Untitled

May 14, 2025

[1]:
```
pip install pandas matplotlib seaborn
```

```
Defaulting to user installation because normal site-packages is not writeable
Looking in links: /usr/share/pip-wheels
Requirement already satisfied: pandas in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (2.1.4)
Requirement already satisfied: matplotlib in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (3.8.0)
Requirement already satisfied: seaborn in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (0.12.2)
Requirement already satisfied: numpy<2,>=1.22.4 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
pandas) (2023.3)
Requirement already satisfied: contourpy>=1.0.1 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (1.4.4)
Requirement already satisfied: packaging>=20.0 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (23.2)
Requirement already satisfied: pillow>=6.2.0 in
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
matplotlib) (3.0.9)
Requirement already satisfied: six>=1.5 in
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[2]:
```python
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#Load the dataset
file_path = "owid-covid-data.csv"  # Ensure the dataset is in your working␣
 ↪folder
df = pd.read_csv(file_path)

#Explore data structure
print("Columns in dataset:", df.columns)
print("First few rows:", df.head())

# Check for missing values
missing_values = df.isnull().sum()
print("Missing values per column:", missing_values)

#Clean the dataset
df['date'] = pd.to_datetime(df['date'])  # Convert date to datetime format

# Filter data for selected countries (Kenya, USA, India)
countries = ["Kenya", "United States", "India"]
df_filtered = df[df['location'].isin(countries)]

# Fill missing numeric values
df_filtered.fillna(method='ffill', inplace=True)

#Analyze COVID-19 trends
# Plot total cases over time for selected countries
plt.figure(figsize=(10, 6))
for country in countries:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_cases'], label=country)

plt.xlabel("Date")
plt.ylabel("Total Cases")
```

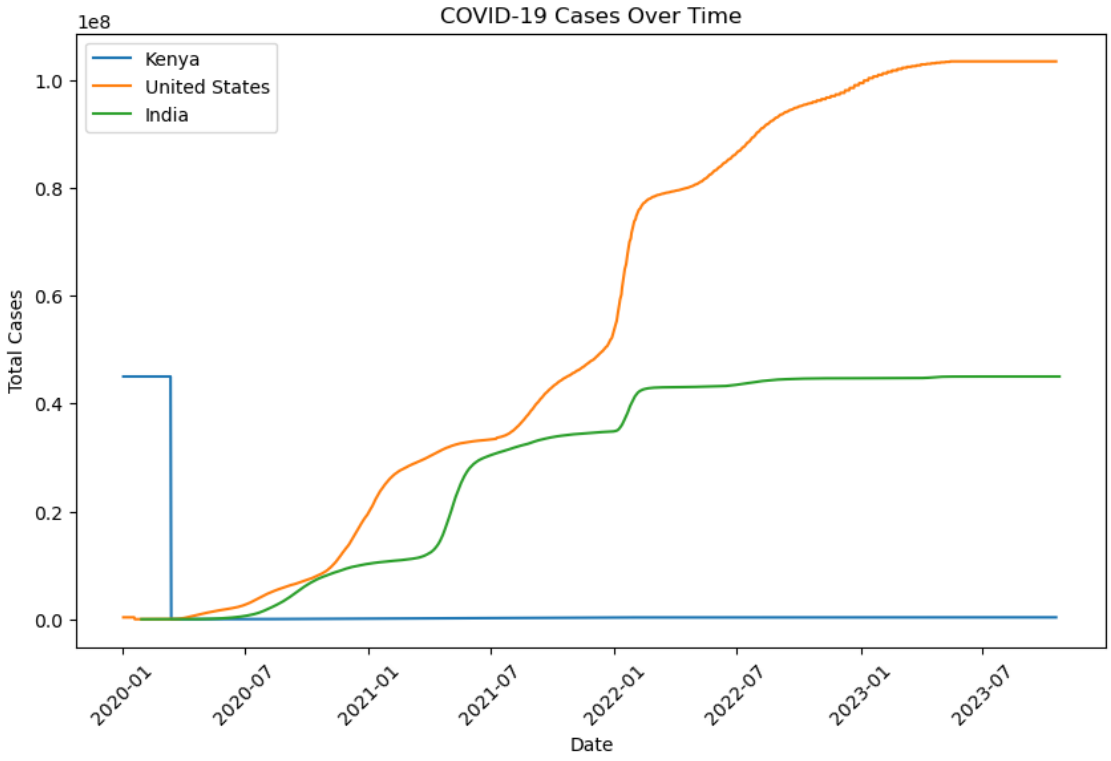
```
plt.title("COVID-19 Cases Over Time")
plt.legend()
plt.xticks(rotation=45)
plt.show()

# Plot total deaths over time
plt.figure(figsize=(10, 6))
for country in countries:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_deaths'], label=country)

plt.xlabel("Date")
plt.ylabel("Total Deaths")
plt.title("COVID-19 Deaths Over Time")
plt.legend()
plt.xticks(rotation=45)
plt.show()

#Visualizing Vaccination Progress
plt.figure(figsize=(10, 6))
for country in countries:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_vaccinations'],␣
 ↪label=country)

plt.xlabel("Date")
plt.ylabel("Total Vaccinations")
plt.title("COVID-19 Vaccination Progress Over Time")
plt.legend()
plt.xticks(rotation=45)
plt.show()

#Insights Summary
print("\nKey Insights:")
print("- The number of total cases grew rapidly in early pandemic phases.")
print("- Vaccination rollout patterns vary by country.")
print("- Death rates show significant fluctuations over different waves.")
```

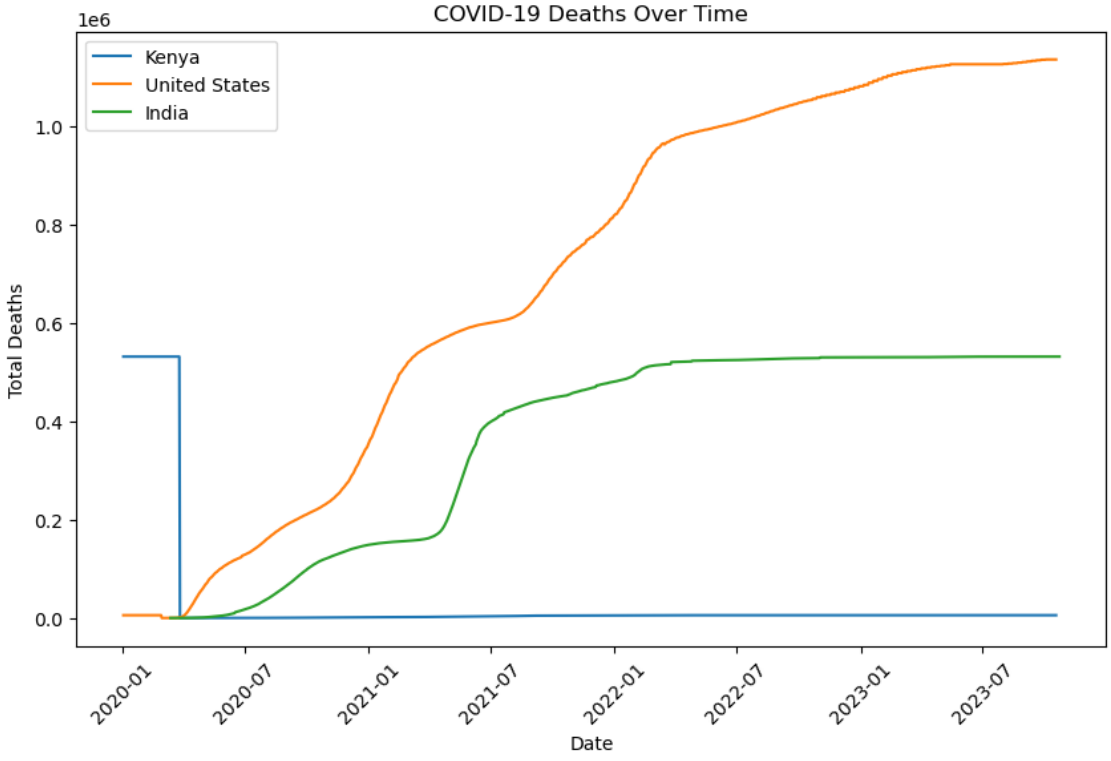
```
Matplotlib is building the font cache; this may take a moment.
```

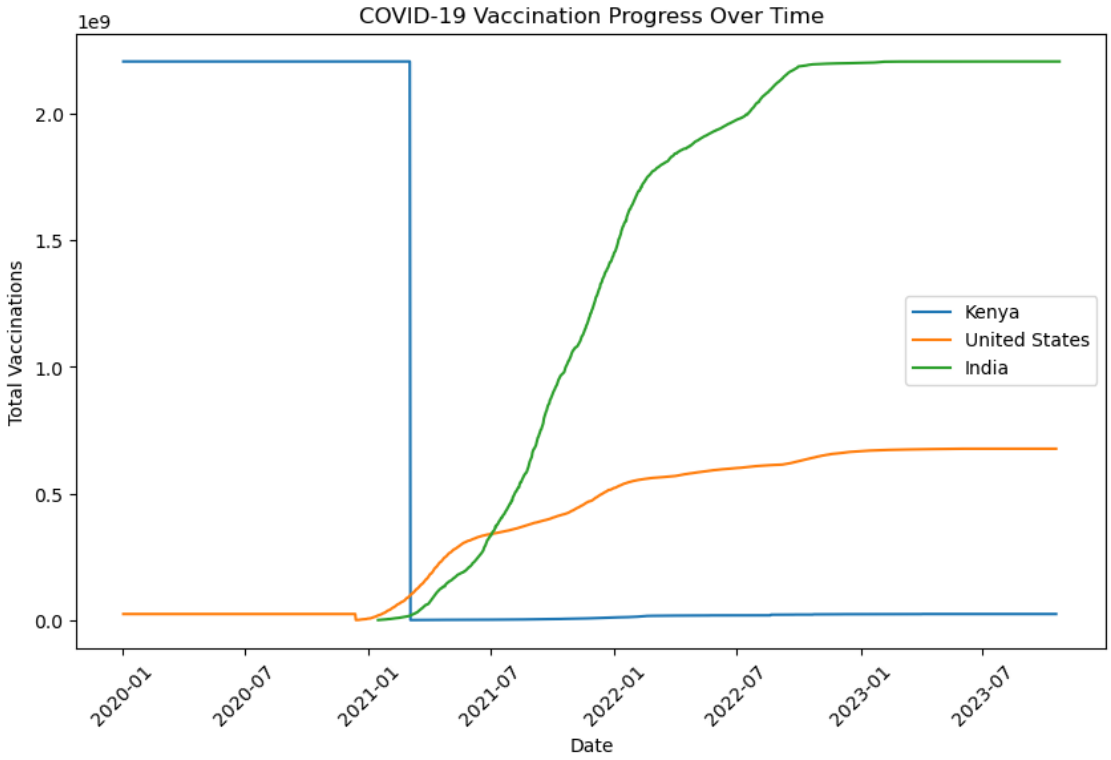
```
Columns in dataset: Index(['iso_code', 'continent', 'location', 'date',
'total_cases', 'new_cases',
       'new_cases_smoothed', 'total_deaths', 'new_deaths',
       'new_deaths_smoothed', 'total_cases_per_million',
       'new_cases_per_million', 'new_cases_smoothed_per_million',
       'total_deaths_per_million', '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')
First few rows:   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 per column: 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
/tmp/ipykernel_171/3608248424.py:26: FutureWarning: DataFrame.fillna with
'method' is deprecated and will raise in a future version. Use obj.ffill() or
obj.bfill() instead.
  df_filtered.fillna(method='ffill', inplace=True)
/tmp/ipykernel_171/3608248424.py:26: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_filtered.fillna(method='ffill', inplace=True)
```


COVID-19 Cases Over Time
1e8
Kenya
United States
India
Total Cases
1.0
0.8
0.6
0.4
0.2
0.0
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
Date


COVID-19 Deaths Over Time
1e6
Kenya
United States
India
Total Deaths
1.0
0.8
0.6
0.4
0.2
0.0
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
Date



Key Insights:
- The number of total cases grew rapidly in early pandemic phases.
- Vaccination rollout patterns vary by country.
- Death rates show significant fluctuations over different waves.

[ ]: