**Project Title: COVID-19 Global Data Tracker**

**Project Description:**

In this project, learners will build a data analysis and reporting notebook (or app) that tracks global COVID-19 trends. The project will analyze **cases, deaths, recoveries, and vaccinations** across countries and time. Learners will clean and process real-world data, perform exploratory data analysis (EDA), generate insights, and visualize trends using Python data tools.

By the end, they’ll have a **data analysis report with visuals and narrative insights**, suitable for presentation or publishing.

**🚩 Project Objectives:**

✅ Import and clean COVID-19 global data  
✅ Analyze time trends (cases, deaths, vaccinations)  
✅ Compare metrics across countries/regions  
✅ Visualize trends with charts and maps  
✅ Communicate findings in a Jupyter Notebook or PDF report

**🗂️ Project Segments (Step-by-Step Guide)**

**1️⃣ Data Collection**

* **Goal:** Obtain a reliable COVID-19 dataset.

✅ **Data Sources:**

* [Our World in Data COVID-19 Dataset (CSV & API)](https://www.kaggle.com/datasets)
* [Johns Hopkins University GitHub Repository](https://github.com/CSSEGISandData/COVID-19)

👉 **Recommended for beginners:** Use the cleaned CSV from **Our World in Data** (easy to load with pandas).

✅ **Action:**

* Download owid-covid-data.csv from the above link.
* Save in your working folder.

**2️⃣ Data Loading & Exploration**

* **Goal:** Load the dataset and explore its structure.

✅ **Tasks:**

* Load data using pandas.read\_csv().
* Check columns: df.columns.
* Preview rows: df.head().
* Identify missing values: df.isnull().sum().

✅ **Tools:**

* pandas

📌 **Key columns:**

* date, location, total\_cases, total\_deaths, new\_cases, new\_deaths, total\_vaccinations, etc.

**3️⃣ Data Cleaning**

* **Goal:** Prepare data for analysis.

✅ **Tasks:**

* Filter countries of interest (e.g., Kenya, USA, India).
* Drop rows with missing dates/critical values.
* Convert date column to datetime: pd.to\_datetime().
* Handle missing numeric values with fillna() or interpolate().

✅ **Tools:**

* pandas

**4️⃣ Exploratory Data Analysis (EDA)**

* **Goal:** Generate descriptive statistics & explore trends.

✅ **Tasks:**

* Plot total cases over time for selected countries.
* Plot total deaths over time.
* Compare daily new cases between countries.
* Calculate the death rate: total\_deaths / total\_cases.

✅ **Visualizations:**

* Line charts (cases & deaths over time).
* Bar charts (top countries by total cases).
* Heatmaps (optional for correlation analysis).

✅ **Tools:**

* matplotlib
* seaborn

**5️⃣ Visualizing Vaccination Progress**

* **Goal:** Analyze vaccination rollouts.

✅ **Tasks:**

* Plot cumulative vaccinations over time for selected countries.
* Compare % vaccinated population.

✅ **Charts:**

* Line charts.
* Optional: Pie charts for vaccinated vs. unvaccinated.

✅ **Tools:**

* matplotlib
* seaborn

**6️⃣ Optional: Build a Choropleth Map**

* **Goal:** Visualize cases or vaccination rates **by country on a world map**.

✅ **Tools:**

* Plotly Express
* Or geopandas (advanced)

✅ **Tasks:**

* Prepare a dataframe with iso\_code, total\_cases for the latest date.
* Plot a choropleth showing case density or vaccination rates.

**7️⃣ Insights & Reporting**

* **Goal:** Summarize findings.

✅ **Tasks:**

* Write 3-5 key insights from the data (e.g., "X country had the fastest vaccine rollout").
* Highlight anomalies or interesting patterns.
* Use markdown cells in Jupyter Notebook to write your narrative.

✅ **Deliverables:**

* A well-documented **Jupyter Notebook** combining:
  + Code
  + Visualizations
  + Narrative explanations
* Optional export: **Notebook → PDF** or a PowerPoint with screenshots.

**🛠️ Recommended Tools:**

✅ Jupyter Notebook (or VS Code with Jupyter extension)  
✅ pandas  
✅ matplotlib & seaborn  
✅ Optional: plotly, geopandas

**🌍 Helpful References:**

1. Kaggle Dataset -<https://www.kaggle.com/datasets>

## 🎯 Optional Stretch Goals:

⭐ Allow user input (choose country & date range)  
⭐ Build an interactive dashboard with **Streamlit** or **Dash**  
⭐ Include hospitalization or ICU data if available

**✅ Final Deliverable:**

A **Jupyter Notebook data report** that:

* Loads, cleans, analyzes, and visualizes COVID-19 data.
* Communicates insights with a clear narrative and visuals.
* It is easy to read, well-commented, and reproducible.

**Key Objectives You May Have Achieved:**

✅ Collected global COVID-19 data from **Our World in Data**  
✅ Loaded and explored the dataset using **pandas**  
✅ Cleaned and prepared the data by handling missing values and filtering relevant countries  
✅ Performed **exploratory data analysis** (EDA) to identify trends in cases, deaths, and vaccinations  
✅ Created visualizations (line charts, bar charts, heatmaps, optional choropleth maps) to illustrate key metrics  
✅ Calculated critical indicators such as death rates and vaccination coverage  
✅ Presented insights through a well-documented **Jupyter Notebook report** combining code, visuals, and narrative explanations

**Submission**

Go to the assignment for this week and submit your project as an assignment.

Happy Coding