In [2]:
```
import pandas as pd
file_path = 'owid-covid-data.csv'
df = pd.read_csv(file_path)
print(df.head())
print(df.info())
```

```
  iso_code continent     location        date  total_cases  new_cases  \
0      AFG      Asia  Afghanistan  2020-01-03          NaN        0.0
1      AFG      Asia  Afghanistan  2020-01-04          NaN        0.0
2      AFG      Asia  Afghanistan  2020-01-05          NaN        0.0
3      AFG      Asia  Afghanistan  2020-01-06          NaN        0.0
4      AFG      Asia  Afghanistan  2020-01-07          NaN        0.0

   new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  ...  \
0                 NaN           NaN         0.0                  NaN  ...
1                 NaN           NaN         0.0                  NaN  ...
2                 NaN           NaN         0.0                  NaN  ...
3                 NaN           NaN         0.0                  NaN  ...
4                 NaN           NaN         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  41128772.0
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                                   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]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 302512 entries, 0 to 302511
Data columns (total 67 columns):
 #   Column                                    Non-Null Count   Dtype
---  ------                                    --------------   -----
 0   iso_code                                  302512 non-null  object
 1   continent                                 288160 non-null  object
 2   location                                  302512 non-null  object
 3   date                                      302512 non-null  object
 4   total_cases                               266771 non-null  float64
 5   new_cases                                 294064 non-null  float64
 6   new_cases_smoothed                        292800 non-null  float64
 7   total_deaths                              246214 non-null  float64
```

```
8   new_deaths                                294139 non-null  float64
9   new_deaths_smoothed                       292909 non-null  float64
10  total_cases_per_million                   266771 non-null  float64
11  new_cases_per_million                     294064 non-null  float64
12  new_cases_smoothed_per_million            292800 non-null  float64
13  total_deaths_per_million                  246214 non-null  float64
14  new_deaths_per_million                    294139 non-null  float64
15  new_deaths_smoothed_per_million           292909 non-null  float64
16  reproduction_rate                         184817 non-null  float64
17  icu_patients                              34764 non-null   float64
18  icu_patients_per_million                  34764 non-null   float64
19  hosp_patients                             35138 non-null   float64
20  hosp_patients_per_million                 35138 non-null   float64
21  weekly_icu_admissions                     9101 non-null    float64
22  weekly_icu_admissions_per_million         9101 non-null    float64
23  weekly_hosp_admissions                    21287 non-null   float64
24  weekly_hosp_admissions_per_million        21287 non-null   float64
25  total_tests                               79387 non-null   float64
26  new_tests                                 75403 non-null   float64
27  total_tests_per_thousand                  79387 non-null   float64
28  new_tests_per_thousand                    75403 non-null   float64
29  new_tests_smoothed                        103965 non-null  float64
30  new_tests_smoothed_per_thousand           103965 non-null  float64
31  positive_rate                             95927 non-null   float64
32  tests_per_case                            94348 non-null   float64
33  tests_units                               106788 non-null  object
34  total_vaccinations                        73561 non-null   float64
35  people_vaccinated                         70411 non-null   float64
36  people_fully_vaccinated                   68149 non-null   float64
37  total_boosters                            42324 non-null   float64
38  new_vaccinations                          60542 non-null   float64
39  new_vaccinations_smoothed                 163536 non-null  float64
40  total_vaccinations_per_hundred            73561 non-null   float64
41  people_vaccinated_per_hundred             70411 non-null   float64
42  people_fully_vaccinated_per_hundred       68149 non-null   float64
43  total_boosters_per_hundred                42324 non-null   float64
44  new_vaccinations_smoothed_per_million     163536 non-null  float64
45  new_people_vaccinated_smoothed            163587 non-null  float64
46  new_people_vaccinated_smoothed_per_hundred 163587 non-null  float64
47  stringency_index                          193194 non-null  float64
48  population_density                        256703 non-null  float64
49  median_age                                238751 non-null  float64
50  aged_65_older                             230391 non-null  float64
51  aged_70_older                             236359 non-null  float64
52  gdp_per_capita                            233979 non-null  float64
53  extreme_poverty                           150700 non-null  float64
54  cardiovasc_death_rate                     234406 non-null  float64
55  diabetes_prevalence                       246348 non-null  float64
56  female_smokers                            175815 non-null  float64
57  male_smokers                              173423 non-null  float64
58  handwashing_facilities                    114817 non-null  float64
59  hospital_beds_per_thousand                206911 non-null  float64
60  life_expectancy                           278219 non-null  float64
61  human_development_index                   227212 non-null  float64
62  population                                302512 non-null  float64
63  excess_mortality_cumulative_absolute      10295 non-null   float64
```

```
 64  excess_mortality_cumulative                10295 non-null   float64
 65  excess_mortality                           10295 non-null   float64
 66  excess_mortality_cumulative_per_million    10295 non-null   float64
dtypes: float64(62), object(5)
memory usage: 154.6+ MB
None
```

In [3]:
```
print(df['location'].unique())
print(f"\nNumber of unique locations: {df['location'].nunique()}")
```

```
['Afghanistan' 'Africa' 'Albania' 'Algeria' 'American Samoa' 'Andorra'
 'Angola' 'Anguilla' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba'
 'Asia' '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' 'Ecuador' 'Egypt' 'El Salvador' 'England'
 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini' 'Ethiopia' 'Europe'
 'European Union' 'Faeroe 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' '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'
 'Lower middle income' '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 America' '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 America' 'South Korea' 'South Sudan'
 'Spain' 'Sri Lanka' 'Sudan' 'Suriname' 'Sweden' 'Switzerland' 'Syria'
 'Taiwan' 'Tajikistan' 'Tanzania' 'Thailand' 'Timor' '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' 'Uruguay'
 'Uzbekistan' 'Vanuatu' 'Vatican' 'Venezuela' 'Vietnam' 'Wales'
 'Wallis and Futuna' 'Western Sahara' 'World' 'Yemen' 'Zambia' 'Zimbabwe']

Number of unique locations: 255
```

