

VISUALIZING US NATURAL DISASTER DECLARATION – TRENDS AND PATTERNS

Week 1 Documentation

Data Cleaning in Power BI (Power Query)

Data Cleaning in Power BI

1. Introduction

Data cleaning is the foundational step in building reliable dashboards. In Week 1, the FEMA Disaster Declarations dataset was imported into Power BI and transformed using **Power Query**. The goal was to ensure consistency, remove duplicates, and prepare the dataset for accurate visualization and analysis.

2. Step-by-Step Procedure

2.1 Importing Data

- **Source:** FEMA Disaster Declarations dataset (CSV/Excel format).
One can download the dataset from
<https://www.fema.gov/openfema-data-page/fema-web-disaster-declarations-v1>
- **Process:**
 - Open Power BI Desktop → Home → Get Data.
 - Select the file format (CSV/Excel).
 - Load the dataset into Power Query Editor

2.2 Removing Unnecessary Columns

- **To Do:** Identify columns not required for analysis (e.g., metadata, unused codes).
- **Procedure:**
 - Right-click column → *Remove*.
 - Document removed fields in a log for reproducibility.
- **Result:** Removed the following columns
 - stateCode
 - disasterPageUrl
 - shapefileUrl
 - kmzfileUrl
 - geoJsonUrl
 - id
 - hash
 - lastRefresh

2.3 Checking Data Types

- **To Do:** Verify that each column in the FEMA dataset has the correct data type assigned.
- **Procedure:**
 - Use Power Query Editor to review column type icons (ABC = Text, 123 = Number, Calendar = Date).
 - Change types where necessary:
 - disasterNumber → Whole Number
 - declarationDate, incidentBeginDate, incidentEndDate, entryDate, closeoutDate, updateDate → Date
 - stateName, declarationType, incidentType, disasterName, designatedIncidentTypes → Text
 - iaProgramDeclared, ihProgramDeclared, paProgramDeclared, hmProgramDeclared → Boolean (True/False)
- **Outcome:** Ensures consistency in calculations, prevents errors in aggregations, and supports accurate visualizations.

2.4 Handling Missing Values

- **To Do:** Check for null values across key fields.
- **Procedure:**
 - Use *Transform* → *Replace Values* for categorical fields.
 - For dates, apply conditional logic:
 - If incidentEndDate is missing, replace with declarationDate as proxy.
 - If stateName is missing, mark as “Unknown”.
 - *The closeOutDate field contained numerous null values, which indicate disasters that are still ongoing. To ensure accuracy, these null values are retained. For visualization purposes, a conditional column was added to categorize records as either ‘Closed’ (with a valid date) or ‘Open’ (null values).*
- **Outcome:** Ensures no blank values exists.

2.5 Creating Derived Columns

- **Fiscal Year (fyDeclared):**
 - Formula: Date.Year([declarationDate]) adjusted for fiscal cutoff (Oct–Sep).

- **Incident Duration:**
 - Formula: Duration.Days([incidentEndDate] - [incidentBeginDate]).
- **Status:**
 - Formula: If [closeoutDate] is null → "Open", else "Closed".

2.6 Remove duplicates

- **To Do:** Remove duplicate rows.
- **Procedure:**
 - *Home → Remove Rows drop-down → Remove Duplicates.*
- **Outcome:** Prevents double counting in dashboards.

2.7 Load Cleaned Data

- **To Do:** Close & Apply changes.
- **Procedure:**
 - Save transformations in Power Query.
 - Load cleaned dataset into Power BI model.
- **Outcome:** Dataset ready for visualization dashboards.

3. Outcome of Week 1

By the end of Week 1:

- The FEMA dataset was cleaned, standardized, and loaded into Power BI.
- Key derived fields (Fiscal Year, Incident Duration, Assistance Flags) were created.
- The dataset is now reliable for building dashboards in subsequent weeks.

4. Next Steps (Week 2 Preview)

In Week 2, the same cleaning workflow will be implemented in Python to validate reproducibility, automate transformations, and prepare for advanced analytics.