

# Finding Hidden Weather Patterns

— *An Unsupervised Learning Exploration*

## Learning Objectives

By the end of this project students will be able to

1. Clean & pre-process real-world sensor data.
2. Engineer meaningful features from raw weather measurements.
3. Run, evaluate and compare **at least two unsupervised algorithms** (K-Means + one more).
4. Visualise clusters in 2-D/3-D and **interpret** them meteorologically.
5. Report findings in a reproducible Jupyter notebook + short PDF summary.

## Data Set at a Glance

- **File:** daily\_weather.xlsx
- **Rows:** 92 days (Dec 2024 → Feb 2025)
- **Columns:** 9 variables
  - DateTime (daily)
  - AirTemperatureMin/Max/Avg (°C)
  - HumidityMin/Max/Avg (%)

- RainSum (mm)
- AirPressureAvg (hPa)

## **Quick EDA Questions to Ask First**

- Are there duplicated dates?
- Missing or zero-values that look suspicious?
- Any obvious seasonal trend in temp/pressure?
- Does rain correlate with humidity spikes?

## **Tooling Stack**

- Python 3.10+
- pandas, numpy, matplotlib, seaborn
- scikit-learn (clustering, preprocessing, metrics)

## **End-to-End Pipeline Outline**

1. Data wrangling & cleaning
2. Feature engineering & scaling
3. Dimensionality reduction (PCA / t-SNE / UMAP)
4. Clustering (K-Means + DBSCAN or Gaussian Mixture)
5. Cluster evaluation (Silhouette, Davies-Bouldin, elbow)
6. Interpretation & visual storytelling
7. Notebook + 2-page PDF report

**Your turn:**

Open the Excel file, run a quick `df.info()` & `df.describe()` and note down **three data-quality issues** you spot.

*Good luck!*