In [4]:
```
df['date'] = pd.to_datetime(df['date'])
print(df['date'].dtype)
```

```
datetime64[ns]
```

In [5]:
```
kenya_data = df[df['location'] == 'Kenya'].copy()
print(kenya_data.head())
print(kenya_data.info())
```

```
       iso_code continent location       date  total_cases  new_cases  \
136231      KEN    Africa    Kenya 2020-01-03          NaN        0.0
136232      KEN    Africa    Kenya 2020-01-04          NaN        0.0
136233      KEN    Africa    Kenya 2020-01-05          NaN        0.0
136234      KEN    Africa    Kenya 2020-01-06          NaN        0.0
136235      KEN    Africa    Kenya 2020-01-07          NaN        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
136231                 NaN           NaN         0.0                  NaN
136232                 NaN           NaN         0.0                  NaN
136233                 NaN           NaN         0.0                  NaN
136234                 NaN           NaN         0.0                  NaN
136235                 NaN           NaN         0.0                  NaN

        ...  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
136231  ...          20.4                  24.651                         1.4
136232  ...          20.4                  24.651                         1.4
136233  ...          20.4                  24.651                         1.4
136234  ...          20.4                  24.651                         1.4
136235  ...          20.4                  24.651                         1.4

        life_expectancy  human_development_index  population  \
136231             66.7                    0.601  54027484.0
136232             66.7                    0.601  54027484.0
136233             66.7                    0.601  54027484.0
136234             66.7                    0.601  54027484.0
136235             66.7                    0.601  54027484.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
136231                                   NaN                          NaN
136232                                   NaN                          NaN
136233                                   NaN                          NaN
136234                                   NaN                          NaN
136235                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
136231               NaN                                     NaN
136232               NaN                                     NaN
136233               NaN                                     NaN
136234               NaN                                     NaN
136235               NaN                                     NaN

[5 rows x 67 columns]
<class 'pandas.core.frame.DataFrame'>
Index: 1196 entries, 136231 to 137426
Data columns (total 67 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   iso_code                                  1196 non-null   object
 1   continent                                 1196 non-null   object
 2   location                                  1196 non-null   object
 3   date                                      1196 non-null   datetime64[ns]
 4   total_cases                               1125 non-null   float64
 5   new_cases                                 1196 non-null   float64
 6   new_cases_smoothed                        1191 non-null   float64
 7   total_deaths                              1112 non-null   float64
```

```
8   new_deaths                                  1196 non-null   float64
9   new_deaths_smoothed                         1191 non-null   float64
10  total_cases_per_million                     1125 non-null   float64
11  new_cases_per_million                       1196 non-null   float64
12  new_cases_smoothed_per_million              1191 non-null   float64
13  total_deaths_per_million                    1112 non-null   float64
14  new_deaths_per_million                      1196 non-null   float64
15  new_deaths_smoothed_per_million             1191 non-null   float64
16  reproduction_rate                           1005 non-null   float64
17  icu_patients                                0 non-null      float64
18  icu_patients_per_million                    0 non-null      float64
19  hosp_patients                               0 non-null      float64
20  hosp_patients_per_million                   0 non-null      float64
21  weekly_icu_admissions                       0 non-null      float64
22  weekly_icu_admissions_per_million           0 non-null      float64
23  weekly_hosp_admissions                      0 non-null      float64
24  weekly_hosp_admissions_per_million          0 non-null      float64
25  total_tests                                 272 non-null    float64
26  new_tests                                   184 non-null    float64
27  total_tests_per_thousand                    272 non-null    float64
28  new_tests_per_thousand                      184 non-null    float64
29  new_tests_smoothed                          832 non-null    float64
30  new_tests_smoothed_per_thousand             832 non-null    float64
31  positive_rate                               827 non-null    float64
32  tests_per_case                              827 non-null    float64
33  tests_units                                 839 non-null    object
34  total_vaccinations                          291 non-null    float64
35  people_vaccinated                           239 non-null    float64
36  people_fully_vaccinated                     259 non-null    float64
37  total_boosters                              89 non-null     float64
38  new_vaccinations                            209 non-null    float64
39  new_vaccinations_smoothed                   759 non-null    float64
40  total_vaccinations_per_hundred              291 non-null    float64
41  people_vaccinated_per_hundred               239 non-null    float64
42  people_fully_vaccinated_per_hundred         259 non-null    float64
43  total_boosters_per_hundred                  89 non-null     float64
44  new_vaccinations_smoothed_per_million       759 non-null    float64
45  new_people_vaccinated_smoothed              759 non-null    float64
46  new_people_vaccinated_smoothed_per_hundred  759 non-null    float64
47  stringency_index                            1075 non-null   float64
48  population_density                          1196 non-null   float64
49  median_age                                  1196 non-null   float64
50  aged_65_older                               1196 non-null   float64
51  aged_70_older                               1196 non-null   float64
52  gdp_per_capita                              1196 non-null   float64
53  extreme_poverty                             1196 non-null   float64
54  cardiovasc_death_rate                       1196 non-null   float64
55  diabetes_prevalence                         1196 non-null   float64
56  female_smokers                              1196 non-null   float64
57  male_smokers                                1196 non-null   float64
58  handwashing_facilities                      1196 non-null   float64
59  hospital_beds_per_thousand                  1196 non-null   float64
60  life_expectancy                             1196 non-null   float64
61  human_development_index                     1196 non-null   float64
62  population                                  1196 non-null   float64
63  excess_mortality_cumulative_absolute        0 non-null      float64
```

