# COVID-19 Global Tracker Project

## 1. Objective

Track the spread, impact, and recovery trends of COVID-19 across countries using ETL workflow and visualization in Power BI & Tableau.

## 2. Data Sources

- Johns Hopkins CSSE COVID-19 Time Series Data (Confirmed, Deaths, Recovered)
- WHO daily situation reports
- Kaggle curated datasets
- Our World in Data (Vaccination Data)
- Population Data (World Bank / UN)

## 3. ETL Workflow

- **Extract:** Download daily CSVs via GitHub API.
- **Transform:** Standardize country names, unpivot date columns, calculate daily new cases and 7-day averages, handle missing data.
- **Load:** Export cleaned data into Power BI / Tableau-friendly fact table.

## 4. Data Model

Star Schema with dimensions and fact table:

| Table | Key Fields |
|---|---|
| dim_date | date_id, year, month, day, quarter |
| dim_country | country_id, country_name, iso3, population |
| fact_covid_cases | date_id, country_id, confirmed, deaths, recovered, new_confirmed, new_deaths, active_cases |

## 5. Visualizations in Power BI

- Global map heatmap (Confirmed per 100k population)
- Line chart of Daily New Cases with 7-day average
- KPI Cards: Total Confirmed, Total Deaths, Recovery Rate, Active Cases
- Bar chart of Top 10 countries by cases
- Drillthrough report for individual countries

## 6. Sample DAX Measures

```
Total Confirmed = SUM(fact_covid_cases[confirmed_cumulative])
Total Deaths = SUM(fact_covid_cases[deaths_cumulative])
```

```
Active Cases = [Total Confirmed] - [Total Deaths] -
SUM(fact_covid_cases[recovered_cumulative])
Case Fatality Rate = DIVIDE([Total Deaths], [Total Confirmed], 0)
Recovery Rate = DIVIDE(SUM(fact_covid_cases[recovered_cumulative]), [Total
Confirmed], 0)
Seven Day Avg =
AVERAGEX(
DATESINPERIOD(Date[date], LASTDATE(Date[date]), -6, DAY),
SUM(fact_covid_cases[new_confirmed])
)
```

## 7. Visualizations in Tableau

Map: Country colored by total confirmed cases or per-100k

Trendline of Daily New Cases + 7-day average overlay

KPIs: Confirmed, Deaths, Recovery Rate

Dashboard with interactive filters (Country, Date range)

## 8. Tableau Calculated Fields

```
Daily New Cases = SUM([Confirmed]) - LOOKUP(SUM([Confirmed]), -1)
7-Day Average = WINDOW_AVG([Daily New Cases], -6, 0)
Active Cases = SUM([Confirmed]) - SUM([Deaths]) - SUM([Recovered])
Case Fatality Rate = SUM([Deaths]) / SUM([Confirmed])
Recovery Rate = SUM([Recovered]) / SUM([Confirmed])
```

## 9. Challenges & Pitfalls

Inconsistent country names across sources

Missing recovered data in some regions

Data gaps and reporting delays

Performance issues with large daily-level datasets

## 10. Extensions / Advanced Ideas

Add vaccination data (Our World in Data)

Forecast trends using machine learning (Prophet / ARIMA)

Per-region drilldowns (states, provinces)

Anomaly detection on daily spikes