

# covid19

May 17, 2025

```
[64]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[2]: df = pd.read_csv("owid-covid-data.csv")
```

```
[3]: df.head(20)
```

```
[3]:
```

	iso_code	continent	location	date	total_cases	new_cases	\
0	AFG	Asia	Afghanistan	05/01/2020	0.0	0.0	
1	AFG	Asia	Afghanistan	06/01/2020	0.0	0.0	
2	AFG	Asia	Afghanistan	07/01/2020	0.0	0.0	
3	AFG	Asia	Afghanistan	08/01/2020	0.0	0.0	
4	AFG	Asia	Afghanistan	09/01/2020	0.0	0.0	
5	AFG	Asia	Afghanistan	10/01/2020	0.0	0.0	
6	AFG	Asia	Afghanistan	11/01/2020	0.0	0.0	
7	AFG	Asia	Afghanistan	12/01/2020	0.0	0.0	
8	AFG	Asia	Afghanistan	13/01/2020	0.0	0.0	
9	AFG	Asia	Afghanistan	14/01/2020	0.0	0.0	
10	AFG	Asia	Afghanistan	15/01/2020	0.0	0.0	
11	AFG	Asia	Afghanistan	16/01/2020	0.0	0.0	
12	AFG	Asia	Afghanistan	17/01/2020	0.0	0.0	
13	AFG	Asia	Afghanistan	18/01/2020	0.0	0.0	
14	AFG	Asia	Afghanistan	19/01/2020	0.0	0.0	
15	AFG	Asia	Afghanistan	20/01/2020	0.0	0.0	
16	AFG	Asia	Afghanistan	21/01/2020	0.0	0.0	
17	AFG	Asia	Afghanistan	22/01/2020	0.0	0.0	
18	AFG	Asia	Afghanistan	23/01/2020	0.0	0.0	
19	AFG	Asia	Afghanistan	24/01/2020	0.0	0.0	

	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	\
0	NaN	0.0	0.0	NaN	...	
1	NaN	0.0	0.0	NaN	...	
2	NaN	0.0	0.0	NaN	...	
3	NaN	0.0	0.0	NaN	...	
4	NaN	0.0	0.0	NaN	...	
5	0.0	0.0	0.0	0.0	...	

6	0.0	0.0	0.0	0.0	...
7	0.0	0.0	0.0	0.0	...
8	0.0	0.0	0.0	0.0	...
9	0.0	0.0	0.0	0.0	...
10	0.0	0.0	0.0	0.0	...
11	0.0	0.0	0.0	0.0	...
12	0.0	0.0	0.0	0.0	...
13	0.0	0.0	0.0	0.0	...
14	0.0	0.0	0.0	0.0	...
15	0.0	0.0	0.0	0.0	...
16	0.0	0.0	0.0	0.0	...
17	0.0	0.0	0.0	0.0	...
18	0.0	0.0	0.0	0.0	...
19	0.0	0.0	0.0	0.0	...

	male_smokers	handwashing_facilities	hospital_beds_per_thousand	\
0	NaN	37.746	0.5	
1	NaN	37.746	0.5	
2	NaN	37.746	0.5	
3	NaN	37.746	0.5	
4	NaN	37.746	0.5	
5	NaN	37.746	0.5	
6	NaN	37.746	0.5	
7	NaN	37.746	0.5	
8	NaN	37.746	0.5	
9	NaN	37.746	0.5	
10	NaN	37.746	0.5	
11	NaN	37.746	0.5	
12	NaN	37.746	0.5	
13	NaN	37.746	0.5	
14	NaN	37.746	0.5	
15	NaN	37.746	0.5	
16	NaN	37.746	0.5	
17	NaN	37.746	0.5	
18	NaN	37.746	0.5	
19	NaN	37.746	0.5	

	life_expectancy	human_development_index	population	\
0	64.83	0.511	41128772.0	
1	64.83	0.511	41128772.0	
2	64.83	0.511	41128772.0	
3	64.83	0.511	41128772.0	
4	64.83	0.511	41128772.0	
5	64.83	0.511	41128772.0	
6	64.83	0.511	41128772.0	
7	64.83	0.511	41128772.0	
8	64.83	0.511	41128772.0	

9	64.83	0.511	41128772.0
10	64.83	0.511	41128772.0
11	64.83	0.511	41128772.0
12	64.83	0.511	41128772.0
13	64.83	0.511	41128772.0
14	64.83	0.511	41128772.0
15	64.83	0.511	41128772.0
16	64.83	0.511	41128772.0
17	64.83	0.511	41128772.0
18	64.83	0.511	41128772.0
19	64.83	0.511	41128772.0

	excess_mortality_cumulative_absolute	excess_mortality_cumulative	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	
7	NaN	NaN	
8	NaN	NaN	
9	NaN	NaN	
10	NaN	NaN	
11	NaN	NaN	
12	NaN	NaN	
13	NaN	NaN	
14	NaN	NaN	
15	NaN	NaN	
16	NaN	NaN	
17	NaN	NaN	
18	NaN	NaN	
19	NaN	NaN	

	excess_mortality	excess_mortality_cumulative_per_million
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	NaN
6	NaN	NaN
7	NaN	NaN
8	NaN	NaN
9	NaN	NaN
10	NaN	NaN
11	NaN	NaN