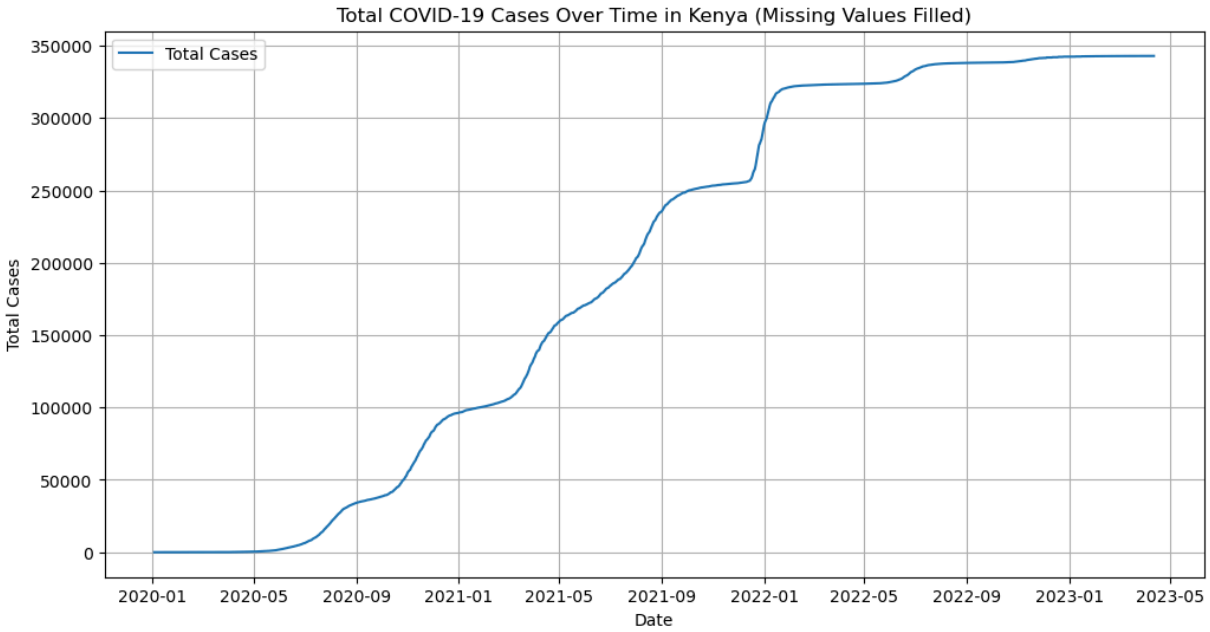
```
 64  excess_mortality_cumulative                0 non-null      float64
 65  excess_mortality                           0 non-null      float64
 66  excess_mortality_cumulative_per_million    0 non-null      float64
dtypes: datetime64[ns](1), float64(62), object(4)
memory usage: 635.4+ KB
None
```

In [6]:
```
print(kenya_data[['date', 'total_cases', 'new_cases', 'total_deaths', 'new_deaths']]
```

```
date            0
total_cases    71
new_cases       0
total_deaths   84
new_deaths      0
dtype: int64
```

In [8]:
```
import matplotlib.pyplot as plt

kenya_data['total_cases'].fillna(0, inplace=True)
kenya_data['total_deaths'].fillna(0, inplace=True)

# Verify that the missing values have been filled
print(kenya_data[['total_cases', 'total_deaths']].isnull().sum())

# Let's also re-plot the total cases to see the effect
plt.figure(figsize=(12, 6))
plt.plot(kenya_data['date'], kenya_data['total_cases'], label='Total Cases')
plt.xlabel('Date')
plt.ylabel('Total Cases')
plt.title('Total COVID-19 Cases Over Time in Kenya (Missing Values Filled)')
plt.grid(True)
plt.legend()
plt.show()
```

```
total_cases     0
total_deaths    0
dtype: int64
```

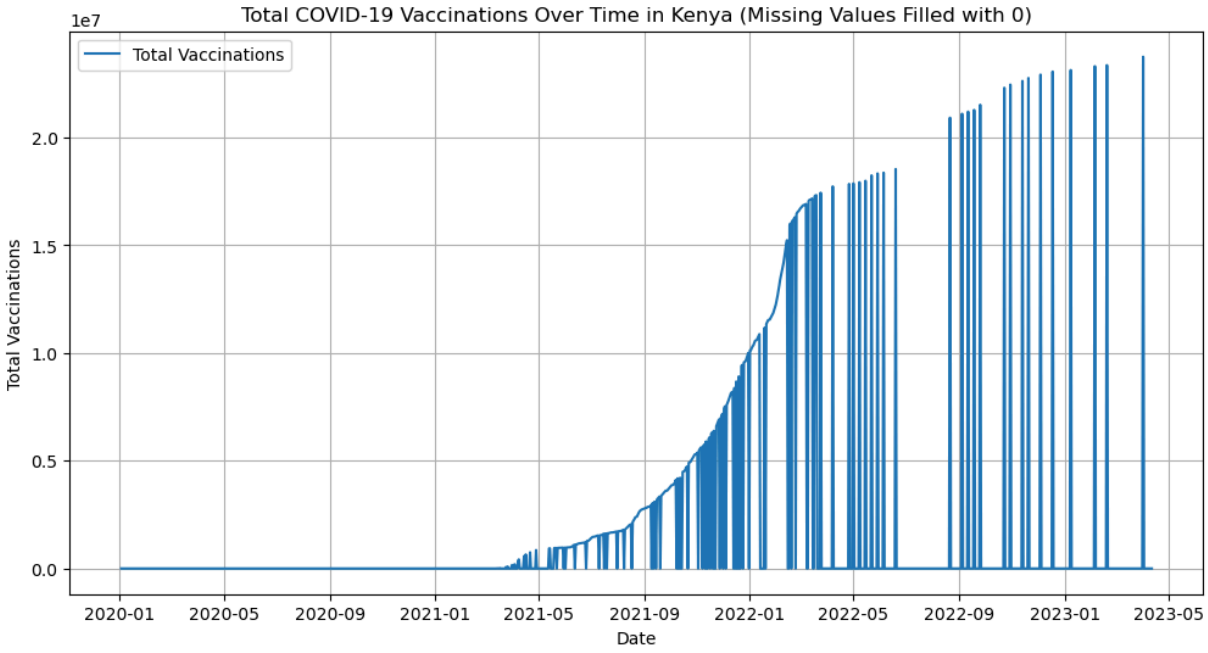


In [9]:
```
print(kenya_data[['total_vaccinations', 'new_vaccinations', 'new_vaccinations_smoot
```

```
total_vaccinations           905
new_vaccinations             987
new_vaccinations_smoothed    437
dtype: int64
```

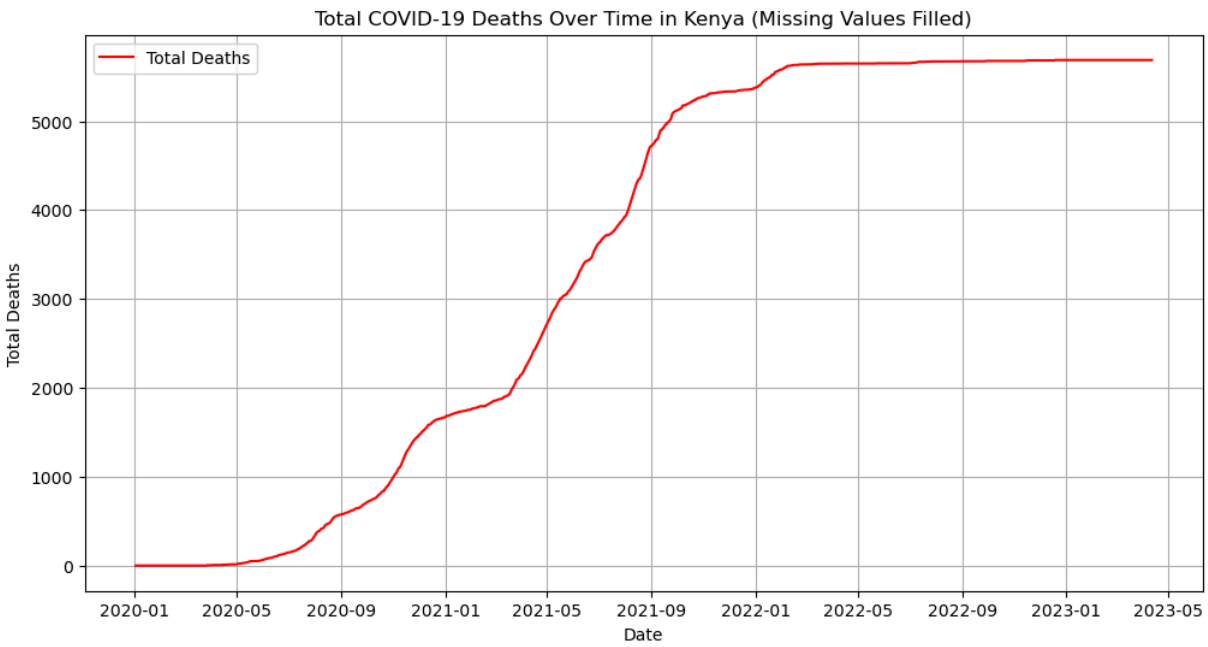
In [10]:
```
kenya_data['total_vaccinations'].fillna(0, inplace=True)
kenya_data['new_vaccinations'].fillna(0, inplace=True)
kenya_data['new_vaccinations_smoothed'].fillna(0, inplace=True)

# Verify that the missing values have been filled
print(kenya_data[['total_vaccinations', 'new_vaccinations', 'new_vaccinations_smoot

# Plot total vaccinations over time
plt.figure(figsize=(12, 6))
plt.plot(kenya_data['date'], kenya_data['total_vaccinations'], label='Total Vaccina
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.title('Total COVID-19 Vaccinations Over Time in Kenya (Missing Values Filled wi
plt.grid(True)
plt.legend()
plt.show()
```

```
total_vaccinations           0
new_vaccinations             0
new_vaccinations_smoothed    0
dtype: int64
```



In [11]:
```
plt.figure(figsize=(12, 6))
plt.plot(kenya_data['date'], kenya_data['total_deaths'], label='Total Deaths', colo
plt.xlabel('Date')
plt.ylabel('Total Deaths')
plt.title('Total COVID-19 Deaths Over Time in Kenya (Missing Values Filled)')
plt.grid(True)
```

```
plt.legend()
plt.show()
```



