Covid _insights

May 10, 2025

[16]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

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

```
       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
…         …       …          …         …          …        …
302507      ZWE    Africa     Zimbabwe  2023-04-08     264578.0        2.0
302508      ZWE    Africa     Zimbabwe  2023-04-09     264582.0        4.0
302509      ZWE    Africa     Zimbabwe  2023-04-10     264584.0        2.0
302510      ZWE    Africa     Zimbabwe  2023-04-11     264584.0        0.0
302511      ZWE    Africa     Zimbabwe  2023-04-12     264584.0        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
…                    …           …         …                  …
302507               4.143        5684.0         1.0                0.571
302508               4.714        5684.0         0.0                0.571
302509               4.143        5684.0         0.0                0.571
302510               3.143        5684.0         0.0                0.571
302511               2.571        5684.0         0.0                0.429

        …  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
…     …           …                     …                         …
302507  …          30.7                   36.791                         1.7
302508  …          30.7                   36.791                         1.7
302509  …          30.7                   36.791                         1.7
302510  …          30.7                   36.791                         1.7
302511  …          30.7                   36.791                         1.7

        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
…                 …                      …          …
302507            61.49                    0.571  16320539.0
302508            61.49                    0.571  16320539.0
302509            61.49                    0.571  16320539.0
302510            61.49                    0.571  16320539.0
302511            61.49                    0.571  16320539.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                        NaN                          NaN
1                                        NaN                          NaN
2                                        NaN                          NaN
3                                        NaN                          NaN
4                                        NaN                          NaN
…                                      …                          …
302507                                   NaN                          NaN
302508                                   NaN                          NaN
302509                                   NaN                          NaN
302510                                   NaN                          NaN
302511                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
0                    NaN                                      NaN
1                    NaN                                      NaN
2                    NaN                                      NaN
3                    NaN                                      NaN
4                    NaN                                      NaN
…                  …                                      …
302507               NaN                                      NaN
302508               NaN                                      NaN
302509               NaN                                      NaN
302510               NaN                                      NaN
302511               NaN                                      NaN
```

```
[302512 rows x 67 columns]
```

[17]:

```
#Filter for countries of interest
countries_of_interest = ['Kenya', 'USA', 'India']
filtered_df = df[df['location'].isin(countries_of_interest)]
```

[18]:

```
#Drop rows with missing critical values and make a safe copy
cleaned_df = filtered_df.dropna(subset=['date', 'total_cases', 'total_deaths']).
 ↪copy()
#Convert 'date' column to datetime
cleaned_df['date'] = pd.to_datetime(cleaned_df['date'])
```

[19]:

```
# Interpolate missing numeric values
numeric_columns = cleaned_df.select_dtypes(include='number').columns
cleaned_df[numeric_columns] = cleaned_df[numeric_columns].
 ↪interpolate(method='linear')
```

[20]:

```
print(cleaned_df.isnull().sum())
```

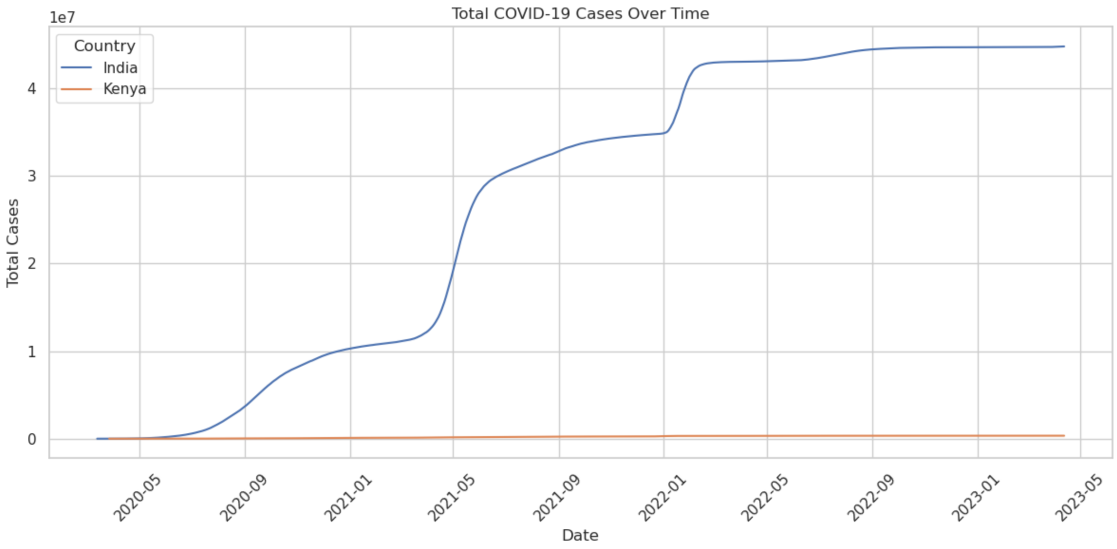
```
iso_code                                      0
continent                                     0
location                                      0
date                                          0
total_cases                                   0
                                           …
population                                    0
excess_mortality_cumulative_absolute       2238
excess_mortality_cumulative                2238
excess_mortality                           2238
excess_mortality_cumulative_per_million    2238
Length: 67, dtype: int64
```

[29]:

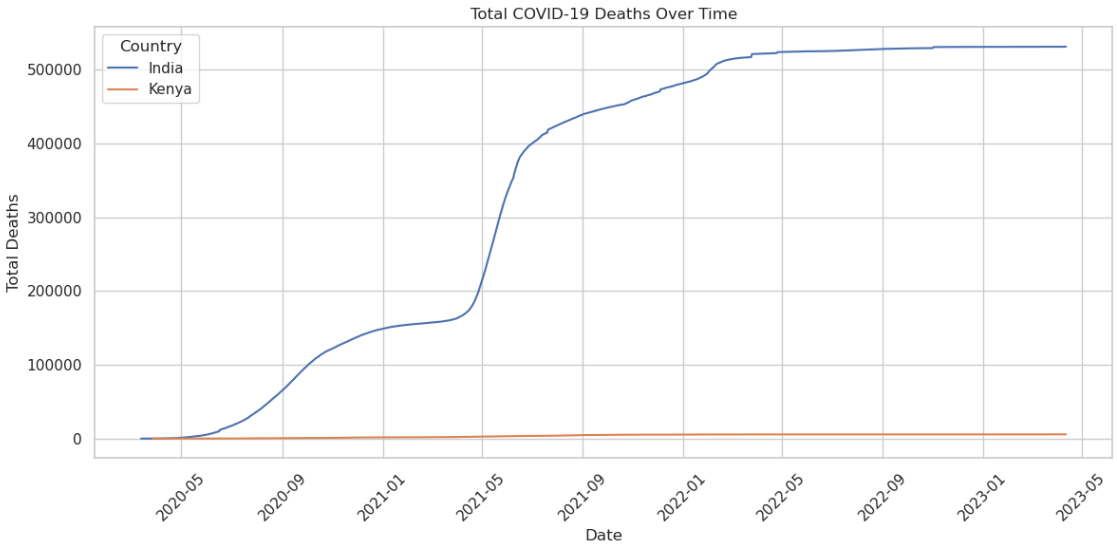
```
# Set style
sns.set(style="whitegrid")
plt.rcParams['figure.figsize'] = (12, 6)
```

