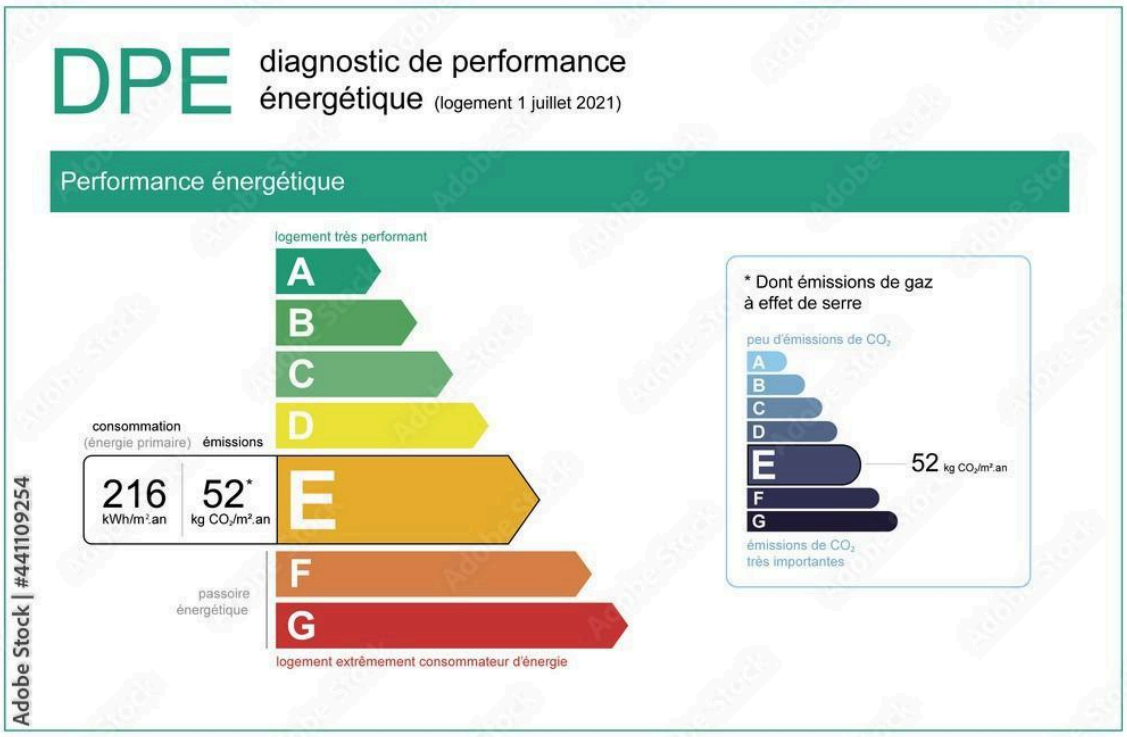
PROJET OPEN DATA UNIVERSITY

Le diagnostic de performance énergétique (DPE)



RÉDIGÉ PAR :

. DAALOUL YOUSSEF
. ABDOUL MOUMOUNI HAMANI
. MOHAMED HABIB SARRAJ
LINDA ATTARI
.MOHAMED ALI MAROUANE

INTERVENANTE :

INSEEC MSc & MBA

Msc Big Data & Manager Marketing

SOMMAIRE

1.INTRODUCTION

2. CONTEXTE
3. Méthodologie
4. Problématique
5. CONSTAT INSIGHT
6. BASES DE DONNÉES
- **Données structurées**
- **Données non-structurées**
- **Résultat d'analyse**

7. RECOMMANDATIONS
8. CONCLUSION

Introduction

Notre intérêt s’est porté sur le sujet des diagnostics de performances énergétiques parce que nous trouvons que c’est un sujet intéressant, d’actualité et possédant ainsi un défi technique en termes de manipulation de données de tailles pour nous et qui nous permettra de se former d’avantage d’un point de vue technique. Pouvoir aussi échanger avec un mentor aussi passionné que le notre (Mr Xavier Duquesnes) était très enrichissant puisqu’il baigne dans le métier et la problématique que nous allons traiter.

CONTEXTE DU PROJET

Dans un monde confronté à des défis environnementaux croissants et à une prise de conscience grandissante de l'importance de la conservation des ressources, la gestion efficace de la consommation d'énergie est devenue une priorité majeure. L'évaluation de la performance énergétique des bâtiments, souvent mesurée à travers les Diagnostics de Performance Énergétique (DPE), est devenue un outil essentiel dans la quête d'une utilisation plus responsable et durable de l'énergie.

La mise en place d'un projet de suivi de la consommation basé sur les DPE s'inscrit dans ce contexte d'urgence climatique et de transition énergétique. Ce projet vise à collecter, analyser et interpréter les données de consommation d'énergie des bâtiments, telles que définies dans les rapports de DPE, afin de fournir des informations précieuses pour la prise de décision, la planification stratégique et la mise en œuvre de mesures d'efficacité énergétique.

L'importance de ce projet réside dans sa capacité à offrir une vue d'ensemble complète de la consommation d'énergie à l'échelle des bâtiments individuels, des quartiers ou même des régions entières. En permettant une analyse approfondie des tendances de consommation, des points chauds énergétiques et des opportunités d'amélioration, ce projet contribue à l'élaboration de politiques énergétiques efficaces, à la réduction des émissions de gaz à effet de serre et à la promotion d'une utilisation plus efficiente des ressources.

Cette introduction jettera les bases nécessaires pour comprendre l'importance et les enjeux associés à la mise en place d'un projet de suivi de la consommation basé sur les DPE, en mettant en lumière les défis actuels et les opportunités futures dans le domaine de la gestion énergétique.

Problématique

Suivant les données que nous avons et le but recherché de la mise en place des DPE, nous avons décidés de nous concentrer sur les gains de consommations entre les différentes classes énergétiques (A, B, C, D, E, F)

Est-ce les données qu’on a nous permettent-ils de voir de manière explicite le gain de consommation entre deux classes de DPE ?

Informations sur le jeu de données

Nous disposons d'une base de données sur la consommation d'énergie fournie par ENEDIS, une filiale du groupe EDF chargée de la gestion et de l'aménagement de 95% du réseau de distribution d'électricité en France. Ce jeu de données en open source, mis à disposition par ENEDIS, présente la consommation électrique annuelle des adresses de 10 logements ou plus. Il comporte une multitude de champs tels que le code IRIS, le nom et le numéro de voie, le code INSEE de la commune, la consommation annuelle de l'adresse, la consommation annuelle par logement de l'adresse, etc.

Le deuxième jeu de données à notre disposition est fourni par l'ADEME, un établissement public à caractère industriel et commercial chargé de mettre en œuvre des politiques dans les domaines de l'environnement et de l'énergie. Ce jeu de données concerne notamment les fameux diagnostics de performance énergétique (DPE), qui permettent d'estimer la consommation d'énergie d'un logement ou d'un bâtiment et d'évaluer ainsi son impact écologique. Suite aux nouvelles politiques de transition énergétique, l'État a mis en place un système de classe énergétique allant de A à F, dans le but de rendre les seuils des DPE plus équitables face à une surreprésentation des passoires énergétiques (étiquette F ou G).

Cette révision du DPE vise à permettre aux propriétaires de gagner du temps pour rénover leur logement, sachant que les biens classés « G » ne pourront plus être mis en location à partir de 2025. Environ 140 000 logements devraient ainsi sortir de la catégorie des passoires thermiques. Le jeu de données de l'ADEME est plus complet que celui d'ENEDIS, comprenant des données sur les différents modèles de DPE ainsi qu'une description plus détaillée du logement et de la consommation.

Méthologie et axe d’analyse

Dans le cadre de ce projet, le but étant de mener une étude quantitative quant a l’efficacité d’un changement de catégorie d’un point de vue financier et de consommation.

Le but premier était de collecter les données avant et après changement de catégorie de DPE afin de voir le gain généré par ce changement, on va comparer principalement la consommation électrique.

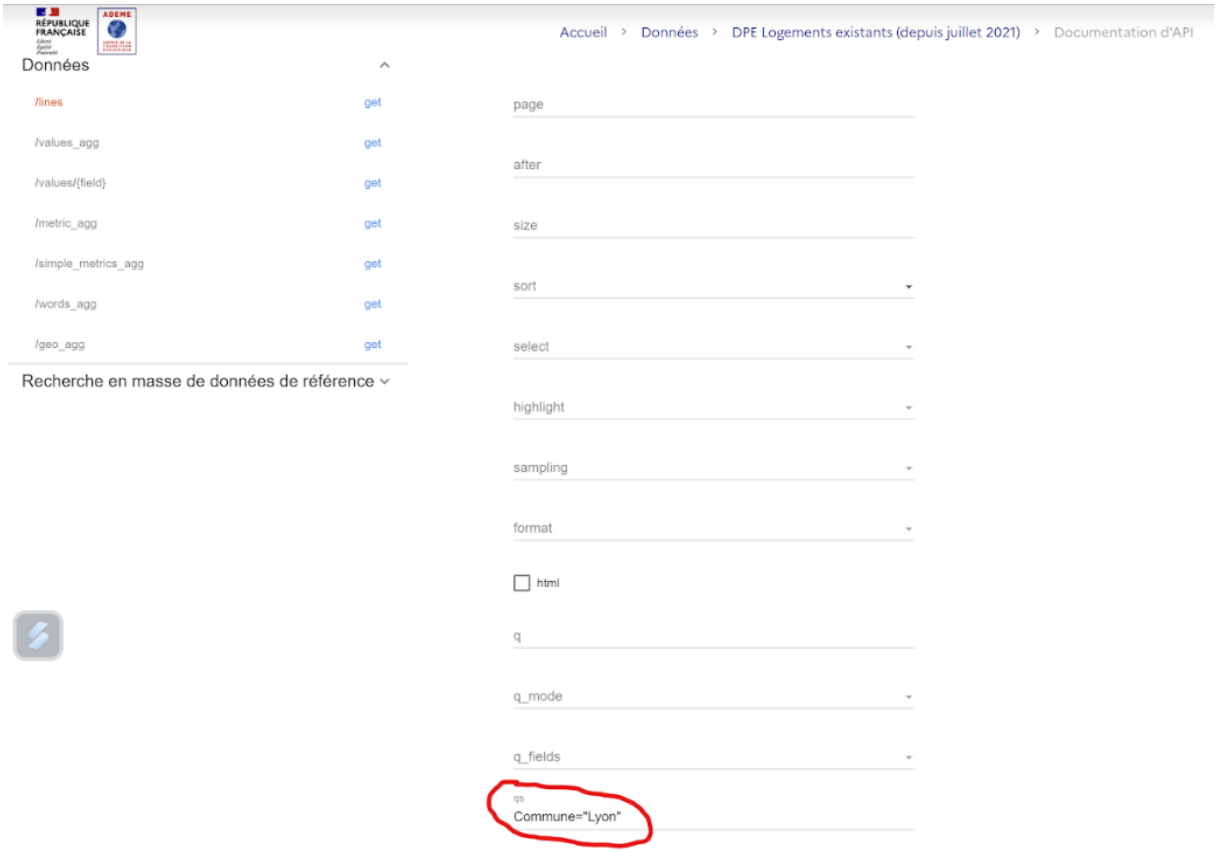
Pour rappel, nous avons deux bases que nous allons rapprocher qui sont une base de consommation issue de ENEDIS et une base de DPE.

Le téléchargement de la base ENEDIS a été facile à faire puisque la base était au format plat et pas volumineuse, en revanche nous avons passés beaucoup pour extraire la base de l'ADAME (DPE) car elle était au format SQL et très volumineuse.

Pour ce faire, nous avons utilisé un Système de Gestion de Base de Données Relationnelle (SGBDR), en l'occurrence PgAdmin, qui offre une interface pour interagir avec des bases de données PostgreSQL et la manipuler à l'aide de python et d'une libraire. Sauf que nous avons eu un souci sur le SGBDR avait atteint ses limites a cause de la faible capacité de nos machines et nous a pas sortit toutes les bases du dump.

La décision a donc été prise de se focaliser sur un échantillon de données plus restreint afin de pouvoir avancer sans barrière technique et d'avoir un traitement plus fluide.

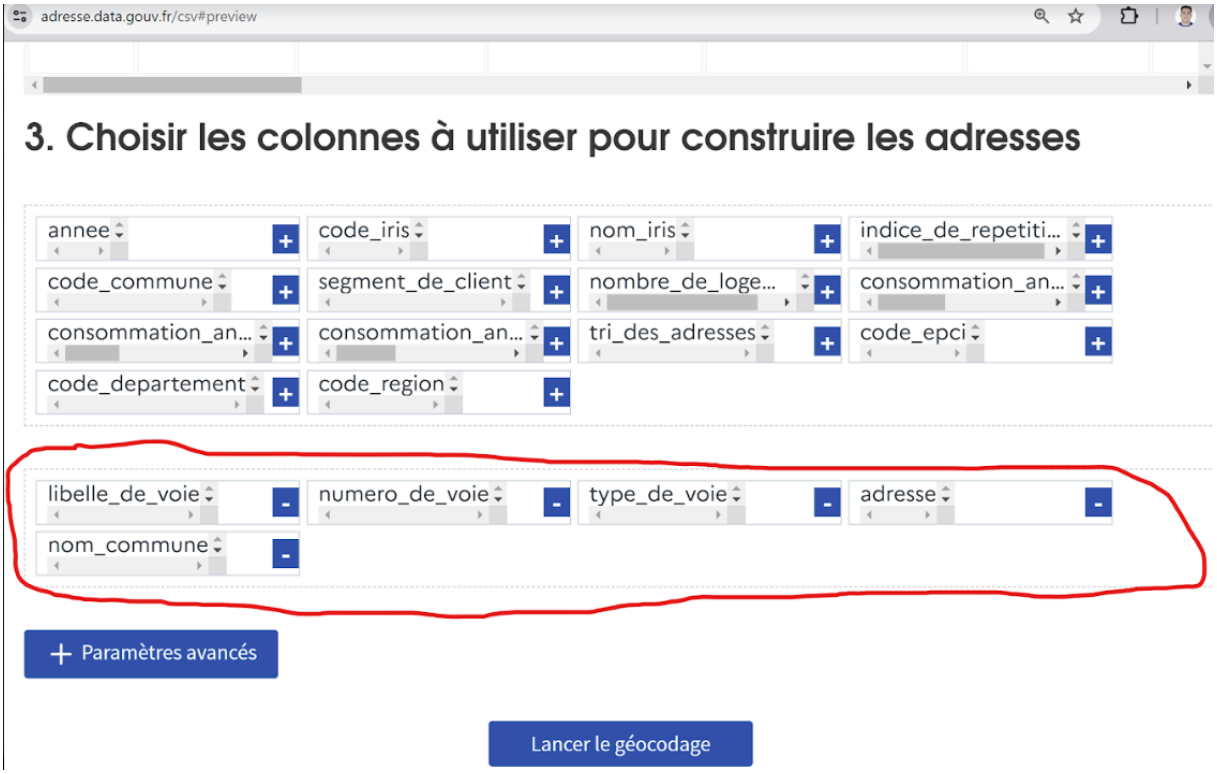
Notre échantillon d'étude se portera sur la ville de Lyon, on est passé a travers l'API de l'ADEME et filtrer sur Lyon (voir screen ci-dessous).



Pour pouvoir faire un rapprochement entre la base de consommation (ENEDIS) et la base DPE (ADEME), il nous fallait une clé unique afin d'avoir des données fiables. Les numéros DPE (unique) n'étaient pas présents dans la base ENEDIS, c'est dans cette optique la que l'on a décidé de rapprocher les données avec les adresses.

Une étape de normalisation des adresses étaient nécessaires puisqu'elles n'étaient pas écrites de la même manière dans les deux bases

On a utilisé l'API adresse du gouvernement qui est un outil qui permet de normaliser une adresse fournit avec les informations sur l'emplacement physique de l'adresse comme les colonnes suivantes : numéro de voie, type de voie, libelle de voie, code commune, nom de commune, code département, code région.



Lorsqu'on rentre les colonnes suivantes à l'API, cette dernière nous renvoie une suite de colonnes avec les données de chaque ligne d'adresses dont la longitude, la latitude, code postale et une colonne nommée « result name » qui semblée être la clé la plus granulaire afin de pouvoir faire notre jointure entre les deux tables et faire ainsi le rapprochement être la table de l'ADEME et la table d'ENEDIS.

Jointures de deux tables :

La jointure de type right joint a été faites entre les deux bases normalisées sur la colonne « result name » afin d'avoir une base finale propre pour faire notre analyse.

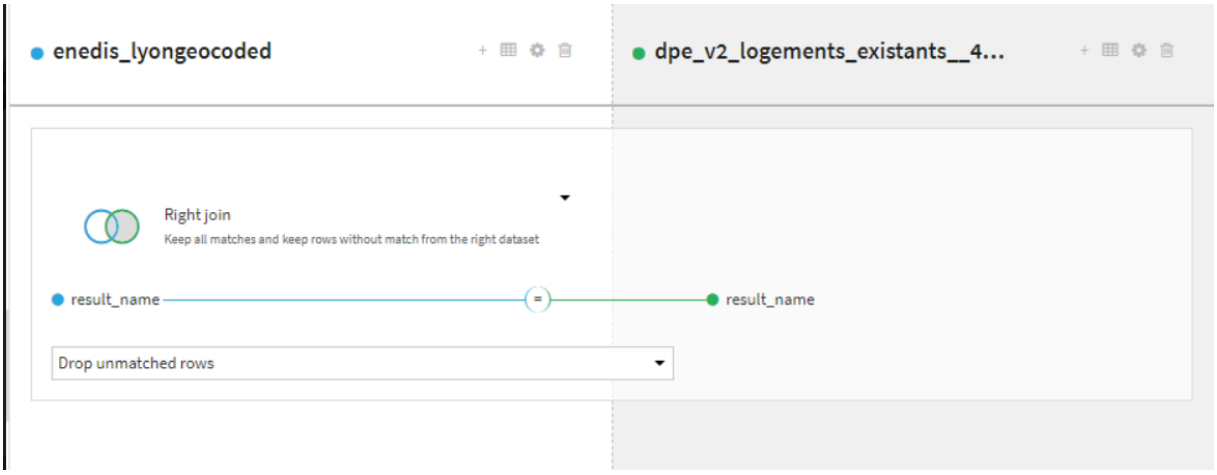
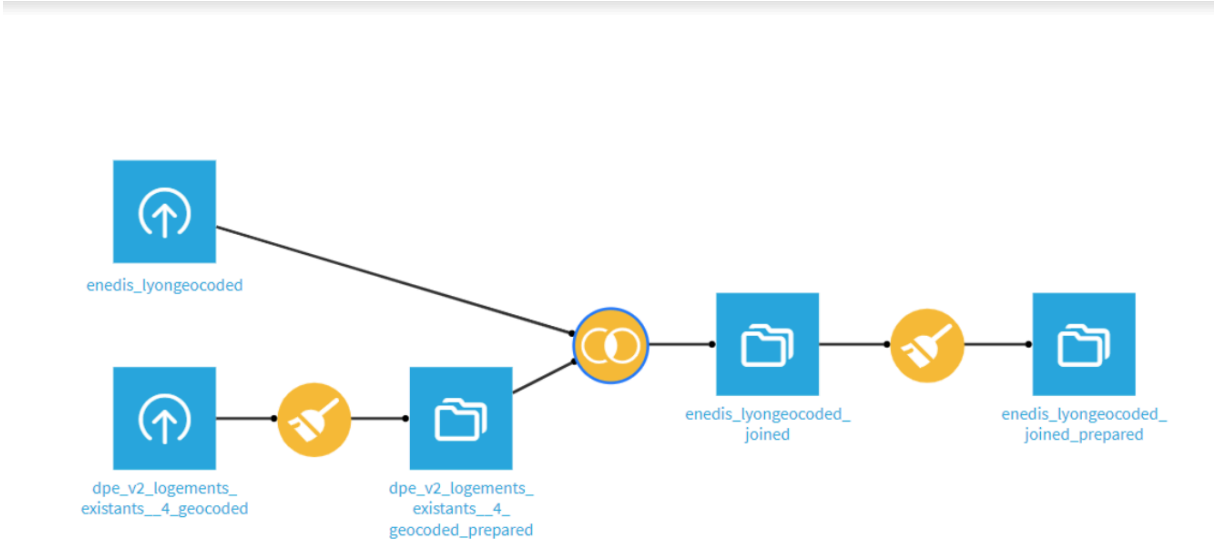
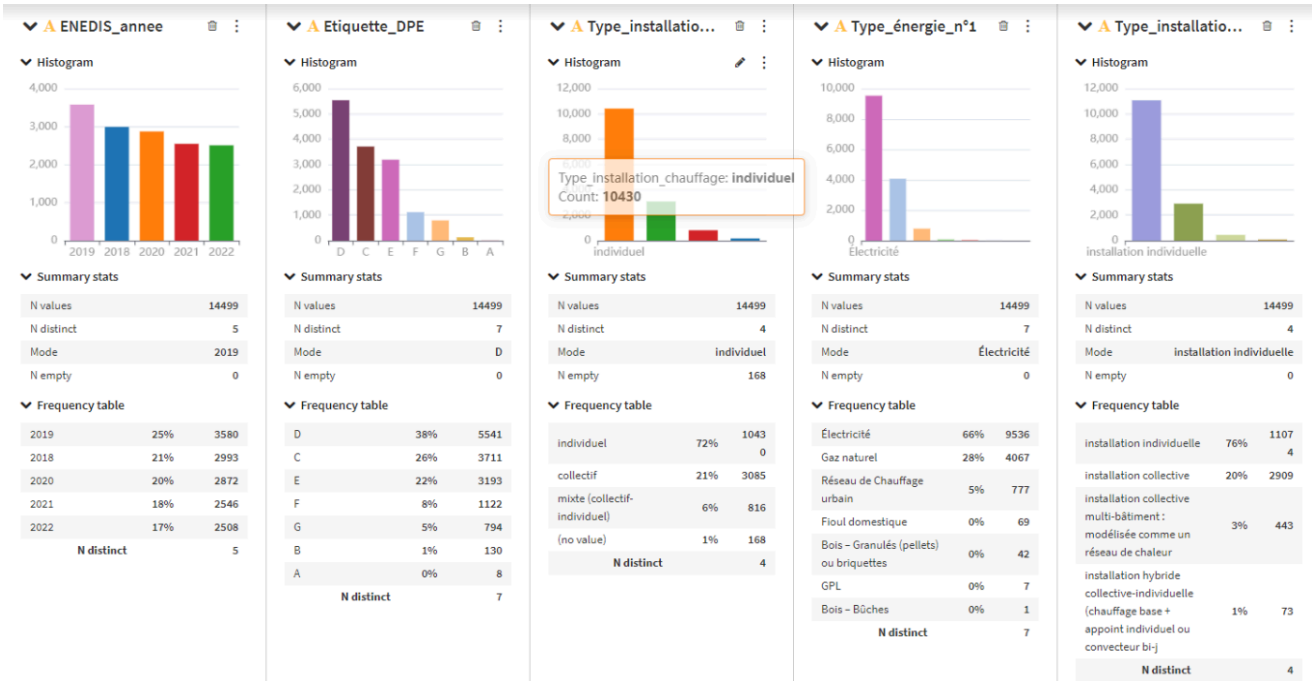


Schéma relationnel



Analyses statistiques :



- **Distribution Annuelle (ENEDIS_annee)** : Taille de données globale sur une période de 2019 à 2022.
- **Étiquettes DPE** : Catégorisation de A à G, à l'efficacité énergétique, avec des hauteurs de barres variables indiquant leur fréquence.
- **Type d'Installation** : Deux catégories 'collectif' et 'individuel' sont présentées, suggérant différents types d'installations énergétiques.

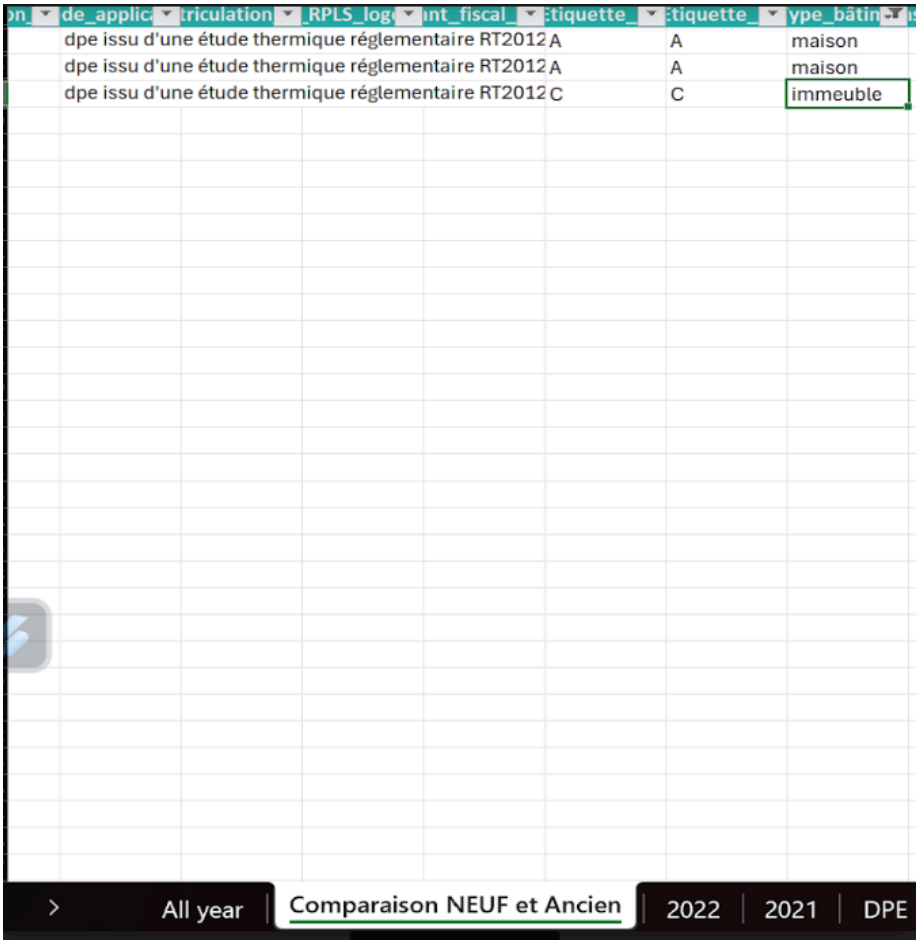
1. **Type d'Énergie n°1** : Différents types d'énergie comme le gaz naturel, l'électricité, le fioul domestique, les granulés de bois et les bûches de bois.

- **Installation Collective vs Individuelle** : Une autre répartition entre 'installation collective', 'multi-bâtiment' et 'installation individuelle'.

Les statistiques sommaires sous chaque histogramme incluent le nombre total de valeurs (N), la mode, le nombre d'éléments distincts et le nombre de cellules vides, ainsi qu'un tableau de fréquence détaillant les comptes ou pourcentages pour chaque catégorie ou année

Lorsque nous avons essayé de comparer notre base avec les DPE de logements neufs pour avoir pour un même logement un upgrade de classe et ainsi pouvoir faire l'analyse du gain, nous avons constaté que le rapprochement ne peut pas être fait puisque on a eu que 3 adresses qui concordé. Le problème étant que c'est impossible de comparer les appartements puisqu'ils partagent la même adresse de l'immeuble et ont un nouveau numéro DPE. Sauf les logements de types maisons et bâtiments, il est possible de faire un rapprochement mais 3 échantillons n'est clairement pas représentatifs.

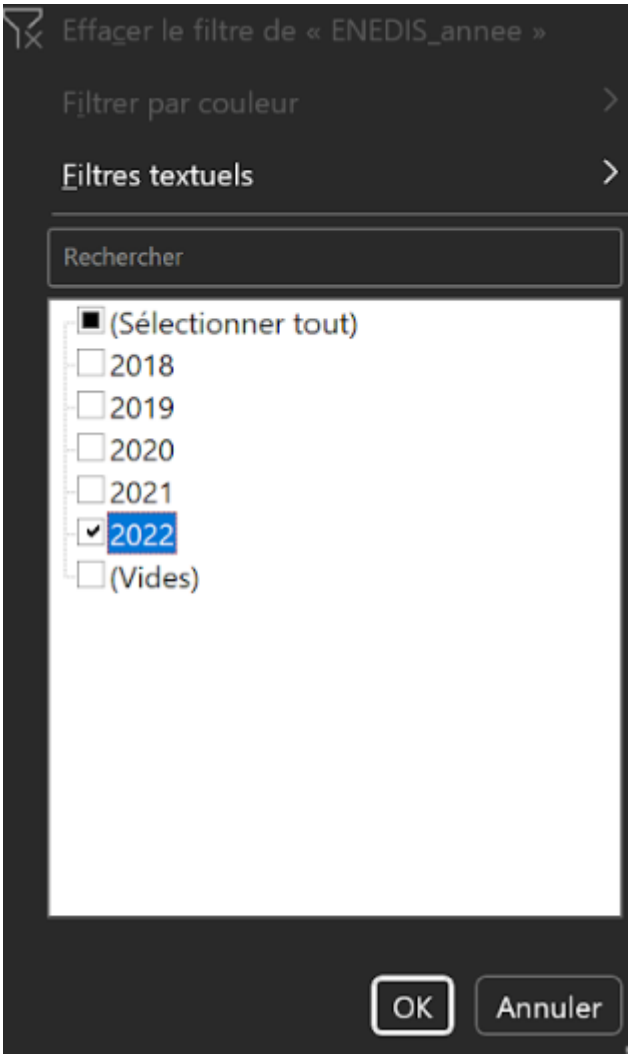
IE	IF	IG
ANCIEN DPE	ANCIEN E_ DPE	Nouveau E_ DPE
2369E2428950X	D	A
2369E2428950X	D	A
2169E0781390D	E	C



on_	de_applic	triculation	_RPLS_log	int_fiscal_	Etiquette_	Etiquette_	ype_bâtin
	dpe issu d'une étude thermique réglementaire RT2012				A	A	maison
	dpe issu d'une étude thermique réglementaire RT2012				A	A	maison
	dpe issu d'une étude thermique réglementaire RT2012				C	C	immeuble

Nous avons donc basculé sur une autre approche mais suivant la même problématique, qui est de comparer la consommation énergétique des différentes classes DPE d'immeuble pour l'année 2022, et voir si les données reflètent la réalité.

Filtre sur l'année 2022



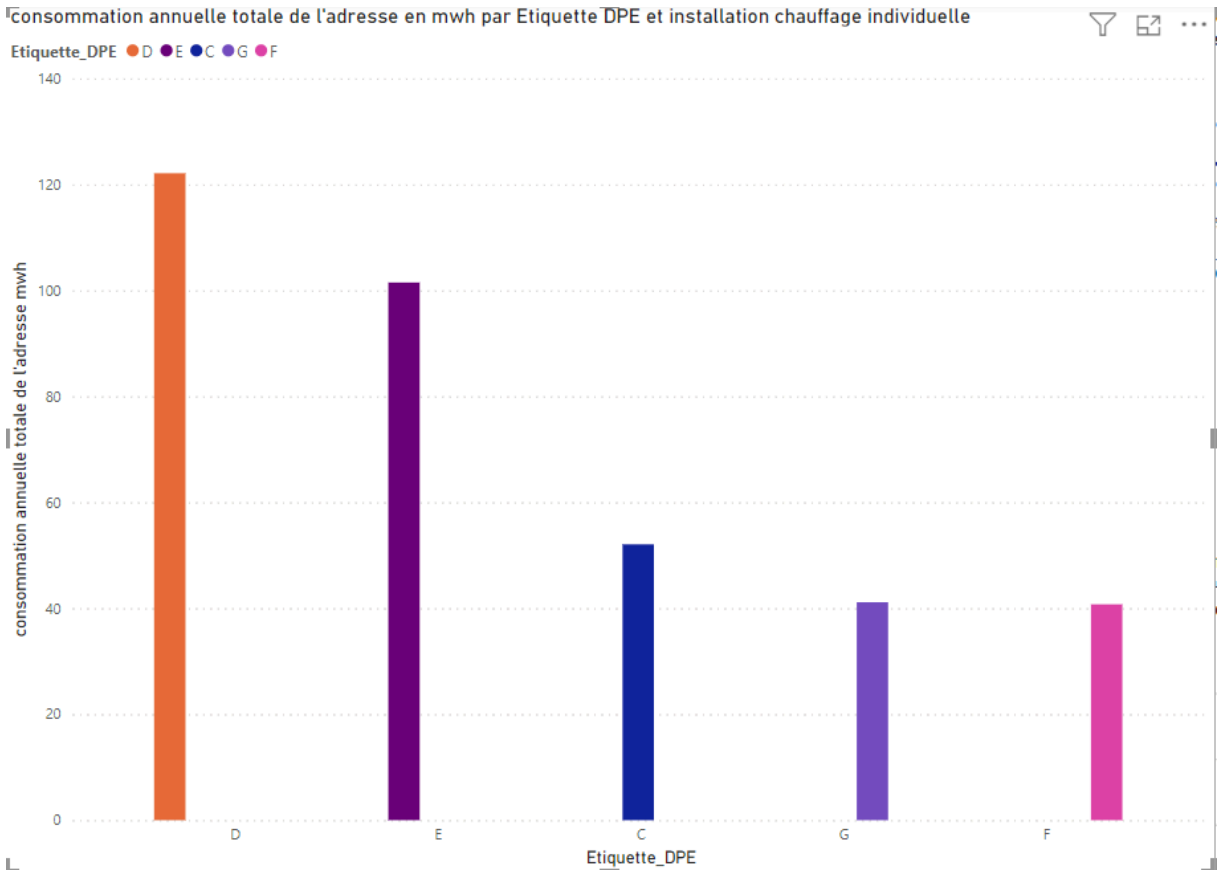
Puisque notre donnée contient des Dpe par appartements et la base Enedis contient la consommation par immeuble, il fallait donc ajouter une colonne qui nous permet de trouver la classe la plus occurrente par appartement et de l'associer a la classe du bâtiment entier

N°_DPE_immeuble_associé	Etiquette_D	dpe immeuble	Classementdpe
2269E1310744Y	D	1	1
2269E2450020G	E	2	1
2269E2995206Y	D	1	1
2369E0047746D	C	6	1
2369E0107898R	C	8	1
2369E0125300Z	D	3	1
2369E0134220B	C	4	1
2369E0147195C	B	3	1
2369E0174077A	E	1	1
2369E0229185O	C	1	1
2369E0229238P	C	1	1
2369E0229238P	D	1	1
2369E0282068N	D	4	1
2369E0284452F	C	4	1
2369E0352085M	C	10	1
2369E0373022T	D	1	1
2369E0374794X	C	1	1
2369E0374794X	D	1	1
2369E0513162T	C	3	1
2369E0561487K	C	4	1
2369E1315554Z	C	1	1
2369E1606046T	D	25	1
2369E1891457C	E	4	1
2369E2288508H	C	27	1
2369E2345972L	C	20	1
2369E2346813U	B	1	1

La colonne tout a droite a été ajouter a l'aide d'une formule DAX qui nous permet de calculer le nombre d'occurrence de classe par appartement groupé par immeuble et ensuite nous avons classé cette occurrence par ordre décroissant comme ca on garde l'étiquette de DPE qui est majoritairement représentées.

Nous avons à disposition maintenant une table complète avec pour chaque ligne un immeuble et chaque champs correspondant (consommation, classe DPE…).

Résultat d’analyse :

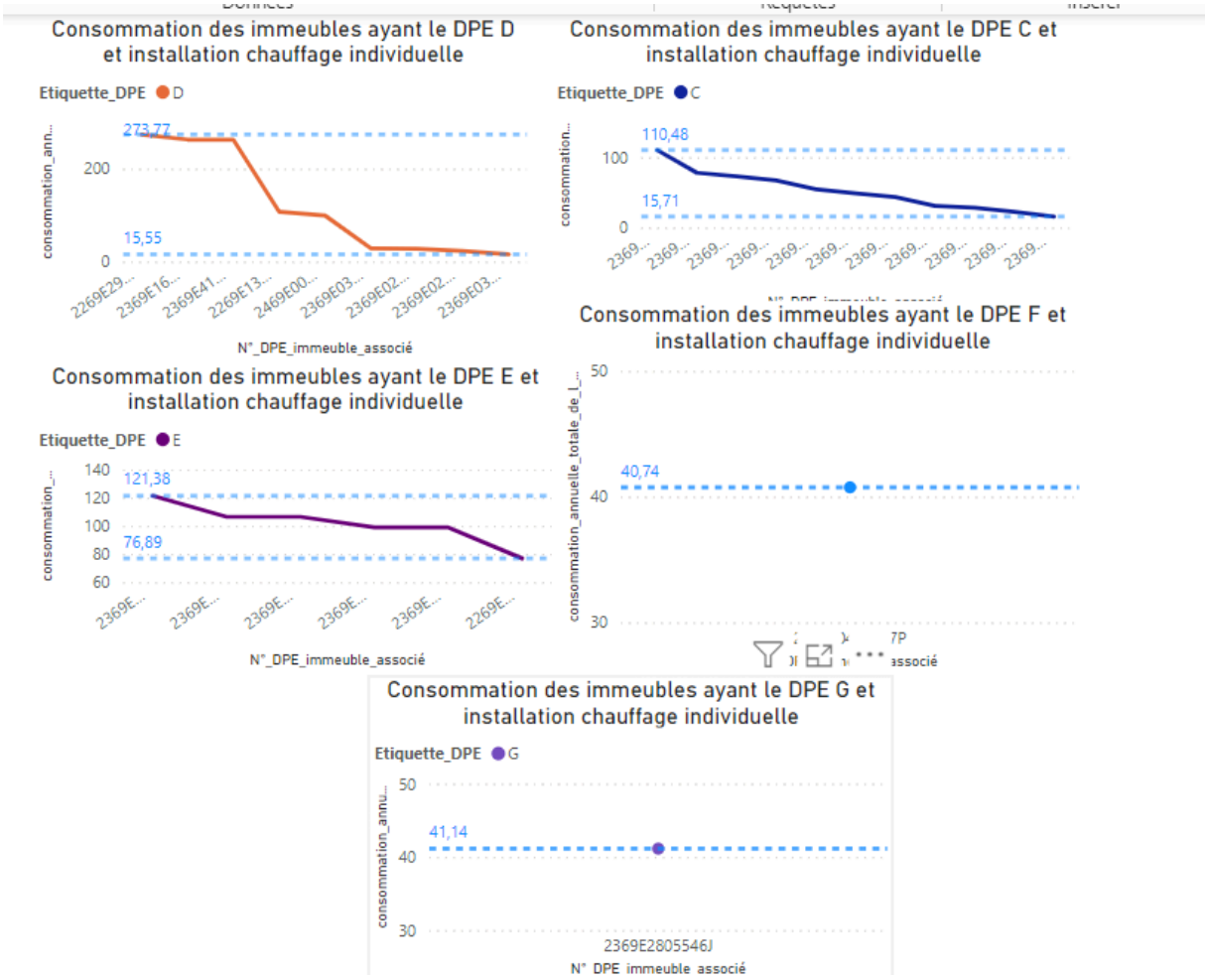


Ce graphique nous montre la consommation moyenne par immeuble ayant les mêmes classes de DPE.

On remarque que la consommation des immeubles dont la performance énergétique correspond à D est supérieur a des consommations moyennes des immeubles qui possèdent des performances énergiques plus gourmandes tel que E, G et F, ce qui n’est pas logique parce que nous attendons à un résultat trié par l’échelle mis en avance par l’ADEME.

Pour comprendre les raisons de cette incohérence, on a fait un suivi des consommations moyennes par immeuble, tout en fixant un caractère qui pourras influencer la consommation qui est le type de chauffage, (on a choisi le type d’installation individuel).

Pour comprendre les raisons de cette incohérence, on a fait un suivi des consommations moyennes par immeuble, tout en fixant un caractère qui pourras influencer la consommation qui est le type de chauffage, (on a choisi le type d’installation individuel).



Pour comprendre les raisons de cette incohérence, on a fait un suivi des consommations moyennes par immeuble, tout en fixant un caractère qui pourras influencer la consommation qui est le type de chauffage, (on a choisi le type d'installation individuel).

Et là, on remarque que dans le Graphique pour l'étiquette D, la consommation minimale enregistrée par un immeuble est égale à 15 mWh.

Or que, la barre minimale pour l'étiquette C = 15, donc on peut conclure qu'il n'y pas de concordance entre la consommation et la performance énergétique.

Donc on ne peut pas estimer le gain ou la perte en passant d'une classe énergétique a une autre en utilisant la consommation.

Conclusion :

Pour conclure, les données reflètent la réalité, c'est-à-dire qu'il est très difficile de catégoriser un bâtiment dans une classe précise de performance énergétique puisque pour un même bâtiment il y a plusieurs paramètres tel que le type de chauffage, le taux de présence des habitant et leur consommation qui viennent biaiser la segmentation.

ANNEXE DPE :

Identifiant BDD	Donnée obligatoire	Type de donnée	Table de références	Modèle concerné	DPE 2012	Description
N°DPE	Oui	Numérique	-	DPE	-	Numéro du Diagnostic de Performance Énergétique.
Date_réception_DPE	Oui	Date	-	DPE	-	Date de réception du DPE.
Date_établissement_DPE	Oui	Date	-	DPE	-	Date d'établissement du DPE.
Date_visite_diagnostiqueur	Oui	Date	-	DPE	-	Date de la visite du diagnostiqueur.
Modèle_DPE	Oui	Texte	-	DPE	-	Modèle du DPE utilisé.
N°_DPE_remplacé	Non	Numérique	-	DPE	-	Numéro du DPE remplacé, si applicable.
Date_fin_validité_DPE	Oui	Date	-	DPE	-	Date de fin de validité du DPE.
Version_DPE	Oui	Texte	-	DPE	-	Version du DPE utilisée.
N°_DPE_immeuble_associé	Non	Numérique	-	DPE	-	Numéro du DPE de l'immeuble associé, si applicable.
Appartement_non_visité_(0/1)	Non	Booléen	-	DPE	-	Indique si l'appartement n'a pas été visité (1) ou a été visité (0).

Méthode_application_DPE	Oui	Texte	-	DPE	-	Méthode d'application du DPE.
N°_immatriculation_copropriété	Non	Numérique	-	DPE	-	Numéro d'immatriculation de la copropriété, si applicable.
N°_RPLS_logement	Non	Numérique	-	DPE	-	Numéro du Répertoire des Logements Sociaux, si applicable.
Invariant_fiscal_logement	Non	Numérique	-	DPE	-	Invariant fiscal du logement, si applicable.
Etiquette_DPE	Oui	Texte	-	DPE	-	Étiquette de performance énergétique du DPE (de A à G).
Etiquette_GES	Oui	Texte	-	DPE	-	Étiquette d'émission de gaz à effet de serre du DPE (de A à G).
Type_bâtiment	Oui	Texte	-	DPE	-	Type de bâtiment (par exemple, résidentiel, tertiaire, etc.).
Type_installation_chauffage	Oui	Texte	-	DPE	-	Type d'installation de chauffage.
Type_installation_ECS_(général)	Oui	Texte	-	DPE	-	Type d'installation d'Eau Chaude Sanitaire.

Hauteur_sous-plafond	Non	Numérique	-	DPE	-	Hauteur sous plafond du logement.
Nombre_appartement	Non	Numérique	-	DPE	-	Nombre d'appartements dans le bâtiment.
Nombre_niveau_immeuble	Non	Numérique	-	DPE	-	Nombre de niveaux dans l'immeuble.
Nombre_niveau_logement	Non	Numérique	-	DPE	-	Nombre de niveaux dans le logement.
Surface_habitable_immeuble	Non	Numérique	-	DPE	-	Surface habitable de l'immeuble.
Surface_tertiaire_immeuble	Non	Numérique	-	DPE	-	Surface tertiaire de l'immeuble.
Surface_habitable_logement	Oui	Numérique	-	DPE	-	Surface habitable du logement.
Classe_inertie_bâtiment	Non	Texte	-	DPE	-	Classe d'inertie du bâtiment.
Typologie_logement	Oui	Texte	-	DPE	-	Typologie du logement (par exemple, appartement, maison, etc.).
Position_logement_dans_immeuble	Non	Texte	-	DPE	-	Position du logement dans l'immeuble (par exemple, rez-de-chaussée, étage, etc.).
Classe_altitude	Non	Texte	-	DPE	-	Classe d'altitude du logement.

Zone_climatique_	Oui	Texte	-	DPE	-	Zone climatique du logement.
Adresse_brute	Oui	Texte	-	DPE	-	Adresse brute du logement.
Nom__commune_(BAN)	Oui	Texte	-	DPE	-	Nom de la commune selon la Base Adresse Nationale (BAN).
Code_INSEE_(BAN)	Oui	Texte	-	DPE	-	Code INSEE de la commune selon la BAN.
N°_voie_(BAN)	Oui	Numérique	-	DPE	-	Numéro de voie selon la BAN.
Identifiant__BAN	Oui	Texte	-	DPE	-	Identifiant unique de l'adresse selon la BAN.
Adresse_(BAN)	Oui	Texte	-	DPE	-	Adresse complète selon la BAN.
Code_postal_(BAN)	Oui	Numérique	-	DPE	-	Code postal selon la BAN.
Score_BAN	Non	Numérique	-	DPE	-	Score de qualité de l'adresse selon la BAN.
Nom__rue_(BAN)	Oui	Texte	-	DPE	-	Nom de la rue selon la BAN.
Coordonnée_cartographique_X_(BAN)	Non	Numérique	-	DPE	-	Coordonnée cartographique X selon la BAN.

Coordonnée_cartographique_Y_(BAN)	Non	Numérique	-	DPE	-	Coordonnée cartographique Y selon la BAN.
Code_postal_(brut)	Oui	Numérique	-	DPE	-	Code postal brut.
N°_étage_appartement	Non	Numérique	-	DPE	-	Numéro d'étage de l'appartement.
Nom_résidence	Non	Texte	-	DPE	-	Nom de la résidence, si applicable.
Complément_d'adresse_bâtiment	Non	Texte	-	DPE	-	Complément d'adresse du bâtiment, si applicable.
Cage_d'escalier	Non	Texte	-	DPE	-	Cage d'escalier, si applicable.
Complément_d'adresse_logement	Non	Texte	-	DPE	-	Complément d'adresse du logement, si applicable.
Statut_géocodage	Non	Texte	-	DPE	-	Statut du géocodage de l'adresse.
Nom__commune_(Brut)	Oui	Texte	-	DPE	-	Nom de la commune brut.
N°_département_(BAN)	Oui	Numérique	-	DPE	-	Numéro de département selon la BAN.
N°_région_(BAN)	Oui	Numérique	-	DPE	-	Numéro de région selon la BAN.

Conso_5_usages_é_finale	Oui	Numérique	-	DPE	-	Consommation finale des 5 usages (chauffage, éclairage, eau chaude sanitaire, refroidissement, auxiliaires).
Conso_5_usages/m²_é_finale	Oui	Numérique	-	DPE	-	Consommation finale des 5 usages par mètre carré.
Conso_chauffage_é_finale	Oui	Numérique	-	DPE	-	Consommation finale de chauffage.
Conso_chauffage_dépensier_é_finale	Non	Numérique	-	DPE	-	Consommation finale de chauffage pour un usage dépensier.
Conso_éclairage_é_finale	Oui	Numérique	-	DPE	-	Consommation finale d'éclairage.

Annexe ENEDIS :

Identifiant BDD	Donnée obligatoire	Type de donnée	Description
Année	Oui	Numérique	L'année de la consommation d'énergie.
Code IRIS	Oui	Texte	Le code IRIS (Ilots Regroupés pour l'Information Statistique) qui est un découpage du territoire français en unités de taille homogène.
Nom IRIS	Non	Texte	Le nom de l'IRIS.
Numéro de voie	Non	Numérique	Le numéro de la voie de l'adresse.

Indice de répétition	Non	Texte	Un indice supplémentaire pour l'adresse (par exemple, bis, ter, etc.).
Type de voie	Non	Texte	Le type de voie de l'adresse (par exemple, rue, avenue, etc.).
Libellé de voie	Non	Texte	Le nom de la voie de l'adresse.
Code INSEE de la commune	Oui	Texte	Le code INSEE de la commune.
Nom de la commune	Non	Texte	Le nom de la commune.
Segment de client	Oui	Texte	Le segment du client (dans ce cas, résidentiel).
Nombre de logements	Oui	Numérique	Le nombre de logements à l'adresse.
Consommation annuelle totale de l'adresse (MWh)	Oui	Numérique	La consommation d'énergie totale de l'adresse pour l'année en mégawattheures (MWh).
Consommation annuelle moyenne par logement de l'adresse (MWh)	Non	Numérique	La consommation d'énergie moyenne par logement à l'adresse pour l'année en MWh.
Consommation annuelle moyenne de la commune (MWh)	Non	Numérique	La consommation d'énergie moyenne de la commune pour l'année en MWh.
Adresse	Non	Texte	L'adresse complète.