

Projet Power BI – Évaluation du Cadre de Vie dans les Communes Françaises

M1GPLA - ALVADO Louis, LAMY Julien, ALLAVENA Leïla

1. Problématique et Contexte

En 2020, le logement représente 28,5 % de la dépense de consommation finale des ménages, ce qui en fait l'un des postes les plus lourds dans le budget d'un foyer (*source : INSEE Portrait social de la France édition 2021*). Dans un contexte de tension immobilière nationale, marqué par une hausse continue des loyers, une rareté croissante des logements disponibles, et une forte disparité territoriale, la notion de cadre de vie est devenue un critère déterminant dans les choix résidentiels.

Ce sujet nous touche directement en tant qu'étudiants en alternance. Avec des budgets restreints, nous sommes déjà confrontés aux réalités du marché immobilier, notamment lors de la recherche de logement à proximité de notre entreprise ou de notre établissement de formation. Mais au-delà de cette expérience actuelle, ce projet nous intéresse aussi à titre prospectif.

En effet, plusieurs d'entre nous envisagent à moyen terme :

- d'investir dans l'immobilier locatif, dans des villes à fort potentiel ;
- ou d'acheter leur résidence principale à l'issue du master, en s'installant dans une commune où la qualité de vie est bonne, les services accessibles, et les prix encore abordables.

Dans cette optique, disposer d'un outil d'analyse territoriale clair, fiable et interactif constitue un véritable atout stratégique. Il permet non seulement de comparer les communes sur des critères essentiels (sécurité, qualité des infrastructures, environnement, coût du logement...), mais aussi d'anticiper des décisions importantes liées à notre avenir personnel et professionnel.

2. Objectifs

Nos objectifs étaient de :

- Fournir un outil d'aide à la décision multicritère sur le cadre de vie communal ;
- Permettre une lecture hiérarchique et interactive des territoires (région > département > commune) ;
- Tester l'utilisation de Power BI pour croiser des données ouvertes de sources multiples ;
- Et enfin, en tant qu'alternants, interroger notre propre réalité : comment choisir un lieu de vie, investir ou s'installer, en tenant compte de critères rationnels et objectifs.

3. Indicateurs sélectionnés et critères d'analyse

Domaine	Indicateurs retenus	Détail d'analyse
Sécurité	Taux de criminalité (pour 1 000 hab.)	Moyenne sur 5 ans si disponible
Internet	Débit 4G et filaire max/min	Qualité de la couverture
Environnement	Niveau de pollution (qualitatif)	Faible, Moyen, Élevé
Économie	Revenu médian annuel, part de chômage	En euros & %
Logement	Loyer moyen au m ² selon type de logement (1-2 pièces, 3+, maison), évolution 2018-2024	Lecture temporelle et typologique
Éducation	Nombre d'établissements, taux de réussite collège et lycée	Taux de réussite 2023
Santé	Nombre d'établissements de santé	Catégorisé avec regroupements
Loisirs / Services	Équipements publics recensés	Par type : stade, site naturel, intérieur, etc.

4. Méthodologie

1) Sélection des données

Nous avons commencé par dresser une liste d'indicateurs clés. Dans un premier nous nous sommes questionnés sur ce que nous recherchons comme informations pour vivre dans une commune et nous avons complété avec l'aide de ChatGPT.

Nous avons ensuite récupéré nos bases de données sur :

- Data.gouv.fr
- INSEE
- SNCF

2) Connexion des données à Power BI

Nous avons ensuite pu connecter nos bases de données quand celles-ci étaient au format Excel ou CSV directement au Power BI via Power Query avec l'URL source. Cela permet une actualisation dynamique.

Les bases aux formats parquet, HTML, ou JSON, ont dû être téléchargées et intégrées manuellement, constituant une limite pour les mises à jour automatiques.

3) Nettoyage & structuration

Le nettoyage a représenté une phase longue et complexe, notamment à cause :

- des noms de communes non homogènes d'une base à l'autre (accents, majuscules, orthographe divergentes),
- des géocodes INSEE parfois absents ou multiples pour des communes similaires.
- Les évolutions de noms et fusions de communes.

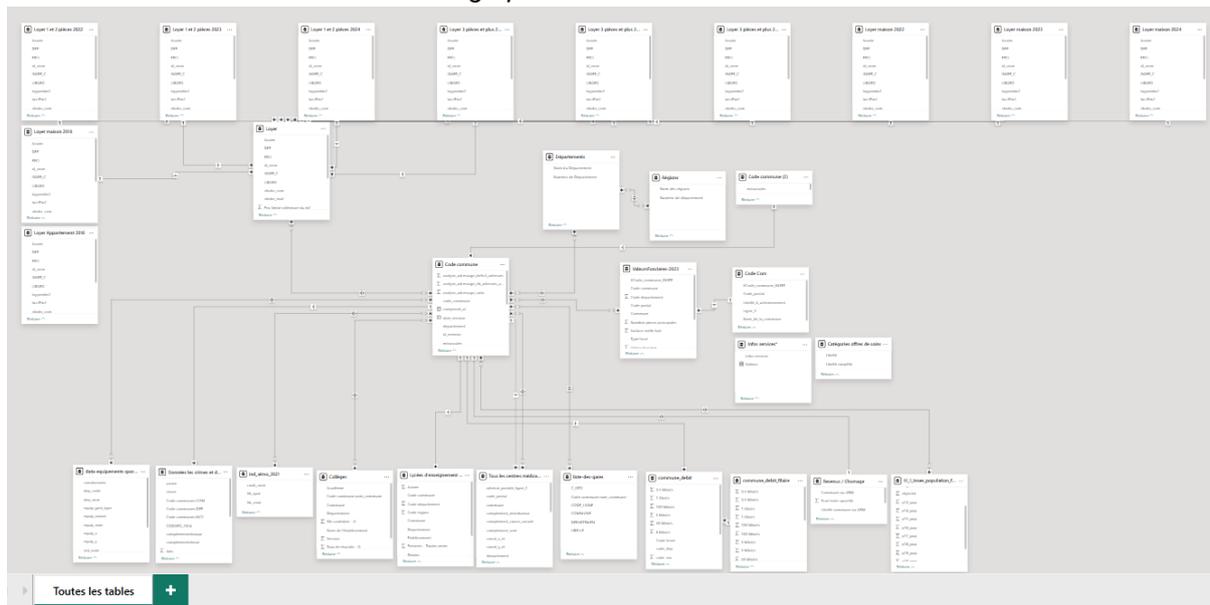
Nous avons donc :

- Supprimer les doublons et colonnes/lignes sans valeur ajoutée pour notre analyse,
- Identifié les points communs (code INSEE) pour relier les jeux de données entre eux,
- Utilisé ChatGPT pour catégoriser certaines bases complexes, notamment celles sur les établissements de santé (147 types initiaux regroupés en 10 catégories).

4) Création du modèle Power BI

- Relations entre tables établies via le code commune INSEE.

Cartographie des connexions du BI



- Création d'indicateurs via des créations de mesures, et formules DAX

🧠 Étape 1 : Créer les mesures DAX individuelles

📁 Nombre d'établissements scolaires

```
DAX
Nb_Etab_Scolaires =
CALCULATE(
    COUNTROWS('Collèges')
) +
CALCULATE(
    COUNTROWS('Lycées d''enseignement supérieur')
)
```

📄 Étape 2 : Créer la table statique avec les indicateurs

```
DAX
KPI_Indicateurs =
DATATABLE(
    "Indicateur", STRING,
    {
        {"🎓", "Nombre total d'établissements scolaires"},
        {"🏆", "Taux de réussite moyen"},
        {"🏥", "Nombre d'établissements de santé"},
        {"🏠", "Nombre total d'équipements publics"}
    }
)
```

🌿 Étape 3 : Créer la mesure dynamique pour afficher la bonne valeur

```
DAX
Valeur_KPI =
SWITCH(
    SELECTEDVALUE(KPI_Indicateurs[Indicateur]),
    "🎓 Nombre total d'établissements scolaires", [Nb_Etab_Scolaires],
    "🏆 Taux de réussite moyen", [Taux_Reussite_Moyen],
    "🏥 Nombre d'établissements de santé", [Nb_Etab_Sante],
    "🏠 Nombre total d'équipements publics", [Nb_Equipements_Publics]
)
```

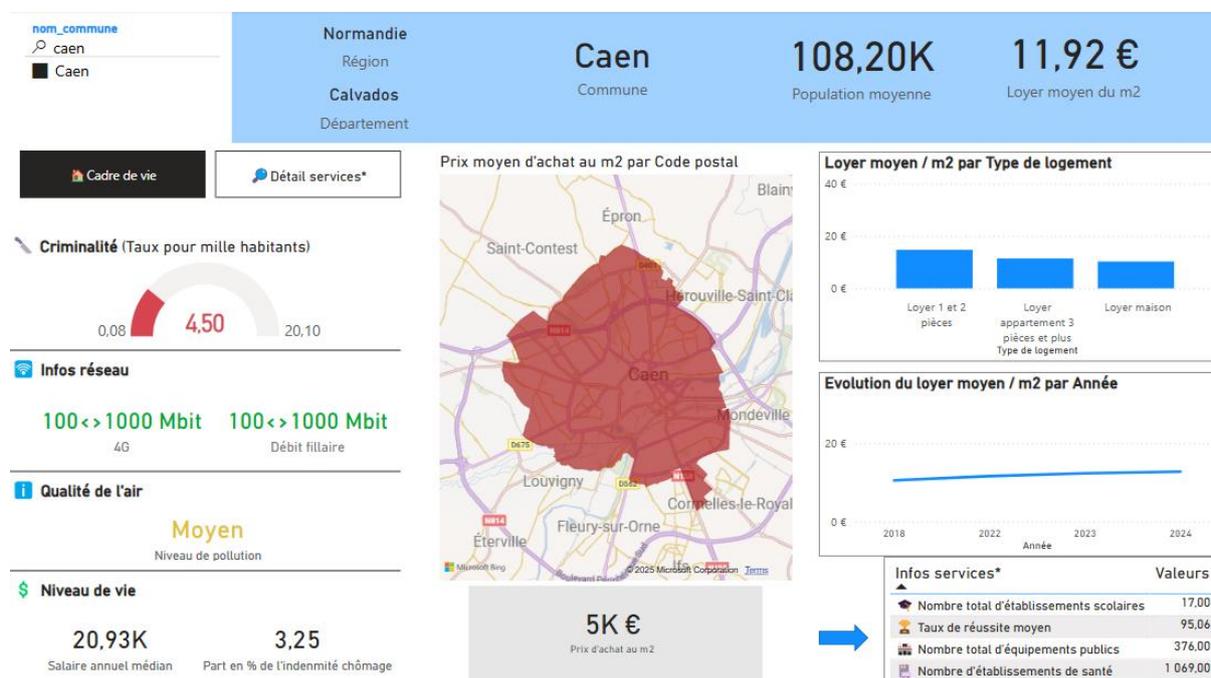
Infos services*	Valeurs
🎓 Nombre total d'établissements scolaires	17.00
🏆 Taux de réussite moyen	95.06
🏥 Nombre total d'équipements publics	376.00
🏠 Nombre d'établissements de santé	1 069.00

5) Résultat

POWER BI lien we transfer téléchargeable jusqu'au 4 avril : <https://we.tl/t-yzXJ5wLjQ1>

Le tableau de bord se structure en deux vues principales :

Vue "Cadre de vie"



Elle présente :

- Les informations de base : Région, Département, Commune (ici : Caen)
- Population moyenne : 108 200 habitants
- Loyer moyen au m² : 11,92 €
- Graphique d'évolution du loyer entre 2018 et 2024
- Criminalité : 4,50‰, soit un niveau modéré
- Connexion Internet : 4G et fibre à 100–1000 Mbit
- Qualité de l'air : Moyenne
- Revenu annuel médian : 20 930 €
- Part de l'indemnité chômage : 3,25 %

Focus sur la carte :

Comment avons-nous fait ? Nous avons pris l'extraction des demandes de valeurs foncières. Nous avons retravaillé la base pour que cette dernière soit à la maille commune.

Puis nous avons fait le lien avec une base commune pivot car il manquait des éléments dans la base DVF. Une perte de 12 % de lignes est à constater.

Comment la comprendre ? Dès lors qu'une ville est sélectionnée, la carte suit l'emplacement géographique (via le code postale). Si aucun nom de ville n'est sélectionné alors une vue sur les départements s'affichent. Il est aussi possible via la flèche au-dessus de la carte de faire un focus sur les départements ou la ville en question.

Un code bi-couleurs Vert pour un prix au m2 < 2000 euros et rouge pour un prix au m2 > 4000 euros.

Vue "Détail services"

On y retrouve :

- Nombre d'établissements scolaires : 17 (collèges et lycées)
- Taux de réussite moyen : 95,06 %
- Équipements publics/loisirs : 376 (dont 203 extérieurs)
- Établissements de santé : 1 069 (catégorisés en 10 types)

Cette vue est dynamique : en cliquant sur un type (équipements ou établissements), les données détaillées apparaissent dans des tableaux interactifs.

The screenshot displays a dynamic data interface for Caen, Normandy. The header shows the location (Normandie, Caen) and key statistics: Population moyenne (108,20K) and Loyer moyen du m2 (11,92 €). The navigation bar includes 'Cadre de vie' and 'Détail services*'. The 'Infos services*' table shows: Nombre total d'établissements scolaires (17,00), Taux de réussite moyen (95,06), Nombre total d'équipements publics (376,00), and Nombre d'établissements de santé (1 069,00). The 'Etablissements de santé' table lists 10 categories with a total of 1069. The 'Equipements publics / Loisirs' table lists 5 categories with a total of 376. The 'Collèges' table lists 6 schools with their 2023 success rates, and the 'Lycées' table lists 8 lycées with their 2023 success rates.

nom_commune	Région	Commune	Population moyenne	Loyer moyen du m2
Caen	Normandie	Caen	108,20K	11,92 €

Infos services*		Valeurs	
🏫	Nombre total d'établissements scolaires	17,00	
📊	Taux de réussite moyen	95,06	
🏠	Nombre total d'équipements publics	376,00	
🏥	Nombre d'établissements de santé	1 069,00	

Etablissements de santé		Equipements publics / Loisirs	
	Nb		Nb
Autres activités médicales	10	Piste d'Athlétisme	3
Centre de ressources	8	Stade d'athlétisme du lycée Malherbe	1
Établissements de formation	3	📦 Découvert	203
Exercice libéral & laboratoires	112	📦 Extérieur couvert	1
Famille, enfance & justice	486	📦 Intérieur	162
Handicap & éducation spécialisée	168	📦 Site artificiel	2
Hébergement & structures collectives	196	📦 Site naturel aménagé	5
Hôpitaux & Centres de soins	7	Total	376
Insertion & social	10		
Lutte contre les maladies	56		
Soins & services à domicile	13		
Total	1069		

Collèges		Taux de réussite en 2023	
COLLEGE GUILLAUME DE NORMANDIE			85
COLLEGE FERNAND LECHANTEUR			92
COLLEGE HENRI BRUNET			93
COLLEGE VILLEY DESMESERETS			94
COLLEGE HASTINGS			97
COLLEGE PRIVE INSTITUT SAINT PIERRE			99

Lycées		Taux de réussite en 2023	
INSTITUT LEMONNIER			94
LYCEE AUGUSTIN FRESNEL			91
LYCEE CHARLES DE GAULLE			96
LYCEE DUMONT D'URVILLE -LAPLACE (GENERAL ET TECHNO.)			93
LYCEE JEAN ROSTAND			90
LYCEE JEANNE D'ARC			100
LYCEE MALHERBE			97
LYCEE NOTRE DAME DE FIDELITE			100
LYCEE SAINTE MARIE			100
LYCEE SAINTE URSULE			98
LYCEE VICTOR HUGO			97

6. Analyse des résultats – Focus sur Caen

Accessibilité & Services

Caen se distingue par un niveau de service élevé :

- 17 établissements scolaires avec un taux de réussite moyen élevé (95,06%), dont plusieurs lycées atteignent 100 % de réussite (ex : Lycée Jeanne d'Arc, Lycée Notre-Dame de Fidélité).
- Plus de 1 000 structures de santé, ce qui reflète une bonne couverture médico-sociale, avec un équilibre entre structures hospitalières, libérales et sociales.

- 376 équipements publics/loisirs, dont 203 extérieurs (pistes, stades, parcs...).

Point de vigilance : Ces données sont malheureusement à prendre avec du recul en raison de la fiabilité des données qui en ressort. La base recensant tous les établissements de santé était très lourde et difficile à traiter de manière optimale, nous l'avons tout de même gardée pour ajouter une dimension essentielle à nos yeux dans le choix.

Marché immobilier

Le loyer moyen de 11,92 €/m² reste raisonnable pour une grande ville régionale.

Évolution chiffrée :

- +1,32 € en 6 ans, soit une hausse de 12,45 % entre 2018 et 2024.
- Une croissance stable et continue, sans baisse sur la période observée.

Ce rythme de progression, sans être excessif, traduit une pression locative croissante, reflet de l'attractivité de la commune. Cette évolution est à mettre en relation avec le développement de l'agglomération Caen la Mer et les projets d'aménagement récents (quartier Presqu'île, mobilité, université...).

Qualité de vie

- La criminalité à 4,5% est contenue pour une ville de cette taille. comparaison
- Notre tableau de bord indique une qualité de l'air de niveau "Moyen" à Caen, sans dépassement des seuils critiques. Toutefois, ce niveau ne garantit pas une qualité optimale, notamment pour les populations sensibles.

La communauté urbaine Caen la Mer a donc lancé plusieurs projets structurants pour réduire les émissions polluantes, en lien avec la Zone à Faibles Émissions mobilité (ZFE-m) :

- Entrée en vigueur de la ZFE : 1er janvier 2025.
- Objectif : restreindre l'accès aux véhicules polluants dans un périmètre couvrant 100 000 habitants, soit la moitié de l'agglomération.
- Mesures envisagées : extension du tramway, développement des pistes cyclables, conversion progressive de la flotte de bus.

Par ailleurs, la ville prévoit une végétalisation accrue de l'espace urbain pour lutter contre les îlots de chaleur et améliorer la filtration de l'air ainsi qu'un plan d'action lié au développement durable adopté dès 2020, intégrant la qualité de l'air comme enjeu prioritaire.

Sources : Rapport ZFE – caenlamer.fr / Rapport Développement Durable 2020 – caen.fr

La connectivité est excellente, avec des débits élevés en 4G et fibre. Pour information, selon la boutique Orange, voici les débits requis en fonction des différentes utilisations.

- Télétravail : 10 Mbit/s pour des usages classiques,
- 500 Mbit/s pour des multi-usages sans latence,
- Visioconférence : entre 1,2 et 4 Mbit/s pour des échanges vidéo en HD,
- Jeux vidéo en ligne : entre 5 et 15 Mbit/s en fonction de la plateforme.

7. Limites et axes d'amélioration

1. Complexité du traitement des bases de données

Nous avons manipulé un grand nombre de bases de données, issues de sources variées, aux formats et structures très différentes. Certaines bases étaient incomplètes, d'autres obsolètes, ce qui a compliqué leur exploitation. L'extraction, le tri, et le nettoyage des données se sont révélés particulièrement fastidieux, et malgré nos efforts, nous avons parfois dû renoncer à certaines informations, faute de qualité suffisante.

Perspective d'amélioration : Mettre en place un pipeline automatisé de traitement de données, via Power Query avancé ou des scripts Python, pour faciliter l'import, la standardisation et le nettoyage systématique des bases avant intégration dans Power BI.

2. Incohérences dans les noms de communes

Nous avons rapidement constaté une hétérogénéité dans les noms des communes, liée à des différences d'écriture (accents, majuscules), à des fusions administratives récentes ou à des critères d'identification différents selon les sources. Cela a constitué un frein majeur à l'appariement des données, et a parfois conduit à l'absence d'informations pour certaines villes, alors même que ces données existaient ailleurs.

Perspective d'amélioration : Utiliser une API de géocodage INSEE ou une table de correspondance normalisée pour associer chaque enregistrement à un code unique, garantissant l'uniformité des jointures entre toutes les bases.

3. Difficulté à relier toutes les bases entre elles

Chaque base que nous avons utilisée avait sa propre logique (nomenclature, granularité, niveau géographique). Relier toutes ces sources a donc été un travail délicat, souvent manuel. Dans certains cas, même après nettoyage, des données restaient manquantes faute de clés de correspondance fiables. Cela montre que l'universalité des résultats affichés doit être relativisée.

Perspective d'amélioration : Structurer le modèle de données autour d'un schéma relationnel robuste, centré sur le code INSEE, et créer une table référentielle de communes pour forcer les correspondances entre bases hétérogènes.

4. Formats non compatibles avec une actualisation dynamique

Certaines bases étaient proposées dans des formats inexploitablement directement dans Power BI (.parquet, .html, .json). Cela nous a contraints à les télécharger manuellement, ce qui limite les possibilités de mise à jour automatique du rapport.

Perspective d'amélioration : Développer des scripts d'import/export automatisés (ex : en Python) pour convertir ces formats vers du `.csv` ou `.xlsx`, et ainsi rendre les bases connectables en direct via Power BI.

5. Absence de pondération par la population

Nos indicateurs sont exprimés en valeur brute (nombre d'équipements, d'écoles, etc.), sans les rapporter au nombre d'habitants. Cela peut biaiser les comparaisons entre communes, notamment entre petites et grandes villes.

Perspective d'amélioration : Intégrer une pondération par la population via des ratios (ex : équipements publics pour 1 000 habitants, médecins pour 10 000 habitants), pour une analyse plus fine et plus juste du niveau de service réel.

6. Difficulté à intégrer des notions d'accessibilité

Nous souhaitons mesurer l'accessibilité réelle aux services, comme la distance jusqu'à une gare, un hôpital ou un centre-ville. Mais cette analyse aurait nécessité des calculs géographiques complexes, basés sur des coordonnées GPS ou des outils de géolocalisation que nous ne maîtrisons pas entièrement.

Perspective d'amélioration : Croiser les données actuelles avec des bases géolocalisées (comme celles de la SNCF ou du ministère de la Santé), et utiliser des fonctions de distance spatiale (via QGIS ou un service d'API géographique) pour intégrer la notion de proximité géographique.

7. Données parfois anciennes ou figées

Plusieurs données exploitées dans notre tableau de bord ne sont pas actualisées régulièrement. Cela peut altérer la fiabilité de l'analyse, surtout dans des communes en pleine mutation ou dans des secteurs comme l'emploi, la pollution ou le logement.

Perspective d'amélioration : Connecter Power BI à des API en temps réel (INSEE, Airparif, etc.), afin de rafraîchir automatiquement les indicateurs les plus sensibles à l'évolution temporelle.

8. La carte interactive

Le point dur majeur de notre BI. En raison de la quantité et de l'hétérogénéité des informations que nous avons liées. La connexion d'une carte interactive sur notre tableau de bord nous a fait perdre "12%" de données.

Le nom de la commune n'est pas toujours localisé sur la carte c'est la raison pour laquelle nous avons utilisé le code postale.

9. Création d'un site

Afin de rendre l'analyse davantage professionnelle, nous avons créé un site se nommant [ATEMA-GPLA](#).

Le power BI étant trop lourd, nous ne pouvions pas connecter le BI en l'état, du moins pas

sans passer sur une version payante. Nous avons donc sur le site une proposition d'analyse à la maille départementale et régionale qui vient compléter et apporter un visuel plus agréable pour l'utilisateur.

La création du site nous a permis de nous projeter dans un véritable cadre professionnel. L'intégration des rapports Power BI se fait via une API directement connectée aux bases de données et à notre Power Query, permettant ainsi aux visuels de s'actualiser automatiquement (toutes les 24 heures dans le cas présent). Cela permet d'avoir toujours le visuel à jour.

L'ensemble du site internet a été codé en HTML, CSS, PHP, JavaScript et Python pour l'implantation des visualisations Power BI.

8. Conclusion & Mise en perspective

Ce projet nous a permis de mobiliser l'ensemble des compétences abordées en cours : connexion de sources hétérogènes, nettoyage et structuration via Power Query, création de mesures DAX, et construction d'un modèle analytique avec Power BI.

Nous avons pris conscience des freins liés à la donnée publique : qualité variable, complexité de croisement, actualisation difficile. Ces limites n'ont pas entamé la valeur de l'outil, mais nous amènent à relativiser certaines conclusions, et à envisager ce tableau comme une base perfectible.

Dans une phase ultérieure, nous pourrions :

- Ajouter des ratios de population pour objectiver les indicateurs,
- Connecter des API temps réel pour garantir des données toujours à jour,
- Intégrer la notion de distance géographique (aux soins, aux transports...),
- Et étendre l'analyse à d'autres cas d'usage : politiques publiques, attractivité territoriale, accompagnement à l'installation, etc.

Ce travail a aussi été utile à titre personnel. En tant qu'alternants et futurs diplômés, nous nous posons des questions concrètes sur où vivre, où investir, et comment arbitrer entre coût et qualité de vie. Ce tableau nous offre une grille de lecture pour mieux comprendre les territoires.

Annexe – Sources des données et contexte de collecte Power BI et site ATEMA.

1. Données sur les loyers

- Carte des loyers – Indicateurs de loyers d’annonce par commune
 - 2018 : <https://www.data.gouv.fr/fr/datasets/carte-des-loyers-indicateurs-de-loyers-dannonce-par-commune-en-2018/>
 - 2022 : <https://www.data.gouv.fr/fr/datasets/carte-des-loyers-indicateurs-de-loyers-dannonce-par-commune-en-2022/>
 - 2023 : <https://www.data.gouv.fr/fr/datasets/carte-des-loyers-indicateurs-de-loyers-dannonce-par-commune-en-2023/>
 - 2024 : <https://www.data.gouv.fr/fr/datasets/carte-des-loyers-indicateurs-de-loyers-dannonce-par-commune-en-2024/>

Contexte :

La "carte des loyers" est une initiative publique lancée en 2018 par la DGALN, en collaboration avec Agrosup Dijon, INRAE, SeLogger et Leboncoin. Depuis 2020, elle est pilotée par l’ANIL. Elle vise à fournir des indicateurs fiables sur les loyers à l’échelle communale à partir d’annonces immobilières (logements vides, non meublés, charges comprises).

2. Données sur les infrastructures publiques et le cadre de vie

- Recensement des équipements sportifs et de loisirs <https://www.data.gouv.fr/fr/datasets/recensement-des-equipements-sportifs-espaces-et-sites-de-pratiques/#/resource>
 - Fichier FINESS (établissements de santé et sociaux) <https://www.data.gouv.fr/fr/datasets/finess-extraction-du-fichier-des-etablissements/>
 - Liste des gares SNCF <https://data.sncf.com/explore/dataset/liste-des-gares/export/>
 - Fichier des départements et régions
Créés manuellement à l’aide de ChatGPT pour faciliter la hiérarchisation géographique (région > département > commune).
-

3. Données socio-économiques

- Revenus, pauvreté et niveau de vie (INSEE, niveau IRIS) <https://www.insee.fr/fr/statistiques/5055909>
- Population municipale des communes (INSEE) <https://www.data.gouv.fr/fr/datasets/population-municipale-des-communes-france-entiere/>
- Géolocalisation des entreprises et secteurs d’activités (INSEE) <https://www.data.gouv.fr/fr/organizations/institut-national-de-la-statistique-et-des-etudes-economiques-insee/>

Contexte :

Ces données sont issues du recensement officiel de la population et sont mises à jour chaque

année par l'INSEE. Elles comprennent la population résidente habituelle, y compris les personnes sans abri et celles en hébergement collectif.

4. Données sur la connectivité internet

- Ma connexion Internet – ARCEP
<https://www.data.gouv.fr/fr/datasets/ma-connexion-internet/>

Contexte :
Il s'agit d'un projet de l'ARCEP permettant d'évaluer l'éligibilité aux réseaux internet (4G, fibre, ADSL) à l'échelle communale. Nous avons utilisé le fichier « tableau de bord communes » mis à disposition.

5. Données sur la qualité de l'air

- Indice de la qualité de l'air (ATMO)
<https://www.data.gouv.fr/fr/datasets/indice-de-la-qualite-de-lair-quotidien-par-commune-indice-atmo/>

Contexte :
L'indice ATMO est un indicateur synthétique mesurant la pollution de l'air au quotidien par commune. Il est catégorisé en "faible", "moyen", ou "élevé".

6. Données sur la sécurité

- Base statistique de la délinquance (police et gendarmerie nationales)
<https://www.data.gouv.fr/fr/datasets/bases-statistiques-communale-departementale-et-regionale-de-la-delinquance-enregistree-par-la-police-et-la-gendarmerie-nationales/>

Contexte :
Ces bases contiennent les infractions constatées dans le cadre de l'activité judiciaire des forces de l'ordre, catégorisées et géolocalisées à l'échelle communale.

7. Données sur les établissements scolaires

- Indices de position sociale des lycées (2016–2021)
<https://www.data.gouv.fr/fr/datasets/indices-de-position-sociale-dans-les-lycees-de-france-metropolitaine-et-drom/>
- Indicateurs de valeur ajoutée des collèges
<https://www.data.gouv.fr/fr/datasets/indicateurs-de-valeur-ajoutee-des-colleges/>

Contexte :

Ces fichiers listent les établissements du secondaire avec leur taux de réussite, leur position sociale moyenne, et d'autres indicateurs contextuels utiles pour évaluer la qualité de l'offre éducative locale.