[30]:

```
# Plot 1: Total Covid-19 Cases Over Time
plt.figure()
sns.lineplot(data=cleaned_df, x='date', y='total_cases', hue='location')
plt.title('Total COVID-19 Cases Over Time')
plt.xlabel('Date')
plt.ylabel('Total Cases')
plt.legend(title='Country')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

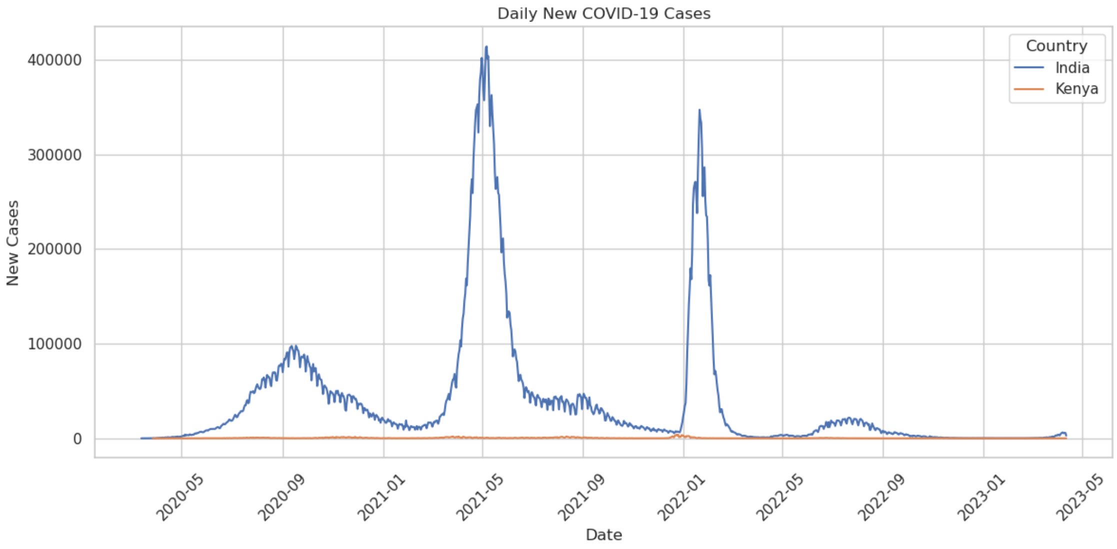


[31]:
```python
# Plot 2: Total Covid-19 Deaths Over Time
plt.figure()
sns.lineplot(data=cleaned_df, x='date', y='total_deaths', hue='location')
plt.title('Total COVID-19 Deaths Over Time')
plt.xlabel('Date')
plt.ylabel('Total Deaths')
plt.legend(title='Country')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[32]:

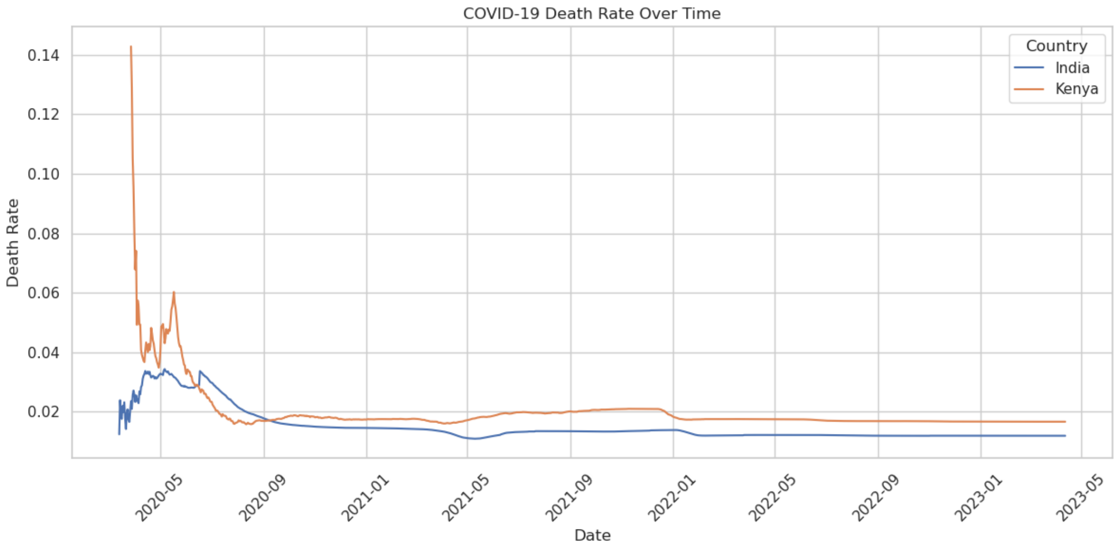
```
# Plot 3: Daily New Cases
plt.figure()
sns.lineplot(data=cleaned_df, x='date', y='new_cases', hue='location')
plt.title('Daily New COVID-19 Cases')
plt.xlabel('Date')
plt.ylabel('New Cases')
plt.legend(title='Country')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[33]:

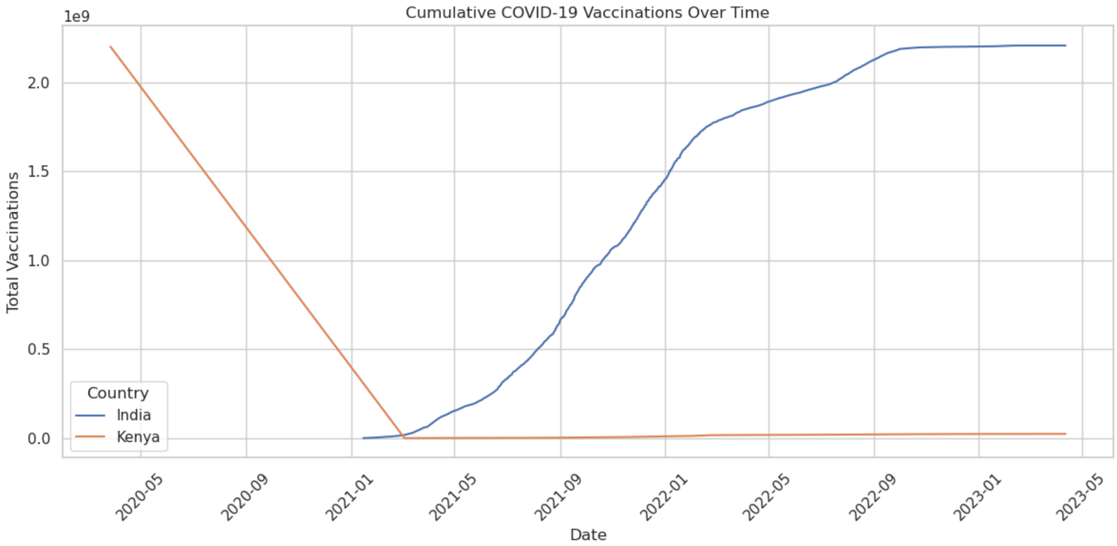
```
# Plot 4: Death Rate Over Time
cleaned_df['death_rate'] = cleaned_df['total_deaths] /␣
 ↪cleaned_df['total_cases']

