**Cleaning Code Summary Report**

This cleaning code is designed to preprocess and normalize a sovereign debt dataset, ensuring data consistency, accuracy, and readiness for subsequent analysis, such as correlation studies and network construction. The following steps summarize the main actions performed by the code:

**1. Country Mapping**

* **Objective:** Replace abbreviated country codes (e.g., 'US', 'JP') with their full names (e.g., 'United States', 'Japan') for better readability and interpretability.
* **Implementation:**
  + A country\_mapping dictionary maps country abbreviations to their full names.
  + Rows with unmapped or unknown country codes are dropped (dropna).

**2. Column Abbreviation Expansion**

* **Objective:** Convert abbreviations in specific columns (e.g., 'G', 'I', 'C') into full descriptive labels for clarity.
* **Implementation:**
  + An abbreviation\_mapping dictionary replaces shorthand codes with full terms like 'Gross Issues' or 'Amounts Outstanding.'
  + Columns processed: 'MEASURE', 'COLLECTION', 'Collateral'.
  + The empty 'COLLECTION' column is dropped as it has no usable data.

**3. Handling Missing Values**

* **Objective:** Ensure time-series continuity and fill gaps in data for accurate analysis.
* **Steps:**
  1. **Missing Flagging:** A 'Missing\_Flag' column is created to indicate rows with any missing values in time-series columns.
  2. **Group-Wise Mean Imputation:** Missing values are replaced with the mean of the corresponding group, grouped by 'ISSUER\_RES' (country).
  3. **Interpolation:** Smooth transitions are ensured by linear interpolation across time-series columns.
  4. **Forward and Backward Fill:** Remaining gaps are filled using forward-fill (ffill) and backward-fill (bfill) methods.

**4. Time-Series Normalization**

* **Objective:** Scale data in each time-series column (e.g., '1962-Q4', '1963-Q1') to a [0, 1] range for uniformity and to remove unit scale bias.
* **Implementation:**
  + Each column is divided by its maximum value if greater than zero.

**5. Currency Conversion**

* **Objective:** Standardize all monetary values to USD for uniform comparisons.
* **Steps:**
  + A dictionary of exchange rates converts currencies like EUR, JPY, etc., to USD.
  + Each time-series value is multiplied by its corresponding exchange rate based on the 'ISSUE\_CUR' column.
  + The 'ISSUE\_CUR' column is updated to reflect that all values are now in USD.

**6. Column Selection**

* **Objective:** Retain only relevant columns for project objectives (e.g., correlation and network analysis).
* **Included Columns:**
  + 'ISSUER\_RES': Indicates the country.
  + 'ISSUE\_CUR': Indicates the currency (post-conversion to USD).
  + 'MEASURE': Provides a descriptive label for the type of debt measure.
  + All time-series columns dynamically selected (e.g., '1962-Q4', '2024-Q2').