### **Global Pollution Analysis and Energy Recovery**

**Objective** The goal is to analyze pollution data across various countries and predict how pollution levels can impact energy recovery. This dataset will be used to explore **clustering and neural networks** for environmental analysis.

#### **Phase 1 - Data Preprocessing and Feature Engineering**

* **Data Import and Cleaning**
  + Load the dataset ([**Global\_Pollution\_Analysis.cs**v](https://drive.google.com/file/d/1AEMcCWzJ24fc26Q761SEOa1sOsZiJBC5/view?usp=sharing)).
  + Handle missing data using imputation or removal techniques.
  + Normalize/scale pollution indices (air, water, and soil) for consistency.
  + Encode categorical features (e.g., country, year) using Label Encoding.
* **Feature Engineering**
  + Create new features like energy consumption per capita.
  + Analyze yearly pollution trends and their impact on energy recovery.

#### **Phase 2 - Clustering using K-Means and Hierarchical Clustering**

1. **K-Means Clustering**
   * **Objective** Cluster countries based on pollution levels (e.g., air, water, and soil pollution) and energy recovery metrics.
   * **Implementation**
     + Apply K-Means to group countries based on pollution indices and energy consumption.
     + Use **Elbow Method** to determine the optimal number of clusters.
   * **Evaluation**
     + Visualize pollution trends and energy recovery by clusters.
     + Identify countries with similar environmental and energy characteristics.
2. **Hierarchical Clustering**
   * **Objective** Perform hierarchical clustering to analyze pollution levels and energy recovery.
   * **Implementation**
     + Use Agglomerative Clustering to build a hierarchical structure based on pollution and energy data.
     + Visualize the dendrogram and determine the number of clusters.
   * **Evaluation**
     + Compare the results of hierarchical clustering with K-Means.
     + Visualize clusters in a hierarchical tree structure.

#### **Phase 3 - Neural Networks for Energy Recovery Prediction**

1. **Introduction to Neural Networks**
   * **Objective** Build a neural network to predict energy recovery from pollution data.
   * **Implementation**
     + Train a simple feedforward neural network using Keras/TensorFlow.
     + Input features: Air Pollution Index, CO2 Emissions, Industrial Waste, etc.
     + Output: Energy recovered in GWh.
   * **Evaluation Metrics** R², Mean Squared Error (MSE), Mean Absolute Error (MAE).
2. **Model Improvement**
   * Tune hyperparameters such as the number of layers, neurons, activation functions, and learning rates to improve model accuracy.
   * Evaluate the neural network’s performance compared to linear regression models for energy recovery.

#### **Phase 4 - Reporting and Insights**

* **Model Comparison**
  + Compare the performance of K-Means and Hierarchical Clustering with the neural network model in predicting energy recovery from pollution data.
  + Analyze the strengths and weaknesses of each model and which one offers better predictions and clustering results.
* **Actionable Insights**
  + Provide insights into how clustering can reveal trends in pollution and energy recovery.
  + Recommend strategies for countries to reduce pollution based on clustering results and neural network predictions.

### **Final Deliverables**

1. **Jupyter Notebook (.ipynb)**
   * Full code for data preprocessing, clustering, and neural network modeling.
2. **Data Visualizations**
   * Visualizations such as Elbow Method graphs, dendrograms, and neural network performance metrics.
3. **Final Report**
   * A comprehensive report summarizing the methodology, model evaluations, key findings, and actionable recommendations.