1. **Project Goals and Requirements**:
   * We are trying to develop a predictive model for natural disasters, which will, based on the specific natural disaster, and geographic location, be able to give us a close prediction of the number of deaths.
   * Although our project is mainly focused on the predictions, we will have also a descriptive model, which we will talk about later.
   * Available natural disasters: drought, earthquake, epidemic, extreme, temperature/weather, Flood, Landslide, Mass Movement, Volcanic Activity, Wildfire

the geographic region of interest: most countries

the timeframe for predictions: 2000🡪2023

1. **Data Collection and ETL Process**:

* Data Collection:

1. Identify Data Sources: Determine the sources of data that are relevant to predicting natural disasters. This led us to the following dataset that we used

* ETL Process

1. Extract data from: International Disaster Database (EM-DAT):

EM-DAT is a comprehensive global database on natural and technological disasters. It offers historical data on a wide range of disasters, including information on casualties, economic losses, and other impacts.

1. Data Cleaning: we erase the rows that have empty values for necessary fields, and remove irrelevant features.
2. Data Documentation:we will maintain documentation that describes the data sources, transformations, and any changes made to the dataset.
3. Automation: Consider automating the ETL process to enable regular updates of your dataset with new data as it becomes available.

------------------------------------------------------------------------------------------------------------------------------------------

1. **Machine Learning Model Development**:
   * We will use mainly neural networks for our predictions, with Keras TensorFlow.
   * We’re working with 2500+ rows of data, which will help us with the prediction of the death in each disaster.

The main factors to test the predictions will be:

1. Magnitude
2. Total Damage
3. Total Affected

this is why, in our case, we’ll be splitting our dataset into multiple sections, which is based each one on the respected disaster type, so that it increases the relevance of the magnitude.

1. **Descriptive Analysis and PowerBI**

In this part, we will revise our dataset, and give conclusions of the distribution of natural disasters around the world, based on our already given data, with no predictions, or any prescription.

Some analysis that might seem useful:

* Continent with the most deathes by natural causes
* Countries with the most OFDA responses
* Continent/country count of each natural disaster type
* Visual Representation of the deaths…. Using map