Report: Data Doomsday Timed Challenge

By: Ved Dharmatti, Daksh Khuttan, Arun Jiju Joseph

1. Introduction

This report details the comprehensive analysis performed using two major datasets: the INFORM Risk Index and the EM-DAT Disaster Database. The primary aim of the project is to examine global disaster risks by validating risk scores against historical disaster records, forecasting future disaster frequencies, and identifying actionable insights for resource allocation and policy recommendations. In addition to traditional exploratory analysis and modeling, the project features creative visualization and interactive dashboard components to enhance the communication of risk insights.

2. Data Overview and Preprocessing

2.1 Dataset Description

- **INFORM Risk Index:**
  – Consists of country-level risk scores and multiple risk indicators such as vulnerability, exposure, and coping capacity.

- **EM-DAT Disaster Database:**
  – Provides detailed disaster records, including disaster counts, types, affected populations, and temporal patterns (from 2000 to 2024).

2.2 Data Cleaning and Preparation

- **Standardization:**
  – Country identifiers and column headers were standardized across both datasets to ensure consistency.

- **Missing Values and Imputation:**
  – Missing numeric values were imputed using appropriate methods, while ensuring that aggregation and modeling steps remain robust.

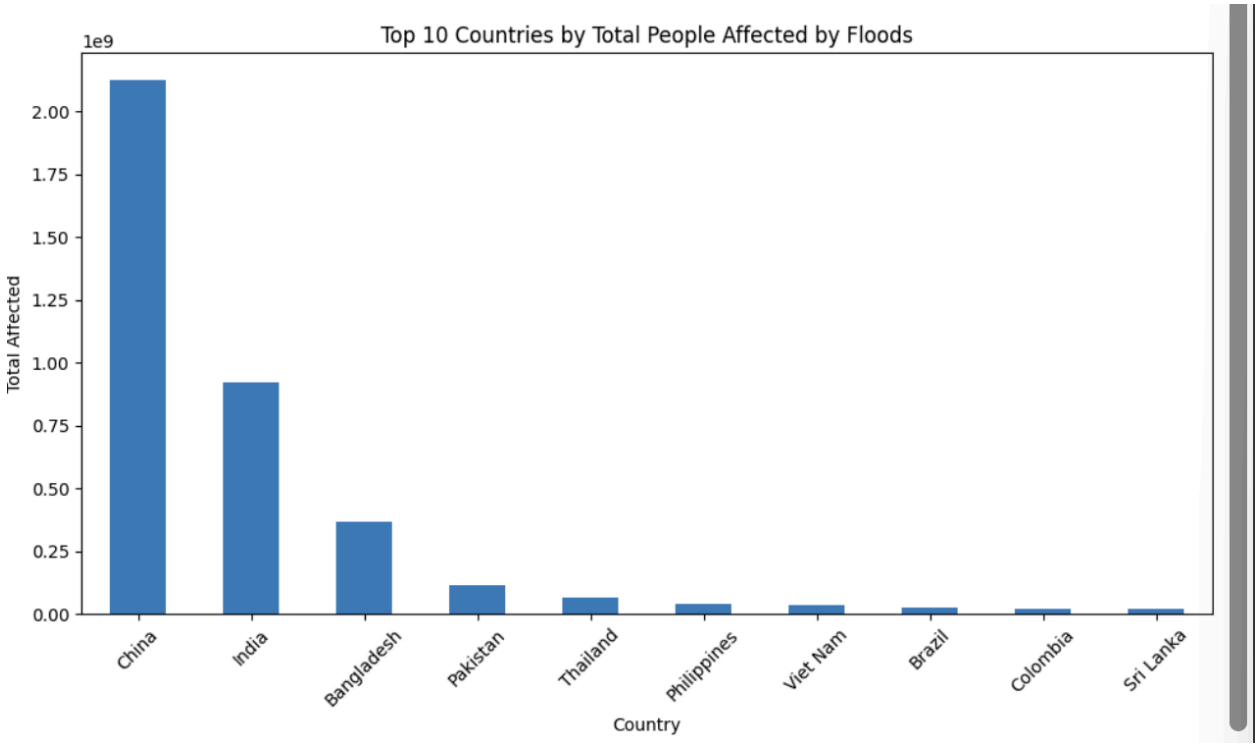
- **Data Integration:**
  – The datasets were merged using common country identifiers, enabling direct correlation of INFORM risk scores with historical disaster events.

3. Exploratory Data Analysis (EDA) & Trend Analysis

3.1 Overview of Risk and Disaster Metrics

- Identification of the top 5 high-risk countries based on INFORM scores.

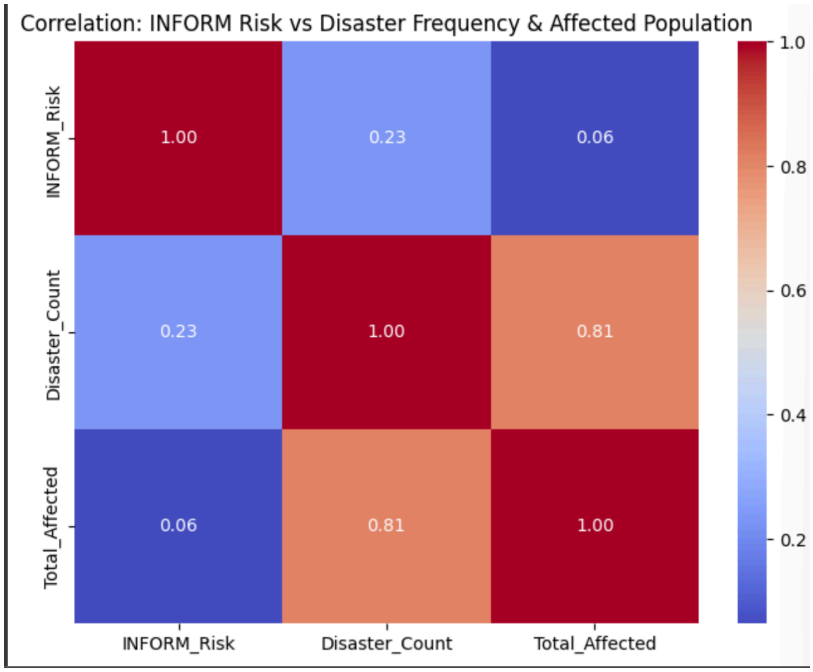
● Computation and visualization of disaster counts (2000–2024) for the top 10 disaster-prone countries as per the EM-DAT database.



3.2 **Correlation and Validation Analysis**

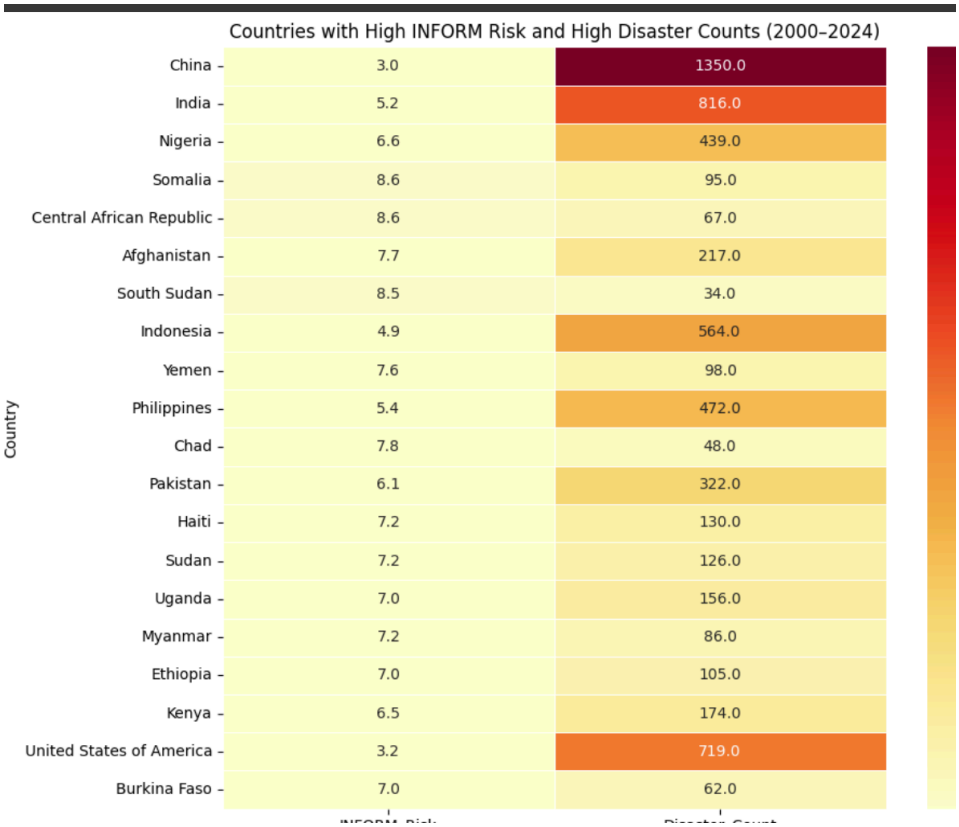
● **Risk vs. Disaster Frequency:**

– Statistical correlation analysis was performed between INFORM scores, disaster frequency, and affected population metrics.



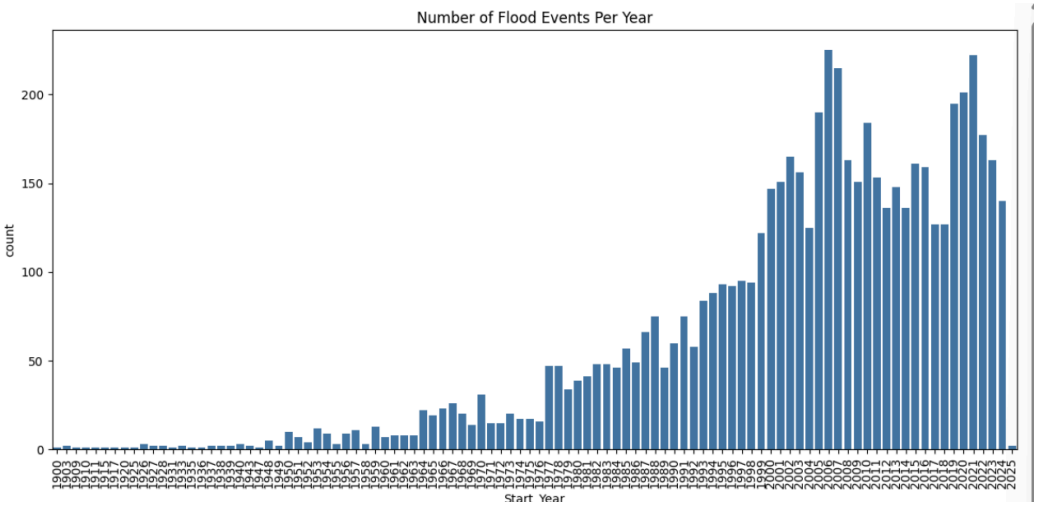
- **Heatmap Visualization:**

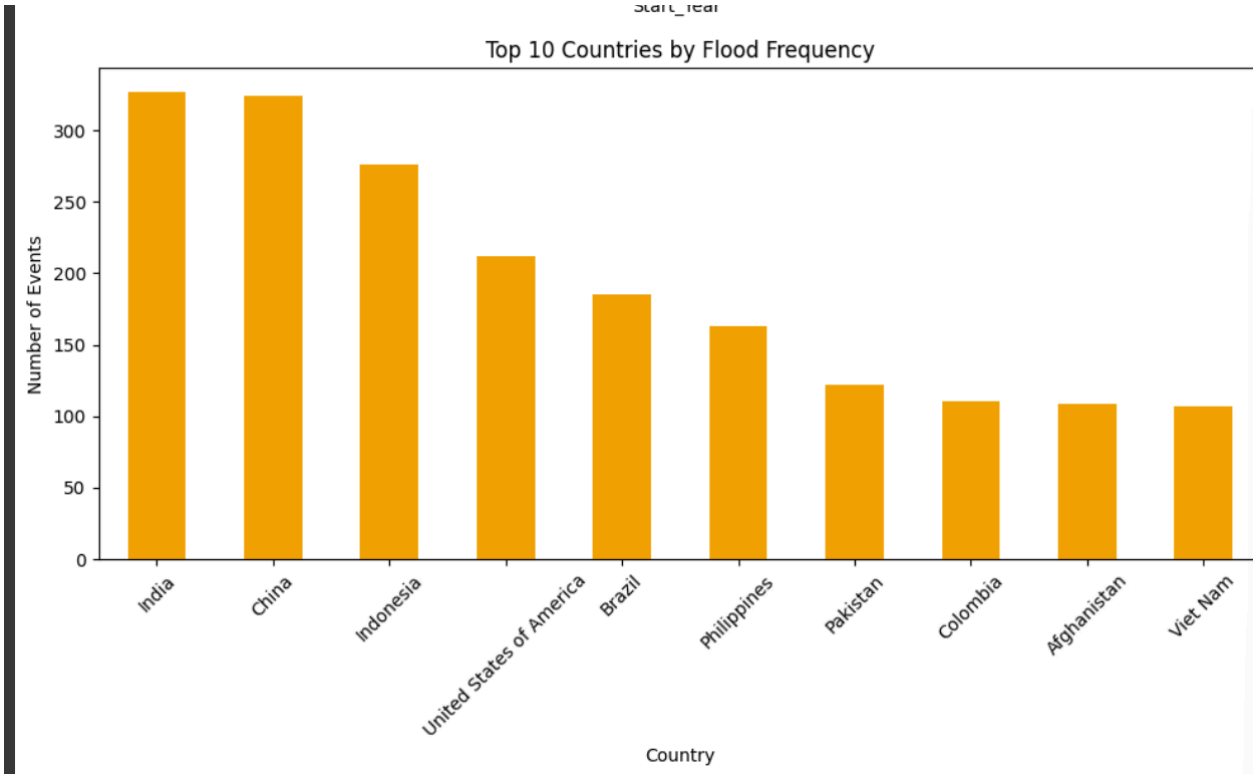
– A heatmap was generated to highlight countries exhibiting both high risk and high historical disaster incidence.



3.3 **Interactive Dashboard Features**

- **Time-Series Plots:**
  - Line charts and bar graphs were used to track disaster trends over the selected period.





- **Geospatial Mapping:**
  - Dashboards allow for interactive examination of regional risk profiles.

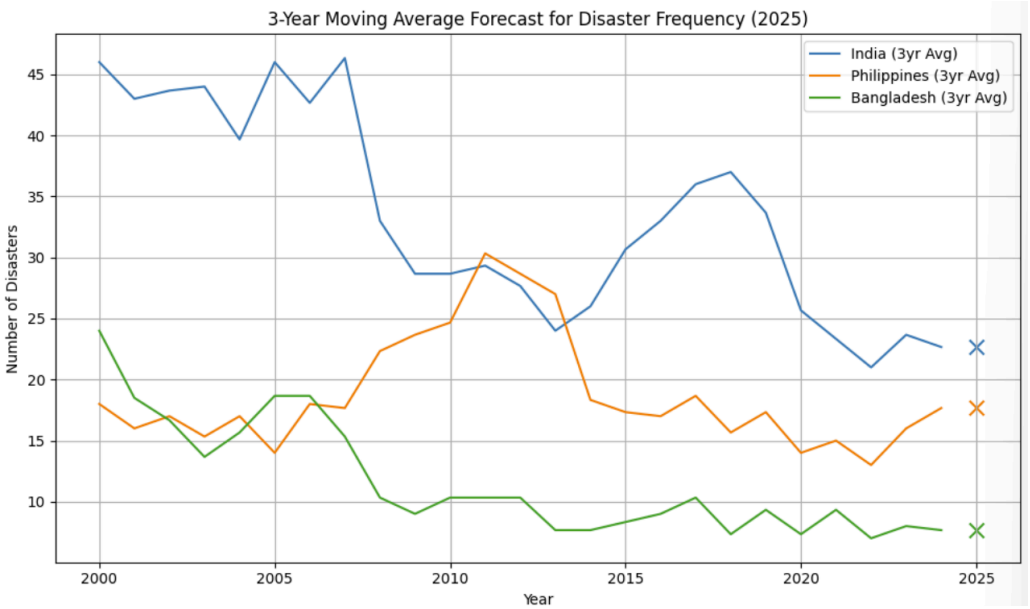
4. **Advanced Analysis and Modeling**

4.1 **Predictive Modeling: Future Disaster Forecasting**

- **Moving Average Forecast:**

– Applied moving average techniques to predict the disaster frequency for 2025 for a selection of key countries.

```
📊 2025 Forecast based on 3-year Moving Average:
       Country  Forecast_2025
0        India          22.67
1  Philippines          17.67
2   Bangladesh           7.67
```



- **Regression Modeling:**

– Developed regression models to forecast disaster-prone regions using INFORM risk indicators.

```
Top 5 Most Predictive INFORM Indicators:
           Feature  Importance
1  Hazard_Exposure    0.474485
3   Lack_of_Coping    0.223861
0      INFORM_Risk    0.181408
2    Vulnerability    0.120246
```

- **Feature Importance and Selection:**

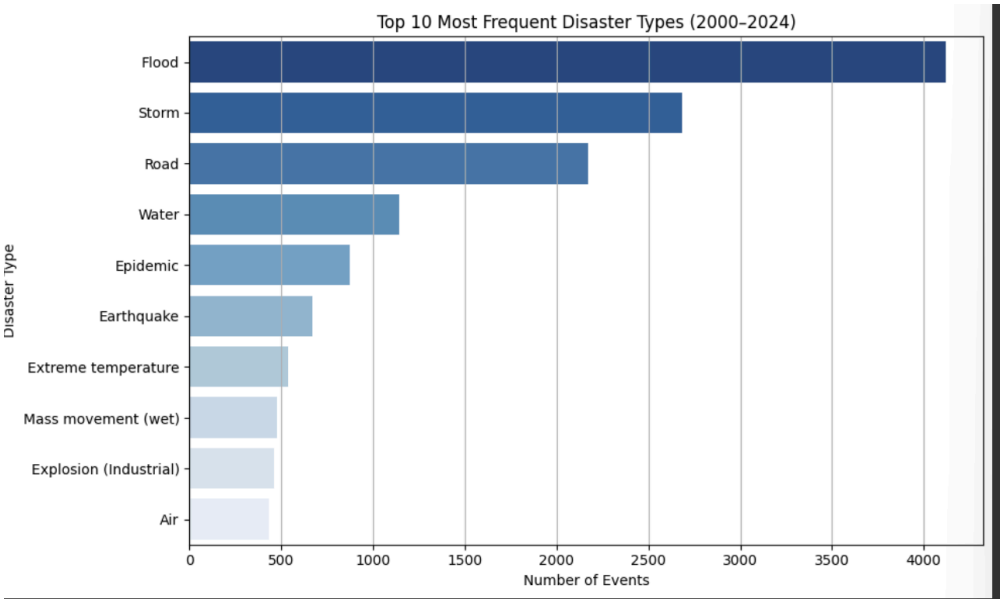
– Employed methods such as decision trees to identify the top 5 most predictive INFORM indicators.

– Compared various feature selection techniques by evaluating performance metrics (accuracy, precision, AUC).

4.2 **Disaster Type-Specific Analysis**

- **Environmental Risk Insights:**

– Analysis of the frequency of disaster types in the EM-DAT dataset provided conclusions regarding prevalent global environmental risks.



- **Clustering Analysis:**

– Clustering countries by disaster type frequency uncovered natural groupings and regional similarities.

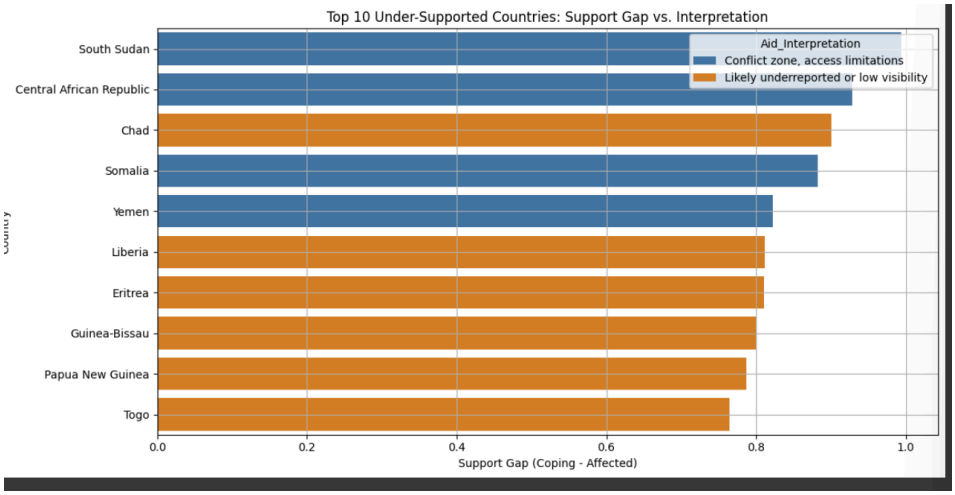
- **Risk Discrepancy Evaluation:**

– Identified 5 countries where high INFORM risk scores were paired with low EM-DAT disaster counts and provided potential explanations (e.g., underreporting, resilience policies, emerging risks).

4.3 **Resource Allocation Optimization**

- **Aid vs. Impact Analysis:**

– Analyzed the disparity between aid allocation and disaster impact to identify regions that may be under-supported or over-supported.

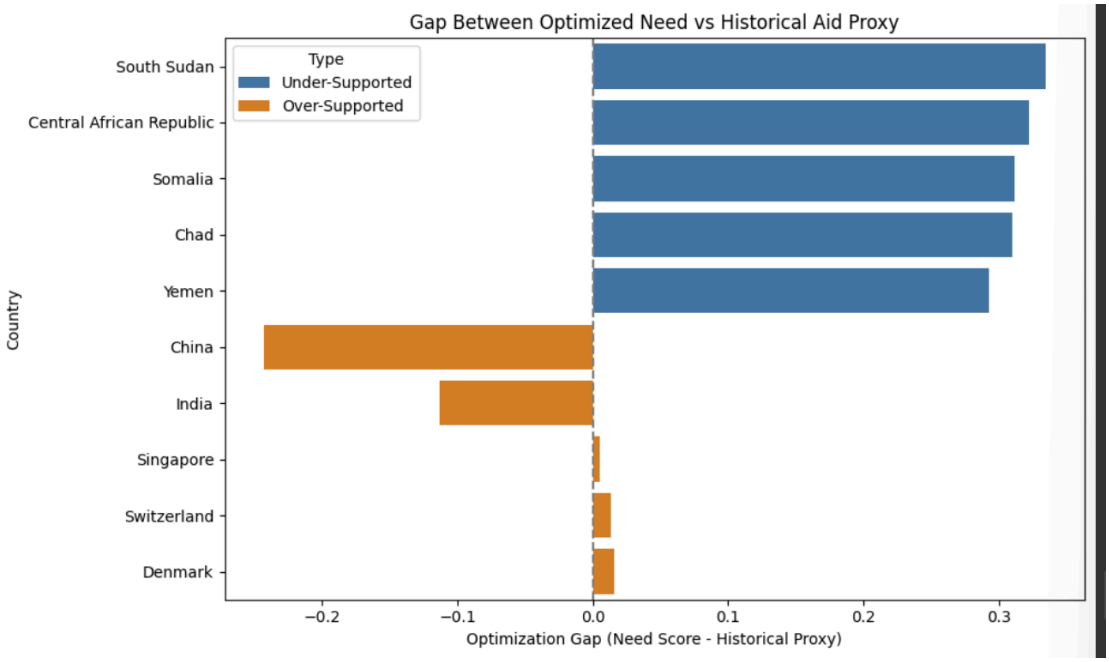


- **Socioeconomic & Geopolitical Insights:**

– Investigated how socioeconomic and geopolitical factors might contribute to imbalances in resource distribution.

- **Optimization Model Evaluation:**

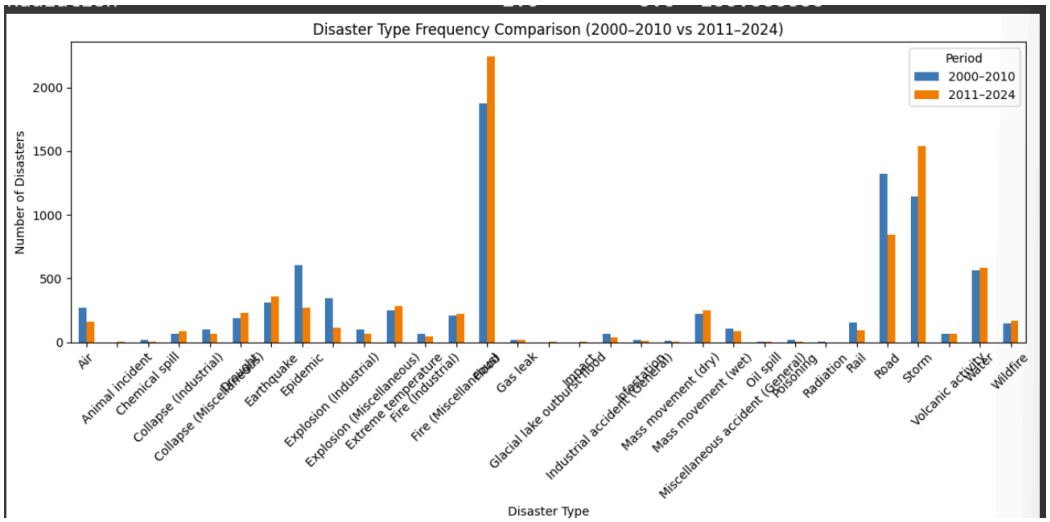
– Compared model-suggested resource allocation shifts against actual past aid distributions to illustrate practical challenges.



4.4 **Climate Change & Disaster Risk Trends**

- **Temporal Trend Analysis:**

– Compared disaster frequencies between two distinct periods (2000–2010 and 2011–2024) to assess the impact of climate change on disaster incidence.



- **Climate-Driven Risk Factors:**

– Analyzed climate-related factors (e.g., urban heat islands, deforestation, glacier melt) that may be driving increasing disaster risks.

