

SOURCING OPEN DATA

1. PROJECT CONTEXT AND OBJECTIVE

Climate change has been a topic of discussion for the last few decades, and its impact is becoming increasingly evident due to the overproduction of greenhouse gases and the overexploitation of natural resources. One effect attributed to climate change, often overlooked, is the migration of people caused by natural disasters worldwide.

This project aims to explore the correlations between key climate change indicators, such as temperature, CO2 emissions, and energy consumption, with global migration patterns. The goal is to gain valuable insights into the complex interconnections between climate change, human migration, and energy practices.

2. THE DATA SETS

Limitations of the data:

- Data availability: The availability of data varies across countries and years, which may limit the scope and comparability of the analysis.
- Data quality: The quality and accuracy of the data may vary depending on the data collection methods and reporting practices of different countries.
- In the migration data set, the reasons for migration patterns are not provided. The data set lacks sufficient information about why people decided to move. Various factors, such as political, economic, and climate-related reasons, could contribute to these migration patterns.

The following data sets have been sourced for the project:

Data set #1. ALL NATURAL DISASTERS 1900-2021 / EOSDIS

Link: <https://www.kaggle.com/datasets/brsdincer/all-natural-disasters-19002021-eosdis>

Description of the Data Source: The data source provides information on natural disasters that occurred between 1900 and 2021.

Reason for Choosing this Data Set: The data set was chosen because it provides a comprehensive overview of natural disasters over a long time period, which can be valuable for understanding the frequency, intensity, and impact of these events.

Summary of the Data Source:

- Source: The data is sourced from the Earth Observing System Data and Information System (EOSDIS)
- Collection Method: The data is collected through various sources, including satellite imagery, ground-based observations, and reports from national and international agencies.
- Contents: Some of the key columns in the data set are the year of the disaster, disaster Type (e.g., earthquake, flood, storm, wildfire), location, country, and duration of the event.

Column Name	Description
Year	The year in which the natural disaster occurred
Month	The month in which the natural disaster occurred
Day	The day on which the natural disaster occurred
Country	The country in which the natural disaster occurred
Disaster Type	The type of natural disaster that occurred
Disaster Subtype	The subtype of natural disaster that occurred
Disaster Subsubtype	The subsubtype of natural disaster that occurred
Associated Disasters	The number of associated disasters
Total Deaths	The total number of deaths caused by the natural disaster
Total Affected	The total number of people affected by the natural disaster
Insured Damages ('000 US\$)	The total insured damages caused by the natural disaster in thousands of US dollars
Total Damages ('000 US\$)	The total damages caused by the natural disaster in thousands of US dollars

Data set #2. Temperature change on land

Links: <https://www.fao.org/faostat/en/#data/ET>
<https://www.kaggle.com/datasets/sevgisarac/temperature-change>

Description of the Data Source: Annual statistics on mean surface temperature change for various countries

Reason for Choosing this Data Set: The temperature change data set was chosen for its relevance to the analysis of the effects of climate change on people's migration worldwide. Temperature change is a key indicator of climate change, and understanding its patterns and trends can provide insights into the potential impacts on migration.

Data ownership: The FAOSTAT data is external, as it is collected and maintained by the Food and Agriculture Organization of the United Nations (FAO). The FAO owns the data, as it is the entity responsible for collecting and maintaining it.

Summary of the Data Source:

- Collection method: The data is collected through a combination of satellite measurements, weather station records, and other sources.
- Contents: The data set includes information on normalized temperature data for individual countries from 1961 to 2022.

Column Name	Description
Area Code	The numerical code
Area	Countries and Territories
Months Code	The numerical code for months
Months	Months, Seasons, Meteorological year
Element Code	The numerical code for the element column
Element	Type of measure: 'Temperature change', 'Standard Deviation'
Unit	Unit of measurement: Celsius degrees °C

Data set #3 Net migration by country.

Link: <https://www.kaggle.com/datasets/ulrikthgepedersen/migration>

Description of the Data Source: The data set provides information on global migration patterns.