12	NaN	NaN
13	NaN	NaN
14	NaN	NaN
15	NaN	NaN
16	NaN	NaN
17	NaN	NaN
18	NaN	NaN
19	NaN	NaN

[20 rows x 67 columns]

```
[4]: df.columns
```

```
[4]: Index(['iso_code', 'continent', 'location', 'date', 'total_cases', 'new_cases',
        'new_cases_smoothed', 'total_deaths', 'new_deaths',
        'new_deaths_smoothed', 'total_cases_per_million',
        'new_cases_per_million', 'new_cases_smoothed_per_million',
        'total_deaths_per_million', 'new_deaths_per_million',
        'new_deaths_smoothed_per_million', 'reproduction_rate', 'icu_patients',
        'icu_patients_per_million', 'hosp_patients',
        'hosp_patients_per_million', 'weekly_icu_admissions',
        'weekly_icu_admissions_per_million', 'weekly_hosp_admissions',
        'weekly_hosp_admissions_per_million', 'total_tests', 'new_tests',
        'total_tests_per_thousand', 'new_tests_per_thousand',
        'new_tests_smoothed', 'new_tests_smoothed_per_thousand',
        'positive_rate', 'tests_per_case', 'tests_units', 'total_vaccinations',
        'people_vaccinated', 'people_fully_vaccinated', 'total_boosters',
        'new_vaccinations', 'new_vaccinations_smoothed',
        'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred',
        'people_fully_vaccinated_per_hundred', 'total_boosters_per_hundred',
        'new_vaccinations_smoothed_per_million',
        'new_people_vaccinated_smoothed',
        'new_people_vaccinated_smoothed_per_hundred', 'stringency_index',
        'population_density', 'median_age', 'aged_65_old', 'aged_70_old',
        'gdp_per_capita', 'extreme_poverty', 'cardiovasc_death_rate',
        'diabetes_prevalence', 'female_smokers', 'male_smokers',
        'handwashing_facilities', 'hospital_beds_per_thousand',
        'life_expectancy', 'human_development_index', 'population',
        'excess_mortality_cumulative_absolute', 'excess_mortality_cumulative',
        'excess_mortality', 'excess_mortality_cumulative_per_million'],
        dtype='object')
```

```
[7]: # checking null values
df.isnull().sum()
```

```
[7]: iso_code      0
continent    16799
```

```

location          0
date              1
total_cases       9339
...
population        1
excess_mortality_cumulative_absolute  249904
excess_mortality_cumulative          249904
excess_mortality                    249904
excess_mortality_cumulative_per_million  249904
Length: 67, dtype: int64

```

```
[6]: # checking number of rows and column
df.shape
```

```
[6]: (258583, 67)
```

```
[24]: # finding index with missing date
null_date_index = df[df['date'].isnull()].index
```

```
[25]: # dropping missing date and getting second dataframe
df2 = df.drop(null_date_index)
```

```
[26]: # original dataframe
df.isnull().sum()
```

```
[26]: iso_code          0
continent             16799
location              0
date                  1
total_cases           9339
...
population            1
excess_mortality_cumulative_absolute  249904
excess_mortality_cumulative          249904
excess_mortality                    249904
excess_mortality_cumulative_per_million  249904
Length: 67, dtype: int64

```

```
[27]: # data with dropped empty data
df2.isnull().sum()
```

```
[27]: iso_code          0
continent             16799
location              0
date                  0
total_cases           9338
...

```

```

population                                0
excess_mortality_cumulative_absolute      249903
excess_mortality_cumulative               249903
excess_mortality                         249903
excess_mortality_cumulative_per_million   249903
Length: 67, dtype: int64

```

```

[28]: # handling missing values
df3 = df2.fillna(value = 0)

```

```

[30]: df3.isnull().sum()

```

```

[30]: iso_code                0
continent                    0
location                     0
date                         0
total_cases                  0
..
population                   0
excess_mortality_cumulative_absolute  0
excess_mortality_cumulative         0
excess_mortality                   0
excess_mortality_cumulative_per_million  0
Length: 67, dtype: int64

```

```

[54]: # changing to datetime
df3['date'] = pd.to_datetime(df3['date'], format="mixed")

```

```

[56]: # checking if change was successful
df3.head()

```

```

[56]: iso_code  continent  location  date  total_cases  new_cases  \
0      AFG      Asia  Afghanistan 2020-05-01         0.0         0.0
1      AFG      Asia  Afghanistan 2020-06-01         0.0         0.0
2      AFG      Asia  Afghanistan 2020-07-01         0.0         0.0
3      AFG      Asia  Afghanistan 2020-08-01         0.0         0.0
4      AFG      Asia  Afghanistan 2020-09-01         0.0         0.0

      new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  ...  \
0                    0.0            0.0         0.0                    0.0  ...
1                    0.0            0.0         0.0                    0.0  ...
2                    0.0            0.0         0.0                    0.0  ...
3                    0.0            0.0         0.0                    0.0  ...
4                    0.0            0.0         0.0                    0.0  ...

      male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0              0.0                37.746                0.5

```

1	0.0	37.746	0.5
2	0.0	37.746	0.5
3	0.0	37.746	0.5
4	0.0	37.746	0.5

	life_expectancy	human_development_index	population \
0	64.83	0.511	41128772.0
1	64.83	0.511	41128772.0
2	64.83	0.511	41128772.0
3	64.83	0.511	41128772.0
4	64.83	0.511	41128772.0

	excess_mortality_cumulative_absolute	excess_mortality_cumulative \
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0

	excess_mortality	excess_mortality_cumulative_per_million
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0

[5 rows x 67 columns]

```
[60]: # filter countries
countries = ["Nigeria", "Kenya", "South Africa", "China", "India",
↳ "Brazil", "United State", "England",]
df3_filtered = df3[df3["location"].isin(countries)].copy()
```

```
[63]: # confirming countries filtered
df3_filtered.head(20)
```

```
[63]:
```

	iso_code	continent	location	date	total_cases	new_cases \
50234	BRA	South America	Brazil	2020-05-01	0.0	0.0
50235	BRA	South America	Brazil	2020-06-01	0.0	0.0
50236	BRA	South America	Brazil	2020-07-01	0.0	0.0
50237	BRA	South America	Brazil	2020-08-01	0.0	0.0
50238	BRA	South America	Brazil	2020-09-01	0.0	0.0
50239	BRA	South America	Brazil	2020-10-01	0.0	0.0
50240	BRA	South America	Brazil	2020-11-01	0.0	0.0
50241	BRA	South America	Brazil	2020-12-01	0.0	0.0
50242	BRA	South America	Brazil	2020-01-13	0.0	0.0
50243	BRA	South America	Brazil	2020-01-14	0.0	0.0

50244	BRA	South America	Brazil	2020-01-15	0.0	0.0
50245	BRA	South America	Brazil	2020-01-16	0.0	0.0
50246	BRA	South America	Brazil	2020-01-17	0.0	0.0
50247	BRA	South America	Brazil	2020-01-18	0.0	0.0
50248	BRA	South America	Brazil	2020-01-19	0.0	0.0
50249	BRA	South America	Brazil	2020-01-20	0.0	0.0
50250	BRA	South America	Brazil	2020-01-21	0.0	0.0
50251	BRA	South America	Brazil	2020-01-22	0.0	0.0
50252	BRA	South America	Brazil	2020-01-23	0.0	0.0
50253	BRA	South America	Brazil	2020-01-24	0.0	0.0

	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	\
50234	0.0	0.0	0.0	0.0	...	
50235	0.0	0.0	0.0	0.0	...	
50236	0.0	0.0	0.0	0.0	...	
50237	0.0	0.0	0.0	0.0	...	
50238	0.0	0.0	0.0	0.0	...	
50239	0.0	0.0	0.0	0.0	...	
50240	0.0	0.0	0.0	0.0	...	
50241	0.0	0.0	0.0	0.0	...	
50242	0.0	0.0	0.0	0.0	...	
50243	0.0	0.0	0.0	0.0	...	
50244	0.0	0.0	0.0	0.0	...	
50245	0.0	0.0	0.0	0.0	...	
50246	0.0	0.0	0.0	0.0	...	
50247	0.0	0.0	0.0	0.0	...	
50248	0.0	0.0	0.0	0.0	...	
50249	0.0	0.0	0.0	0.0	...	
50250	0.0	0.0	0.0	0.0	...	
50251	0.0	0.0	0.0	0.0	...	
50252	0.0	0.0	0.0	0.0	...	
50253	0.0	0.0	0.0	0.0	...	

	male_smokers	handwashing_facilities	hospital_beds_per_thousand	\
50234	17.9	0.0	2.2	
50235	17.9	0.0	2.2	
50236	17.9	0.0	2.2	
50237	17.9	0.0	2.2	
50238	17.9	0.0	2.2	
50239	17.9	0.0	2.2	
50240	17.9	0.0	2.2	
50241	17.9	0.0	2.2	
50242	17.9	0.0	2.2	
50243	17.9	0.0	2.2	
50244	17.9	0.0	2.2	
50245	17.9	0.0	2.2	
50246	17.9	0.0	2.2	



