Visualizing US Natural Disaster Declarations — Milestone 1

Overview

Milestone-1 focuses on importing, cleaning, and preparing the FEMA Disaster Declarations dataset so that it can be used for meaningful analysis in later stages of the project. The work completed in this phase ensures that the data is accurate, consistent, and free from major quality issues before performing Exploratory Data Analysis (EDA) and visualizations.

The tasks carried out mainly include dataset loading, datatype correction, handling missing values, removing duplicate records, standardizing categorical attributes, and generating initial summary outputs.

Objectives Completed

The following milestones were successfully achieved in this phase:

- Imported the FEMA Disaster Declarations dataset into Pandas
- Examined dataset shape, column details, and datatypes
- Converted disaster-related date fields into datetime format
- Identified and handled missing and duplicate values
- Standardized and normalized the incidentType column
- Exported the cleaned dataset for later milestones
- Performed basic exploratory checks to validate data quality

These steps ensure that the dataset is reliable and suitable for further analysis.

Data Pre-Processing and Cleaning

The dataset was cleaned and prepared using the following steps:

❖ **Conversion of Date Fields**

Columns such as
declarationDate, incidentBeginDate, and incidentEndDate
were converted into proper datetime format to support time-based analysis and trend evaluation.

❖ **Removal of Duplicate Records**

Duplicate entries were detected and removed to avoid repeated disaster counts and to maintain data accuracy.

❖ **Handling Missing Values**

Records missing essential identifiers were eliminated, and nullable fields were treated appropriately so that analysis results are not affected by incomplete data.

❖ **Standardization of incidentType**

To avoid inconsistent categories:

- extra spaces were removed
- all values were converted to uppercase
- similar disaster labels were grouped together

This improves clarity and prevents category duplication during analysis.

Initial Exploratory Checks

After cleaning, several verification checks were performed, including:

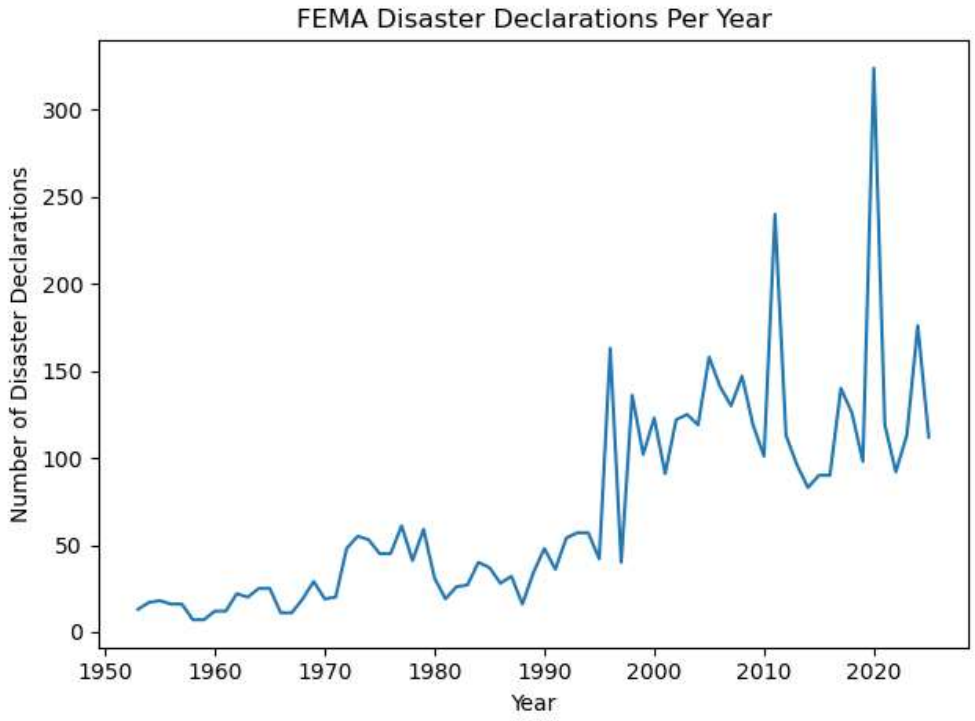
- reviewing dataset shape
- inspecting null values
- validating datatypes
- examining frequency distributions
- confirming column consistency

From these checks, the dataset was confirmed to be **clean, structured, and ready for analysis**.

