Global Climate Trends Analysis

Ollscoil Teicneolaíochta an Atlantaigh

Atlantic Technological University

Author: Aaron Darcy

Module: Data Operations & Management

Date: December 18th, 2025

Abstract

Climate change is one of the most pressing global challenges, and long-term temperature trends play a central role in shaping policy, industry planning, and public understanding (United Nations 2025). This project analyses historical global temperature anomaly data from Berkeley Earth to identify long-term warming patterns at both global and country levels (Berkley Earth 2025) A Python-based data pipeline was developed to programmatically access, clean, and preprocess monthly country-level temperature data. To ensure analytical reliability, a global coverage analysis was performed, and the dataset was restricted to the first year in which at least 95% of countries had available data. Exploratory data analysis and feature engineering were then carried out, including seasonal classification, baseline comparison, country-level z-scores, long-term warming trends, and anomaly volatility. The processed datasets were visualised using an interactive Power BI dashboard to support comparative and exploratory analysis. The results confirm a clear and sustained increase in global temperature anomalies over time, with meaningful variation in warming rates and variability across countries and seasons. The project also reflects on key data governance considerations, including data quality, transparency, licensing, and responsible use of climate data in decision-making contexts.

1. Introduction

Global surface temperatures have increased significantly since the late nineteenth century, with recent decades showing accelerated warming (NASA Science 2010). This rise has far-reaching implications, affecting extreme weather events, sea-level rise, food systems, energy demand, and public health (WHO 2025). For industry and government alike, understanding not just that the climate is warming, but how warming varies across regions and seasons, is increasingly important for risk assessment and long-term planning.

This project focuses on the analysis of historical temperature anomaly data. Temperature **anomalies represent deviations from a long-term average** rather than absolute temperature values, making them suitable for comparing regions with very different baseline climates. As a result, anomaly-based datasets are widely used in climate research and by major institutions such as Berkley (Meegle.Com 2025).

1.1 Aims & Objectives

The aims of this project are to:

- Collect monthly temperature anomaly data for all available countries from Berkeley Earth.
- Clean and standardise the dataset to ensure consistency and reliability.
- Assess global data coverage over time and restrict analysis to periods with sufficient coverage.
- Perform exploratory data analysis and feature engineering to reveal long-term trends and variability.
- Present results using an interactive Power BI dashboard.
- Appraise the data governance implications of using public climate data in analytical projects.

1.2 Literature Context

Global temperature records are typically constructed by combining land station measurements and sea surface temperature observations into long-term anomaly series. Major datasets such as HadCRUT5 (Morice *et al.* 2021) and Berkeley Earth's land temperature records (Berkley Earth n.d.) consistently show a strong upward trend in global mean temperatures. While there are methodological differences between datasets, the overall conclusion is: the planet is warming, and recent decades are the warmest on record. This project does not aim to challenge these findings, but rather to explore them through transparent, reproducible data analytics.

2. Methodology and Implementation

2.1 Data Source and Access

The dataset used in this project was sourced from Berkeley Earth, a publicly available and widely cited provider of climate data. A list of countries was scraped from the Berkeley Earth website (Berkley Earth n.d.), and monthly temperature anomaly files were downloaded directly from their online repository. Each file contains metadata followed by monthly records consisting of year, month, and anomaly values.

Python was used to automate data access, ensuring that all countries were processed consistently and that the analysis could be reproduced.

