In [45]:

```python
import pandas as pd

df = pd.read_csv('owid-covid-data.csv')
df['date'] = pd.to_datetime(df['date'])

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

```
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', 'new_tests', 'total_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', 'population_density', 'median_age', 'aged_65_older',
       'aged_70_older', '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',
       'excess_mortality_cumulative_absolute', 'excess_mortality_cumulative',
       'excess_mortality', 'excess_mortality_cumulative_per_million'],
      dtype='object')
  iso_code continent     location       date  total_cases  new_cases  \
0      AFG      Asia  Afghanistan 2020-02-24          5.0        5.0   
1      AFG      Asia  Afghanistan 2020-02-25          5.0        0.0   
2      AFG      Asia  Afghanistan 2020-02-26          5.0        0.0   
3      AFG      Asia  Afghanistan 2020-02-27          5.0        0.0   
4      AFG      Asia  Afghanistan 2020-02-28          5.0        0.0   

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

   female_smokers  male_smokers  handwashing_facilities  \
0             NaN           NaN                  37.746   
1             NaN           NaN                  37.746   
2             NaN           NaN                  37.746   
3             NaN           NaN                  37.746   
4             NaN           NaN                  37.746   

   hospital_beds_per_thousand  life_expectancy  human_development_index  \
0                         0.5            64.83                    0.511   
1                         0.5            64.83                    0.511   
2                         0.5            64.83                    0.511   
3                         0.5            64.83                    0.511   
4                         0.5            64.83                    0.511   

   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]
```

In [46]:

```
countries = ["Nigeria", "South Africa", "Kenya", "Uganda", "Tanzania", "Ghana",
df_filtered = df[df["location"].isin(countries)].copy()
df_filtered["new_cases_smoothed"] = df_filtered["new_cases_smoothed"].fillna(0)
df_filtered["total_deaths"] = df_filtered["total_deaths"].fillna(df_filtered["to


df_filtered.to_csv("cleaned_owid_covid_data.csv", index=False)

print(df_filtered.head())
```

```
      iso_code continent  location        date  total_cases  new_cases  \
26411      CMR    Africa  Cameroon 2020-03-06          1.0        1.0   
26412      CMR    Africa  Cameroon 2020-03-07          1.0        0.0   
26413      CMR    Africa  Cameroon 2020-03-08          2.0        1.0   
26414      CMR    Africa  Cameroon 2020-03-09          2.0        0.0   
26415      CMR    Africa  Cameroon 2020-03-10          2.0        0.0   

       new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  ...  \
26411                 0.0   5938.034418         NaN                  NaN  ...   
26412                 0.0   5938.034418         NaN                  NaN  ...   
26413                 0.0   5938.034418         NaN                  NaN  ...   
26414                 0.0   5938.034418         NaN                  NaN  ...   
26415                 0.0   5938.034418         NaN                  NaN  ...   

       female_smokers  male_smokers  handwashing_facilities  \
26411             NaN           NaN                   2.735   
26412             NaN           NaN                   2.735   
26413             NaN           NaN                   2.735   
26414             NaN           NaN                   2.735   
26415             NaN           NaN                   2.735   

       hospital_beds_per_thousand  life_expectancy  human_development_index  \
26411                         1.3            59.29                    0.563   
26412                         1.3            59.29                    0.563   
26413                         1.3            59.29                    0.563   
26414                         1.3            59.29                    0.563   
26415                         1.3            59.29                    0.563   

       excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
26411                                   NaN                          NaN   
26412                                   NaN                          NaN   
26413                                   NaN                          NaN   
26414                                   NaN                          NaN   
26415                                   NaN                          NaN   

       excess_mortality  excess_mortality_cumulative_per_million  
26411               NaN                                      NaN  
26412               NaN                                      NaN  
26413               NaN                                      NaN  
26414               NaN                                      NaN  
26415               NaN                                      NaN  

[5 rows x 67 columns]
```

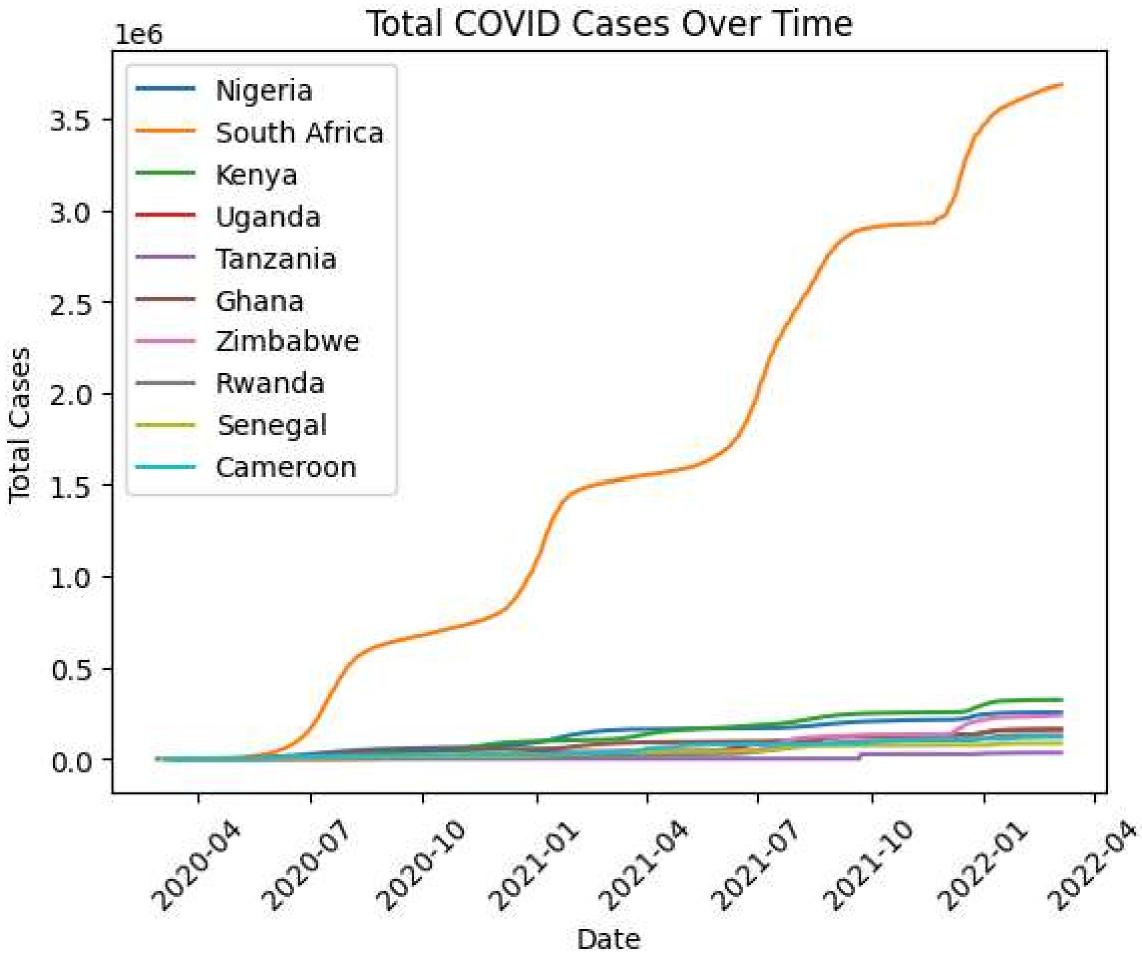
In [47]:

```
cleaned_df = pd.read_csv("cleaned_owid_covid_data.csv")
print(cleaned_df.head())
```

```
  iso_code continent  location        date  total_cases  new_cases  \
0      CMR    Africa  Cameroon  2020-03-06          1.0        1.0   
1      CMR    Africa  Cameroon  2020-03-07          1.0        0.0   
2      CMR    Africa  Cameroon  2020-03-08          2.0        1.0   
3      CMR    Africa  Cameroon  2020-03-09          2.0        0.0   
4      CMR    Africa  Cameroon  2020-03-10          2.0        0.0   

   new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  ...  \
0                 0.0   5938.034418         NaN                  NaN  ...   
1                 0.0   5938.034418         NaN                  NaN  ...   
2                 0.0   5938.034418         NaN                  NaN  ...   
3                 0.0   5938.034418         NaN                  NaN  ...   
4                 0.0   5938.034418         NaN                  NaN  ...   

   female_smokers  male_smokers  handwashing_facilities  \
0             NaN           NaN                   2.735   
1             NaN           NaN                   2.735   
2             NaN           NaN                   2.735   
3             NaN           NaN                   2.735   
4             NaN           NaN                   2.735   

   hospital_beds_per_thousand  life_expectancy  human_development_index  \
0                         1.3            59.29                    0.563   
1                         1.3            59.29                    0.563   
2                         1.3            59.29                    0.563   
3                         1.3            59.29                    0.563   
4                         1.3            59.29                    0.563   

   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]
```

In [48]:

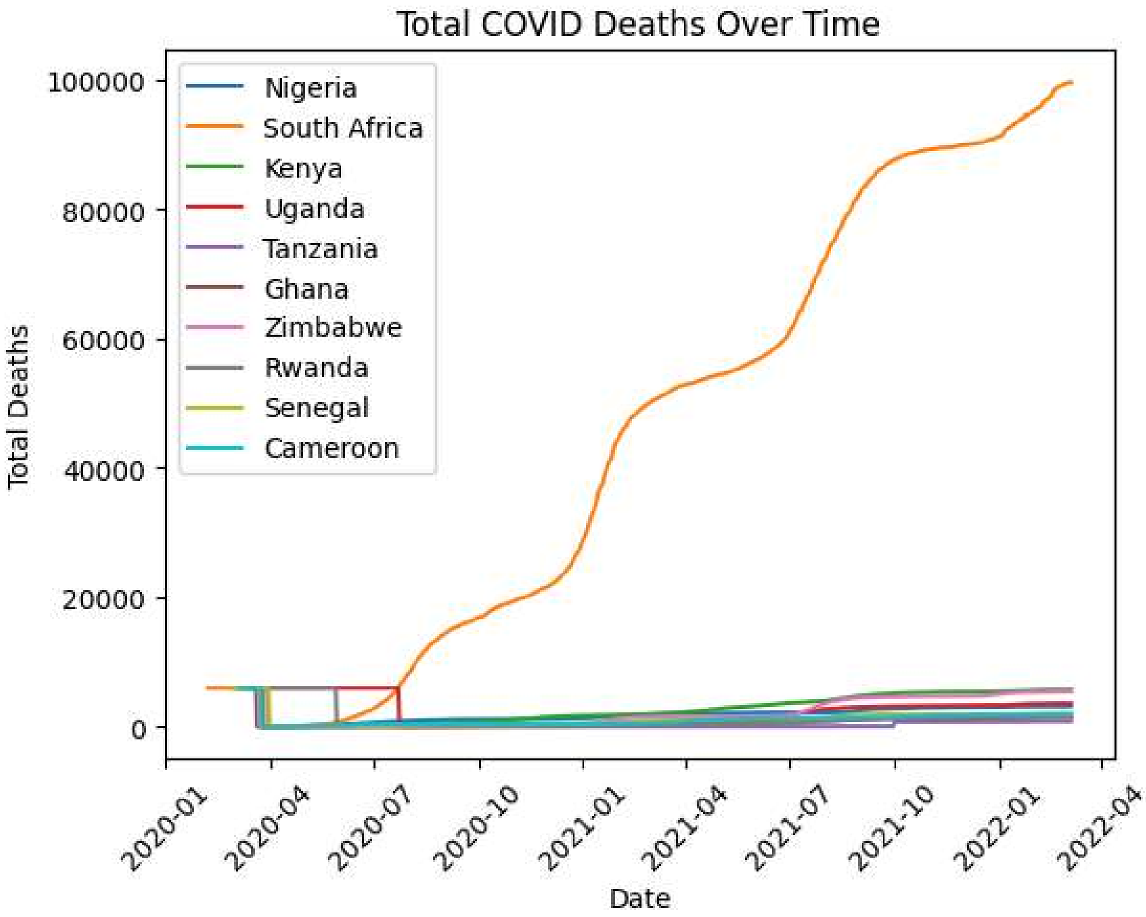
```python
import matplotlib.pyplot as plt