50247	17.9	0.0	2.2
50248	17.9	0.0	2.2
50249	17.9	0.0	2.2
50250	17.9	0.0	2.2
50251	17.9	0.0	2.2
50252	17.9	0.0	2.2
50253	17.9	0.0	2.2

	life_expectancy	human_development_index	population \
50234	75.88	0.765	215313504.0
50235	75.88	0.765	215313504.0
50236	75.88	0.765	215313504.0
50237	75.88	0.765	215313504.0
50238	75.88	0.765	215313504.0
50239	75.88	0.765	215313504.0
50240	75.88	0.765	215313504.0
50241	75.88	0.765	215313504.0
50242	75.88	0.765	215313504.0
50243	75.88	0.765	215313504.0
50244	75.88	0.765	215313504.0
50245	75.88	0.765	215313504.0
50246	75.88	0.765	215313504.0
50247	75.88	0.765	215313504.0
50248	75.88	0.765	215313504.0
50249	75.88	0.765	215313504.0
50250	75.88	0.765	215313504.0
50251	75.88	0.765	215313504.0
50252	75.88	0.765	215313504.0
50253	75.88	0.765	215313504.0

	excess_mortality_cumulative_absolute	excess_mortality_cumulative \
50234	0.0	0.0
50235	0.0	0.0
50236	0.0	0.0
50237	0.0	0.0
50238	0.0	0.0
50239	0.0	0.0
50240	0.0	0.0
50241	0.0	0.0
50242	0.0	0.0
50243	0.0	0.0
50244	0.0	0.0
50245	0.0	0.0
50246	0.0	0.0
50247	0.0	0.0
50248	0.0	0.0
50249	0.0	0.0

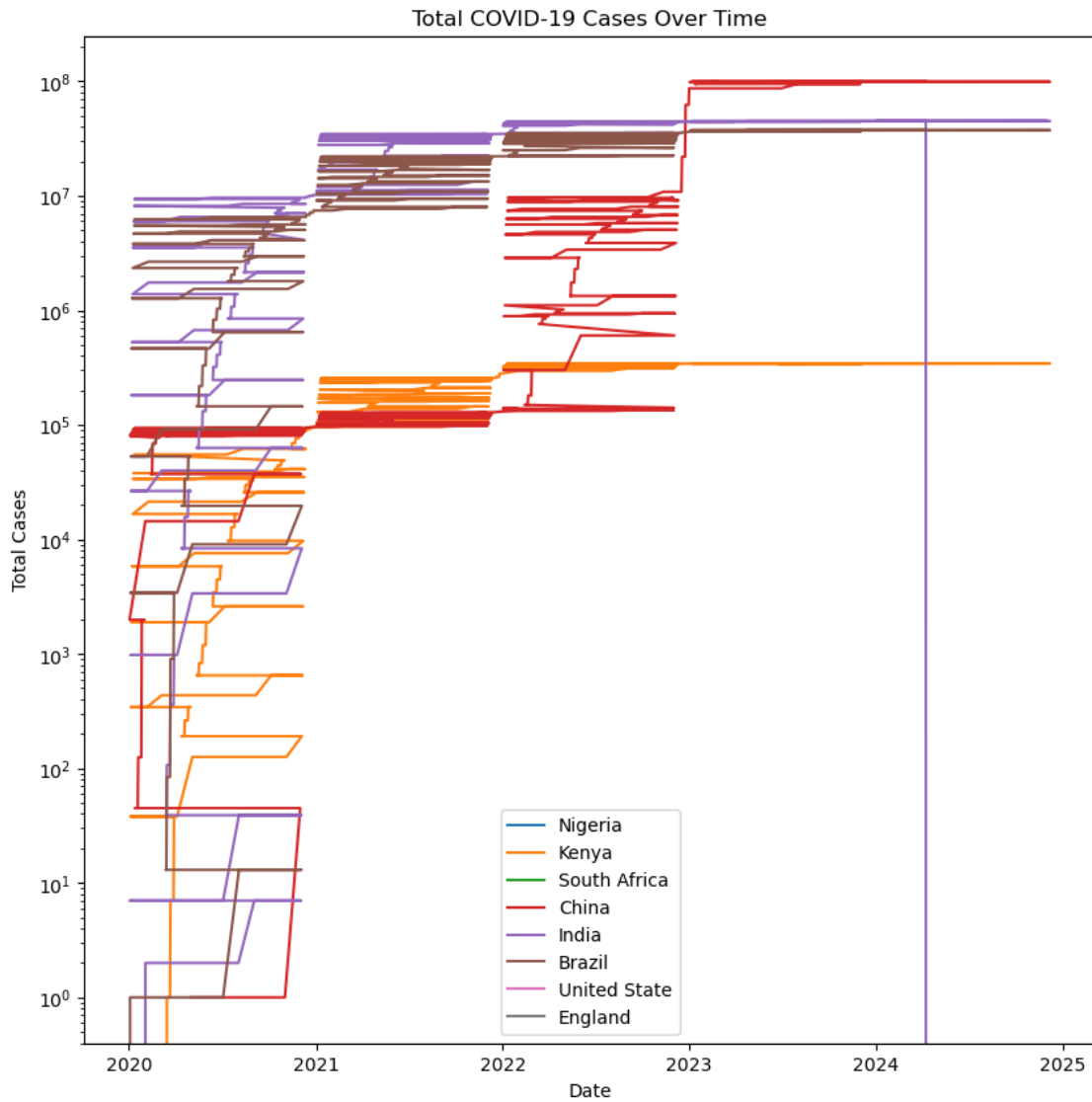
50250	0.0	0.0
50251	0.0	0.0
50252	0.0	0.0
50253	0.0	0.0

	excess_mortality	excess_mortality_cumulative_per_million
50234	0.0	0.0
50235	0.0	0.0
50236	0.0	0.0
50237	0.0	0.0
50238	0.0	0.0
50239	0.0	0.0
50240	0.0	0.0
50241	0.0	0.0
50242	0.0	0.0
50243	0.0	0.0
50244	0.0	0.0
50245	0.0	0.0
50246	0.0	0.0
50247	0.0	0.0
50248	0.0	0.0
50249	0.0	0.0
50250	0.0	0.0
50251	0.0	0.0
50252	0.0	0.0
50253	0.0	0.0

[20 rows x 67 columns]

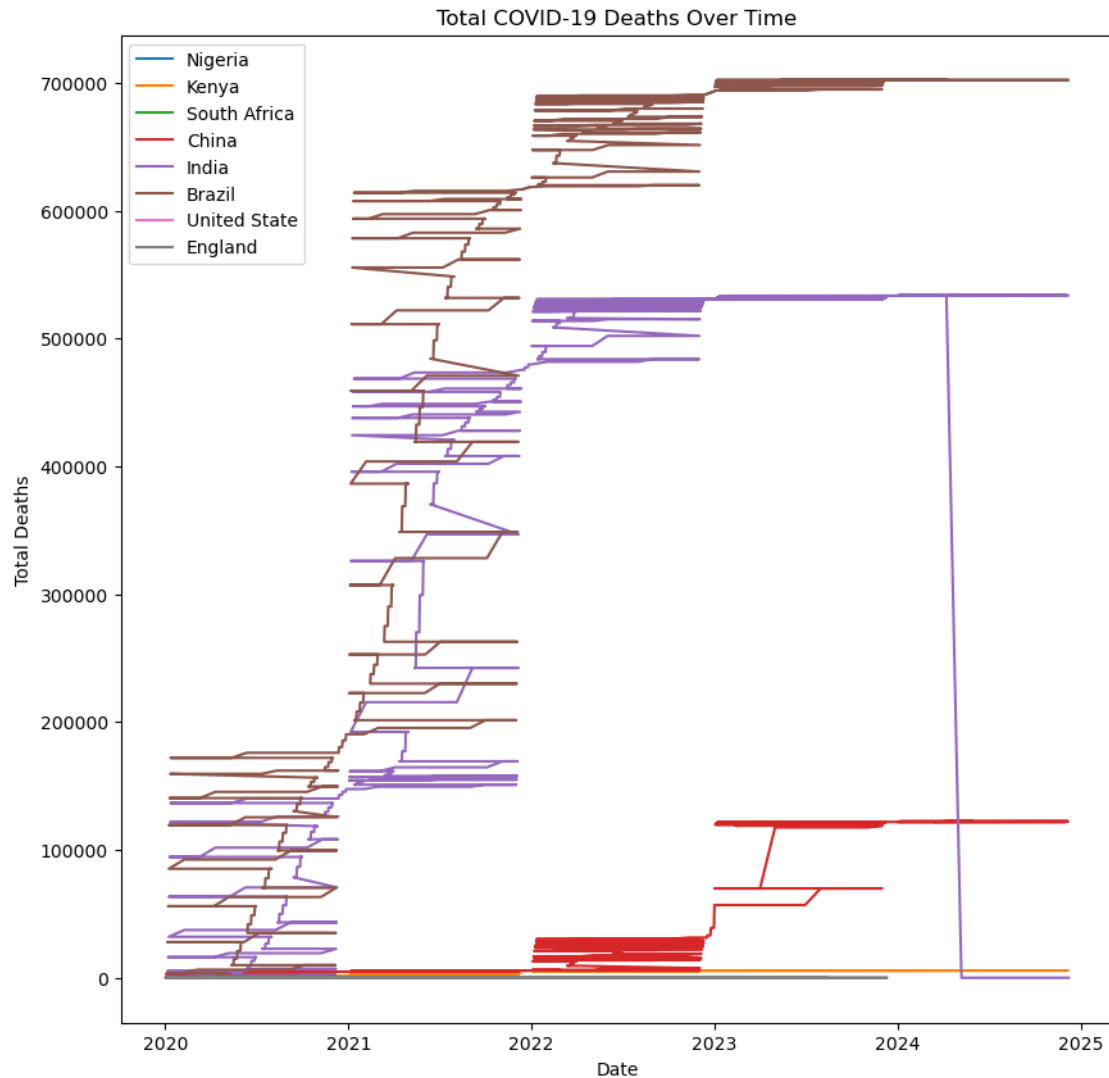
```
[85]: plt.figure(figsize=(10, 10))
      for country in countries:
          country_data = df3_filtered[df3_filtered['location'] == country]
          plt.plot(country_data['date'], country_data['total_cases'], label=country)

      plt.title('Total COVID-19 Cases Over Time')
      plt.xlabel('Date')
      plt.ylabel('Total Cases')
      plt.legend()
      plt.yscale('log')
      plt.show()
```



```
[87]: plt.figure(figsize=(10, 10))
      for country in countries:
          country_data = df3_filtered[df3_filtered['location'] == country]
          plt.plot(country_data['date'], country_data['total_deaths'], label=country)

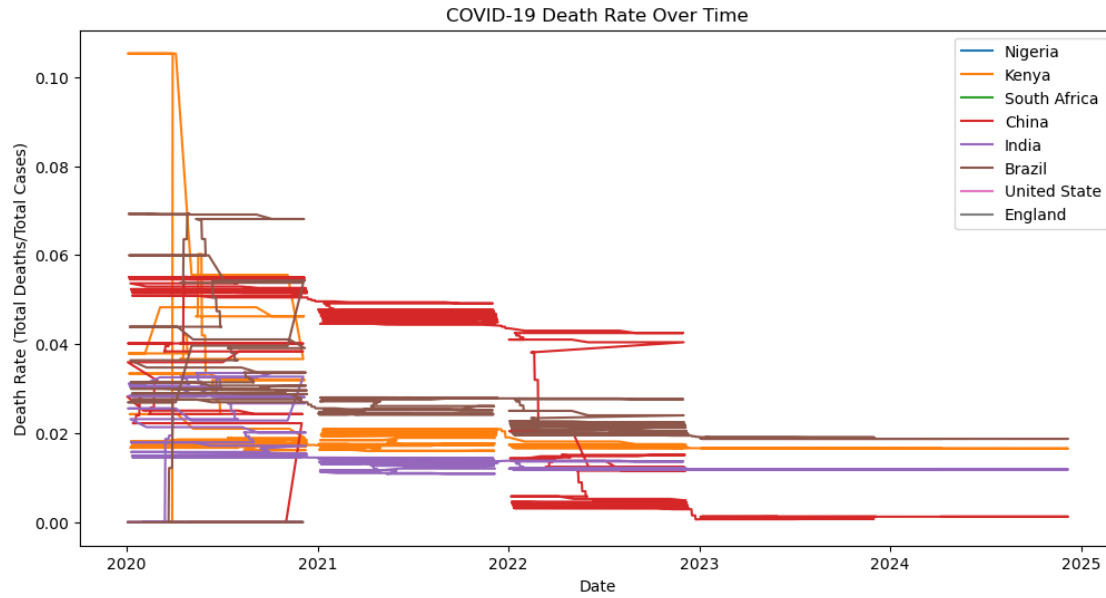
      plt.title('Total COVID-19 Deaths Over Time')
      plt.xlabel('Date')
      plt.ylabel('Total Deaths')
      plt.legend()
      plt.show()
```



```
[74]: df3_filtered['death_rate'] = df3_filtered['total_deaths'] / \
      ↪ df3_filtered['total_cases']

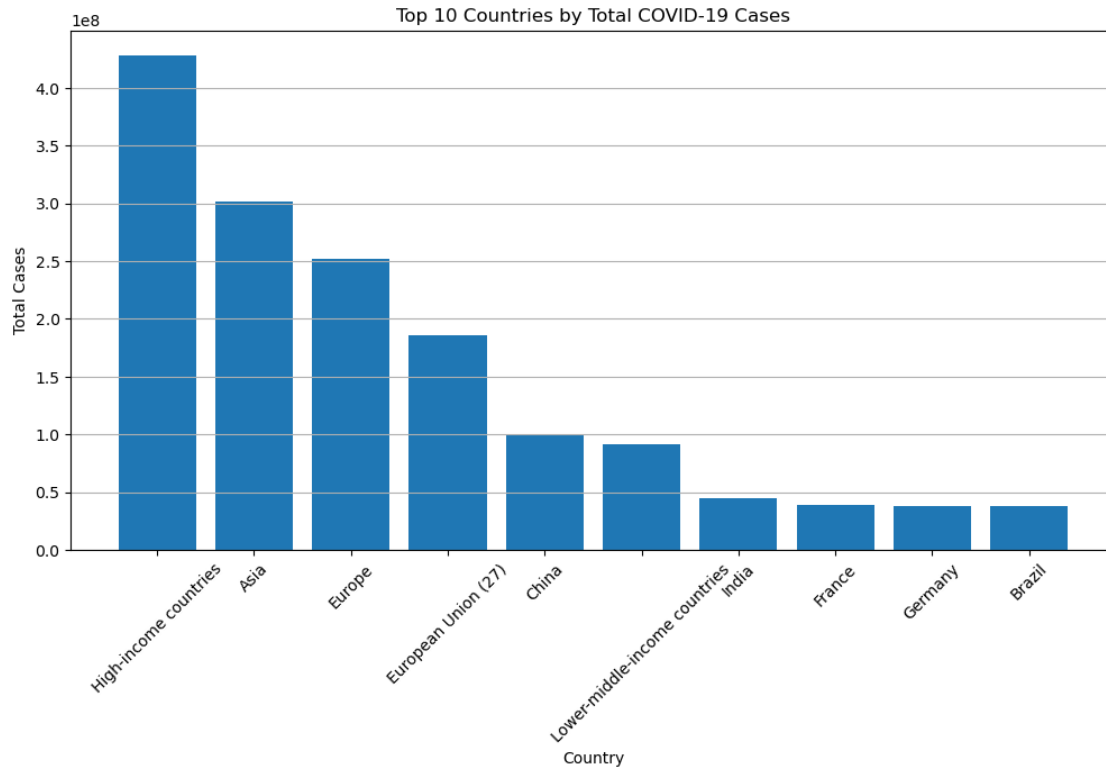
plt.figure(figsize=(12, 6))
for country in countries:
    country_data = df3_filtered[df3_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['death_rate'], label=country)

plt.title('COVID-19 Death Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Death Rate (Total Deaths/Total Cases)')
plt.legend()
plt.show()
```