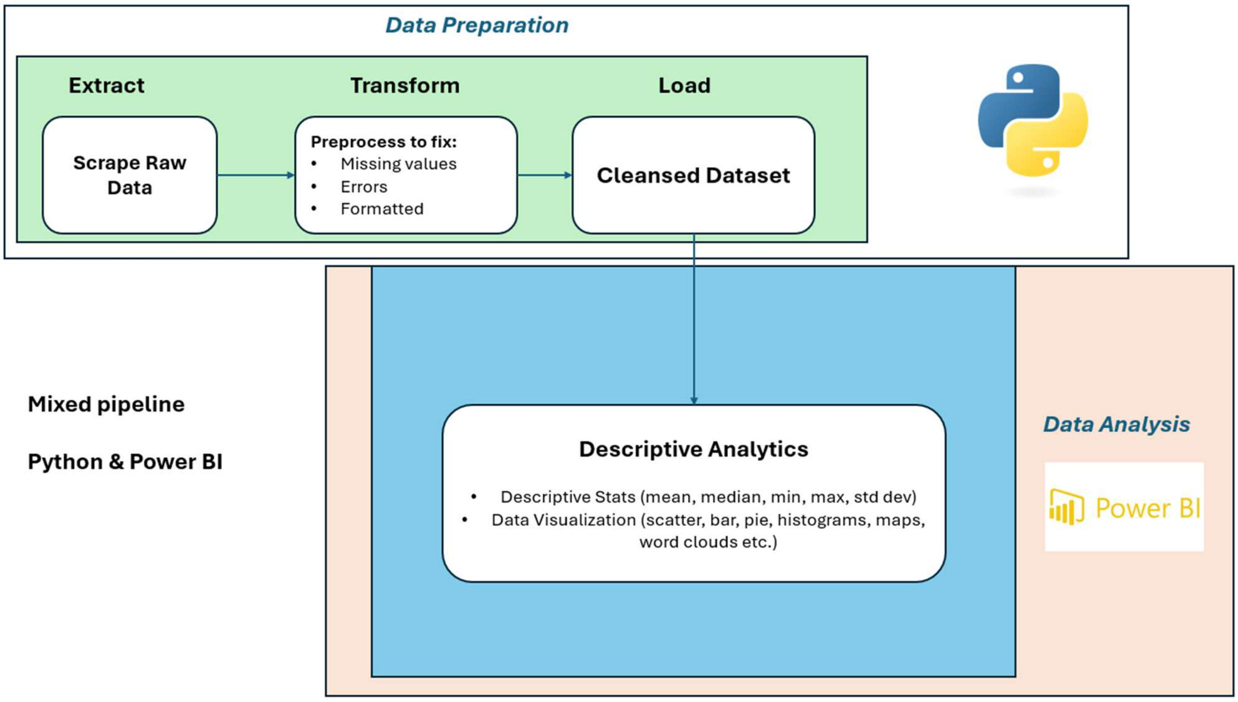

Figure 1 – Mixed Pipeline of Python & Power BI Displaying the ETL Process

2.2 Data Preprocessing

Several preprocessing steps were required before analysis:

- Metadata and empty lines were removed from each country file.
- Data types were standardised to ensure numerical consistency.
- Country names were normalised to match file naming conventions.
- All country-level datasets were combined into a single structured table.

Basic validation checks were performed, including row counts, missing value inspection, and verification of the number of unique countries.

2.3 Coverage Analysis

Historical climate data often suffers from uneven geographic coverage, particularly in earlier years. To address this, the number of countries reporting data for each year was calculated. The dataset was then restricted to the first year where at least **95% of countries** had valid data. This step reduces bias and improves the interpretability of global averages.