Graphs & Visual Outputs (Milestone-1 Results)

The following sections are reserved for the visualizations generated in this milestone.

Graph-1: Disaster Declarations Per Year (Trend Plot)



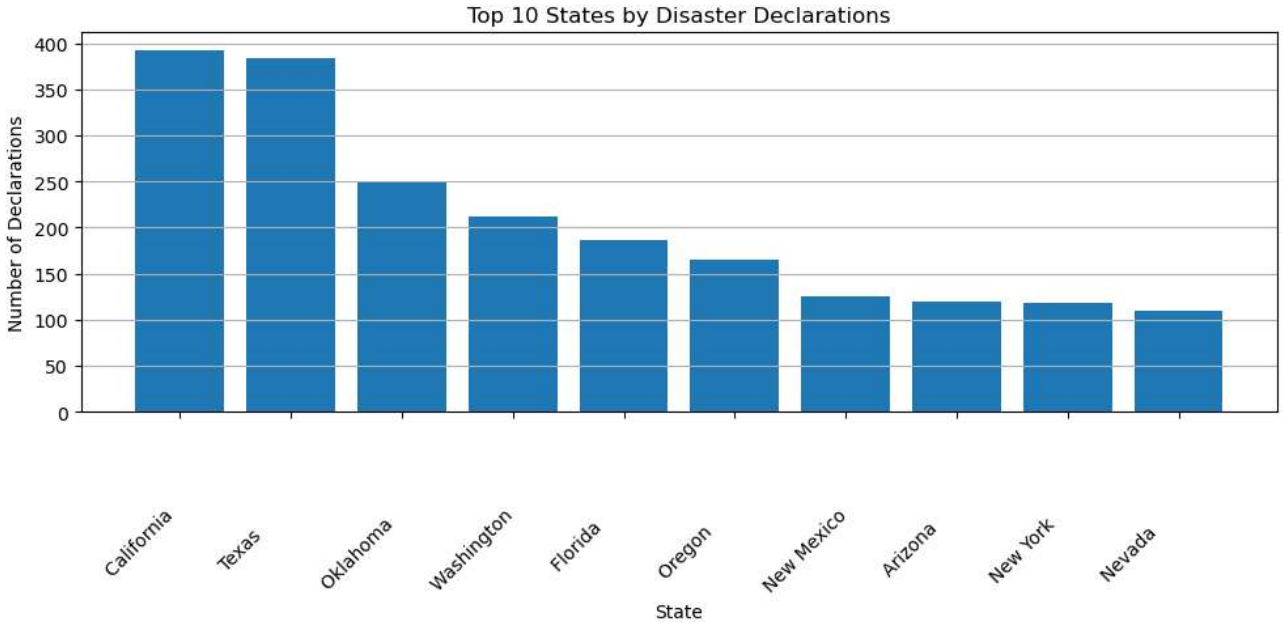
Explanation

This graph displays the number of FEMA disaster declarations recorded each year. It helps in understanding:

- changes in disaster occurrences over time
- periods with higher disaster frequency
- long-term disaster behavior patterns

The visualization provides an overview of yearly trends across different time periods.

Graph-2: Top 10 States with Highest Disaster Declarations (Bar Chart)



Explanation

This bar chart shows the states that reported the highest number of disaster declarations. It helps identify:

- regions more prone to disasters
- states with repeated disaster exposure
- areas requiring greater disaster management focus

This visualization supports geographical comparison and risk-based analysis.

Conclusion

Milestone-1 completed the essential data preparation phase of the project. Through systematic cleaning, validation, and standardization steps, the FEMA Disaster Declarations dataset has been converted into an organized and analysis-ready form. The prepared dataset and visual placeholders will support trend analysis and deeper insights in **Milestone-2**, where detailed exploratory visualizations and interpretations will be performed.
