COVID-19 Global Data Tracker

May 11, 2025

[1]:
```
import pandas as pd

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

# Checking Columns
print("Columns:", df.columns)

# Preview Rows
print("\nPreview:")
print(df.head())

# Identifying Missing Values
print("\nMissing values:")
print(df.isnull().sum())
```

```
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')

Preview:
  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]

Missing values:
iso_code                                      0
continent                                 14352
location                                      0
date                                          0
total_cases                               35741
                                          …
population                                    0
excess_mortality_cumulative_absolute     292217
excess_mortality_cumulative              292217
excess_mortality                         292217
excess_mortality_cumulative_per_million  292217
Length: 67, dtype: int64
```

[5]:
```python
import pandas as pd

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

# Filtering Countries
df = df[df['location'].isin(['Kenya', 'USA', 'India', 'Uganda'])]

# Dropping rows with missing dates/critical values.
df = df.dropna(subset=['date', 'total_cases', 'new_cases'])

# Converting date column to datetime: pd.to_datetime().
df['date'] = pd.to_datetime(df['date'])

# Handling missing numeric values
df.interpolate(method='linear', inplace=True)

# Previewing cleaned data
print(df.head())
```

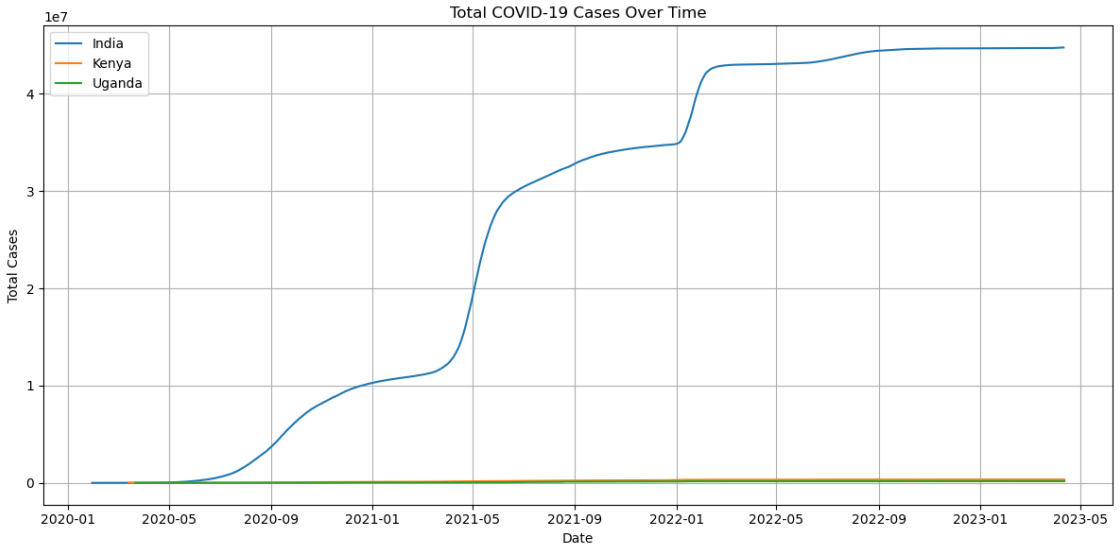
```
       iso_code continent location       date  total_cases  new_cases  \
120710      IND      Asia    India 2020-01-30          1.0        1.0
120711      IND      Asia    India 2020-01-31          5.0        4.0
120712      IND      Asia    India 2020-02-01          5.0        0.0
120714      IND      Asia    India 2020-02-03          3.0        1.0
120715      IND      Asia    India 2020-02-04          7.0        4.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
120710               0.143           NaN         0.0                  0.0
```

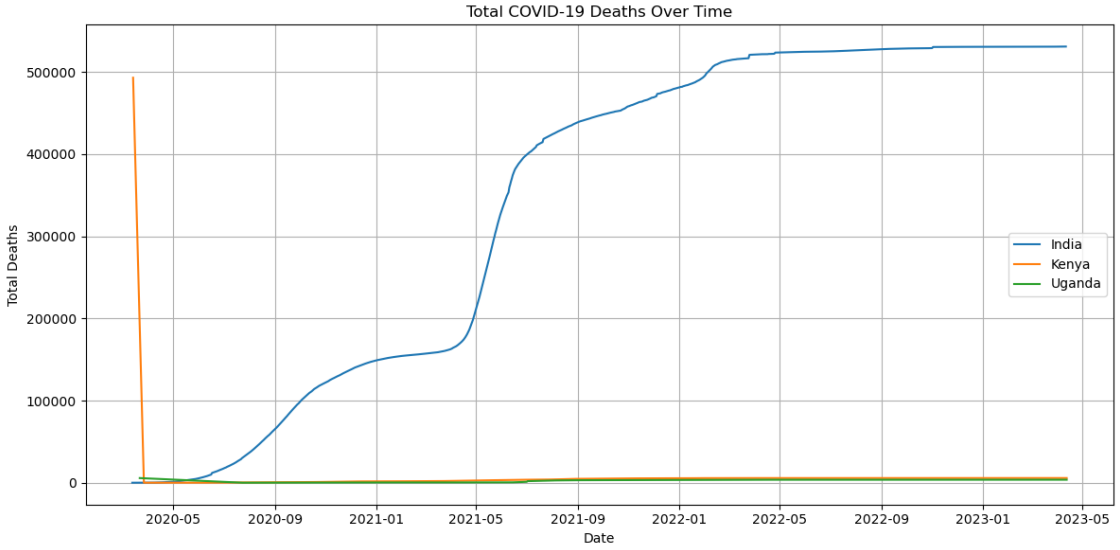
```
120711               0.714           NaN         0.0                   0.0
120712               0.714           NaN         0.0                   0.0
120714               0.857           NaN         0.0                   0.0
120715               1.429           NaN         0.0                   0.0

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
120710  …          20.6                   59.55                        0.53
120711  …          20.6                   59.55                        0.53
120712  …          20.6                   59.55                        0.53
120714  …          20.6                   59.55                        0.53
120715  …          20.6                   59.55                        0.53

        life_expectancy  human_development_index    population  \
120710            69.66                    0.645  1.417173e+09
120711            69.66                    0.645  1.417173e+09
120712            69.66                    0.645  1.417173e+09
120714            69.66                    0.645  1.417173e+09
120715            69.66                    0.645  1.417173e+09

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
120710                                   NaN                          NaN
120711                                   NaN                          NaN
120712                                   NaN                          NaN
120714                                   NaN                          NaN
120715                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
120710               NaN                                      NaN
120711               NaN                                      NaN
120712               NaN                                      NaN
120714               NaN                                      NaN
120715               NaN                                      NaN

[5 rows x 67 columns]
```

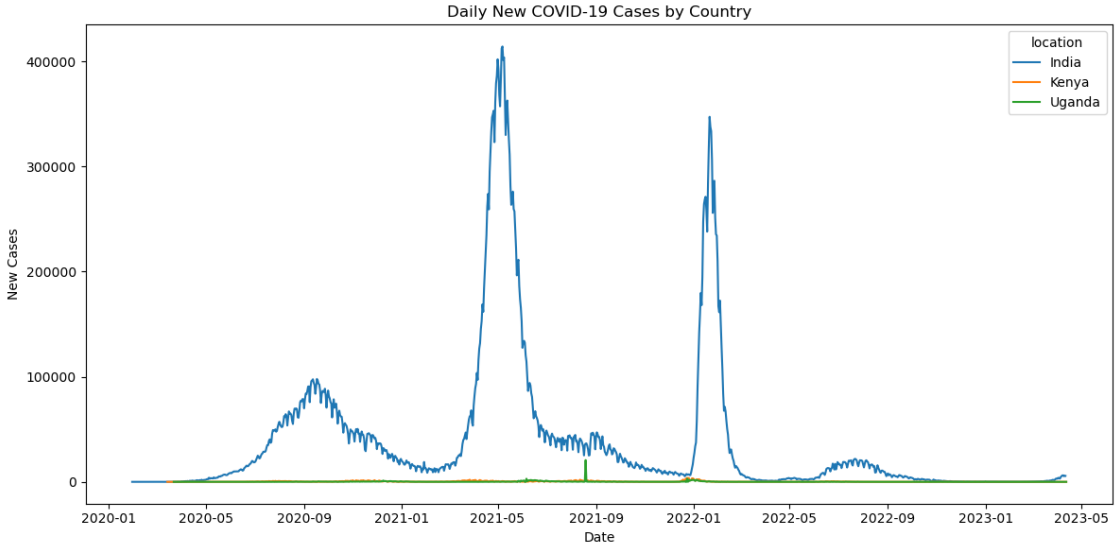
[6]:
```python
#  Plotting total cases over time (Line chart)
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12, 6))
for country in df['location'].unique():
    country_data = df[df['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')
```

