covid 19 datatracker

May 14, 2025

1 Global COVID-19 Vaccination Analysis

This report presents an analysis of COVID-19 vaccination rollouts across various countries. It includes data exploration, visualizations, and key insights drawn from recent trends.

[ ]:

```
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px


df = pd.read_csv('owid-covid-data.csv')

#Checking if I have the right columns
print(df.columns)
```

[17]:

```
print(df.head())
```

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

   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   
```

```
   life_expectancy  human_development_index  population  \
0            64.83                    0.511    41128772
1            64.83                    0.511    41128772
2            64.83                    0.511    41128772
3            64.83                    0.511    41128772
4            64.83                    0.511    41128772

   excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                   NaN                          NaN
1                                   NaN                          NaN
2                                   NaN                          NaN
3                                   NaN                          NaN
4                                   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 rows x 67 columns]
```

[18]:

```
# Define key columns
key_columns = ["date", "location", "total_cases", "total_deaths", "new_cases",␣
 ↪"new_deaths", "total_vaccinations"]

# Check for missing columns
missing_cols = [col for col in key_columns if col not in df.columns]

if missing_cols:
    print(f"Warning: These columns are missing: {missing_cols}")
else:
    print("All key columns are present!")
```

```
All key columns are present!
```

[19]:

```
# Filter for Kenya and making a copy
kenya_df = df[df["location"] == "Kenya"].copy()

# Drop rows with missing dates or critical values
critical_columns = ["date", "total_cases", "total_deaths", "new_cases",␣
 ↪"new_deaths", "total_vaccinations"]
kenya_df.dropna(subset=critical_columns, inplace=True)

# Display cleaned data
print(kenya_df.head())
```

```
       iso_code continent location        date  total_cases  new_cases  \
195746      KEN    Africa    Kenya  2021-03-04     105648.0        0.0
195752      KEN    Africa    Kenya  2021-03-10     108362.0        0.0
195756      KEN    Africa    Kenya  2021-03-14     112805.0     4443.0
195759      KEN    Africa    Kenya  2021-03-17     112805.0        0.0
195766      KEN    Africa    Kenya  2021-03-24     120163.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
195746             236.429        1854.0         0.0                5.286
195752             387.714        1874.0         0.0                2.857
195756             634.714        1908.0        34.0                4.857
195759             634.714        1908.0         0.0                4.857
195766            1051.143        1994.0         0.0               12.286

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
195746  …          20.4                  24.651                         1.4
195752  …          20.4                  24.651                         1.4
195756  …          20.4                  24.651                         1.4
195759  …          20.4                  24.651                         1.4
195766  …          20.4                  24.651                         1.4

        life_expectancy  human_development_index  population  \
195746             66.7                    0.601    54027484
195752             66.7                    0.601    54027484
195756             66.7                    0.601    54027484
195759             66.7                    0.601    54027484
195766             66.7                    0.601    54027484

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
195746                                   NaN                          NaN
195752                                   NaN                          NaN
195756                                   NaN                          NaN
195759                                   NaN                          NaN
195766                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
195746               NaN                                      NaN
195752               NaN                                      NaN
195756               NaN                                      NaN
195759               NaN                                      NaN
195766               NaN                                      NaN

[5 rows x 67 columns]
```

[20]:

```
# Filter the DataFrame for the three East African countries
ea_countries = ['Kenya', 'Uganda', 'Tanzania']
df_ea = df[df['location'].isin(ea_countries)]
```

```python
# Generate the statistical summary
summary = df_ea[['location', 'total_cases', 'total_deaths']].
 ↪groupby('location').describe()

# Display the summary
print(summary)
```

```
         total_cases                                                        \
               count           mean            std  min       25%       50%
location
Kenya         1674.0  237047.882915  131389.050269  0.0  103993.0  323609.0
Tanzania      1674.0   25383.257467   18860.738776  0.0     509.0   33864.0
Uganda        1674.0  114375.297491   69890.469285  0.0   40213.0  163994.0

                            total_deaths                                \
               75%       max        count         mean          std  min
location
Kenya     343312.0  344106.0       1674.0  4079.434886  2202.393488  0.0
Tanzania   43078.0   43230.0       1674.0   541.675627   384.370224  0.0
Uganda    171324.0  172154.0       1674.0  2422.534648  1593.157437  0.0


             25%     50%     75%     max
location
Kenya     1817.0  5649.0  5688.0  5689.0
Tanzania    21.0   803.0   846.0   846.0
Uganda     330.0  3597.0  3632.0  3632.0
```

[21]:

```python
# Define key columns to check for missing values
critical_columns = ["date", "total_cases", "total_deaths", "new_cases",␣
 ↪"new_deaths", "total_vaccinations"]

# Drop rows with missing values in any of these columns
kenya_df.dropna(subset=critical_columns, inplace=True)

# Display cleaned dataset
print(kenya_df.head())
```

```
       iso_code continent location        date  total_cases  new_cases  \
195746        KEN    Africa    Kenya  2021-03-04     105648.0        0.0
195752        KEN    Africa    Kenya  2021-03-10     108362.0        0.0
195756        KEN    Africa    Kenya  2021-03-14     112805.0     4443.0
195759        KEN    Africa    Kenya  2021-03-17     112805.0        0.0
195766        KEN    Africa    Kenya  2021-03-24     120163.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
195746             236.429        1854.0         0.0                5.286
195752             387.714        1874.0         0.0                2.857
195756             634.714        1908.0        34.0                4.857
195759             634.714        1908.0         0.0                4.857
195766            1051.143        1994.0         0.0               12.286

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
195746  …          20.4                  24.651                         1.4
195752  …          20.4                  24.651                         1.4
195756  …          20.4                  24.651                         1.4
195759  …          20.4                  24.651                         1.4
195766  …          20.4                  24.651                         1.4

        life_expectancy  human_development_index  population  \
195746             66.7                    0.601    54027484
195752             66.7                    0.601    54027484
195756             66.7                    0.601    54027484
195759             66.7                    0.601    54027484
195766             66.7                    0.601    54027484

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
195746                                   NaN                          NaN
195752                                   NaN                          NaN
195756                                   NaN                          NaN
195759                                   NaN                          NaN
195766                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
195746               NaN                                      NaN
195752               NaN                                      NaN
195756               NaN                                      NaN
195759               NaN                                      NaN
195766               NaN                                      NaN

[5 rows x 67 columns]
```

[22]:

```python
# Convert the 'date' column to datetime format
kenya_df["date"] = pd.to_datetime(kenya_df["date"])

# Display first few rows to confirm changes
print(kenya_df.head())
print(kenya_df.dtypes)  # Verify 'date' is now datetime
```

```
       iso_code continent location       date  total_cases  new_cases  \
195746      KEN    Africa    Kenya 2021-03-04     105648.0        0.0
195752      KEN    Africa    Kenya 2021-03-10     108362.0        0.0
195756      KEN    Africa    Kenya 2021-03-14     112805.0     4443.0
195759      KEN    Africa    Kenya 2021-03-17     112805.0        0.0
```

```
195766      KEN    Africa    Kenya 2021-03-24     120163.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
195746             236.429        1854.0         0.0                5.286
195752             387.714        1874.0         0.0                2.857
195756             634.714        1908.0        34.0                4.857
195759             634.714        1908.0         0.0                4.857
195766            1051.143        1994.0         0.0               12.286

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
195746  …          20.4                  24.651                         1.4
195752  …          20.4                  24.651                         1.4
195756  …          20.4                  24.651                         1.4
195759  …          20.4                  24.651                         1.4
195766  …          20.4                  24.651                         1.4

        life_expectancy  human_development_index  population  \
195746             66.7                    0.601    54027484
195752             66.7                    0.601    54027484
195756             66.7                    0.601    54027484
195759             66.7                    0.601    54027484
195766             66.7                    0.601    54027484

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
195746                                   NaN                          NaN
195752                                   NaN                          NaN
195756                                   NaN                          NaN
195759                                   NaN                          NaN
195766                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
195746               NaN                                      NaN
195752               NaN                                      NaN
195756               NaN                                      NaN
195759               NaN                                      NaN
195766               NaN                                      NaN

[5 rows x 67 columns]
iso_code                                           object
continent                                          object
location                                           object
date                                       datetime64[ns]
total_cases                                       float64
                                                …
population                                          int64
excess_mortality_cumulative_absolute              float64
excess_mortality_cumulative                       float64
excess_mortality                                  float64
```

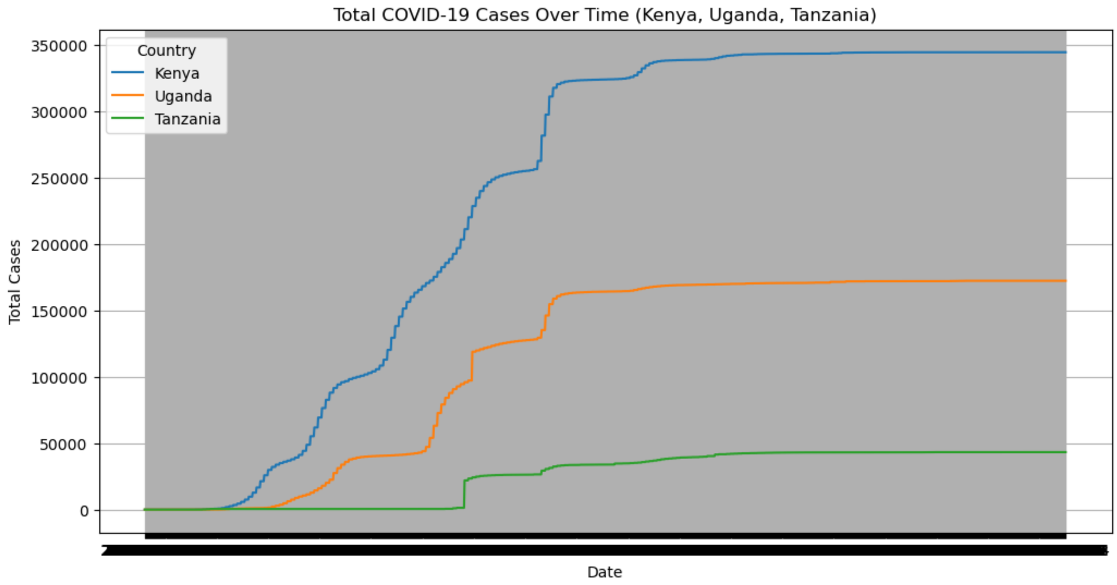
```
excess_mortality_cumulative_per_million          float64
Length: 67, dtype: object
```

[23]:

```
kenya_df.interpolate(method="linear", inplace=True)
```

[26]:

```
# Filter for Kenya, Uganda, and Tanzania
import matplotlib.pyplot as plt

countries = ["Kenya", "Uganda", "Tanzania"]
df_filtered = df[df["location"].isin(countries)]

# Plot total cases over time
plt.figure(figsize=(12, 6))
for country in countries:
    country_df = df_filtered[df_filtered["location"] == country]
    plt.plot(country_df["date"], country_df["total_cases"], label=country)

