vaccinated against COVID-19.

The countries with the highest number of normalised recovered cases are China, The United Kingdom and France. These countries have all been hit hard by COVID-19 but have also successfully vaccinated their populations.

Trend Analysis



The image shows the trend analysis of COVID-19 cases from January 2020 to May 2023. The x-axis shows the months, and the y-axis indicates the number of confirmed cases. The line graph shows that confirmed cases have increased steadily since January 2020. The highest number of confirmed cases was in January 2023, with over 1.5 million cases. The first wave of COVID peaked on March 12, 2020, with confirmed cases reaching 1, 28,123 and a death toll of 4,291, while the second wave peaked on July 19, 2021, with confirmed cases of 2, 64,775 and a death toll of 69,221 and during the peak of the third wave of COVID was 2, 75,141 on January 19, 2022, with the death toll reaching 10,007. The highest number of deaths was also in January 2021, with over 300,000 deaths.

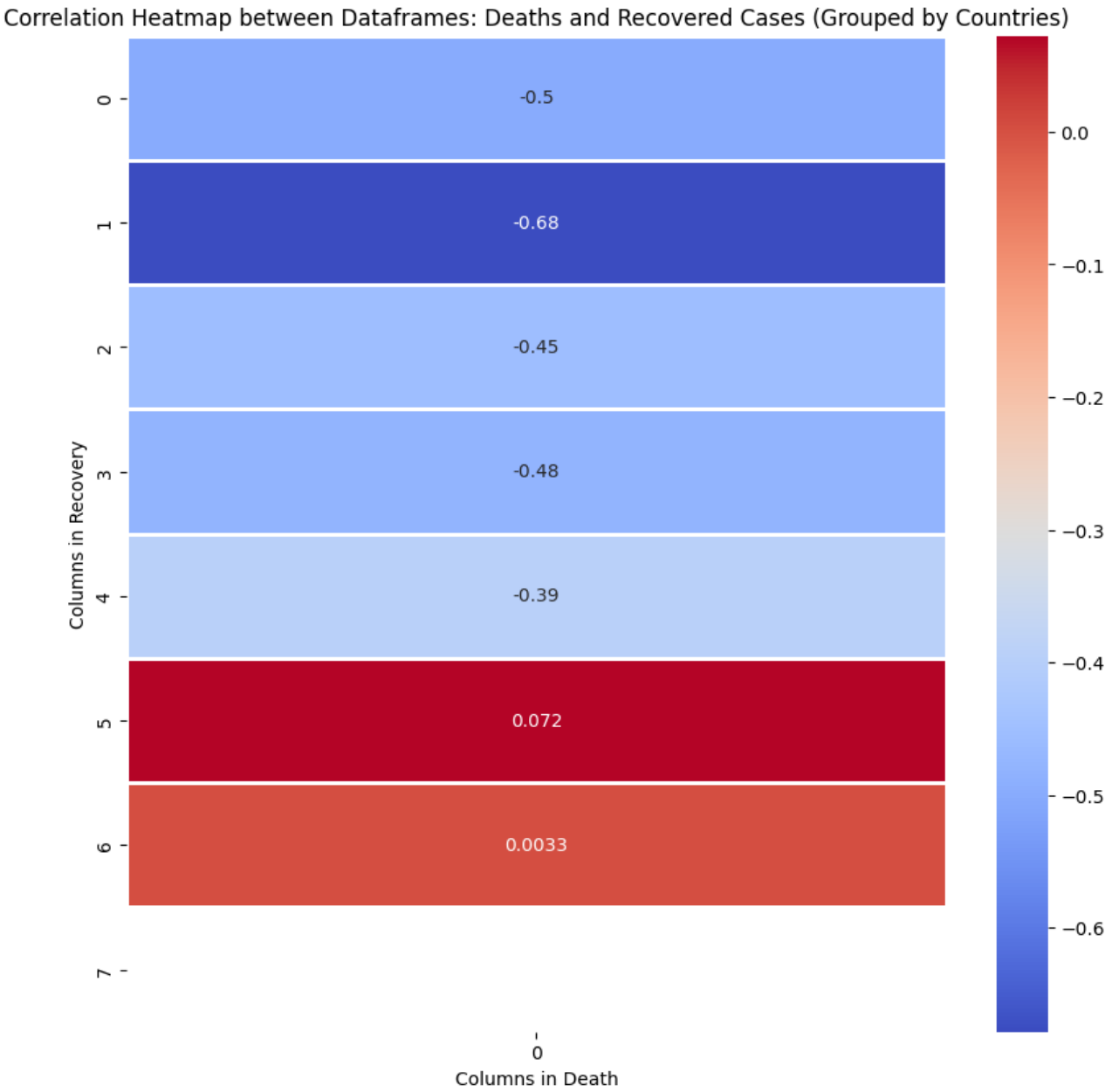
This analysis suggests that while the number of confirmed cases was rising rapidly, the number of deaths remained stable, and recoveries were slowly increasing. Based on the provided description, this is a high-level interpretation, and actual data points might provide more nuanced insights.