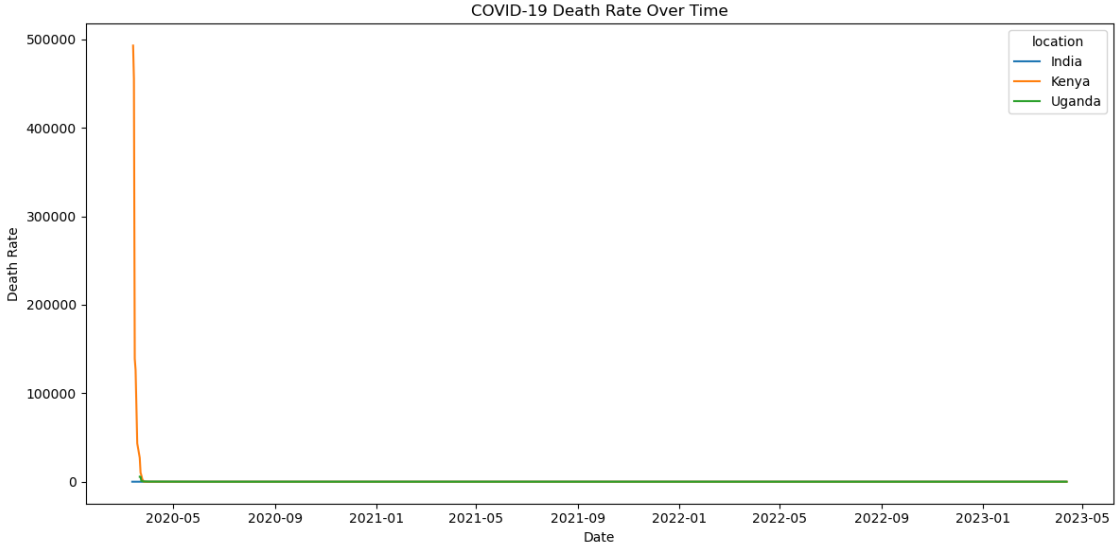
```
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

# Plotting total deaths over time (Line chart)
plt.figure(figsize=(12, 6))
for country in df['location'].unique():
    country_data = df[df['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.grid(True)
plt.tight_layout()
plt.show()

# Comparing daily new cases between countries (Line chart)
plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='date', y='new_cases', hue='location')

plt.title('Daily New COVID-19 Cases by Country')
plt.xlabel('Date')
plt.ylabel('New Cases')
plt.tight_layout()
plt.show()

# Calculating and Plotting Death Rate
df['death_rate'] = df['total_deaths'] / df['total_cases']

plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='date', y='death_rate', hue='location')

plt.title('COVID-19 Death Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Death Rate')
plt.tight_layout()
plt.show()

# Correlation Heatmap
correlation = df[['total_cases', 'total_deaths', 'new_cases', 'new_deaths']].
 ↪corr()

plt.figure(figsize=(8, 6))
sns.heatmap(correlation, annot=True, cmap='coolwarm', fmt='.2f')
```

