

final python project

May 13, 2025

```
[ ]: import pandas as pd

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

print(df.columns)

print(df.head())

print(df.isnull().sum())

[ ]: countries = ['Kenya', 'United States', 'India']
df_countries = df[df['location'].isin(countries)]
print(df_countries['location'].unique())

df_countries = df_countries.dropna(subset=['date', 'total_cases', 'total_deaths'])

df_countries['date'] = pd.to_datetime(df_countries['date'])

numeric_cols = ['total_cases', 'total_deaths', 'new_cases', 'new_deaths', 'total_vaccinations']
df_countries[numeric_cols] = df_countries[numeric_cols].interpolate()

[ ]: import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12,6))
sns.lineplot(data=df_countries, 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.show()

plt.figure(figsize=(12,6))
sns.lineplot(data=df_countries, 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.show()

plt.figure(figsize=(12,6))
sns.lineplot(data=df_countries, 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.show()

df_countries['death_rate'] = df_countries['total_deaths'] /
    ↪df_countries['total_cases']

plt.figure(figsize=(12,6))
sns.lineplot(data=df_countries, 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.show()

```

```

[ ]: plt.figure(figsize=(12,6))
sns.lineplot(data=df_countries, 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.show()

if 'people_vaccinated_per_hundred' in df_countries.columns:
    plt.figure(figsize=(12,6))
    sns.lineplot(data=df_countries, x='date',
    ↪y='people_vaccinated_per_hundred', hue='location')
    plt.title('Percentage of Population Vaccinated Over Time')
    plt.xlabel('Date')
    plt.ylabel('People Vaccinated per Hundred')

```

```

plt.legend(title='Country')
plt.show()
else:
    print("Column 'people_vaccinated_per_hundred' not found in dataset.")

latest = df_countries.sort_values('date').groupby('location').tail(1)
for idx, row in latest.iterrows():
    vaccinated = row.get('people_vaccinated', None)
    population = row.get('population', None)
    if vaccinated and population:
        labels = ['Vaccinated', 'Unvaccinated']
        sizes = [vaccinated, population - vaccinated]
        plt.figure(figsize=(5,5))
        plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140)
        plt.title(f'Vaccinated vs. Unvaccinated in {row["location"]}')
        plt.show()

```

```

[ ]: latest = df.sort_values('date').groupby('location').tail(1)
top_countries = latest.nlargest(10, 'total_cases')

plt.figure(figsize=(12,6))
sns.barplot(data=top_countries, x='location', y='total_cases')
plt.title('Top 10 Countries by Total COVID-19 Cases')
plt.xlabel('Country')
plt.ylabel('Total Cases')
plt.xticks(rotation=45)
plt.show()

plt.figure(figsize=(10,6))
sns.heatmap(df_countries[['total_cases', 'total_deaths', 'new_cases',
    ↪ 'new_deaths', 'total_vaccinations']].corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()

```

```

[ ]: import plotly.express as px

latest = df.sort_values('date').groupby('location').tail(1)

latest = latest[latest['iso_code'].str.len() == 3]

fig = px.choropleth(
    latest,

```

```

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

fig = px.choropleth(
    latest,
    locations="iso_code",
    color="people_vaccinated_per_hundred",
    hover_name="location",
    color_continuous_scale="Greens",
    title="COVID-19 Vaccination Rate (%) by Country (Latest)"
)
fig.show()

```