# Customize the plot
plt.title("Total COVID-19 Cases Over Time (Kenya, Uganda, Tanzania)")
plt.xlabel("Date")
plt.ylabel("Total Cases")
plt.legend(title="Country")
plt.grid(True)

# Show the plot
plt.show()
```



2 This graph illustrates total COVID-19 cases over time for Kenya, Uganda, and Tanzania. Here are some key insights :

1. Uganda's Sharp Increase: Uganda shows a significant surge in total cases at certain points, possibly due to mass outbreaks, increased testing, or delayed reporting.
2. Kenya's Consistent Growth: Kenya displays a steady rise in total cases, suggesting ongoing transmission but without extreme spikes. This could reflect continuous community spread with stable case identification.
3. Tanzania's Lower Case Numbers: Tanzania reports fewer total cases than Kenya and Uganda, which might be due to differences in testing strategies, reporting policies, or government interventions.

Overall, Uganda seems to have experienced a major outbreak or policy shift leading to high case numbers, while Kenya follows a more predictable trajectory. Tanzania's lower numbers warrant a closer look into how data collection and public health responses shaped its trends.

[ ]:

```
#plot total deaths over time
plt.figure(figsize = (12,6))
for country in countries:
    country_df = df_filtered[df_filtered['location'] == country]
    plt.plot(country_df["date"], country_df["total_deaths"], label=country)

# Customize the plot
plt.title("Total COVID-19 Deaths Over Time (Kenya, Uganda, Tanzania)")
plt.xlabel("Date")
plt.ylabel("Total Deaths")
plt.legend(title="Country")
plt.grid(True)

# Show the plot
plt.show()
```

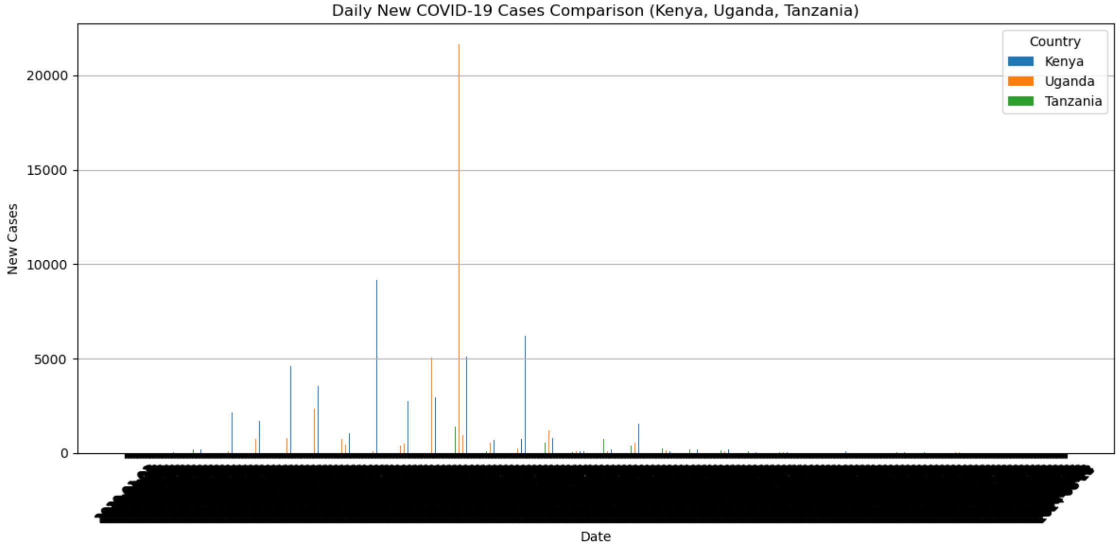
3 This graph presents total COVID-19 deaths over time for Kenya, Uganda, and Tanzania. Here are some insights based on the trends:

1. Uganda's Sudden Increase: Uganda shows a sharp rise in total deaths at certain points, potentially corresponding to significant outbreaks or delayed reporting of fatalities. This suggests periods of high transmission and severe impact on public health.
2. Kenya's Gradual Growth: Kenya exhibits a steady increase in total deaths, indicating ongoing transmission with a sustained but moderate death toll. This might reflect effective healthcare interventions reducing extreme spikes.
3. Tanzania's Lower Numbers: Compared to Kenya and Uganda, Tanzania maintains the lowest reported death toll. However, this could be influenced by differences in testing, reporting policies, or approaches to managing COVID-19.

These trends highlight key differences in how each country experienced and reported fatalities. The spikes in Uganda's data might warrant further investigation into causes, such as overwhelmed healthcare systems or a change in reporting practices.

DAILY NEW COVID 19 CASES COMPARISON

[51]:
```
import numpy as np
plt.figure(figsize=(12, 6))

df_filtered["date"] = pd.to_datetime(df_filtered["date"])
df_filtered = df_filtered.sort_values(by="date")
dates = df_filtered["date"].unique()
bar_width = 0.25
x = np.arange(len(dates))

for i, country in enumerate(countries):
    country_df = df_filtered[df_filtered["location"] == country]
    country_df = country_df.set_index("date").reindex(dates).fillna(0).
  ↪reset_index()
    plt.bar(x + i * bar_width, country_df["new_cases"], width=bar_width,␣
  ↪label=country)

plt.xticks(x + bar_width, [date.strftime('%Y-%m-%d') for date in dates],␣
  ↪rotation=45)
plt.title("Daily New COVID-19 Cases Comparison (Kenya, Uganda, Tanzania)")
plt.xlabel("Date")
plt.ylabel("New Cases")
plt.legend(title="Country")
plt.grid(True, axis='y')
plt.tight_layout()
plt.show()
```



4 key insights

Looking at this graph comparing daily new COVID-19 cases in Kenya, Uganda, and Tanzania, some interesting insights emerge:

1. Uganda's Sudden Surge: Uganda shows a dramatic spike in new cases compared to Kenya and Tanzania. This could indicate a testing anomaly, a super-spreader event, or a change in reporting methods.
2. Kenya and Tanzania's Stability: While both Kenya and Tanzania appear to have fluctuating case numbers, neither displays the sudden spike Uganda does. This could suggest more stable transmission rates or differences in testing capacity.
3. Regional Spread Considerations: If Uganda's spike is real and not due to reporting inconsistencies, it might mean a major outbreak occurred, potentially affecting neighboring countries.

Policy Implications: Authorities in all three countries might need to investigate trends closely—Uganda could require more interventions, while Kenya and Tanzania might maintain their existing strategies

ANALYSIS OF THE THREE COUNTRIES AS COMPARED TO THE WORLD

[33]:

```
countries = ["Kenya", "Uganda", "Tanzania"]
df_filtered = df[df["location"].isin(countries)].copy()  # This ensures␣
  ↪df_filtered is defined

# Calculate death rate
df_filtered["death_rate"] = df_filtered["total_deaths"] /␣
  ↪df_filtered["total_cases"]
print(df_filtered.pivot_table(index="date", columns="location",␣
  ↪values="death_rate"))
```

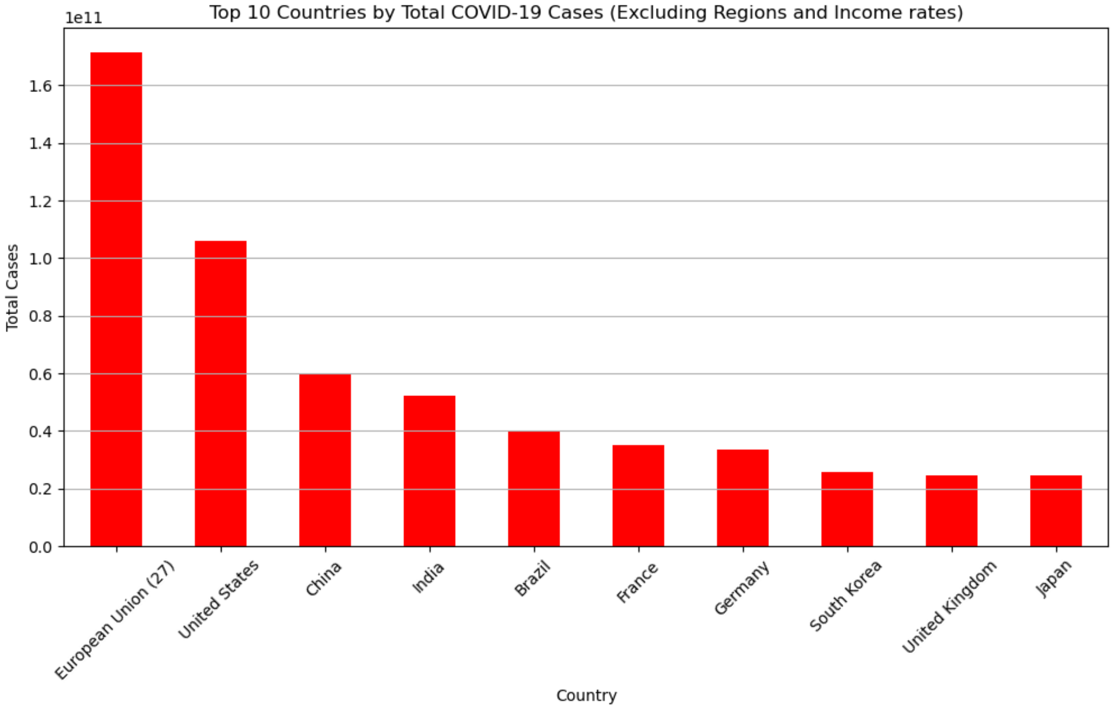
```
location       Kenya  Tanzania    Uganda
date
2020-03-15  0.000000       NaN       NaN
2020-03-16  0.000000       NaN       NaN
2020-03-17  0.000000       NaN       NaN
2020-03-18  0.000000       NaN       NaN
2020-03-19  0.000000       NaN       NaN
…                …       …       …
2024-07-31  0.016533   0.01957  0.021097
2024-08-01  0.016533   0.01957  0.021097
2024-08-02  0.016533   0.01957  0.021097
2024-08-03  0.016533   0.01957  0.021097
2024-08-04  0.016533   0.01957  0.021097

[1604 rows x 3 columns]
```

[50]:

```
# List of regions and income categories to exclude
regions_to_exclude = ["World", "European Union", "Asia", "Europe", "Africa",␣
 ↪"North America", "South America",
                      'High income', 'Upper middle income', 'Lower middle␣
 ↪income']


df_filtered = df[~df["location"].str.strip().isin(regions_to_exclude)]


print("Unique values in 'location' after exclusion:")
print(df_filtered["location"].unique())  # To see what is left in the␣
 ↪'location' column

)
df_filtered = df_filtered[~df_filtered["location"].str.
 ↪contains("income|region", case=False, na=False)]


print("Unique values in 'location' after further filtering:")
print(df_filtered["location"].unique())