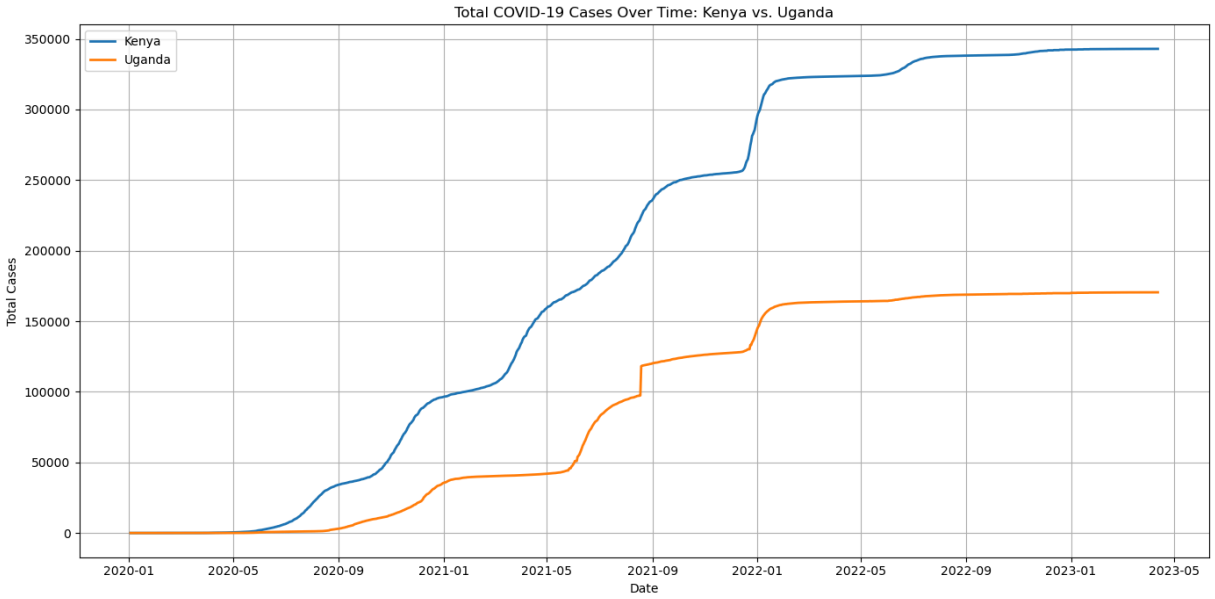
In [12]:
```
uganda_data = df[df['location'] == 'Uganda'].copy()

# Convert the 'date' column to datetime if we haven't already (should be done globa
if uganda_data['date'].dtype != 'datetime64[ns]':
    uganda_data['date'] = pd.to_datetime(uganda_data['date'])

# Fill missing values in total_cases and total_deaths with 0 for Uganda as well
uganda_data['total_cases'].fillna(0, inplace=True)
uganda_data['total_deaths'].fillna(0, inplace=True)

# Display the first few rows of the Uganda-specific data
print(uganda_data.head())

# Get some info about the Uganda-specific data
print(uganda_data.info())
```

```
       iso_code continent location       date  total_cases  new_cases  \
279883      UGA    Africa   Uganda 2020-01-03          0.0        0.0
279884      UGA    Africa   Uganda 2020-01-04          0.0        0.0
279885      UGA    Africa   Uganda 2020-01-05          0.0        0.0
279886      UGA    Africa   Uganda 2020-01-06          0.0        0.0
279887      UGA    Africa   Uganda 2020-01-07          0.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
279883                 NaN           0.0         0.0                  NaN
279884                 NaN           0.0         0.0                  NaN
279885                 NaN           0.0         0.0                  NaN
279886                 NaN           0.0         0.0                  NaN
279887                 NaN           0.0         0.0                  NaN

        ...  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
279883  ...          16.7                  21.222                         0.5
279884  ...          16.7                  21.222                         0.5
279885  ...          16.7                  21.222                         0.5
279886  ...          16.7                  21.222                         0.5
279887  ...          16.7                  21.222                         0.5

        life_expectancy  human_development_index  population  \
279883            63.37                    0.544  47249588.0
279884            63.37                    0.544  47249588.0
279885            63.37                    0.544  47249588.0
279886            63.37                    0.544  47249588.0
279887            63.37                    0.544  47249588.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
279883                                   NaN                          NaN
279884                                   NaN                          NaN
279885                                   NaN                          NaN
279886                                   NaN                          NaN
279887                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
279883               NaN                                     NaN
279884               NaN                                     NaN
279885               NaN                                     NaN
279886               NaN                                     NaN
279887               NaN                                     NaN

[5 rows x 67 columns]
<class 'pandas.core.frame.DataFrame'>
Index: 1196 entries, 279883 to 281078
Data columns (total 67 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   iso_code                                  1196 non-null   object
 1   continent                                 1196 non-null   object
 2   location                                  1196 non-null   object
 3   date                                      1196 non-null   datetime64[ns]
 4   total_cases                               1196 non-null   float64
 5   new_cases                                 1195 non-null   float64
 6   new_cases_smoothed                        1190 non-null   float64
 7   total_deaths                              1196 non-null   float64
```

