



▣ CAHIER DES CHARGES – PROJET FIL ROUGE DATA ANALYST

**Projet Data Analytics
Fil Rouge**

Réalisé par : Mawada Ennaciri

Encadré par : Yassine Ammami

Année universitaire : 2025 – 2026



1 Intitulé du projet

Analyse des logements énergivores (“passoires thermiques”) afin de prioriser les stratégies de rénovation énergétique à partir des données territoriales et socio-économiques

2 Contexte métier

La transition énergétique est une priorité nationale en France.

Les logements classés **F et G au DPE** (appelés *passoires thermiques*) représentent un enjeu majeur :

- précarité énergétique des ménages,
- émissions élevées de CO₂,
- coûts importants pour les collectivités.

Les pouvoirs publics doivent **prioriser les rénovations** en fonction :

- de la performance énergétique,
 - du niveau de revenu des habitants,
 - de la densité de population,
 - du territoire.
-

3 Problématique analytique

Quelles communes présentent la plus forte concentration de logements énergivores, combinée à une population vulnérable économiquement, et doivent être prioritaires pour les politiques de rénovation énergétique ?

4 Objectifs du projet

- Identifier les communes à forte proportion de passoires thermiques
 - Croiser DPE ↔ revenus ↔ population ↔ territoire
 - Construire des **indicateurs de priorité de rénovation**
 - Fournir une **analyse data-driven** exploitable par une collectivité
 - Produire des **dashboards clairs et cartographiques**
-

5 Périmètre géographique

📍 **France – échantillon de 2 à 3 régions uniquement** (exemple) :

- Île-de- France
- Hauts-de-France
- Auvergne-Rhône-Alpes

6 Datasets – Sources EXACTES & types

1 Données DPE des logements

- **Type** : données énergétiques
- **Format** : fichiers CSV / base SQL / API exportée
- **Contenu** :
 - classe énergétique des logements (A à G)
 - type de logement (maison, appartement)
 - année de construction
 - code INSEE de la commune
- **Utilité** :
 - identifier les logements énergivores (passoires thermiques F/G)
 - base principale de l'analyse énergétique

2 Données population communale

- **Type** : données démographiques
- **Format** : CSV / fichiers tabulaires
- **Contenu** :
 - population totale par commune
 - évolution démographique
 - code INSEE de la commune
- **Utilité** :
 - rapporter les passoires thermiques à la population
 - comparer les communes entre elles

3 Données revenus communaux

- **Type** : données socio-économiques
- **Format** : CSV / Excel
- **Contenu** :
 - revenu médian par commune
 - niveau de vie des ménages
 - indicateurs de précarité
 - code INSEE de la commune
- **Utilité** :
 - analyser la capacité financière des ménages
 - prioriser les aides à la rénovation

4 Données

logements – Parc immobilier

- **Type** : données de logement
- **Format** : CSV / Parquet
- **Contenu** :
 - nombre total de logements
 - résidences principales
 - caractéristiques du parc immobilier
 - code INSEE de la commune
- **Utilité** :
 - contextualiser les données DPE
 - calculer des ratios (passoires / logements)

5 Données géographiques communales

- **Type** : données géospatiales
- **Format** : GeoJSON / Shapefile
- **Contenu** :
 - contours géographiques des communes
 - code INSEE de la commune
- **Utilité** :
 - cartographie des résultats
 - visualisation territoriale dans Power BI ou Python

6 Données comparatives énergie-climat (optionnel)

- **Type** : données environnementales agrégées
- **Format** : Excel / CSV
- **Contenu** :
 - indicateurs énergie / climat par territoire
 - données de comparaison régionale
- **Utilité** :
 - enrichir l'analyse
 - mise en perspective territoriale

7 Clé de jointure (très important)

Clé commune à TOUS les datasets :

code_insee_commune

Dataset	Nom de la colonne
Population	CODGEO
Revenus	Code_commune

Dataset	Nom de la colonne
Logements RP2022	codgeo
DPE	code_insee_commune
Géographie (GeoJSON)	code ou insee_com

8 Indicateurs & KPI envisagés

- % de logements classés F/G par commune
 - Nombre de passoires thermiques / 1 000 habitants
 - Revenu médian vs classe énergétique
 - Indice de priorité de rénovation (score composite)
-

9 Outils & technologies envisagés

- **Python** : Pandas, NumPy, GeoPandas
 - **SQL** : PostgreSQL / SQLite
 - **Power BI** : dashboards & cartes
 - **VS Code** : développement
 - **Git / GitHub** : versioning
 - **JIRA**
 - **CONFLUENCE**
 - **API REST** : DPE
 - (*Databricks optionnel si volumétrie élevée*)
-

10 Livrables attendus

- Base de données nettoyée et consolidée
- Notebook EDA
- Scripts de fusion & préparation
- Dashboard Power BI
- Slides
- Documentation du data
- Repository GitHub documenté