DQLab

Disasters in Indonesia

Analysis of the regional (province) distribution of natural and technological disasters, 2000-2023

Capstone Project TETRIS Batch 4

raihan.qodri1217@gmail.com

https://www.linkedin.com/in/raihan-qodri

https://github.com/raihanqodri

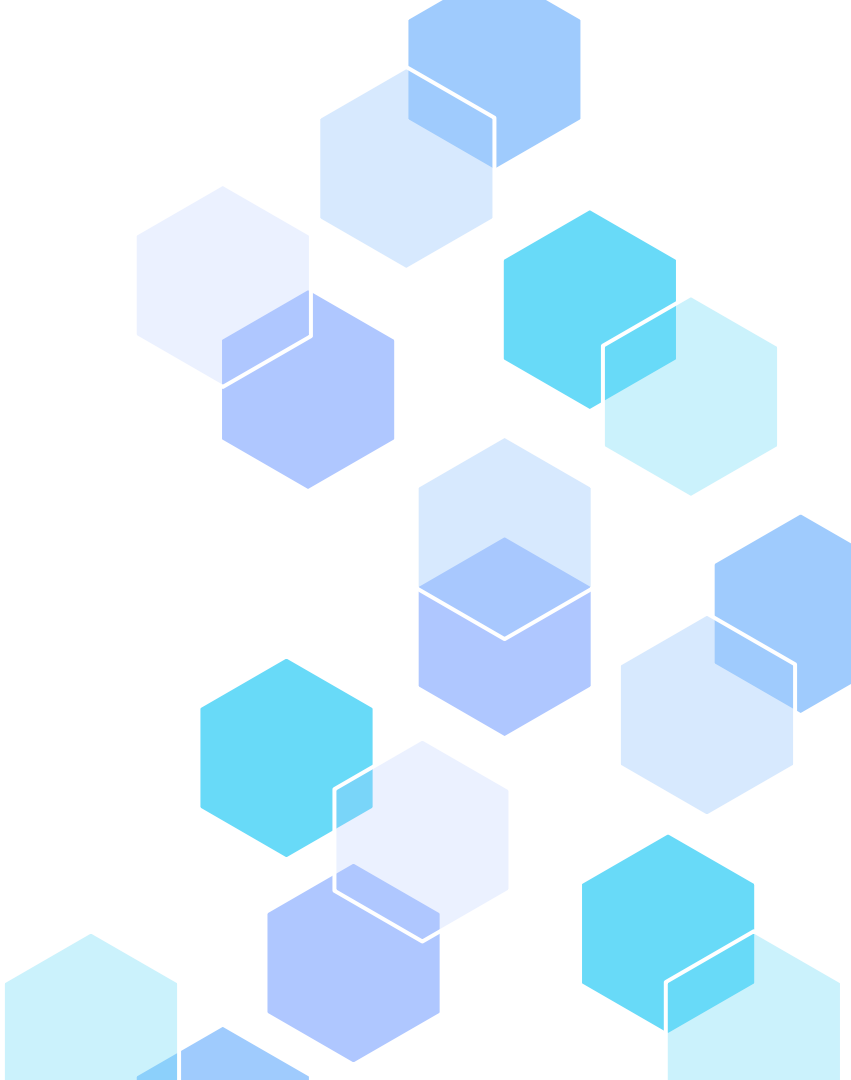
Outline Scope of Work

There are several steps in analysis disaster data in Indonesia

Data Collection & Data Integration	Search for reliable sources of disaster data, namely from https://public.emdat.be/data
Data Understanding	Examine the form of data, variables, and purpose of data
Data Cleansing	Cleaning data, filtering unused data, and transforming data with spreadsheet and SQL
Data Exploration	Exploring data to understand the data thoroughly with SQL and Python
Data Visualization	Visualization clean data with Tablue
Insight Analysis	Draw insight from visualised data

Introduction

Know the background and problem formulation



Introduction

Indonesia is located in a geographical region with seismic and volcanic activity, and has a diverse topography, ranging from high mountains to lowlands. In addition, Indonesia consists of a huge archipelago with a large river network and tropical weather patterns, so it is not uncommon for human activities to result in uncontrolled deforestation and extreme climate change.

Starting from the conditions that exist and are created**, Indonesia becomes one of the disaster-prone countries, such as volcanic activity, seismic, tsunamis, landslides, flash floods and floods caused by human hands, and many other damages**. In addition to natural disasters, **Indonesia also faces technological disasters** which include **industrial accidents, factory fires, chemical leaks**. These technological disasters can have serious impacts on the environment, public health and the economy.

Stakeholders:
Disaster management agencies that are serious in monitoring disaster activities in Indonesia, as well as work safety and transport safety agencies

Assumption:

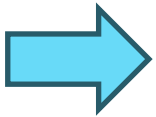
- There are differences in the impact and casualties caused by natural and technological disasters.
- The highest damage and casualties are caused by natural disasters

With this assumption, stakeholders **can understand what causes the number of victims and losses due to natural and technological disasters**, and **can evaluate preventive actions taken to minimize the impact of losses and casualties**.

Problem Statment

Key Question

1. What is the impact of natural and technological disasters from year to year ?
2. Which disaster has the most impact on casualties ?



Goals

The results of this analysis can **assist in evaluating preventive** measures to minimise casualties and resulting losses.

Data analysis activities

Data Collection

Get data from EM-DAT source:
The number of disasters that occurred in the Indonesian region from 2000-2023

Data Preprocessing

Processing data focuses on data filters and data transformation

1. Determining the variables used for analysis
2. Imputing null data
3. Transforming region data into each province

Exploratory

From the first question: The impact of natural and technological disasters (casualties, losses) will be visualized based on the years 2000-2023. Based on the data, **natural disasters have a higher impact on losses and casualties** than technological disasters.

Visualization

From the second question: Will be filtered based on the highest total casualties and losses to see which disasters cause the most impact. Based on the data, **there are significant impacts caused by the disaster**

Data collection

Survey

Conduct **survey** by category. Then selected the category on humanity

Explore

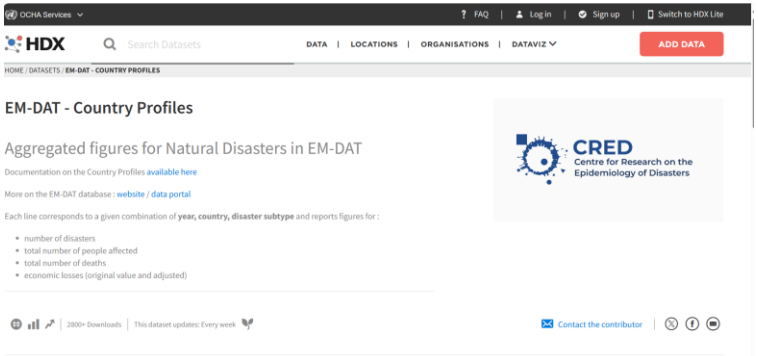
Explore humanitarian data on HDX (Humanitarian Data Exchange)

Find

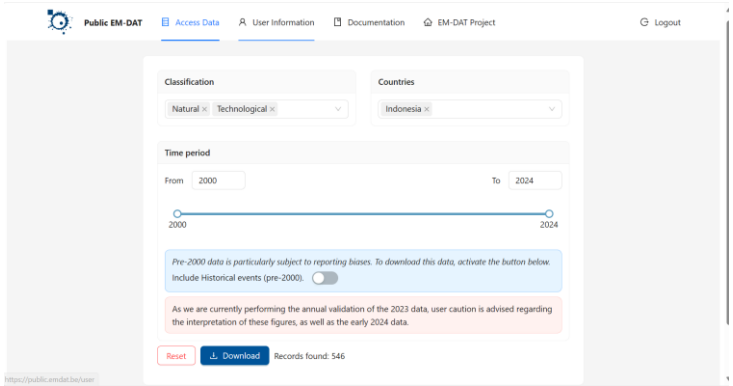
Find disaster data and trace the source of the data

Access

Accessing data from the EM-DAT source, then selecting Indonesia with the classification of natural and technological disasters.



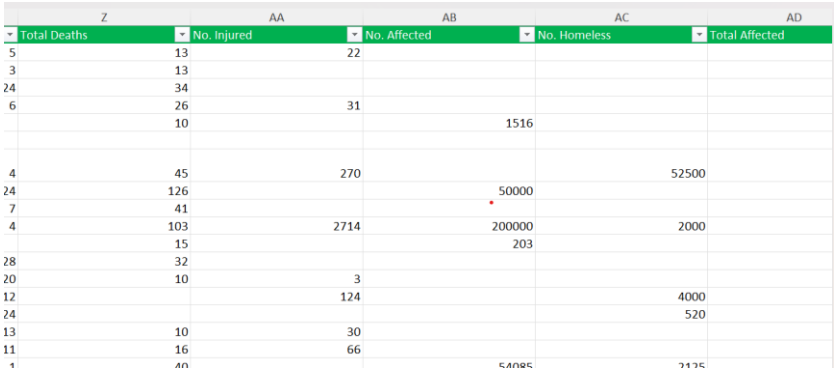
Source: EM-DAT - Country Profiles - Humanitarian Data Exchange (humdata.org)



Source: Public EM-DAT platform (emdat.be)

Data Preprocessing

Part 1



Z	AA	AB	AC	AD
Total Deaths	No. Injured	No. Affected	No. Homeless	Total Affected
5	13	22		
3	13			
24	34			
6	26	31		
	10	1516		
4	45	270	52500	
24	126	50000		
7	41			
4	103	2714	200000	2000
	15	203		
28	32			
20	10	3		
12		124	4000	
24			520	
13	10	30		
11	16	66		
1	40	54085	2125	

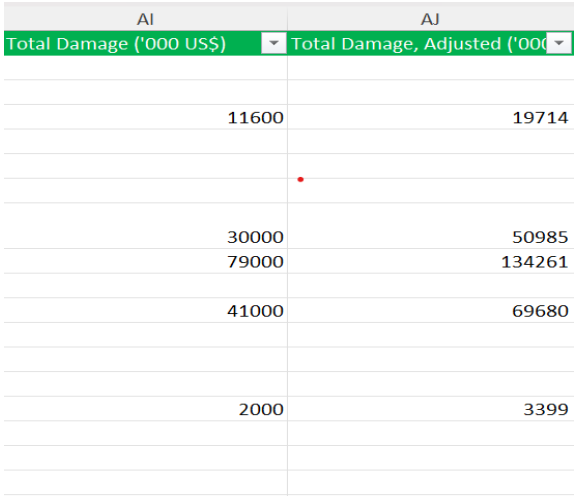
There are **missing values** in total death and total affected

Possible Cause:

1. No access to the number of victims affected
2. Indeed, no victims died or were affected

Addressing the situation:

Fill in the blank data with the KNN method



AI	AJ
Total Damage ('000 US$)	Total Damage, Adjusted ('000
11600	19714
30000	50985
79000	134261
41000	69680
2000	3399

There are **missing values** in **Total Damage and Total Damage, Adjusted**

Possible Cause:

1. Disasters with technological classification do not include total damage and total damage, adjusted
2. Not getting access to information or data on total damage and total damage, adjusted from Indonesia

Addressing the situation:

Since the data is about the range of disaster impact losses, and every year there are economic fluctuations, we filled the data with a value of 0

Data Preprocessing

Part 2

Location
Woosobo (Java)
Java
Brebes district (Jawa Tengah province)
Java Isl.
Jakarta
Riau province (Sumatra Isl.), Kalimatan Barat, Kalimatan Tengah provinces (Borneo Isl.)
Totikum, Tinangkung, Liang villages (Banggai Kepulauan district, Sulawesi Tengah province), Banggai district (Sulawesi Tengah province)
Malaka Tengah, Malaka Barat areas (Belu district, Nusatenggara Timur province)
Ambon Island
Enggano island (Bengkulu Utara district, Bengkulu province)
Ngada district (Flores Isl.)
Padang Panjang, Sumatera Barat
Dukuh Karak (Cilacap district, Java Isl.)
Ciranggon village (Karawang district, Jawa Barat province)
Banggai city (Banggai Kepulauan district, Sulawesi Tengah province)
Jakarta
Near Katanggan village (Java Isl.)
Cilacap, Banyumas districts (Jawa Tengah province)

The data format in the location variable column is **still too random**

Addressing the situation:

In order to simplify the analysis of the distribution of disaster areas, we standardize the location variables by province

	B	C	D	E	F	G	H	I	J	K	L
1	Year	Month	Disaster Gro	Disaster Subgroup	Disaster Type	Disaster Subtype	Total Deat	Total Affect	Total Damage ('000 US	Total Damage, Adjusted ('000 US	Location
2	2000	Januari	Technological	Transport	Road	Road	13	22	0	0	Jawa Tengah
3	2000	Januari	Technological	Transport	Road	Road	13	333	0	0	Jawa Tengah
4	2000	Februari	Natural	Hydrological	Mass movement (wet)	Mudslide	34	84377	11600	19714	Jawa Tengah
5	2000	Maret	Technological	Transport	Road	Road	26	31	0	0	Jawa Tengah
6	2000	Januari	Natural	Biological	Epidemic	Viral disease	10	1516	0	0	Jakarta
7	2000	Februari	Natural	Climatological	Wildfire	Forest fire	585	511572	0	0	Riau
8	2000	Februari	Natural	Climatological	Wildfire	Forest fire	585	511572	0	0	Kalimantan Barat
9	2000	Februari	Natural	Climatological	Wildfire	Forest fire	585	511572	0	0	Kalimantan Tengah
10	2000	Mei	Natural	Geophysical	Earthquake	Ground movement	45	52770	30000	50985	Sulawesi Tengah
11	2000	Mei	Natural	Hydrological	Flood	Flash flood	126	50000	79000	134261	Nusa Tenggara Timur
12	2000	Mei	Technological	Transport	Water	Water	41	12005	0	0	Maluku
13	2000	Juni	Natural	Geophysical	Earthquake	Ground movement	103	204714	41000	69680	Bengkulu
14	2000	Mei	Natural	Biological	Epidemic	Viral disease	15	203	0	0	Nusa Tenggara Timur
15	2000	Juni	Technological	Transport	Rail	Rail	32	24750	0	0	Sumatera Barat
16	2000	Juli	Technological	Transport	Road	Road	10	3	0	0	Jawa Tengah
17	2000	Juli	Natural	Geophysical	Earthquake	Ground movement	25	4124	2000	3399	Jawa Barat

The following data is already in a clean state

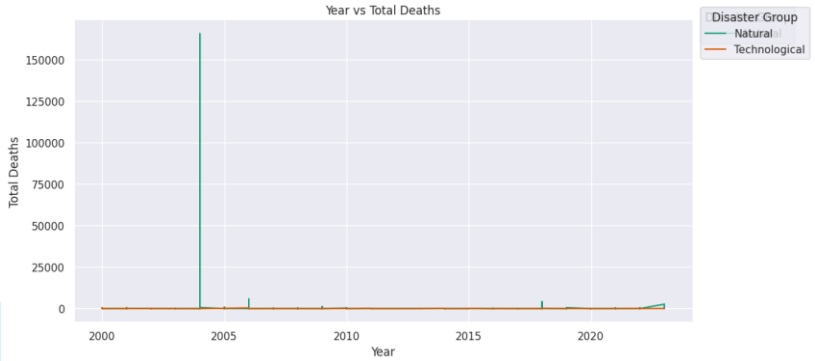
Then the data is ready to be explored in SQL and Python

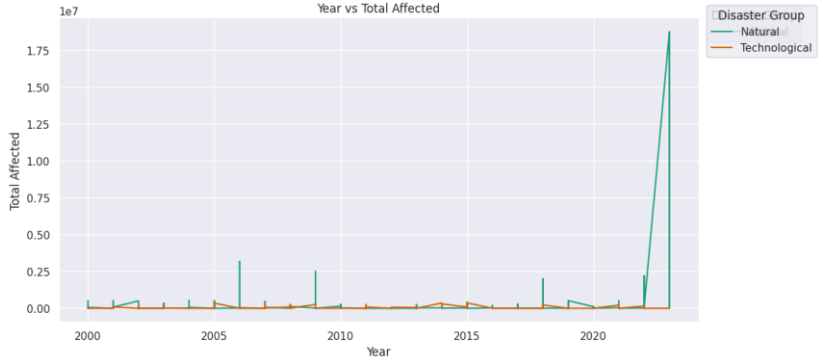
EDA 1 | Answer the Question

What is the impact of natural and technological disasters from year to year ?

Disaster Group	jumlah_bencana
Technological	170
Natural	672

jumlah_tewas	jumlah_terdampak
442,002	387,498,281

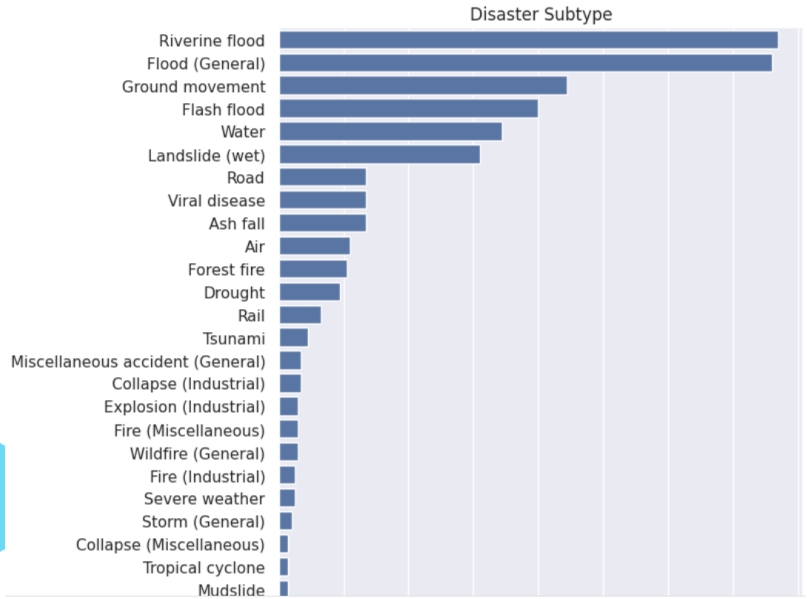




1. The data show that the consequences of natural and technological disasters in the period 2000-2023 have **caused many deaths and injuries.**
2. The number of victims of **natural disasters appears to be higher** each year than the number of victims **of technological disasters**, because the frequency of **natural disasters is higher** and **more difficult to predict**.

EDA 2 | Answer the Question

Which disaster has the most impact on casualties ?



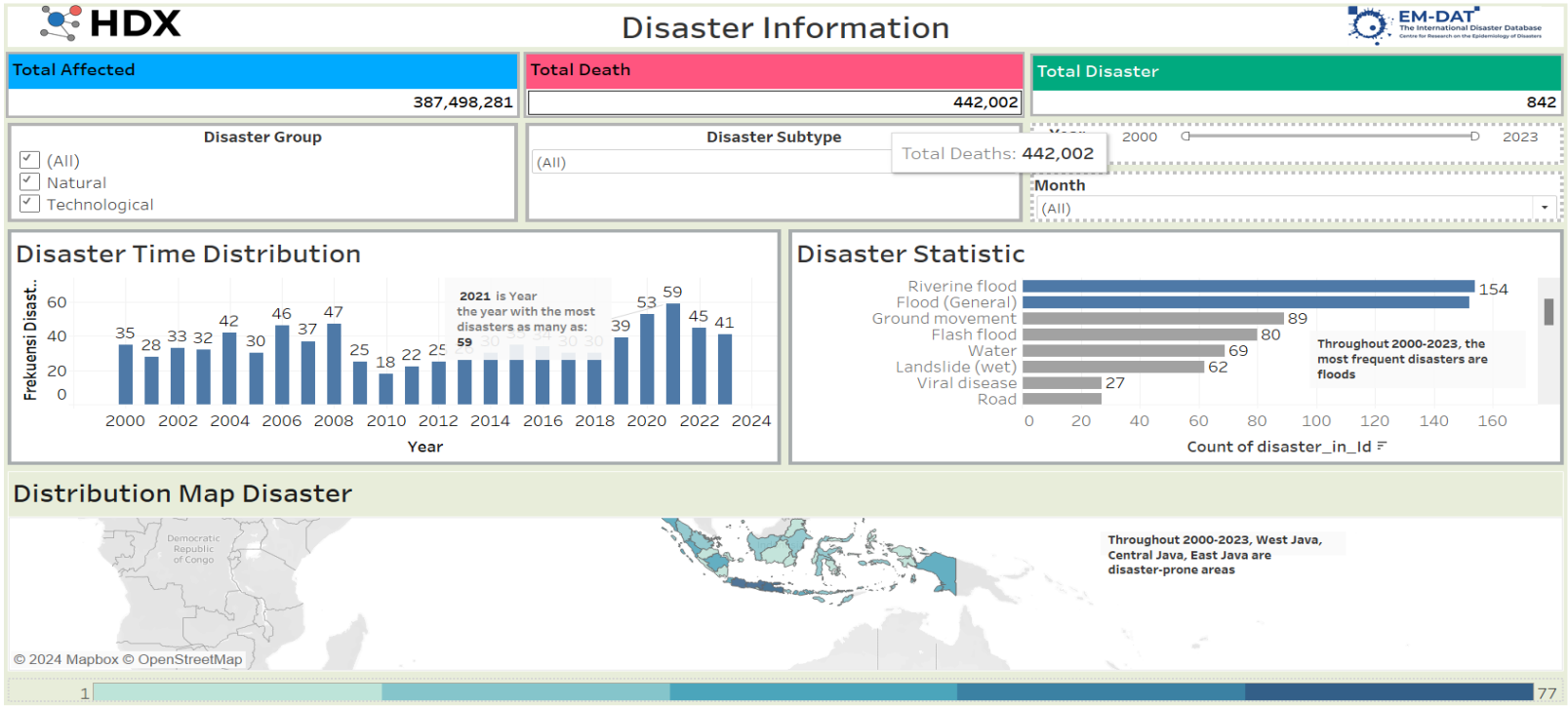
	Disaster Subtype	jumlah_tewas	jumlah_terdampak	total_kerusakan_usd
1	Tsunami	337,702	3,159,125	15,644,179
2	Drought	45,329	318,799,817	1,591
3	Ground movement	19,142	19,190,320	20,721,350
4	Viral disease	7,187	623,499	0
5	Riverine flood	5,024	13,168,252	15,336,709
6	Forest fire	4,369	6,040,588	7,510,072
7	Flash flood	4,261	6,107,567	7,461,960
8	Flood (General)	4,229	7,050,814	589,751

1. From the data exploration, it can be seen that the **most frequent disaster is flooding** throughout 2000-2023.
2. However, the disaster that caused the most impact during 2000-2023 was **the tsunami** with the **highest total deaths** and **material losses.**

Data Visualization| Deepening the Analysis

Disasters in Indonesia

Disaster Information | Quick Chats Diasaster

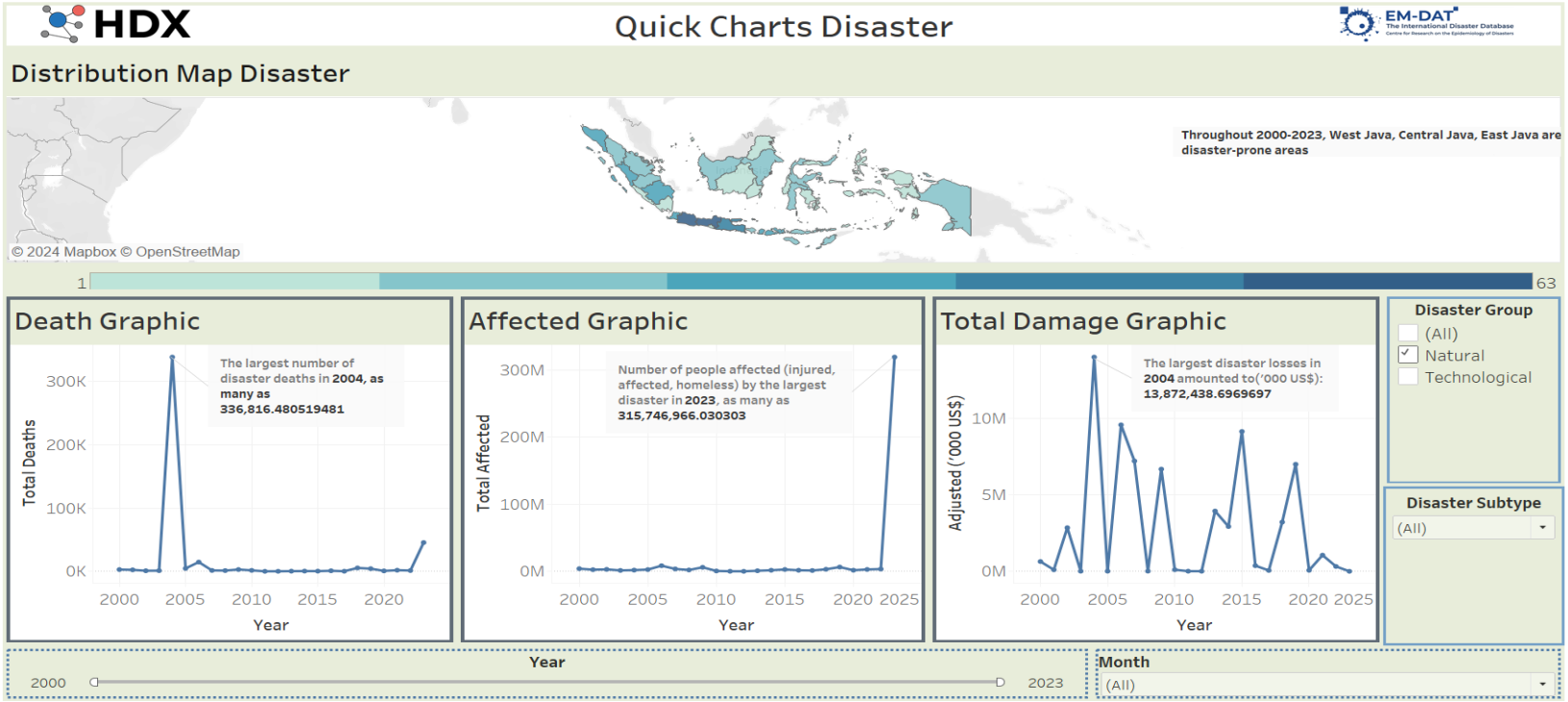


Summary: Of all the disasters that occurred throughout 2000-2023, **the greatest frequency of disasters in 2021**, then the region with the greatest frequency of disasters is **Java Island**

Data Visualization| Deepening the Analysis

Disasters in Indonesia

Disaster Information | Quick Chats Diasaster



Summary: The largest **total deaths** and **losses occurred** in **2004 in December due to the tsunami in Aceh Province**, then **the largest total impact in 2023** due to **strong winds** that hit **almost all parts of Indonesia.**

Report Feedback

Report Feedback

Summary Insight	1. The impact of natural and technological disasters in the period 2000-2023 has caused many deaths and injuries, but the number of victims of **natural disasters is higher every year**, because **the frequency of natural disasters is higher and more difficult to predict**. 2. Then the **most frequent disaster was flooding**, but the disaster that caused the **most impact was the tsunami with the highest total casualties and material losses**. 3. Of all the disasters that occurred throughout 2000-2023, **the largest disaster frequency was in 2021**, then the region with the **largest disaster frequency was Java Island**. In addition, **the largest total casualties and losses occurred in 2004 in December** due to **the tsunami in Aceh Province**, then **the largest total impact in 2023** due to **strong winds** that hit **almost all parts of Indonesia**.

Feedback	**Reconfirm** the data with **the relevant official institutions**, so that the accuracy of the data is better maintained. Then it is important to **update the data**, so that **the implementation of policies** is more measurable based on the latest phenomena.

Suggestions for related institutions

In accordance with the insights gained, improving natural disaster mitigation as a preventive measure such as providing safety insights in the event of a disaster

In addition, Equalization of disaster preparedness across Indonesia, such as monitoring with support tools, management of areas away from disaster sites, and disaster-resilient infrastructure.

Stricter implementation of safety training and education, regular inspections, and monitoring and evaluation by relevant agencies to minimise the number of industrial accidents.

Strengthening regulations and law enforcement related to transport safety standards, then improving infrastructure and routine maintenance, as well as campaigning for public awareness of transport safety.

27 Februari 2024

Thanks!

Capstone Project TETRIS Batch 4

raihan.qodri1217@gmail.com

https://www.linkedin.com/in/raihan-qodri

https://github.com/raihanqodri

Raihan Qodri