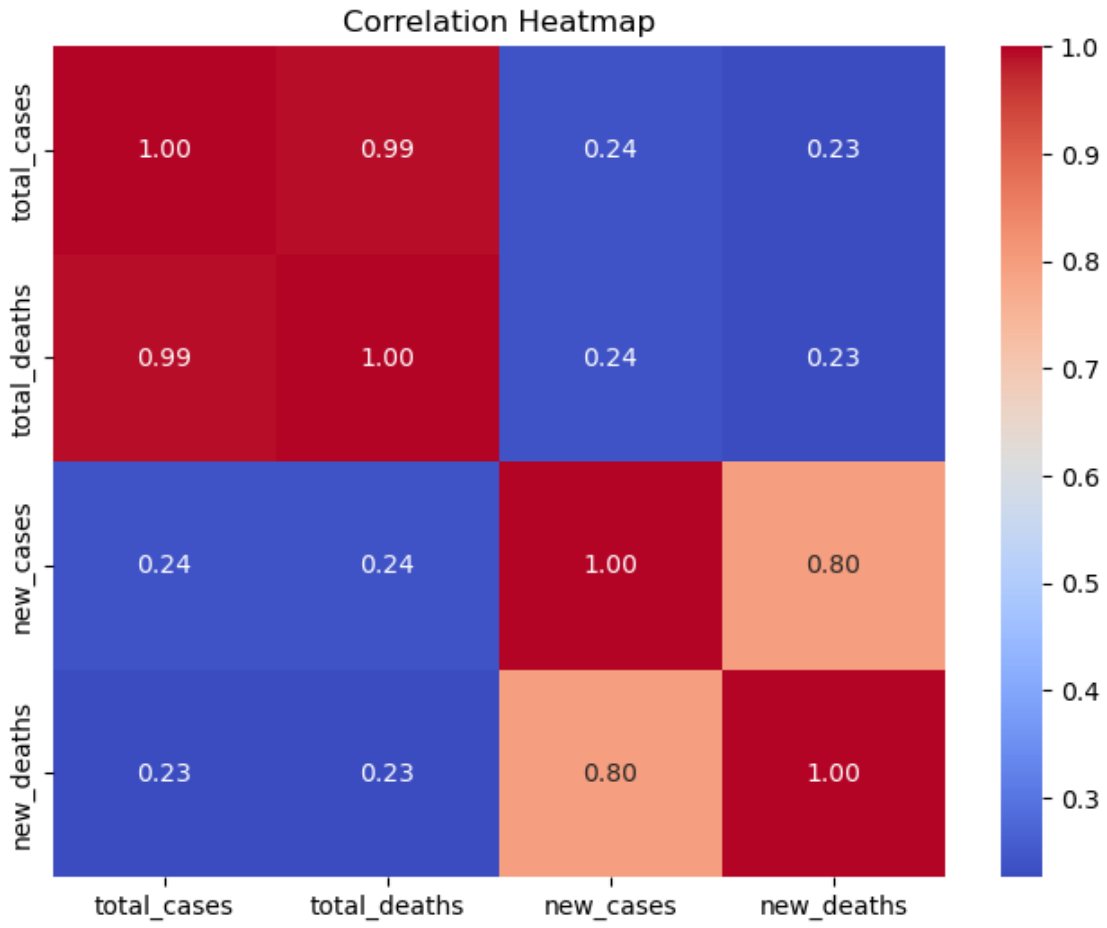
```
plt.title('Correlation Heatmap')
plt.show()
```