# Group by date and country
for country in countries:
    subset = df_filtered[df_filtered["location"] == country]
    plt.plot(subset["date"], subset["total_cases"], label=country)

plt.xlabel("Date")
plt.ylabel("Total Cases")
plt.title("Total COVID Cases Over Time")
plt.legend()
plt.xticks(rotation=45)
plt.show()
```



## Total COVID cases over time

- South Africa has the highest rate of increase in COVID cases over the period of 2020 - 2024
- South Africa also has the highest nomber of recorded cases(over 3.5million)

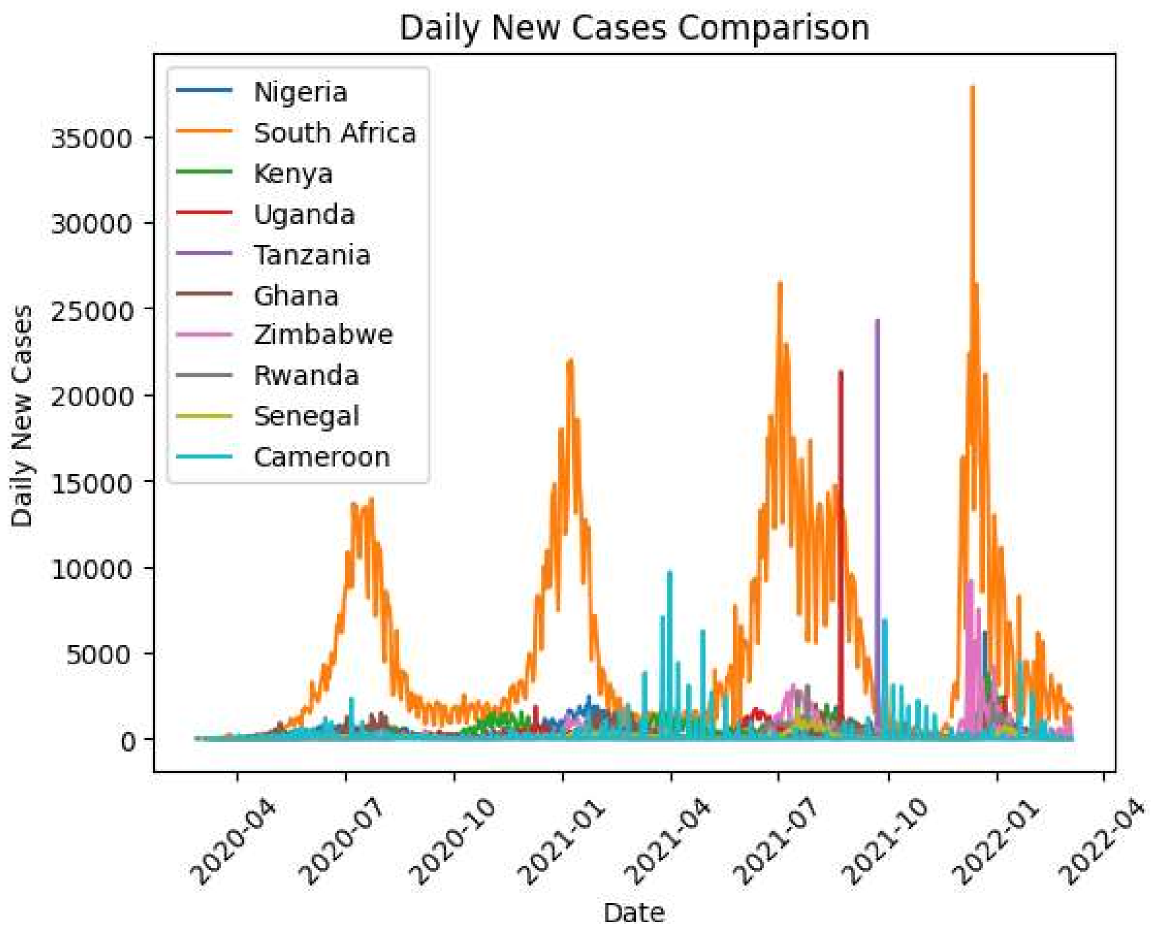
In [49]:
```
for country in countries:
    subset = df_filtered[df_filtered["location"] == country]
    plt.plot(subset["date"], subset["total_deaths"], label=country)

