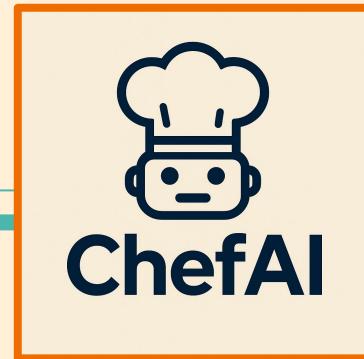


# GenHack 2025

---

Phase: Warmup

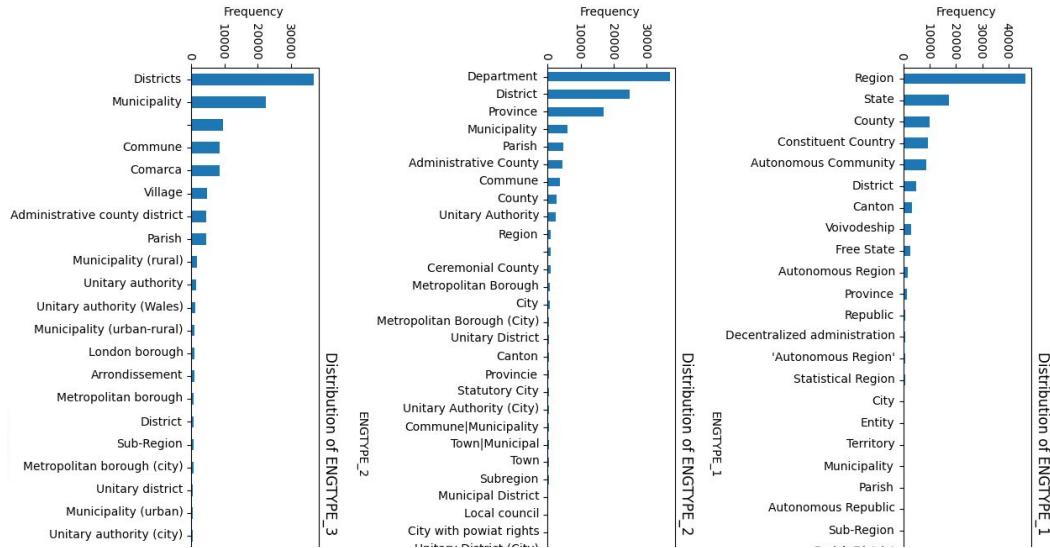
## Data Exploration



# Dataset Exploration - GADM

The data is in the form of a **GeoPackage file**, a standard format for storing geospatial data.

The GADM dataset follows a **hierarchical structure of administrative divisions**, using numerical suffixes to denote different levels: (GID\_0: Country level, GID\_1: First-level administrative division, ...)



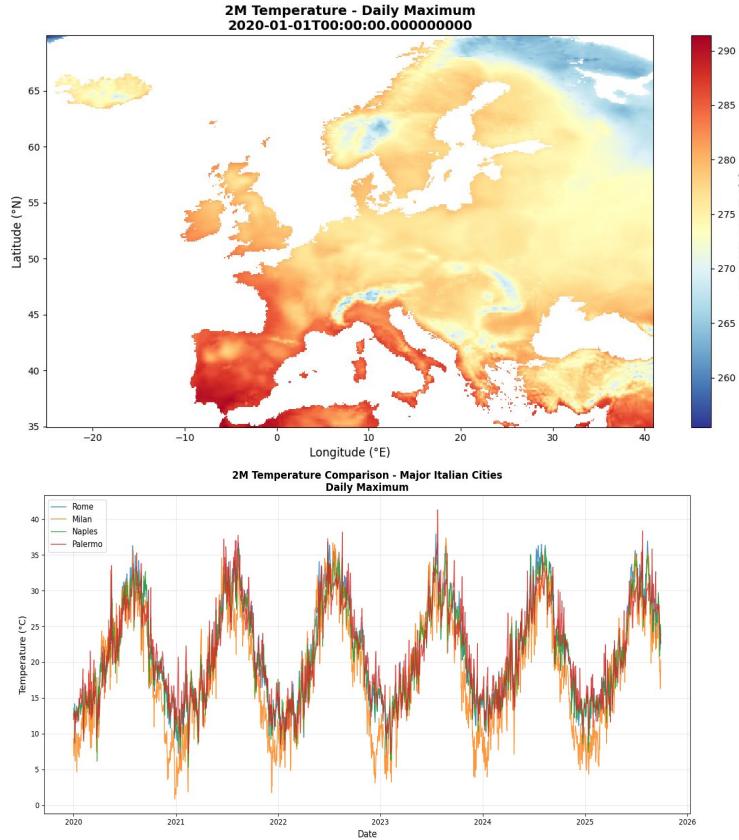
**Different countries may have varying numbers of administrative levels**, and the types of divisions can differ significantly.

