# **First Project: Research on Natural Disasters | Risk Control Organization- SafeHorizon**

## **Introduction**

## I used this dataset, retrieved via the API Disaster SAFE, which can be found at <https://dashboard.disastercheckin.app/documentation>, to enhance my skills in data cleaning, analysis, and problem-solving.

## Throughout the process, I applied various techniques to ensure the data was accurate and meaningful for analysis. This project involved developing a business model based on the data insights, with a focus on practical applications and strategic decision-making. Working independently, I handled each stage, from research to technical execution, showcasing my ability to work with real-world data and draw actionable conclusions.

## **Objective of the Analysis**

My objective was to creatively and strategically approach the data cleaning process for a disorganized dataset. Focusing on the top natural disaster occurring worldwide on the past month, as well the top countries most affected by them.

I developed a hypothesis focused on identifying countries vulnerable to natural disasters, aiming to help prevent them and protect communities. This hypothesis guided my data wrangling efforts. The overall goal of the analysis was to ensure data accuracy and integrity while identifying key trends and patterns in disaster occurrence and severity. By extracting actionable insights, I aimed to inform a risk management strategy that addresses these vulnerabilities, driving informed decision-making and supporting effective disaster prevention and community protection initiatives.

## **Working Hypothesis**

**Hypothesis 1:** "The prevalence of certain disaster types, such as floods, is significantly associated with specific geographical and climatic conditions, particularly in warm and dry zones"

**Supporting Ideas:**

As the Southern Hemisphere enters its warmest seasons while the Northern Hemisphere moves into winter, we observe a **predominance of fires** due to extremely high temperatures in southern regions.

The rise in global temperatures, combined with **less rainfall and increased dryness**, creates ideal conditions for fires. This correlation suggests that the **frequency of fires** may be linked to climate change, as rising temperatures dry out vegetation and decrease humidity levels. *(More information can be found at* [*https://wmo.int/media/news/global-temperature-record-streak-continues-climate-change-makes-heatwaves-more-extreme*](https://wmo.int/media/news/global-temperature-record-streak-continues-climate-change-makes-heatwaves-more-extreme)*, sources by World Meteorological Organization)*

**Company Perspective:**

My company suggests that the current surge in fire incidents serves as a call to action for long-term climate resilience strategies that extend beyond immediate fire control measures.

**Recommendations:**

* Advocate for broader climate action initiatives, such as reducing greenhouse gas emissions and promoting reforestation efforts
* Establish systems for continuous monitoring and data collection regarding fire incidents and environmental conditions, which can involve:
  + Remote Sensing Technology like using satellite image to monitor endangered and vulnerable areas for forest fires.
  + Implementing early warning systems to alert communities and authorities about potential fire outbreaks.
  + Conduct controlled burns to reduce fuel loads in forests.
  + Investment in Fire Management Resources like providing local fire departments with resources for effective firefighting and prevention.

**Observations:**

* The **high incidence of forest fires** in regions like **Australia** and **Brazil** poses significant risks to **ecosystems, wildlife, and human settlements**. These fires are likely aggravated by climatic factors such as **drought, high temperatures, and prolonged dry seasons**.

**Conclusion:**

* Countries like **Australia** and **Brazil** are currently facing elevated fire risks, particularly as they enter their **summer months**. The combination of **extreme heat, drought, and reduced rainfall** contributes to the increasing occurrence of wildfires, highlighting the urgent need for proactive disaster management and climate resilience measures.

**Hypothesis 2**: "The countries most frequently affected by natural disasters are not necessarily the same as those experiencing the highest severity of incidents, indicating a complex relationship between disaster occurrence and impact"

**Supporting Ideas:**

* **Difference Between Frequency and Severity**: Some countries may experience a high number of natural disasters (ex:. frequent storms or floods), but the severity and impact of these events can vary greatly. In contrast, other countries may face fewer disasters, but when they do occur, they result in more severe damage, higher death tolls, or greater economic disruption.
* **Underlying Factors**: The relationship between disaster frequency and severity is influenced by various factors such as **geographical location**, **infrastructure quality**, **economic preparedness**, and **response capacity**. For example, countries with **good and strong infrastructure** and **efficient disaster response systems** may experience frequent disasters but can mitigate the impact, while others with limited resources may suffer significantly from fewer but more destructive incidents.
* **Global Disparities**: Wealthier nations may experience fewer casualties and less severe damage due to better preparedness and resilience strategies, while developing nations might be hit harder by single events due to fragile infrastructure and limited resources.

**Company Perspective:**

* The goal of this analysis is to **identify and understand the differences** between countries frequently affected by natural disasters and those suffering from high-severity incidents. This knowledge is essential for creating **targeted strategies** that address not only disaster frequency but also the potential severity and long-term impacts of these events.

**Recommendations:**

1. **Customized Risk Management Plans**:
   * Develop personalized disaster management strategies that focus not only on frequency but also on preparing for **high-severity events**. This involves recognizing that some countries may need to prioritize **infrastructure resilience** and **emergency response systems**, even if they are not frequently impacted by disasters.
2. **Disaster Severity Index**:
   * Create a **Disaster Severity Index** that takes into account not only the number of incidents but also the power of the damage, human impact, and economic disruption caused by each disaster. This helps provide a clearer picture of countries that are vulnerable to high-severity events and help prioritize action on them.
3. **Infrastructure and Community Resilience**:
   * For countries experiencing **high-severity disasters**, focus on investing in **strong infrastructure** upgrades and **community resilience** programs. This could include reinforcing buildings, establishing flood defenses, improving **evacuation routes**, and ensuring that communities are trained and prepared for disaster scenarios.

**Observations:**

* **Countries like Portugal and Zimbabwe** may not experience natural disasters as frequently as others, but when they do occur, the severity of incidents—whether in terms of economic loss, environmental damage, or human casualties—can be very high compared to countries were this disasters may occur more frequently.
* **Portugal’s recent trends** in disaster severity (such as severe fires or floods) highlight how a single, intense event can have significant consequences, especially when climatic and geographical factors aggravate the situation.
* **Zimbabwe**, on the other hand, faces challenges due to **underdeveloped infrastructure** and **limited disaster response resources**, meaning that even moderate natural disasters can result in widespread devastation, since it is a developing country that has not much resources in order to have a fast response to these situations.

**Conclusion:**

* There is a **complex relationship between the frequency and severity of natural disasters**. Countries experiencing frequent disasters, like **Japan** or **the United States**, may have developed robust systems to mitigate their effects, while others, like **Paraguay** or **Zimbabwe**, might face fewer disasters but struggle with the severity of each incident due to limited preparedness.
* Understanding this complexity allows for more effective, **data-driven disaster management strategies**. It emphasizes the need for **targeted interventions** that consider not only the frequency of the disaster but also the **potential severity** and the country’s ability to cope, recover, and rebuild.

## **Risk Control Organization- What is?**

A Risk Control Organization is an entity focused on identifying, assessing, and mitigating risks that threats people, infrastructure, and communities.

In the context of natural disasters, its primary goal is to minimize the exposure of populations to these risks by implementing proactive strategies and enhancing disaster preparedness.

Here it shown in more detail what this organization do to prevent devastating situations:

* **Study the locations**: By analysing data on possibly affected regions, the organization gains insights into patterns and risk factors in specific areas. This allows to define and personalize the contingency plans for each area, ensuring that responses are quick, effective, and aligned with the unique risks of each location.
* **Improve infrastructure**: Depending on the types of disasters most common in a region (floods, earthquakes, fires, etc.), the organization works on enhancing infrastructure. For example, in flood-prone areas, they might advocate for improved drainage systems or flood barriers, while in earthquake zones, they could focus on strengthening building codes to withstand seismic activity.
* **Increase security and resilience**: The organization develops strategies to increase the safety and resilience of communities, ensuring that they are better prepared to face disasters. This can include everything from public awareness campaigns, emergency training, and drills to setting up early warning systems that give people more time to act when a disaster strikes.
* **Reduce vulnerability**: By addressing weaknesses in both infrastructure and community preparedness, the organization works to reduce the overall vulnerability of populations. This means minimizing the potential impact of disasters and ensuring communities recover more quickly and efficiently when they do occur.

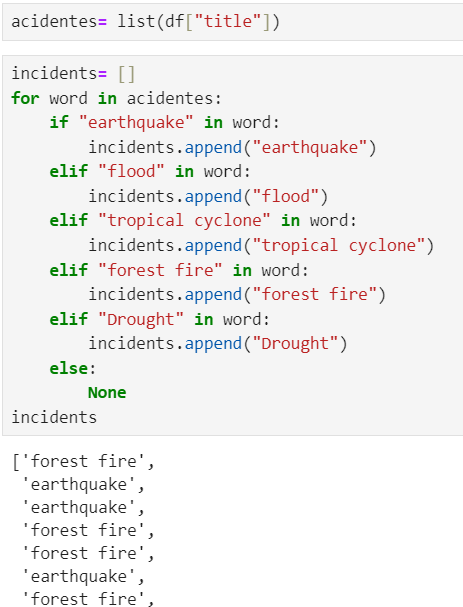
In summary, a Risk Control Organization plays a crucial role in disaster prevention and management by studying risks, improving infrastructure, enhancing community safety, and reducing overall vulnerability to natural disasters. With this project my company wants to implement better contingency planning to countries that need prevention like anti sismic houses, better forest cleaning in danger areas, coast protection for flooding, and more.

## **Data Cleaning and Preparation**

In order to get enough data to draw accurate and truthful insights I made 60 requests to the chosen API Disasters SAFE. Each request made concerns one page of information, ended up with an initial data set with 1500 rows × 12 columns.

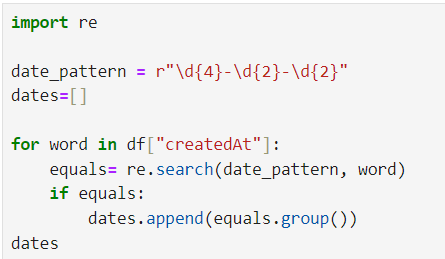
With big pieces of data comes also an intense process of cleaning so let’s break it down step by step.

1. **Dropped Unnecessary columns**: I analysed the original data set regarding my hypotheses and dropped the columns that were unnecessary for my insights. Therefore, I chose to eliminate *'url', 'id', '\_deletedAt', 'latlong', 'guid', 'description', 'updatedAt’, 'magnitudeUnit'*.
2. **Standardization of Columns**: I renamed the columns for better understanding of its meaning and the data that was being shown. I got this by doing the following: *df.columns=["incident", "alert\_level", "alert\_score", "date\_incident", "magnitude\_value", "type", "country", "severity\_score"].*
3. **Data Cleaning**:
   1. **Incident**: Initially called ‘title’, I started by iterating over this column to get the name of the disaster that occurred in each row. To be able to do it I transformed the column title into a list.



I used a for loop to iterate to each row first, then impose conditions inside the loop. If the word I was looking for was present in that row, it would be added to the empty list I created before for this purpose called incidents. So, after this I got a list with only the incidents presents in the data Frame. Lastly, I just assigned the new cleaned list to the title columns substituting the values for a better understanding of the data.

* 1. **Alert\_level**: To this column I started by checking the NaN’s, which was shown to not exist, so I proceed to the next column.
  2. **Alert\_score & Severity\_score**: Alert score was also standardized and well cleaned; however, I decide to create a new column based on this one. I wanted to make it as simple to read as possible as 0,1 and 2 didn’t specify what was the meaning behind it. So in the new column called “severity\_score” I associate the values with rankings (0- minor, 1- moderate, 2-severe). To get these results I created a dictionary called mapping\_severity= {0: “minor”, 1: “moderate”, 2: “severe”}, then replace the integers with these corresponding values by doing df[“severity\_score”]= df[“severity\_score”].replace(mapping\_severity).
  3. **Date\_incident**: Initially called “createdAt”, it had date and specific hour mixed for each incident. Since I only needed the specific date to be able to identify which period of time I was analysing I decided to use Regex. Started by importing the library with import re and defining the pattern I wanted to search. Similar to the incident column I used a for loop to iterate over each row of the column “createdAt”. So, resuming it, in each row I searched for the date pattern “yyyy/mm/dd”. If that pattern was found in the row, it would be added to the empty list I created prior called dates.



Having my list with the dates, I replaced the cleaned values over the “createdAt” now called date\_incident.

* 1. **Magnitude\_value & Type**: Similar to alert\_level, I started by checking the NaN’s using the method isna().sum(), which was shown to not exist, so I proceed to the next column
  2. **Country**: Initially calls “landCountry”, I checked the NaN’s which were significantly high considering that in the biggening I started with 250 rows. However as stated before, I decided to get more data in order to deal with the NaN’s. After requesting the 60 pages I was able to drop the NaN values without missing big chunks of significant data since the countries column was my priority and was not possible to fill it with the fillna() method.

## **EDA- Visualization**

After some research on possible libraries and graphics I could use to present my data. I came across plotly.express and geopandas which allows easy manipulation and analysis of geographic data in Python.

I follow the steps to install these libraries, starting by using “!pip install geopandas” since I did not had this library installed previously and then I imported plotly.express.

To start I wanted to know what the predominant Natural Disaster in the past month was Worldwide so I could focus on that point. I decided to select the incidents column and count the values to see the frequency of each incident. Turned out to be Forest Fires, and I created a simple barplot to show it. From there I analyse the top 10 countries affected by forest fires.

To do so, I selected countries and incident columns, counted the values to see the frequency of the incidents by country and reset the index in order to bring it back as a DataFrame. Lastly I requested only the top 10 by using “.head(10)”.



The second visualization I wanted to create was concerning my second hypothesis, therefore I used a tool for data aggregation called groupby(), selected countries and alert\_score. Since I am dealing with integers, I calculated the mean of the alert\_scores and checked the countries with higher values consequently the ones with the most severe disasters.

Finally, I reset the index to bring it back to a DataFrame shape and sort the values by alert score on a descending order, using “.head(10)” to display only the top 10 countries.



## **Conclusion**

This project highlighted the complex nature of natural disasters, showing how critical it is to consider both the frequency of these events and their severity. By analysing the data, I was able to identify important trends that support the hypotheses and offer valuable insights for improving risk control strategies. These findings can help guide decisions aimed at reducing vulnerability and enhancing community resilience in high-risk regions.

The first hypothesis was confirmed by the data. As we analysed regions in the Southern Hemisphere, such as Australia, Brazil, and Mozambique, the data highlighted an increased prevalence of forest fires, particularly as these countries enter their hottest seasons. This suggests that global climate change, through rising temperatures and prolonged dry seasons, is contributing to an increase fire condition.

The second hypothesis was also confirmed. Countries like Portugal, Zimbabwe, and Paraguay demonstrated high disaster severity, despite not appearing at the top of the list for disaster frequency. This underscores the critical need for personalized risk management approaches that address both frequent low-impact disasters and occasional high-severity events.

To conclude, the analysis shows the importance of investing in infrastructure resilience, early warning systems, and community preparedness, particularly in countries that may be less frequently impacted but suffer greater consequences when disasters do strike.