

# GenHack 2025 : Analyse des Îlots de Chaleur Urbains (UHI)

Équipe : NoName (Semaine 1)

Basé sur l'analyse approfondie des notebooks

17 novembre 2025

# Contexte et Objectif

## Contexte du Challenge (Semaine 1)

- ▶ Explorer l'effet d'Îlot de Chaleur Urbain (UHI).
- ▶ Utiliser les données de température ERA5, les cartes NDVI de Sentinel-2 et les observations des stations au sol.

## Objectif Principal

- ▶ Démontrer que l'ajout du NDVI (végétation) améliore la prédiction de la température au sol ( $T_{\text{Station}}$ ).
- ▶ Prouver la capacité à généraliser un modèle local (downscaling).

# Sources de Données Utilisées

Ce notebook fusionne quatre sources de données distinctes :

## 1. Température au Sol (Vérité Terrain)

- ▶ Source : ECA\_blend\_tx/ (Données ECA&D)
- ▶ Fichiers : stations.txt et TX\_...txt

## 2. Occupation du Sol (Végétation)

- ▶ Source : sentinel2\_ndvi/ (Données Sentinel-2)
- ▶ Fichiers : Images GeoTIFF (.tif)

## 3. Température Globale (Météo)

- ▶ Source : derived-era5-land-daily.../ (Données ERA5)

## 4. Géométrie (Frontières)

- ▶ Source : gadm\_410\_europe.gpkg

# Définition de la Zone d'Intérêt (AoI) : La France

L'analyse est explicitement filtrée pour se concentrer sur la France.

- ▶ 1. Stations (Vérité Terrain) :
  - ▶ Le fichier global stations.txt est filtré :  
`stations_df [ 'CN' ] == 'FR'`
  - ▶ Résultat : L'analyse se limite à **44 stations** situées en France.
- ▶ 2. Données Météo (ERA5) :
  - ▶ Le fichier NetCDF chargé est déjà pré-filtré pour la France  
(...\_FRANCE.nc).
- ▶ 3. Données NDVI (Végétation) :
  - ▶ L'échantillonnage ne se fait **qu'aux coordonnées** des 44 stations françaises.

# Méthodologie de Fusion et d'Analyse

## 1. Fusion Spatiale (Fonction `create_snapshot`) :

- ▶ Crée un DataFrame unique (`T_Station`, `T_ERA5_C`, `NDVI`) pour les 44 stations à des dates clés.

## 2. Analyse de Corrélation (Preuve UHI) :

- ▶ Vérification de la corrélation négative entre `NDVI` et `Temperature_C`.

## 3. Modélisation par Cluster (Downscaling) :

- ▶ Entraînement d'un Modèle B (`T_ERA5 + NDVI`) sur un cluster local (ex : Marseille).
- ▶ Test de sa capacité à "généraliser" sur une station non vue (ex : Toulouse).

## 4. Analyse de performance par Altitude :

- ▶ Segmentation des stations en "Basse" ( $\leqslant 500m$ ) et "Haute" ( $> 500m$ ) Altitude.

# Difficultés et Points Méthodologiques Clés

- ▶ **Gestion des Données (Drive) :**
  - ▶ Navigation dans un Google Drive partagé (chemins et filtres glob.glob).
- ▶ **Complexité Géospatiale :**
  - ▶ Gestion de Systèmes de Coordonnées de Référence (CRS) différents (EPSG :4326 vs EPSG :3035).
- ▶ **Hétérogénéité des Données :**
  - ▶ Fusion de données tabulaires (CSV), matricielles NetCDF (ERA5) et matricielles GeoTIFF (NDVI).

# Analyse Approfondie - Preuve de l'UHI

Concerne : Cellule 20 (Nuage de points)

**Objectif : C'est le graphique clé pour l'hypothèse.**

- ▶ Vise à prouver visuellement l'effet UHI dans le snapshot du 15 juillet 2022.

**Analyse Détailnée :**

- ▶ Le graphique trace Temperature\_C (Axe Y) en fonction du NDVI (Axe X).
- ▶ Il montre une **corrélation négative** évidente.
- ▶ **Interprétation :**
  - ▶ **NDVI faible (gauche)** → Peu de végétation → Températures élevées.
  - ▶ **NDVI élevé (droite)** → Beaucoup de végétation → Températures **basses**.
- ▶ **Coefficient de régression (Modèle B) : -2.203 °C.**

# Résultats de la Modélisation (Downscaling)

Validation de l'ajout du NDVI et de la généralisation

## 1. Amélioration du Modèle (Cluster Marseille - Été) :

- ▶ Modèle A (T\_ERA5 seul) :  $R^2 = 0.632$
- ▶ Modèle B (T\_ERA5 + NDVI) :  $R^2 = 0.643$
- ▶ Conclusion : L'ajout du NDVI (facteur local) améliore la prédiction.

## 2. Validation Externe (Robustesse) :

- ▶ Le modèle B (entraîné sur Marseille) est utilisé pour prédire une station non vue (Toulouse).
- ▶ Résultat : Erreur moyenne (RMSE) d'environ  $1.5 \text{ } ^\circ\text{C}$  sur les quatre saisons.
- ▶ Conclusion : Le modèle B réussit à généraliser.

# Conclusion Stratégique - L'Effet de l'Altitude

Le besoin de correction (downscaling) dépend de l'altitude

L'altitude est le meilleur prédicteur de la faiblesse de l'ERA5 :

Basse Altitude ( $\leq 500m$ ) :

- ▶ Le Modèle ERA5 est déjà excellent.
- ▶  $R^2$  (ERA5 seul)  $\approx 0.900$
- ▶ Besoin : Le NDVI (correction locale) est peu utile.

Haute Altitude ( $> 500m$ ) :

- ▶ Le Modèle ERA5 est **faible**.
- ▶  $R^2$  (ERA5 seul)  $\approx 0.35$
- ▶ **Besoin** : Le NDVI (correction locale) est **crucial** pour capturer les microclimats.

# Prochaines Étapes (Semaine 2)

## 1. Feature Engineering (Priorité 1) :

- ▶ Introduire l'\*\*altitude\*\* (HGHT) comme variable explicative.
- ▶ Objectif : Cibler directement la faiblesse de l'ERA5 et réduire l'erreur de prédiction dans les zones montagneuses.

## 2. Modèles Avancés :

- ▶ Tester des modèles non linéaires (Random Forest, Gradient Boosting).
- ▶ Objectif : Capturer des relations plus complexes (ex : interaction entre NDVI, altitude et ensoleillement).