

# Processus de Traitement des Données et Justification des Nettoyages

## 1 Fusion et Sélection des Données

Nous avons combiné les données de deux sources (dpe-v2-logements-neufs et dpe-v2-logements-existants) pour obtenir une vue d'ensemble complète. Après cette fusion, nous avons sélectionné uniquement les colonnes pertinentes afin de simplifier l'analyse.

## 2 Suppression des Colonnes Inutiles

Certaines colonnes ont été supprimées pour plusieurs raisons :

- **Colonnes redondantes** : Des informations similaires étaient déjà présentes sous d'autres formes (ex : Code\_postal\_(BAN) et Code\_INSEE\_(BAN)).
- **Colonnes incomplètes** : Certaines variables contenaient trop de valeurs manquantes (Production\_électricité\_PV\_kWh avec des valeurs mixtes).
- **Colonnes non pertinentes pour l'analyse** : Certaines informations n'étaient pas utiles pour notre étude sur la consommation énergétique et les émissions de GES (ex : N°DPE).

## 3 Nettoyage des Données Textuelles

Nous avons identifié des erreurs d'encodage dans les noms de communes (Nom\_\_commune\_(BAN)) et appliqué un remplacement automatique des caractères mal encodés (Ã, Å, etc.) pour harmoniser les valeurs.

## 4 Filtrage des Données

- **Filtrage des Années de Construction** : Nous avons conservé uniquement les bâtiments construits après 1900 pour éviter des valeurs aberrantes ou non pertinentes.
- **Filtrage des Valeurs Erronées** : Les entrées où Production\_électricité\_PV\_kWh contenait du texte ont été supprimées pour assurer la cohérence des calculs.

## 5 Création des Mesures et Indicateurs

Nous avons défini plusieurs mesures pour l'analyse :

- **Consommation Totale** : Somme des consommations énergétiques de chaque logement.
- **Consommation par m<sup>2</sup>** : Rapport entre la consommation totale et la surface.

- **Émissions de GES Totales et par m<sup>2</sup>** : Calcul des émissions en fonction des surfaces et des types de bâtiments.
- **Énergie Renouvelable (%)** : Indicateur du ratio entre l'énergie produite en interne (Production\_électricité\_PV\_kWh) et la consommation totale.

## **6 Structuration et Visualisation des Données**

Nous avons structuré le rapport en trois pages :

- **Bilan Énergétique Global** : Vue d'ensemble avec indicateurs clés, cartes et graphiques de distribution.
- **Comparaison Énergétique Communale** : Analyse détaillée par commune avec comparaison de la consommation et des émissions de GES.
- **Sources et Émissions Énergétiques** : Focus sur les sources d'énergie et leur impact sur les émissions.