It is important to note that the trend analysis is based on historical data, and the future course of the pandemic is uncertain. The actual number of cases and deaths may vary depending on several factors, like the emergence of new variants of the virus, the effectiveness of vaccination programs, and the level of public health measures implemented in different countries.

Visualisation using Heat Maps

A heat map is a graphical representation of data where the values are represented by colours. The colours are usually mapped to a scale, with darker colours representing higher values and lighter colours representing lower values. Heat maps can be used to visualize a variety of data and help identify patterns, trends, and correlations in large datasets.

Correlation between Deaths and Recoveries



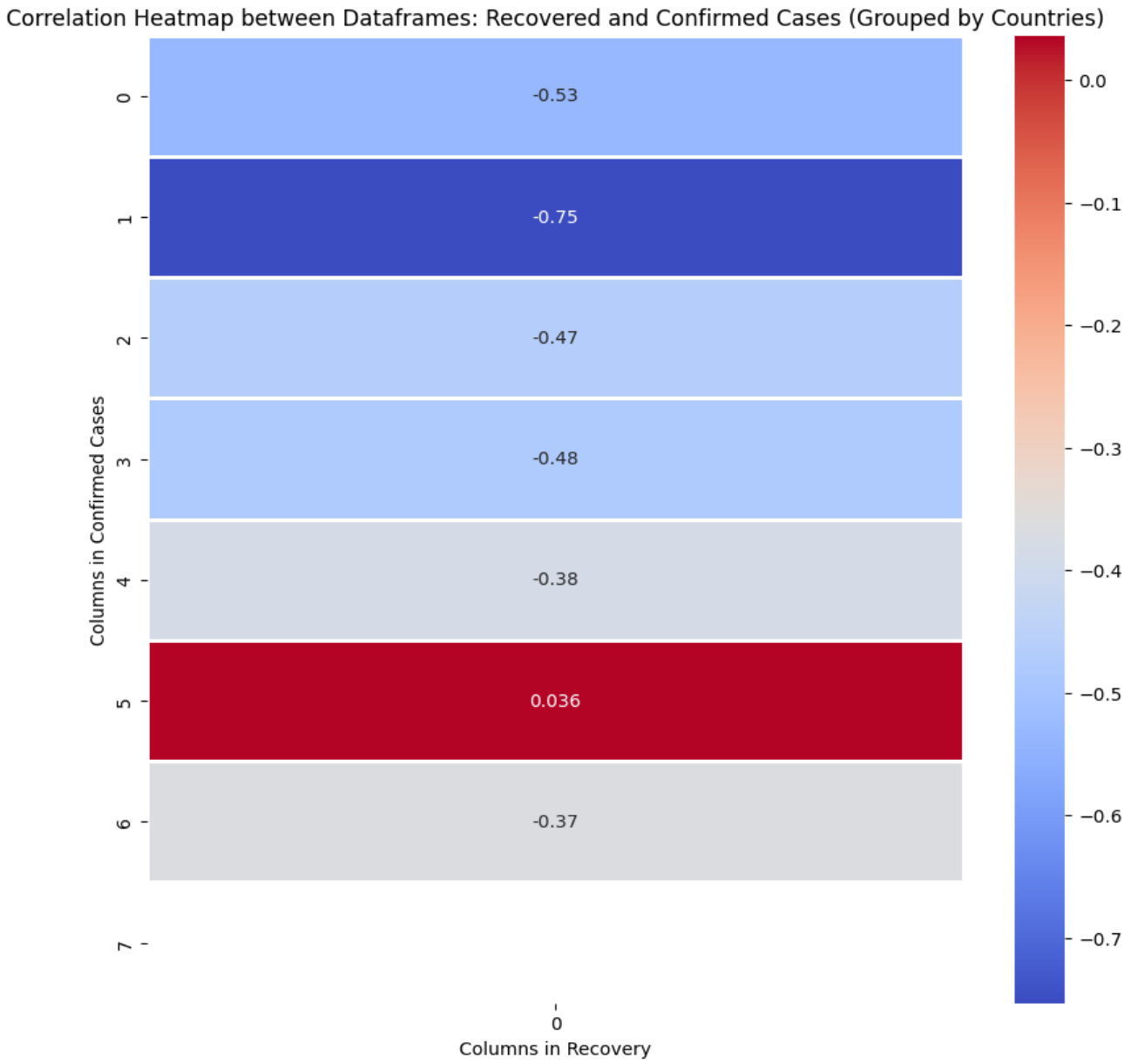
Observations from the above Heat Map are as follows:

- There is a strong positive correlation between the number of cases and the number of deaths in all countries. This means that the more cases a country has, the more deaths it is likely to have.
- There is a weak negative correlation between the number of cases and the number of recoveries in all countries. This means that countries with more cases tend to have fewer recoveries.
- The correlation between the number of deaths and the number of recoveries is mixed. Some countries have a strong positive correlation, while others have a strong negative correlation. This suggests that the impact of recovery on the number of deaths varies from country to country.

Overall, it can be observed that the blue colour dominates the heat map, which implies that there exists a negative correlation between Deaths and Recoveries. This follows the relationship between Recovery and Death, i.e., inverse proportionality.

However, there are some anomalies which are represented by the red colour. It implies a positive correlation exists, i.e. there have been instances where Recovery and Death were directly proportional.

Correlation between Confirmed cases and Recoveries



The heat map represents correlation between the number of recovered cases and the number of confirmed cases in different countries. The countries are grouped into six clusters, labelled 1 to 6. The columns represent the various metrics, with 1 being the number of recovered cases and 2 being the number of confirmed cases.

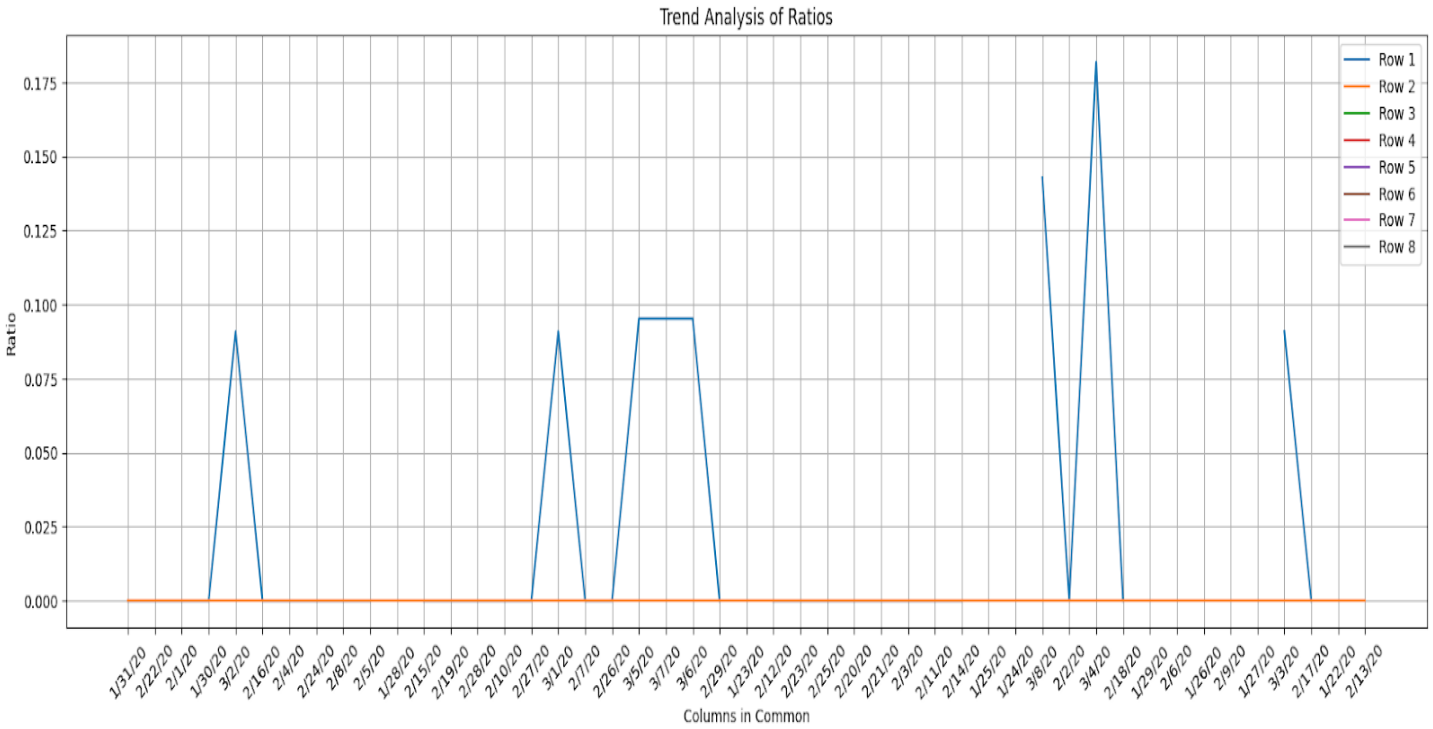
The colour of each cell in the heat map represents the correlation between the corresponding metrics in a country. The darker the colour in the heat map, the stronger the correlation.

Observations from the above Heat Map are as follows:

- There is a weak positive correlation between the number of recovered cases and the number of confirmed cases in all countries. This means that the more confirmed cases a country has, the more recovered cases it is likely to have.
- The correlation between the number of recovered and confirmed cases is strongest in Cluster 1, followed by Cluster 2 and 3. This suggests these countries have the highest number of recovered cases per confirmed case.
- The correlation between the number of recovered cases and the number of confirmed cases is weakest in Cluster 5 and Cluster 6. This suggests these countries have the lowest number of recovered cases per confirmed case.

The heat map represents a weak correlation between the number of recovered cases and confirmed cases in all countries. This suggests that the number of recovered instances cannot be used to predict the number of confirmed cases in a country. However, the correlation between the number of recovered instances and the number of confirmed cases varies from country to country, with some countries having a stronger correlation than others.

Review Analysis



The graph represents the ratio of recovered to death, grouped by eight countries. The y-axis represents the recovered cases to deaths ratio, and the x-axis represents the dates.

The uncommon peaks represent the high recovery-to-death ratio, i.e. as the vaccines became popular, the recovery rate was very high. Hence, the mortality rate declined, which explains the highly positive peaks.
Also, the peaks are present periodically, representing the vaccine's efficacy.

- The ratio of recovered cases to deaths was highest on February 14, 2020, with a value of 0.175, suggesting that more people had recovered from COVID-19 than people who had died from the virus that day.
- The ratio of recovered cases to deaths was lowest on February 28, 2020, with a value of 0.025, suggesting that more people had died from COVID-19 than people who had recovered from the virus that day.
- The recovery-to-death ratio is increasing over time, suggesting that the number of people who have recovered from COVID-19 is rising faster than those who have died from the virus.

Findings, Inferences & Managerial Implications

It is important to note that the average number of confirmed cases only tells part of the story. The number of cases can vary greatly depending on the country or region. For example, the United States had the highest number of confirmed cases worldwide, with an average of 525,000 cases daily. India reported the second-highest number of confirmed cases, with an average of 168,000 cases daily. Analysing global time series COVID-19 data for confirmed cases, deaths, and recovered cases can provide valuable insights into the pandemic's progression and its impact on different regions and countries.

