**🧩 Project #2: COVID-19 Global Analysis**

**🎯 Goal:** Analyze confirmed cases, deaths, recoveries, vaccination trends.  
**📦 Dataset:** [Our World in Data — COVID-19](https://github.com/owid/covid-19-data)  
(Most use the owid-covid-data.csv file)

**✅ Main Tools for This Project**

* **Python:** pandas, numpy, plotly
* **Optional:** SQL to practice storing & querying
* **Optional:** Tableau if you want extra visuals for your portfolio

**📌 Key Problems to Solve & What You’ll Learn**

Below is a practical breakdown, **stepping up from the Netflix project**:

**🔑 1️⃣ Data Cleaning & Preparation**

**Problems to Solve:**

* Are there missing values for cases, deaths, or vaccinations?
* Are date columns consistent?
* Are country names consistent?
* Are daily vs cumulative values clear?

**Expectations:**  
✔️ Handle missing data (fill with 0 or NaN carefully)  
✔️ Convert date columns to datetime  
✔️ Drop any duplicate rows  
✔️ Keep only columns you really need: e.g., country, date, cases, deaths, vaccinations.

**🔑 2️⃣ Descriptive Analysis**

**Problems to Solve:**

* Total cases & deaths worldwide
* Top 10 countries by total cases & deaths
* Calculate **daily new cases** and **daily new deaths**
* Highest single-day spike globally
* Average daily new cases over time

**Expectations:**  
✔️ Use groupby, sum, mean, max  
✔️ Learn diff() for daily change  
✔️ Handle outliers or data errors

**🔑 3️⃣ Time-Series Trends**

**Problems to Solve:**

* Global trend of cases over time (line chart)
* Compare top 5 countries’ trends on one chart
* Vaccination trend over time
* Mortality rate over time (deaths / cases)

**Expectations:**  
✔️ Use plotly for interactive line charts  
✔️ Add hover info, labels, and filters

**🔑 4️⃣ Optional SQL Storage**

**Problems to Solve:**

* Store daily COVID-19 stats in PostgreSQL
* Practice SELECT, GROUP BY, ORDER BY
* Query total cases by continent or region

**Expectations:**  
✔️ Good practice for joins if you add population or region table  
✔️ Try window functions for running totals

**🔑 5️⃣ Optional Tableau Dashboard**

If you’d like a **Tableau version**, try:

* Global trend line
* Top countries by cases & deaths
* World map with total cases (color gradient)
* Filters by country and date

**📦 Deliverables**

✅ Cleaned CSV  
✅ Jupyter Notebook with:

* Data cleaning
* Descriptive stats
* Plotly charts  
  ✅ SQL script to create & load the table  
  ✅ Tableau .twbx file (if you do it)  
  ✅ 1-page PDF or slide with key findings

**📈 Next-Level Skills**

✔️ You’ll learn **time-series handling**  
✔️ Interactive plots with plotly (step up from static seaborn)  
✔️ Daily rolling averages & trend smoothing  
✔️ Optionally explore **animated charts** in Plotly

**✅✅ Suggested Tableau Layout**

📌 **Charts to build:**

* Global daily new cases (line)
* Top 10 countries (bar)
* Map with total cases by country
* Vaccinations over time (line)
* Optional: Mortality rate trend

Combine them into a **single interactive dashboard** with filters by date and country.