US Natural Disaster Data Cleaning & Analysis Report (FEMA Dataset)

Power BI | Power Query | Data Preparation

Prepared By: Harika Vavilapalli
Internship/Program: Data Science Intern
Organization: Infosys Springboard 6.O
Mentor: C.Akshay
Submission Date: 23.01.2026

Tools Used: Kaggle, Power BI Desktop, Power Query Editor
Dataset Source: FEMA Disaster Declarations Dataset

Data Cleaning Summary

Objective:

The goal of cleaning was to prepare the FEMA Natural Disaster dataset for analysis by:

- fixing incorrect data types (especially date columns),
- removing unwanted spaces and hidden characters,
- handling missing values,
- converting Yes/No fields into proper logical values,
- making the dataset ready for reporting and visualization.

Step-wise Data Cleaning Process:

Step 1: Import Dataset

1. Opened Power BI Desktop
2. Clicked Get Data → Text/CSV
3. Selected the file: database.csv
4. Clicked Transform Data to open Power Query Editor

Result: Dataset loaded into Power Query for cleaning.

Step 2: Promote First Row as Headers

1. In Power Query Editor, selected:
   Transform → Use First Row as Headers

Result: The first row was converted into proper column names.

Step 3: Check Column Data Types

Initially, many columns were detected incorrectly (especially date columns showing as Text (ABC).

We verified all column types and corrected them for accurate analysis.

Step 4: Convert Date Columns from Text to Date

The dataset contained important date columns like:

- Declaration Date
- Start Date
- End Date
- Close Date

Actions performed:

1. Selected each date column
2. Changed datatype to Date
3. Used Using Locale when needed:
    - Data Type: Date
    - Locale: English (United States)

Reason: The dataset follows MM/DD/YYYY format, and using locale avoids date conversion errors.

Result: All date columns became true Date values (calendar icon).

Step 5: Clean Text Columns (Trim + Clean)

Text columns sometimes contain:

- extra spaces,
- invisible characters,
- formatting issues.

Columns cleaned:

- State
- Disaster Type
- Disaster Title
- Declaration Type
- (and other text columns)

Actions performed:

1. Selected text columns
2. Applied:
    - Transform → Format → Trim
    - Transform → Format → Clean

Result: Removed unwanted spaces and hidden characters for better grouping and filtering.

Step 6: Keep Declaration Number as Text

Declaration Number values are like:

- DR-1, DR-2, DR-3…

Action:

- Kept datatype as Text

Reason: It is an identifier, not a numeric field.

Step 7: Handle Missing Values

Some columns (example: County) contained null values.

Action:

- Kept nulls as they represent missing/unknown values in the original dataset
  (Optional: could replace null with "Unknown" if needed for reporting)

Result: Dataset remains accurate without forcing incorrect values.

Step 8: Convert Program Columns (Yes/No → True/False)

These columns contain Yes/No values:

- Individual Assistance Program
- Individuals & Households Program
- Public Assistance Program
- Hazard Mitigation Program

Actions performed:

1. Replaced:
   - Yes → TRUE
   - No → FALSE
2. Changed datatype to True/False

Reason: Boolean format improves:

- filtering,
- KPI calculations,
- report visuals.

Result: These columns are now logical fields suitable for analysis.

Step 9: Fix Column Width Visibility (UI Cleaning)

Some columns appeared blank due to narrow width.

Actions performed:

- Used Auto Size Column Width
  OR
- Double clicked the column border to expand width

Result: Data values became clearly visible in Power Query preview.

Step 10: Verify Cleaned Dataset

Final verification checks:

✅ Date columns show correct Date format
✅ No datatype mismatch
✅ Text values are consistent
✅ Yes/No fields converted properly
✅ Dataset ready for visualization

Step 11: Apply Changes

1. Clicked Close & Apply
2. Loaded cleaned dataset into Power BI for reporting

Result: Cleaned dataset is successfully applied and ready for analysis.

Final Output:

The dataset is now:

- structured,
- consistent,
- error-free,
- ready for dashboards and insights.

Conclusion

In this project, the FEMA US Natural Disaster dataset was successfully cleaned and prepared using Power BI (Power Query). All major issues such as incorrect data types, inconsistent date formats, and unwanted spaces in text fields were corrected. Date columns were converted into proper *Date* format using locale settings, and categorical fields were standardized for better filtering and analysis. Overall, the dataset is now structured, accurate, and ready for visualization and further insights.

Key Outcomes

- Converted all date columns (Declaration Date, Start Date, End Date, Close Date) from text to proper **Date** format.
- Cleaned text columns using **Trim** and **Clean** to remove extra spaces and hidden characters.
- Kept unique identifiers like **Declaration Number** as text to preserve correct values (ex: DR-1).
- Converted program-related columns (Yes/No) into **True/False** for easier analysis and reporting.
- Improved readability by adjusting column widths and verifying final dataset consistency.

Recommendations / Next Steps

- Use the cleaned dataset to create an interactive Power BI dashboard showing trends and patterns.
- Add visuals such as disasters by year, top disaster types, and state-wise disaster frequency.
- Create slicers for Year, State, and Disaster Type for better user interaction.
- If the dataset gets updated in future, automate refresh and maintain a raw vs processed dataset structure in the repository.

References

- FEMA Disaster Declarations Dataset
- Microsoft Power BI & Power Query Documentation