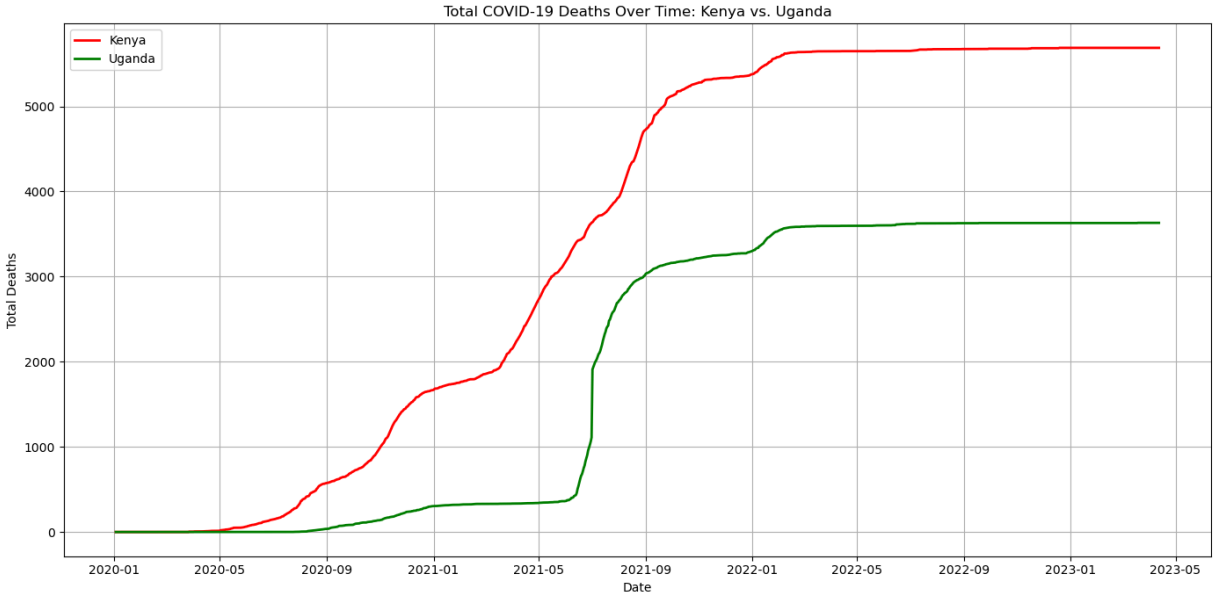
```
8   new_deaths                                  1196 non-null   float64
9   new_deaths_smoothed                         1191 non-null   float64
10  total_cases_per_million                     1117 non-null   float64
11  new_cases_per_million                       1195 non-null   float64
12  new_cases_smoothed_per_million              1190 non-null   float64
13  total_deaths_per_million                    995 non-null    float64
14  new_deaths_per_million                      1196 non-null   float64
15  new_deaths_smoothed_per_million             1191 non-null   float64
16  reproduction_rate                           971 non-null    float64
17  icu_patients                                0 non-null      float64
18  icu_patients_per_million                    0 non-null      float64
19  hosp_patients                               0 non-null      float64
20  hosp_patients_per_million                   0 non-null      float64
21  weekly_icu_admissions                       0 non-null      float64
22  weekly_icu_admissions_per_million           0 non-null      float64
23  weekly_hosp_admissions                      0 non-null      float64
24  weekly_hosp_admissions_per_million          0 non-null      float64
25  total_tests                                 644 non-null    float64
26  new_tests                                   588 non-null    float64
27  total_tests_per_thousand                    644 non-null    float64
28  new_tests_per_thousand                      588 non-null    float64
29  new_tests_smoothed                          806 non-null    float64
30  new_tests_smoothed_per_thousand             806 non-null    float64
31  positive_rate                               804 non-null    float64
32  tests_per_case                              792 non-null    float64
33  tests_units                                 813 non-null    object
34  total_vaccinations                          139 non-null    float64
35  people_vaccinated                           98 non-null     float64
36  people_fully_vaccinated                     44 non-null     float64
37  total_boosters                              21 non-null     float64
38  new_vaccinations                            61 non-null     float64
39  new_vaccinations_smoothed                   747 non-null    float64
40  total_vaccinations_per_hundred              139 non-null    float64
41  people_vaccinated_per_hundred               98 non-null     float64
42  people_fully_vaccinated_per_hundred         44 non-null     float64
43  total_boosters_per_hundred                  21 non-null     float64
44  new_vaccinations_smoothed_per_million       747 non-null    float64
45  new_people_vaccinated_smoothed              747 non-null    float64
46  new_people_vaccinated_smoothed_per_hundred  747 non-null    float64
47  stringency_index                            1075 non-null   float64
48  population_density                          1196 non-null   float64
49  median_age                                  1196 non-null   float64
50  aged_65_older                               1196 non-null   float64
51  aged_70_older                               1196 non-null   float64
52  gdp_per_capita                              1196 non-null   float64
53  extreme_poverty                             1196 non-null   float64
54  cardiovasc_death_rate                       1196 non-null   float64
55  diabetes_prevalence                         1196 non-null   float64
56  female_smokers                              1196 non-null   float64
57  male_smokers                                1196 non-null   float64
58  handwashing_facilities                      1196 non-null   float64
59  hospital_beds_per_thousand                  1196 non-null   float64
60  life_expectancy                             1196 non-null   float64
61  human_development_index                     1196 non-null   float64
62  population                                  1196 non-null   float64
63  excess_mortality_cumulative_absolute        0 non-null      float64
```

```
 64  excess_mortality_cumulative                0 non-null      float64
 65  excess_mortality                           0 non-null      float64
 66  excess_mortality_cumulative_per_million    0 non-null      float64
dtypes: datetime64[ns](1), float64(62), object(4)
memory usage: 635.4+ KB
None
```

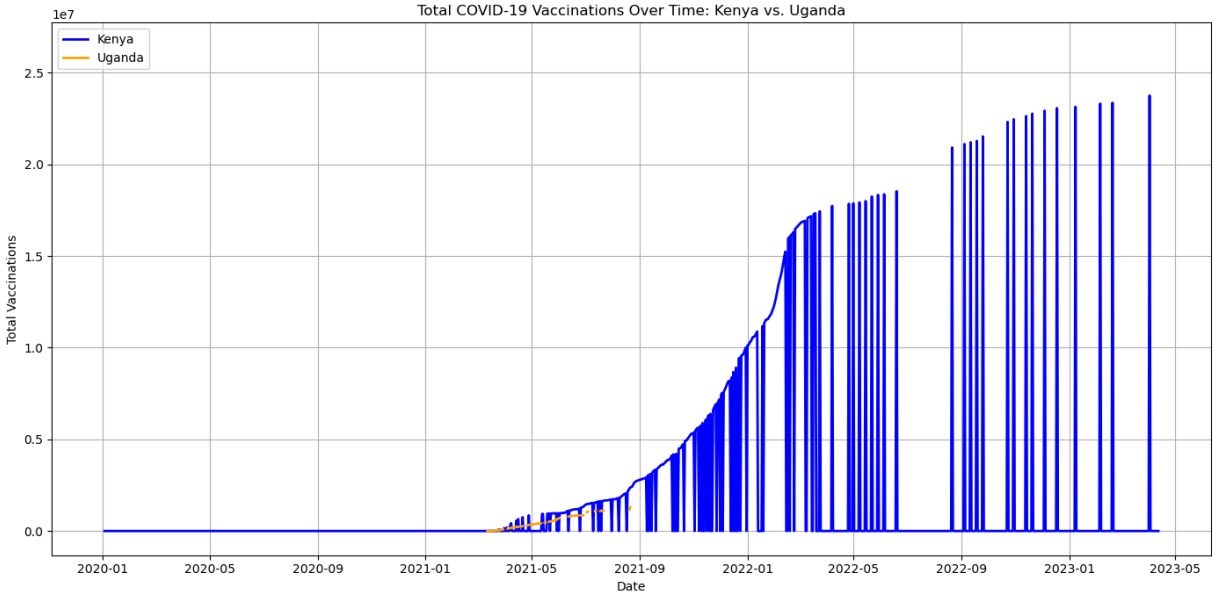
In [13]:
```
plt.figure(figsize=(14, 7))
plt.plot(kenya_data['date'], kenya_data['total_cases'], label='Kenya', linewidth=2)
plt.plot(uganda_data['date'], uganda_data['total_cases'], label='Uganda', linewidth
plt.xlabel('Date')
plt.ylabel('Total Cases')
plt.title('Total COVID-19 Cases Over Time: Kenya vs. Uganda')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```



In [14]:
```
plt.figure(figsize=(14, 7))
plt.plot(kenya_data['date'], kenya_data['total_deaths'], label='Kenya', color='red'
plt.plot(uganda_data['date'], uganda_data['total_deaths'], label='Uganda', color='g
plt.xlabel('Date')
plt.ylabel('Total Deaths')
plt.title('Total COVID-19 Deaths Over Time: Kenya vs. Uganda')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```



