VISUALIZING US NATURAL DISASTER DECLARATION – TRENDS AND PATTERNS

Week 3 Documentation

Exploratory Data Analysis in Python

Exploratory Data Analysis (EDA) in Python

1. Introduction

Week 3 focused on **Exploratory Data Analysis (EDA)** of the FEMA Disaster Declarations dataset. EDA is a critical step in understanding the dataset's structure, identifying patterns, and uncovering relationships between variables. The goal was to move beyond cleaning and begin analyzing trends, distributions, and correlations that can inform meaningful visualizations.

2. Step by Step Procedure

2.1 Volume and Structure of Data

- **To Do**: Assess the size and structure of the dataset.
- **Procedure**:

```
df.shape   # returns (rows, columns)
df.info()  # overview of column types and non-null counts
```

- **Findings**:
  - The cleaned FEMA dataset contained **5139 rows** and **21 columns**.
  - Key columns included: disasterNumber, stateName, declarationDate, incidentType, closeoutDate, and program flags.
- **Outcome**: Established the dataset's scale.

2.2 Distribution of Data

- **To Do**: Explore how values are distributed across key fields.
- **Procedure**:

```
df["incidentType"].value_counts()
```

- **Findings**:
  - **Incident Types**: Fire, Severe Storms, and Hurricanes were the most frequent categories.
  - **State**: California, Texas, and Oklahoma are the top 3 states hit by disasters.
  - **Outliers**: Some records had unusually long durations or missing end dates, highlighting ongoing disasters.
- **Outcome**: Identified typical ranges and extreme cases, which are important for visualization.

2.3 Comparison Across Columns

- **To Do**: Compare one variable against another to uncover insights.
- **Procedure**:
    - Compare **incident type vs. state**:

        df.groupby("stateName")["incidentType"].value_counts()

    - Compare **program declarations vs. incident type**:

        df.groupby("incidentType")[["iaProgramDeclared","paProgramDeclared"]].mean()
- **Findings**:
    - Certain states (e.g., coastal states) had higher hurricane declarations, while inland states had more floods and tornadoes.
    - Public Assistance (PA) programs were more frequently declared for floods and hurricanes, while Individual Assistance (IA) was common in severe storms.
- **Outcome**: Comparisons revealed geographic patterns in disaster declarations.

3. Visualizations (Examples)

- **Histogram**: Distribution of incident durations.
- **Bar Chart**: Incident types by state.
- **Heatmap**: Correlation between assistance programs and incident types.

4. Outcome of Week 3

By the end of Week 3:

- The FEMA dataset's **volume and structure** were fully understood.
- **Distributions** revealed typical ranges and highlighted outliers.
- **Comparisons** across states, incident types, and program flags uncovered geographic and programmatic trends.

EDA provided the foundation for building meaningful dashboards and predictive models in subsequent weeks.

5. Next Steps (Week 4 Preview)

In Week 4, we will discuss about connecting **live data sources** into Power BI for dynamic dashboards and designing an **ETL**, and setting up a clear **GitHub repository** to organize project for collaborative project sharing.