

Use Case 1: Data Exploration and Cleaning

Objective:

Explore the dataset to understand its structure, identify missing values, and clean the data by handling missing values and outliers.

Tasks:

1. **Inspect the Data:** Check the first few rows, data types, and basic statistics.
2. **Handle Missing Values:** Identify and handle missing values in the 'Confirmed', 'Deaths', and 'Recovered' columns.

ex: Fill missing values in 'Confirmed', 'Deaths', 'Recovered' with 0

3. **Remove or Cap Outliers:** Identify extreme outliers in the 'Confirmed' column and handle them appropriately.

ex: 3. Identify and handle outliers in 'Confirmed' column , Assume any value greater than 500,000 is an outlier

Deliverables:

- Cleaned dataset without missing values or extreme outliers.
- Summary of how missing values and outliers were handled.

Use Case 2: Country-Level COVID-19 Analysis

Objective:

Analyze the COVID-19 situation at the country level by aggregating the data to find the total confirmed cases, deaths, and recovered cases for each country.

Tasks:

1. **Group Data by Country:** Aggregate data to calculate total confirmed cases, deaths, and recovered cases for each country.
2. **Rank Countries by Cases:** Identify the top 10 countries with the highest number of confirmed cases.

3. **Calculate Fatality Rate:** For each country, calculate the fatality rate (deaths/confirmed cases)

Deliverables:

- A summary of total confirmed cases, deaths, and recoveries for each country.
- A list of the top 10 countries by confirmed cases.
- A calculated fatality rate for each country.

Use Case 3: Trend Analysis for Specific Country

Objective:

Analyze the trend of confirmed cases, deaths, and recoveries over time for a specific country (e.g., India).

Tasks:

1. **Filter Data by Country:** Extract data for a specific country (e.g., India).
2. **Group by Date:** Aggregate data by date to calculate the daily confirmed, deaths, and recovered cases.
3. **Visualize Trends:** Plot the trend of confirmed cases, deaths, and recoveries over time using a line plot.

Deliverables:

- A line plot showing the trend of confirmed cases, deaths, and recoveries in India over time.

Use Case 4: Time-Series Resampling and Rolling Averages

Objective:

Resample the data to a weekly frequency and calculate rolling averages to smoothen fluctuations in daily confirmed cases.

Tasks:

1. **Resample Data:** Convert the daily data into weekly aggregated data (sum of cases per week).
2. **Calculate Rolling Average:** Compute the 7-day rolling average for confirmed cases.
3. **Visualize the Rolling Average:** Plot the rolling average to show smoothed trends over time

Deliverables:

- A weekly aggregated dataset.
- A plot showing the 7-day rolling average of confirmed cases, smoothing out daily fluctuations

Use Case 5: Merging External Data

Objective:

Merge external population data to the existing COVID-19 dataset and calculate the confirmed cases per capita for each country.

Tasks:

1. **Merge Population Data:** Load external population data and merge it with the COVID-19 dataset.
2. **Calculate Cases per Capita:** Compute confirmed cases per capita for each country.
3. **Rank Countries:** Rank countries by confirmed cases per capita.

Deliverables:

- A merged dataset containing population and COVID-19 data.
- A list of the top 10 countries by confirmed cases per capita

Use Case 6: Creating a Final Report

Objective:

Create a final summary report of key metrics, including total confirmed cases, deaths, recovered cases, and fatality rates per country. Export the results to a CSV file.

Tasks:

1. **Summarize Key Metrics:** For each country, calculate total confirmed cases, deaths, recoveries, and fatality rate.
2. **Export to CSV:** Save the final summarized data to a CSV file for further analysis

Deliverables:

- A CSV file (`yourname_hashInclude.csv`) containing key metrics (confirmed cases, deaths, recovered cases, and fatality rates) for each country.

