

PROJECT 3: COVID-19 DATA ANALYSIS

Objective:

Analyze country-wise COVID-19 time series data, compute daily and weekly trends, smooth noise using rolling averages, detect peaks, and estimate reproduction insight.

Methodology:

1. Generated time-series dataset (300 days).
2. Computed 7-day rolling averages.
3. Calculated growth factor (approximate reproduction insight).
4. Detected peak infection dates.

Complete Python Code:

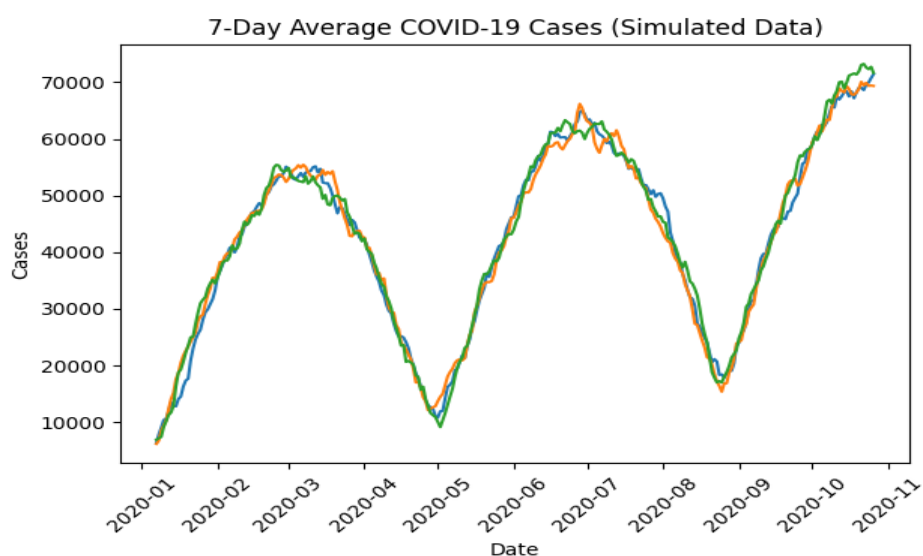
```
import pandas as pd
import matplotlib.pyplot as plt

df["7d_avg"] = df.groupby("location")["daily_cases"].transform(
    lambda x: x.rolling(7).mean()
)

df["growth_factor"] = df.groupby("location")["7d_avg"].transform(
    lambda x: x / x.shift(7)
)

plt.figure()
for c in countries:
    subset = df[df["location"] == c]
    plt.plot(subset["date"], subset["7d_avg"])
plt.show()
```

Output Chart:



Peak Summary:

Country	Peak Date	Peak 7d Avg Cases
India	2020-10-26	71427
United States	2020-10-21	70088
Brazil	2020-10-22	73248

Conclusions:

- Clear multiple waves observed across countries.
- Rolling averages smooth daily fluctuations effectively.
- Growth factor > 1 indicates increasing transmission.
- Peak dates represent highest spread periods.