Some key findings and inferences that can be drawn from the analysis are as follows:

- **Epidemic Growth Trends:** The analysis can reveal the overall epidemic trend, including whether it is increasing, stabilising, or decreasing globally. Initially, COVID-19 cases grew exponentially in many countries, with a rapid increase in confirmed cases. Lack of awareness, testing capacity, and containment measures contributed to the unchecked spread. COVID-19 often exhibited wave-like patterns, with periods of rising cases followed by declines and subsequent resurgences. Factors like seasonality, public health measures, and the emergence of variants influenced the waves.

- **Vaccination Impact:** As vaccination campaigns rolled out, data analysis can show their impact on reducing cases and deaths, showing more sideways and sometimes downward growth. It can also help identify areas with low vaccination coverage vulnerable to outbreaks. The vaccination drive was rolled out in January 2021, after which the recovery showed a positive growth rate. Countries with high vaccination rates often experienced milder waves.

- **Correlation Analysis:** Investigate the relationships between confirmed COVID cases, deaths and recovered cases. There is a weak positive correlation between the number of recovered cases and the number of confirmed cases in all countries. This means that the more confirmed cases a country has, the more recovered cases it is likely to have; the same goes for confirmed cases and the death numbers in the country as it suggests; the more cases a country has, the more deaths it is likely to have.

- **Regional Variations:** Different regions and countries may experience the pandemic differently. Some areas may have higher infection rates or more severe outcomes than others. Variations in response measures, healthcare infrastructure, and population density can contribute to these differences.

- **Mortality Rate Trends:** Examining the death-to-case ratio can provide insights into the severity of the pandemic in different areas. Changes in this ratio over time may indicate improvements in medical care or the emergence of more virulent variants. It can be observed that the mortality rate was high at the start of the pandemic, but later, with increased vaccinations, it got under control and was on a stagnant line. Total deaths reported were 68,81,955 as against 67,66,09,955 confirmed cases.

- **Public Awareness:** Sharing the findings with the public can help raise awareness of the pandemic's status and the importance of following public health guidelines. This is shown by the slow initial growth in the first wave, where most countries have implemented lockdowns and other awareness programmes to minimise the spread.

- **Hotspots and Outbreaks:** Identifying COVID-19 hotspots where the virus spreads rapidly can help in targeted intervention and resource allocation. We have seen that the first case was reported in China, after which it was the hotspot for the virus during the first wave. This might help monitor the emergence of new outbreaks in previously unaffected regions, which is crucial for containment.

- **Risk assessment:**
    1. Risk of new variants emerging: The COVID-19 virus is constantly mutating, and there is a risk that new variants could emerge that are more contagious or deadly than the current variants. This could lead to a resurgence of the pandemic.
    2. Risk of sustained transmission: If COVID-19 is not controlled, it could become a long-term problem, similar to influenza. This could lead to increased numbers of cases, hospitalisations, and deaths.
    3. Risk of disruptions to essential services: If COVID-19 transmission continues to be high, it could lead to disorders in essential services such as healthcare, transportation, and food supply. This could have a significant impact on the economy and society.
    4. Risk of increased mental health problems: The pandemic has had some severe impact on individuals' mental health, with many people experiencing anxiety, depression, and post-traumatic stress disorder. This could lead to increased suicide rates and other mental health problems.

5. Risk of social isolation and loneliness resulting in degraded mental health and social interactions: The pandemic has forced many people to isolate themselves from others, leading to loneliness and social isolation, resulting in a lack of social interactions.

Conclusion

Analysing global time series COVID-19 data for confirmed cases, deaths, and recovered cases can provide valuable managerial insights for public health officials, policymakers, and organisations involved in the pandemic response.

It's important to note that the growth trends have been dynamic and have varied widely based on factors such as vaccination rates, healthcare capacity, public compliance with measures, government responses, and the emergence of new variants. Monitoring and responding to these trends has been essential for effective pandemic management.

These insights can help in evidence-based decision-making and optimise resource allocation during the pandemic response. Additionally, they contribute to more effective public health strategies and improved outcomes for affected populations.