#  top 10 countries by total cases
top_countries = df_filtered.groupby("location")["total_cases"].sum().
 ↪nlargest(10)

# Plot bar chart
plt.figure(figsize=(12, 6))
top_countries.plot(kind="bar", color="red")

# Customize plot
plt.title("Top 10 Countries by Total COVID-19 Cases (Excluding Regions and␣
 ↪Income rates)")
plt.xlabel("Country")
plt.ylabel("Total Cases")
plt.xticks(rotation=45)
plt.grid(axis="y")

# Show plot
plt.show()
```

```
Unique values in 'location' after exclusion:
['Afghanistan' 'Albania' 'Algeria' 'American Samoa' 'Andorra' 'Angola'
 'Anguilla' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba'
 'Australia' 'Austria' 'Azerbaijan' 'Bahamas' 'Bahrain' 'Bangladesh'
 'Barbados' 'Belarus' 'Belgium' 'Belize' 'Benin' 'Bermuda' 'Bhutan'
 'Bolivia' 'Bonaire Sint Eustatius and Saba' 'Bosnia and Herzegovina'
 'Botswana' 'Brazil' 'British Virgin Islands' 'Brunei' 'Bulgaria'
 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Canada' 'Cape Verde'
 'Cayman Islands' 'Central African Republic' 'Chad' 'Chile' 'China'
 'Colombia' 'Comoros' 'Congo' 'Cook Islands' 'Costa Rica' "Cote d'Ivoire"
 'Croatia' 'Cuba' 'Curacao' 'Cyprus' 'Czechia'
 'Democratic Republic of Congo' 'Denmark' 'Djibouti' 'Dominica'
 'Dominican Republic' 'East Timor' 'Ecuador' 'Egypt' 'El Salvador'
 'England' 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini' 'Ethiopia'
 'European Union (27)' 'Faroe Islands' 'Falkland Islands' 'Fiji' 'Finland'
 'France' 'French Guiana' 'French Polynesia' 'Gabon' 'Gambia' 'Georgia'
 'Germany' 'Ghana' 'Gibraltar' 'Greece' 'Greenland' 'Grenada' 'Guadeloupe'
 'Guam' 'Guatemala' 'Guernsey' 'Guinea' 'Guinea-Bissau' 'Guyana' 'Haiti'
 'High-income countries' 'Honduras' 'Hong Kong' 'Hungary' 'Iceland'
 'India' 'Indonesia' 'Iran' 'Iraq' 'Ireland' 'Isle of Man' 'Israel'
 'Italy' 'Jamaica' 'Japan' 'Jersey' 'Jordan' 'Kazakhstan' 'Kenya'
 'Kiribati' 'Kosovo' 'Kuwait' 'Kyrgyzstan' 'Laos' 'Latvia' 'Lebanon'
 'Lesotho' 'Liberia' 'Libya' 'Liechtenstein' 'Lithuania'
 'Low-income countries' 'Lower-middle-income countries' 'Luxembourg'
 'Macao' 'Madagascar' 'Malawi' 'Malaysia' 'Maldives' 'Mali' 'Malta'
 'Marshall Islands' 'Martinique' 'Mauritania' 'Mauritius' 'Mayotte'
 'Mexico' 'Micronesia (country)' 'Moldova' 'Monaco' 'Mongolia'
 'Montenegro' 'Montserrat' 'Morocco' 'Mozambique' 'Myanmar' 'Namibia'
 'Nauru' 'Nepal' 'Netherlands' 'New Caledonia' 'New Zealand' 'Nicaragua'
 'Niger' 'Nigeria' 'Niue' 'North Korea' 'North Macedonia'
 'Northern Cyprus' 'Northern Ireland' 'Northern Mariana Islands' 'Norway'
 'Oceania' 'Oman' 'Pakistan' 'Palau' 'Palestine' 'Panama'
 'Papua New Guinea' 'Paraguay' 'Peru' 'Philippines' 'Pitcairn' 'Poland'
 'Portugal' 'Puerto Rico' 'Qatar' 'Reunion' 'Romania' 'Russia' 'Rwanda'
 'Saint Barthelemy' 'Saint Helena' 'Saint Kitts and Nevis' 'Saint Lucia'
 'Saint Martin (French part)' 'Saint Pierre and Miquelon'
 'Saint Vincent and the Grenadines' 'Samoa' 'San Marino'
 'Sao Tome and Principe' 'Saudi Arabia' 'Scotland' 'Senegal' 'Serbia'
 'Seychelles' 'Sierra Leone' 'Singapore' 'Sint Maarten (Dutch part)'
 'Slovakia' 'Slovenia' 'Solomon Islands' 'Somalia' 'South Africa'
 'South Korea' 'South Sudan' 'Spain' 'Sri Lanka' 'Sudan' 'Suriname'
 'Sweden' 'Switzerland' 'Syria' 'Taiwan' 'Tajikistan' 'Tanzania'
 'Thailand' 'Togo' 'Tokelau' 'Tonga' 'Trinidad and Tobago' 'Tunisia'
 'Turkey' 'Turkmenistan' 'Turks and Caicos Islands' 'Tuvalu' 'Uganda'
 'Ukraine' 'United Arab Emirates' 'United Kingdom' 'United States'
 'United States Virgin Islands' 'Upper-middle-income countries' 'Uruguay'
 'Uzbekistan' 'Vanuatu' 'Vatican' 'Venezuela' 'Vietnam' 'Wales'
 'Wallis and Futuna' 'Western Sahara' 'Yemen' 'Zambia' 'Zimbabwe']
Unique values in 'location' after further filtering:
['Afghanistan' 'Albania' 'Algeria' 'American Samoa' 'Andorra' 'Angola'
 'Anguilla' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba'
 'Australia' 'Austria' 'Azerbaijan' 'Bahamas' 'Bahrain' 'Bangladesh'
 'Barbados' 'Belarus' 'Belgium' 'Belize' 'Benin' 'Bermuda' 'Bhutan'
 'Bolivia' 'Bonaire Sint Eustatius and Saba' 'Bosnia and Herzegovina'
```