In [15]:
```
plt.figure(figsize=(14, 7))
plt.plot(kenya_data['date'], kenya_data['total_vaccinations'], label='Kenya', color
plt.plot(uganda_data['date'], uganda_data['total_vaccinations'], label='Uganda', co
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.title('Total COVID-19 Vaccinations Over Time: Kenya vs. Uganda')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```



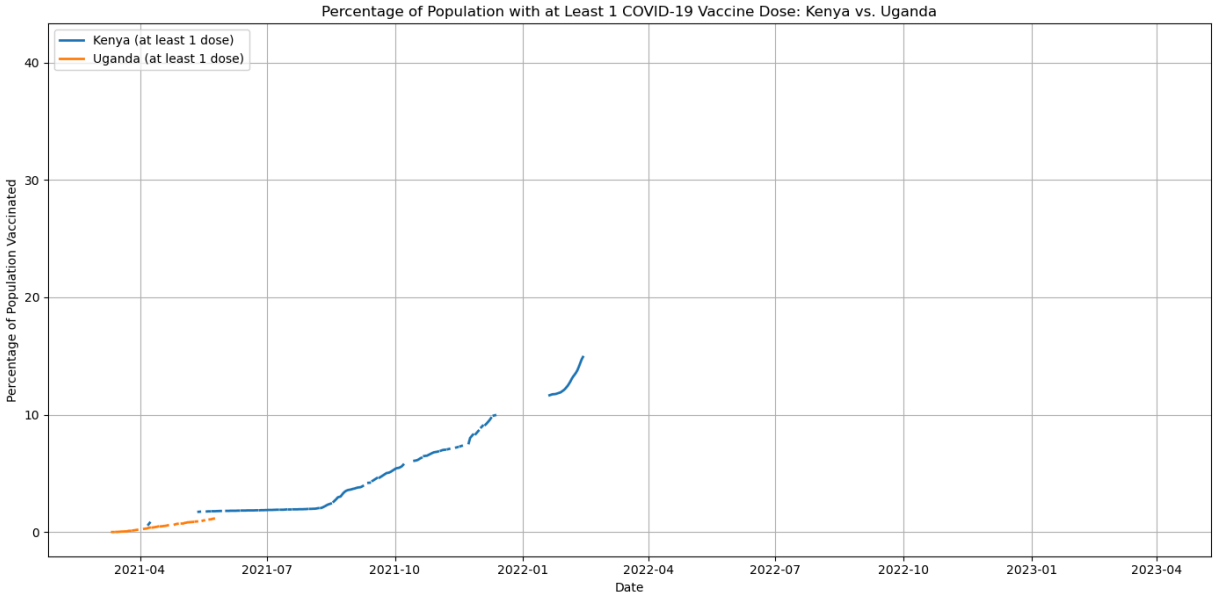
In [16]:
```
print(kenya_data.columns)
```

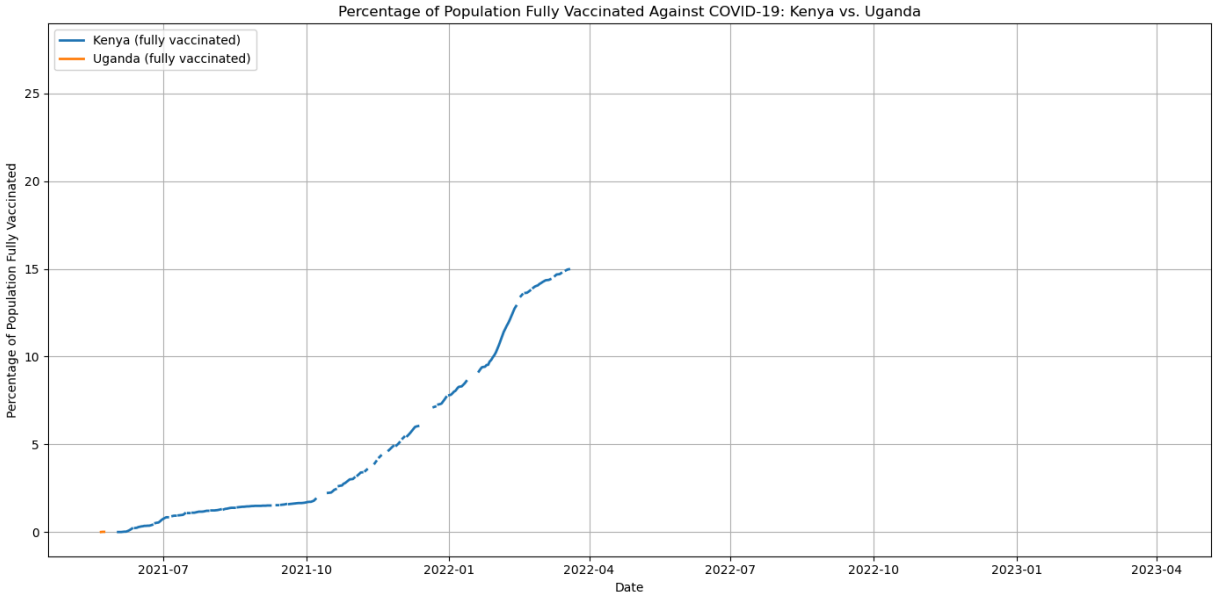
```
Index(['iso_code', 'continent', 'location', 'date', 'total_cases', 'new_cases',
       'new_cases_smoothed', 'total_deaths', 'new_deaths',
       'new_deaths_smoothed', 'total_cases_per_million',
       'new_cases_per_million', 'new_cases_smoothed_per_million',
       'total_deaths_per_million', '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_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', 'population',
       'excess_mortality_cumulative_absolute', 'excess_mortality_cumulative',
       'excess_mortality', 'excess_mortality_cumulative_per_million'],
      dtype='object')
```

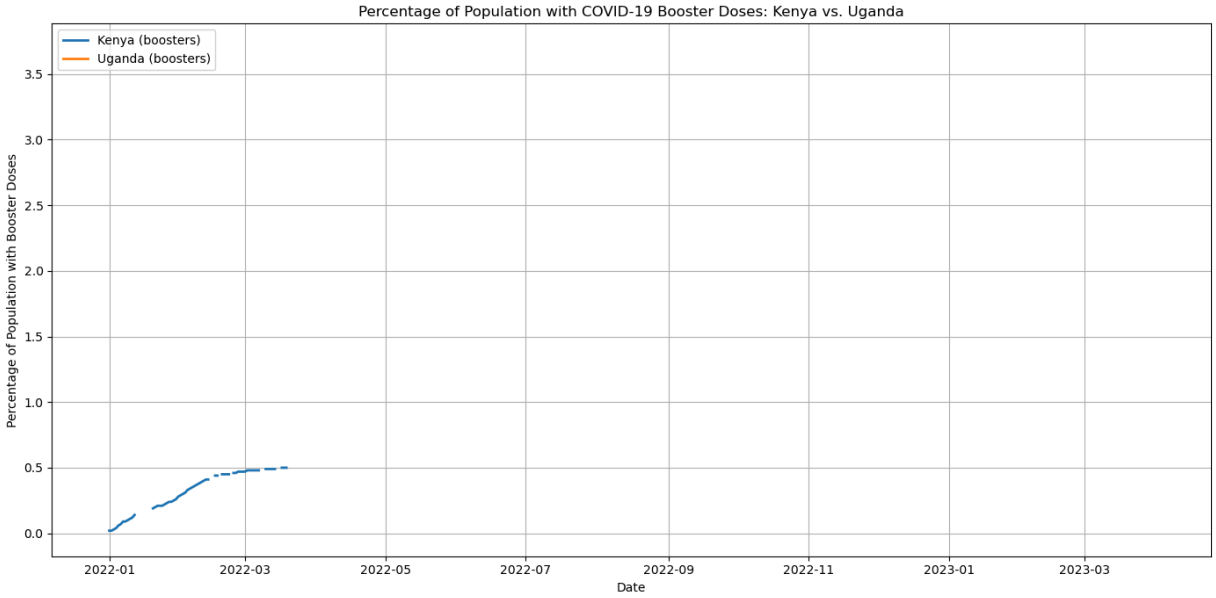
In [17]:

```
plt.figure(figsize=(14, 7))
plt.plot(kenya_data['date'], kenya_data['people_vaccinated_per_hundred'], label='Ke
plt.plot(uganda_data['date'], uganda_data['people_vaccinated_per_hundred'], label='
plt.xlabel('Date')
plt.ylabel('Percentage of Population Vaccinated')
plt.title('Percentage of Population with at Least 1 COVID-19 Vaccine Dose: Kenya vs
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()

plt.figure(figsize=(14, 7))
plt.plot(kenya_data['date'], kenya_data['people_fully_vaccinated_per_hundred'], lab
plt.plot(uganda_data['date'], uganda_data['people_fully_vaccinated_per_hundred'], l
plt.xlabel('Date')
plt.ylabel('Percentage of Population Fully Vaccinated')
plt.title('Percentage of Population Fully Vaccinated Against COVID-19: Kenya vs. Ug
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()

# Optional: Plot boosters as well
plt.figure(figsize=(14, 7))
plt.plot(kenya_data['date'], kenya_data['total_boosters_per_hundred'], label='Kenya
plt.plot(uganda_data['date'], uganda_data['total_boosters_per_hundred'], label='Uga
plt.xlabel('Date')
```

```
plt.ylabel('Percentage of Population with Booster Doses')
plt.title('Percentage of Population with COVID-19 Booster Doses: Kenya vs. Uganda')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```







In [18]:
```
# Sort the DataFrame by location and date
df_sorted = df.sort_values(['location', 'date'])

# Get the last record for each location (which will be the latest date)
latest_data = df_sorted.groupby('location').tail(1)

# Select only the iso_code and total_cases columns
map_data = latest_data[['iso_code', 'total_cases']].copy()

# Print the first few rows of the prepared data
print(map_data.head())
```

```
     iso_code  total_cases
1195      AFG     211630.0
2391  OWID_AFR   13074358.0
3587      ALB     333897.0
4783      DZA     271613.0
5979      ASM       8326.0
```

In [19]:
```
import plotly.express as px

fig = px.choropleth(map_data,
                    locations='iso_code',
                    locationmode='ISO-3',
                    color='total_cases',
                    hover_name='iso_code',
                    color_continuous_scale=px.colors.sequential.Plasma,
                    title='Total Confirmed COVID-19 Cases (Latest Data)')
fig.show()
```

Key Insights from the COVID-19 Data Analysis

1. **Global Spread:** The choropleth map of total confirmed COVID-19 cases illustrates the widespread impact of the pandemic across the globe, with varying levels of cumulative cases observed in different regions.

2. **Comparison of Kenya and Uganda:** Analysis of the trends in Kenya and Uganda reveals that Kenya experienced a higher overall number of confirmed cases compared to Uganda. Furthermore, Kenya's COVID-19 vaccination campaign appears to have commenced earlier and progressed at a faster rate in terms of total doses administered.

3. **Vaccination Progress:** Based on the total vaccination data, Uganda showed a slower initial uptake and a lower overall number of vaccinations administered compared to Kenya within the analyzed period.

Anomalies and Interesting Patterns

- The total vaccination data for Kenya exhibits some sharp, vertical increases over time. This pattern might suggest inconsistencies in the reporting of vaccination data, where large numbers of records were added at once, rather than a smooth daily or weekly progression.

Further Considerations

- This analysis provides a preliminary overview. Further investigation could explore factors contributing to the differences observed between countries, such as population density, public health policies, and access to resources.
- Analyzing the trends in new cases and new deaths could provide additional insights into the dynamics of the pandemic in these regions.
- Exploring vaccination rates in more detail, including the distribution of first, second, and booster doses, could offer a more comprehensive understanding of the vaccination progress.