Data Cleaning & Preparation (Power BI)

This section describes the step-by-step data cleaning process performed using **Power BI – Power Query Editor** to ensure data accuracy, consistency, and readiness for analysis.

---

1. Data Import

The dataset was imported into **Power BI Desktop** using the *Get Data* option. The **Transform Data** feature was used to access Power Query for data cleaning and preprocessing before loading it into the data model.

---

2. Initial Data Assessment

An initial review of the dataset was conducted to understand:

- Number of rows and columns
- Column names and structure
- Data types
- Presence of missing values and duplicates

Power BI features such as **Column Quality, Column Distribution, and Column Profile** were enabled to identify data issues.

---

3. Column Renaming

Column names were standardized to improve readability and usability:

- Removed spaces and special characters
- Used meaningful, business-friendly names
- Followed consistent naming conventions

Example:
Declaration Date → Declaration_Date

---

4. Data Type Standardization

Each column was assigned an appropriate data type to prevent calculation errors:

- Dates → Date
- Numeric fields → Whole Number / Decimal
- Categorical fields → Text

Incorrect data types were corrected early in the process.

---

5. Handling Missing Values

Missing values were treated based on business relevance:

- Rows with critical missing values were removed
- Non-critical null values were replaced with default values such as 0 or "Unknown"

This ensured data completeness without distorting insights.

---

6. Removal of Duplicate Records

Duplicate rows were identified and removed using unique identifiers to prevent data duplication and inaccurate aggregations.

---

7. Filtering Irrelevant Data

Unnecessary or out-of-scope records were filtered out, such as:

- Blank state or category values
- Incomplete or invalid records
- Data not relevant to the analysis timeframe

---

8. Removal of Unwanted Columns

Columns that did not contribute to analysis were removed, including:

- Metadata and system-generated fields
- Redundant identifiers
- Free-text or high-null columns
- Fields not used in KPIs or visuals

This improved data model performance and clarity.

---

9. Data Standardization

Text values were standardized to maintain consistency:

- Corrected inconsistent spellings and cases
- Unified categorical values (e.g., state names, disaster types)
- Applied proper casing where required

---

10. Feature Engineering (Derived Columns)

New columns were created to support analysis, such as:

- Year and Month extracted from date fields
- Time-based grouping fields
- Categorization columns where required

---

11. Data Validation

Final validation checks were performed to ensure:

- No critical null values
- Correct data types across all columns
- Accurate and consistent values
- Dataset aligned with analysis requirements

---

12. Apply & Load

After completing all transformations, the cleaned dataset was loaded into the Power BI data model using **Close & Apply**, making it ready for KPI creation and dashboard visualization.

Summary

The data was cleaned using Power Query by standardizing column names and data types, handling missing values, removing duplicates and irrelevant columns, filtering invalid records, and creating derived fields to ensure accurate and reliable analysis.