Disaster Risk Modeling and Analysis Report

Project Overview:  
This project integrates INFORM risk data and EM‑DAT historical disaster data to perform a comprehensive disaster risk analysis. The analysis covers data cleaning and merging, risk validation, regression modeling, predictive forecasting, resource allocation optimization, climate change and disaster risk trends, policy and government readiness assessment, and creative interactive visualizations.

# 1. Data Acquisition and Preparation

1.1 INFORM Risk Data Processing:  
- Uploaded the INFORM Risk Excel file.  
- Selected the appropriate sheet using fuzzy matching (e.g., 'INFORM Risk 2025 (a-z)').  
- The file contained two header rows; the first row with variable names and the second with units/metadata. The first row was promoted to headers, and the second row was dropped.  
  
1.2 EM‑DAT Data Processing:  
- Uploaded the EM‑DAT data file and selected the 'EM-DAT Data' sheet.  
- Converted the 'Entry Date' to datetime and extracted the 'Year'.  
- Filtered the data for the period 2000–2024 and aggregated disaster frequency and impact metrics by ISO code.  
  
1.3 Merging Datasets:  
- Merged the INFORM dataset (using the 'ISO3' column) with the EM‑DAT metrics (using the 'ISO' column) to obtain a combined dataset.

# 2. Risk Validation & Correlation Analysis

Q1. Top 5 High-Risk Countries and Disaster Count Plot:  
- Sorted the merged dataset by INFORM risk score and identified the top 5 high-risk countries.  
- For the top 3, extracted annual disaster counts from EM‑DAT data and plotted a time-series to observe trends.  
  
Q2. Correlation Analysis:  
- Computed the correlation matrix among 'INFORM RISK', disaster frequency, and impact metric.  
- Visualized these correlations with a heatmap to reveal relationships.  
  
Q3. Regression Analysis for Disaster-Prone Countries:  
- Built a regression model using INFORM indicators (e.g., Hazard & Exposure, Vulnerability, Lack of Coping Capacity) to predict disaster frequency.  
- Identified anomalies where high risk did not translate into high disaster occurrence, indicating possible underreporting or effective mitigation.

# 3. Predictive Modeling: Future Disaster Forecasting

Q1. Moving Average Forecast:  
- Used a 3-year moving average on EM‑DAT time series data for selected high-risk countries to forecast disaster frequency for 2025.  
  
Q2 & Q3. Feature Importance and Feature Selection Techniques for Disaster Type Prediction:  
- Implemented a Random Forest classifier to derive feature importances from INFORM indicators.  
- Employed Recursive Feature Elimination (RFE) and Lasso-based logistic regression for feature selection, evaluating performance via cross-validation.  
- The analysis indicated that traditional indicators (e.g., Hazard & Exposure, Vulnerability) have modest predictive power, suggesting that additional variables might be required.

# 4. Disaster Type-Specific Analysis

Q1. Global Environmental Risks:  
- Disaster type frequencies were analyzed to understand global trends, with floods and storms frequently dominating the dataset.  
  
Q2. Clustering by Disaster Type Frequency:  
- A pivot table of disaster type counts by country was created, and K-Means clustering along with PCA was applied to reveal regional patterns.  
  
Q3. Discrepancies between INFORM Predicted Risks and Historical Records:  
- For some countries, high INFORM risk scores did not match high historical disaster counts. Possible explanations include data underreporting, effective mitigation measures, or emerging risks.

# 5. Resource Allocation Optimization

Q1 & Q2. Aid vs. Disaster Impact:  
- Simulated (or actual) aid data was merged with disaster impact metrics to create a scatter plot, identifying countries where aid does not align with disaster impact.  
  
Q3. Optimization Model:  
- A simple linear programming model (using PuLP) was developed to reallocate additional aid in order to minimize a proxy risk function. The model highlights potential redistributions but also points to real-world implementation challenges (e.g., political and logistical constraints).

# 6. Climate Change & Disaster Risk Trends

Q1. Comparison of Disaster Frequency (2000–2010 vs. 2011–2024):  
- The average disaster frequency was computed for two periods. Interestingly, the average in the earlier period was higher than in the later period.  
 This result prompts further investigation into reporting practices, mitigation improvements, and localized climate changes.  
  
Q2 & Q3. Discussion of Climate Factors and Emerging Hotspots:  
- Factors such as glacier melt, urban heat islands, and deforestation were considered.  
- Despite an overall lower frequency in the later period, regional analysis suggests that certain areas remain highly vulnerable and could be emerging hotspots.

# 7. Policy & Government Readiness Assessment

Q1. High Risk vs. Low Readiness:  
- A composite readiness score was calculated using governance, DRR, infrastructure, and communication indicators.  
- Countries showing high INFORM risk but low readiness were identified, potentially due to economic constraints, governance issues, or conflict.  
  
Q2. Policy Recommendations:  
- Recommendations include investments in resilient infrastructure, enhanced early warning systems, economic support, and improved governance.  
  
Q3. Case Studies of Successful Mitigation:  
- By comparing high-risk countries with low historical disaster counts, the study highlights cases where effective policies have reduced disaster impacts, offering best practices for disaster mitigation.

# 8. Bonus: Creativity & Visualization

Q1. Retro-Style Mission Briefing:  
A mission briefing was scripted to engage the audience in a retro narrative style:  
"""Attention all units:  
Our analysis has identified that Afghanistan (AFG) faces extremely high disaster risks with frequent natural hazards.  
Key data indicate an INFORM Risk Score of 7.8 and an alarming 283 recorded disasters.  
Immediate action is required to strengthen response measures and allocate resources effectively.  
Prepare to mobilize and coordinate with local agencies.  
Mission briefing over."""  
  
Q2. Retro Game-Inspired Interface Concept:  
- An interactive dashboard using Dash (or JupyterDash) was prototyped. The interface features retro styling with a dark background, green monospaced text, and interactive elements (e.g., a risk slider and real-time graph updates) to engage the user.  
  
Q3. Interactive Storytelling Experience:  
- A storyboard was developed for an interactive narrative, allowing users to make decisions regarding disaster preparedness. The storyboard includes multiple decision points (e.g., early warning investments and resource allocations) with branching outcomes, ultimately demonstrating how preparedness impacts disaster outcomes.

# Conclusion

This project presents a holistic approach to disaster risk analysis by integrating INFORM risk indicators and EM‑DAT historical data. The analysis covered data cleaning, merging, risk validation, predictive forecasting, resource allocation optimization, and policy assessments. Creative visualization components such as a retro-style dashboard and interactive storytelling were also developed to effectively communicate complex disaster risk information. Future work includes refining predictive models, incorporating more socioeconomic variables, and further developing interactive tools for broader stakeholder engagement.