**🧩 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?no
* Are date columns consistent? yes
* Are country names consistent? yes
* Are daily vs cumulative values clear? no

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

**🔑 2️⃣ Descriptive Analysis**

**Problems to Solve:**

* Total cases & deaths worldwide, 🌍 Total cases worldwide: 775,866,783
* 🌍 Total deaths worldwide: 7,057,132
* Top 10 countries by total cases & deaths(

|  |  |  |
| --- | --- | --- |
| location | total\_cases | total\_deaths |
| United States | 103436829 | 1193165 |
| China | 99373219 | 122304 |
| France | 38997490 | 168091 |
| Germany | 38437756 | 174979 |
| Brazil | 37511921 | 702116 |
| South Korea | 34571873 | 35934 |
| Japan | 33803572 | 74694 |
| United Kingdom | 24974629 | 232112 |
| Russia | 24268728 | 403188 |
| Turkey | 17004718 | 101419 |

* Calculate **daily new cases** and **daily new deaths, date new\_cases new\_deaths**
* **1683 31-10-2023 0 0**
* **1684 31-12-2020 0 0**
* **1685 31-12-2021 0 0**
* **1686 31-12-2022 0 0**
* **1687 31-12-2023 1138766 16621**
* Highest single-day spike globally, 📈 Highest single-day spike: 177,772,819 cases on 2022-12-25
* Average daily new cases over time, 📈 Average daily new cases worldwide: 1,948,100

**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), yes did year 2020-80317671, year 2021-200278727, year 2022-424013822,year 2023-69238094, year 2024-2063363
* Compare top 5 countries’ trends on one chart, yes did,

|  |  |  |
| --- | --- | --- |
| location | total\_cases | total\_deaths |
| World | 775866783 | 7057132 |
| High-income countries | 429044049 | 2997359 |
| Asia | 301499099 | 1637249 |
| Europe | 252916868 | 2102483 |
| Upper-middle-income countries | 251753518 | 2824452 |
| European Union (27) | 185822587 | 1262988 |
| North America | 124492666 | 1671178 |
| United States | 103436829 | 1193165 |
| China | 99373219 | 122304 |
| Lower-middle-income countries | 91954400 | 1188026 |

* And

|  |  |  |
| --- | --- | --- |
| location | total\_cases | total\_deaths |
| United States | 103436829 | 1193165 |
| China | 99373219 | 122304 |
| France | 38997490 | 168091 |
| Germany | 38437756 | 174979 |
| Brazil | 37511921 | 702116 |
| South Korea | 34571873 | 35934 |
| Japan | 33803572 | 74694 |
| United Kingdom | 24974629 | 232112 |
| Russia | 24268728 | 403188 |
| Turkey | 17004718 | 101419 |

* Vaccination trend over time, yes did line chart, year 2020-7147010366, year 2021- 8071561189, year 2022- 2703126698, year 2023-75771266, year 2024- 1203277
* Mortality rate over time (deaths / cases), year 2020-2.360, year 2021- 1.940, year 2022 – 0.950, year 2023 – 0.910, year 2024 – 0.910

**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, did
* Practice SELECT, GROUP BY, ORDER BY, did
* Query total cases by continent or region, did

**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, did
* Top countries by cases & deaths, did
* World map with total cases (color gradient), did
* Filters by country and date, did

**📦 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), did
* Top 10 countries (bar), did
* Map with total cases by country, did
* Vaccinations over time (line), did
* Optional: Mortality rate trend, did

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

KPIs. Global total case – 775900191

Global total deaths – 7058885

Global total vaccinations – 13754616202

Made dashboard of these. Global daily new cases (line), did

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

And made individual graphs of these

* Global daily new cases
* global KPI values
* Global mortality rate yearly trend
* Global trend line
* Map with total cases by country
* Mortality rate trend per country over time yearly
* Top 10 countries by total cases (latest date)
* Top countries by cases & deaths
* Total cases & deaths globally by date (time trend)
* Vaccination progress per continent
* Vaccinations over time