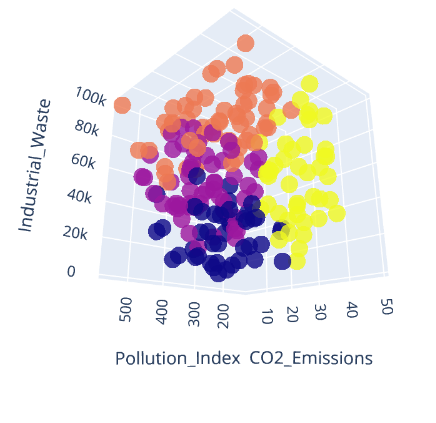
**Report: Model Comparison and Insights**

**1. Clustering Analysis**

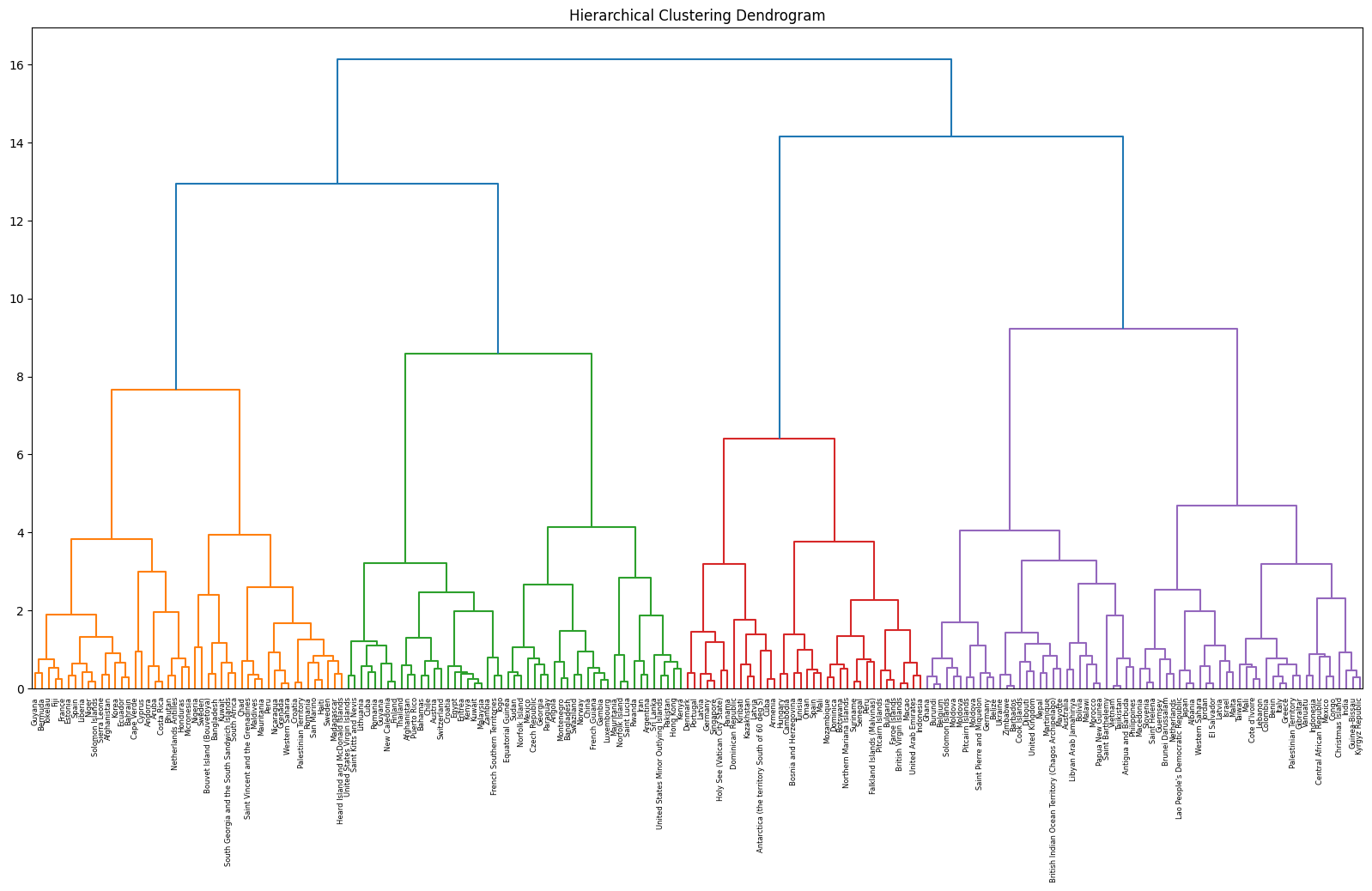
**K-Means**

* Optimal number of clusters (k) determined using **Silhouette score**.
* Produced **4 clusters** of countries based on pollution and industrial metrics.
* 3D scatter visualisation showed separation between countries based on:
  + **CO₂ Emissions**, **Pollution Index**, **Industrial Waste**



