**Week 1 – Lab Assignment**

**NOTES:**

1. **Before attempting this lab assignment, make sure you have completed the exercise in week 1 - lab exercise**
2. **Upon completing this exercise, you’ll need to upload your python (jupyter .ipynb file) to canvas before the end of this week.**

**Assignment details:**

**Advanced Lab Exercise: Real-World Data Visualisation Challenge**

**Instructions:**

* Use **Python and Jupyter Notebook** to complete the tasks.
* You may need to **research** certain techniques to fully solve the exercise.
* Submit your **Jupyter Notebook (.ipynb)** file on Canvas.
* Ensure all code cells are **well-documented** with comments.

**Exercise: Analysing and Visualising Global Temperature Trends**

**Dataset:**

Download the **Global Land Temperature Data** from:  
Global Temperature Dataset (Kaggle): <https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>

Note: you only need the GlobalLandTemperaturesByCountry.csv file.

**Task 1: Data Preprocessing**

1. Load the dataset into a **Pandas DataFrame**.
2. Select data from **a specific country of your choice**.
3. Convert the **date column** to a datetime format and filter for data from the year **1900 onward**.
4. Handle missing values: Decide on an appropriate strategy (e.g., drop, interpolate, or fill).
5. Create a **new column** for the **average yearly temperature** per country.

**Task 2: Data Visualisation**

1. **Line Chart**:
   * Plot the **temperature trend over time** (yearly average temperature).
   * Customise the chart with a **title, axis labels, and markers**.
2. **Rolling Average Chart**:
   * Compute and plot a **10-year rolling average temperature**.
   * Overlay this rolling average on the original line chart.
3. **Comparing Multiple Countries**:
   * Select **three different countries** and compare their temperature trends on the **same chart**.
   * Use different **line styles and colours** for each country.

**Task 3: Extreme Weather Events Analysis**

1. Identify the **top 5 hottest and coldest years** in your chosen country.
2. Mark these years on your temperature trend chart with **annotations**.
3. Plot a **bar chart** showing the **temperature difference** between the hottest and coldest years.

**Task 4: Investigate Correlation with CO₂ Emissions**

1. **Download the CO₂ Emissions Dataset** from:  
   🔗 <https://github.com/owid/co2-data>
   * Click on **"owid-co2-data.csv"** and download the file.
2. **Load and Preprocess the Data:**
   * Read the CO₂ dataset into a **Pandas DataFrame**.
   * Select relevant columns: **year, country, and CO₂ emissions (co2 column)**.
   * Filter the dataset for **the same country** used in your temperature analysis.
3. **Merge the Temperature Data with the CO₂ Data**:
   * Merge both datasets based on **year** and **country**.
   * Handle any missing values appropriately.
4. **Create a Scatter Plot**:
   * Plot **CO₂ emissions vs. average temperature** for the selected country.
   * Label the axes and add a title.
5. **Fit a Linear Regression Line**:
   * Use **Seaborn's regplot** to fit a regression line.
   * Interpret the trend: Does CO₂ emission increase correlate with rising temperatures?

**Submission Guidelines**

* Ensure your Jupyter Notebook **includes all required charts and explanations**.
* Save your notebook as **Week1\_lab\_assignment\_YourID.ipynb**.
* Upload your **.ipynb file** to Canvas before the deadline.