plt.xlabel("Date")
plt.ylabel("Total Deaths")
plt.title("Total COVID Deaths Over Time")
plt.legend()
plt.xticks(rotation=45)
plt.show()
```



# Total COVID deaths over time

- South Africa experienced an increase in deaths over the period of 2020 - 2024 with total deaths rising to 100,000 in 2024.
- Nigeria's total deaths at 2024 was less than 3500

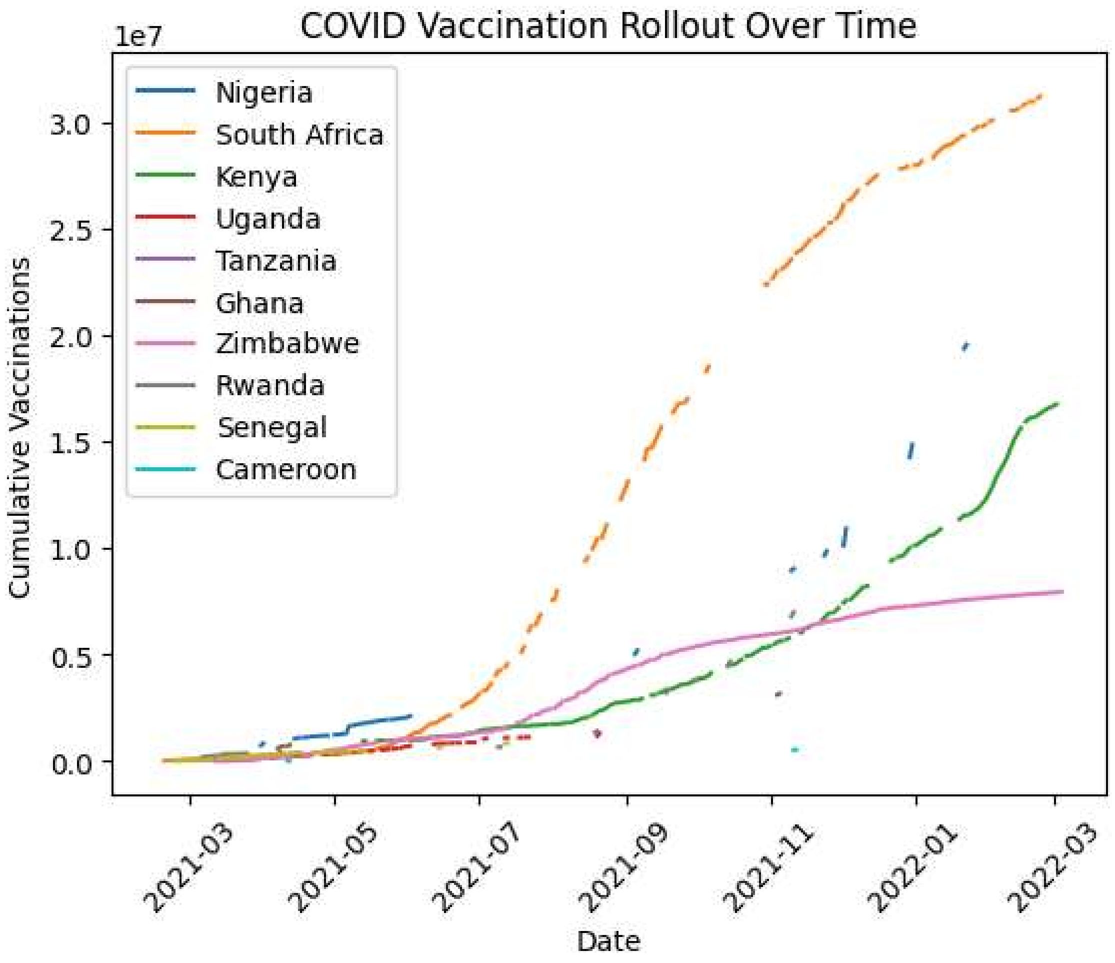
In [50]:
```
for country in countries:
    subset = df_filtered[df_filtered["location"] == country]
    plt.plot(subset["date"], subset["new_cases"], label=country)

plt.xlabel("Date")
plt.ylabel("Daily New Cases")
plt.title("Daily New Cases Comparison")
plt.legend()
plt.xticks(rotation=45)
plt.show()
```



In [51]:
```
df_filtered["death_rate"] = df_filtered["total_deaths"] / df_filtered["total_cas
print(df_filtered[["location", "date", "death_rate"]].head())
```

```
       location        date   death_rate
26411  Cameroon  2020-03-06  5938.034418
26412  Cameroon  2020-03-07  5938.034418
26413  Cameroon  2020-03-08  2969.017209
26414  Cameroon  2020-03-09  2969.017209
26415  Cameroon  2020-03-10  2969.017209
```

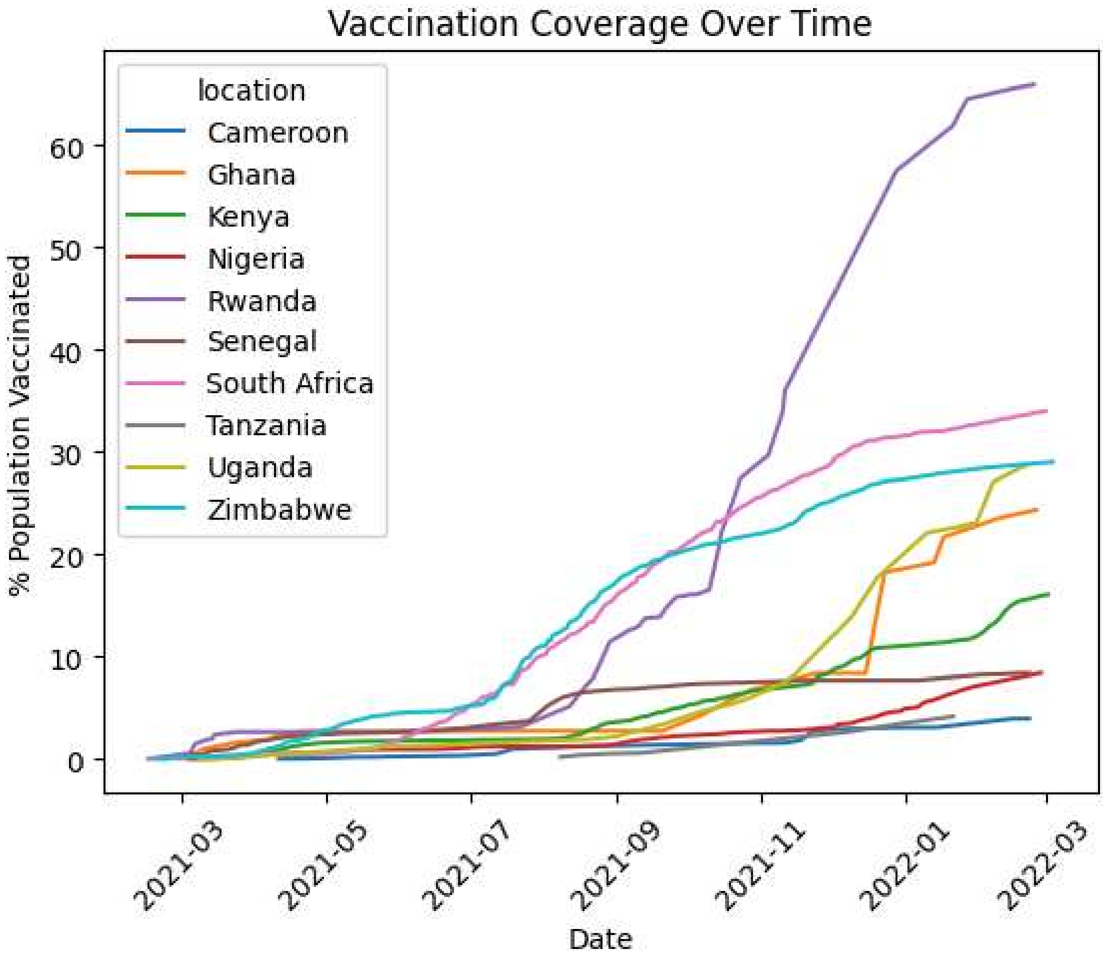
In [52]:
```
# Plot cumulative vaccinations over time
for country in countries:
    subset = df_filtered[df_filtered["location"] == country]
    plt.plot(subset["date"], subset["total_vaccinations"], label=country)