Reason for Choosing this Data Set: This data set is relevant to the analysis of the effects of climate change on people's migration worldwide as it can help understand how climate-related factors, such as extreme weather events and environmental degradation, contribute to population movements.

Data ownership: The data set is owned by the United Nations and the World Bank, which are external entities that collect and provide global data on various topics, including migration.

Summary of the Data Source:

- Collection method: The data is collected through surveys, censuses, and administrative records by the respective national statistical offices and international organizations.
- Contents: The data set includes information on migration flows

Column Name	Description
country_code	The ISO country code.
country_name	The name of the country.
year	The year of the migration data
value	Net Migration per country per year

Data set #4. Migration & Population Density

Link: https://www.kaggle.com/datasets/eliasdabbas/migration-data-worldbank-1960-2018?select=migration_population

Description of the Data Source: The data set provides information on migration by country.

Reason for Choosing this Data Set: This data set is chosen because it's a reliable source of information on migration by country from 1960 to 2018.

Data ownership: The data set is owned by the United Nations and the World Bank

Summary of the Data Source:

- Collection method: The data is collected through surveys, censuses, and administrative records by the respective national statistical offices and international organizations.
- Contents:

Column Name	Description
Year	Year
population	population
pop_density	number of people living in the country per kilometer squared
net_migration	net migrants (can be negative), and available once per five years
migration_perc	net migration as a percentage of population
iso3c	3-letter code for countries

iso2c	2-letter code for countries
region	name of the region the country belongs to
adminregion	adminregion

Data set #5 CO2 Emissions

Link: <https://www.kaggle.com/datasets/ulrikthgepedersen/co2-emissions-by-country>

Description of the Data Source: CO2 emissions by country.

Reason for Choosing this Data Set: This data set is relevant to the analysis of the effects of climate change on people's migration worldwide as it can help understand the relationship between greenhouse gas emissions and climate change, which can in turn affect migration patterns.

Data ownership: The data set is owned by the various international organizations that collect and provide the data, including the World Bank, the International Energy Agency (IEA), and the United Nations Framework Convention on Climate Change (UNFCCC).

Summary of the Data Source:

- Collection method: The data is collected through various sources, including national statistical offices, international organizations, and research institutions.
- Contents: The data set includes information on CO2 emissions by country

Column Name	Description
Country Name	The name of the country
Country Code	The ISO country code
Year	The year of the CO2 emissions data
Value	The value of the CO2 emissions indicator for the specific country and year in kt (kilotons)

Data set #6 Energy Consumption

Link: <https://ourworldindata.org/energy#explore-data-on-energy>

Description of the Data Source: The chosen data source is from Our World in Data and provides information on energy consumption by country. **Primary energy consumption** measures the total energy demand of a country. It covers consumption of the energy sector itself, losses during transformation (for example, from oil or gas into electricity) and distribution of energy, and the final consumption by end users. It excludes energy carriers used for non-energy purposes (such as petroleum not used not for combustion but for producing plastics).

Reason for Choosing this Data Set: This data set is chosen because it provides a comprehensive and reliable source of information on energy consumption, a factor that is related to climate change.

Data ownership: The data source is owned by Our World in Data.

Summary of the Data Source:

- Collection method: The data is collected from various sources, including international organizations, research institutions, and government reports.
- Contents:

Column Name	Description
Entity	The name of the country
Code	The ISO country code
Year	The year of the energy consumption data
Primary energy consumption (TWh)	Energy consumption, measured in terawatt-hours per year.

3. QUESTIONS TO EXPLORE:

- How have natural disasters influenced migration patterns in affected countries? Are there regions where increased migration follows periods of frequent disasters?
- Which specific areas globally are experiencing the most significant impact from climate change, and does this align with their levels of greenhouse gas emissions and energy expenditure?
- Is there a correlation between changes in temperatures, the frequency of natural disasters, and higher CO2 emissions in certain regions?