# Dataset Exploration - ERA5

ERA5-Land is a high-resolution, land-focused reanalysis dataset derived from ERA5, providing a consistent view of land variables over several decades by integrating model data with global observations using the laws of physics.

## Features

- ERA5-Land uses  $0.1^\circ \times 0.1^\circ$  grid
- Since the dataset uses a regular latitude-longitude grid and the Earth is a geoid, **the physical size of each grid cell varies with latitude**: the east-west extent decreases toward the poles, while the north-south extent stays roughly constant.

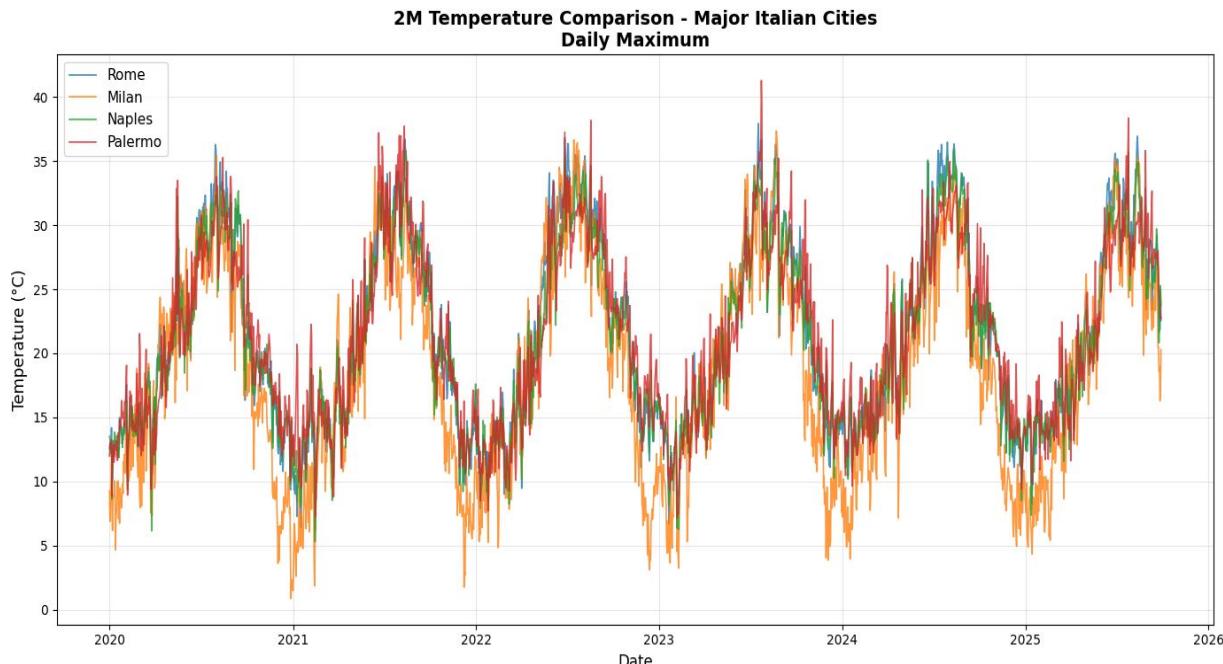


Similar seasonal patterns, with variations in temperature levels due to geographical locations.

# Dataset Exploration - ERA5 - Point-Based Time Analysis

Other than spatial analysis, we can also perform point-based time series analysis for specific locations.

Let's extract time series for major Italian cities:



We can see that the curve trends for the different cities show similar seasonal patterns, with variations in temperature levels due to their geographical locations. For example, Milan has a generally lower temperature compared to Palermo.

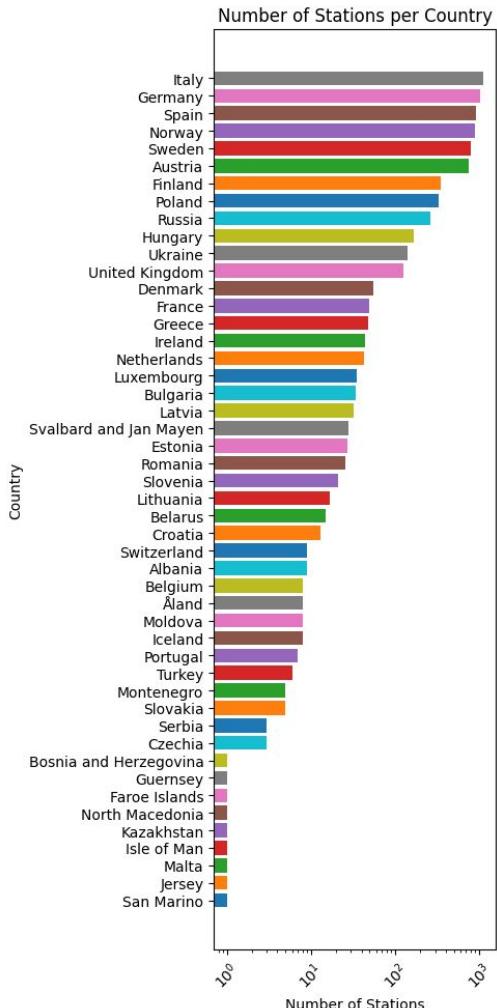
# Dataset Exploration - ECA Blend

The **ECA dataset** includes daily station series from national meteorological services across Europe and the Mediterranean Area.

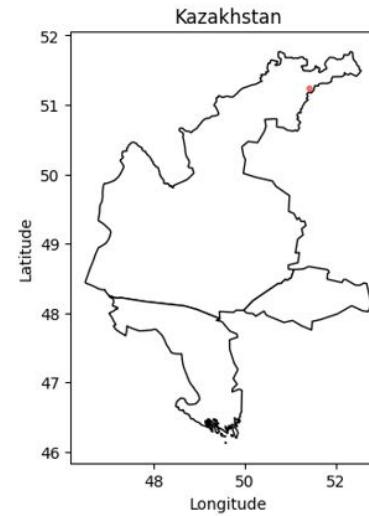
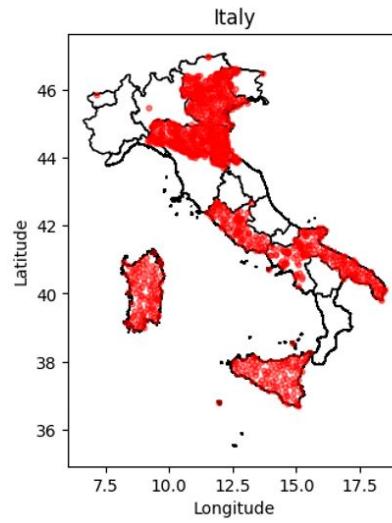
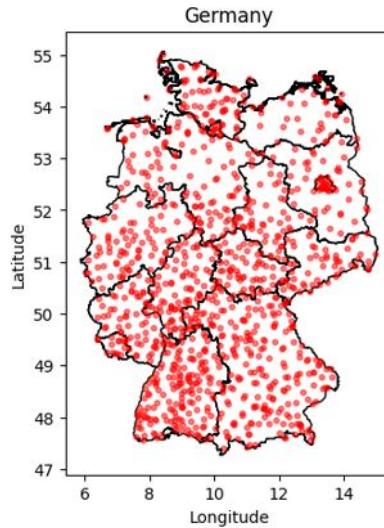
**Non-blend** series contain only the original data from these sources, while **blend series** supplement missing or recent data using near real-time SYNOP reports from the Global Telecommunication System (GTS) to create more complete time series.

## Features

- Station coordinates are in DMS (Degrees, Minutes, Seconds) format and need to be converted to decimal degrees for analysis.
- Each station has a unique identifier (*STNID*) that can be used to link metadata with the actual climate data files.
- The series are **quality controlled** and flags ("OK", "suspect" or "missing") for individual data are attached.



# Dataset Exploration - ECA Blend - Station Density



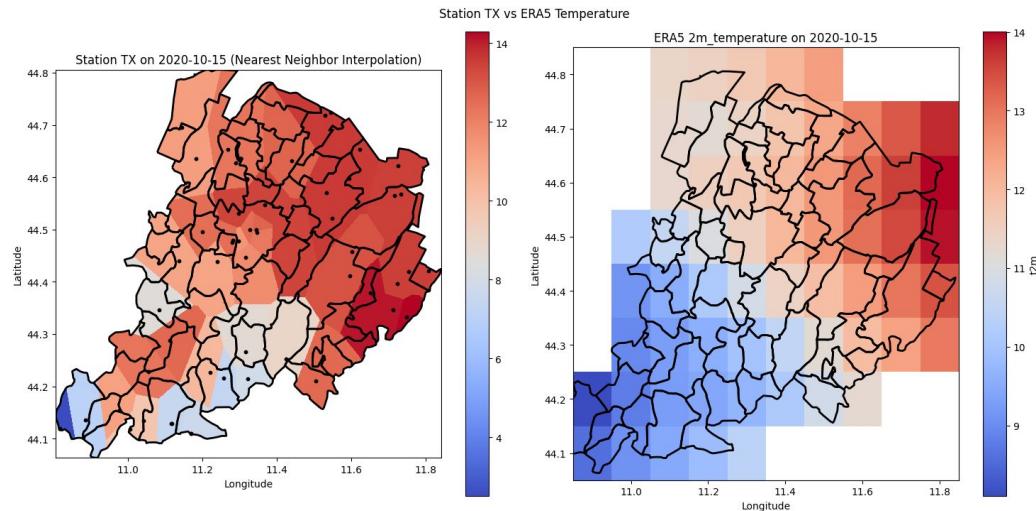
The weather stations included in the dataset are **not uniformly distributed across all countries**. There are cases of sufficiently large states with a substantial number of stations, states that are too small to be meaningful for territorial analysis, and states in which the number of available weather stations is not significant.

The **oldest records** date to **1760s**, an improbable starting point that is too remote to be relevant, with a frequency too sparse to support accurate real-time analysis of atmospheric phenomena.

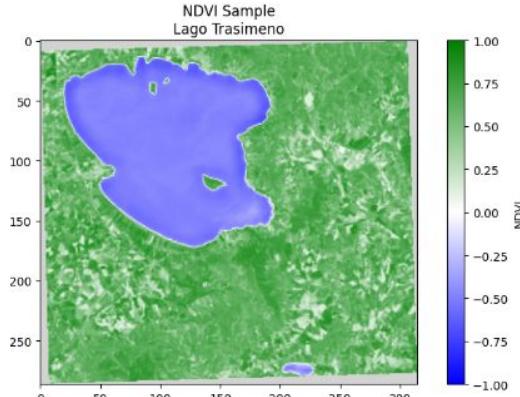
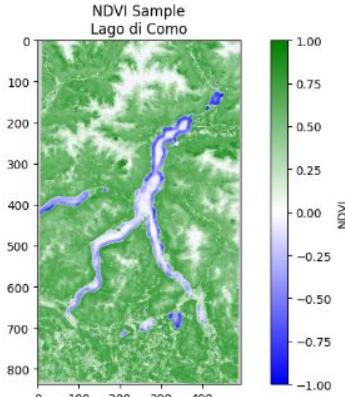
# ECA Interpolation Comparison With ERA5

Using the ECA dataset, we crafted an **interpolated temperature grid** over our Region of Interest (Bologna) and then compared these interpolated values with the corresponding temperature data from the ERA5 dataset.

While both maps show a general tendency for higher temperatures in the eastern part of the region and lower temperatures in the southwest, **the magnitude and positioning of these patterns differ substantially**. The interpolated dataset suggests **stronger local contrasts** that ERA5 smooths out.



# Dataset Exploration - NDVI



The dataset is provided in **GeoTIFF** format, which is a standard format for storing geospatial raster data.

In this format each pixel in the image corresponds to a specific geographic location and contains a value representing the NDVI at that location.



On the Lago di Como we can observe some areas with **higher NDVI values inside the lake**.

This could be due to the **presence of atmospheric effects** that influenced the reflectance values recorded by the satellite, leading to anomalous NDVI readings in those areas.