plt.xlabel("Date")
plt.ylabel("Cumulative Vaccinations")
plt.title("COVID Vaccination Rollout Over Time")
plt.legend()
plt.xticks(rotation=45)
plt.show()
```



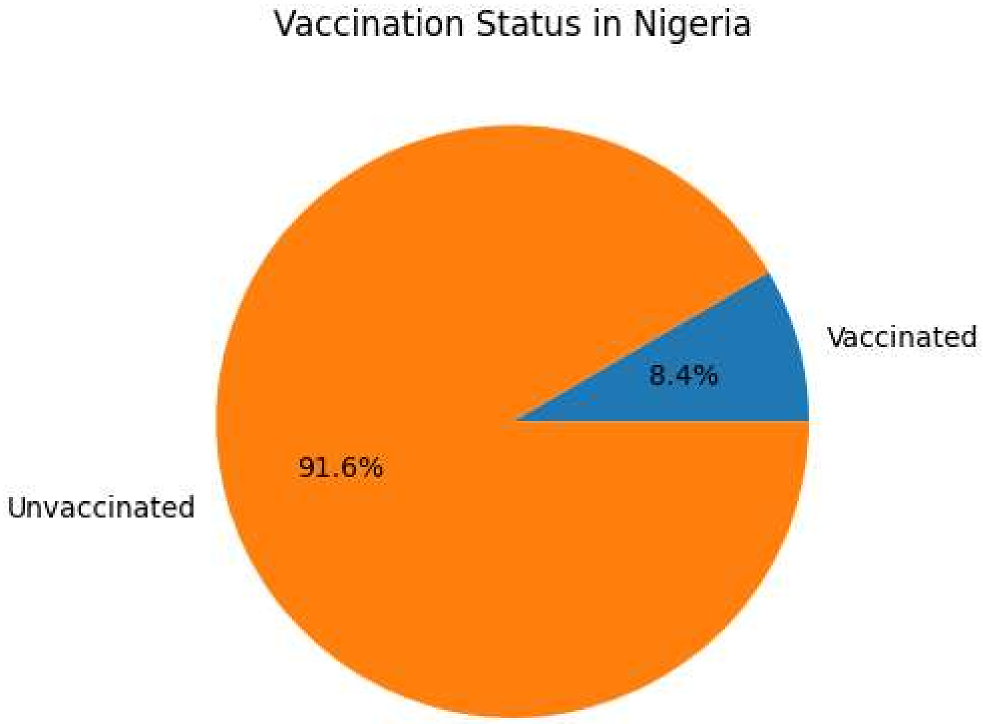
## COVID vaccination roll-out over time

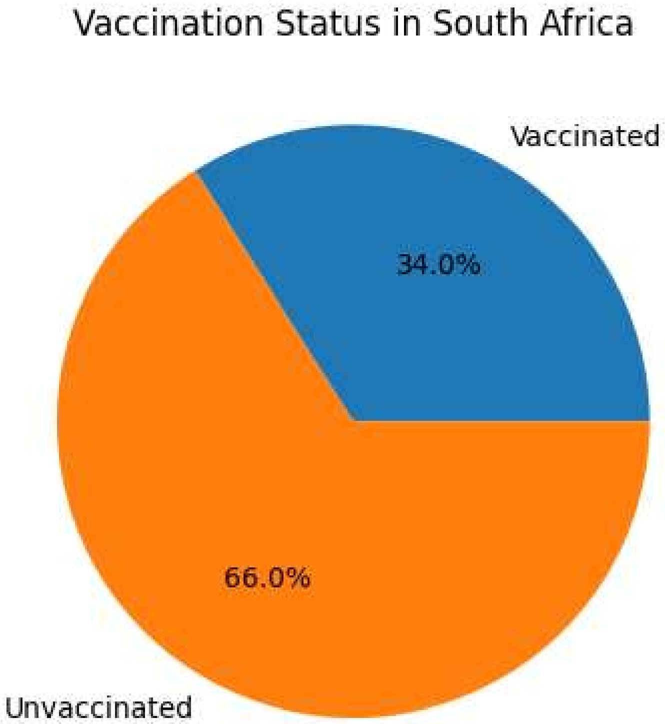
- South Africa has the highest vaccination roll out by 2024, yet it still has the higest number of deaths

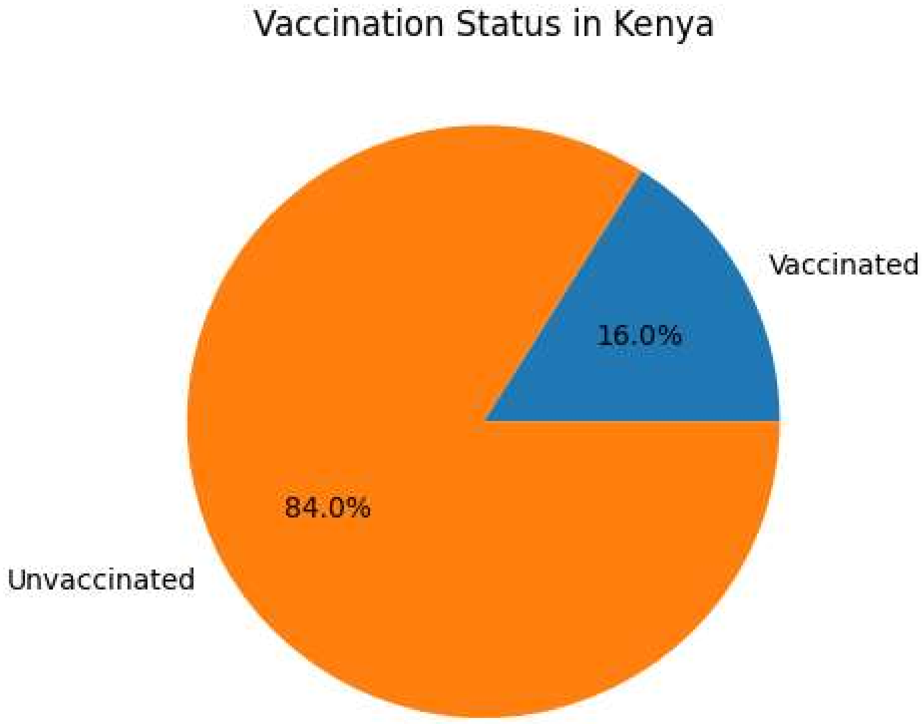
In [53]:
```
import seaborn as sns