plt.figure()
sns.lineplot(data=cleaned_df, x='date', y='death_rate', hue='location')
plt.title('COVID-19 Death Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Death Rate')
plt.legend(title='Country')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[34]:

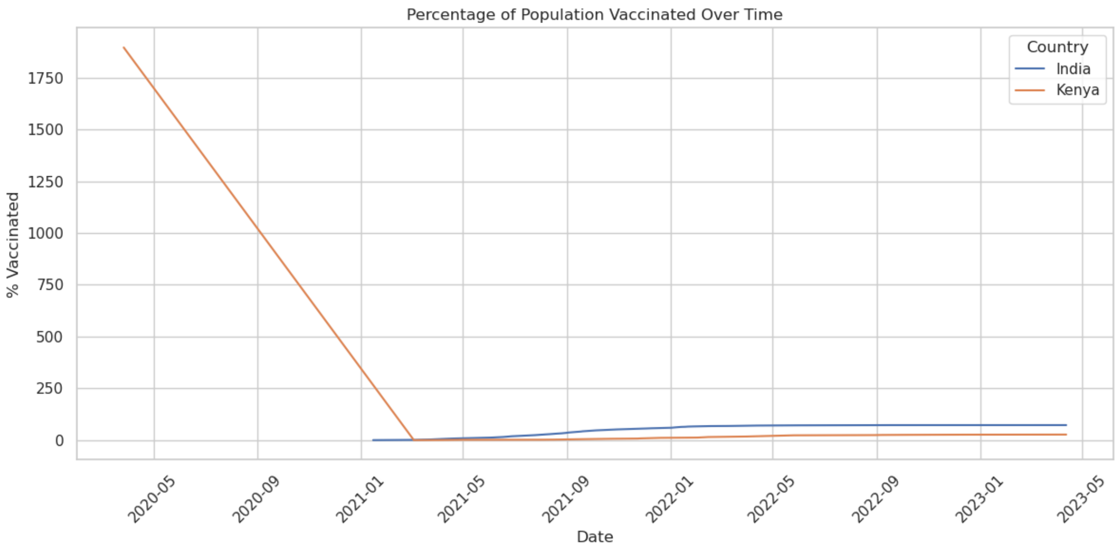
```
 #Visualizing Vaccination Progress
# Plot 1: Cumulative Total Vaccinations Over Time
plt.figure()
sns.lineplot(data=cleaned_df, x='date', y='total_vaccinations', hue='location')
plt.title('Cumulative COVID-19 Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.legend(title='Country')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[36]:

```
# Plot 2: Percentage of Population Vaccinated Over Time
# Create percent_vaccinated column
cleaned_df.loc[:, 'percent_vaccinated'] = (cleaned_df['people_vaccinated'] /␣
 ↪cleaned_df['population']) * 100

plt.figure()
sns.lineplot(data=cleaned_df, x='date', y='percent_vaccinated', hue='location')
plt.title('Percentage of Population Vaccinated Over Time')
plt.xlabel('Date')
plt.ylabel('% Vaccinated')
plt.legend(title='Country')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



[35]:

```
# Prepare the dataset
df['date'] = pd.to_datetime(df['date'])
df = df.sort_values('date')
df = df.dropna(subset=['iso_code'])
```

[ ]:

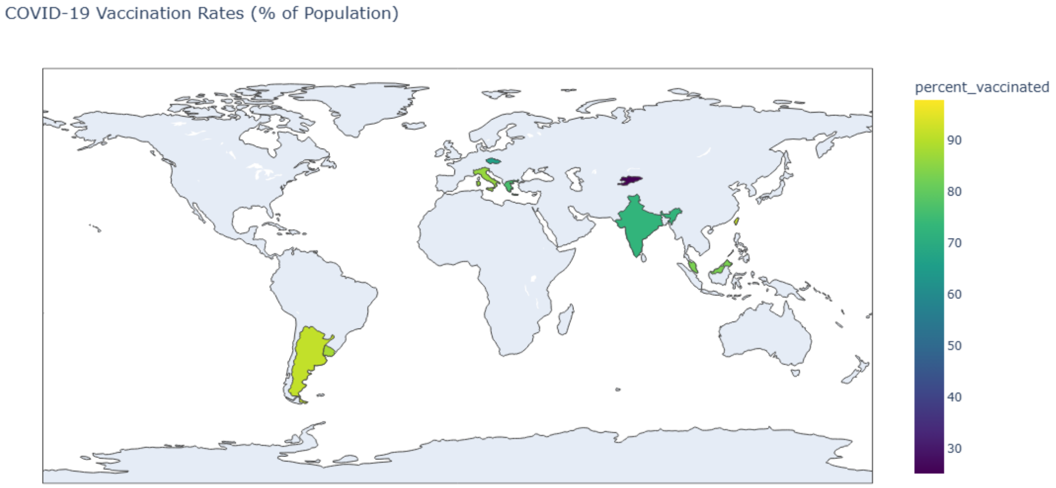
```
# Get the latest entry for each country
latest_df = df.groupby('location').tail(1).copy()
```

[ ]:

```
# Calculate percent vaccinated
latest_df['percent_vaccinated'] = (latest_df['people_vaccinated'] /␣
 ↪latest_df['population']) * 100
```

[27]:

```
# Choropleth Map: Percent Vaccinated by Country
fig = px.choropleth(
    latest_df,
    locations="iso_code",
    color="percent_vaccinated",
    hover_name="location",
    color_continuous_scale="Viridis",
    title="COVID-19 Vaccination Rates (% of Population)",
    width=1000,
    height=600
)
fig.show()
```



Key Insights from the COVID-19 Data

India Experienced the Highest Number of Total Cases Among the Selected Countries From the trend of total COVID-19 cases over time, India consistently showed a steep rise, eventually surpassing the USA and Kenya. This reflects both its large population and broad testing/reporting scale.

The USA Reported the Highest Death Toll While India had more total cases, the USA reported significantly higher total deaths, suggesting either a more vulnerable population or differences in healthcare response and data reporting.

Kenya Had the Lowest Vaccination Coverage By the end of the observed period, Kenya's percent vaccinated remained considerably lower than that of the USA and India, highlighting disparities in vaccine access or uptake.

Death Rates Showed Decline Over Time Across All Countries The calculated death rate (total_deaths / total_cases) decreased over time for all three countries, indicating improved treatment protocols, earlier detection, and vaccine effectiveness.

Sharp Surges in New Cases Suggest Variant Waves The daily new cases plot shows periodic spikes, likely corresponding to COVID-19 variant waves (such as Delta or Omicron), especially prominent in India and the USA.

[ ]: