# **Overview**

This project primarily deals with the analysis of temperature data of German weather stations over the timeframes 1996-2021. The analyses performed include data preprocessing, visualization, and anomaly detection using Python and Gradio-based interactive dashboards.

# **Dataset**

File: german\_temperature\_data\_1996\_2021\_from\_selected\_weather\_stations.csv

# **Description**

This dataset contains temperature readings recorded at various weather stations in Germany over a period of 25 years. It serves to interrogate temperature trends, seasonal variations, and anomalies in temperature data.

## Columns

• date: The date of the recorded temperature.

• station\_<id>: Temperature measurements from specific weather stations.

• Additional derived columns:

* month: Extracted from date for analysis of monthly trends.
* year: Extracted from date for yearly trend analysis.

# **Data processing**

Steps

1. Import the dataset: Read the CSV file in a Pandas DataFrame.

2. Handle missing values:

* Forward-fill (ffill) missing values to fill continuation.
* For station\_298, fill in with mean temperature.

3. Data Cleaning:

* Convert date column to datetime format.
* Ensure all station columns are numeric.
* Remove duplicate rows if any.

4. Feature Engineering:

* Creation of month and year columns to incorporate time in the analysis.

# **Visualization and Analysis**

Implemented Visualizations

• Time Series Analysis: Charts the variation of temperature against time.

• Monthly Heatmap: Appropriate heatmap displays the variation in temperature by month and year.

• Seasonal Box Plot: Seasonal distribution of temperatures.

• Yearly Trend Analysis: Uses moving averages to identify trends.

• Station Comparison: Attempts to compare the temperatures of various stations.

• Anomaly Detection: Z-score based algorithm to find anomalies in the temperature assay data.

• Correlation Heatmap: Compare all stations' readings.

• Histogram. Temperature distribution.

# **Installation and Requirements**

Dependencies

To run the analysis and visualization, the following required python libraries should be installed:

pip install pandas numpy matplotlib sea-born plotly folium geopandas statsmodels gradio

Running The Application

Enter the following command to launch the interactive Gradio-based dashboard:

python script.py

**User guide**

1. Select Visualization Type: Choose from the dropdown list.

2. Select a Weather Station: Select a weather station for analysis.

3. Set Date Range: Specify Start and End dates.

4. Explore Additional Options: Set parameters for anomaly detection, correlation, and comparisons.

5. Generate Visualizations: Click the button to create the desired plot.