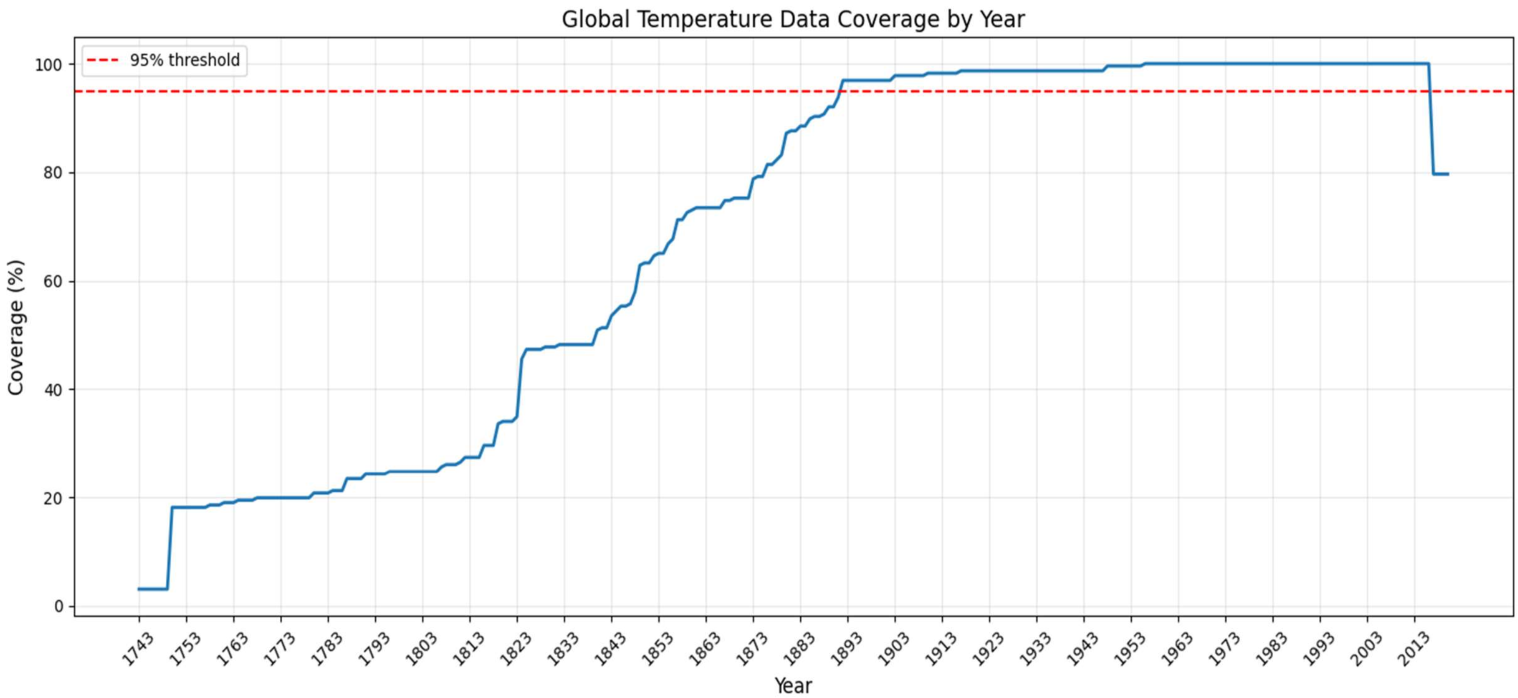

Figure 2 – A line chart which displays coverage of data by year giving a systematic approach to pick the first year in which we start processing records for

2.4 Feature Engineering

To support deeper analysis, several additional features were engineered:

- Season (Winter, Spring, Summer, Autumn)
- Baseline anomaly per country
- Change from baseline
- Country-level anomaly z-scores
- Long-term warming trend, estimated using linear regression
- Volatility, measured as the standard deviation of anomalies

The final cleaned and engineered datasets were exported as CSV files for use in Power BI

3. Results & Data Governance

The exploratory analysis shows a clear and sustained increase in global temperature anomalies over time. Annual global averages demonstrate a strong upward trend, particularly from the mid-twentieth century onwards.

Seasonal analysis indicates that warming is not uniform throughout the year, with some seasons exhibiting larger anomaly increases. At country level, the results show meaningful differences in both the rate of warming and the variability of temperatures.

The Power BI dashboard allows users to:

- Explore global and country-level trends interactively
- Compare warming rates between countries
- Filter results by season and period
- Identify countries with higher volatility or faster warming

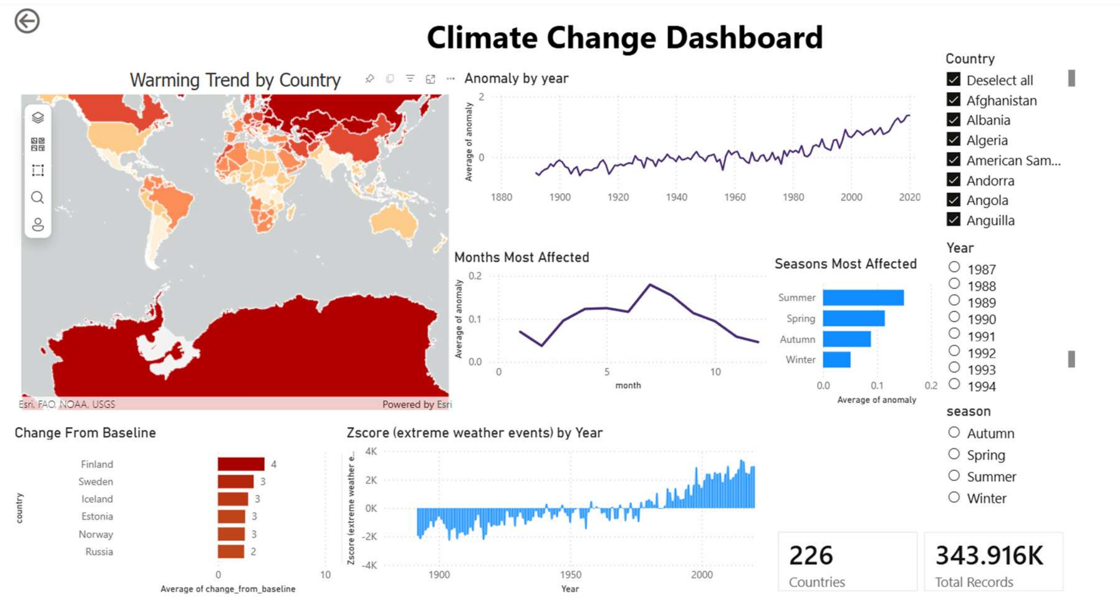

Figure 3 – Complete interactive dashboard implementation in PowerBI

The findings align closely with established climate research, reinforcing evidence of long-term global warming. However, the country-level and seasonal breakdowns highlight that climate change is not experienced uniformly. From an industry perspective, this matters:

infrastructure planning, agriculture, insurance, and energy systems all depend on local and seasonal climate behaviour rather than global averages alone.

Interactive visualisation adds practical value by allowing stakeholders to explore patterns themselves, rather than relying solely on summary statistics.

