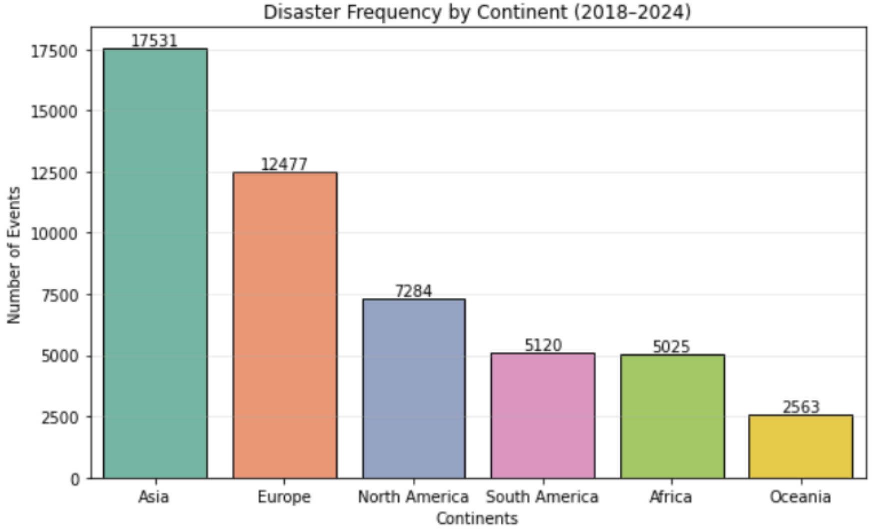
Data Visualization Pitch

Data Visualization Pitch

Student Individual Assignment



*The analysis quantifies the frequency of **disaster events** across continents from **2018 to 2024**, revealing a significant geographic disparity in incident volume. **Asia** emerges as the **predominant** region, exhibiting the highest count of 17,531 events, which substantially outpaces **Europe** (12,477) and **North America** (7,284). This trend indicates a distinct concentration of disaster activity in the **Eastern** Hemisphere, while **Oceania** records the **lowest** incidence (2,563), highlighting a steep decline in frequency compared to the leading **regions**.*

From Natural Hazards to Human Impact: A Trend Analysis of Global Disasters (2018–2024)

Student name: Syed Muhammad Shoaib
Student ID: 946917
Student e-mail address: s.shoaib@campus.unimib.it

About Data

Initial Considerations About the Data

This project is based on the Global Disaster Response Analysis (2018–2024) dataset, which captures major natural disasters across countries worldwide. The dataset covers a wide range of **disaster types**, including earthquakes, floods, hurricanes, wildfires, extreme heat, and storm surges, and provides key indicators like the data of disaster, **country** name with **latitude** and **longitude** values, **severity index** from (1-10), **casualties**, **economic loss**, **aid amount** they have received, It also includes **response efficiency score** (e.g., 30–100), and response metrics such as **response time in hours** (e.g., 1–63) and **recovery duration** in days (e.g., 2–112) enabling spatial analysis of disaster events globally.

Issues Faced While Searching for Data

Finding a suitable dataset was challenging, as many disaster datasets lack important response and recovery information. The selected dataset was mostly clean, with no major null values or outliers. It includes disaster statistics for 20 countries from 2018 to 2024, along with latitude and longitude data.

Dataset Chosen

I selected a global disaster dataset from Kaggle that includes severity, casualties, economic loss, response time, severity index and recovery indicators, making it appropriate for great storytelling

License and Intended Use

The dataset is shared under an open-data license on Kaggle, allowing its use for educational and research purposes. It is used strictly for academic coursework.

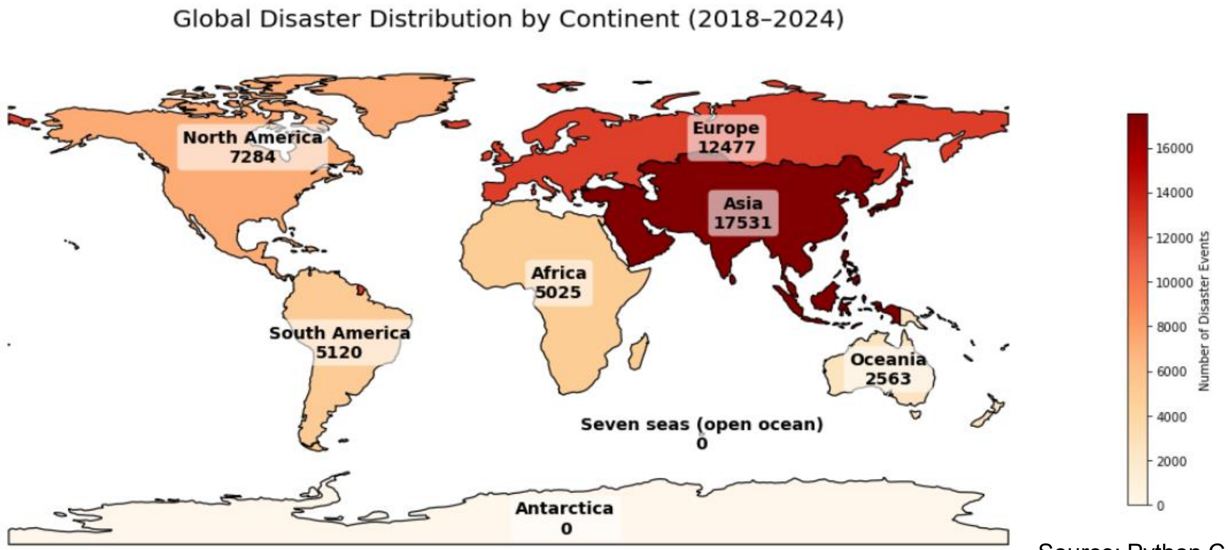
Research questions

The analysis aims to provide an overview of the frequency and impact of global natural disasters, relating quantitative trends in economic loss and severity to response efficiency and aid allocation in the period 2018–2024. Some of the research questions considered are

1. To what extent does continental geography dictate the frequency of natural disasters, and where is the global "risk epicenter" located?
2. Does a nation’s development status act as a shield against mass-casualty events, or is vulnerability universal?
3. Is the intensity of natural disasters following a predictable linear trajectory, or are we entering an era of chaotic volatility?
4. How did the global economic burden of disasters shift during the critical transition from the pandemic era to the post-pandemic recovery?
5. Do wealthy, developed nations possess a significant operational advantage in disaster response efficiency compared to developing regions?
6. Is the duration of recovery driven primarily by national economic capability, or do geological disasters impose a "universal baseline" of reconstruction time that infrastructure cannot easily overcome?
7. Is the global disaster landscape defined by specific regional hazards, or is there a "universal distribution" of risk?
8. What is the precise decay rate of response efficiency, and is there a "golden hour" for disaster intervention?

The Geography of Vulnerability: Global Natural Disaster Trends (2018–2024)

The analysis of spatial distribution reveals a profound geographic imbalance. **Asia** establishes itself as the dominant epicenter of global instability, recording **17,531** events, which accounts for a significantly higher volume than any other region.



Source: Python Code

- Data from 2018–2024 highlights a profound geographic imbalance, with Asia’s incident count exceeding North America’s **(7,284) by 2.5 times** and Oceania’s (2,563) by nearly **sevenfold**, reinforcing the Eastern Hemisphere as the primary locus of hazard density
- Similarly, Europe recorded nearly **double** the frequency of North America, revealing a distinct "Atlantic Gap" of over **5,000** events that suggests longitudinal position is a stronger determinant of risk than development status
- Meanwhile, South America (5,120) and Africa (5,025) demonstrate remarkable **statistical** parity; their near-identical counts imply a uniform "risk floor" across the global south, irrespective of geological differences.

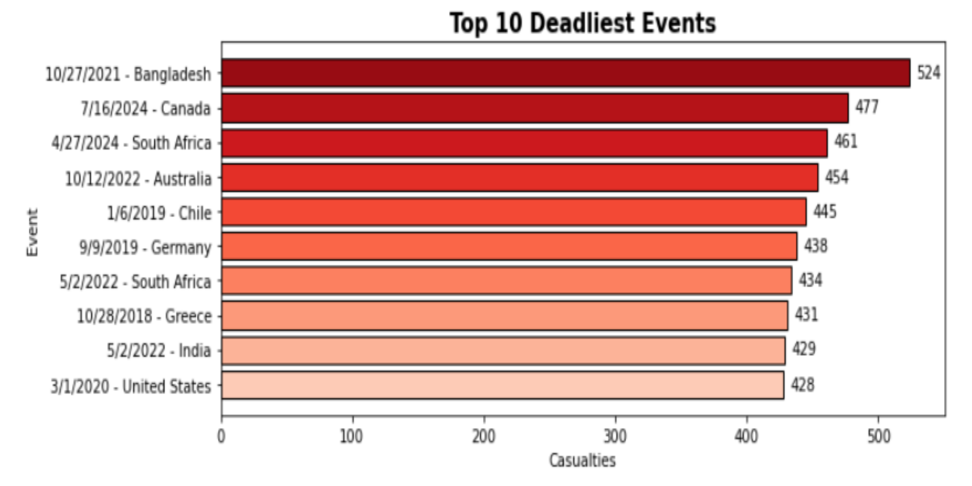
Evaluating the Disparity of Human Loss Across Developed and Developing Economies (2018–2024)

❑ The analysis challenges the assumption of Western immunity, as developed nations like **Canada (477 deaths)** and **Australia (454 deaths)** rank prominently among the top deadliest events alongside developing regions.

❑ Data reveals a disturbing parity in human cost, where the mortality gap between **India (429 casualties)** and the United States **(428 casualties)** is virtually non-existent, rendering GDP a negligible variable for survival in extreme scenarios.

❑ Recent high-mortality incidents in **2024** within **G7** nations like Canada indicate that the intensity of modern hazards is outpacing the historical design parameters of advanced infrastructure

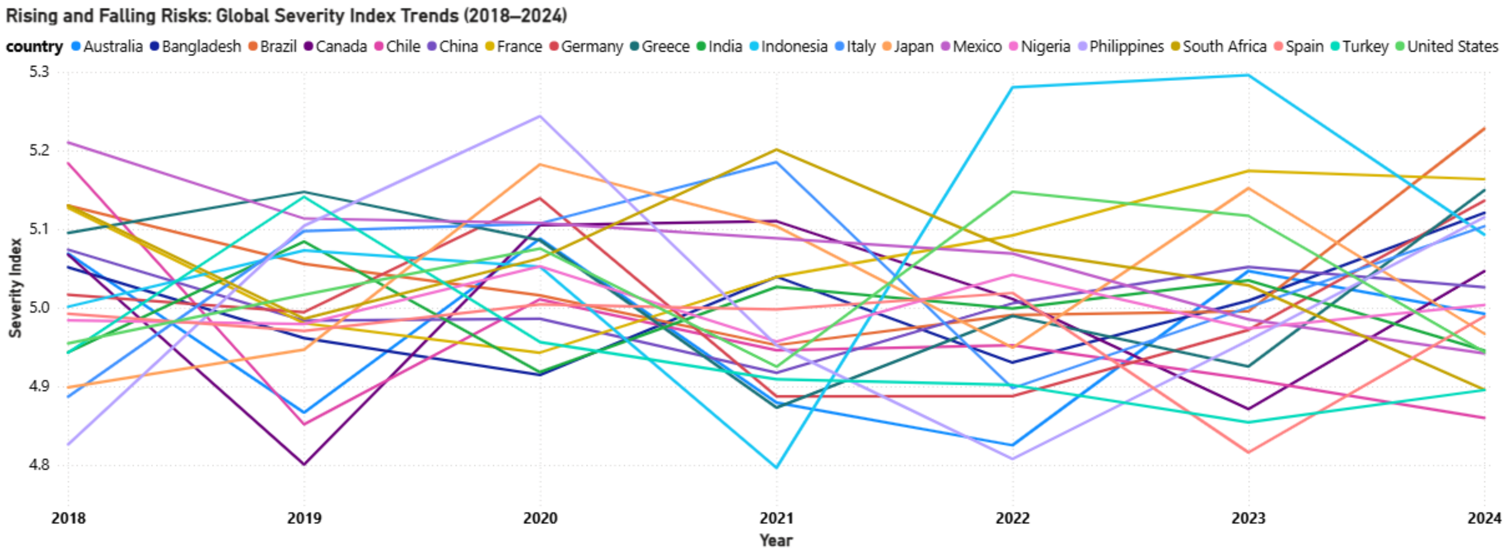
❑ Ultimately, the data proves that vulnerability is universal, with the single deadliest event in **Bangladesh (524)** only marginally exceeding losses in the most developed economies.



Source: Python Code

Top 10 Deadliest Events from the year 2018-2024 from the Global Disaster Data

The Pulse of Instability: Asynchronous Volatility in Global Severity Index



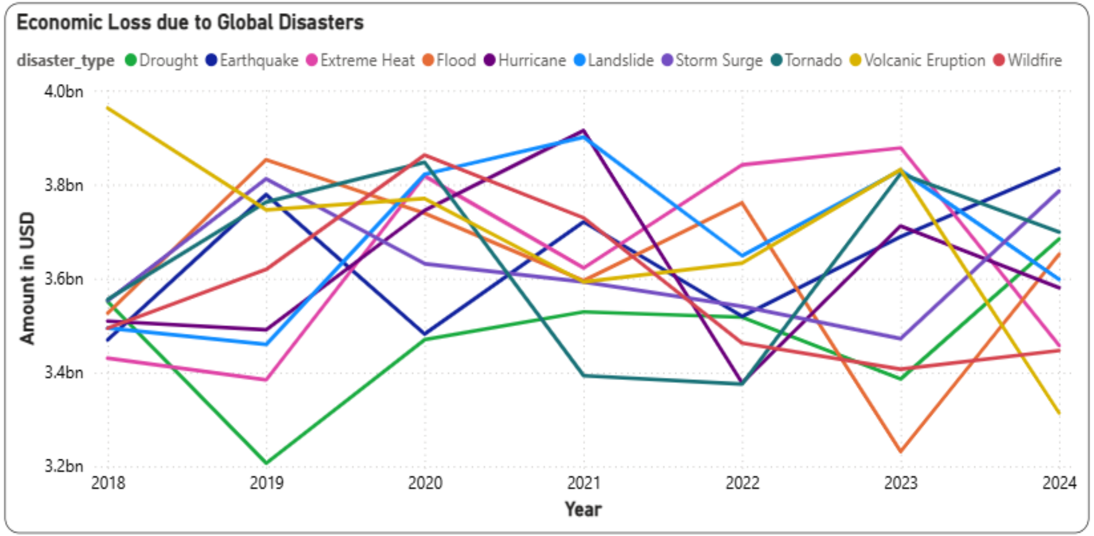
Source: Power BI

- ❑ The data proves that high-intensity risks shift geographically rather than striking globally in unison; the **Philippines** defined the crisis era with a peak severity of **~5.25** in **2020**, only for its risk to subside just as **Indonesia** surged to a global maximum of **~5.3** in **2022** and even more high in **2023.**
- ❑ Moving into **2024** this extreme variance suggests that the global system underwent a period of high instability, likely where specific disaster types intensified while others remained dormant. It marks **2020** not just as a year of high impact, but of **uneven** impact, and in **2022** and **2023** the impact of disasters were usually high in most countries like France, Japan, united States, China etc)
- ❑ Between 4.86 and 5.23, the lines converge but remain intertwined, indicating that toward the end of the study period the world is not becoming safer or more predictable; instead, several risks persist simultaneously at moderate to high levels

Emerging Threats and the "Episodic" Nature of Disaster Costs

- The **2018** fiscal landscape was defined by a singular statistical anomaly: while most disaster types remained within standard variance, **Volcanic Eruptions** exhibited a vertical surge, manifesting as a **'Super-Event'** with economic losses reaching **~3.9 billion USD**. In contrast, the pandemic interval (**2019–2022**) revealed a shift in the hierarchy of risk. During this period, global economic attrition was not driven by a single outlier but was sustained by a cluster of high-frequency **hazards**—specifically **Floods**, **Wildfires**, **Hurricanes**, and **Landslides**—which collectively maintained a massive loss plateau between ~**3.8** and ~**3.9 billion USD**

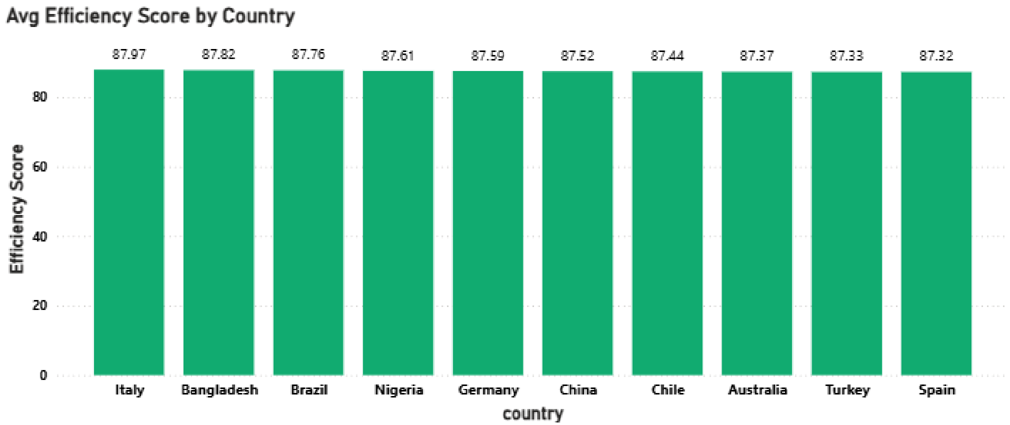
- As environmental factors rapidly evolve; new disaster types are overtaking historical norms. **Extreme Heat** offers a stark example of this volatility: after years of dormancy **(2018–2021),** it surged violently between **2022** and **2023**, incurring a staggering ~**3.8 billion USD** in losses each **year**. while hazards like storm surges and earthquakes remained moderate. This shows that high-cost disasters are episodic that strike with overwhelming intensity before receding, making **long-term** prediction increasingly complex



Source: Power BI

A Comparative Assessment of National Response Capabilities (2018–2024)

- The data reveals a "**flat**" landscape of competence. The difference between the highest-performing nation, **Italy** (**87.97**%), and the lowest in this top-tier group, **Spain** (**87.32**%), is a negligible **0.65**%.
- This statistical flatness implies that disaster response—whether managed by the **EU** or **local agencies** in the Global **South**—has hit a "**competence ceiling**."
- **Bangladesh vs. The G7**: **Bangladesh**, Bangladesh achieved an efficiency score of 87.82%, ranking second globally and surpassing major economies such as Germany (87.59%) and Australia (87.37%)
- Similarly, **Nigeria** (**87.61**%) surpassed **China** (**87.52**%) and **Spain** (**87.32**%). This indicates that nations with frequent exposure to disasters (like Bangladesh and Nigeria) may have developed "**battle-hardened**" response mechanisms that are just as effective than wealthier, more bureaucratic nations.

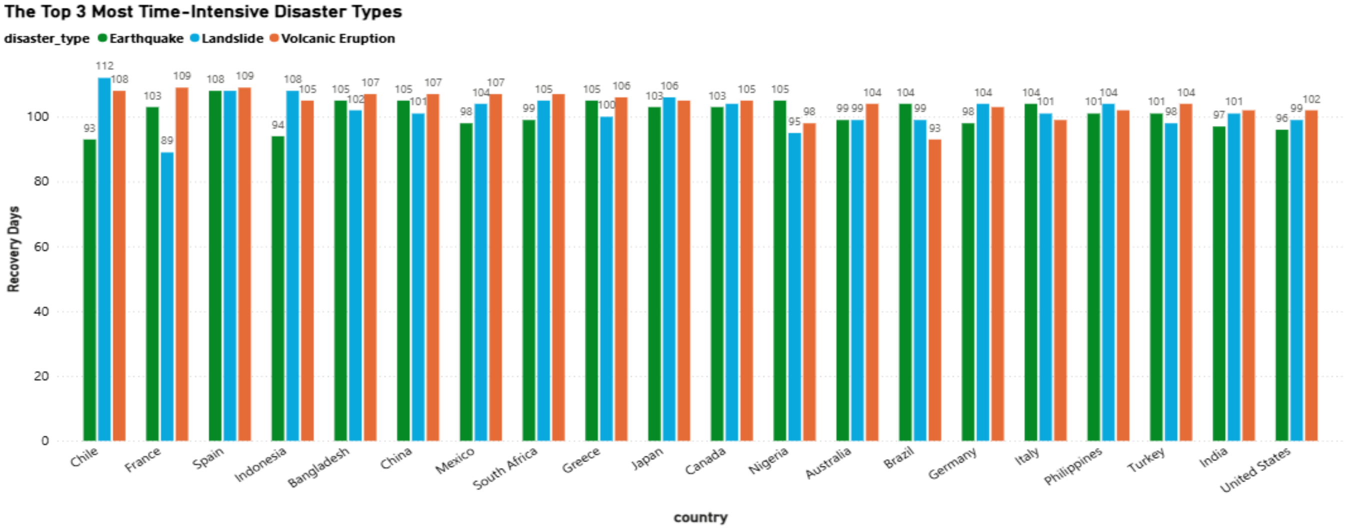


Source: Power BI

The "**Efficiency Score**" reflects a globalized **humanitarian** system where resources and management techniques are standardized, effectively levelling the playing field between the Global North and South

Why Geological Disasters Defy Quick Resolution Regardless of Geography (2018–2024)

- The **"Geological Triad"** of Delay This analysis isolates the three specific disaster types identified as the most time-intensive: Earthquakes, Landslides, and Volcanic Eruptions. unlike transient weather events (like storms that pass quickly), these geological hazards inflict structural damage that necessitates a prolonged **"reconstruction phase"** rather than a simple cleanup. The data shows that across almost all nations frequently peaking near or above **89–112** days
- One might expect the **United States** or **Germany** to recover significantly faster than **Brazil** or Nigeria. However, the data refutes this. The United States shows recovery times ranging from **96** to **102** days, which is statistically almost identical to **Nigeria** (95–105 days) and **Brazil** (99–104 days).

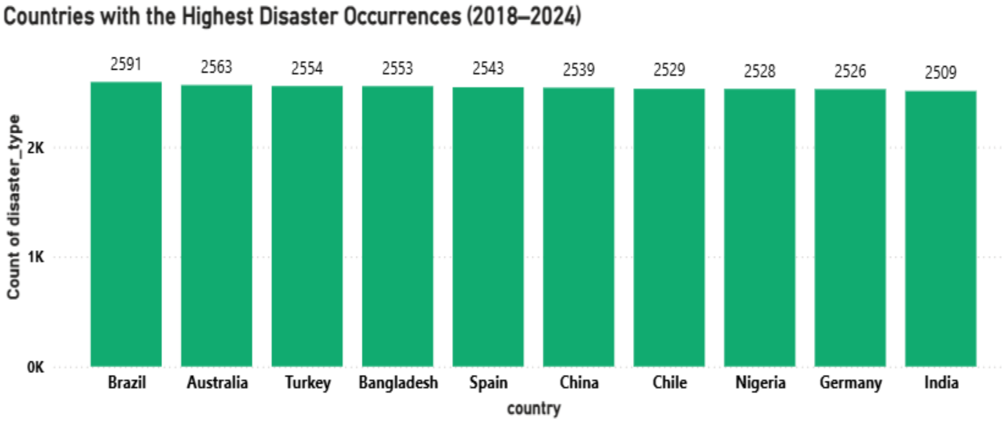


Source: Power BI

Chile presents a unique data point. While it suffers significant delays for Landslides (112 days, the highest on the chart) but the fastest recovery for **Earthquakes (93 days)**

The Statistical Parity of Disaster Frequency (2018–2024)

- This suggests a **"Universal Hazard"** paradigm. It implies that over the **2018–2024** period, the global system generated environmental stress in equal measure across all vectors
- **Statistical Parity:** The variation in disaster frequency between the most affected nation, **Brazil** (2,591 events), and the **tenth** most affected, **India** (2,509 events), is a negligible **3.1%**.
- Despite contrasting climates, countries such as Germany (2,526) and Nigeria (2,528), as well as Australia (2,563) and Spain (2,543), experienced nearly identical disaster frequencies.
- When viewed alongside previous analysis (Severity and Efficiency), a critical insight emerges. While frequency is evenly distributed (everyone faces ~2,500 events) between (2018-2024)
- This implies that the "**risk**" of an event occurring is universal. The difference in human loss , economic cost and disaster type is therefore driven by the **magnitude** of the specific event.

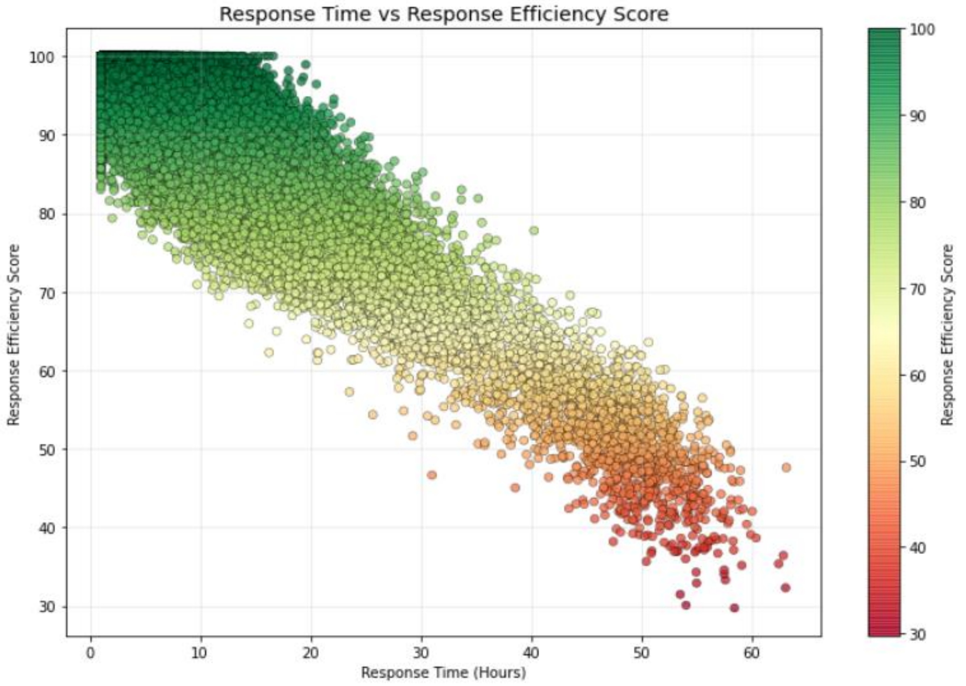


Source: Power BI

Geographic location, latitude, and development status do not insulate a nation from the number of events it faces; risk exposure has been **globalized** and **homogenized**

The Clock is the Enemy: The Quantifying the Hourly Erosion of Response Capabilities

- **Deterministic Correlation:** A rigid inverse relationship exists where operational speed dictates success; efficiency scores degrade linearly from **99 to ~32** as response times lengthen, **proving delay** is the primary **driver** of **failure**.
- The **"Golden Window"**: Superior performance or well efficiency score of >80 is exclusively achievable within the first **0–20** hours, creating a critical timeframe for effective intervention.
- **Measurable Decay:** Operational competence decays at a precise rate of approximately **~1.2** points per hour, meaning a **10-hour delay** reliably costs a nation **~12%** of its response quality.
- **Terminal Tipping Point:** The 40-hour mark serves as a **hard limit;** beyond this threshold, no event achieves a score above **70**, confirming that late interventions suffer from **irreversible** inefficiency.



Source: Python Code

This proves that late intervention is inherently flawed; **no amount** of money or manpower can **compensate** for the initial **loss** of **time**

Insights from the Data

- **The Myth of Western Immunity**: Analysis reveals a disturbing key indicators where developed nations like Canada, Germany, Spain and Australia rank alongside developing regions in death tolls, proving that high GDP offers negligible protection against the sheer intensity of modern hazards.
- **Golden Response Window:** Response efficiency declines by ~1.2% per hour, with a critical tipping point at 40 hours after which effective intervention becomes statistically unlikely.
- **Global Efficiency Ceiling:** Response efficiency converges around ~87% worldwide, with countries like Italy leading and developing nations such as Bangladesh often outperforming G7 economies, indicating standardized global aid capacity.
- **Bounded Severity Volatility:** Disaster severity exhibits region-specific spikes but remains globally constrained within a narrow 4.8–5.3 range, reflecting controlled yet asynchronous risk dynamics
- **Episodic Economic Shocks:** Financial loss is not chronic but driven by "Super-Events," where singular outliers (like the 2018 Volcanic Eruptions) or sudden surges in Extreme Heat (2022–2023) dwarf the baseline costs of routine hazards.
- **The Universality of Risk:** Structural analysis debunks the idea of "disaster zones," revealing a near-perfect statistical uniformity where every major nation faces a consistent barrage of ~2,500 incidents, making risk exposure a geographic constant rather than a regional variable.

Methodology

- **Data Collection:** The primary dataset was acquired from **Kaggle**, ensuring access to a comprehensive repository of historical disaster events for global analysis.

- **Data Processing & Tools:** Data cleaning and exploratory analysis were conducted using Python (Jupyter) with Pandas, NumPy, Matplotlib, Seaborn, and Geopandas was used for cleaning, analysis, and visualization, while Power BI enabled interactive dashboards and deep-dive analytics.

- **Data Transformation:** A key transformation involved the aggregation of geospatial data using **pycountry_convert** library from **python**. Initially, plotting individual disaster locations resulted in visual clutter due to the high density of affected regions. To resolve this, **countries** were **grouped** by **continent**, allowing for a clearer, continent-wise visualization that significantly improved the interpretability of spatial trends.

- **Use of AI Tools: ChatGPT** assisted in generating and debugging Python code,. However, the logic construction, visual design, and analytical interpretation within Power BI were performed entirely manually without AI intervention.

- **Analytical Techniques:** The study combined Descriptive Statistics, Comparative Analysis, and Trend Analysis to summarize risks, compare nations, and track disaster severity and frequency over time.

- **Reproducibility:** Full reproducibility is ensured through a public **GitHub** repository. The complete codebase **(Jupyter Notebooks)** and the **Power BI** source file (.pbix) are available for download, allowing third-party researchers to replicate the findings using the same dataset.

LICENCE

Slides are shared with the following license:

Developed using data from the Kaggle, with charts and analyses produced in **Python** and **Power BI**.

About Me

This is **Syed Muhammad Shoaib**, a Software Engineering graduate and Google Certified Data Analyst, currently pursuing a Master's degree in Data Science at the University of Milan-Bicocca, Italy.

I have over **two years** of professional experience as a Business Intelligence Analyst at Keenu, along with global freelance experience as a Data Analyst on Fiverr, where I have delivered data-driven solutions to more than 30 clients across 20+ countries.

My technical expertise includes Python, C#, SQL, SSIS, SSRS, SSAS, Power BI (DAX, Power Query, RLS), Excel, ELK Stack (Elasticsearch, Logstash, Kibana), Microsoft Fabric Dataflows, Microsoft Azure, Snowflake, and Big Data Analytics. I am experienced in data visualization, data processing and cleaning, ETL pipeline development, data warehousing and data marts, as well as requirements gathering and technical documentation (SOPs, analysis reports, and specifications).I am a quick learner with a strong ability to absorb new concepts, ask effective questions, and solve complex problems through data-driven decision-making and stakeholder-focused communication.

Contact Me

https://www.linkedin.com/in/syed-muhammad-shoaib/

https://shah-1510.github.io/Portfolio/