- **Emerging Hotspots:**

– Identified and discussed emerging climate-related disaster hotspots based on the analysis of INFORM scores and EM-DAT records.

Emerging Climate Disaster Hotspots with Policy Guidance:

	Country	Climate_Driver	Socio_Environmental_Risk	Policy_Recommendation
0	Libya	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
1	Tunisia	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
2	Italy	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
3	Myanmar	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
4	Colombia	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
5	South Sudan	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
6	Sri Lanka	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
7	Serbia	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
8	Malaysia	General climate risk	Emerging or mixed vulnerabilities	Expand climate data, integrate risk in local p...
9	Brazil	Deforestation	Forest loss, ecosystem disruption	Enforce forest protection, fund reforestation...

4.5 **Policy & Government Readiness Assessment**

- **Readiness Analysis:**

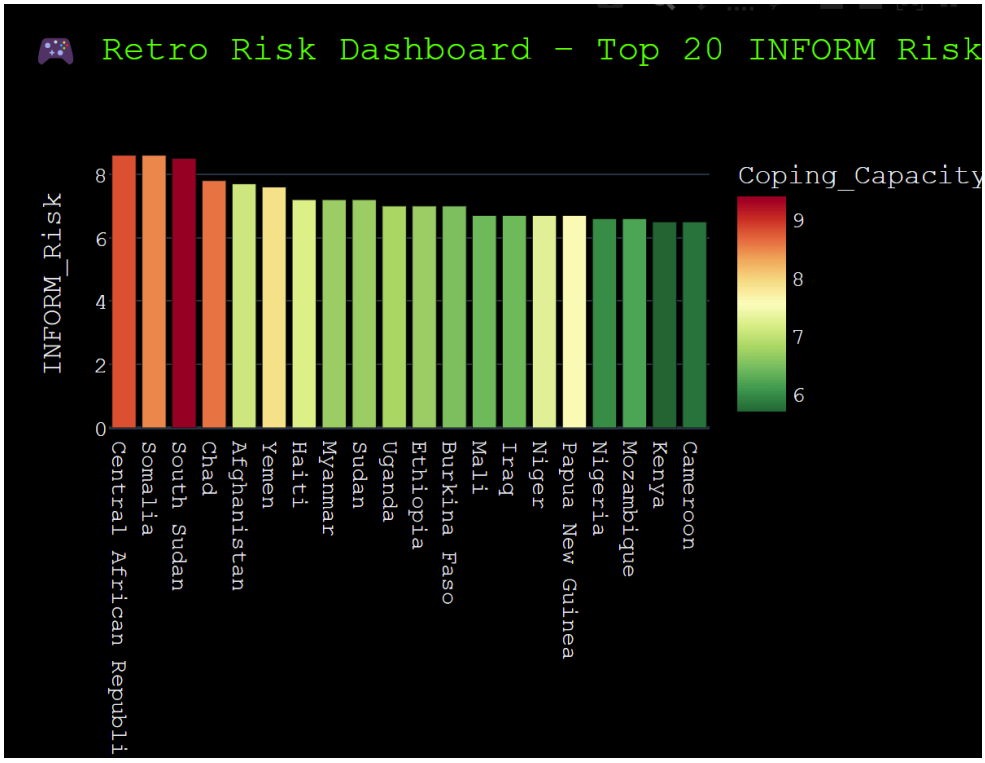
– Evaluated countries with high risk yet low disaster preparedness due to factors such as economic challenges and poor governance.

	Country	INFORM_Risk	Coping_Capacity	Vulnerability	Hazard_Exposure	V
32	Central African Republic	8.6	8.8	8.8	8.1	
158	Somalia	8.6	8.5	8.4	8.9	
160	South Sudan	8.5	9.4	9.2	7.2	
33	Chad	7.8	8.6	7.9	7.1	
1	Afghanistan	7.7	7.1	7.4	8.7	
39	Congo DR	7.7	8.0	7.9	7.2	
189	Yemen	7.6	7.9	8.4	6.5	
163	Sudan	7.2	6.7	7.8	7.2	
72	Haiti	7.2	7.2	6.5	8.0	
118	Myanmar	7.2	6.7	6.3	9.0	

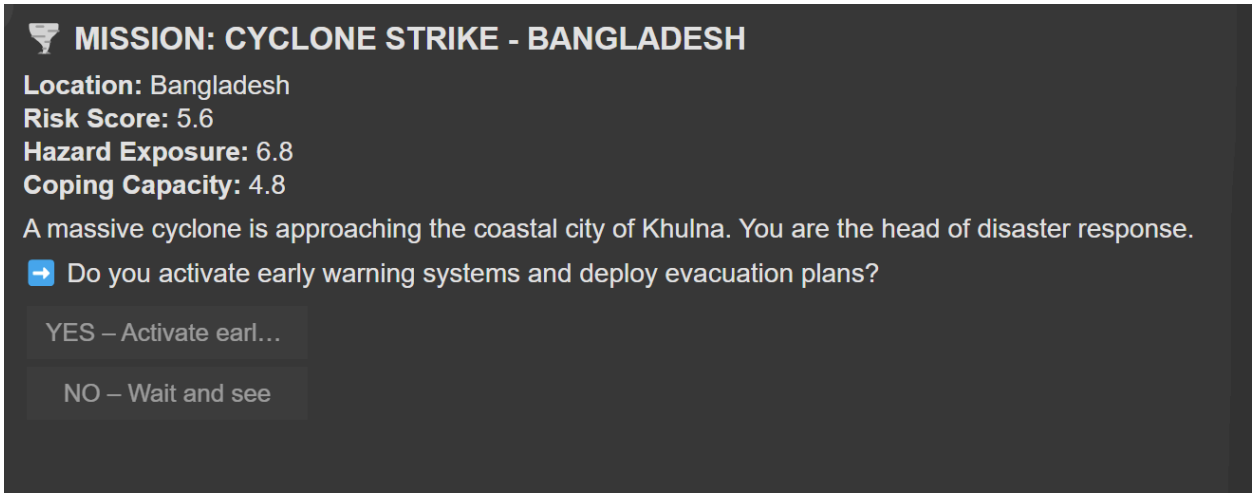
- **Infrastructure Recommendations:**
  – Proposed investments and policy changes aimed at reducing vulnerabilities and bolstering resilience.
- **Case Studies:**
  – Highlighted instances where certain countries have achieved effective mitigation despite high risk, offering best practice insights.

4.6 **Bonus: Creativity & Visualization**

- **Narrative Storytelling:**
  – Developed a retro-style mission briefing narrative to succinctly convey critical disaster risks for a selected country or region.
- **Interactive, Game-Inspired Interface:**
  – Designed an interactive interface featuring tooltips and visual explanations for risk scores and disaster trends.



- **Engaging Storytelling Experience:**
  – Created a branching narrative or timeline animation that guides users through a disaster scenario while illustrating the impact of preparedness levels on outcomes.



5. **Key Findings**

- **Risk Validation and Forecasting:**
  – The strong correlation between INFORM risk scores and historical disaster records

validates the risk assessment framework.

– Forecasting models indicate that moving average and regression techniques can provide valuable short-term disaster predictions.

- **Visual and Analytical Insights:**

– Interactive dashboards and geospatial visualizations enhance the interpretation of complex disaster data.

– Clustering and correlation analyses reveal significant regional variations and potential underreporting issues.

- **Policy and Resource Allocation Implications:**

– Analysis of aid versus impact underscores the need for more targeted resource distribution based on empirical risk assessments.

– Policy evaluations suggest that increased investment in infrastructure and proactive governance can mitigate high-risk vulnerabilities.

6. **Conclusions and Recommendations**

- **Enhance Disaster Preparedness:**

– Reinforce governmental and organizational capacity in high-risk areas through targeted infrastructure improvements and training.

– Increase coordination between disaster monitoring and policy formulation bodies to better align aid with risk.

- **Utilize Data-Driven Insights:**

– Leverage the findings from both risk validation and predictive modeling to inform short-term and long-term disaster management strategies.

– Encourage the continuous updating and integration of additional datasets (socioeconomic, demographic, climate) to refine models.

- **Innovate Communication Methods:**

– Adopt interactive dashboards and storytelling techniques to make complex risk information accessible to policymakers and the public.

– Explore game-inspired interfaces and narrative experiences to raise awareness and educate on disaster preparedness.

7. **Future Work**

- **Model Refinement:**

– Further improve predictive accuracy by incorporating advanced machine learning techniques and additional features.

- **Expanded Spatial Analysis:**
  – Enhance geospatial mapping to enable finer-grained analysis of disaster hotspots and regional vulnerabilities.

- **Broaden Data Integration:**
  – Integrate socioeconomic, environmental, and demographic data to provide a more holistic view of disaster risk and resilience.

- **Interactive Policy Simulations:**
  – Develop interactive simulation models that allow stakeholders to experiment with different resource allocation strategies under varying disaster scenarios.