# Plot % of population vaccinated
sns.lineplot(data=df_filtered, x="date", y="people_vaccinated_per_hundred", hue=

plt.xlabel("Date")
plt.ylabel("% Population Vaccinated")
plt.title("Vaccination Coverage Over Time")
plt.xticks(rotation=45)
plt.show()
```

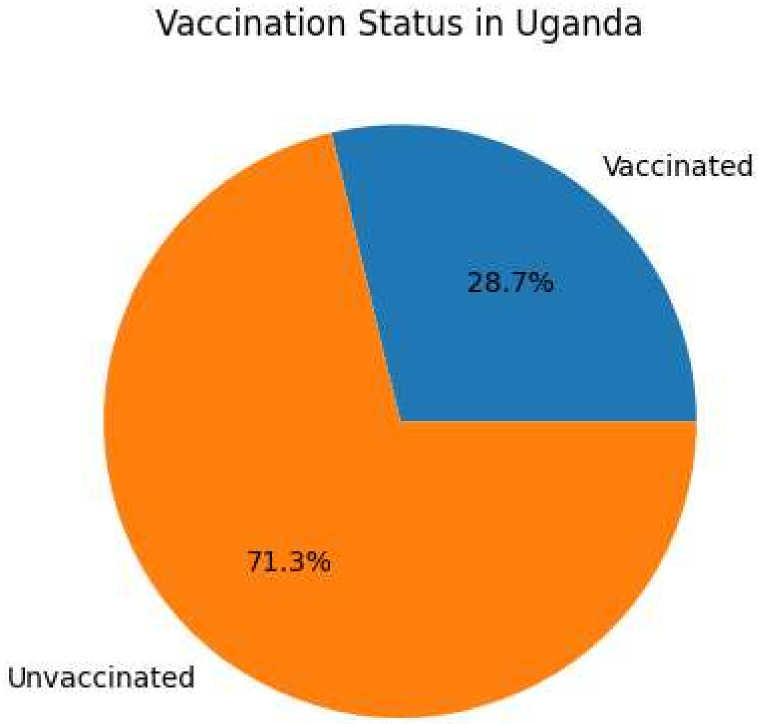


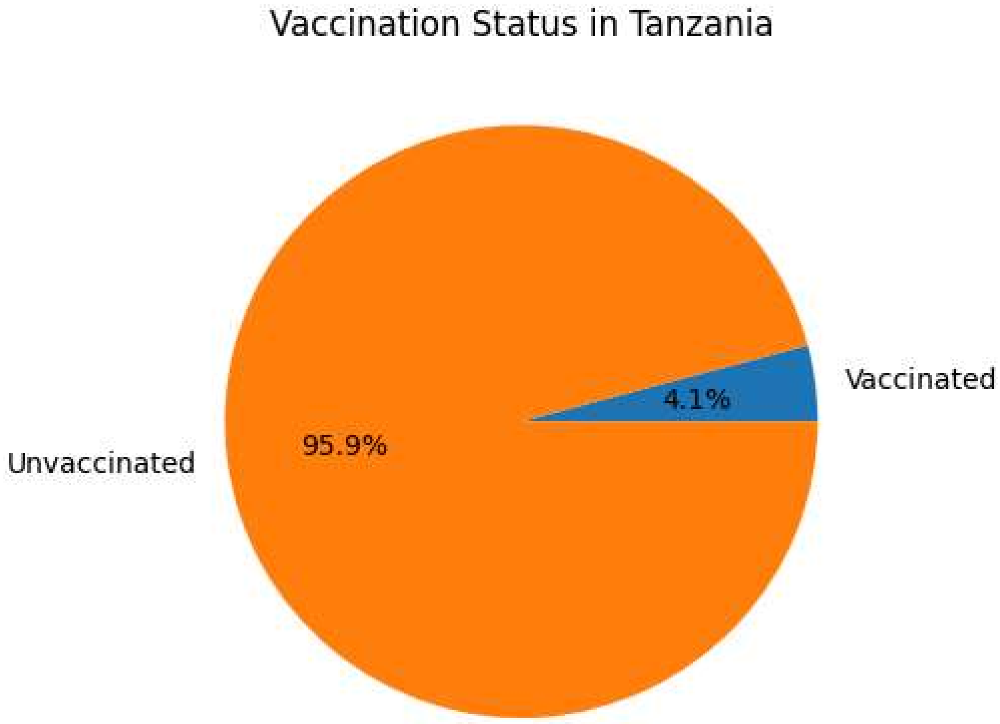
In [54]:
```
# Select latest data for each country
latest_data = df_filtered.groupby("location").last()

