

Cleaning Log



Data Source:	About Analyst
Kaggle.com Link:	Kamara Rowe-Ciftci LinkedIn: https://www.linkedin.com/in/meryack View Dashboard: Link

Step	Activity	Details
1	Gather 5 CSV files from Kaggle	The files represented Foreign Direct Investment, Private Capital Inflows, Remittances, Net Official Development Assistance, Import and Export (total trade)
2	Transform from wide to long format	To enable more flexible filtering and visualization, each CSV was unpivoted/reshaped into a long format with Indicator, Country, Year, and Values
3	Combine into a single table	All 5 files were merged together into a unified long-form table in Power Query
4	Remove duplicates	Duplicate rows were removed to avoid double counting
5	Handle missing values	Any row with a missing "Values" field was removed to avoid inaccuracies in measures

6	Trim and clean	All text fields were trimmed and cleaned to a consistent format (such as country names)
7	Standardize country IDs	A function was applied to country names to convert to lower case, hyphenate multi-word names, and remove special characters
8	GitHub URLs for flags	The cleaned country IDs were used to construct URLs for country flags , stored in a new column (Flag URL)
9	Final enrichment	The new column “Flag URL” enriched the data , allowing Power BI to pull flags directly from GitHub
10	Prepare for Power Query	The final cleaned, enriched, long-form table was ready to be loaded into Power Query and Power BI for visualization and analysis

Data Cleaning and Preparation Summary for Country Flag Images

Objective:

Prepare a consistent and clean list of country names to correctly match flag image files stored in a GitHub folder, enabling automated linking in Excel and Power BI.

Steps Completed:

1. Country Name Standardization:

- Converted all country names to lowercase.
- Replaced spaces and multi-word separators with hyphens (e.g., "saint-lucia").
- Removed or replaced special characters (e.g., accents removed in "cote-divoire").
- Adjusted naming for countries with suffixes like "(Republic of)" to a consistent prefix format (e.g., "republic-of-iran").
- Ensured multi-word names use hyphens consistently (e.g., "bosnia-and-herzegovina").

2. **File Naming Convention:**

- Decided to use .png as the standard image file extension for all flags.
- Planned to save all flag images in one flat GitHub folder with filenames matching the cleaned country names.

3. **URL Formation:**

- Created a formula structure to generate flag image URLs dynamically in Excel by concatenating the GitHub base URL with cleaned country names and .png extension.
- Considered special cases with a possible mapping table for exceptions.

4. **Data Format Considerations:**

- Recognized that the dataset is in a long format with repeated country names.
- Planned to ensure flag URL generation respects grouping by country (not jumping between names).

5. **Next Steps:**

- Upload flag images to GitHub with the finalized naming convention.
- Generate flag URLs in Excel using the prepared formulas.

Flag Image Workflow Documentation (Demonstrating Process Discipline and System Compatibility)

Objective:

Prepare standardized flag image files and corresponding file references for integration into Excel and Power BI dashboards, while demonstrating structured workflow skills relevant to systems like Salesforce.

Step-by-Step Process**1. Sourcing Flag Images**

- Collected high-quality country flag images from a reliable external website.
- Ensured comprehensive coverage (193+ UN-recognized countries and relevant territories).

2. Preparing Naming Convention

- Developed a standardized naming structure:
 - All lowercase
 - Words separated by hyphens (e.g., saint-kitts-and-nevis)
 - Removed special characters and standardized republic/state suffixes/prefixes.

- Matched country names with HDI dataset for accurate flag-to-country mapping.

3. Image Standardization via Canva

- Imported all flags into Canva.
- Standardized dimensions (e.g., consistent width and height) to maintain visual uniformity in Power BI and other tools.
- Exported images in .png format (chosen for compatibility and quality).

4. Renaming Files According to Final Format

- Renamed each image using the agreed convention (e.g., cote-divoire.png, republic-of-korea.png).
- Ensured consistency across all files to prevent lookup errors in automated scripts.

5. File Organization and Upload

- Created a new folder for finalized flag images.
- Uploaded all standardized and renamed images to a GitHub repository with a flat structure (no nested folders).
- Verified URLs work correctly and match dataset names in Excel.

Outcome

- Clean, uniform, and scalable asset structure for easy lookup and visualization.
- Demonstrated ability to follow structured, multi-step procedures—relevant to CRM and ERP systems like

Salesforce.

- Positioned the project to seamlessly connect data rows to country flag images in Excel/Power BI using dynamic URLs.