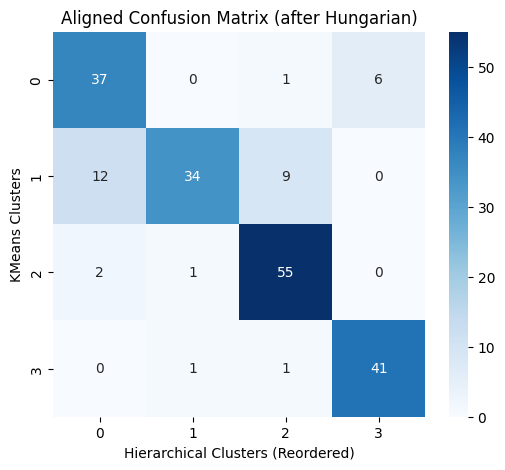
**Hierarchical Clustering**

* Dendrogram visualisation suggested a hierarchical grouping into **4 main clusters**.
* More interpretable for understanding **nested relationships** between countries.
* Countries with similar energy & pollution patterns appear grouped closer.



**Comparison (K-Means vs Hierarchical)**

* Direct cluster labels differ (due to arbitrary numbering), so we aligned them using the Hungarian algorithm.
* Confusion Matrix (Aligned) shows strong agreement:
  + Example: 55 countries matched in Cluster 2 (K-Means) and Cluster 2 (Hierarchical).



* Agreement Metrics:
  + Adjusted Rand Index (ARI): 0.627
  + Normalised Mutual Information (NMI): 0.644
* Interpretation:
  + The two methods are moderately consistent, indicating that both are capturing similar pollution-energy recovery structures in the data.

**2. Predictive Modelling**

Neural Network

* MSE: 22,555
* MAE: 135.73
* R²: 0.0135

Linear Regression

* R²: 0.0294 (slightly higher than Neural Network)

Comparison

* Both models performed poorly in predicting *Energy Recovered (in GWh)* from pollution and industrial features.
* Very low R² values (close to 0) indicate poor feature selection that does not explain much variance in energy recovery.
* Neural Network did not outperform Linear Regression because:
  + Problem may be too linear/simple for NN benefits.
  + Input features may lack enough predictive signal.
  + Possible data limitations (noise, missing features, scale differences).

**3. Insights from Clustering**

* **Clusters reveal trends** that prediction models failed to capture:
  + Certain groups of countries have **high industrial waste but low energy recovery**, suggesting inefficiency.
  + Some clusters show **better pollution control with higher energy recovery**, indicating stronger waste-to-energy policies.
* Clustering helps in **categorising countries** and identifying **outliers**.

**4. Recommendations**

1. **For Modelling**:
   * Add more predictive features (e.g., renewable energy investment, technology adoption, government policy data).
   * Perform **feature engineering** (ratios like *energy recovered per ton of waste*).
   * Try **regularised regression (Ridge/Lasso)** or **tree-based models (Random Forest, XGBoost)**.
2. **For Policy & Strategy**:
   * Countries in clusters with **high waste & low energy recovery** should:
     + Invest in **waste-to-energy technology**.
     + Strengthen **pollution control regulations**.
   * Countries in **efficient clusters** can serve as **benchmarks**.

**Summary**

* **Clustering (K-Means & Hierarchical)** showed **moderate agreement** (ARI = 0.63, NMI = 0.64), successfully grouping countries with similar pollution-energy profiles.
* **Prediction models (NN & Linear Regression)** performed poorly, with **very low R²** values, showing that current features do not strongly explain **Energy recovered**.
* **Clustering provided more useful insights** than prediction models, revealing clear country groupings and inefficiencies.
* **Next steps:** Enrich the dataset, try advanced models, and use clustering insights to guide environmental strategies.