# Create pie chart for each country
for country in countries:
    vaccinated = latest_data.loc[country, "people_vaccinated_per_hundred"]
    unvaccinated = 100 - vaccinated

    plt.pie([vaccinated, unvaccinated], labels=["Vaccinated", "Unvaccinated"], a
    plt.title(f"Vaccination Status in {country}")
    plt.show()
```

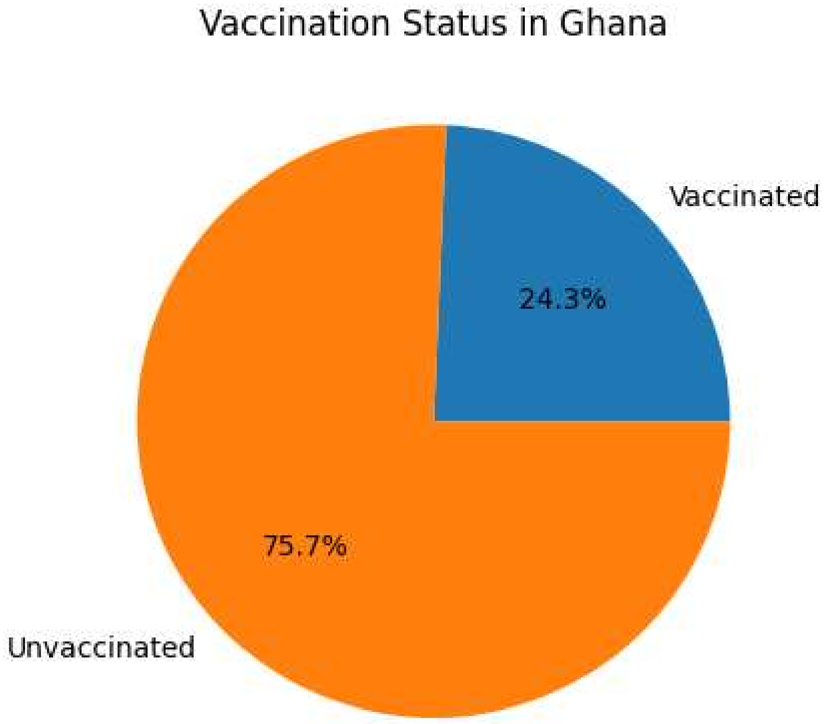

Vaccination Status in Nigeria
Vaccinated
8.4%
91.6%
Unvaccinated

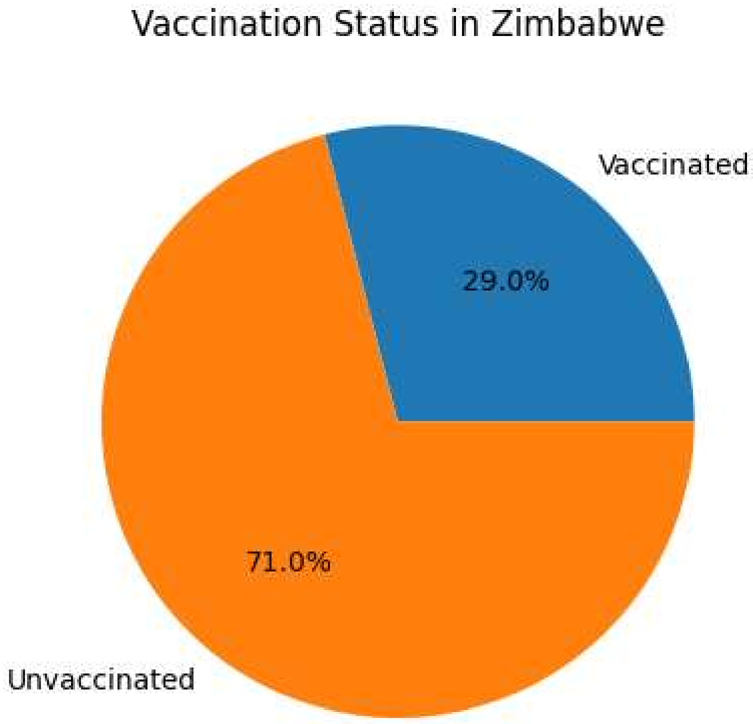

Vaccination Status in South Africa
Vaccinated
34.0%
66.0%
Unvaccinated

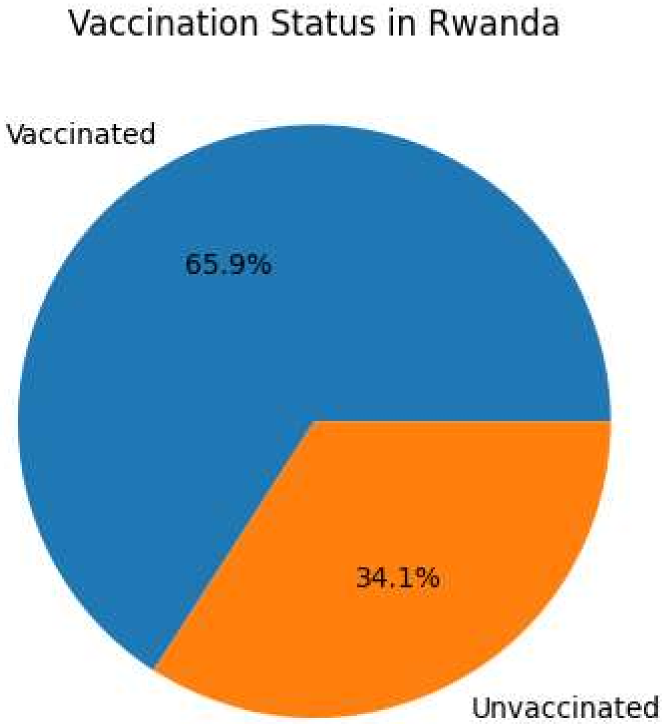

Vaccination Status in Kenya
Vaccinated
16.0%
84.0%
Unvaccinated

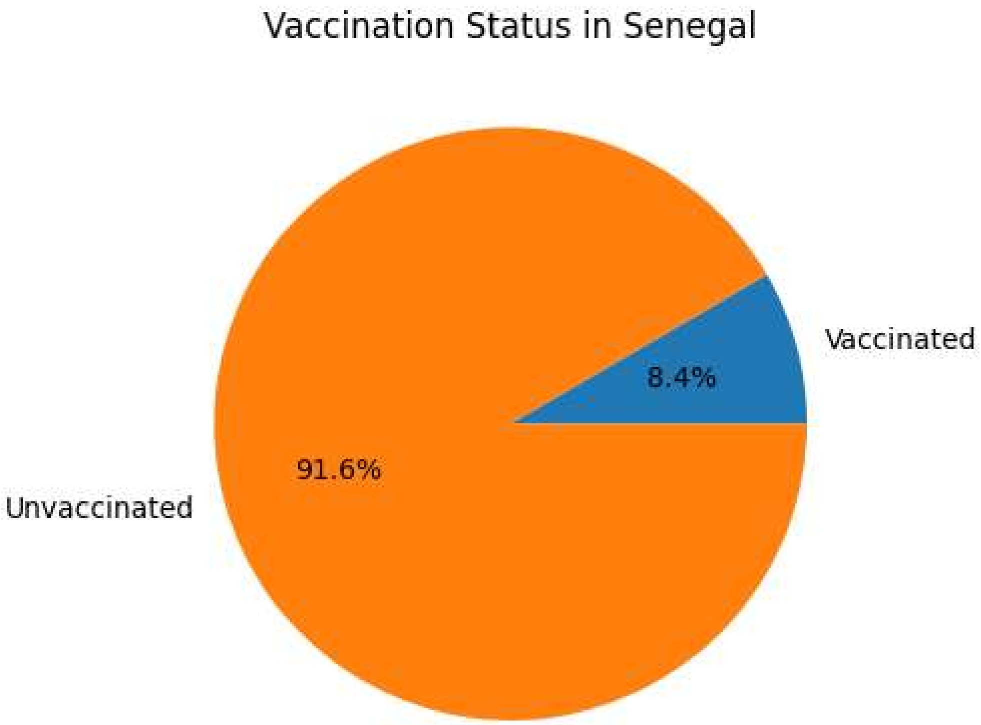

Vaccination Status in Uganda
Vaccinated
28.7%
71.3%
Unvaccinated

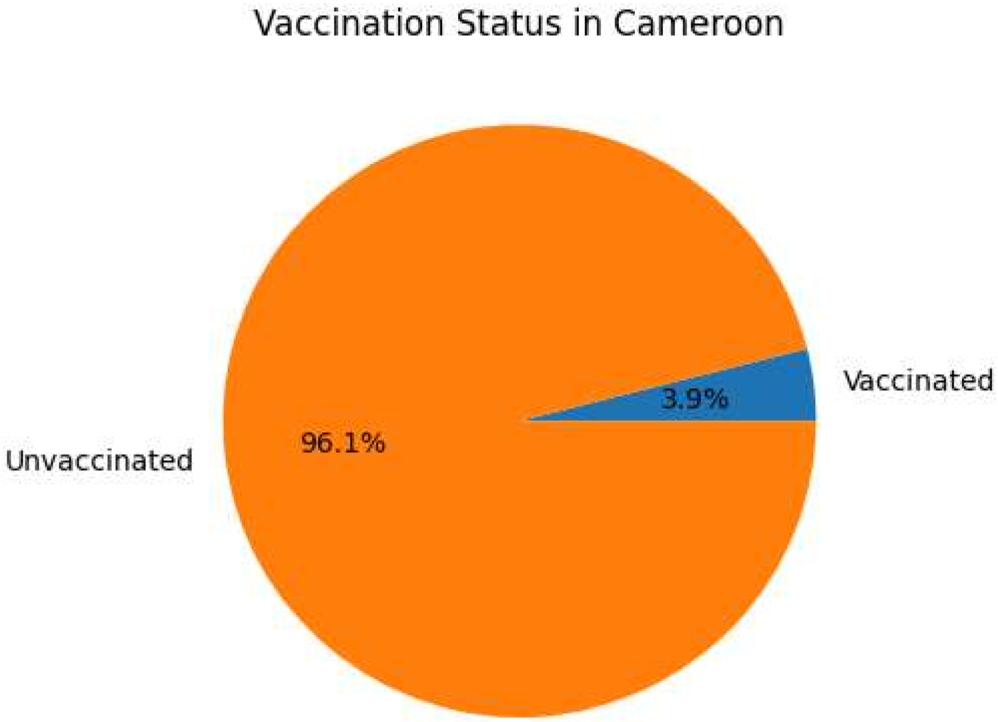

Vaccination Status in Tanzania
Vaccinated
4.1%
Unvaccinated
95.9%


Vaccination Status in Ghana
Vaccinated
24.3%
Unvaccinated
75.7%


Vaccination Status in Zimbabwe
Vaccinated
29.0%
71.0%
Unvaccinated


Vaccination Status in Rwanda
Vaccinated
65.9%
34.1%
Unvaccinated





# Vaccination status across Africa

- Rwanda has the highest percentage of vaccinated people

In [55]:
```
# import pandas as pd

# # Load the dataset
# df = pd.read_csv("owid-covid-data.csv")
# Get the latest data per country
latest_data = cleaned_df.groupby("iso_code").last().reset_index()

# Keep only necessary columns
df_map = latest_data[["iso_code", "location", "total_cases", "people_vaccinated_
```

In [56]:

```
import plotly.express as px

# plotting total cases

fig = px.choropleth(
    df_map,
    locations="iso_code",
    color="total_cases",
    hover_name="location",
    color_continuous_scale="Reds",
    title="COVID-19 Case Density by Country"
)

fig.show()
```

In [57]:

```
# Plotting vaccination rates

fig = px.choropleth(
    df_map,
    locations="iso_code",
    color="people_vaccinated_per_hundred",
    hover_name="location",
    color_continuous_scale="Blues",
    title="COVID-19 Vaccination Rates by Country"
)

fig.show()
```