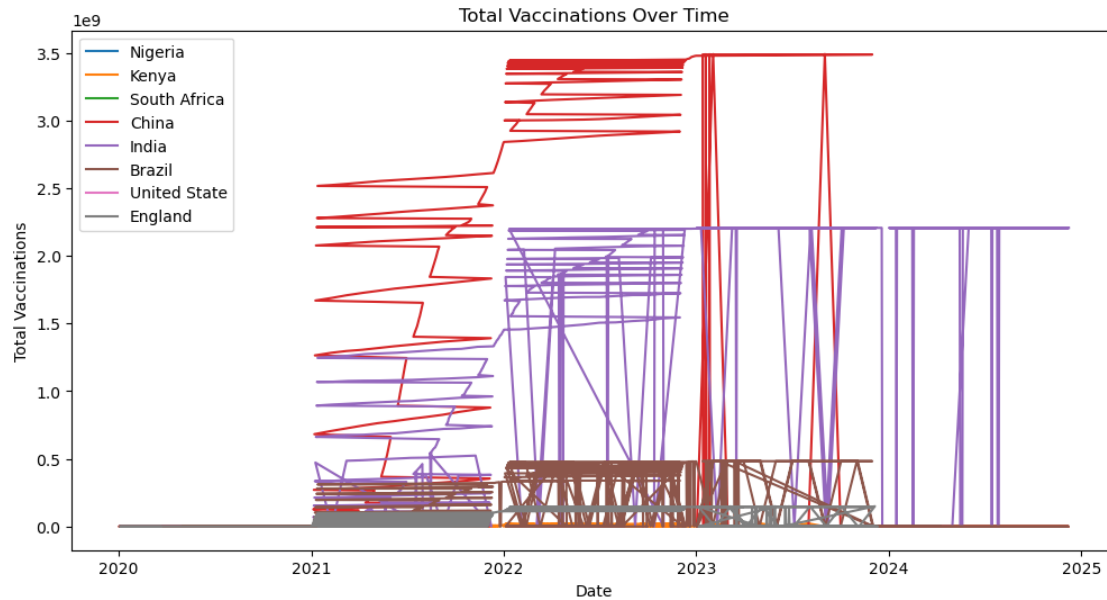
```
[76]: latest_data = df.sort_values('date').groupby('location').last().reset_index()
top_countries = latest_data.nlargest(10, 'total_cases')[['location', 'total_cases']]

plt.figure(figsize=(12, 6))
plt.bar(top_countries['location'], top_countries['total_cases'])
plt.title('Top 10 Countries by Total COVID-19 Cases')
plt.xlabel('Country')
plt.ylabel('Total Cases')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.show()
```



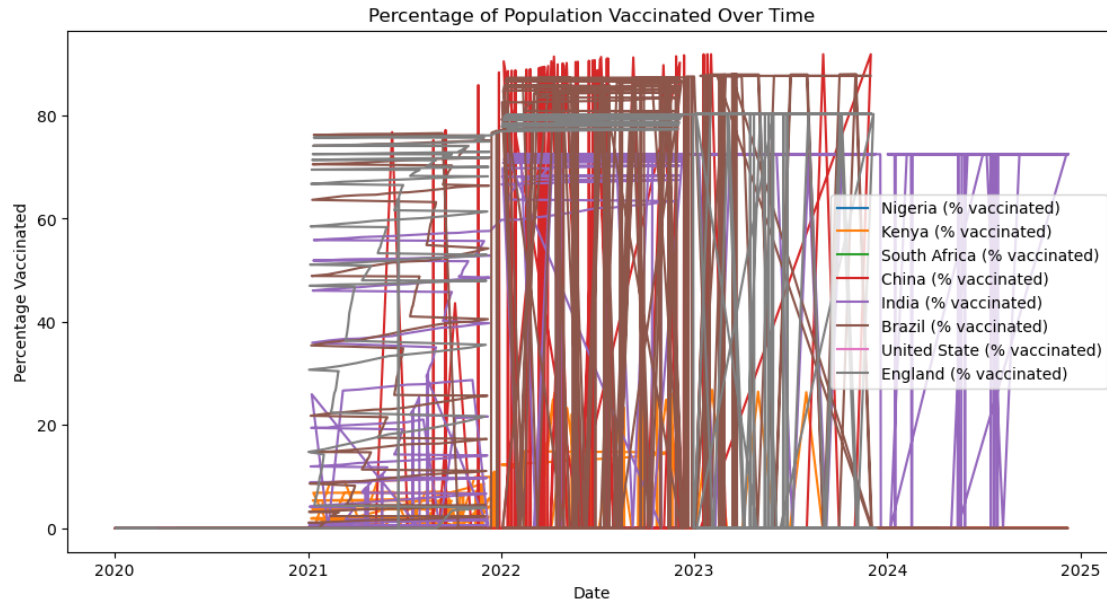
```
[77]: # Plot cumulative vaccinations over time for selected countries
plt.figure(figsize=(12, 6))
for country in countries:
    country_data = df3_filtered[df3_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_vaccinations'],
             label=country)

plt.title('Total Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.legend()
plt.show()
```



```
[79]: # Plot vaccinations per hundred
plt.figure(figsize=(12, 6))
for country in countries:
    country_data = df3_filtered[df3_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['people_vaccinated_per_hundred'],
             label=f'{country} (% vaccinated)')

plt.title('Percentage of Population Vaccinated Over Time')
plt.xlabel('Date')
plt.ylabel('Percentage Vaccinated')
plt.legend()
plt.show()
```

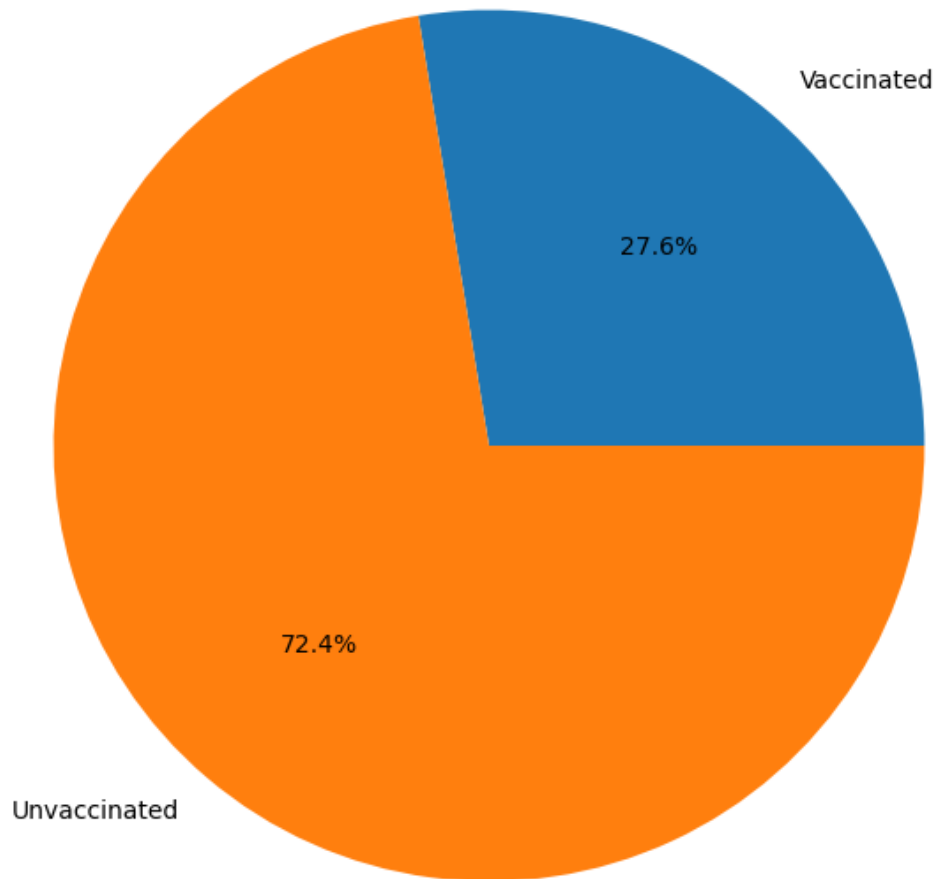


```
[82]: total_vaccinated = df['people_vaccinated'].sum()
world_population = df['population'].sum()
unvaccinated = world_population - total_vaccinated

plt.figure(figsize=(8, 8))
plt.pie([total_vaccinated, unvaccinated],
        labels=['Vaccinated', 'Unvaccinated'],
        autopct='%1.1f%%')
plt.title('Global Vaccination Status')
plt.show()
```



### Global Vaccination Status



USING DATA FROM "OWID-COVID-DATA.CSV". Because of the large data size, just a few countries was filtered for analysis. Countries data used are United State, India, Brazil, Nigeria, kenya, South Africa, China and England. Analysis were based on data from the csv file and the selected countries 1. COVID CASES OVER TIME In terms of covid case over time, India and Brazil had the highest case from 2020 till 2022. China case increased from 2020 till 2021 so did Kenya.

2. DEATH OVER TIME The highest for "death over time" was Brazil from 2020-2023. India was also high from 2021 but still lower than compare to Brazil. China death cases increased from 2022.
3. TOTAL DEATH/TOTAL CASE Comparing the number of covid cases to death recorded for cases, Kenya had the highest followed by Brazil by 2020. It is also noted that by 2021-

2023, Kenya death cases has reduced very well. China and Brazil also had high death from 2020-2023.

4. **VACCINATED AGAINST UNVACCINATED** Looking at the pie chart concerning the world population vaccination, it shows majority of the population are yet to be vaccinated

[ ]:

