

Name: **Muley, Tushar**

Assignment: **DSC 680 - Week 1 Project Proposal Milestone 1**

Date: **December 5, 2021**

## **Project Proposal Milestone – 1**

### **What is the Impact of Natural Disasters?**

#### **Topic:**

In looking for data that would help me expand my data science experience. The topic selected was natural disasters around the world.

#### **Question to research:**

In using the data from Our World in Data site. I want to research the question of what affects different natural disasters have on different variables. These variables include death, injuries, homelessness and cost to the economy. If time permits I want to explore modeling some of this data on a specific natural disaster. Using the data to predict if this type of disaster is increasing. Or model the impact of a specific type of disaster is increasing deaths or other variables. An example, are hurricanes increasing deaths or economic impacts?

#### **Dataset:**

Obtaining the data from Our World in Data[1] site and the subsite called Natural Disasters[2]. It appears the data is pretty robust and descriptive. I am not solely relying on Our World in Data site as my only source. Once I delve into the data I want to make sure I have raw numbers. That might require going to other sites like World Health Organization[3], World Bank[4] or Kaggle[5]. The main data contains the field listed in the appendix section of this paper.

#### **Methods:**

The methods I plan to use. I will determine if I can use the Our World in Data as is. The Our World in Data has a lot of different breakouts in the data. Below is my rough outline of the plan.

1. Profile the data
  - a. Assess missing values
  - b. Assess values types
  - c. Perform graphical analysis (histogram, scatter plots and correlation) to better understand what kind of data is available and how best to use this data.
2. Clean up the data
  - a. Fill in any missing values
  - b. Delete N/A rows

- c. Update numeric strings to numbers
  - d. Remove columns that are not needed
  - e. Pivot any data that needs to be accomplished
3. Connect additional data if needed
  - a. Missing values or connect other values that might help the analysis and modeling
4. Based on the data select a model
5. Run model
  - a. Review data
  - b. Determine changes
6. Pull all my analysis together with a visuals and PowerPoint presentation

### **Ethical Considerations:**

This data does not appear to have a lot of ethical challenges. I have scanned the data and there is no personally identifiable information. Everything is based at country level and aggregated numbers. The data appears to have been collected from government agencies like National Oceanic and Atmospheric Administration (NOAA), Global Health Data Exchange (GHDx) and World Bank.

On a side thought my analysis could be used to increasing insurance premiums. This might affect individual living closer to the proverbial line or financially challenged. If my model works to show an increase in the rate of natural disasters. Companies and municipalities could use it to drive people out of certain areas of the city or country. This might cause migration issues not just across borders but human migration leading to more condensed living areas. That in turn might cause other human made disasters.

### **Challenges/Issues:**

Challenges with the data are going to be limited to what information I might need to run a model on the data. I tend to have a habit of going down rabbit holes with the data. Especially data I connect with or start analyzing and work turns into curiosity. If I end up down the rabbit hole I will need numbers for deaths (compared to death rates as an example), numbers for actual disasters (how many hurricanes, tornados, or other disasters). I am going to try and contain the extra work due to limited time.

Some issues that might be encountered with the data is the possible lack of detail. I noticed in my early analysis that the data contains rates for global data. The initial data has been combined with information from other agencies data to derive it. I might not have access to or require setting up an API, which might need approvals similar to Twitter. That took almost a week to complete. The other big concern is my model does not work.

**Reference:**

1. "Our World in Data" (November 2014).  
<https://ourworldindata.org/>
2. Ritchie H. & Roser M. (November 2014). "Natural Disasters". From OurWorldInData.org.  
Retrieved from: <https://ourworldindata.org/natural-disasters>
3. Global Health Data Exchange (November 2021). From University of Washington.  
<http://ghdx.healthdata.org/gbd-results-tool>
4. Pielke R. (October 2018). "Tracking progress on the economic costs of disasters under the indicators of the sustainable development goals" From Taylor & Francis Online.  
<https://www.tandfonline.com/doi/abs/10.1080/17477891.2018.1540343?journalCode=tenh20>
5. Ache', M. (December 2021). "Natural Disaster Data Explorer". From Kaggle.com  
<https://www.kaggle.com/mathurinache/natural-disasters-data-explorer>

**Appendix:**

Columns Name/Data Description	Columns Name/Data Description
Number of deaths from drought	Insured damages against storms
Number of people injured from drought	Total economic damages from storms
Number of people affected from drought	Death rates from storms
Number of people left homeless from drought	Injury rates from storms
Number of total people affected by drought	Number of people affected by storms per 100,000
Reconstruction costs from drought	Homelessness rate from storms
Insured damages against drought	Total number of people affected by storms per 100,000
Total economic damages from drought	Number of deaths from landslides
Death rates from drought	Number of people injured from landslides
Injury rates from drought	Number of people affected by landslides
Number of people affected by drought per 100,000	Number of people left homeless from landslides
Homelessness rate from drought	Number of total people affected by landslides
Total number of people affected by drought per 100,000	Reconstruction costs from landslides
Number of deaths from earthquakes	Insured damages against landslides
Number of people injured from earthquakes	Total economic damages from landslides
Number of people affected by earthquakes	Death rates from landslides

Number of people left homeless from earthquakes	Injury rates from landslides
Number of total people affected by earthquakes	Number of people affected by landslides per 100,000
Reconstruction costs from earthquakes	Homelessness rate from landslides
Insured damages against earthquakes	Total number of people affected by landslides per 100,000
Total economic damages from earthquakes	Number of deaths from fog
Death rates from earthquakes	Number of people injured from fog
Injury rates from earthquakes	Number of people affected by fog
Number of people affected by earthquakes per 100,000	Number of people left homeless from fog
Homelessness rate from earthquakes	Number of total people affected by fog
Total number of people affected by earthquakes per 100,000	Reconstruction costs from fog
Number of deaths from disasters	Insured damages against fog
Number of people injured from disasters	Total economic damages from fog
Number of people affected by disasters	Death rates from fog
Number of people left homeless from disasters	Injury rates from fog
Number of total people affected by disasters	Number of people affected by fog per 100,000
Reconstruction costs from disasters	Homelessness rate from fog
Insured damages against disasters	Total number of people affected by fog per 100,000
Total economic damages from disasters	Number of deaths from wildfires
Death rates from disasters	Number of people injured from wildfires
Injury rates from disasters	Number of people affected by wildfires
Number of people affected by disasters per 100,000	Number of people left homeless from wildfires
Homelessness rate from disasters	Number of total people affected by wildfires
Total number of people affected by disasters per 100,000	Reconstruction costs from wildfires
Number of deaths from volcanic activity	Insured damages against wildfires
Number of people injured from volcanic activity	Total economic damages from wildfires
Number of people affected by volcanic activity	Death rates from wildfires
Number of people left homeless from volcanic activity	Injury rates from wildfires
Number of total people affected by volcanic activity	Number of people affected by wildfires per 100,000
Reconstruction costs from volcanic activity	Homelessness rate from wildfires

Insured damages against volcanic activity	Total number of people affected by wildfires per 100,000
Total economic damages from volcanic activity	Number of deaths from extreme temperatures
Death rates from volcanic activity	Number of people injured from extreme temperatures
Injury rates from volcanic activity	Number of people affected by extreme temperatures
Number of people affected by volcanic activity per 100,000	Number of people left homeless from extreme temperatures
Homelessness rate from volcanic activity	Number of total people affected by extreme temperatures
Total number of people affected by volcanic activity per 100,000	Reconstruction costs from extreme temperatures
Number of deaths from floods	Insured damages against extreme temperatures
Number of people injured from floods	Total economic damages from extreme temperatures
Number of people affected by floods	Death rates from extreme temperatures
Number of people left homeless from floods	Injury rates from extreme temperatures
Number of total people affected by floods	Number of people affected by extreme temperatures per 100,000
Reconstruction costs from floods	Homelessness rate from extreme temperatures
Insured damages against floods	Total number of people affected by extreme temperatures per 100,000
Total economic damages from floods	Number of deaths from glacial lake outbursts
Death rates from floods	Number of people injured from glacial lake outbursts
Injury rates from floods	Number of people affected by glacial lake outbursts
Number of people affected by floods per 100,000	Number of people left homeless from glacial lake outbursts
Homelessness rate from floods	Number of total people affected by glacial lake outbursts
Total number of people affected by floods per 100,000	Reconstruction costs from glacial lake outbursts
Number of deaths from mass movements	Insured damages against glacial lake outbursts
Number of people injured from mass movements	Total economic damages from glacial lake outbursts
Number of people affected by mass movements	Death rates from glacial lake outbursts

Number of people left homeless from mass movements	Injury rates from glacial lake outbursts
Number of total people affected by mass movements	Number of people affected by glacial lake outbursts per 100,000
Reconstruction costs from mass movements	Homelessness rate from glacial lake outbursts
Insured damages against mass movements	Total number of people affected by glacial lake outbursts per 100,000
Total economic damages from mass movements	Total economic damages from disasters as a share of GDP
Death rates from mass movements	Total economic damages from drought as a share of GDP
Injury rates from mass movements	Total economic damages from earthquakes as a share of GDP
Number of people affected by mass movements per 100,000	Total economic damages from extreme temperatures as a share of GDP
Homelessness rate from mass movements	Total economic damages from floods as a share of GDP
Total number of people affected by mass movements per 100,000	Total economic damages from landslides as a share of GDP
Number of deaths from storms	Total economic damages from mass movements as a share of GDP
Number of people injured from storms	Total economic damages from storms as a share of GDP
Number of people affected by storms	Total economic damages from volcanic activity as a share of GDP
Number of people left homeless from storms	Total economic damages from volcanic activity as a share of GDP
Number of total people affected by storms	Entity
Reconstruction costs from storms	Year