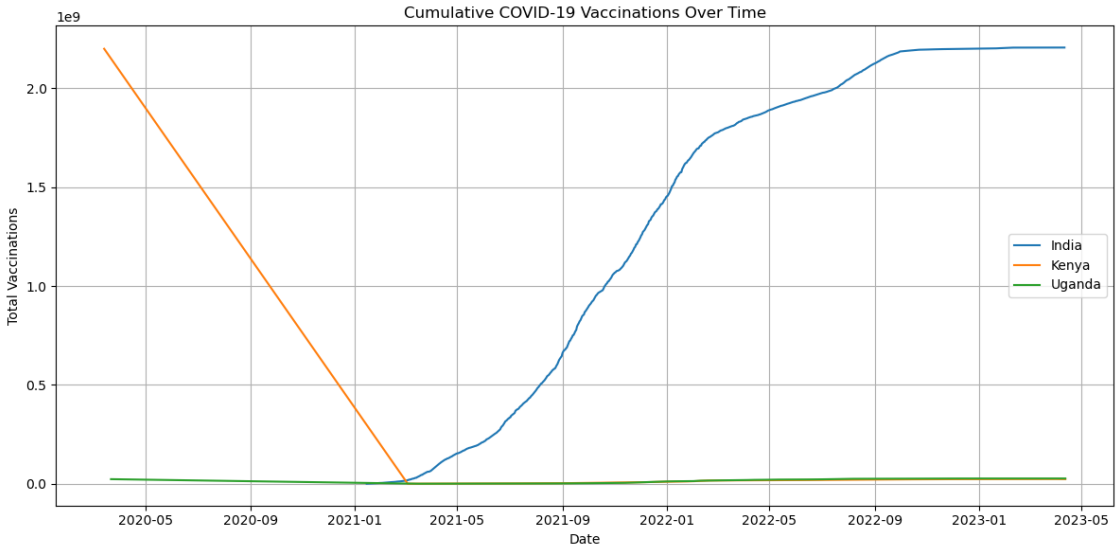

Daily New COVID-19 Cases by Country
location
India
Kenya
Uganda
New Cases
400000
300000
200000
100000
0
2020-01
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01
2023-05
Date

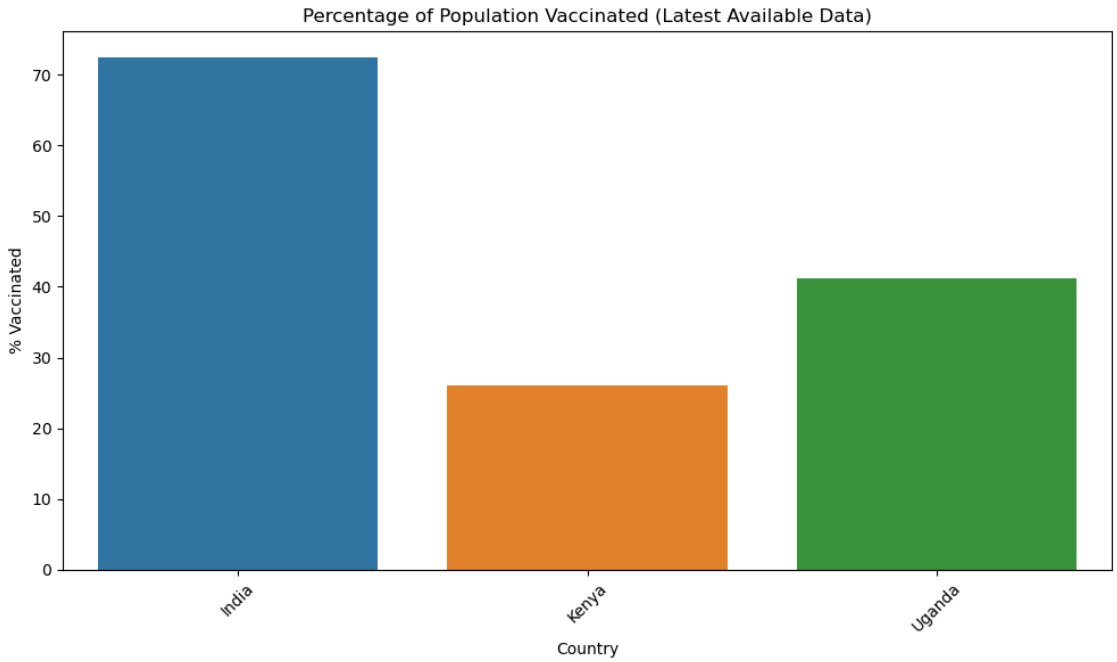

COVID-19 Death Rate Over Time
location
India
Kenya
Uganda
Death Rate
500000
400000
300000
200000
100000
0
2020-05
2020-09
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01
2023-05
Date