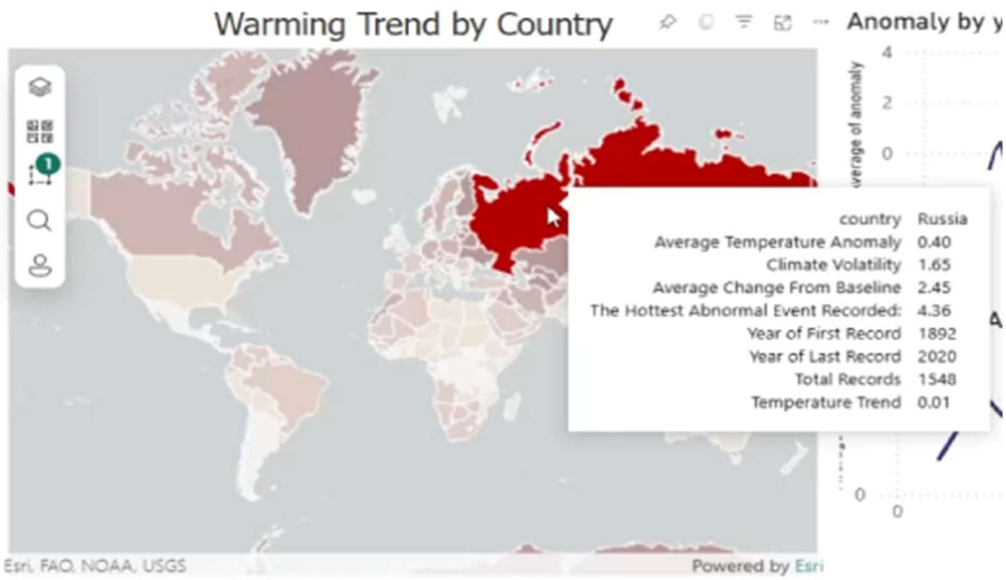

Figure 4 - Interactive visualisation of selecting a country displaying more detail on the region

3.1 Data Governance Implications

This project also highlights several important data governance considerations relevant to both academic analysis and real-world data operations.

3.1.1 Data Source and Licensing:

The data originates from Berkeley Earth and is publicly available for research and educational use. While no restrictive licensing barriers were encountered, proper citation and source transparency remain essential to ensure ethical reuse of the data. Relying on the original source repository also reduces the risk of working with outdated, modified, or unofficial datasets, which is particularly important when analysing data that informs public discourse and policy decisions.

3.1.2 Data Quality:

Key data quality challenges include uneven historical coverage across countries and inconsistencies in file structure and formatting. These issues were addressed through explicit coverage analysis, filtering the dataset to periods with at least 95% global coverage, parsing logic, and systematic validation checks. Treating data quality as something to be actively measured, monitored, and documented, rather than assumed, improves analytical reliability and helps prevent misleading conclusions drawn from incomplete or biased data.

3.1.3 Data Lineage and Transparency:

The full analytical pipeline, from raw source files through cleaning, feature engineering, and final visualisation outputs, is documented and reproducible. Clear intermediate datasets and transparent transformation steps allow others to trace how results were produced and verify analytical decisions. This level of traceability supports reproducibility, accountability, and trust, which are critical in both academic research and industry decision-making contexts, particularly when working with high-impact datasets such as climate records (Dataversity n.d.).

4. Conclusion

This project developed an end-to-end data analytics pipeline to explore global climate trends using historical temperature anomaly data. The analysis confirms a clear global warming signal, with notable variation across countries and seasons. By combining transparent preprocessing, coverage validation, feature engineering, and interactive visualisation, the project demonstrates how public climate data can be used responsibly and effectively.

Limitations include the focus on descriptive analysis and reliance on a single climate variable. Future work could extend this analysis through predictive modelling, integration of additional climate indicators, or linking temperature trends to socio-economic impacts. Overall, the project shows the value of combining sound data governance with modern analytics tools to support informed understanding of climate change.

5. References

Nations, U. (2025) Climate Change [online], *United Nations*, available: https://www.un.org/en/global-issues/climate-change [accessed 17 Dec 2025].

Berkley Earth (2025) 'Data Overview', *Berkeley Earth*, available: https://berkeleyearth.org/data/ [accessed 17 Dec 2025].

NASA Science (2010) 'World of Change: Global Temperatures - NASA Science', available: https://science.nasa.gov/earth/earth-observatory/world-of-change/global-temperatures/ [accessed 18 Dec 2025].

WHO (2025) Climate Change [online], available: [illegible] [accessed 18 Dec 2025].

Meegle.Com (2025) Anomaly Detection In Climate Change [online], available: [illegible] [accessed 18 Dec 2025].

Dataversity (n.d.) 'Harnessing Data Lineage to Enhance Data Governance Frameworks', *Dataversity*, available: https://www.dataversity.net/articles/[illegible]/ [accessed 18 Dec 2025].