```
'Botswana' 'Brazil' 'British Virgin Islands' 'Brunei' 'Bulgaria'
'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Canada' 'Cape Verde'
'Cayman Islands' 'Central African Republic' 'Chad' 'Chile' 'China'
'Colombia' 'Comoros' 'Congo' 'Cook Islands' 'Costa Rica' "Cote d'Ivoire"
'Croatia' 'Cuba' 'Curacao' 'Cyprus' 'Czechia'
'Democratic Republic of Congo' 'Denmark' 'Djibouti' 'Dominica'
'Dominican Republic' 'East Timor' 'Ecuador' 'Egypt' 'El Salvador'
'England' 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini' 'Ethiopia'
'European Union (27)' 'Faroe Islands' 'Falkland Islands' 'Fiji' 'Finland'
'France' 'French Guiana' 'French Polynesia' 'Gabon' 'Gambia' 'Georgia'
'Germany' 'Ghana' 'Gibraltar' 'Greece' 'Greenland' 'Grenada' 'Guadeloupe'
'Guam' 'Guatemala' 'Guernsey' 'Guinea' 'Guinea-Bissau' 'Guyana' 'Haiti'
'Honduras' 'Hong Kong' 'Hungary' 'Iceland' 'India' 'Indonesia' 'Iran'
'Iraq' 'Ireland' 'Isle of Man' 'Israel' 'Italy' 'Jamaica' 'Japan'
'Jersey' 'Jordan' 'Kazakhstan' 'Kenya' 'Kiribati' 'Kosovo' 'Kuwait'
'Kyrgyzstan' 'Laos' 'Latvia' 'Lebanon' 'Lesotho' 'Liberia' 'Libya'
'Liechtenstein' 'Lithuania' 'Luxembourg' 'Macao' 'Madagascar' 'Malawi'
'Malaysia' 'Maldives' 'Mali' 'Malta' 'Marshall Islands' 'Martinique'
'Mauritania' 'Mauritius' 'Mayotte' 'Mexico' 'Micronesia (country)'
'Moldova' 'Monaco' 'Mongolia' 'Montenegro' 'Montserrat' 'Morocco'
'Mozambique' 'Myanmar' 'Namibia' 'Nauru' 'Nepal' 'Netherlands'
'New Caledonia' 'New Zealand' 'Nicaragua' 'Niger' 'Nigeria' 'Niue'
'North Korea' 'North Macedonia' 'Northern Cyprus' 'Northern Ireland'
'Northern Mariana Islands' 'Norway' 'Oceania' 'Oman' 'Pakistan' 'Palau'
'Palestine' 'Panama' 'Papua New Guinea' 'Paraguay' 'Peru' 'Philippines'
'Pitcairn' 'Poland' 'Portugal' 'Puerto Rico' 'Qatar' 'Reunion' 'Romania'
'Russia' 'Rwanda' 'Saint Barthelemy' 'Saint Helena'
'Saint Kitts and Nevis' 'Saint Lucia' 'Saint Martin (French part)'
'Saint Pierre and Miquelon' 'Saint Vincent and the Grenadines' 'Samoa'
'San Marino' 'Sao Tome and Principe' 'Saudi Arabia' 'Scotland' 'Senegal'
'Serbia' 'Seychelles' 'Sierra Leone' 'Singapore'
'Sint Maarten (Dutch part)' 'Slovakia' 'Slovenia' 'Solomon Islands'
'Somalia' 'South Africa' 'South Korea' 'South Sudan' 'Spain' 'Sri Lanka'
'Sudan' 'Suriname' 'Sweden' 'Switzerland' 'Syria' 'Taiwan' 'Tajikistan'
'Tanzania' 'Thailand' 'Togo' 'Tokelau' 'Tonga' 'Trinidad and Tobago'
'Tunisia' 'Turkey' 'Turkmenistan' 'Turks and Caicos Islands' 'Tuvalu'
'Uganda' 'Ukraine' 'United Arab Emirates' 'United Kingdom'
'United States' 'United States Virgin Islands' 'Uruguay' 'Uzbekistan'
'Vanuatu' 'Vatican' 'Venezuela' 'Vietnam' 'Wales' 'Wallis and Futuna'
'Western Sahara' 'Yemen' 'Zambia' 'Zimbabwe']
```



5 TOP 10 COUNTRIES BY TOTAL CASES OF COVID 19

5.1 KEY INSIGHTS

Key Insights on Top COVID-19 Case Totals 1. European Union (27 countries) reported the highest cumulative COVID-19 cases when considered as a single entity, reflecting the aggregate impact across its member nations.

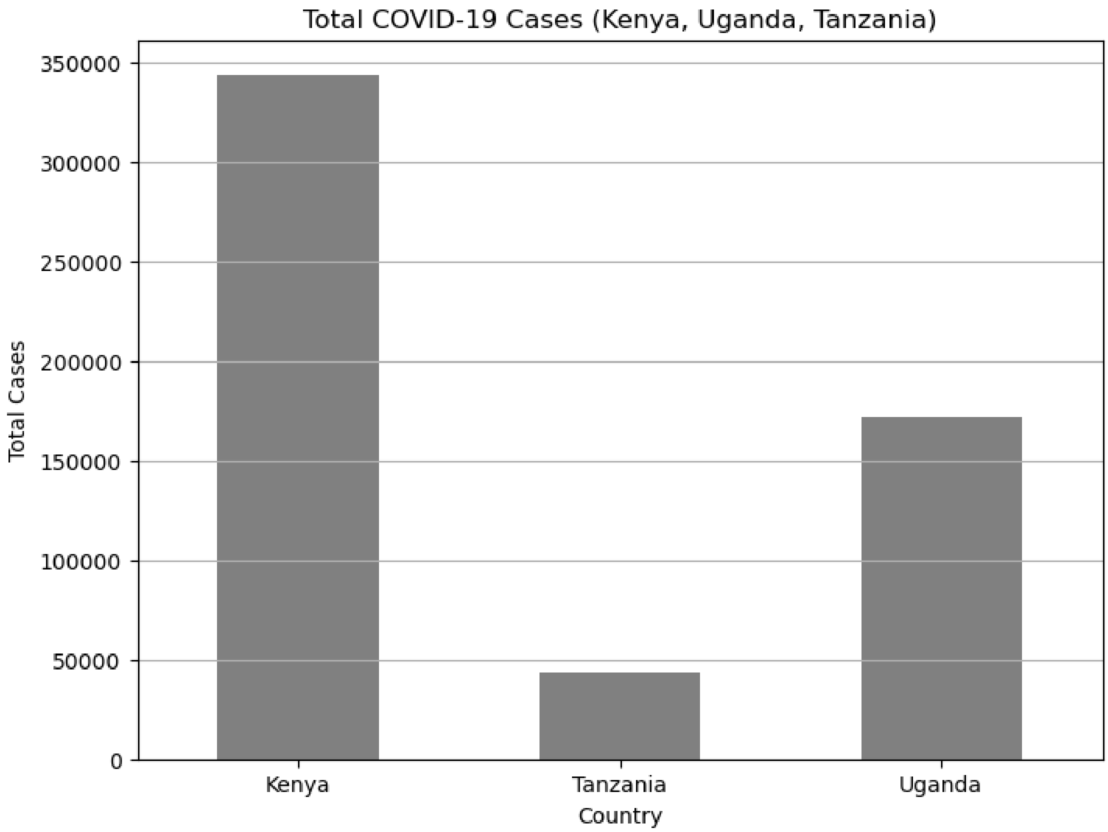
2. United States remains the single country with the highest number of total cases, underscoring its significant exposure throughout the pandemic despite extensive vaccination campaigns.
3. China, where the outbreak began, also recorded high cumulative numbers, likely reflecting both early spread and later waves influenced by population size and dense urban centers.
4. India shows extremely high case totals, driven by its large population and severe waves (such as the Delta variant surge), despite rapid scaling of healthcare responses.
5. Brazil had the highest number of cases in South America, attributed to delayed containment measures and large urban populations.
6. France and Germany represent the top two European countries with the highest national case counts, showing how Western Europe was heavily affected during multiple waves.
7. South Korea, initially successful in controlling early outbreaks, later experienced large waves likely due to newer variants and easing of restrictions.
8. United Kingdom had high totals due to early and widespread community transmission, but later became a leader in vaccine rollouts.

9. Japan maintained relatively low case numbers early in the pandemic but eventually saw surges, especially during and after the Tokyo Olympics and later variant waves.

TOTAL COVID-19 CASES (Kenya, Uganda, Tanzania)

[37]:

```
# Filter for Kenya, Uganda, and Tanzania
countries = ["Kenya", "Uganda", "Tanzania"]
df_filtered = df[df["location"].isin(countries)].copy()

# Select latest available data
df_latest = df_filtered.sort_values("date").groupby("location").last()

# Extract total cases
total_cases = df_latest["total_cases"]

# Plot bar chart
plt.figure(figsize=(8, 6))
total_cases.plot(kind="bar", color=["grey"])
# Customize plot
plt.title("Total COVID-19 Cases (Kenya, Uganda, Tanzania)")
plt.xlabel("Country")
plt.ylabel("Total Cases")
plt.xticks(rotation=0)
plt.grid(axis="y")

# Show plot
plt.show()
```



6 KEY INSIGHTS.

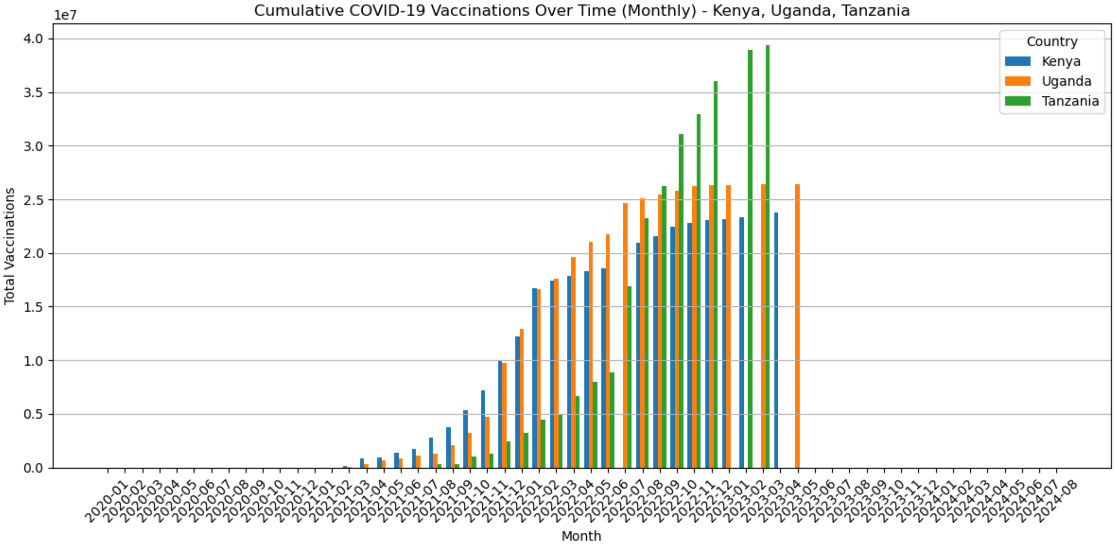
1.Kenya shows highest total covid 19 cases, reaching approximately 350000.This suggests more widespread transmission, higher testing rates, or better case reporting. 2.Uganda follows with around 150,000 cases, showing a significant impact but lower than Kenya.3. Tanzania reports the lowest case count, roughly 50,000 cases. This could be due to limited testing, underreporting, or different COVID-19 management strategies:

# CUMULATIVE COVID-19 VACCINATIONS OVER TIME-EAST AFRICA

[40]:

```
plt.figure(figsize=(12, 6))

months = df_grouped["month"].unique()
bar_width = 0.25
x = np.arange(len(months))

for i, country in enumerate(countries):
    country_df = df_grouped[df_grouped["location"] == country]
    country_df = country_df.set_index("month").reindex(months).fillna(0).
  ↪reset_index()
```

```
    plt.bar(x + i * bar_width, country_df["total_vaccinations"],␣
  ↪width=bar_width, label=country)

plt.xticks(x + bar_width, months, rotation=45)
plt.title("Cumulative COVID-19 Vaccinations Over Time (Monthly) - Kenya,␣
  ↪Uganda, Tanzania")
plt.xlabel("Month")
plt.ylabel("Total Vaccinations")
plt.legend(title="Country")
plt.grid(True, axis='y')
plt.tight_layout()
plt.show()
```



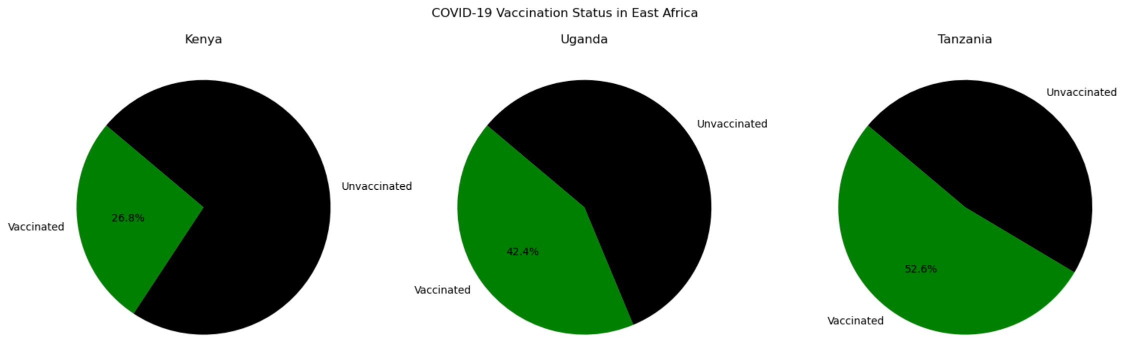
7 Key Insight:

in the above graph: Tanzania showed the most significant rise in total COVID-19 vaccinations over time, surpassing both Kenya and Uganda by a substantial margin. While Uganda initially led in vaccination rollout, Tanzania accelerated quickly and reached the highest cumulative vaccinations. Kenya maintained steady progress but ended with the lowest total vaccinations among the three countries. This suggests Tanzania's vaccination campaign became more aggressive and effective in later phases.

8 COVID-19 VACCINATION STATUS IN EAST AFRICA(PERCENTAGE AGAINST POPULATION)

[42]:

```
# Filter for the three countries
countries = ['Kenya', 'Uganda', 'Tanzania']
df_ea = df[df['location'].isin(countries)].dropna(subset=['people_vaccinated',␣
 ↪'population'])

# Ensure date column is datetime
df_ea['date'] = pd.to_datetime(df_ea['date'])

# Get the latest data per country
latest = df_ea.sort_values('date').groupby('location').last()

# Setup subplots
fig, axes = plt.subplots(1, 3, figsize=(15, 5))  # 1 row, 3 columns

colors = ['green', 'black']  # Green for vaccinated, red for unvaccinated

for i, country in enumerate(countries):
    vaccinated = latest.loc[country, 'people_vaccinated']
    population = latest.loc[country, 'population']
    unvaccinated = population - vaccinated

    sizes = [vaccinated, unvaccinated]
    labels = ['Vaccinated', 'Unvaccinated']

    axes[i].pie(sizes, labels=labels, colors=colors, autopct='%.1f%%',␣
 ↪startangle=140)
    axes[i].set_title(f'{country}')
    axes[i].axis('equal')  # Equal aspect ratio ensures pie is a circle

plt.suptitle('COVID-19 Vaccination Status in East Africa')
plt.tight_layout()
plt.show()
```



9 Key Insight:

Among the three East African countries analyzed:

Tanzania has the highest vaccination rate, with approximately 52.6% of its population vaccinated.

Uganda follows with a vaccination rate of around 42.4%.

Kenya has the lowest vaccination rate at 26.8%, meaning nearly three-quarters of its population remain unvaccinated.

This suggests that Tanzania has made more progress in vaccinating its population compared to its regional neighbors, while Kenya may need to intensify its vaccination campaigns to catch up.

[ ]: