

# Projet Energie - Rendu FINAL

Cohorte : Bootcamp Data Analyst mai 2025

Date : 30/06/2025

Groupe de travail : Asmae AZIZI, Nicolas JACQUET, Nicolas TRIOLET, Laurie BESINET

Mentor Projet : Aléaume DRUEL

---

## Contexte

Dans un contexte de dérèglement climatique, les enjeux de la transition énergétique et sa mise en œuvre suscitent beaucoup d'interrogations. L'objectif de ce projet est d'analyser la consommation et la production énergétique à l'échelle nationale et régionale, puis d'apporter un focus sur les énergies renouvelables.

Le tableau de bord interactif que nous avons créé pour illustrer ces problématiques est destiné à un public de décideurs (type députés ou sénateurs). Cela doit leur permettre de saisir l'essentiel du sujet afin de prendre les décisions stratégiques adéquates, pour prioriser des investissements dans les différentes filières ou régions.

Dans ce document nous avons décrit les étapes par lesquelles nous sommes passées pour répondre à cette problématique. Nous commencerons par analyser le jeu de données, puis nous présenterons une première exploration grâce à de la data visualisation. Nous décrirons ensuite le travail de pre-processing des données, qui nous permet d'obtenir un jeu de données "propre". Ce dernier sera utilisé dans des visualisations avancées et nous décrirons le processus de création d'un tableau de bord interactif.

NB : Un mode d'emploi est disponible en annexe avec les indications pour initialiser les fichiers.

## Sommaire

Contexte.....	1
Sommaire.....	2
<b>Introduction.....</b>	<b>3</b>
Objectif.....	3
Cadre.....	3
<b>Présentation du jeu de données.....</b>	<b>5</b>
Présentation des variables.....	5
Volumétrie.....	6
<b>Pre-processing et feature engineering.....</b>	<b>7</b>
Doublons.....	7
Cohérence du type des données.....	7
Nettoyage des lignes et colonnes inutiles.....	7
Gestion des valeurs manquantes.....	7
Valeurs aberrantes.....	8
Ajout de colonnes pertinentes.....	9
Optimisation des données.....	9
<b>Visualisations et Statistiques.....</b>	<b>10</b>
Statistiques descriptives.....	10
Evolution de la production et de la consommation.....	12
Répartition régionale.....	14
Qualification du mix énergétique.....	17
Focus sur les énergies renouvelables.....	20
<b>Conception du tableau de bord.....</b>	<b>24</b>
Import des données.....	24
Modélisation du dataset.....	24
Étapes de transformation Power Query.....	26
Mise en place de mesures avec le langage DAX.....	28
Elaboration du rapport dynamique.....	29
Optimisation du rapport.....	33
<b>Conclusion.....</b>	<b>34</b>
Bilan projet.....	34
Difficultés rencontrées.....	34
Atteinte des objectifs.....	34
Compléments d'analyse.....	35
Améliorations.....	35
Interprétation.....	35
<b>Bibliographie.....</b>	<b>37</b>
<b>Annexes.....</b>	<b>38</b>
Liste des fichiers joints.....	38
Mode d'emploi.....	39

# Introduction

---

Le jeu de données étudié concerne la production et la consommation d'énergie en France, ventilée par régions et filières de production (nucléaire, thermique, éolien, solaire, etc.).

## Objectif

Le projet vise à améliorer la compréhension de la production et de la consommation par région, par exemple pour aider à prendre des décisions stratégiques. On cherchera notamment à :

- Identifier les filières sous ou sur-utilisées afin d'optimiser les investissements en infrastructure,
- Détecter des écarts structurels (régions en surplus ou déficit d'énergie) pour renforcer l'autonomie énergétique régionale,
- Identifier les opportunités d'optimisation énergétique et augmenter la part d'EnR (énergies renouvelables),
- Anticiper la consommation afin de produire au plus proche du besoin sans surplus.

## Cadre

### Jeu de données utilisé

Le jeu de données analysé est mis à disposition par RTE (Réseau de Transport d'Électricité). Cet organisme est le gestionnaire du réseau de transport d'électricité français. Sa mission est d'optimiser le système électrique français, opérer la transition énergétique et éclairer les décisions des pouvoirs publics. RTE a créé l'outil [éCO2mix](#) pour donner un accès facile aux données de production et consommation d'énergie. C'est à partir de cet outil que les données sont extraites.

Le jeu de données est accessible en open data via [ce lien](#) et présente la consommation par filière et par régions, du 1er janvier 2013 au 31 janvier 2023.

Nous verrons au fur et à mesure de ce rapport que des données complémentaires sont venues compléter notre jeu de données initial (cf récapitulatif en annexe).

### Accès aux données

Le jeu de données est sous Licence Ouverte v2.0 (Etalab), en accès libre en ligne sous différents formats (CSV, JSON, Excel, Parquet, API).

Nous avons récupéré le fichier CSV directement au lieu de passer par une récupération via API. En effet le fichier comporte les données entre le 01/01/2013 et le 31/01/2023, il est donc fixe. De plus, vu le volume du jeu de données, le temps de chargement serait bien trop long par API.

## Organisation du travail collaboratif

Réalisé dans le cadre de notre formation de Data Analyst chez DataScienTest, ce projet nous permettra de mettre en application les compétences que nous avons acquises tout au long de la formation :

- Traitement de données massives et hétérogènes (valeurs manquantes, erreurs de format, traitement des valeurs aberrantes ....),
- Analyse temporelle (évolution horaire et/ ou journalière),
- Visualisation de données énergétiques selon la consommation et production
- Utilisation de bibliothèques Python (Pandas, Matplotlib, Seaborn).

Pour ce projet collaboratif, nous avons mis en place un dossier partagé sur google drive dans lequel est stocké le fichier CSV source et les Notebooks de travail. Nous avons chargé le dataset et les librairies (pandas, numpy, seaborn, matplotlib) sur google colab dans un Notebook partagé. Nous avons aussi mis en place un gitHub du projet afin de versionner régulièrement notre avancement (accessible [via ce lien](#)).

Au sein de l'équipe, une seule personne a une connaissance métier du secteur de l'énergie, et nous n'avons pas eu l'opportunité de prendre contact avec des experts du métier. Néanmoins la question est primordiale et il est intéressant de savoir comment la France gère l'énergie. Le projet représente donc une bonne opportunité pour mieux connaître le secteur tout en mettant en pratique nos nouvelles compétences en data.

# Présentation du jeu de données

---

Ce chapitre vise à présenter les variables qui composent le jeu de données et la volumétrie de celui-ci.

## Présentation des variables

Les variables sont décrites ci-dessous regroupées en 4 thématiques :

### **Des données de contexte géographique et temporel :**

- Les données sont ventilées par région française, 12 au total : Île-de-France, Centre-Val de Loire, Bourgogne-Franche-Comté, Normandie, Hauts-de-France, Grand Est, Pays de la Loire, Bretagne, Nouvelle-Aquitaine, Occitanie, Auvergne-Rhône-Alpes, Provence-Alpes-Côte d'Azur ;
- La date et l'heure du relevé ;
- La nature, qui indique si les données ont été consolidées (vérifiées et complétées) ou définitives (lorsque les partenaires ont transmis et vérifié les données).

### **Les informations sur la consommation d'énergie :**

- La consommation totale en Mégawatts (MW), qui indique la puissance instantanée ;
- la consommation des pompes dans les Stations de Transfert d'Energie par Pompage (STEP) en MW. Les STEP sont des installations qui puisent aux heures creuses de l'eau dans un bassin inférieur afin de remplir une retenue en amont. L'eau sera ensuite turbinée aux heures pleines pour produire de l'électricité.

### **La production d'énergie par filière :**

- Les filières représentées sont : Thermique (usines de gaz et charbon), Nucléaire, Éolien, Solaire, Hydraulique, Bioénergies (usines d'incinérations de déchets, méthanisation, bois), Éolien terrestre, Éolien offshore ;
- Pour chacune de ces filières il est indiqué :
  - sa production en Mégawatts (MW), qui indique la puissance instantanée ;
  - le Taux de COuverture (TCO) en % : part de la production de cette filière dans la consommation de la région ;
  - le Taux de CHarge (TCH) en % : volume de production de cette filière par rapport à sa capacité de production.

### **Le stockage de l'énergie et les échanges :**

- Le solde des échanges avec les régions limitrophes en MW ;
- le stockage et déstockage batterie.

## Volumétrie

Le jeu de données initial comporte 2 121 408 lignes et 32 colonnes. Il représente 312 Mo à l'import brut en format csv. La période couverte va de janvier 2013 à janvier 2023 avec une mesure toutes les 30 min pour chacune des régions.

Nous avons fait une première analyse pour identifier les colonnes avec des valeurs manquantes, représentées sur ce graphe (*échelle logarithmique*) :



On voit déjà que les colonnes de contexte géographique et temporel sont bien remplies. Puis on identifie 6 colonnes qui contiennent exactement le même nombre de NA (de *Bioénergies (MW)* à *Hydraulique (MW)* dans le graphe). Cela pourrait indiquer des similitudes entre ces colonnes, que l'on pourra traiter ensemble par la suite.

De même, les colonnes de *TCH Bioénergies (%)* à *TCO Nucléaire (%)* dans le graphe ont exactement le même nombre de manquants, très élevé ici (1 472 256).

Globalement on voit qu'un nombre conséquent de colonnes ont plus de 50% manquants (à partir de la colonne *TCH Bioénergies (%)* dans le graphe). Pour qu'elles soient exploitables il faudra les traiter.

Pour conclure cette partie, on peut voir que ce jeu de données est conséquent mais qu'il y a un grand nombre de valeurs manquantes, ce qui pourrait nous gêner dans l'analyse.

# Pre-processing et feature engineering

---

Ce chapitre décrit le processus de traitement du jeu de données que nous avons réalisé sur Python (avec l'aide de la librairie Pandas notamment). Il s'agit de nettoyer les données afin d'être sûr que l'interprétation qui suit ne sera pas faussée, et de le compléter si nécessaire.

Les fichiers Jupyter Notebooks fournis en annexe contiennent le code détaillé.

## Doublons

Nous n'avons pas trouvé de doublons donc il n'y a pas eu de traitement nécessaire.

## Cohérence du type des données

Nous avons converti les colonnes suivantes

- date, heure et date-heure au format datetime adapté ;
- code INSEE en *string* car cet identifiant ne sert pas pour des calculs.

## Nettoyage des lignes et colonnes inutiles

La colonne 30 et les 12 premières lignes ne contiennent pas de valeurs, nous les supprimons.

On constate que le fichier contient des valeurs pour les années complètes de 2013 à 2022, mais l'année 2023 ne comporte qu'un mois. Cela peut fausser les analyses qui suivront car cette année-là ne sera pas représentative. Nous supprimons les valeurs de cette année, soit 17 856 lignes.

*Note : Cette étape est supprimée dans la 2ème phase du projet car nous serons amenés à compléter les données avec un deuxième jeu de données, donc nous aurons besoin de ces valeurs pour qu'il n'y ait pas de trou dans les données.*

La colonne Nature apporte peu d'informations car il n'y a que 2 valeurs possibles, *définitive* ou *consolidée* et cette dernière représente moins de 1%. Nous la supprimons.

## Gestion des valeurs manquantes

Il y a des valeurs 'Nd' et des '-' présentes dans le jeu de données, on précise lors du chargement du CSV qu'elles sont à considérer en que valeurs manquantes.

Comme vu plus tôt, certaines colonnes présentent un taux de valeurs manquantes particulièrement élevé :

- En regardant de plus près les colonnes *Stockage batterie*, *Déstockage batterie*, *Eolien terrestre* et *Éolien offshore* on voit qu'elles contiennent que des NaN et des 0. Ces colonnes n'apportent aucune information donc on les supprime.
- Les colonnes de *TCH Bioénergies (%)* à *TCO Nucléaire (%)* dans le graphe ont exactement le même nombre de manquants, très élevé ici (1 472 256). En regardant la répartition des valeurs vides dans le temps, on constate que les colonnes qui concernent le TCO et le TCH sont totalement vides avant 2020. Ces mesures ont dû être ajoutées au jeu de données à ce moment-là. Nous choisissons tout de même de conserver ces colonnes et de remplacer les valeurs vides par 0.

Pour les valeurs vides qui restent, nous avons appliqué une interpolation linéaire temporelle sur toutes les colonnes numériques (après les avoir triées dans l'ordre chronologique et regroupées par région). En effet pour des données chronologiques cela reste cohérent avec une évolution progressive dans le temps comme c'est le cas ici.

## Valeurs aberrantes

Les colonnes TCO (%) et TCH (%) de chaque filière expriment des taux en pourcentages. Ces valeurs ne peuvent donc pas être négatives, on vérifie qu'il n'y en ait pas.

Le TCO représente la couverture des besoins d'une région par une filière donnée dans cette même région. Il peut dépasser 100 quand la filière produit plus que la région ne consomme d'énergie.

