

06 juillet 2025

# MODULE 6 : Projet final de formation

Analyse du parc immobilier du département du Territoire  
de Belfort (90)

# SOMMAIRE

<b>1. INTRODUCTION ET OBJECTIF DE L'ETUDE .....</b>	<b>2</b>
A. Présentation du sujet .....	2
B. Objectif de l'étude .....	2
<b>2. DONNEES ET PREPARATION .....</b>	<b>3</b>
A. Présentation du jeu de données .....	3
B. Analyse exploratoire et préparation des données .....	3
a. Analyse du fichier « documentation » .....	3
b. Analyse exploratoire des données .....	7
<b>3. ANALYSE ET RESULTATS .....</b>	<b>19</b>
a. Formes urbaines et occupations du Territoire de Belfort .....	19
b. Performances énergétiques .....	21
c. Valeurs financières.....	24
<b>4. CONCLUSION .....</b>	<b>26</b>
<b>5. ANNEXES .....</b>	<b>27</b>
A. Jeu de données .....	27
B. Détail du contenu des tables retenues pour l'étude .....	27
C. Notebook Python .....	27
D. Script MySQL et résultats des requêtes .....	27
E. Fichier PowerBI et tableaux de bord.....	27

# 1. INTRODUCTION ET OBJECTIF DE L'ETUDE

## A. Présentation du sujet

La présente étude porte sur l'analyse du parc immobilier du département du Territoire de Belfort (90).

L'enjeu pour le département est de disposer d'une meilleure vision de son territoire pour orienter ses futures actions en termes de développement.



## B. Objectif de l'étude

Sur la base des données à sa disposition, **le département souhaiterait consolider la connaissance de son territoire notamment sur les thématiques suivantes :**

- Les formes urbaines et la nature des occupations.
- Les performances énergétiques de son parc de logement.
- Les tendances de valeurs financières de son patrimoine immobilier.

## 2. DONNEES ET PREPARATION

### A. Présentation du jeu de données

Le jeu de données sur laquelle la présente étude repose est issue de la Base de Données Nationale des Bâtiments (BNDB), établie et gérée par le Centre Scientifique Technique des Bâtiments (CSTB), un établissement public français.

La BNDB est une agrégation de données issues de plusieurs bases de données fournissant différentes informations sur près de 32 millions de bâtiments français, telles que la forme des bâtiments (surface, hauteur, etc.), leurs usages (résidentiel, tertiaire, etc.), leurs performances énergétiques, etc. Ces informations sont en partie en libre-accès et sont segmentées par département.

Le jeu de données s'organise en deux principaux dossiers :

- Un dossier « csv » qui contient plusieurs fichiers .csv qui regroupent les différentes informations des bâtiments du département.
- Un dossier « doc » qui contient notamment un fichier Excel « documentation » qui présente le contenu des différents fichiers du dossier « csv ».

### B. Analyse exploratoire et préparation des données

#### a. Analyse du fichier « documentation »

**Le point de départ de cette analyse exploratoire porte sur celle du fichier « documentation » afin de mieux cerner les informations disponibles sur les bâtiments du département.** Ce fichier est organisé en plusieurs onglets :

- Onglet « source », qui présente les différentes sources à partir desquelles la BDNB est construite. Seules les sources en open data sont en libre-accès au travers de la BDNB.
- Onglet « modèle détaillé », qui présente sommairement les différents fichiers (ou tables) du dossier « csv ». Ces tables sont regroupées selon 8 volets : adresse, bâtiment, bâtiment\_construction, bâtiment\_groupe, iris, parcelle, propriétaire et tables\_sources.
- Un onglet par volet, qui présente par table son contenu détaillé (ci-dessous un extrait de l'onglet « bâtiment\_groupe » concernant le contenu de la table « bâtiment\_groupe\_bdtopo\_bat »).

batiment_groupe_bdtopo_bat		Informations de la BDTopo, couche bâti, agrégées à l'échelle du bâtiment			
Nom de la colonne	Base source	Type	Définition	Niveau d'accès	Modalités
batiment_groupe_id	bdtopo	varchar(22)	Identifiant du groupe de bâtiment au sens de la BDNB	BDNB OPEN	
code_departement_insee	bdtopo	varchar(2)	Code département INSEE	BDNB OPEN	[01', 02', 03', 04', 05', 06', 07', 08', 09', 10', 11', 12', 13', 14', 15', 16', 17', 18', 19', 21', 22', 23', 24', 25', 26', 27', 28', 29', 2A', 2B', 30', 31', 32', 33', 34', 35', 36', 37', 38', 39', 40', 41', 42', 43', 44', 45', 46', 47', 48', 49', 50', 51', 52', 53', 54', 55', 56', 57', 58', 59', 60', 61', 62', 63', 64', 65', 66', 67', 68', 69', 70', 71', 72', 73', 74', 75', 76', 77', 78', 79', 80', 81', 82', 83', 84', 85', 86', 87', 88', 89', 90', 91', 92', 93', 94', 95']
l_nature	bdtopo	text[]	(ign) Catégorie de nature du bâtiment	BDNB OPEN	
l_usage_1	bdtopo	text[]	(ign) Usage principal du bâtiment	BDNB OPEN	
l_usage_2	bdtopo	text[]	(ign) Usage secondaire du bâtiment	BDNB OPEN	
l_etat	bdtopo	text[]	(ign) Etat des bâtiments	BDNB OPEN	
hauteur_mean	bdtopo	int4	(ign) Hauteur moyenne des bâtiments [m]	BDNB OPEN	
max_hauteur	bdtopo	int4	(ign) Hauteur maximale des bâtiments [m]	BDNB OPEN	
altitude_sol_mean	bdtopo	int4	(ign) Altitude au sol moyenne [m]	BDNB OPEN	

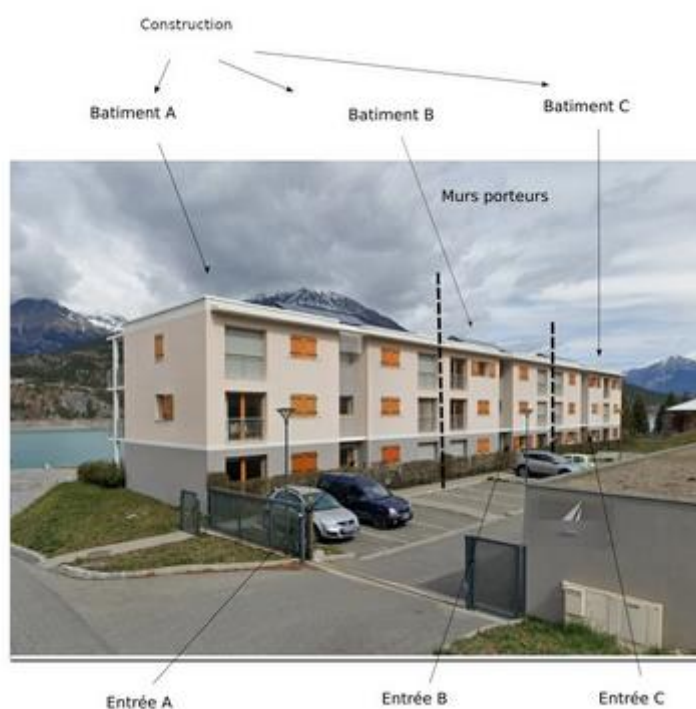
Extrait de l'onglet « batiment\_groupe » portant sur le contenu de la table « batiment\_groupe\_bdtopo\_bat » (source : fichier « documentation » de la BDNB)

Aussi, ce fichier précise via un code couleur la disponibilité des données (remplissage blanc = données en libre-accès / remplissage gris foncé = données en accès restreint). La présente étude porte uniquement sur les données en libre-accès.

**Concernant la distinction entre les volets bâtiment, bâtiment\_construction et bâtiment\_groupe, celle-ci correspond à des logiques de périmètre et est définie ainsi** (source : site internet de la BDNB) :

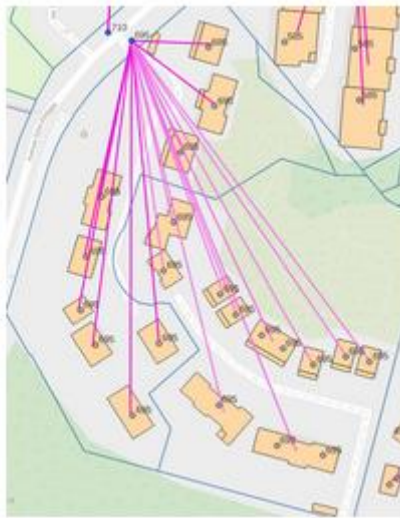
« Le cœur de la BDNB est construit autour de trois niveaux de bâtiments :

- La construction, ensemble constructif cohérent, potentiellement constitué de plusieurs bâtiments séparés par des murs porteurs et desservis chacun par une entrée.
- Le bâtiment, aligné avec les bâtiments fiscaux, assimilable à l'ensemble des logements accessibles depuis une entrée / cage d'escalier.



- Le groupe de bâtiment, regroupement de l'ensemble des constructions d'une parcelle lorsque leur identification n'a pas été possible (bâtiments trop similaires par exemple) avec les sources de données actuellement disponibles.

En effet, les cas aboutissant parfois à des situations complexes rendant actuellement l'identification individuelle impossible sont nombreux. Nous rebasculons alors les données à une échelle agrégée.



*Dans le cas de maisons individuelles, ces trois objets sont représentés par le même contour géométrique. »*

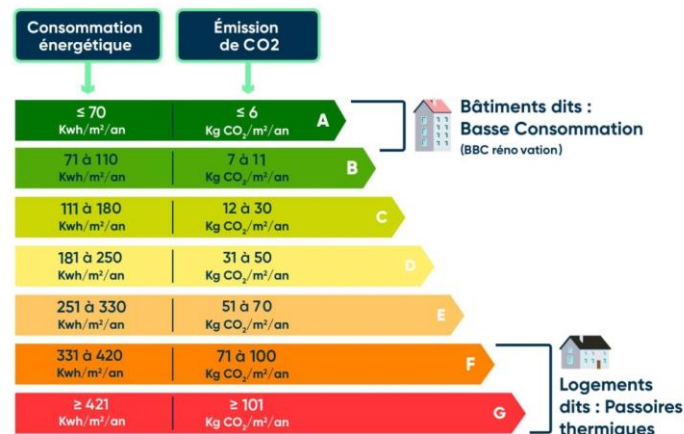
Il faut savoir que les données relatives au périmètre bâtiment ne sont pas disponibles en libre-accès, et que celles relatives au périmètre bâtiment\_construction sont assez limitées.

**L'analyse de ce document permet donc de déterminer les tables a priori pertinentes en vue de l'étude. Il s'agit des tables 15 suivantes (sur 48 disponibles) dont le contenu est présenté sommairement ici** (le contenu détaillé de ces tables est présenté en annexe) :

- Table « adresse », qui fournit des informations relatives aux adresses des bâtiments.
- Table « rel\_batiment\_groupe\_adresse », qui est une table de relation entre les adresses et les groupes de bâtiments.
- Table « batiment\_groupe », qui fournit des informations relatives à l'identification des groupes de bâtiments ainsi que sur leur surface au sol.

- Table « batiment\_groupe\_bdtopo\_bat », qui fournit des informations relatives à la nature et à la morphologie des groupes de bâtiments (usages, hauteur, altitude par rapport au sol, etc.).
- Table « batiment\_groupe\_bdtopo\_equ », qui fournit des précisions concernant la nature des groupes de bâtiments lorsqu'il s'agit d'équipements spécifiques (exemple : borne de rechargement électrique pour véhicules).
- Table « batiment\_groupe\_bdtopo\_zoac », qui fournit des précisions concernant l'affiliation des groupes de bâtiments à une zone d'activités (exemple : bâtiment situé en zone industrielle).
- Table « batiment\_groupe\_dle\_reseaux\_multimillesime », qui fournit des informations relatives aux consommations des groupes de bâtiments lorsque celui-ci est raccordé à un réseau de chaleur (qui est un système de chauffage à une échelle urbaine et non à l'échelle d'un simple bâtiment ou logement).
- Table « batiment\_groupe\_dpe\_representatif\_logement », qui contient les Diagnostics de Performance Énergétique (DPE) représentatifs de chaque groupe de bâtiment lorsqu'il s'agit de logement (les DPE étant uniquement disponibles pour les logements). Ceci veut dire que pour un groupe de bâtiment, il est retenu et présenté un DPE réalisé sur un logement qui est estimé comme représentatif de tous les autres. Si le groupe de bâtiment est une maison individuelle, alors le DPE représentatif correspond au DPE de cette maison.
- Table « batiment\_groupe\_dpe\_statistique\_logement », qui fournit le nombre de DPE par étiquette (cf. exemple d'étiquette ci-après) sur les groupes de bâtiments de logement.
- Table « batiment\_groupe\_dvf\_open\_statistique », qui fournit des données statistiques relatives aux mutations (une mutation étant une transaction immobilière) que les groupes de bâtiments ont connues selon une période de référence. Cette table fournit donc des informations sur le nombre de mutations sur cette période de référence, sur la valeur foncière, sur le prix au m<sup>2</sup>, etc.
- Table « batiment\_groupe\_indicateur\_reseau\_chaud\_froid », qui fournit des informations relatives au raccordement des groupes de bâtiments à un réseau urbain de chaleur et/ou de froid.
- Table « batiment\_groupe\_qpv », qui fournit des précisions concernant l'affiliation des groupes de bâtiments à un Quartier Prioritaire de la Ville (QPV) (exemple : bâtiment situé en zone industrielle). Les quartiers prioritaires de la politique de la ville sont des territoires d'intervention de l'État et des collectivités territoriales définis par la loi de programmation pour la ville et la cohésion urbaine du 21 février 2014 dans l'objectif commun de réduire les écarts de développement entre les quartiers défavorisés et leurs unités urbaines (source : INSEE).
- Table « batiment\_groupe\_synthese\_propriete\_usage », qui est une table de synthèse sur l'usage des bâtiments (résidentiel, tertiaire, etc.).
- Table « parcelle », qui fournit des informations relatives à l'identification des parcelles cadastrales ainsi que sur leur surface.

- Table « rel\_batiment\_groupe\_parcelle », qui est une table de relation entre les groupes de bâtiments et les parcelles.



Exemple d'étiquette relative à un DPE

## b. Analyse exploratoire des données

### i. Exploration préliminaire des données sous Python

La première étape de l'analyse exploratoire des données est réalisée sous Python. Le Notebook correspondant est joint en annexe du présent rapport et présente en détail la manière dont les données ont été explorées. Il est présenté ici de manière synthétique cette analyse exploratoire sous Python.

Tout d'abord, après chargement des bibliothèques nécessaires à l'exploration des données (notamment Pandas), les tables pertinentes pour l'étude sont importées.

**Il est ensuite procédé à une visualisation des données, qui consiste à afficher pour chaque table un extrait de celle-ci et ses informations générales** (nombre de lignes, nombre de colonnes, noms des colonnes, nombre de valeurs manquantes, format des données). **De cette visualisation, il en ressort les principaux éléments suivants :**

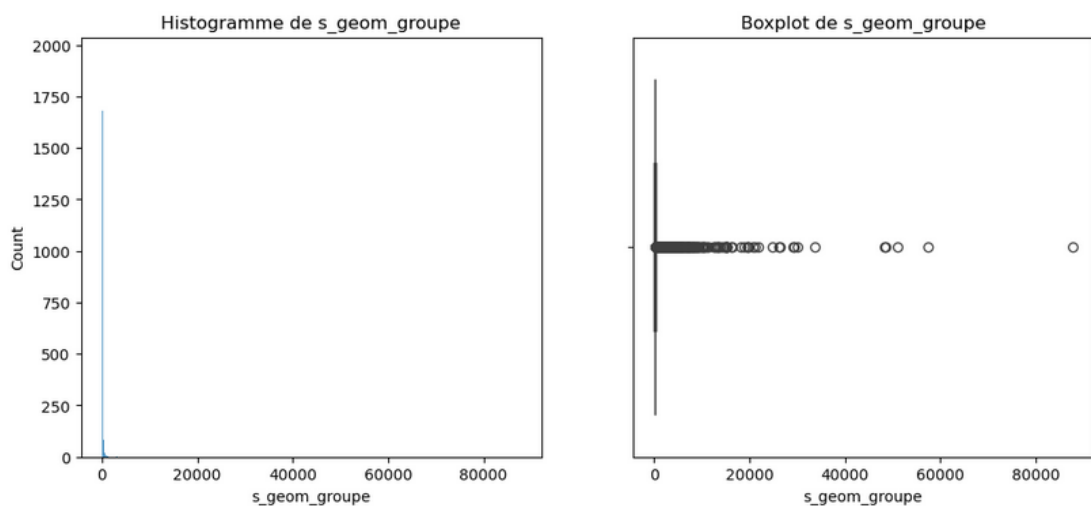
- Table « batiment\_groupe\_batopo\_bat » (nature et morphologie des groupes de bâtiments) : des valeurs manquantes sont relevées pour les colonnes hauteur\_mean (hauteur moyenne) et max\_hauteur (hauteur maximum) – 4 760 valeurs manquantes pour chacune de ces colonnes, ainsi que pour la colonne altitude\_sol\_mean (altitude moyenne par rapport au sol) – 553 valeurs manquantes pour cette colonne. Cependant, le nombre de valeurs manquantes reste relativement faible par rapport au nombre total d'enregistrements (54 356), ce qui permet d'assurer que les données disponibles restent suffisamment représentatives du territoire.



- Table « batiment\_groupe\_dle\_reseaux\_multimillesime » (consommations des groupes de bâtiments raccordés à un réseau de chaleur) : la table est vide, elle ne comporte aucun enregistrement. Ce qui pourrait vouloir dire qu'aucun bâtiment du département n'est raccordé à un réseau de chaleur (peu probable) ou que les données ne sont pas suffisamment consolidées. Ceci est plus probable au regard des indications dans la « documentation » pour cette table : *« Attention les données sur les réseaux de chaleurs sont aujourd'hui bien moins fiables que les vecteurs gaz ou électricité. »*.
- Table « batiment\_groupe\_dpe\_representatif\_logement » (DPE représentatifs de chaque groupe de bâtiment de logement) : ce fichier contient de très nombreuses informations (plus de 10 000 lignes et 106 colonnes), puisqu'il fournit le détail des informations du DPE représentatif pour chaque groupe de bâtiments de logements (classe énergétique, caractéristiques techniques du bâtiment et de ses équipements, etc.). Il semble donc normal que de nombreuses valeurs soient manquantes en fonction du bâtiment et de ses composants, de la manière dont le DPE a été réalisé, de sa date de réalisation (la réglementation ayant évolué au cours du temps), etc. Il est difficile de traiter ces valeurs manquantes au regard de leur nombre (plus de 5 000 valeurs manquantes sur environ 10 000 pour certaines colonnes) et de leur hétérogénéité (exemple : 4 470 occurrences différentes pour une des variables du jeu de données). En effet, supprimer les enregistrements comportant des valeurs manquantes ou remplacer ces valeurs manquantes par des valeurs de tendance (moyenne, médiane) risqueraient de fortement altérer la représentativité initiale des données et donc cela biaiserait les résultats. Il convient donc de travailler en présence de ces valeurs manquantes pour cette table.
- Table « batiment\_groupe\_dvf\_open\_statistique » (valeurs foncières et prix au m<sup>2</sup>) : des valeurs manquantes sont relevées principalement pour les colonnes relatives aux données sur le prix au m<sup>2</sup> (plus de 3 000 valeurs manquantes pour certaines de ces colonnes sur la base de 11 486 enregistrements). Les données manquantes pourront peut-être être remplacées par des valeurs de tendance sous réserve de la distribution des données.
- Selon les tables, d'autres valeurs manquantes sont ponctuellement relevées, mais restent non problématiques vis-à-vis de la portée de l'étude.
- Pour les tables relatives au volet groupes de bâtiments (dont chaque groupe de bâtiments est identifié par un ID unique), le nombre d'enregistrements varie d'un fichier à un autre. Ceci voudrait donc dire que le niveau d'informations par groupe de bâtiments n'est pas équivalent. Certains groupes de bâtiments disposent d'informations spécifiques/complémentaires par rapport à d'autres.

Ensuite, **il est procédé à une analyse des statistiques descriptives, et à une visualisation graphique de la distribution des données** le cas échéant, pour certaines tables en fonction de la pertinence de leurs informations. Il en ressort les principaux résultats suivants :

- Table « `batiment_groupe` » (identification des groupes de bâtiments et de leur surface au sol) : la valeur maximale de la surface au sol de la géométrie du bâtiment groupe s'élève à 87 689 m<sup>2</sup> et semble être une valeur aberrante au regard du reste de la distribution (les valeurs moyenne et médiane s'élevant respectivement à 191 et 110 m<sup>2</sup>). La visualisation de la distribution (cf. graphiques ci-dessous : histogramme à gauche et boîte à moustache à droite) montre bien une concentration des surfaces des groupes de bâtiments vers des valeurs faibles (qu'on pourrait qualifier de normales car attendues) et la présence de valeurs extrêmes qui pourraient correspondre à de grands ensembles immobiliers (centre commercial, plateforme logistique, etc.). Cette hypothèse pourra être vérifiée en recoupant le ou les enregistrements correspondants à ces valeurs aberrantes avec les informations contenues dans d'autres tables, comme celles relatives aux adresses et aux usages des groupes de bâtiments par exemple.

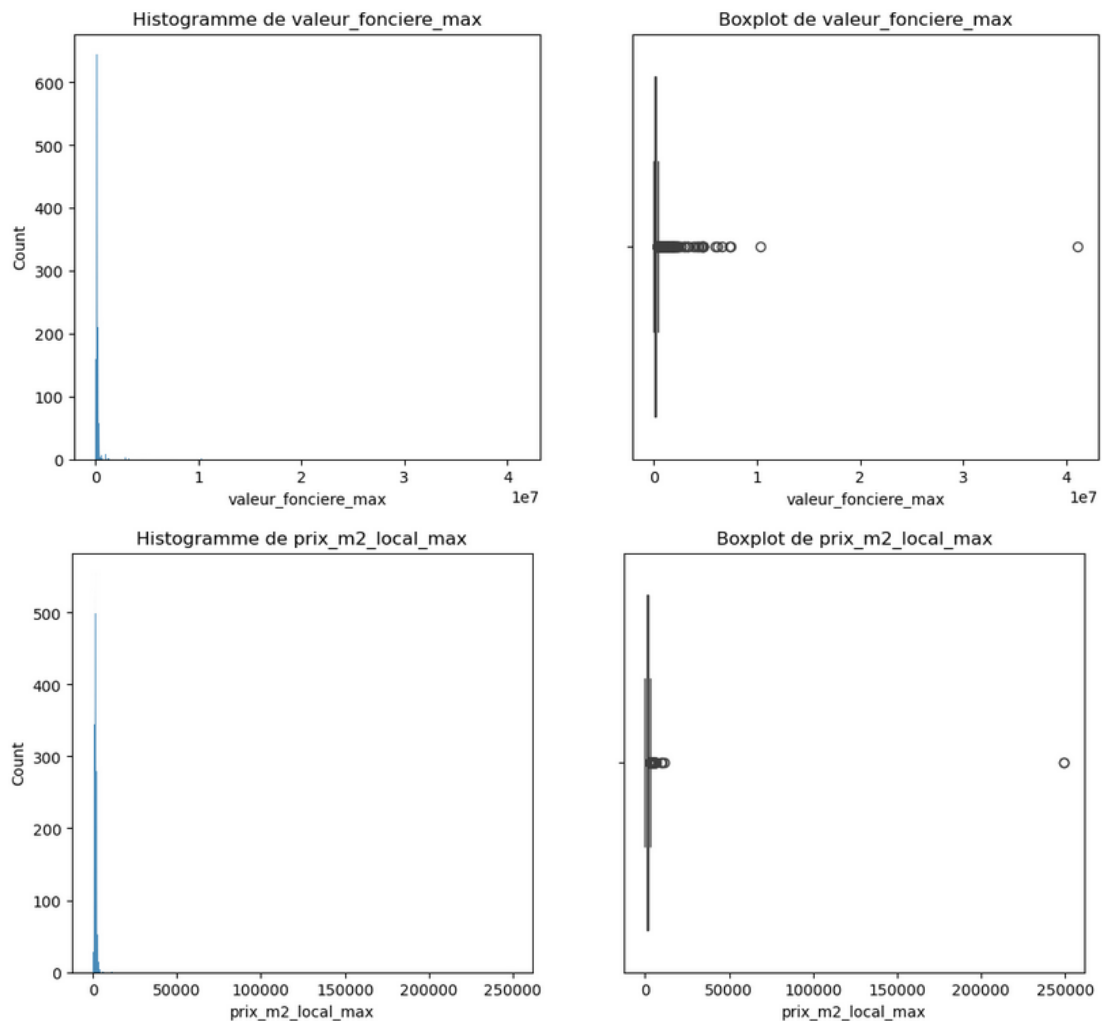


- Table « `batiment_groupe_bdtopo_bat` » (nature et morphologie des groupes de bâtiments) : la présence de valeurs nulles comme valeurs minimales pour les hauteurs questionne la nature des groupes de bâtiments correspondants. Cette information devra être recoupée avec les informations contenues dans d'autres tables, comme celles relatives aux adresses et aux usages des groupes de bâtiments par exemple, pour comprendre leur nature. Les autres valeurs ne présentent pas d'anomalie particulière.
- Table « `batiment_groupe_dpe_statistique_logement` » (nombre de DPE par étiquette pour les groupes de bâtiments de logement) : les valeurs de tendance centrale et les valeurs des quartiles pour les différentes variables sont souvent égales à 0 (cf. extrait de tableau ci-dessous). Le nombre de catégories d'étiquettes (classe a, classe b, etc.) s'élevant à 15 et le nombre d'étiquettes par groupe de bâtiment pouvant être faible (exemple : 1 seule étiquette dans le cas d'une maison individuelle), ceci tire les valeurs de tendances (moyenne, quartiles, etc.) vers des valeurs nulles. Ces valeurs a priori aberrantes ne le sont donc pas en réalité.

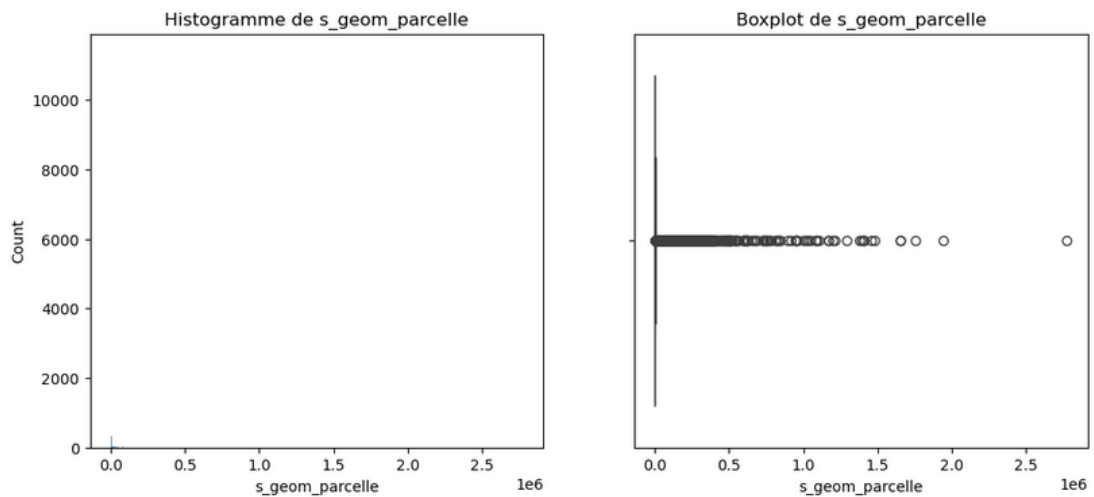
	nb_classe_bilan_dpe_a	nb_classe_bilan_dpe_b	nb_classe_bilan_dpe_c	nb_classe_bilan_dpe_d	nb_classe_bilan_dpe_e
<b>count</b>	10113.0	10113.0	10113.0	10113.0	10113.0
<b>mean</b>	0.0	0.0	0.0	0.0	0.0
<b>std</b>	0.0	2.0	4.0	3.0	3.0
<b>min</b>	0.0	0.0	0.0	0.0	0.0
<b>25%</b>	0.0	0.0	0.0	0.0	0.0
<b>50%</b>	0.0	0.0	0.0	0.0	0.0
<b>75%</b>	0.0	0.0	0.0	0.0	0.0
<b>max</b>	11.0	208.0	191.0	131.0	232.0

- Table « batiment\_groupe\_dvf\_open\_statistique » (valeurs foncières et prix au m<sup>2</sup>) : en ce qui concerne les données relatives aux valeurs foncières et aux prix au m<sup>2</sup>, il est relevé de grandes disparités dans la distribution de ces données (cf. extrait de tableau ci-dessous), avec des valeurs d'écart-type généralement élevées et des valeurs maximales souvent extrêmes. Certaines valeurs maximales semblent même aberrantes, avec par exemple des prix au m<sup>2</sup> bâti de l'ordre de 200 000 €/m<sup>2</sup>. Les visualisations de la distribution pour les valeurs foncières et les prix au m<sup>2</sup> (cf. graphiques ci-dessous) montrent bien une concentration vers des valeurs faibles (qu'on pourrait qualifier de normales car attendues). Cependant, il est également relevé la présence d'une ou plusieurs valeurs extrêmes selon les variables. Et une valeur particulièrement aberrante est relevée pour chacune des variables de valeur foncière et de prix du local au m<sup>2</sup>. Il conviendra donc de recouper les enregistrements correspondant à ces valeurs aberrantes avec les informations contenues dans d'autres tables, comme celles relatives aux adresses et aux usages des groupes de bâtiments par exemple, pour comprendre leur nature.

	valeur_fonciere_max	valeur_fonciere_min	prix_m2_local_median	prix_m2_local_moyen	prix_m2_local_max
<b>count</b>	11481.0	11481.0	10447.0	10447.0	10447.0
<b>mean</b>	192282.0	166854.0	1539.0	1539.0	1625.0
<b>std</b>	459328.0	432836.0	3106.0	3106.0	3480.0
<b>min</b>	0.0	0.0	0.0	0.0	0.0
<b>25%</b>	110000.0	82950.0	1072.0	1075.0	1153.0
<b>50%</b>	158190.0	143400.0	1462.0	1459.0	1553.0
<b>75%</b>	220000.0	205000.0	1870.0	1867.0	1949.0
<b>max</b>	41076960.0	41076960.0	221675.0	221675.0	249000.0



- Table « parcelle » (identification des parcelles cadastrales et de leur surface) : la valeur maximale de la surface au sol de la géométrie de la parcelle s'élève à 277 ha et semble être une valeur aberrante au regard du reste de la distribution (les valeurs moyenne et médiane s'élevant respectivement à 4 297 m<sup>2</sup> et 964 m<sup>2</sup>). La visualisation de la distribution (cf. graphiques ci-dessous) montre bien une concentration des surfaces des parcelles vers des valeurs faibles et la présence de valeurs extrêmes qui pourraient correspondre à des parcelles particulières (terrains agricoles, bois/forêts, etc.). Cette hypothèse pourra être vérifiée en recoupant le ou les enregistrements correspondants à ces valeurs aberrantes avec les informations contenues dans d'autres tables, comme celles relatives aux adresses et aux usages des groupes de bâtiments par exemple, pour comprendre leur nature.

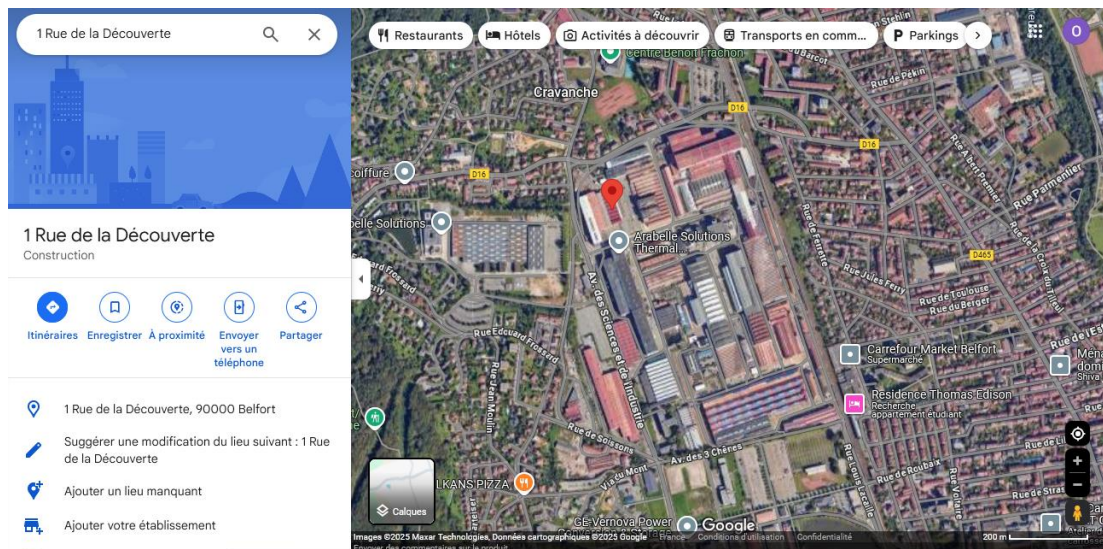


## ii. Vérifications et contrôles sous MySQL

Afin de vérifier les précédentes hypothèses et de contrôler les valeurs aberrantes rencontrées lors de l'analyse exploratoire sous Python, une analyse est ensuite réalisée sous MySQL. Celle-ci a pour but, en chargeant les tables de données nécessaires, de réaliser des jointures entre elles afin de comprendre la nature des groupes de bâtiments qui comportent des valeurs extrêmes et/ou aberrantes. Le script MySQL correspondant ainsi que les résultats des différentes requêtes sont joints en annexe du présent rapport.

**Premièrement, il est procédé à la vérification des valeurs aberrantes pour la variable de surface au sol de la géométrie des groupes de bâtiments :**

- Le 1<sup>er</sup> groupe de bâtiments présentant la plus grande surface au sol (qui correspond également au 2<sup>ème</sup> groupe de bâtiments présentant la plus grande surface – ID identique), avec une surface de 87 689 m<sup>2</sup>, est situé au 1 rue de la Découvert à Belfort (cf. photo aérienne GoogleMaps ci-dessous). Cette zone correspond à la zone d'activités industrielles Techn'Hom et cette adresse aux locaux d'une entreprise de logistique. Cette information est donc cohérente avec la surface de la géométrie du bâtiment groupe.



- Les groupes de bâtiments qui présentent la plus grande surface au sol suivants font également partis de la zone d'activités Techn'Hom.

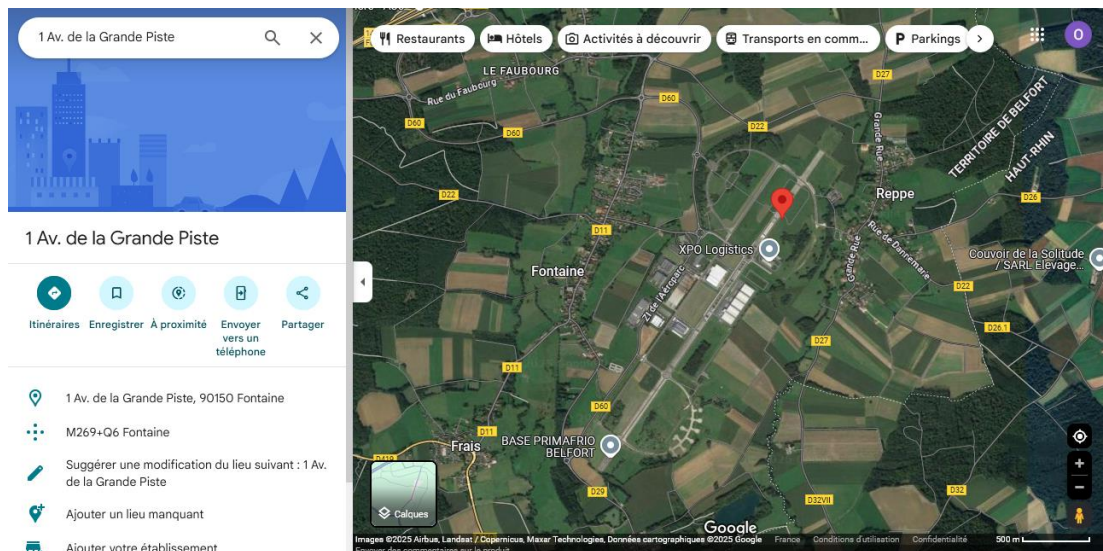
**Il est donc possible d'en conclure que les valeurs extrêmes relevées correspondent bien à de vastes ensembles immobiliers du type bâtiment industriel/tertiaire de parcs d'activités, centre-commercial, etc.**

**Deuxièmement, il est procédé à la vérification des valeurs aberrantes pour les variables de hauteur des bâtiments :** 12 enregistrements sont concernés par une hauteur maximale de 0 m. Cependant, les informations disponibles ne permettent d'identifier aucun groupe de bâtiment. En effet, les adresses des groupes de bâtiments ne sont pas renseignées, MySQL renvoie des valeurs manquantes (null) pour les colonnes des tables adresse et de relation entre les adresses et les groupes de bâtiments. Il est étonnant de s'apercevoir que certains groupes de bâtiments ne sont pas rattachés à une adresse. De plus, même en interrogeant d'autres tables sur ces groupes de bâtiments (comme en effectuant une jointure avec celle relative à l'usage principal), MySQL renvoie également des valeurs manquantes. **La nature de ces groupes de bâtiments pose donc question, peut-être pourrait-il s'agir de bâtiments qui n'existent plus mais dont l'ID unique est conservé dans la BDNB.**

**Troisièmement, il est procédé à la vérification des valeurs aberrantes pour la variable de valeurs foncières :**

- Le 1<sup>er</sup> et le 3<sup>ème</sup> groupes de bâtiments qui présentent la plus grande valeur foncière (valeurs foncières maximums respectives de 41 076 960 € et 7 485 000 €) correspondent à des bâtiments de l'Aéroparc de Fontaine qui est une des plus grandes zones d'activité de l'est de la France et qui accueille plusieurs entreprises dans les secteurs de l'industrie et du transport (cf. photo aérienne GoogleMaps ci-dessous). Ces données semblent donc cohérentes avec les valeurs foncières de ces bâtiments groupes.





- Concernant le 2<sup>ème</sup> groupe de bâtiments avec la valeur foncière maximum la plus importante (10 300 000 €), les informations disponibles ne permettent pas d'identifier précisément le groupe de bâtiment correspondant. En effet, MySQL renvoie plusieurs valeurs manquantes (null) pour différentes variables qui permettraient d'identifier ce groupe de bâtiment (comme l'adresse par exemple).

**Il est donc possible d'en conclure que les valeurs extrêmes relevées correspondraient bien à de vastes ensembles immobiliers du type bâtiment industriel/tertiaire de parcs d'activités, centre-commercial, etc.**

**Quatrièmement, il est procédé à la vérification des valeurs aberrantes pour la variable de prix au mètre carré :**

- Concernant les deux premiers groupes de bâtiments qui présentent la plus grande valeur de prix au mètre carré (prix au mètre carré maximums de 249 000 €/m<sup>2</sup> pour les deux) et qui semblent correspondre au même groupe de bâtiments puisque plusieurs valeurs de variables coïncident entre ces deux groupes (les ID sont néanmoins différents), les informations disponibles ne permettent pas d'identifier précisément les groupes de bâtiments correspondant. En effet, MySQL renvoie plusieurs valeurs manquantes (null) pour différentes variables qui permettraient d'identifier ces groupes de bâtiments (comme l'adresse par exemple).
- Concernant le 3<sup>ème</sup> et le 4<sup>ème</sup> groupe de bâtiments présentant les plus hautes valeurs de prix au mètre carré, ces valeurs ne relèvent pas de valeurs extrêmes. En effet, les montants du prix au mètre carré sont respectivement de 11 233 et 10 284 €/m<sup>2</sup> pour ces groupes de bâtiments. Le 3<sup>ème</sup> groupe de bâtiment correspondrait à un bâtiment situé sur un espace public et l'autre à une caserne de pompiers.

**Il est donc possible d'en conclure que, à part les deux premiers groupes de bâtiments présentant les prix au mètre carré les plus importants, les autres valeurs correspondent à des montants « normaux ». La nature des premiers groupes de bâtiments resteraient donc à confirmer pour mieux saisir leurs prix au mètre carré qui semblent aberrants.**

**Enfin, il est procédé à la vérification des valeurs aberrantes pour la variable de surface des parcelles** (via un aperçu des 10 premières parcelles présentant les surfaces les plus élevées). Malheureusement, comme pour le sujet des hauteurs de bâtiments, les informations disponibles ne permettent pas d'identifier les parcelles. En effet, les adresses des parcelles ne sont pas renseignées, MySQL renvoie des valeurs manquantes (null) pour les colonnes des tables adresse et de relation entre les adresses et les groupes de bâtiments lors de la jointure. Il est étonnant de s'apercevoir que certaines parcelles ne sont pas rattachées à une adresse. De plus, même en interrogeant d'autres tables sur ces parcelles (comme en effectuant une jointure avec celle relative aux usages), MySQL ne renvoie quasiment que des valeurs nulles. La seule parcelle dont la nature est définie n'apporte pas de précision particulière dans la mesure où il est indiqué « Indifférencié » comme usage du groupe de bâtiments situé sur la parcelle. Néanmoins, il est à noter que pour des parcelles présentant des surfaces « normales », l'ensemble des données liées à la parcelle (usage du bâtiment groupe de la parcelle, adresse, etc.) sont disponibles. **Il est donc difficile de cerner le problème lié aux parcelles qui présentent des surfaces extrêmement élevées, mais il est fort probable qu'ils s'agissent de vastes étendues en milieu rural (bois, forêts), ce qui expliquerait l'absence d'adresse précise comme de groupe de bâtiment sur la parcelle.**

### iii. Préparation et traitement des données avant analyse sous PowerBI

En vue d'analyser les données en utilisant PowerBI, il doit tout d'abord être procédé à une phase de préparation de ces données. Le fichier PowerBI correspondant est joint en annexe du présent rapport.

Il est tout d'abord procédé au chargement des différentes tables de données.

Ensuite, **une première étape consiste à la vérification du format des données et aux modifications nécessaires depuis Power Query :**

- Ajout de la valeur « null » pour les cellules vides.
- Conversion des valeurs numériques en texte pour les variables appropriées (exemple : code postal).
- Conversion des 0/1 en true/false pour les variables appropriées.
- Conversion des formats date+heure en date, l'heure n'apportant aucune information pertinente.
- Conversion des caractères « , » en « . » pour les colonnes avec des valeurs numériques mais identifiées initialement en tant que valeurs textuelles.



- Conversion des valeurs textuelles en valeurs numériques pour les colonnes appropriées.

**Puis il est procédé à une vérification des relations entre les tables à partir de la vue du modèle. Les ajustements réalisés sont les suivants :**

- Ajout de la relation entre les tables « parcelle » et « rel\_batiment\_groupe\_parcelle », non identifiée initialement.
- Modification de la direction du filtre croisé avec une direction à double-sens entre les tables « adresse » et « rel\_batiment\_groupe\_adresse » ainsi que les tables « parcelle » et « rel\_batiment\_groupe\_parcelle ». Autrement, les actions de filtrage n'auraient porté que sur certaines tables ce qui aurait biaisé les résultats.

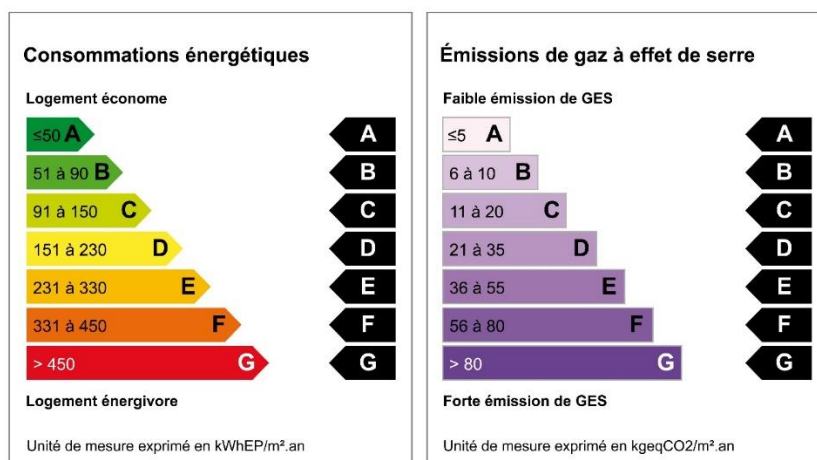
**Ensuite, en vue des analyses demandées, il est procédé à l'ajout de nouvelles données :**

- Ajout d'une colonne « batiment\_groupe\_indicateur\_reseau\_chaud\_froid » indiquant si le bâtiment groupe est réellement raccordé à un réseau de chaleur urbain ou s'il y a seulement un potentiel de raccordement car proximité avec un réseau de chaleur existant.
- Ajout d'une colonne dans la table « batiment\_groupe\_qpv » indiquant que le bâtiment est bien intégré à un Quartier Prioritaire de la Ville (valeur TRUE) pour permettre des actions de filtrage.

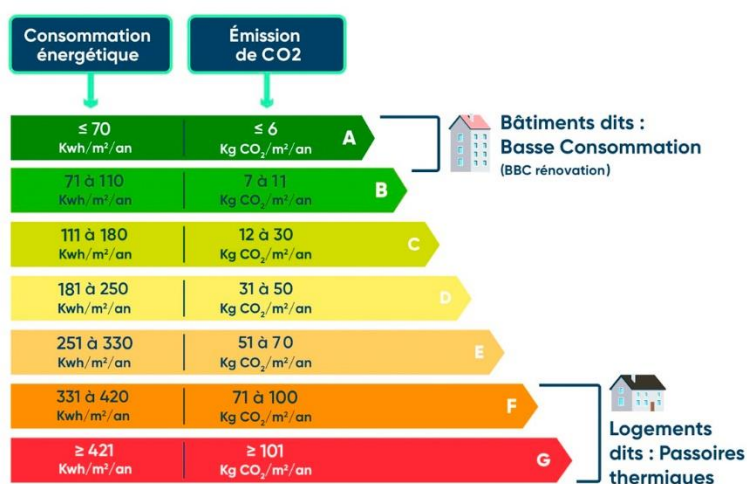
**Enfin, le dernier sujet porte sur la catégorisation des logements entre trois catégories (logement économe, moyennement économe et énergivore) en fonction de leur classe énergétique selon leur DPE.** Un écueil rencontré porte sur la différence entre les DPE issus de l'arrêté de 2012 et ceux issus de l'arrêté de 2021 (cf. ci-dessous étiquettes correspondant à chaque arrêté) :

- Arrêté 2012 (ancien arrêté) : il définissait la performance du logement selon deux étiquettes, une étiquette énergétique basée sur la consommation d'énergie du logement, qui indiquait la classe finale du logement (A, B, C, etc.), et une étiquette liée aux émissions de gaz à effet de serre, indicative. Le calcul des indicateurs était réalisé sur la base de 3 vecteurs de consommation (chauffage, eau chaude sanitaire, climatisation).
- Arrêté 2021 (arrêté en vigueur) : il définit la performance du logement selon un double-seuil, la consommation d'énergie et les émissions de gaz à effet de serre selon 5 vecteurs de consommation (chauffage, ECS, climatisation, éclairage et auxiliaires). Le moins bon des deux indicateurs détermine la classe finale. A

noter également que les valeurs seuils entre chaque classe (A, B, C, etc.) ont été modifiées par rapport à l'ancien arrêté.



Etiquettes DPE arrêté 2012



Etiquette DPE arrêté 2021

Compte-tenu de cette évolution réglementaire, la base de données contient en partie des logements concernés par l'arrêté de 2012 et en partie des logements concernés par l'arrêté de 2021.

Afin de comparer entre eux l'ensemble de ces logements selon un même référentiel (celui de l'arrêté de 2021 par exemple), il était souhaité de repartir des relevés de consommations d'énergie et d'émissions de gaz à effet de serre disponibles via la table « batiment\_groupe\_dpe\_representatif\_logement ». Cependant, il n'est pas possible de procéder ainsi puisque les méthodes de calculs ont évolué entre les deux arrêtés. En effet, comme indiqué précédemment, le calcul des consommations

énergétiques et des émissions de gaz à effet de serre de l'ancien arrêté était réalisé selon 3 vecteurs de consommation, tandis que pour le dernier arrêté il est réalisé selon 5 vecteurs de consommation. Les relevés de consommations et d'émissions disponibles dans le jeu de données ne sont donc pas comparables entre les logements concernés par l'ancien arrêté et ceux concernés par l'actuel.

**Donc le classement entre logements économes, moyennement économes et énergivores a été réalisé sur la base du postulat suivant** (pour les logements concernés par l'ancien arrêté, seule la classe énergétique est retenue dito ancien arrêté) :

- Logement économe : classes A à B incluses (tout arrêté confondu).
- Logement moyennement économe : classes C à E incluses (tout arrêté confondu).
- Logement énergivore : classes F à G incluses (tout arrêté confondu).

Il a été ensuite ajouté aux tables « `batiment_groupe_dpe_statistique_logement` » et « `batiment_groupe_dpe_representatif_logement` » une colonne qui indique la catégorie du logement selon sa classe.

### 3. ANALYSE ET RESULTATS

Pour plus de détails, il conviendra de se rapporter au fichier PowerBI joint au présent rapport, qui comprend les différents tableaux de bord relatifs aux thématiques de l'étude dont les résultats présentés ci-après sont issus.

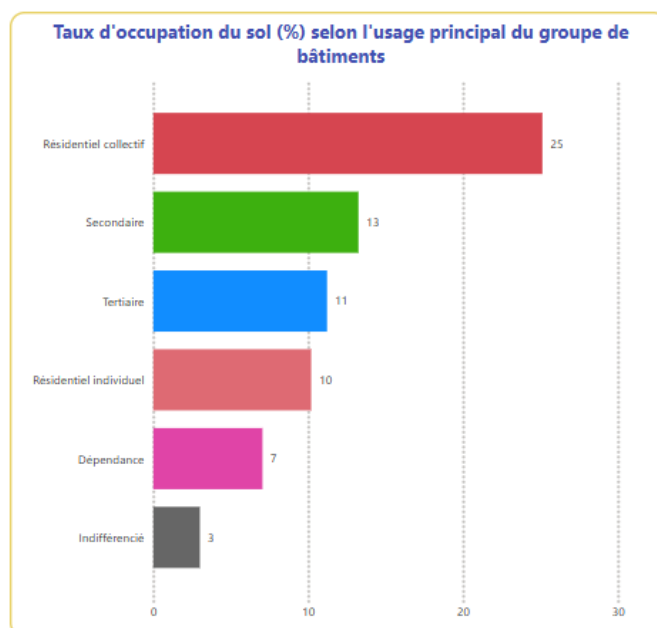
#### a. Formes urbaines et occupations du Territoire de Belfort

**Le taux d'occupation du sol (qui correspond au rapport entre la surface d'emprise au sol des bâtiments et celle du terrain) s'élève à environ 1,8%. Ce taux d'occupation est relativement faible et indique que le département est globalement peu urbanisé** (nota : ce taux intègre uniquement la surface d'emprise au sol des bâtiments, et non d'autres ouvrages du type zone de stationnement, voirie, etc.). **Alors que le taux d'occupation du sol pour la commune de Belfort, chef-lieu du département, s'élève à environ 28%.**

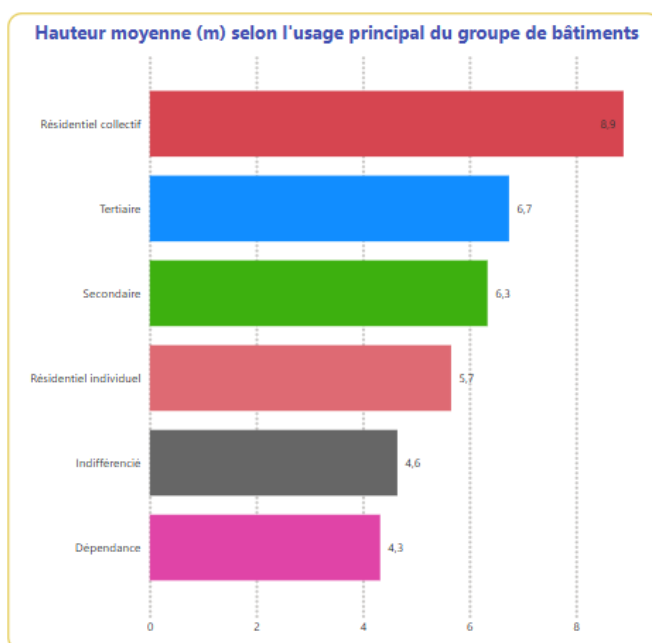
A noter que d'après les données de la BDNB, la surface totale de parcelles (équivalente à la surface des terrains au sol) pour ce département s'élève à environ 587 km<sup>2</sup>. Alors que selon le site officiel du département, sa superficie s'élève à 610 km<sup>2</sup>. **Il pourrait donc y avoir des parcelles non répertoriées dans la BDNB.**

Le graphique ci-dessous présente le taux d'occupation du sol en fonction de l'usage principal du groupe de bâtiments. **Le résidentiel collectif (assimilable à des appartements) est l'usage qui présente le taux d'occupation du sol le plus élevé (25%), alors que le résidentiel individuel (assimilable à des maisons individuelles) présente un taux d'occupation du sol plus faible (10%).** Ce résultat est globalement cohérent car, pour une parcelle donnée, un immeuble de logements va généralement occuper une grande partie de cette parcelle. Tandis que pour une maison individuelle située dans un environnement périurbain/rural, la maison n'occupera généralement qu'une faible partie de la parcelle sur laquelle elle est construite (le reste du terrain pouvant être dédié à des espaces verts). Pour information :

- Les usages secondaire et indifférencié correspondent généralement à des locaux d'activité type industriel (atelier, usine, etc.) ou tertiaire spécifique (exemple : logistique).
- L'usage dépendance correspond généralement à de petits bâtiments isolés et dont l'activité peut varier.

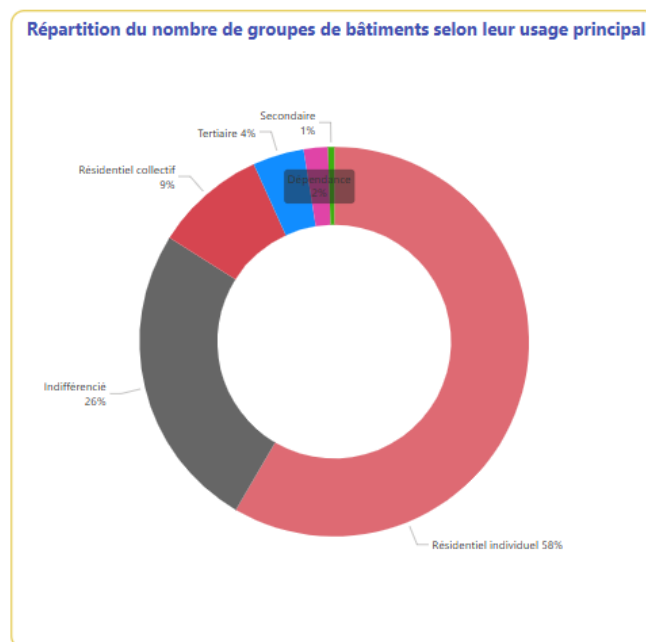


Concernant la hauteur moyenne des groupes de bâtiments selon leur usage principal (cf. graphique ci-dessous), il y a dans l'ordre décroissant le résidentiel collectif (8,9 m, équivalent d'un R+2), le tertiaire (6,7 m), la catégorie secondaire (6,3 m) le résidentiel individuel (5,7 m), la catégorie indifférenciée (4,6 m) et enfin la catégorie dépendance (4,3 m). Cette hiérarchie est peu surprenante, et **les hauteurs moyennes reflètent un parc immobilier globalement de faible hauteur dans le département.**



Enfin, concernant la répartition du nombre de groupes de bâtiments en fonction de leur usage principal, il est constaté selon le graphique ci-dessous que **le résidentiel individuel représente la majeure partie des bâtiments du territoire (environ 58%). Vient ensuite la catégorie indifférenciée (26%) puis le résidentiel collectif (9%).** La forte part

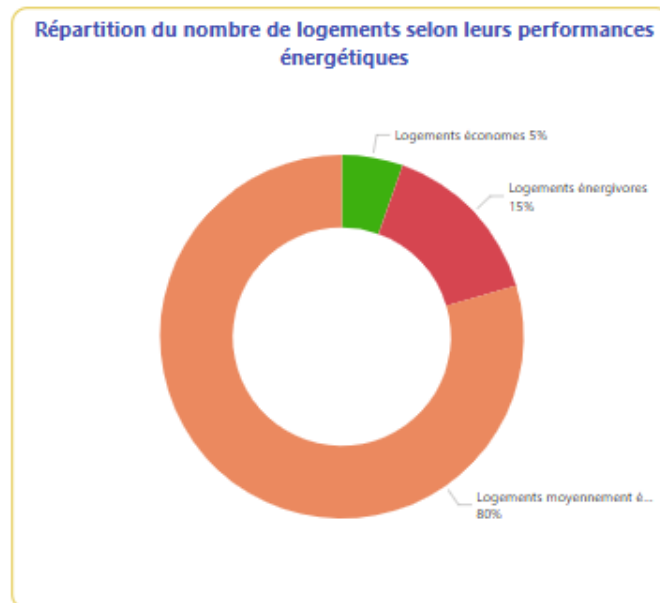
de bâtiments type résidentiel individuel dans le département confirme sa faible urbanisation et sa faible densification, ainsi que son caractère résidentiel



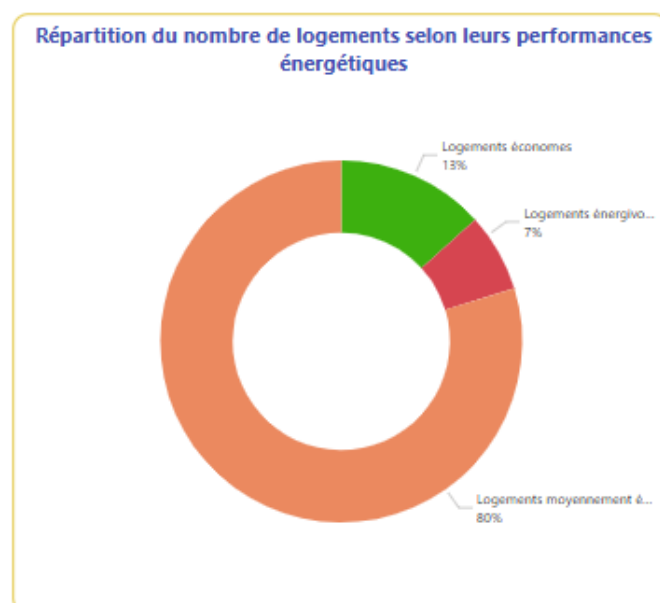
## b. Performances énergétiques

### i. Classement des logements selon leurs performances

D'après le graphique ci-dessous (réalisé sur la base des DPE représentatifs par groupe de logements, donc qui ne reflète pas réellement l'ensemble des logements du département) qui porte sur l'ensemble du parc de logements du département, **la part de logements moyennement économes s'élève à environ 80%, vient ensuite les logements énergivores qui représentent près de 15%, et enfin les logements économes qui en représentent environ 5%.** Ce résultat est relativement cohérent dans la mesure où près de 65% du parc de logements du département date d'avant 1974. Or, c'est essentiellement à partir des premiers chocs pétroliers (1973 et 1979) que les pouvoirs publics ont pris des mesures visant au renforcement des performances énergétiques des bâtiments.

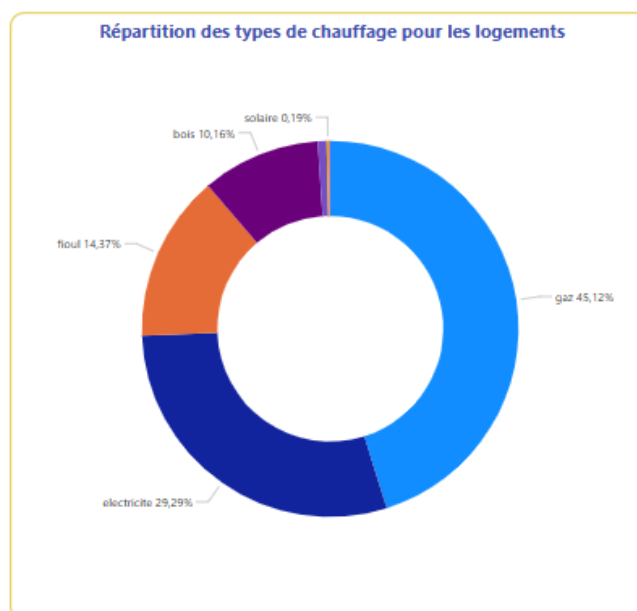


Concernant les logements situés dans des quartiers prioritaires de la ville, près de 82% des logements datent d'avant 1974. Cependant, **bien que la part des logements moyennement économes s'élève également à environ 80% (comme dans le reste du département), la part des logements énergivores est de seulement 7% environ et celle des logements économes est de 13% environ** (cf. graphique ci-dessous). Alors qu'on pourrait s'attendre à des logements moins performants dans les quartiers prioritaires de la ville (quartier en situation de précarité), il s'avère finalement que les logements dans ces zones sont globalement plus performants qu'à l'échelle des autres zones du département. **Ceci pourrait indiquer que des actions propres à la politique de la ville ont déjà été mises en œuvre dans ces zones, dont la rénovation thermique des logements.**



## ii. Modes de chauffage

D'après le graphique suivant (également réalisé sur la base des DPE représentatifs par groupe de logements), **les principaux types de chauffage du parc de logements sont dans l'ordre décroissant le gaz (45%), l'électricité (29%) puis le fioul (14%)**. Bien que le gaz et le fioul sont avantageux d'un point de vue notation énergétique comparés à l'électricité, puisque le coefficient de conversion d'énergie finale à énergie primaire pour le gaz et le fioul est de 1 tandis qu'il est de 2,3 pour l'électricité (ce coefficient reflète le principe que la consommation finale de 1 unité d'énergie revient à en consommer en réalité 2,3 unités, compte-tenu des pertes liées à la production, au transport, etc.), ils le sont moins d'un point de vue émission de gaz à effet de serre. En effet, pour l'usage chauffage, le fioul et le gaz sont fortement émissifs (respectivement 324 gCO<sub>2e</sub> / kWh et 227 gCO<sub>2e</sub> / kWh) tandis que l'électricité l'est beaucoup moins (147 gCO<sub>2e</sub> / kWh) dans la mesure où une part importante de la production d'électricité en France provient du nucléaire, qui est une énergie très peu émissive (sources : [legifrance.gouv.fr](http://legifrance.gouv.fr) et [notre-environnement.gouv.fr](http://notre-environnement.gouv.fr)).



**La part des réseaux de chaleur dans le chauffage est quant à elle négligeable (environ 0,08%).** Or, il s'agit d'un type de chauffage qui influe généralement de manière positive sur les étiquettes énergétique (coefficient de conversion d'énergie finale à énergie primaire de 1) et gaz à effet de serre (116 gCO<sub>2e</sub> / kWh). D'après la table « *batiment\_groupe\_indicateur\_reseau\_chaud\_froid* » qui fournit des informations relatives au raccordement des groupes de bâtiments à un réseau urbain de chaleur, le nombre de groupes de bâtiments raccordés s'élèverait à seulement 31, ce qui semble particulièrement faible à l'échelle d'un département. Il faut tout de même rappeler qu'il est indiqué dans la documentation : « *Attention les données sur les réseaux de chaleurs sont aujourd'hui bien moins fiables que les vecteurs gaz ou électricité.* ». Le nombre de groupes de bâtiments qui pourraient potentiellement se raccorder à des réseaux de chaleur car situés à proximité s'élève à 52, ce qui reste



très faible également. **Néanmoins, ce type de raccordement resterait à privilégier si faisable et pertinent d'un point de vue technico-économique.**

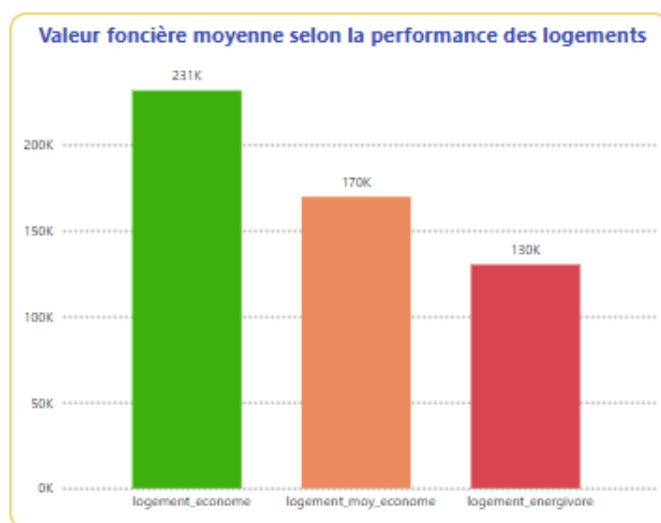
### c. Valeurs financières

#### i. Valeur foncière moyenne selon la performance des logements

En lien avec la précédente thématique, il est procédé à un recoupement entre les valeurs foncières (qui caractérise le prix de vente d'un bien immobilier) et les performances énergétiques des logements. D'après le graphique ci-dessous, **plus le logement est performant, plus sa valeur foncière est importante, ce qui est un résultat cohérent.** En effet :

- Pour les logements économes, la valeur foncière moyenne s'élève à environ 231 k€.
- Pour les logements moyennement économes, la valeur foncière moyenne s'élève à environ 170 k€.
- Pour les logements énergivores, la valeur foncière moyenne s'élève à environ 130 k€.

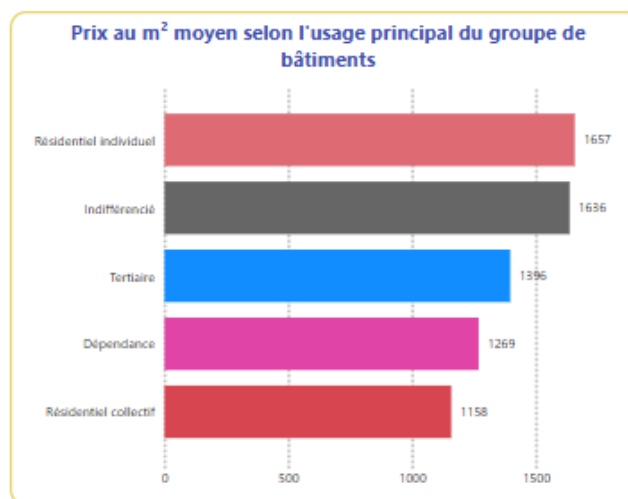
Il y aurait donc une corrélation positive entre la performance du logement et sa valeur foncière.



#### ii. Prix au mètre carré selon l'usage du bâtiment

**En ce qui concerne l'évaluation du prix au mètre carré moyen selon l'usage principal du groupe de bâtiments, le prix moyen au m<sup>2</sup> d'une maison individuelle** (assimilable à la catégorie résidentiel individuel) **est d'environ 1 657 €/m<sup>2</sup> contre environ 1 158 €/m<sup>2</sup> pour un appartement** (assimilable à la catégorie résidentiel collectif) – cf. graphique ci-dessous. Ce résultat semble globalement cohérent à l'échelle du département, mais il est fort probable que cela puisse varier en fonction des zones étudiées (exemple entre un appartement en centre-ville d'une grande ville et une maison en zone rurale). En ce qui concerne les autres catégories, les prix au m<sup>2</sup> sont les suivants :

- Indifférencié : 1 636 €/m².
- Tertiaire : 1 396 €/m².
- Dépendance : 1 269 €/m².



## 4. CONCLUSION

Le Territoire de Belfort est un département faiblement urbanisé (taux d'occupation du sol : 1,8%), peu dense (hauteur moyenne des constructions : 5,8 m) et essentiellement résidentiel (67% des groupes de bâtiments).

Le parc de logements est globalement vieillissant (65% des logements datés d'avant 1974) et peu performant (80% des logements sont évalués comme moyennement performants). Néanmoins, des actions de rénovation énergétique semblent être à l'œuvre dans les Quartiers Prioritaires de la Ville, où la part de logements performants (13%) est supérieure à celle dans le reste du département (5%). Or, plus un logement est performant, plus sa valeur foncière est importante (231 k€ pour un logement performant contre 130 k€ pour un logement énergivore).

A ce jour, les modes de chauffage privilégiés dans les logements sont le gaz (45%), l'électricité (29%) et le fioul (14%). Malgré le caractère performant du gaz et du fioul d'un point de vue énergétique, il est important de rappeler que ces modes de chauffage sont aussi très émissifs en gaz à effet de serre.

Enfin, les tendances de prix au mètre carré dans le département sont les suivantes :

- Résidentiel individuel : 1 660 €/m<sup>2</sup>.
- Résidentiel collectif : 1 160 €/m<sup>2</sup>.
- Tertiaire et industriel : 1 400 à 1 640 €/m<sup>2</sup>.

## 5. ANNEXES

### A. Jeu de données

Le jeu de données ainsi que le fichier « documentation » sont disponibles via le lien de téléchargement suivant :

[https://open-data.s3.fr-par.scw.cloud/bdnb\\_millesime\\_2024-10-a/millesime\\_2024-10-a\\_dep90/open\\_data\\_millesime\\_2024-10-a\\_dep90\\_csv.zip](https://open-data.s3.fr-par.scw.cloud/bdnb_millesime_2024-10-a/millesime_2024-10-a_dep90/open_data_millesime_2024-10-a_dep90_csv.zip)

### B. Détail du contenu des tables retenues pour l'étude

Cf. fichier « documentation\_tables etude » en annexe du présent rapport.

### C. Notebook Python

Cf. fichier « Module 6\_Notebook\_BROYELLE Olivier » en annexe du présent rapport.

### D. Script MySQL et résultats des requêtes

**Script MySQL** : cf. fichier « Module 6\_ScriptSQL\_BROYELLE Olivier » en annexe du présent rapport.

**Résultats des requêtes** : cf. dossier « Résultat requêtes » en annexe du présent rapport.

### E. Fichier PowerBI et tableaux de bord

Cf. fichier « Module 6\_PowerBI\_BROYELLE Olivier » en annexe du présent rapport.