COVID-19 Data Analysis Project (Intermediate Level)

1. Data Loading & Cleaning

- Load the CSV using pandas.
- Check for missing values. Handle them properly (e.g., fill, drop, or flag).
- Convert any date/time columns to datetime format.
- Check the data types of all columns. Fix any incorrect ones.

Hint: What's one surprising thing you noticed about the raw data?

2. Basic Analysis

- Total number of cases, deaths, and recoveries.
- Countries with the highest number of cases and deaths.
- Mortality rate = (deaths / cases) 100 for each country.
- Recovery rate = (recoveries / cases) 100.

Question: Which country had the highest mortality rate?

3. Time Series Trends

- Plot daily new cases and deaths.
- Plot cumulative cases over time.
- Calculate 7-day moving averages for smoothing trends.

Hint: Use rolling(window=7).mean()

4. Visualizations (matplotlib & seaborn)

- Heatmap: Correlation between numerical columns.
- Barplot: Top 10 countries by total cases.
- Lineplot: Cases over time for 3 selected countries.
- KDE Plot or Histogram: Distribution of daily new cases.

5. Advanced Task (Intermediate+)

Choose one:

- Compare the growth rate of cases between two countries.
- Cluster countries based on total cases, deaths, and recoveries (using KMeans).
- Forecast future case numbers using a basic moving average model.