# Climate Chaos: Understanding the Correlations between Atmospheric Pressure, Temperature and Precipitation with the Patterns of Extreme Weather

## Introduction

The idea behind this project is to find relationships (if they exist) between surface temperature

change and disasters. Using a combination of datasets we are going to try and understand

whether the change in surface temperature over the course of a few decades has had an impact

on the number of natural disasters and extreme weather events occurring around the world.

## Research Question

How do changes in atmospheric pressure, temperature, and precipitation correlate with the

occurrence of severe weather events (e.g., tornadoes, hail, lightning) and how has the frequency

and intensity of extreme weather events changed over the past century?

## Data Sources

**Data source 1:** Climate Related Disaster Frequency

• Metadata URL: IMF Forest Data

• Source of data: International Monetary Fund (IMF)

• What data it contains: Country wise climate related disasters over the years.

• Data Quality: The dataset is mostly complete, consistent, and accurate, sourced from

the IMF, covering nearly three decades.

**Data source 2:** Annual Surface Temperature Change Dataset

• Metadata URL: IMF Surface Temperature Data

• Source of data: International Monetary Fund (IMF)

• What data it contain: Country wise change in annual surface temperature over the

course of several decades.

• Data Quality: The dataset is mostly complete, consistent, and accurate, sourced from

the IMF, providing a long-term perspective on temperature changes.

## Licenses

The datasets are available for use under the open-data licenses that permit its use with proper citation. IMF Data Terms of Use.

## Data Pipeline

### Data Pipeline Description

• An automated data pipeline is used to download, clean and transform the data. The pipeline was created using Python with the help of libraries such as pandas and sqlite3 to manipulate and store the data. The data is stored in a SQLite database.

### Pipeline Steps

1. Extract/Download Data: The data is retrieved from the respetive URLs and stored locally.

2. Data Cleaning: Remove columns and handle missing values.

3. Data Transformation: Once cleaned, merge the datasets into one.

4. Data Storage: Store the cleaned and transformed data in an SQLite database.

### Data Pipeline Diagram

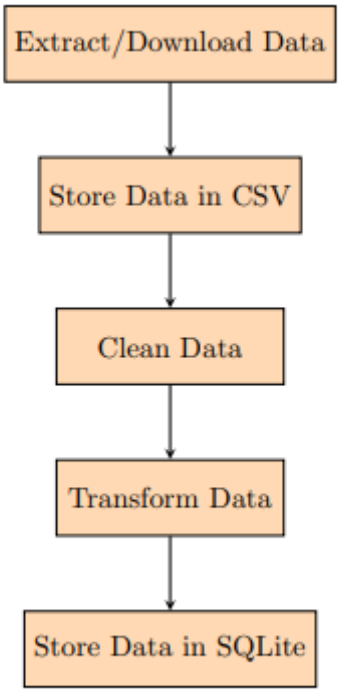


Figure 1: Data Pipeline Diagram

### Transformation and Cleaning Steps

Download and Read CSVs: The first step is to download the datasets from the respective URLs and store them into pandas dataFrames.

Filter Relevant Data:

• Columns Removed:

– For the Annual Surface Temperature Change dataset: ObjectId, ISO2, Unit, Source, CTS Code, CTS Name, and CTS Full Descriptor.



– For the Forest and Carbon dataset: ObjectId, ISO2, Unit, Source, CTS Code, CTS Name, and CTS Full Descriptor.

• Reason for Removal: The columns were removed to allow us to derive meaningful insights without causing too much clutter.

• Columns Preserved:

– For the Annual Surface Temperature Change dataset: Country, ISO3, Indicator, and annual data columns (F1961 to F2022 ).

– For the Climate Related Disaster dataset: Country, Indicator, and annual data columns (F1980 to F2022 ).

Drop Missing Values: Rows with missing values in the climate related disaster dataset were dropped, however, in the surface temperature dataset there were a lot of null values that were essentially treated as 0, therefore the better approach was to add the rows country wise.

Set Index: The Country column was set as the index for both datasets since both the datasets contain the column it would be easier to align both the datasets.

Store in SQLite Database: Once all the operations were performed, the transformed data was stored into a SQLite database.

**Problems Encountered**

• Missing values: Addressed by dropping rows with NaNs.

**Error Handling**

The pipeline includes checks for data integrity and log issues that might be encoutered during the processing.

**Result and Limitations**

**Output Data**

The output data of the pipeline consists of cleaned and transformed datasets stored in an SQLite

database. The datasets include:

• Annual Surface Temperature Change data, containing the mean surface temperature

changes for various countries over several decades.

• climate-related disaster dataset, containing the various natural disasters that occurred

over the years in different countries.

• A merged dataset that will allow to derive insights and correlations much more easily.

**Data Structure and Quality**

Data Structure:

• Format: The data is stored in tables within an SQLite database.

• Tables:

– Annual Surface Temperature Change: Contains columns for Country, Indicator, and

annual temperature change data from 1961 to 2022.

– Climate-Related Disaster: Contains columns for Country, Indicator, and annual num□ber of climate disasters per country from 1980 to 2022.

**Data Quality:**

• Completeness: Once the missing values were handled the data was rendered mostly

complete.

• Consistency: The data was thoroughly cleaned and filtered to ensure consistency.

• Accuracy: Since the data was sources from IMF a highly reputable institution therefore

it is quite accurate.

**Data Format**

Data Format: The cleaned and transformed data is stored in an SQLite database.

Reason for Choosing SQLite:

• Efficiency: SQLite provides the perfect vessel for storage and its querying capabilities

make handling large datasets much easier..

**Critical Reflection and Potential Issues**

Reflection on Data:

• Strengths:

– The pipeline insures data reliablility and querying capabilities.

• Potential Issues:

– Data Completeness: Handling missing values might lead to the loss of important information.

– Geographical and Temporal Coverage: The data covers specific regiion which may affect the analysis..

– Future Adaptations: Highlighting the need for potential adaptations to the pipeline to accommodate future changes in data sources.