Le TCH en revanche doit rester sous 100 car il représente la production de la filière par rapport à sa capacité. Le tableau ci-dessous montre le nombre de valeurs au-dessus de 100 pour chaque colonne :

Nom de la colonne	Nombre de valeurs > 100
TCH Nucléaire (%)	1530
TCH Thermique (%)	0
TCH Eolien (%)	695
TCH Hydraulique (%)	89
TCH Solaire (%)	1
TCH Bioénergies (%)	74

Les valeurs qui dépassent 100 sont remplacées par 100.

Concernant la production, on considère qu'elle ne peut pas être négative. Le tableau ci-dessous montre le nombre de valeurs au-dessous de 0 dans les colonnes de production :

Nom de la colonne	Nombre de valeurs < 0
Thermique (MW)	78327
Nucléaire (MW)	456
Eolien (MW)	619
Solaire (MW)	33437



Hydraulique (MW)	8
Bioénergies (MW)	0

Les valeurs inférieures à 0 sont remplacées par 0 dans ces colonnes.

## Ajout de colonnes pertinentes

On ajoute les colonnes suivantes, qui seront pertinentes pour l'analyse :

- création de colonnes temporelles : *Heure*, *Mois*, *Année* et *Année-mois* ;
- ajout d'une colonne *PROD* qui totalise les valeurs de productions des filières ;
- création d'une colonne *CONSO* qui somme la consommation de la région et l'énergie utilisée pour le pompage (en valeur absolue).

**Le jeu de données est maintenant nettoyé (lignes et colonnes inutiles supprimées), corrigé (types et valeurs aberrantes), sans doublons sans valeurs manquantes (corrigées par interpolation ou remplissage par 0).**

## Optimisation des données

Comme nous l'avons vu, les relevés des données sont effectués toutes les 30 min. Un pas de temps horaire suffira pour l'analyse ultérieure. Pour ce faire nous regroupons les données par *Code INSEE région*, *région*, *date* et nous agrégeons les colonnes quantitatives (somme pour les quantités en MW et moyenne pour les pourcentages).

**On finit cette phase de nettoyage en générant un nouveau fichier csv "propre" que nous pourrions exploiter dans les prochaines étapes.**

Pour conclure cette partie de pre-processing des données, on notera que la qualité du jeu de données laisse à désirer. Malgré un prétraitement il reste une grande part de valeurs nulles pour certaines catégories de données. Par exemple, entre 2021 et 2025 quasiment la moitié des valeurs sont nulles pour la consommation des régions Ile-de-France, Haut-de-France, Nouvelle-Aquitaine, Normandie, Centre-Val-de-Loire et Pays de la Loire.

Par manque de temps nous ne pourrions pas tout reprendre mais il serait pertinent de trouver un jeu de données complémentaire pour fiabiliser les analyses.

## Visualisations et Statistiques

Cette phase d'analyse a pour but de nous permettre de dégager des tendances et de formuler des hypothèses de travail.

Nous comparerons la consommation et la production dans le temps et par régions, puis nous analyserons la composition du mix énergétique, et enfin nous zoomerons sur les filières de production d'énergies renouvelables.

### Statistiques descriptives

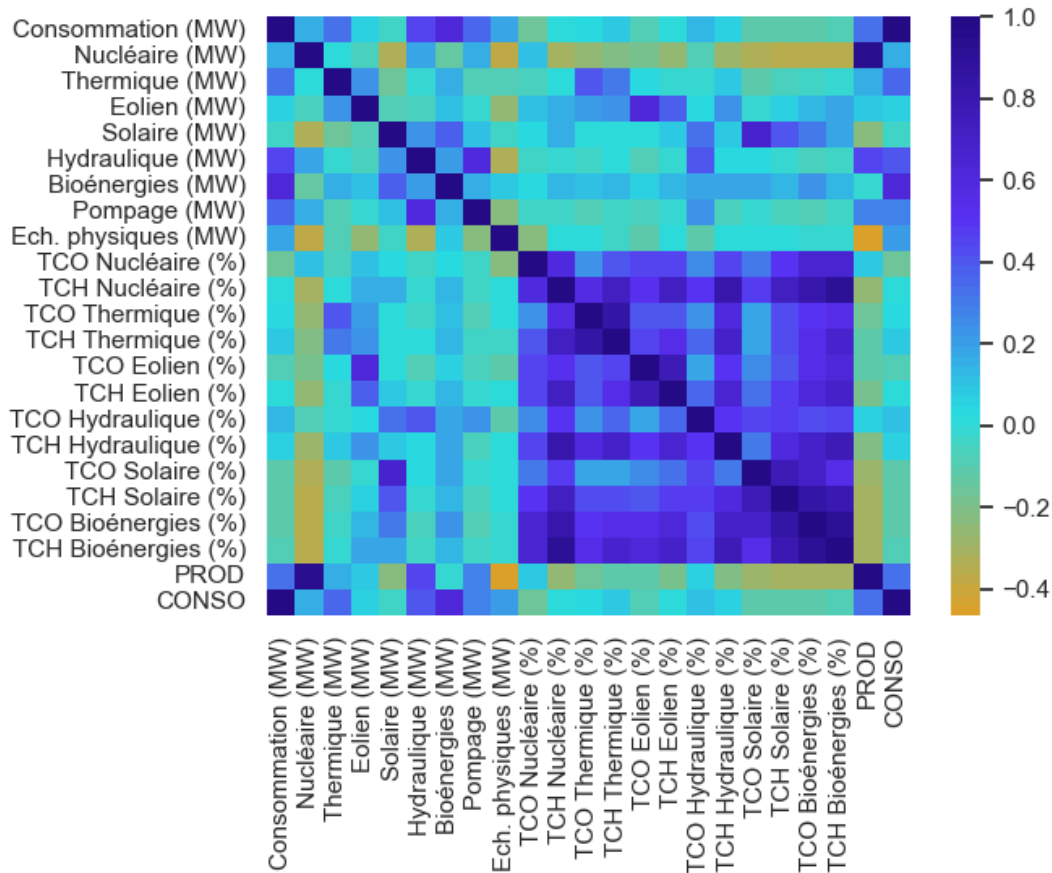
Le tableau suivant montre les statistiques des variables numériques :

	mean	std	min	25%	50%	75%	max
Consommation (MW)	214414.70	101102.91	59592.00	131977.75	196810.00	275905.25	656890.00
Nucléaire (MW)	<b>283949.52</b>	<b>159269.70</b>	0.00	164010.62	294206.23	411308.79	651959.00
Thermique (MW)	18284.33	24950.62	0.00	1267.75	8093.50	24374.25	183730.00
Eolien (MW)	12603.57	<b>19704.85</b>	0.00	1994.00	6028.50	14918.50	230038.00
Solaire (MW)	4666.12	6320.41	0.00	805.00	2039.00	5951.00	51206.00
Hydraulique (MW)	28248.56	<b>46455.78</b>	0.00	294.00	3439.00	41859.25	281005.00
Bioénergies (MW)	4071.84	2154.63	236.00	2356.00	3692.50	5613.00	14916.00
Pompage (MW)	3938.80	7905.11	0.00	0.00	278.85	4793.25	75233.00
Ech. physiques (MW)	<b>-21188.16</b>	<b>204246.58</b>	-533963.00	-199393.75	25573.00	104621.75	585849.00
TCO Nucléaire (%)	31.01	77.73	0.00	0.00	0.00	0.00	664.94
TCH Nucléaire (%)	18.63	30.06	0.00	0.00	0.00	46.73	100.00
TCO Thermique (%)	2.53	6.57	0.00	0.00	0.00	0.45	48.44
TCH Thermique (%)	6.87	14.96	0.00	0.00	0.00	1.47	89.86
TCO Eolien (%)	2.82	7.48	0.00	0.00	0.00	0.79	102.19
TCH Eolien (%)	7.35	15.46	0.00	0.00	0.00	6.69	99.93
TCO Hydraulique (%)	3.15	9.42	0.00	0.00	0.00	0.07	80.33
TCH Hydraulique (%)	8.15	15.30	0.00	0.00	0.00	9.87	86.76
TCO Solaire (%)	1.05	2.75	0.00	0.00	0.00	0.39	30.04
TCH Solaire (%)	4.17	7.67	0.00	0.00	0.00	5.03	34.22
TCO Bioénergies (%)	0.72	1.20	0.00	0.00	0.00	1.62	6.28
TCH Bioénergies (%)	16.65	26.30	0.00	0.00	0.00	43.54	83.14
PROD	351823.95	174931.74	6183.00	224407.24	350575.23	464394.10	935484.00
CONSO	210475.91	98626.56	55129.00	130474.00	194317.00	269903.50	656650.00

Les différentes filières de production présentent de grandes disparités, par exemple la moyenne du nucléaire très élevée et les valeurs assez dispersées, les énergies éolien et

hydraulique très dispersées aussi, ... Nous analyserons la distribution des productions en détail plus loin. On voit aussi que les échanges physiques moyens sont négatifs. Cela signifie qu'il y a beaucoup d'exportations, mais on voit aussi que cette colonne présente de grandes variations.

**La matrice de corrélation des valeurs numériques est présentée ci-dessous :**



Elle montre en bleu foncé qu'il y a des corrélations fortes à modérées entre :

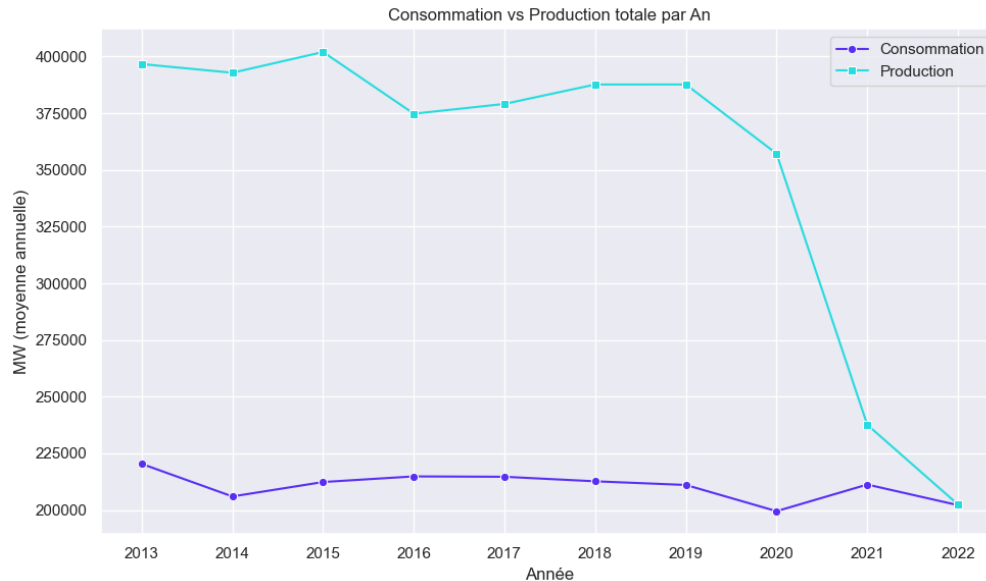
- la production totale et la production nucléaire, et hydraulique dans une moindre mesure, ce qui pourrait être expliqué par l'importance de ces deux filières de production dans la production totale ;
- le TCO éolien et la production éolienne ou le TCO nucléaire et la production nucléaire, ce qui s'explique par le fait que lorsque la production augmente, le taux de couverture des régions par cette filière est meilleur ; les paires TCO - TCH de la plupart des filières sont corrélées, pour la même raison.

L'éolien et le solaire, des énergies renouvelables, ne sont quasiment pas corrélés. Cela peut signifier qu'elles sont indépendantes et aléatoires, fortement influencées par la météo.

Globalement, les corrélations sont plutôt faibles dans notre jeu de données, ce qui pourrait témoigner de la diversité et la complémentarité des sources d'énergies.

## Evolution de la production et de la consommation

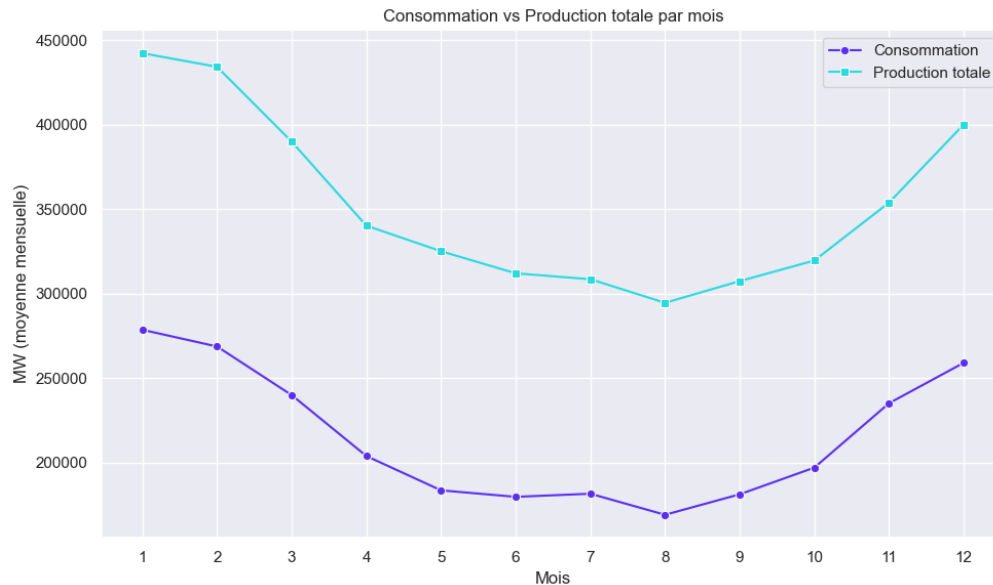
L'objectif de cette partie est de comprendre les variations de ces deux indicateurs clés (Consommation et production) à différentes échelles temporelles (an, mois, heure) et d'identifier les relations éventuelles.



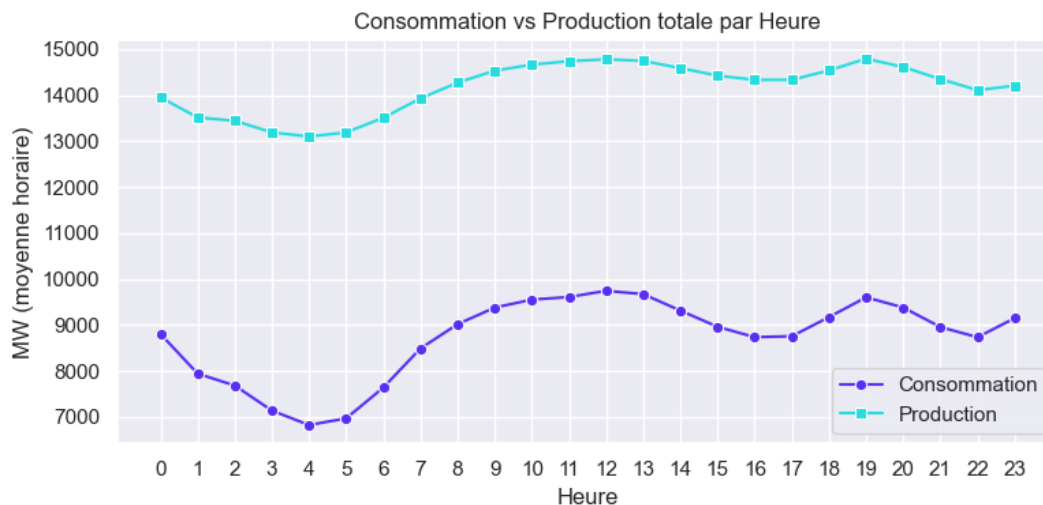
Ce graphique montre au premier abord que la France a une production largement supérieure à sa consommation jusqu'en 2020. Le pays est en effet le plus gros exportateur d'électricité en Europe.

On constate une première baisse de la consommation et de la production en 2020, qui peut être expliquée par la crise du Covid-19. En effet cette année-là les confinements et la mise à l'arrêt d'activités économiques énergivores (notamment la sidérurgie, la construction automobile, le transport ferroviaire) ont impacté les besoins à la baisse. L'activité de production ralentie et la fermeture de la centrale nucléaire de Fessenheim expliquent aussi cette baisse.

La production continue de chuter largement à partir de 2020. Elle baisse de 30% en 2021 puis de 15% supplémentaires en 2022. Cette production historiquement basse peut être expliquée par plusieurs événements concomitants, notamment l'indisponibilité de plusieurs réacteurs nucléaires, la production hydraulique impactée par les sécheresses sévères et le contexte géopolitique tendu. Nous approfondirons l'évolution de la production pour chaque filière plus tard.



Les courbes de consommation et de production suivent une tendance saisonnière parallèle. La consommation est au plus haut pendant les mois d'hiver, et baisse en été. Cela suit notamment les besoins d'énergie pour le chauffage et l'éclairage qui sont plus élevés en hiver. Les moyennes de production et de consommation au mois d'août sont particulièrement basses. La période de congés et le ralentissement des industries énergivores explique peut-être la baisse de la demande.



La consommation augmente fortement le matin (entre 6h et 9h), se stabilise en journée, puis atteint un second pic en début de soirée (entre 18h et 21h) et redescend pendant la nuit.

Cela correspond au rythme de vie des foyers et des entreprises.

Ces trois graphes montrent que des tendances de périodes de consommation se dégagent, expliquées par le rythme de vie des consommateurs, les saisons ou la météo, et à plus large échelle le contexte géopolitique.

La connaissance des tendances de consommation permet d'ajuster la production. Ceci est d'autant plus important que l'électricité produite est difficilement stockée. Il faut donc

produire juste ce qu'il faut, tout en gardant suffisamment de marge pour parer à un pic de consommation imprévu ou un problème technique. Ce couplage entre la production et la consommation apparaît bien sur nos graphes.

### Test de corrélation

Le test de corrélation de Pearson nous permet d'évaluer la force de la relation entre la consommation et la production :

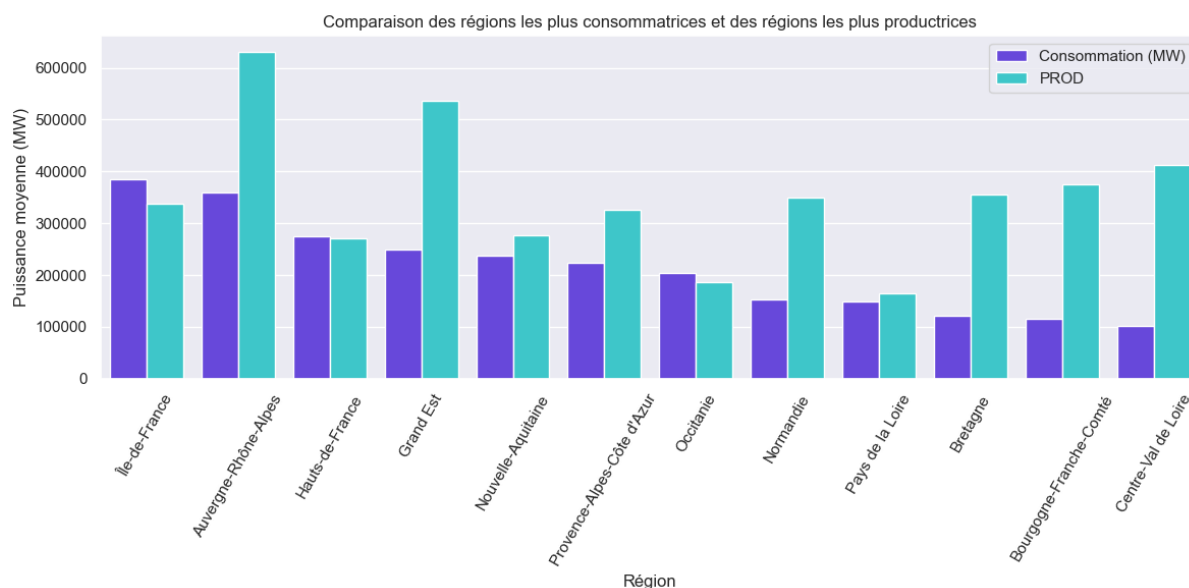
- La corrélation globale ( $r = 0.35$ ) est faible. Elle peut être influencée par des facteurs non modélisés.
- Même avant 2020, la relation linéaire reste modeste ( $r = 0.37$ ), ce qui suggère que la production n'est pas uniquement pilotée par la consommation.
- En revanche, les corrélations par mois ( $r = 0.99$ ) et par heure ( $r = 0.97$ ) montrent que la production suit très précisément la consommation à court terme, ce qui est cohérent avec un pilotage opérationnel à la demande.

## Répartition régionale

La production et la consommation sont fournies pour chaque région de France dans notre jeu de données. Dans ce chapitre nous analyserons leur répartition et formulerons des hypothèses pour expliquer les différences entre les régions.

À noter que les données présentées dans ces graphiques couvrent toutes les années du jeu de données.

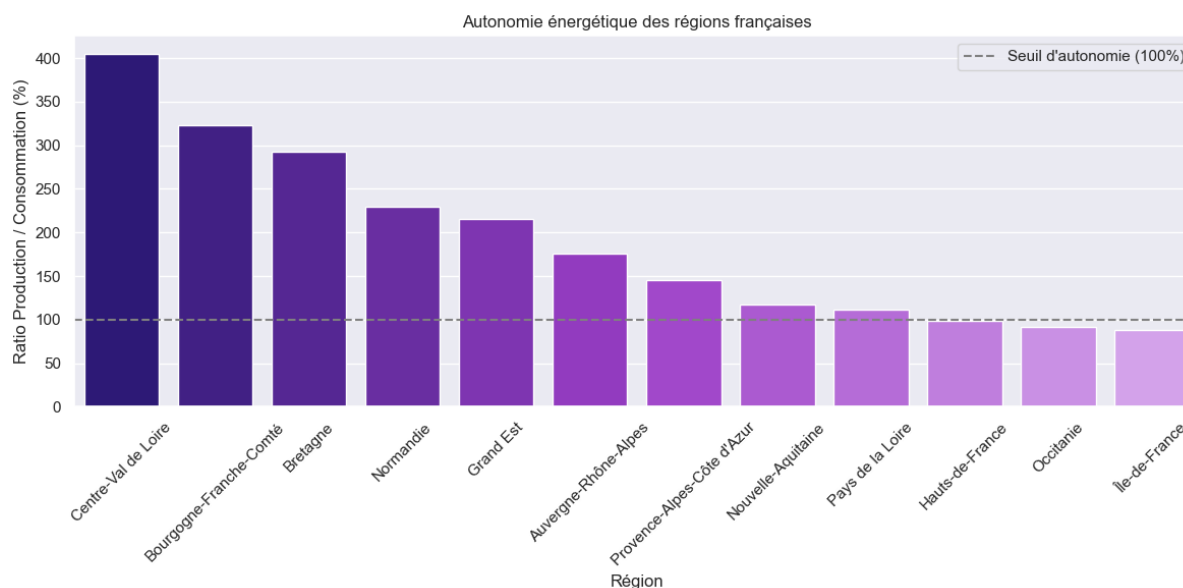
### Production et consommation par région :



La région qui consomme le plus est l'**Ile-de-France**, suivie d'**Auvergne-Rhône-Alpes**. Celles qui consomment le moins sont la **Bretagne**, la **Bourgogne-Franche-Comté** et le **Centre-Val de Loire**.

La production la plus élevée est en **Auvergne-Rhône-Alpes**, suivie du **Grand-Est** et du **Centre-Val de Loire**. Quant aux productions les plus basses, elles se situent en **Occitanie** et en **Ile-de-France**.

A partir de ces valeurs, on peut calculer l'autonomie énergétique des régions françaises :

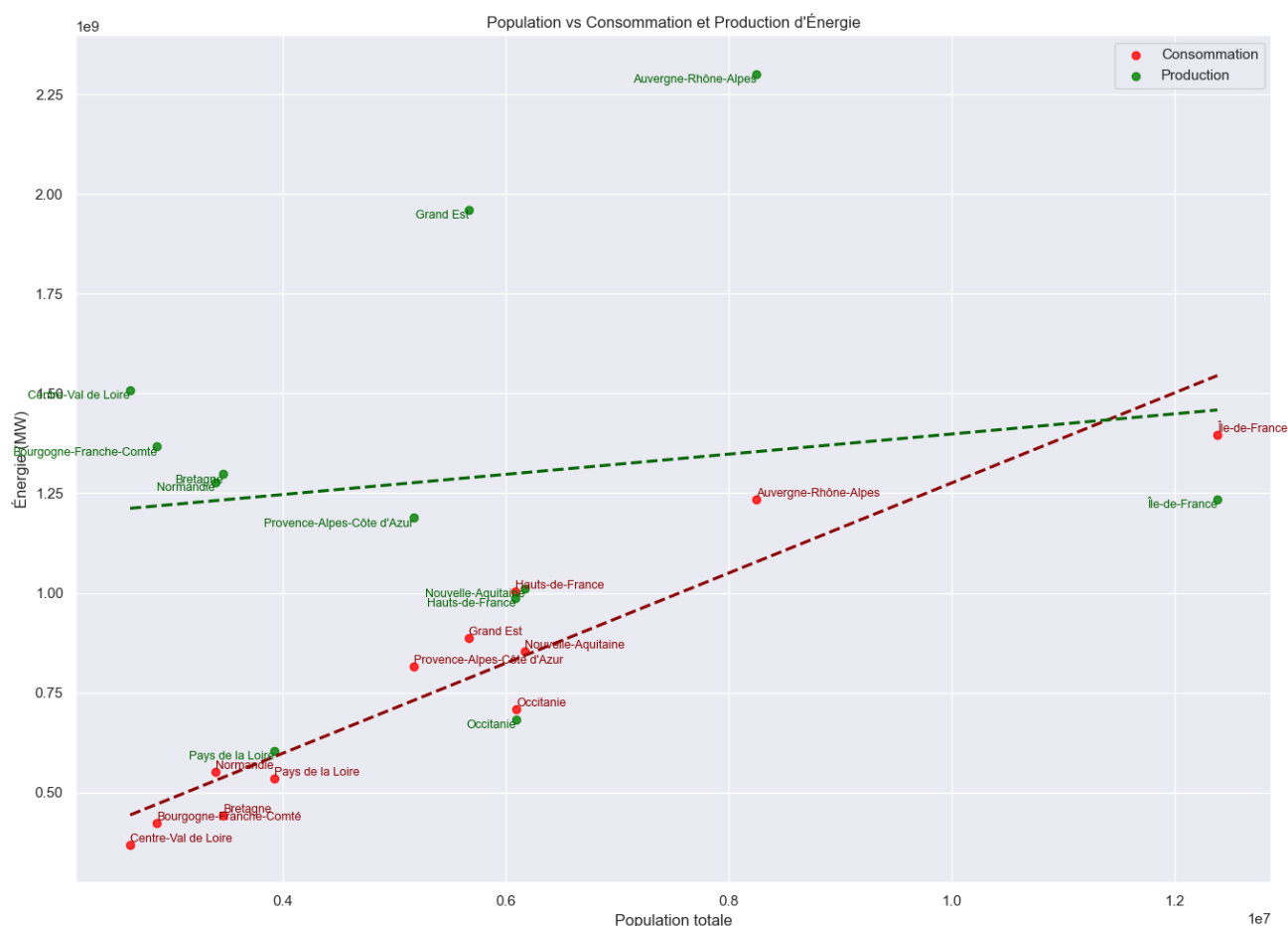


Le **Centre-Val de Loire** et la **Normandie** produisent largement plus que leurs besoins, avec un ratio de couverture supérieur à 400 % et 300% respectivement. En revanche, l'**Ile-de-France** et l'**Occitanie** ont un ratio inférieur à 100% et dépendent donc des apports extérieurs pour couvrir leur consommation.

Ceci nous laisse supposer une corrélation entre le nombre d'habitants et la consommation. Pour comparer ces deux mesures, nous importons le jeu de données de la Population des régions 2012-2023 généré par l'OFGL à partir de données INSEE ([disponible ici](#)).

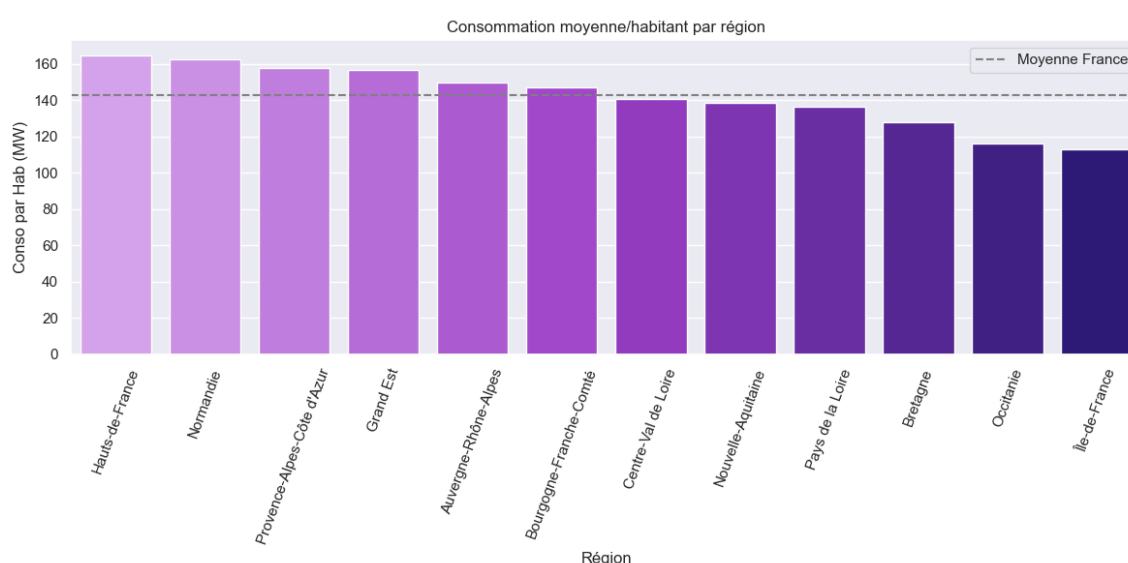
Avant 2015, les données régionales présentent des incohérences, probablement liées au redécoupage administratif ayant réduit le nombre de régions métropolitaines de 22 à 13. Le jeu de données brut pose donc problème pour les années antérieures à 2015. En croisant nos deux sources de données, nous obtenons un suivi continu à partir de 2015. L'analyse portera ainsi sur la période allant de 2015 à 2022. Les codes régionaux ont été regroupés et ajustés afin de correspondre aux régions actuelles.

Le graphique suivant révèle un décalage important entre production (points verts) et consommation (points rouges) selon les régions :



Ce graphique nous montre que la consommation et la population sont corrélées, mais pas la production car elle dépend d'autres facteurs : présence de centrales, ressources naturelles (barrages, vent, soleil), politiques publiques.

On peut maintenant calculer la consommation moyenne par habitant dans chaque région :



On retrouve maintenant l'**Occitanie** et l'**Île-de-France** à droite du graphe. Au final, ce sont les régions ayant la consommation par habitant les plus basses. Cette information n'était pas évidente dans les premiers graphiques de ce chapitre.



A l'inverse, les régions **Haut-de-France**, **Auvergne-Rhône-Alpes**, et **Normandie** s'avèrent être celles qui ont la consommation par habitant les plus élevées.

On voit donc que la densité de population a un faible impact sur la consommation par habitant, ce n'est pas les régions les plus peuplées qui consomment le plus par habitant. On peut émettre quelques hypothèses pour expliquer les différences de consommation par habitant entre les régions :

- un climat plus froid en hiver pour les régions montagneuses ou au nord,
- la présence d'industries énergivores,
- l'impact du tourisme dans certaines régions,
- le statut de carrefour de circulation à l'échelle européenne des Hauts-de-France.

Ces hypothèses pourraient être approfondies à l'aide de jeux de données pertinents, mais nous ne le ferons pas ici.

Ce chapitre a mis en lumière les disparités de consommation et production entre régions :

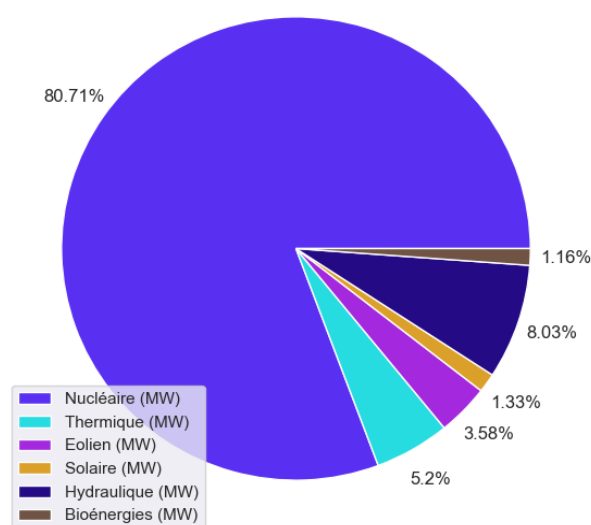
- Certaines régions, comme l'Île-de-France, sont très vulnérables car fortement dépendantes des autres pour leur approvisionnement,
- D'autres régions, peu peuplées mais fortement équipées, jouent un rôle stratégique dans l'équilibre énergétique du pays,

Cette connaissance est cruciale pour répartir la production d'énergie à l'endroit où elle est nécessaire et ainsi limiter les risques de black out.

Le chapitre suivant s'intéressera à la répartition de la production dans les différentes filières.

## Qualification du mix énergétique

Ce chapitre vise à analyser la répartition de la production au sein des différentes filières et leurs particularités.



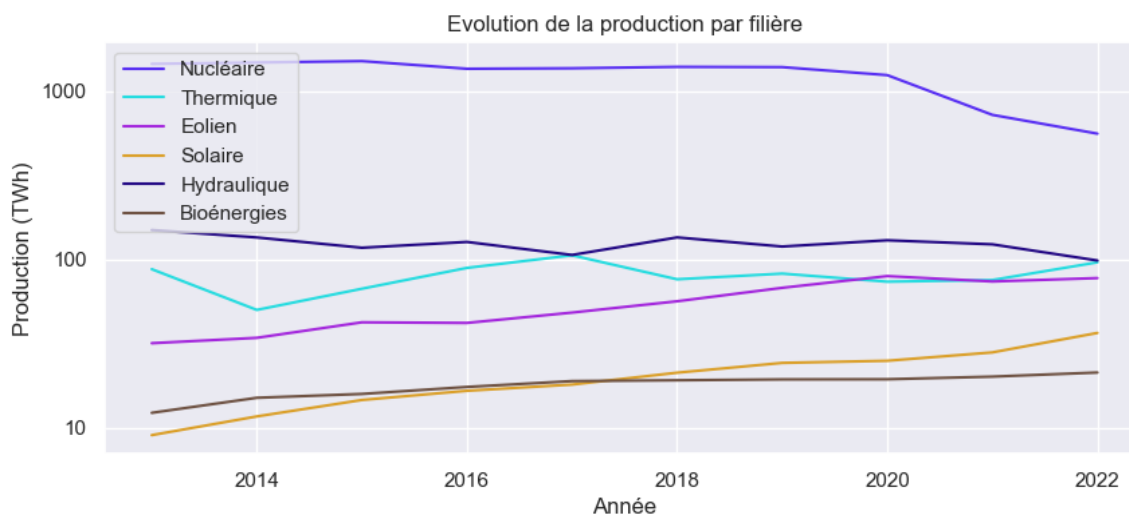
Le premier graphique montre la part de chaque filière (nucléaire, hydraulique, solaire, etc.) dans la production d'énergie sur la période 2012-2023 :

Ce graphe nous montre que la production électrique Française repose principalement sur le **nucléaire**, qui représente 80 % du mix.

Les énergies **hydraulique** et **thermique** ont un rôle de soutien important et représentent ensemble 13% de la production.

L'**éolien**, le **solaire** et les **bioénergies** sont faiblement présents, ils représentent 6% de la production.

Regardons si cette répartition évolue au fil des ans (*échelle logarithmique*) :



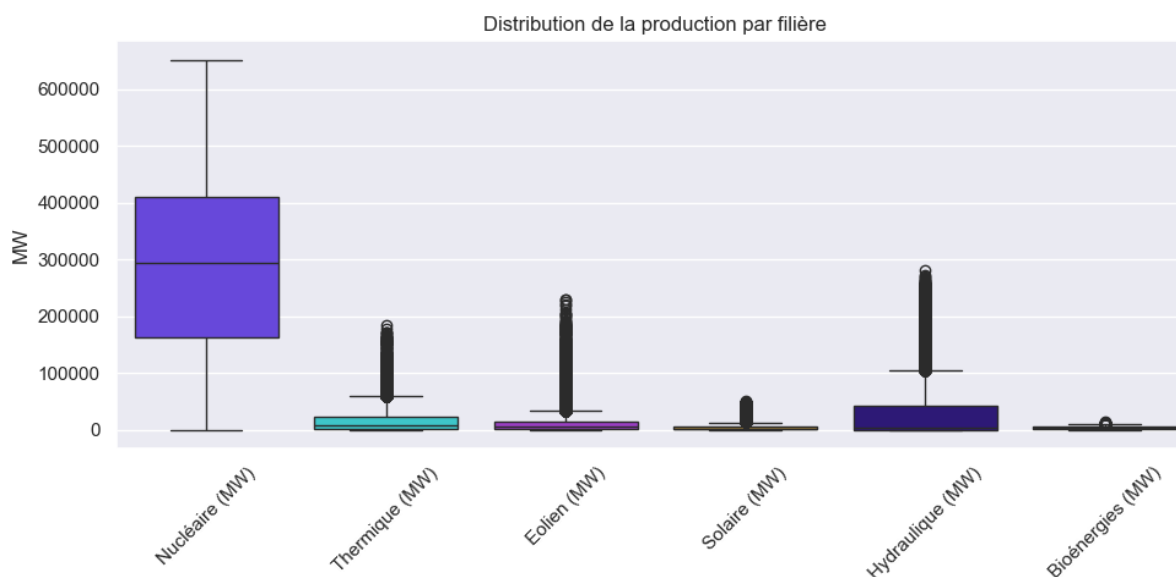
Comme on l'a évoqué plus haut, la production **nucléaire** a été fortement impactée en 2020 par la crise du Covid-19 et la fermeture de la centrale nucléaire de Fessenheim, puis en 2021 et 2022 du fait de l'indisponibilité de plusieurs réacteurs nucléaires. La production atteint un record historiquement bas en 2022 avec 63% de la production totale. La filière reste néanmoins la première source d'énergie en France.

L'énergie **hydraulique** est la 2ème filière, mais elle peut être impactée par des années de sécheresses (en 2017 et 2022). Ces années-là, la filière thermique vient compenser.

La filière **thermique** est la 3ème la plus productive, à l'exception de 2020 où l'éolien lui est passé devant. Même si les émissions de gaz à effet de serre et les tarifs du charbon et gaz sont dissuasifs, cette filière est nécessaire pour soutenir les productions d'énergies nucléaire et renouvelables, comme 2017 et 2022.

Les productions des filières **éolienne**, **solaire** et **bioénergies** augmentent au fil des ans, particulièrement le solaire qui est passé devant les bioénergies en 2018.

Le graphique qui suit montre la distribution des valeurs de production de chaque filière :



Ce graphique, qui confirme la position dominante du nucléaire que l'on a vue précédemment, apporte des informations complémentaires :

Il montre une forte dispersion des valeurs pour le **nucléaire**, ce qui signifie une production qui varie beaucoup, mais sans valeurs extrêmes. On pourrait s'attendre à plus de stabilité de cette filière pilotable. Cela peut être dû à des périodes de maintenance ou bien à la modulation de la production en fonction du besoin.

Les valeurs pour les filières **hydraulique** et **thermique**, présentent des valeurs extrêmes assez élevées, ce qui correspond à de forts pics de production. On peut émettre l'hypothèse qu'elles sont utilisées pour compléter la filière nucléaire. En effet, elles se pilotent avec plus de flexibilité qu'une centrale nucléaire qui est difficile à démarrer ou arrêter.

Les filières **solaire** et **éolienne** présentent aussi de forts pics de production. Cela peut s'expliquer par la dépendance aux conditions météorologiques, qui peuvent varier fortement.

La filière **bioénergie** est celle qui présente la production moyenne la plus basse mais aussi la plus constante.

Le graphe ci-dessous montre l'évolution du taux de charge (TCH) moyen de chaque filière au fil des mois entre 2020 et 2022 (il n'y a pas de valeurs avant 2020) :



Des tendances saisonnières se détachent, mais elles sont différentes selon les filières :

Le TCH **nucléaire** monte chaque hiver, lorsque les besoins d'énergie augmentent. Un pic de 84 % a été atteint en janvier 2021. Le TCH baisse ensuite avec l'arrivée de journées plus longues et plus douces, jusqu'à passer en dessous de 40% en juillet 2022.

Le TCH **thermique** augmente aussi en hiver pour compléter le nucléaire, avec des valeurs autour de 40%. Il descend fortement au printemps-été avec un minimum en avril 2020 à 4%. On remarque que la baisse n'a pas été aussi forte à l'été 2022 où la filière a été plus mobilisée que d'autres années pour compenser la faible production nucléaire.

Le TCH **éolien** est plus irrégulier, car dépendant de la météo, avec un pic à 47% en février 2020. Il a tendance à baisser en été, jusqu'à 11% en juin 2021.

Le TCH **hydraulique** est assez élevé en hiver, pour accompagner la hausse de la consommation. Il subit un creux en fin d'été lors des mois les plus chauds, ce qui peut signifier que les retenues d'eau sont à sec.

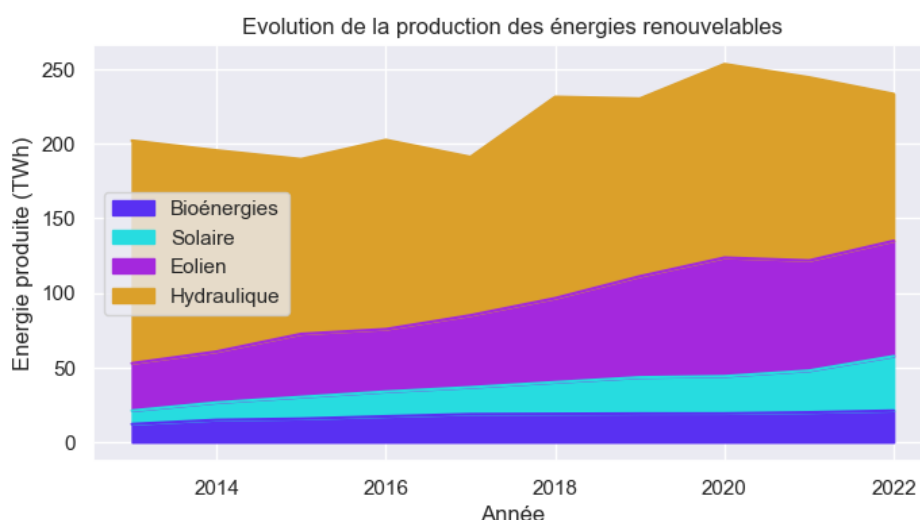
Le TCH **solaire** augmente principalement au printemps et en été. Avec les journées plus longues et la météo plus clément, il passe fréquemment au-dessus de 20%. Il descend fortement en hiver, en dessous de 10%, avec un minimum de 3,3 % en décembre 2022.

Le TCH des **bioénergies** est le plus stable, avec des valeurs comprises entre 43 et 60%.

## Focus sur les énergies renouvelables

Bien que la production d'énergie en France soit dominée par le nucléaire, les objectifs d'augmentation de la part d'énergies renouvelables de la France sont ambitieux : 33 % de production d'électricité de source renouvelable en 2030. Ce chapitre se concentrera donc sur la production d'**énergies renouvelables** (hydraulique, éolien, solaire et bioénergies)

Ce premier graphique montre l'évolution de la production de chaque filière au fil des ans :

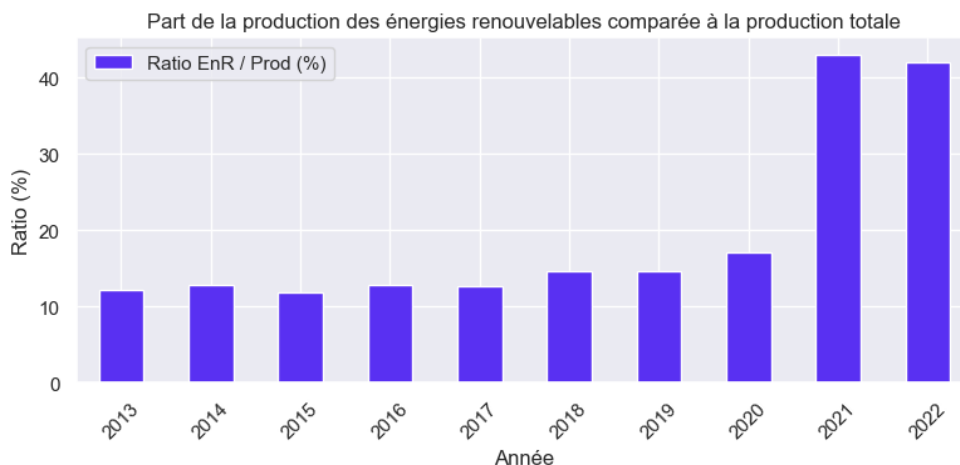


En moyenne l'hydraulique est le plus présent avec plus de la moitié de la production, suivi de l'éolien (25%), du solaire (9 %) et des bioénergies (8%).

La tendance générale de production d'énergies renouvelables est à la hausse :

- la production de **bioénergies** reste relativement stable avec une légère augmentation au fil des ans ;
- la production d'énergie **solaire** s'accroît progressivement, avec une augmentation significative de 30% entre 2021 et 2022 ;
- l'**éolien** est en croissance aussi, malgré des conditions défavorables en 2021 et 2022;
- l'**hydraulique** est plus instable, elle est impactée par les années de sécheresses.

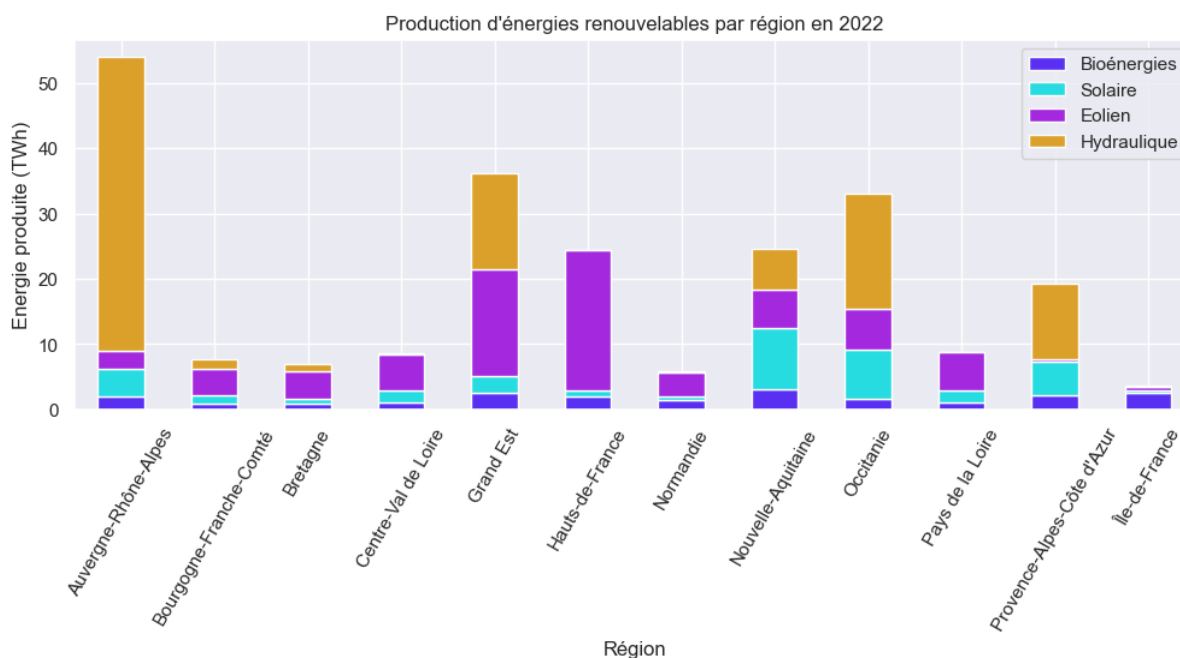
Le graphique suivant précise la part de la production des énergies renouvelables dans la production totale du pays :



La part des énergies renouvelables augmente petit à petit de 2017 à 2020 (de 12% à 17%), puis elle s'envole en 2021 et 2022 en dépassant les 40%.

Ce pic est à modérer : comme on l'a vu précédemment, la production d'énergie nucléaire et la production totale a été particulièrement basse ces deux années-là, ce qui fait particulièrement augmenter le ratio.

Se pose maintenant la question de la localisation des énergies renouvelables. Le graphe suivant présente la production d'énergies renouvelables par région sur la dernière année du jeu de données, **en 2022** :



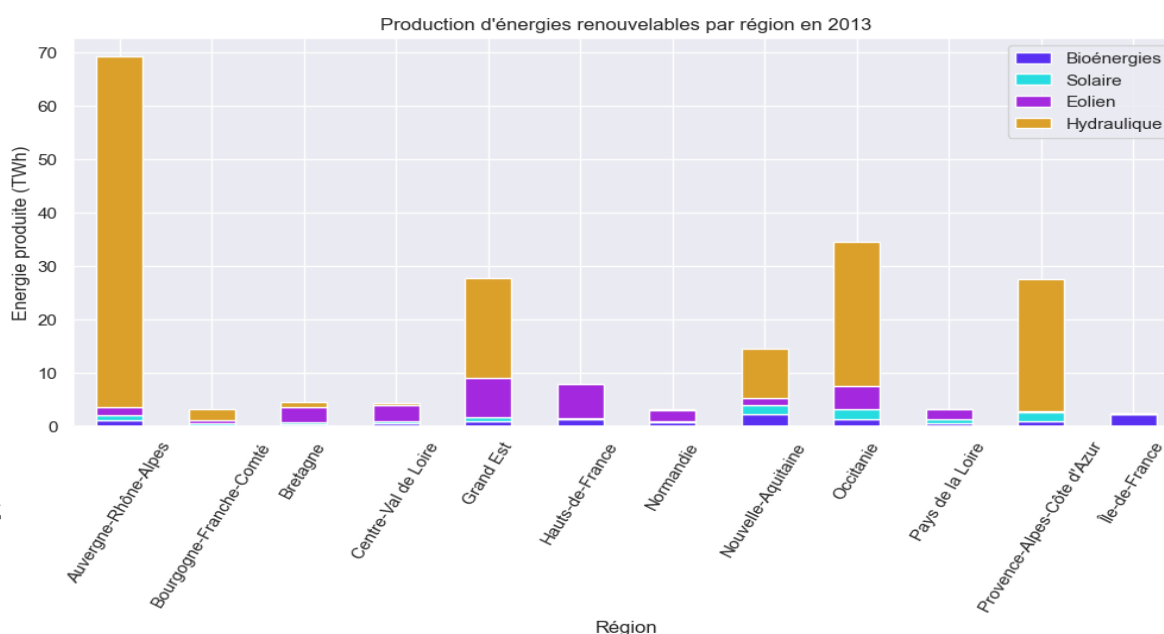
Ce graphique montre les disparités entre les régions. La première région productrice est **Auvergne-Rhône-Alpes** (plus de 50 TWh), puis **Grand-Est** et **Occitanie** (plus de 30 TWh). Ces trois régions présentent une forte proportion de production **hydraulique**.

La production **éolienne** est fortement concentrée dans deux régions : **Grand-Est** et **Haut-de-France**.

Quant à la production **solaire**, elle est plutôt présente dans les régions **Nouvelle-Aquitaine**, **Occitanie** et **Provence-Alpes-Côte d'Azur**. Ceci peut s'expliquer par la localisation de ces régions, les plus au sud de la métropole.

L'**Ile-de-France** et la **Normandie** sont les régions qui produisent le moins d'énergies renouvelables.

Pour mettre en perspective l'évolution de la répartition régionale des énergies renouvelables, nous avons comparé **l'année 2022 à l'année 2013**, le graphique suivant présente la production des énergies renouvelables pour **l'année 2013** :



En 2013, la production est très majoritairement **hydraulique**, concentrée dans quelques régions comme **Auvergne-Rhône-Alpes, Grand Est et Occitanie**. À cette époque, l'éolien et le solaire restent peu développés, et leur contribution au mix énergétique est marginale.

Cette comparaison met en lumière une **évolution notable de la structure de production** : entre 2013 et 2022, la part des énergies renouvelables **s'est diversifiée** avec une montée en puissance visible de l'éolien (notamment dans le Grand Est et les Hauts-de-France) et du solaire (dans le sud de la France).

Elle illustre également une **meilleure répartition territoriale**, bien que des disparités régionales persistent.

On a vu dans ce chapitre que les énergies renouvelables s'installent petit à petit, encore de façon inégale. Elles ont montré leur capacité à soutenir la production les années où le nucléaire vacille.

Elles sont toutefois soumises aux aléas climatiques ce qui les rends moins fiables et plus compliquées à piloter.

## Conception du tableau de bord

---

Ce tableau de bord a vocation à fournir une vision synthétique mais approfondie des dynamiques énergétiques, afin de favoriser des décisions éclairées en matière de politique énergétique.

### Import des données

L'importation des données constitue une étape essentielle, réalisée dans Power Query afin de garantir un nettoyage structuré dès l'origine du processus.

Pour intégrer les données issues de l'application éCO2mix, nous avons utilisé la méthode d'importation par **dossier (Folder)**. Deux fichiers CSV y sont regroupés :

- Un premier fichier couvrant la période de **2013 à 2023**,
- Un second fichier, contenant les données de **2023 à aujourd'hui**, a été téléchargé manuellement depuis le site open data de RTE via [ce lien](#), afin de compléter le jeu initial. Ce fichier permet de garantir une couverture temporelle à jour. Le format CSV a été choisi pour faciliter son intégration avec le premier fichier, déjà au même format, et ainsi permettre une fusion cohérente dans Power Query

L'utilisation d'un dossier permet de centraliser les fichiers dans un répertoire unique (**c:\ENERGIE\data**) et d'automatiser l'importation, rendant le processus dynamique : tout ajout futur de fichier dans ce dossier sera automatiquement pris en compte par Power BI.

Ces fichiers sont ensuite combinés automatiquement via Power Query. Un nettoyage est appliqué (normalisation des types, renommage des colonnes, filtrage, etc.), suivi d'une **agrégation par jour** pour optimiser les performances et la lisibilité.

En complément, un fichier CSV distinct nommé **population.csv** (situé à **c:\ENERGIE\population.csv**) est importé pour enrichir les analyses. Ce fichier contient les données de population par région, ce qui permet d'introduire des indicateurs relatifs (par habitant) dans le tableau de bord final.

Ces différentes sources sont ensuite transformées en tables prêtes à l'analyse, puis fusionnées lorsque nécessaire pour assurer la cohérence des visualisations.

### Modélisation du dataset

Afin de structurer les données de manière optimale pour l'analyse et la visualisation, nous avons adopté une **modélisation en étoile**. Celle-ci permet de séparer les données en une



**table de faits** centrale et plusieurs **tables de dimensions** associées. Cela améliore la lisibilité, la performance et la flexibilité des visualisations dans Power BI.

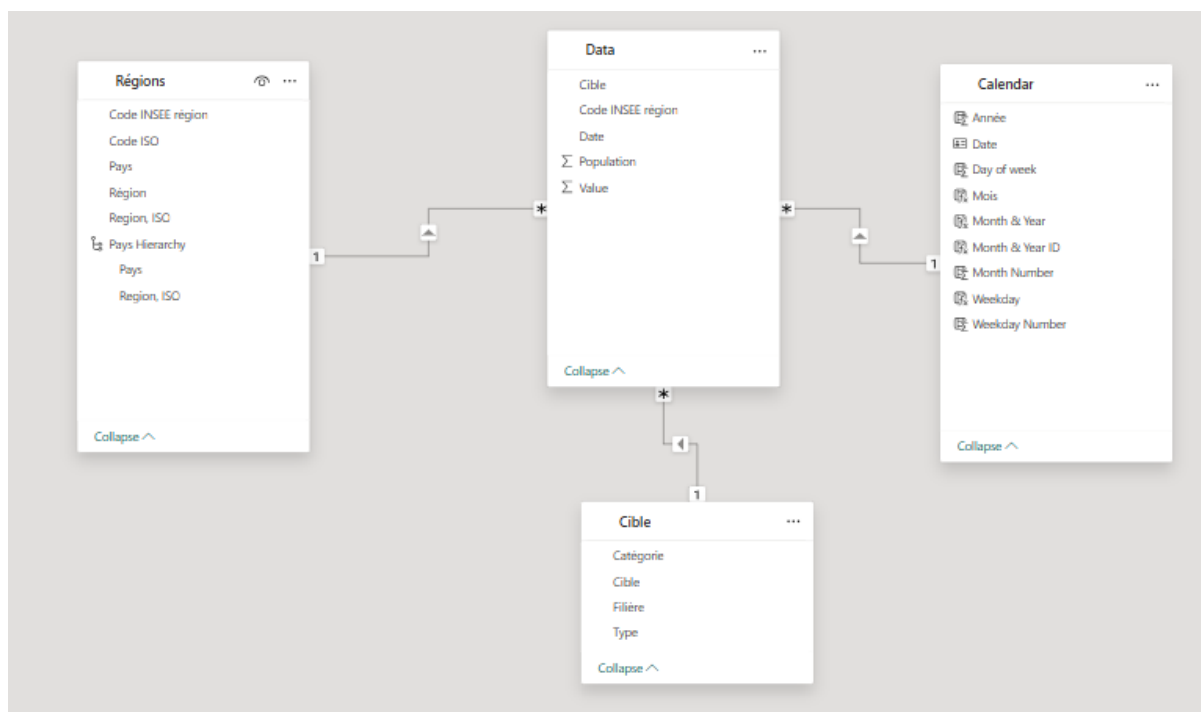
- **Data**, qui sera la table de fait, avec les colonnes suivantes :
  - Date : la date de la mesure au format JJ/MM/AAAA,
  - code INSEE région : identifiant unique de la région (clé de jointure avec la table *Regions*),
  - cible : type de mesure réalisée (clé de jointure avec la table *Cible*),
  - value : la mesure,
  - Population : population associée à la région et à la date;
- **Régions**, la table de dimension qui rassemble les informations sur les régions :
  - Région : le nom de la région,
  - Code INSEE région : l'identifiant de la région (clé primaire),
  - Code ISO : les codes ISO de certaines région ont été nécessaires pour que certains visuels puissent bien interpréter la région concernée,
  - Region, ISO :concaténation des noms de région et codes ISO,
  - Pays : pays associé à la région ;
- **Cible**, la table de dimension centralise les métadonnées relatives aux types de mesures énergétiques. Elle permet d'interpréter correctement les valeurs chiffrées présentes dans la table de faits (**Data**) et de faciliter la création de visualisations dynamiques par catégorie, type ou filière. Elle est composée des colonnes suivantes:
  - Cible : Nom complet de la mesure telle qu'elle apparaît dans les données sources (ex. Consommation (MW), TCO Nucléaire (%), TCH Solaire (%), etc.). Cette colonne permet une identification précise et directe de l'indicateur à analyser,
  - Catégorie :Classe la mesure selon son caractère **Thermique**, **Nucléaire** ou **Renouvelable**. Elle permet de distinguer les grandes familles d'énergie,
  - Type : répartition des catégories de mesures selon si c'est la consommation, la production, les échanges, les taux de charges ou les taux de couvertures,
  - Filière : indique de quelle filière est issue la mesure (thermique, nucléaire, éolien, solaire, hydraulique ou bioénergies).

Une fois les tables de faits et de dimensions préparées, nous passons à la **construction du modèle relationnel** dans Power BI.

On crée une table supplémentaire en DAX, la table **Calendar** qui nous sera nécessaire pour visualiser l'évolution des valeurs à différentes échelles de temps.

Cette table **Calendar** est essentielle pour permettre une analyse temporelle structurée. Elle contient différentes colonnes dérivées (année, mois, jour, semaine, etc.) permettant d'analyser les données à plusieurs granularités de temps.

Dans la **vue modèle**, nous relierons la table de faits (**Data**) aux différentes tables de dimensions (**Régions**, **Cible**, **Calendar**) à l'aide de **relations logiques basées sur des clés primaires et étrangères**. Ci-après notre schéma en étoile :



Sur le schéma du modèle de données Power BI, on observe les relations suivantes de type **1 à \*** (**un-à-plusieurs**) entre les tables, établies comme suit :

- **Régions (1) → Data (\*)** : Relation sur le champ Code INSEE région,
- **Calendar (1) → Data (\*)** : Relation sur le champ Date,
- **Cible (1) → Data (\*)** : Relation sur le champ Cible.

Pour finir de nettoyer le jeu de données les données inutiles au créateur de tableau de bord sont masquées et les colonnes obsolètes supprimées.

Quelques améliorations supplémentaires amélioreront l'expérience d'exploration du modèle :

- création de hiérarchies dans les tables de temps (année, trimestre, mois, jour) et de région (pays, région) ;
- définition d'une colonne de tri pour les colonnes mois et jour de la semaine, afin qu'elles soient triées par leur numéro et pas par ordre alphabétique.

## Étapes de transformation Power Query

### Table **Data** :

Dans Power Query, nous avons nettoyé notre table de base en ne conservant qu'une seule colonne de date, en supprimant les autres ainsi que toutes les colonnes de production et de consommation, ces dernières étant destinées à être traitées sous forme de mesures dans Power BI. Nous avons ensuite effectué un **dépivotage** afin d'isoler les indicateurs dans une colonne "**Cible**", et leurs valeurs associées dans une autre colonne "**Value**", en modifiant le séparateur décimal pour le rendre compatible (virgule). Une colonne "**Year**" a été insérée à partir de la date, uniquement pour permettre une opération de **fusion (Merge**

**Queries)** avec la table **Population**. Une fois la colonne population récupérée, la colonne "Year" a été supprimée, puis nous avons renommé les colonnes pour assurer cohérence et lisibilité du modèle. Cette préparation garantit une structure optimisée pour les analyses temporelles et démographiques.

### Table **Régions** :

Nous avons dupliqué la requête de base afin de créer une table dédiée aux données régionales. À l'aide de l'étape **"Removed Columns"**, nous avons supprimé toutes les colonnes relatives aux cibles ainsi que celles contenant des informations de dates. Seules les colonnes essentielles ont été conservées, notamment **"Code INSEE région"** (utilisé comme clé primaire) et **"Région"**. Nous avons ensuite utilisé **"Added Custom Column"** pour créer deux nouvelles colonnes personnalisées : l'une pour le **code ISO** et l'autre pour le **pays**. Ces transformations permettent d'obtenir une table de dimension claire et normalisée, prête à être utilisée dans les modèles d'analyse géographique.

### Table **Cible** :

Nous avons dupliqué la table source afin de créer une table dédiée à l'analyse des cibles énergétiques. Après avoir appliqué un **changement de type** sur les colonnes nécessaires, nous avons utilisé l'étape **"Unpivoted Other Columns"** pour transformer les différentes mesures en une structure normalisée à deux colonnes : **"Cible"** (ancienne colonne "Attribute") et **"Value"**, facilitant ainsi les analyses dynamiques. Nous avons ensuite supprimé les colonnes non pertinentes, en ne conservant que **Cible** et **Value**. À l'aide de plusieurs étapes **"Added Custom Column"**, nous avons enrichi les données avec trois nouvelles dimensions : une colonne **"Type de cible"** (production, consommation, échanges, TCA, TCO), une **"Catégorie"** (nucléaire, renouvelable, thermique), et une colonne **"Filière"** pour identifier précisément chaque source d'énergie. Les types de données ont été ajustés et les valeurs nettoyées via des remplacements ciblés (**"Replaced Value"**) pour garantir la qualité et la cohérence du modèle.

Dans Power Query, les tables **"population"** et **"RégionsF"** ont été utilisées exclusivement pour enrichir la base principale. La table **"population"** a été nettoyée via les étapes classiques de transformation : promotion des en-têtes (**Promoted Headers**), conversion des types de données (**Changed Type**), filtrage et suppression de colonnes non nécessaires, puis tri des valeurs (**Sorted Rows**). Elle a ensuite été agrégée par regroupement pour harmoniser les doublons régionaux avant d'être fusionnée dans la table principale via une jointure (**Merge Queries**).

De son côté, la table **"RégionsF"** a été générée à partir de la source brute et transformée pour ne conserver que les colonnes pertinentes : code INSEE, nom de région, année et population. Une colonne personnalisée a été ajoutée (**Added Custom**) pour concaténer le code INSEE et l'année en vue d'une correspondance avec les autres données.

Une fois les données de population intégrées aux bonnes observations de la base, ces deux requêtes ont été **désactivées dans le modèle** afin d'optimiser la structure et éviter toute redondance.

## Mise en place de mesures avec le langage DAX

Dans Power BI, le langage **DAX** est utilisé pour créer des mesures dynamiques qui facilitent la manipulation, l'analyse et la visualisation des données dans le rapport. Ces mesures sont mises à disposition du concepteur de rapport pour répondre à des besoins précis de calculs et de filtrages.

Nous avons développé un ensemble important de mesures permettant de présenter les données sous différents angles : sommes, pourcentages, évolutions, filtrages spécifiques (par filière, région, type d'énergie, etc.), comparaisons temporelles (ex. : année précédente), et bien plus encore.

Pour plus de clarté, les mesures sont classées dans des dossiers thématiques : Autonomie, Consommation, Production, Production EnR et Autres.

### **Exemple concret de mesure** : *Consommation moyenne par habitant et par jour (KW)*

Cette mesure calcule la consommation quotidienne moyenne d'électricité par habitant, à partir des valeurs de consommation exprimées en **MW** et de la population.

Code DAX utilisé :

```
1 Consommation moyenne par hab/jour =  
2 IF (  
3     AVERAGE(Valeurs[Population]) > 0,  
4     CALCULATE(  
5         AVERAGEX(  
6             Valeurs,  
7             [Consommation]/Valeurs[Population] *1000  
8         )  
9     ),  
10    0  
11 )
```

### **Interprétation :**

- on vérifie si la population est renseignée, afin d'éviter des erreurs de division, si non la valeur sera 0 ;
- on se base sur des valeurs de *consommation* et *population* déjà calculées dans des mesures
- on divise la valeur de consommation par la population (convertie en KW) pour obtenir une moyenne par habitant.

Ce type de mesure permet des comparaisons pertinentes entre régions ou périodes.

Ci après la liste de nos mesures :

- ☐
 Consommation
    - ☐
 Consommation France
      - ☐
 Consommation moyenne par hab/jour
        - ☐
 Consommation moyenne par hab/jour France
  - ☐
 Production
    - ☐
 Production
      - ☐
 Production / Production totale France (%)
    - ☐
 Production France
      - ☐
 Production France LY
    - ☐
 Production LY
      - ☐
 Production moy par hab/jour (KW)
    - ☐
 Production Totale
      - ☐
 Production Totale France
  - ☐
 Production EnR
    - ☐
 Production EnR
      - ☐
 Production EnR (%)
    - ☐
 Production EnR (%) LY
      - ☐
 Production EnR France
    - ☐
 Production EnR France (%)
      - ☐
 Production EnR LY
    - ☐
 Production EnR toutes filières (%)

- ☐
 Autonomie
- ☐
 Autonomie
  - ☐
 Autonomie LY
- ☐
 Autonomie France
  - ☐
 Couverture besoins (%)
  - ☐
 Couverture des besoins (%) LY
- ☐
 Autres
- ☐
 1
  - ☐
 Année du visuel
  - ☐
 Dernière année de la période
  - ☐
 Moyenne
  - ☐
 Objectif taux EnR 2030 (%)
  - ☐
 Taux de Charge (%)
- ☐
 Consommation
- ☐
 Consommation
  - ☐
 Consommation France
  - ☐
 Consommation moyenne par hab/jour
  - ☐
 Consommation moyenne par hab/jour France

## Elaboration du rapport dynamique

On passe maintenant à la conception du rapport, la partie visuelle et interactive qui sera accessible au grand public.

Une bonne pratique veut que le rapport soit séparé du dataset afin de permettre à plusieurs rapport de venir s'y connecter. Etant donné qu'un seul rapport est créé avec ce dataset, et que l'outil Power services n'est pas utilisable avec un compte personnel, cette méthode ne sera pas mise en place.

Le rapport est séparé en 5 onglets, présentés ci-dessous :

## 1. Introduction

Le premier onglet décrit brièvement le tableau de bord. Il avertit aussi l'utilisateur que la qualité des données n'est pas optimale.



## 2. Général

L'onglet **Général** constitue une **vue d'ensemble** de la situation énergétique en France. Il synthétise les principales données disponibles de manière visuelle et interactive, afin de faciliter la compréhension globale des dynamiques de **consommation** et de **production**.

**KPIs clés :**

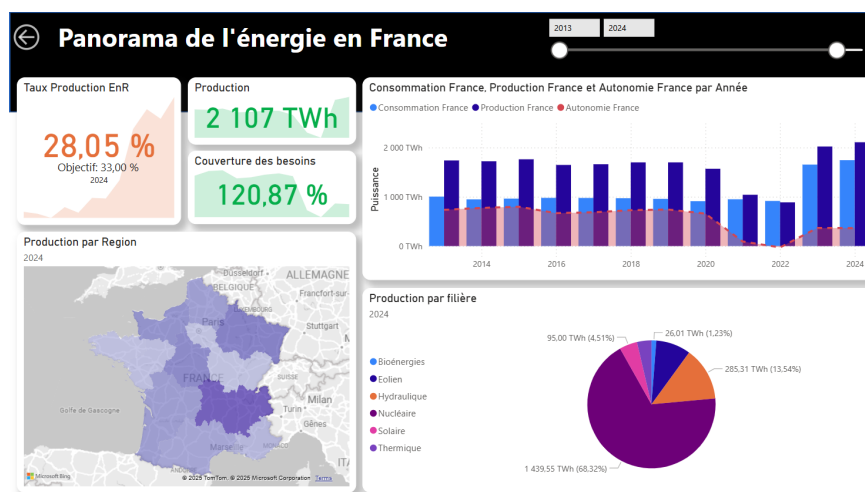
- Taux de production EnR (énergies renouvelables),
- Volume de production total en TWh,
- Taux de couverture des besoins énergétiques.

**Évolution temporelle :** Graphique illustrant la consommation, la production et l'autonomie énergétique **année par année**

**Répartition géographique :** Carte de la **production par région**, permettant une lecture territoriale des données.

**Répartition par filière :** Camembert représentant la part de chaque filière (thermique, nucléaire, hydraulique, etc.) dans la production nationale.

L'utilisateur peut **ajuster la période d'analyse** à l'aide d'un **curseur temporel** en haut de page (2013–2025).



## 3. Régions

Cet onglet permet une **analyse détaillée par région** de la situation énergétique, en offrant une vision localisée des dynamiques de production et de consommation.

**KPIs clés :**

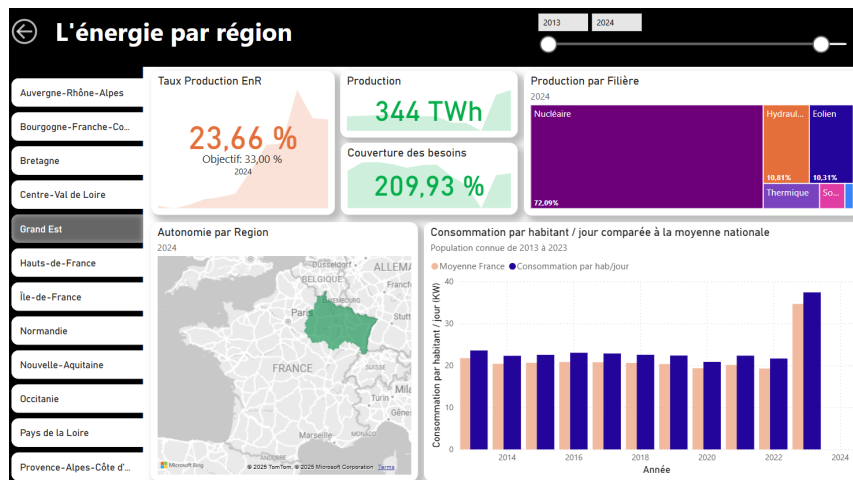
- Taux de production EnR par région,
- Production totale (en TWh),
- Taux de couverture des besoins énergétiques.

**Suivi temporel :** Graphique de l'évolution de la **consommation par habitants** comparée à celle de la France par année, filtré selon la région sélectionnée.

**Répartition géographique** : Carte de l'autonomie énergétique régionale, permettant d'identifier les zones excédentaires ou déficitaires en énergie.

**Analyse par filière** : Histogramme de la répartition de la production par type d'énergie (nucléaire, hydraulique, éolien, etc.) à l'échelle régionale.

- L'utilisateur peut sélectionner la région de son choix via un menu latéral.
- La période d'analyse peut être ajustée dynamiquement avec un curseur temporel.



#### 4. Filières

Cet onglet permet une **exploration détaillée de chaque filière de production** (nucléaire, hydraulique, thermique, éolien, solaire, bioénergies) grâce à une navigation latérale par bouton. L'objectif de cet onglet :

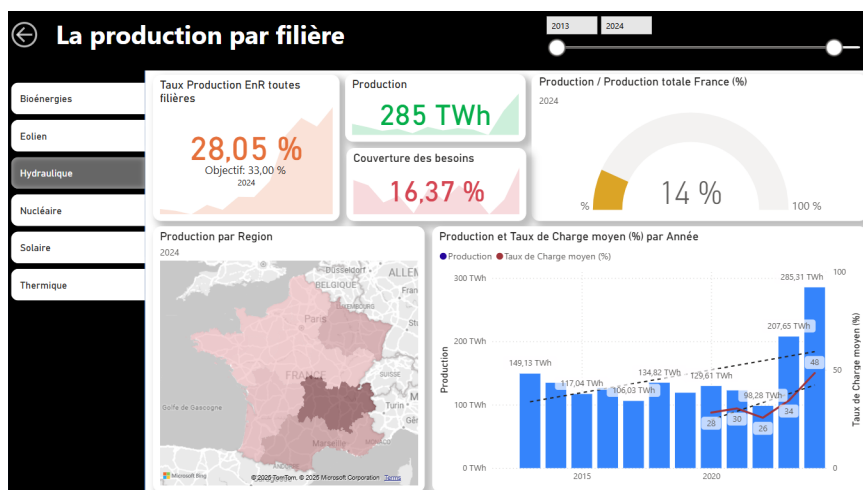
**KPIs clés** :

- Volume de production pour la filière sélectionnée,
- Taux de couverture des besoins par la filière sélectionnée,
- Part de la filière sélectionnée dans la production nationale.

Carte régionale qui affiche la **contribution par région** à la production de la filière choisie.

Graphique combiné : Courbes et barres illustrant la **production annuelle** et le **taux de charge moyen (%)**.

- L'utilisateur peut sélectionner la **filière** à analyser via le menu latéral.
- La période d'analyse peut être ajustée dynamiquement avec un curseur temporel.



## 5. Energies Renouvelables (EnR)

Cet onglet propose un **focus dédié aux énergies renouvelables** : éolien, solaire, hydraulique et bioénergies. Il permet de suivre leur **évolution**, leur **répartition géographique**, ainsi que leur **poids dans le mix énergétique français**.

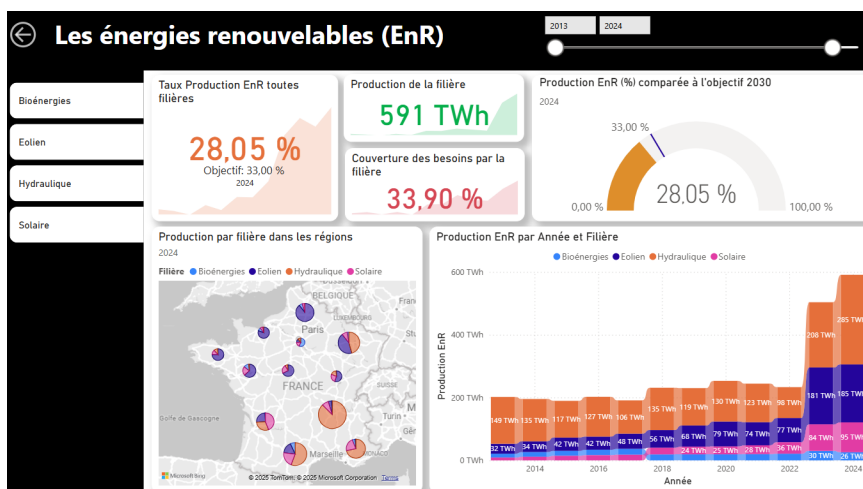
### KPIs clés :

- Taux de production EnR actuel,
- Volume total produit par la filière sélectionnée,
- Taux de couverture des besoins pour la filière sélectionnée,
- Part de la filière sélectionnée dans la production nationale (**28,44 % en 2025**), comparée à l'**objectif national de 33 % en 2030** ([source ministère de l'écologie](#)) ,

**Cartographie régionale** : Représentation de la production EnR **par filière et par région**, via des camemberts superposés sur la carte de France.

**Évolution temporelle** : Graphique montrant la **progression de chaque filière**,

- L'utilisateur peut sélectionner la **filière** EnR à analyser via le menu latéral.
- La période d'analyse peut être ajustée dynamiquement avec un curseur temporel.

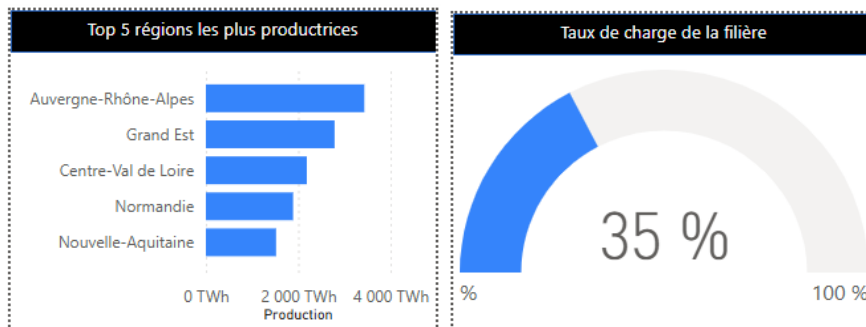




## 6. Utilisation des tooltips

Dans le but d'améliorer l'interprétation des données tout en gardant une interface épurée, nous avons mis en place des tooltips personnalisés dans Power BI. Ces éléments interactifs permettent d'afficher automatiquement des visuels complémentaires lorsque l'utilisateur survole un graphique.

- Taux de charge de la filière
- Top 5 des régions les plus productives



## Optimisation du rapport

Plusieurs axes d'optimisation du rapport ont été travaillés :

- Une attention particulière a été apportée à la mise en forme des onglets afin que l'expérience de l'utilisateur soit la plus fluide possible,
- Les couleurs sont cohérentes avec l'information véhiculée (vert pour du positif, orange ou rouge pour le négatif, différents niveaux de saturation pour les données plus ou moins élevées, ...),
- Des filtres sont à disposition de l'utilisateur afin d'accéder facilement à l'information qui l'intéresse,
- Des "drills-downs" permettent de naviguer dans différents niveaux hiérarchiques, notamment dans les visualisations temporelles, pour adapter les visuels à la précision souhaitée.

Afin d'avoir des performances intéressantes, les bonnes pratiques suivantes ont été mises en places :

- les jeux de données ont été limités à des données journalières (au lieu de la demi-heure), ce qui permet de réduire la taille du fichier et le temps de chargement,
- l'utilisation des mesures pour faire des calculs dynamiques, plutôt que les colonnes calculées.

Cela pourrait être approfondi grâce à l'outil d'analyse de performances intégré à powerBI

# Conclusion

---

La première phase de projet nous aura permis de nous organiser en tant qu'équipe, de prendre en main le jeu de données, le nettoyer et faire les premières analyses.

La seconde phase a été consacrée à la conception d'un tableau de bord interactif pour illustrer ces problématiques.

## Bilan projet

### Difficultés rencontrées

Une des principales problématiques que nous avons rencontrées est la qualité du jeu de données. En effet, il manque beaucoup de valeurs, ce qui risque de fausser nos interprétations.

Une autre limite de notre travail réside dans le manque d'expertise sur le sujet au sein de l'équipe. Un travail complémentaire avec un expert serait idéal pour aider à l'interprétation du jeu de données.

Le périmètre couvert par le jeu de données n'est pas complet :

- Toutes les régions de France ne sont pas représentées : il manque les données pour la Corse et l'outre-mer. Cela s'explique par le fait que les données viennent de l'opérateur du réseau électrique de la métropole, qui est différent dans ces régions;
- Certaines données ne sont présentes que depuis 2020, donc nous aurons moins de recul pour les analyser.
- On regrettera aussi que les données ne précisent pas les capacités installées par région, ce qui limite les possibilités de normalisation des productions par filière.

Certains d'entre nous ont eu des soucis techniques ou de performance avec leur matériel ce qui a pu freiner notre avancement.

### Atteinte des objectifs

Les objectifs de notre projet étaient de *"Constater le phasage entre la consommation et la production énergétique au niveau national et au niveau régional (risque de black out notamment)"*

- *Analyse au niveau régional pour en déduire une prévision de consommation ;*
- *Analyse par filière de production : énergie nucléaire / renouvelable ;*
- *Focus sur les énergies renouvelables (où sont-elles implantées ?).*

Nous avons répondu à cette demande dans la mesure où nous savons visualiser l'autonomie énergétique de chaque région, identifier la répartition de la production pour chaque filière et l'implantation des filières renouvelables.

En revanche nous n'avons pas travaillé la partie prévision de la consommation, n'ayant pas eu de cours à ce sujet (cette problématique s'adressait plutôt à des profils de data Scientists).

## Compléments d'analyse

Au fur et à mesure de l'analyse nous avons formulé des hypothèses de causes qui pourraient impacter la production ou la consommation d'énergie.

Dans le temps imparti nous en avons approfondie une : le lien entre la consommation d'énergie d'une région et le nombre d'habitants. Pour cela nous avons complété les données initiales avec la population des régions de 2012 à 2023, un jeu de données généré par l'OFGL à partir de données INSEE ([accessible via ce lien](#) au format CSV).

Plusieurs autres axes d'étude pourraient être analysés : la météo, la surface des régions, leurs industries, ...

## Améliorations

Concernant le jeu de données et le rapport, différentes améliorations pourraient être apportées :

- Le principal souci réside dans le manque de données, un travail pour reprendre les données nulles serait à entreprendre. L'avantage de power BI est que si le jeu de données change mais que sa structure reste la même, les visuels du tableau de bord seront mis à jour de façon automatique.
- Une fois les données mises en cohérence, la mise à disposition du data set via power BI service permettrait à un grand nombre de personnes d'y accéder pour faire des tableaux de bords,
- Il serait intéressant d'avoir les retours des utilisateurs pour identifier d'autres pistes d'améliorations.

## Interprétation

Il ressort de notre tableau de bord et de nos analyses que la production d'énergie en France reste dominée par la production nucléaire, mais les énergies renouvelables prennent leur place.

La force du parc énergétique Français réside dans sa diversité et la complémentarité des filières de production. Elles peuvent prendre le relais les unes des autres afin de suivre au mieux la demande.

Cette demande évolue dans le temps en suivant des tendances qui peuvent être exploitées afin de prédire les besoins d'énergies. On peut ainsi activer la bonne filière de production au bon moment.

Le couplage entre la consommation et la production est essentiel car l'énergie est difficilement stockable, il faut donc être en mesure de suivre le besoin au mieux pour éviter les pertes.

On voit notamment que, sauf exception, la France produit plus que nécessaire et que la production d'énergies renouvelables est en augmentation.

Mais la production d'énergies renouvelables est limitée par sa dépendance aux conditions météorologiques qui la rendent difficile à piloter. Cela s'ajoute à la complexité de stocker l'énergie, et des pics de production sont parfois perdus.

Verdir notre production d'électricité est un bel objectif, ceci étant dit, il ne faut pas oublier que l'électricité n'est qu'une petite partie de l'énergie consommée en France (27%). Le pétrole et le gaz représentent 57% de l'énergie consommée par les français en 2023, dont une grande partie est utilisée pour les transports. Il reste encore du chemin pour changer notre consommation au sens large.

# Bibliographie

---

Tout au long de ce projet, nous nous sommes appuyés sur différents articles et publications pour mieux comprendre le sujet :

**RTE, publication du 14 février 2023 - Bilan électrique 2022**

Bilan de l'année 2022 qui décrit à posteriori les événements de l'année et en tire des analyses et conclusions pertinentes.

<https://www.rte-france.com/actualites/bilan-electrique-2022>

**Ministère de l'écologie, publication mise à jour le 30 août 2024 - Les énergies renouvelables**

Panorama des énergies renouvelables en France et perspectives.

<https://www.ecologie.gouv.fr/politiques-publiques/energies-renouvelables>

**Science et Vie, article du 13 juin 2025 - Face à la surproduction d'électricité, la France cherche des solutions d'urgence**

[https://www.science-et-vie.com/nature-et-environnement/energie-nature-et-environnement/face-a-la-surproduction-delectricite-la-france-cherche-des-solutions-durgence-201968.html?utm\\_source=Facebook\\_Nonli&utm\\_medium=Social&utm\\_campaign=Nonli](https://www.science-et-vie.com/nature-et-environnement/energie-nature-et-environnement/face-a-la-surproduction-delectricite-la-france-cherche-des-solutions-durgence-201968.html?utm_source=Facebook_Nonli&utm_medium=Social&utm_campaign=Nonli)

**Electricity map**

Carte interactive qui représente les émissions CO2 dues à la consommation électrique à l'échelle mondiale.

<https://app.electricitymaps.com>

**SDES (Service des données et études statistiques) du ministère de la transition écologique, Edition 2024 - Chiffres clés de l'énergie**

Publication qui fournit des informations sur la production, la consommation et les prix pour les différentes formes d'énergie en France et dans le monde, et met en lumière les principales évolutions à l'œuvre.

<https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-energie/donnees-cles>

# Annexes

---

## Liste des fichiers joints

Les fichiers de travail que nous avons utilisés sont disponibles dans le dossier joint au rapport :

### Jeux de données brutes :

- *eco2mix-regional-cons-def.csv* : Données éCO2mix régionales consolidées et définitives (janvier 2013 à janvier 2023) ([lien](#)). Ce jeu de données fourni par RTE donne la consommation et production par filière et par régions, de 2013 à janvier 2023 ;
- *eco2mix-regional-tr* : Données éCO2mix régionales temps réel ([lien](#)). Ce sont les données de consommation et production par filière et par régions de 2023 à aujourd'hui, fournies par RTE ;
- *populations-ofgl-regions.csv* : les données de la Population des régions 2012-2023 généré par l'OFGL à partir de données INSEE ([lien](#)).

### Jeux de données transformés et nettoyés :

- *eco2mix\_clean.csv* : données de consommation et production régionales de 2013 à 2023, généré à l'issue du Notebook Jupyter *Nettoyage* ;
- *eco2mix\_clean\_2.csv* : données de consommation et production régionales de 2023 à aujourd'hui, généré à l'issue du Notebook Jupyter *Nettoyage\_2*.

### Notebooks Jupyter de manipulation et analyse des données

- *Nettoyage* : notebook qui utilise le fichier de données brutes *eco2mix-regional-cons-def.csv* et génère le fichier nettoyé *eco2mix\_clean.csv* ;
- *Nettoyage\_2* : notebook qui utilise le deuxième fichier de données brutes *eco2mix-regional-tr* et génère le fichier nettoyé *eco2mix\_clean\_2.csv* ;
- *DataViz* : notebook qui nous a permis de générer les visuels d'analyse présents dans le rapport, dans la partie Visualisations et Statistiques.

### Rapport powerBI

- *Energie* : notre rapport réalisé avec power BI, qui utilise les jeux de données nettoyés (*eco2mix\_clean.csv* et *eco2mix\_clean\_2.csv*) ainsi que *populations-ofgl-regions.csv*

## Mode d'emploi

Nous avons fourni un dossier zip ENERGIE.

Si possible, placer ce dossier ENERGIE à la racine du disque 'C:' (**c:\ENERGIE**).  
L'actualisation des données dans le rapport Power BI sera automatique.

Si ce n'est pas possible, il faudra modifier l'adresse des fichiers compris dans le sous-dossier **ENERGIE\data** dans le power BI, dans power Query.

A l'ouverture du rapport power BI, selon la langue du logiciel, il arrive qu'une erreur apparaisse à l'actualisation des données. Si cela arrive, il faut relancer l'actualisation jusqu'à 3 fois.