[8]:
```
# Plotting cumulative vaccinations over time (Line chart)
plt.figure(figsize=(12, 6))
for country in df['location'].unique():
    country_data = df[df['location'] == country]
    plt.plot(country_data['date'], country_data['total_vaccinations'],␣
 ↪label=country)

plt.title('Cumulative COVID-19 Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

# Comparing % vaccinated population
latest_data = df.sort_values('date').
 ↪dropna(subset=['people_vaccinated_per_hundred']).groupby('location').tail(1)
```

```
plt.figure(figsize=(10, 6))
sns.barplot(data=latest_data, x='location', y='people_vaccinated_per_hundred')
plt.title('Percentage of Population Vaccinated (Latest Available Data)')
plt.xlabel('Country')
plt.ylabel('% Vaccinated')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Pie chart (vaccinated vs. unvaccinated)
country = 'United States'
country_data = df[(df['location'] == country) &␣
 ↪(df['people_vaccinated_per_hundred'].notna())]

if not country_data.empty:
    latest = country_data.sort_values('date').iloc[-1]
    vaccinated = latest['people_vaccinated_per_hundred']
    unvaccinated = 100 - vaccinated

    plt.figure(figsize=(6, 6))
    plt.pie([vaccinated, unvaccinated],
            labels=['Vaccinated', 'Unvaccinated'],
            autopct='%1.1f%%',
            colors=['green', 'red'])
    plt.title(f'Vaccination Distribution in {country}')
    plt.show()
else:
    print(f"No vaccination data available for {country}.")
```





No vaccination data available for United States.

[15]:
```
import pandas as pd
import plotly.express as px

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

# Converting 'date' to datetime
df['date'] = pd.to_datetime(df['date'])

# Getting the latest data for each country
latest_data = df.sort_values('date').drop_duplicates('location', keep='last')

# Dropping rows with missing values in total_cases or iso_code
choropleth_df = latest_data[['location', 'iso_code', 'total_cases']].dropna()

fig = px.choropleth(
    choropleth_df,
    locations="iso_code",
    color="total_cases",
    hover_name="location",
    color_continuous_scale="Reds",
    title="Total COVID-19 Cases by Country (Latest Data)"
```

```
)

fig.show()

# Visualizing Vaccination Rates Instead
choropleth_df = latest_data[['location', 'iso_code',␣
 ↪'people_vaccinated_per_hundred']].dropna()

fig = px.choropleth(
    choropleth_df,
    locations="iso_code",
    color="people_vaccinated_per_hundred",
    hover_name="location",
    color_continuous_scale="Greens",
    title="Vaccination Rate by Country (Latest Data)"
)

fig.show()
```

1 Key Insights

- The United Arab Emirates and Portugal had the highest vaccination rates, with over 95% of the population receiving at least one dose per 100 people.
- India and the United States reported the highest total number of COVID-19 cases, with India showing a strong vaccination effort despite a large population.

- Several African countries reported significantly lower vaccination rates, highlighting regional disparities in vaccine accessibility and distribution.
- European countries like the UK, France, and Italy showed high case numbers and vaccination coverage, suggesting effective national health campaigns.
- Some small countries and territories appeared as outliers, showing extremely high or low figures due to small population sizes or incomplete data reporting.

2 Anomalies & Interesting Patterns

- Missing Data: Countries like North Korea and some African nations have incomplete or missing vaccination data, which may skew global visualizations.
- Small Population Bias: Microstates (e.g., San Marino, Gibraltar) show extremely high per-hundred vaccination rates due to small population sizes.
- High Cases but Low Vaccination: Some countries (e.g., Ukraine, Russia) show a high number of total cases but relatively low vaccination per hundred, indicating possible hesitancy or supply issues.