What each dataset contains & how to use it

1) `score1_global.csv` — Global coal rents share of global GDP

- **year** – calendar year.
- **global_coal_rents_usd** – summed coal rents in USD (constructed from country data).
- **global_gdp_usd** – summed world GDP in current USD for that year.
- **score1_coal_rent_share_percent** – 100×global_coal_rents_usd/global_gdp_usd100 × global_coal_rents_usd / global_gdp_usd100×global_coal_rents_usd/global_gdp_usd. *Meaning:* macro share of world GDP that corresponds to coal rents (resource surplus). Coal rents are defined by the World Bank as the surplus value from coal mining (market price – extraction cost, excluding labor & O&M) and are often used as a conservative indicator of resource dependence.

Use: context only (not in the per-country composite).

---

2) `score1a_country_year.csv` — National coal rents (% of GDP)

- **country_iso3, country_name**
- **year**
- **coal_rents_pct** – coal rents as **% of national GDP** (World Bank indicator `NY.GDP.COAL.RT.ZS`).
- **gdp_usd** – national GDP (current US$).
- **coal_rents_usd** – implied level (coal_rents_pct × gdp_usd / 100).

Small score:
`Score1a = coal_rents_pct` (already in the file).
Meaning: conservative measure of direct macroeconomic surplus from coal mining.

---

3) `score2_global.csv` — Global GDP share of coal-fired electricity (proxy)

- **year**
- **world_coal_share_electricity_pct** – coal's share of global electricity generation (WB `EG.ELC.COAL.ZS`).
- **assumed_electricity_sector_share_gdp** – proxy share of "utilities" (electricity/gas/water) in global GDP. In the absence of consistent global value-added by the electricity & heat sector for all years, we used a reasonable constant approximation of **~4% of GDP** (utilities' average weight across economies).
- **score2_global_pct** – assumed_share×world_coal_share_electricity_pctassumed_share × world_coal_share_electricity_pctassumed_share×world_coal_share_electricity_pct.

Use: context only (not in the per-country composite). Where possible, replace the proxy with proper "electricity & heat value added" from national accounts.

---

4) `score2a_country_year.csv` — National GDP share of coal-fired electricity (proxy)

- **country_iso3, country_name, year**
- **coal_share_electricity_pct** – coal's share of national electricity generation (WB `EG.ELC.COAL.ZS`).
- **assumed_electricity_sector_share_gdp** – proxy for each country (default 0.04 = 4%).
- **score2a_coal_power_gdp_share_pct** – assumed_share×coal_share_electricity_pctassumed_share × coal_share_electricity_pctassumed_share×coal_share_electricity_pct.

Small score:
`Score2a = score2a_coal_power_gdp_share_pct` (already computed).
Meaning: a first-order proxy for the weight of coal-power activity in the economy; replace the 4% assumption with **actual electricity & heat value added** where available (e.g., UN SNA/AMA, IEA). The 4% figure is a stylized average indicating utilities' weight in GDP; use with caution.

---

5) `score3_country_year.csv` — Coal export dependency

- **country_iso3, country_name**
- **coal_exports_usd** – value of coal exports (mainly HS 2701; 2024 values compiled from World's Top Exports).
- **total_exports_usd** – total exports of goods & services (WB `NE.EXP.GNFS.CD`; latest available year, often 2023–2024). Example indicator page used for verification.
- **total_exports_year** – reference year for total exports.
- **coal_export_share_percent** – 100×coal_exports_usd/total_exports_usd100 × coal_exports_usd / total_exports_usd100×coal_exports_usd/total_exports_usd.

Small score:
`Score3 = coal_export_share_percent` (already in the file).
Meaning: the extent to which national exports rely on coal.

> *Note:* If you need full coal-related exports (HS 2701 + 2702 + 2704, etc.), expand the coal basket and recompute.

From datasets to scores (formulas summary)

- **Score1a (country, year)** = `coal_rents_pct`.
- **Score2a (country, year)** = `coal_share_electricity_pct × assumed_electricity_sector_share_gdp`.
- **Score3 (country, year)** = `100 × coal_exports_usd / total_exports_usd`.

Normalization (per-year, cross-country):

- **Min–max (default):** (x−min)/(max−min)(x - \min)/(\max - \min)(x−min)/(max−min) on each year's cross-section.

- **Z-score (optional):** (x−μ)/σ(x - \mu)/\sigma(x−μ)/σ, optionally winsorized.

Composite (country, year):

- Normalize each **small score** → n_Score1an_Score1an_Score1a, n_Score2an_Score2an_Score2a, n_Score3n_Score3n_Score3.
- Weighted average (default equal):
  Economic Composite=w1n_S1a+w2n_S2a+w3n_S3w1+w2+w3.\text{Economic Composite} = \frac{w_1 n_S1a + w_2 n_S2a + w_3 n_S3}{w_1 + w_2 + w_3}.Economic Composite=w1+w2+w3w1n_S1a+w2n_S2a+w3n_S3.

Download: Python module (metrics, normalization, composite, dashboard)

Download the Python module

What it contains:

- **Loaders** for `score1a_country_year.csv`, `score2a_country_year.csv`, `score3_country_year.csv`.
- **Small-score calculators** (if you pass raw fields).
- **Normalization** (`minmax`, `zscore`) per year.
- **Composite builder** with customizable weights.
- **Plots** (matplotlib only; each plot in its own window):
  - `plot_country_trends(df, 'AUS')` – lines for S1a, S2a, S3, and composite.
  - `plot_top_countries(df, year=2023)` – top-N bar chart.
  - `plot_metric_distribution(df, metric='score1a', year=2021)` – histogram.

Minimal usage example

python
CopyEdit

```
# 0) pip install pandas numpy matplotlib
from coal_scores_pipeline import load_small_scores, normalize_by_year,
build_composite
from coal_scores_pipeline import plot_country_trends,
plot_top_countries, plot_metric_distribution

# 1) Load (adjust paths)
df = load_small_scores(
    'data/score1a_country_year.csv',
    'data/score2a_country_year.csv',
    'data/score3_country_year.csv'
)

# 2) Normalize (per year)
dfn = normalize_by_year(df, metrics=['score1a','score2a','score3'],
method='minmax',
                        lower_clip=0.01, upper_clip=0.99)

# 3) Composite (equal weights; change as needed)
comp = build_composite(dfn, metrics=['score1a','score2a','score3'],
                       weights={'score1a':1,'score2a':1,'score3':1})

# 4) Plots (dashboard-like set of charts)
plot_country_trends(comp, country_iso3='AUS')
plot_top_countries(comp, year=2023, metric='economic_composite',
top_n=15)
plot_metric_distribution(comp, metric='score1a', year=2021)
```

Notes, limits, and how to improve

- **Score2a proxy:** Using a constant utilities share (≈4%) is a simplifying assumption; replace with measured **electricity & heat value added** where available to sharpen accuracy. The 4% figure reflects the approximate utilities weight across economies in 2022; it's a reasonable placeholder but not a substitute for country-level SNA data.

- **Score3 coverage:** Current coal export values draw on HS-2701 (coal/briquettes) for 2024. For fuller coverage, include HS-2702 (lignite) and HS-2704 (coke), recompute `coal_exports_usd`, then `Score3`.

- **Global scores (Score1/2):** Provided as context; they are not used in the per-country composite. If you want a *global-tilt* composite, you can use Score1/2 to scale country weights or as additional features.