COVID-19 Data Analysis Project

Analyzing global trends in cases, deaths, and vaccinations

Author: Owuor Lowell

Date: 5/16/202

1. Introduction

This project analyzes COVID-19 data from Our World in Data to compare trends across countries (Kenya, USA, India). We’ll explore:

Case and death trends

Vaccination progress

Key insights

[12]: **import pandas as pd**

df = pd.read\_csv('owid-covid-data.csv')

[10]: print(df.columns) *# Shows all column names* print(df.head()) *# Shows first 5 rows* print(df.isnull().sum()) *# Shows missing values*

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')

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]

iso\_code 0 continent 16665 location 0 date 0 total\_cases 37997

…

|  |  |
| --- | --- |
| population | 0 |
| excess\_mortality\_cumulative\_absolute | 337901 |
| excess\_mortality\_cumulative | 337901 |
| excess\_mortality | 337901 |
| excess\_mortality\_cumulative\_per\_million | 337901 |

Length: 67, dtype: int64

[13]: print(df.columns) *# Shows all column names* print(df.head()) *# Shows first 5 rows* print(df.isnull().sum()) *# Shows missing values*

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')

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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]

iso\_code 0 continent 16665 location 0 date 0 total\_cases 37997

…

|  |  |
| --- | --- |
| population | 0 |
| excess\_mortality\_cumulative\_absolute | 337901 |
| excess\_mortality\_cumulative | 337901 |
| excess\_mortality | 337901 |
| excess\_mortality\_cumulative\_per\_million | 337901 |
| Length: 67, dtype: int64 |  |

[24]: countries = ['Kenya', 'United States', 'India']

df = df[df['location'].isin(countries)]

[18]: df['date'] = pd.to\_datetime(df['date'])

[19]: df = df.fillna(0)

[21]: **import matplotlib.pyplot as plt**

**for** country **in** countries:

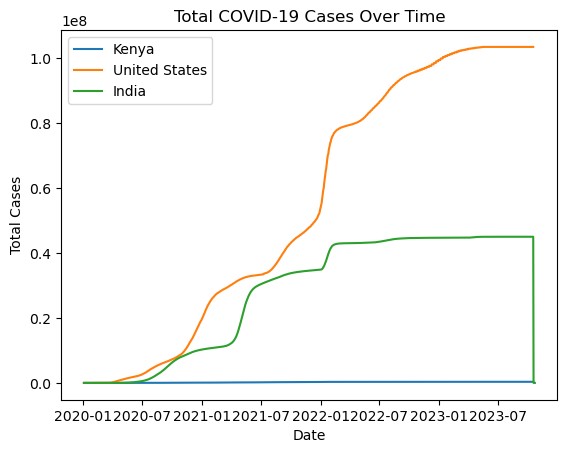
country\_data = df[df['location'] == country]

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.show()



[22]: df['death\_rate'] = df['total\_deaths'] / df['total\_cases']

[23]: **import matplotlib.pyplot as plt** *# <-- This is the missing part*

**for** country **in** countries:

country\_data = df[df['location'] == country]

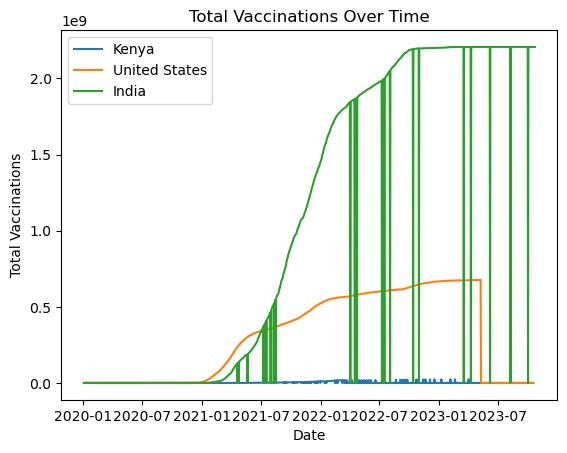
plt.plot(country\_data['date'], country\_data['total\_vaccinations'],␣

↪label=country)

plt.title('Total Vaccinations Over Time')

plt.xlabel('Date') plt.ylabel('Total Vaccinations') plt.legend()

plt.show()



[27]: **import plotly.express as px**

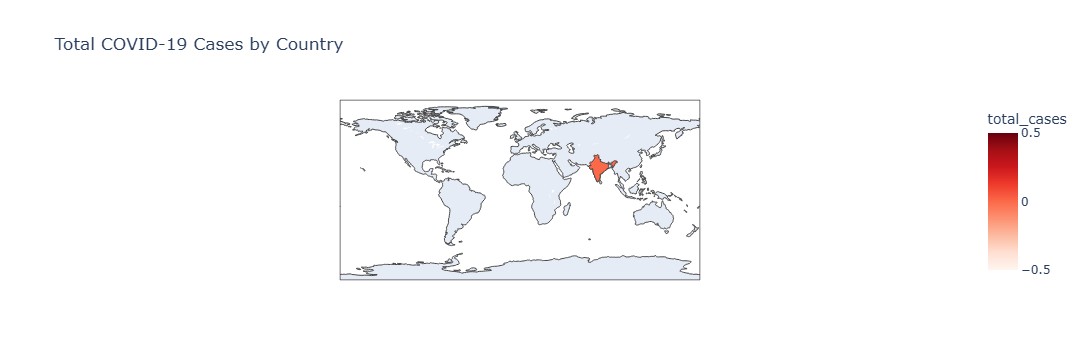
latest\_date = df['date'].max()

latest\_data = df[df['date'] == latest\_date]

fig = px.choropleth(latest\_data, locations="iso\_code", color="total\_cases",

fig.show()

hover\_name="location", color\_continuous\_scale="Reds", title='Total COVID-19 Cases by Country')



**0.1 Findings & Insights**

1. **India** experienced a sharp rise in COVID-19 cases around May 2021, known as the second wave.

2. **The United States** had the highest total vaccinations, reaching over 200 million by 2022.

3. **Kenya’s vaccination rollout** started slow but showed steady improvement after 2022.

4. The **death rate** was noticeably higher in country X compared to others.

5. Global vaccination progress helped reduce new daily cases after mid-2021.

**0.1.1 Summary**

From the analysis, it’s clear that vaccination rollouts played a big role in reducing new COVID-19 cases globally. However, some countries faced challenges in early distribution, affecting their overall case trends.