**THÈME : OPTIMISATION DE LA CONSOMMATION D'ÉNERGIE**

# **INTRODUCTION**

L'optimisation de la consommation d'énergie est un enjeu majeur dans le contexte actuel de développement durable et de responsabilité environnementale. Les entreprises, les gouvernements et les particuliers sont de plus en plus conscients de l'importance de gérer efficacement leur consommation d'énergie pour minimiser leur impact sur l'environnement et réduire leurs coûts. Dans ce rapport de projet Big Data, nous explorerons comment les données peuvent être utilisées pour optimiser la consommation d'énergie, en exploitant les avancées de l'analyse de données à grande échelle. Nous présenterons les objectifs, la méthodologie, les résultats et les conclusions de notre projet, mettant en lumière les opportunités et les défis de l'utilisation du Big Data pour améliorer l'efficacité énergétique et contribuer à un avenir plus durable.

# 

# **I - ANALYSE DES BESOINS**

## **1) Pourquoi optimiser la consommation d’énergie ?**

Il y a plusieurs raisons pour lesquelles il est important d'optimiser la consommation d'énergies :

* Économie d'argent : une consommation d'énergie réduite signifie une facture d'électricité moins élevée. En économisant de l'énergie, vous économisez également de l'argent.
* Protection de l'environnement : La production d'énergie, en particulier la production d'électricité, peut avoir un impact environnemental important, notamment en termes de production de gaz à effet de serre. En optimisant la consommation d'énergie, on peut réduire la quantité d'énergie produite, ce qui contribue à la réduction des émissions de gaz à effet de serre et à la protection de l'environnement.
* Réduction de la dépendance énergétique : Les pays qui dépendent fortement de l'importation de combustibles fossiles sont vulnérables aux fluctuations des prix de l'énergie. En optimisant la consommation d'énergie, les pays peuvent réduire leur dépendance énergétique et devenir plus autonomes.

* Amélioration de la qualité de vie : L'optimisation de la consommation d'énergie peut conduire à une meilleure qualité de vie. Par exemple, une maison bien isolée nécessitera moins de chauffage en hiver, ce qui peut conduire à des factures d'énergie moins élevées, mais aussi à une maison plus confortable et plus agréable à vivre.

En somme, l'optimisation de la consommation d'énergies peut avoir des avantages aussi bien économiques, environnementaux que sociaux. Il est donc important de prendre des mesures pour réduire notre consommation d'énergies et maximiser notre efficacité énergétique.

## **2) Quel est le but de notre étude ?**

Dans un premier temps, le but de notre étude est principalement d’analyser et visualiser les données de consommation d’énergie en France et dans le monde.

Dans un second temps (si le temps le permet), nous pourrons mettre en place des modèles de machine learning ayant pour but de prédire la consommation d’énergie.

## **3) Quel est le périmètre de notre étude ?**

Dans un premier temps, nous faisons une étude comparative de la consommation d’énergies entre les différents pays pour identifier quels sont les pays qui consomment le plus, mais aussi, quels sont les pays qui consomment le moins.

Nous réalisons ensuite une étude centrée cette fois sur la France, dans laquelle nous ferons une étude comparative de la consommation d’énergies entre les différentes régions, départements. Nous analyserons également la consommation d’énergie selon le secteur d’activité, le type d’énergie (électricité ou gaz), l’opérateur qui fournit l’énergie et l’année.

Nous allons également analyser l’influence de la température et de la population sur la consommation d’énergie en France, notamment d’électricité.

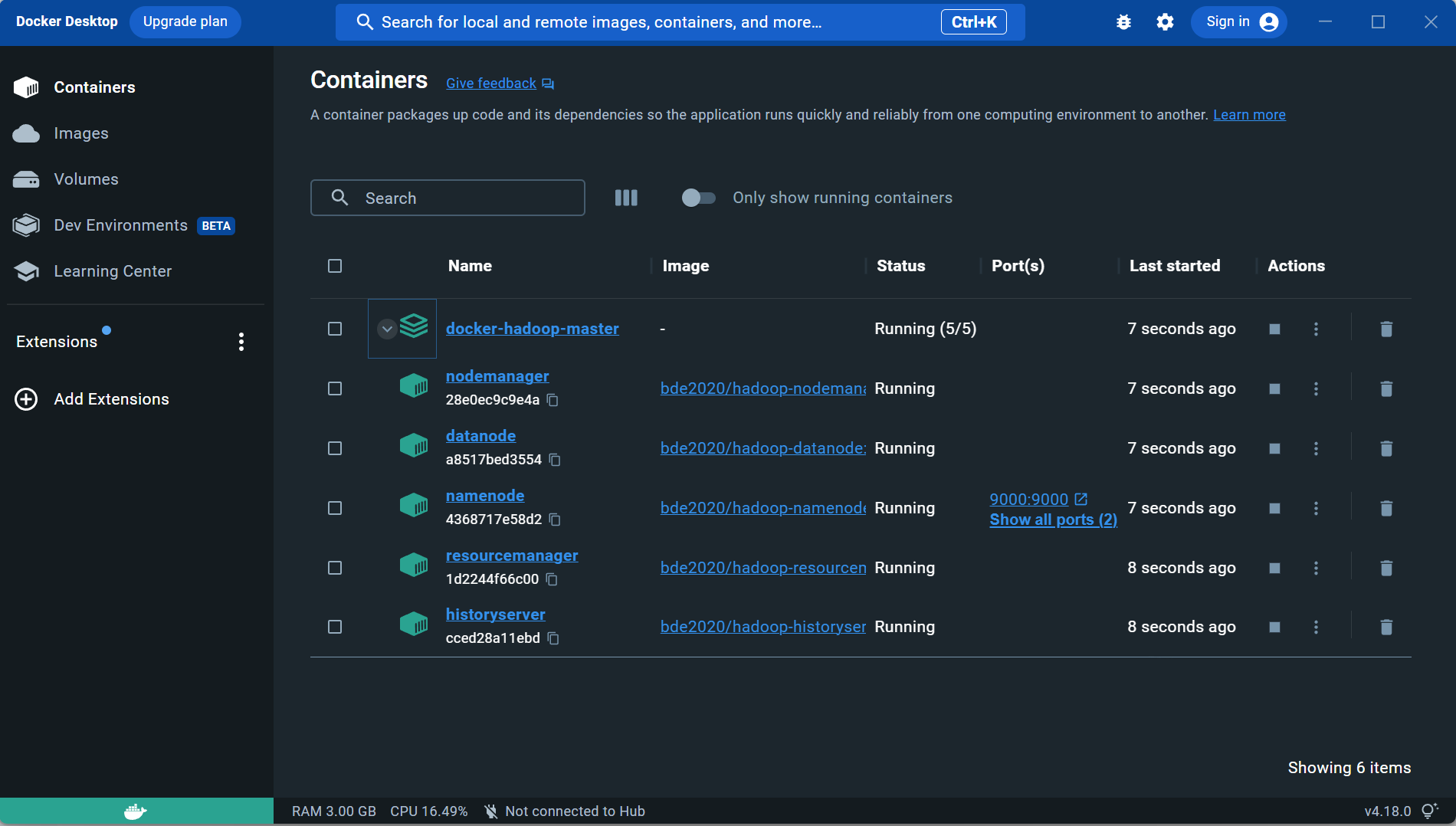
## **4) Quelles données allons-nous utiliser dans le cadre de notre étude ?**

* (données comprises entre 1946 et 2020) Consommation d'énergies des différents pays du monde <https://www.kaggle.com/code/yomnanasseryounis/energy-consumption-data-visualization-tutorial/input>
* (données comprises entre 2012 et 2023) France, Consommation d'électricité dans les grandes métropoles Françaises <https://odre.opendatasoft.com/explore/dataset/eco2mix-metropoles-tr/information/?flg=fr&disjunctive.libelle_metropole&disjunctive.nature&sort=-date_heure&refine.date_heure=2021>
* (données comprises entre 2011 et 2021) France, Consommation annuelle d’electricité et de gaz par département et par secteur d’activité <https://opendata.agenceore.fr/explore/dataset/conso-elec-gaz-annuelle-par-secteur-dactivite-agregee-departement/export/>
* (données comprises entre 2012 et 2023) France, Pic journalier de la consommation brute d'électricité avec les données de température et de date.

<https://odre.opendatasoft.com/explore/dataset/pic-journalier-consommation-brute/information/?dataChart=%3D%3D>

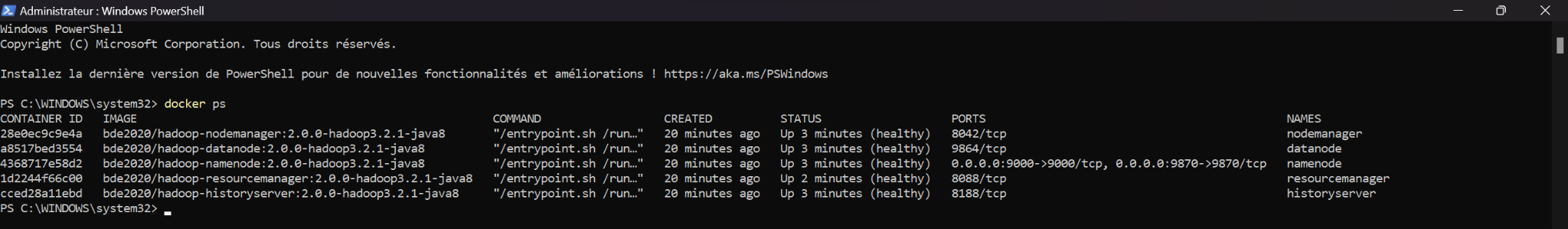
# **II - MISE EN PLACE DE L’ENVIRONNEMENT DE STOCKAGE ET CONSTITUTION DU DATA LAKE**

Nous allons stocker nos datasets (données de travail) dans le système de fichiers HDFS. Pour cela, nous utiliserons une image Hadoop dans Docker.



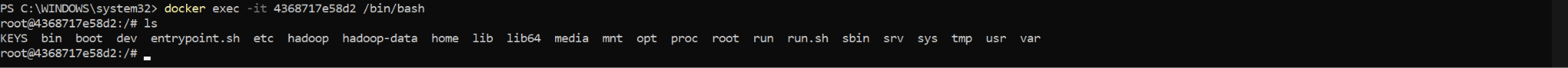
Une fois que nous avons téléchargé nos datasets, nous les mettons dans un premier temps dans un dossier sur la machine en local. L’idée sera ensuite de les stocker par la suite dans HDFS.

Pour stocker les données dans HDFS, nous utiliserons le namenode. Il nous faut donc connaître l’identifiant de son conteneur.



On voit que l’identifiant du namenode est id = 4368717e58d2.

A présent, nous allons accéder à l’id de ce conteneur.

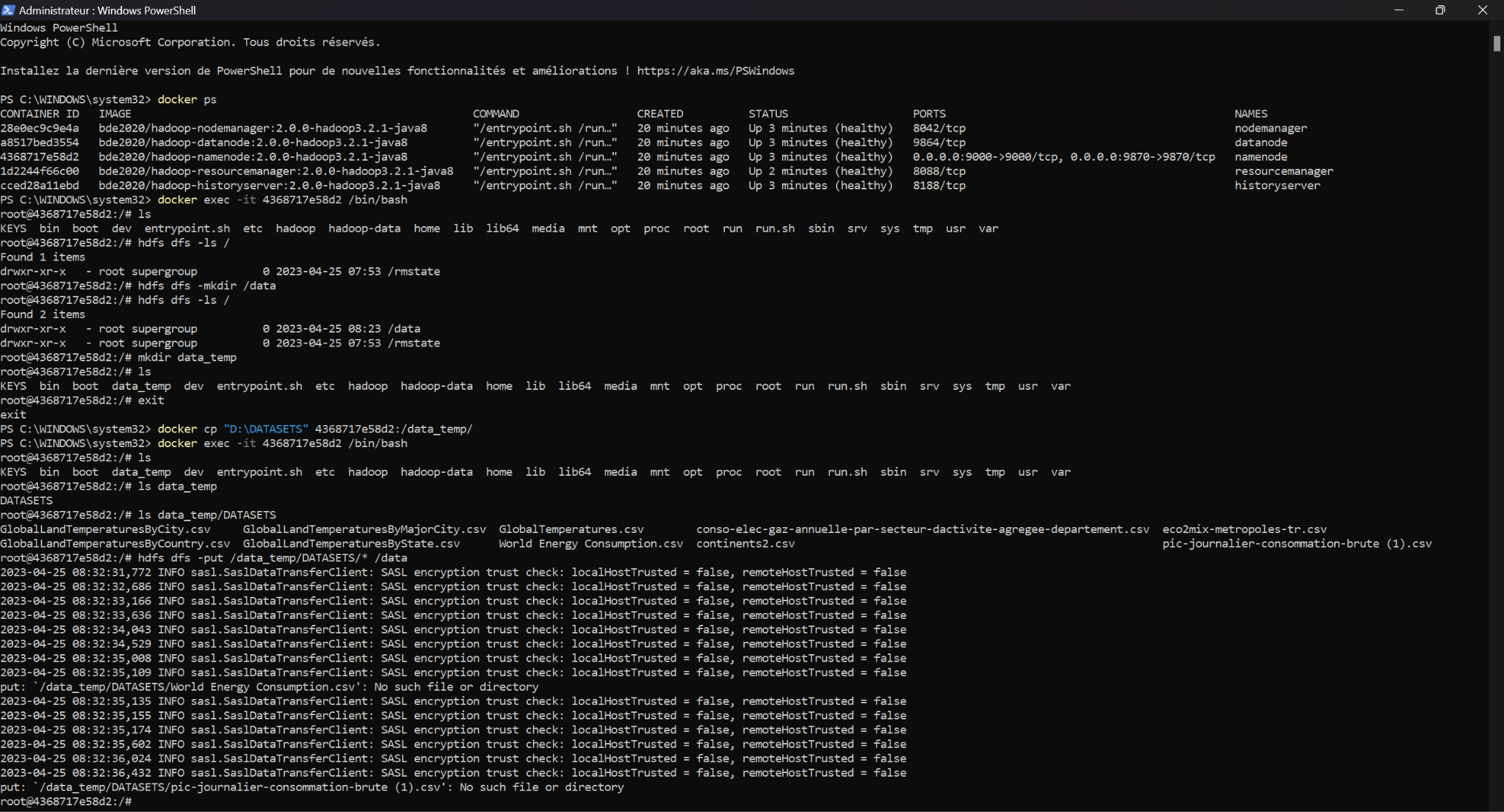


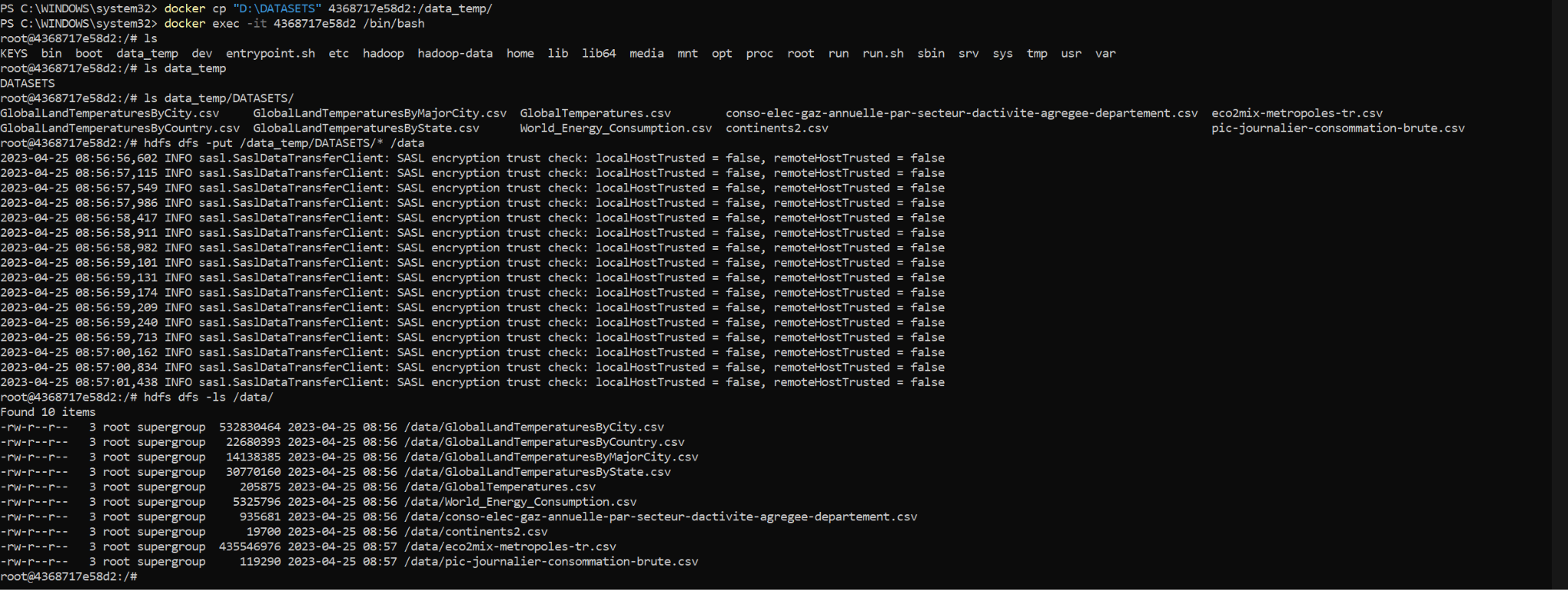


A présent nous allons créer un répertoire *data* dans HDFS.



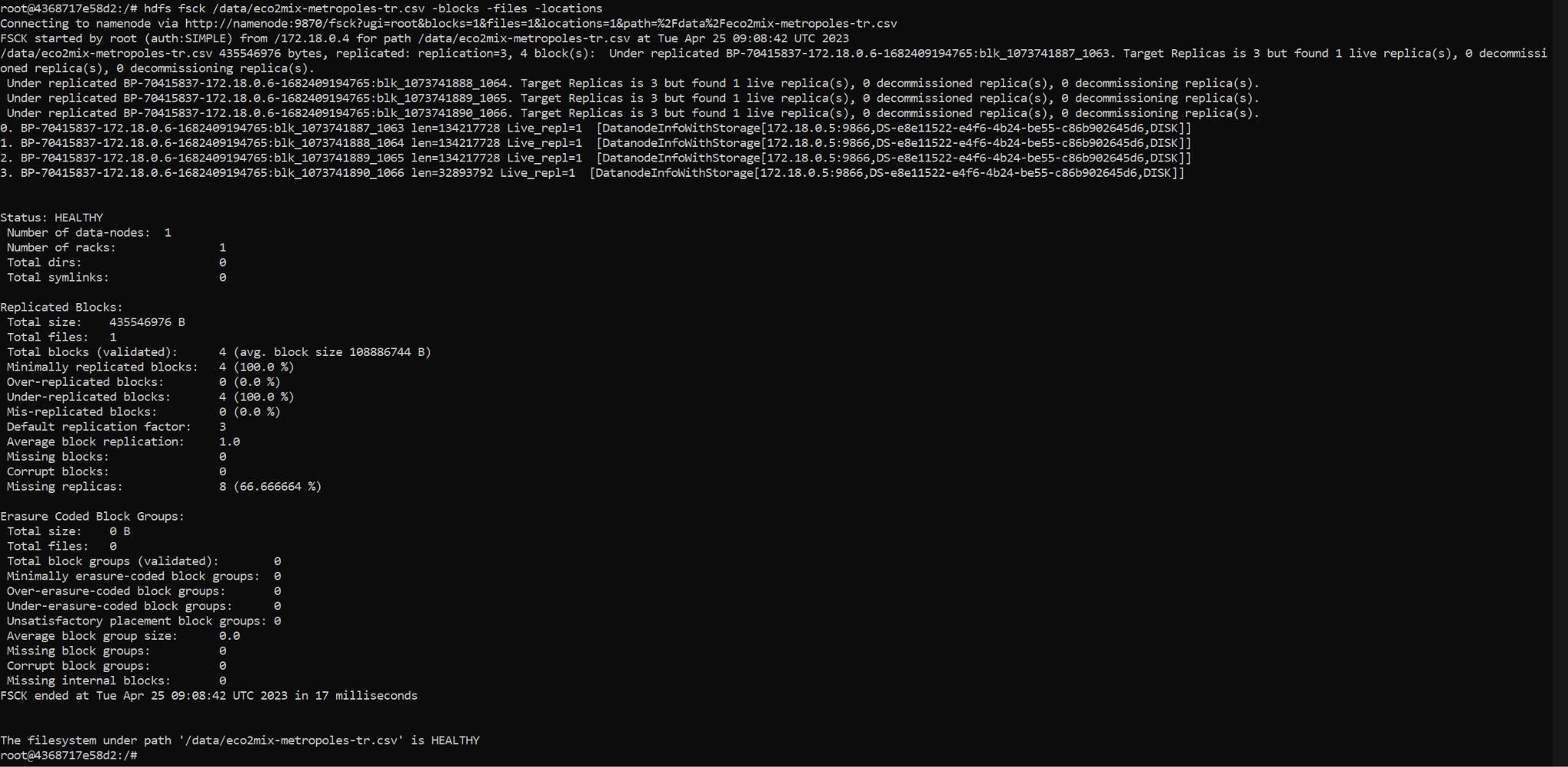
A présent, nous allons créer un répertoire *data\_temp* sur le namenode et copier les datasets du dossier local vers ce répertoire *data\_temp* du namenode

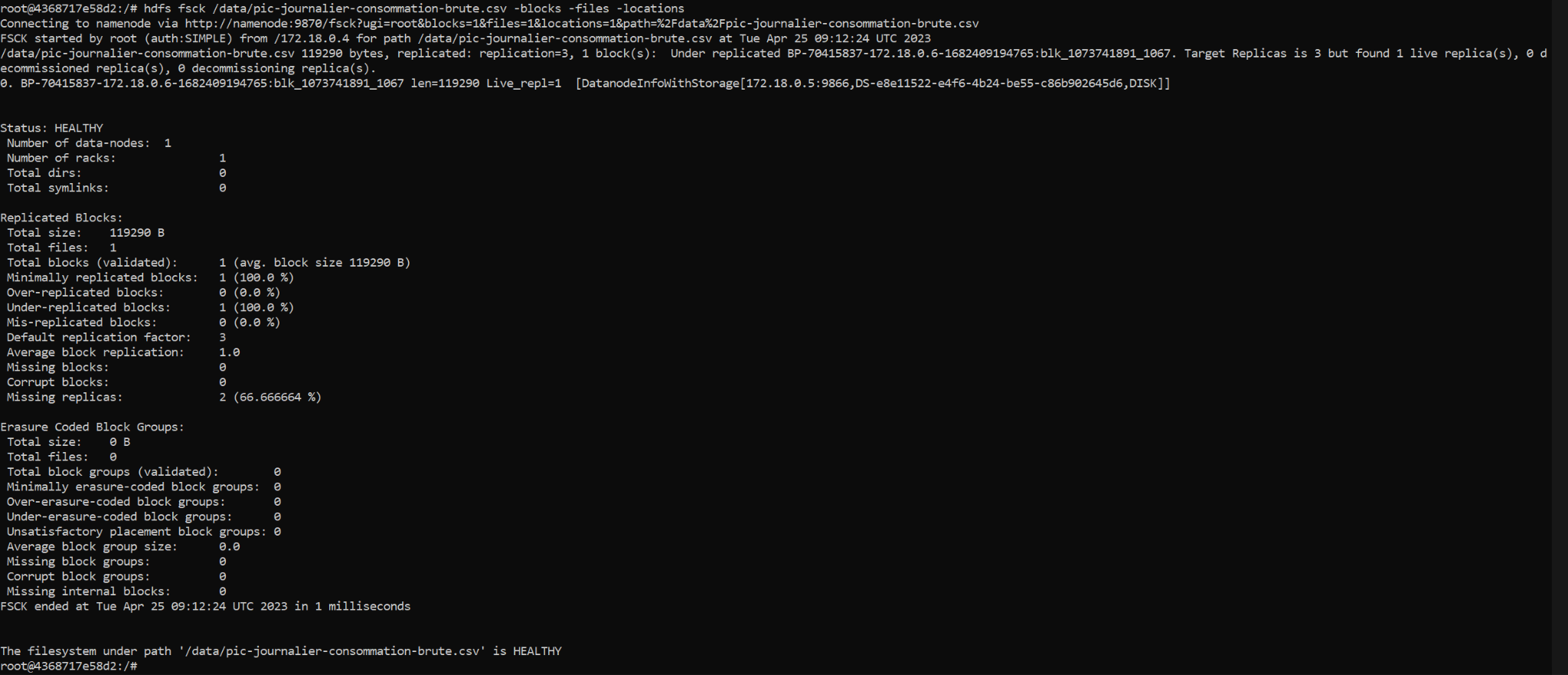




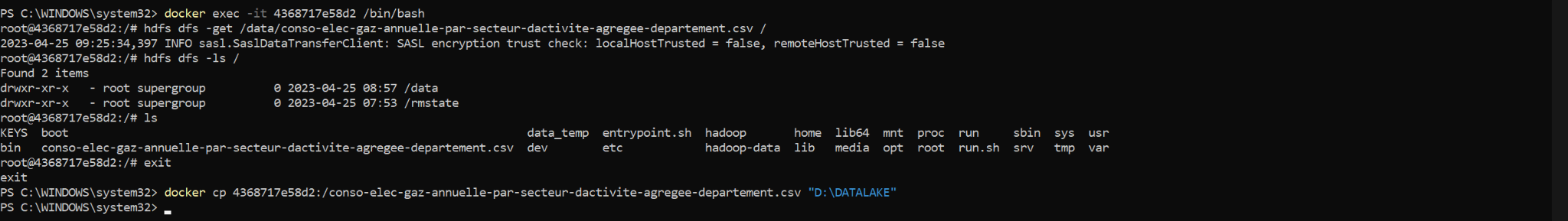
Notre datalake est donc ainsi constitué et stocké dans HDFS dans le dossier *data*.

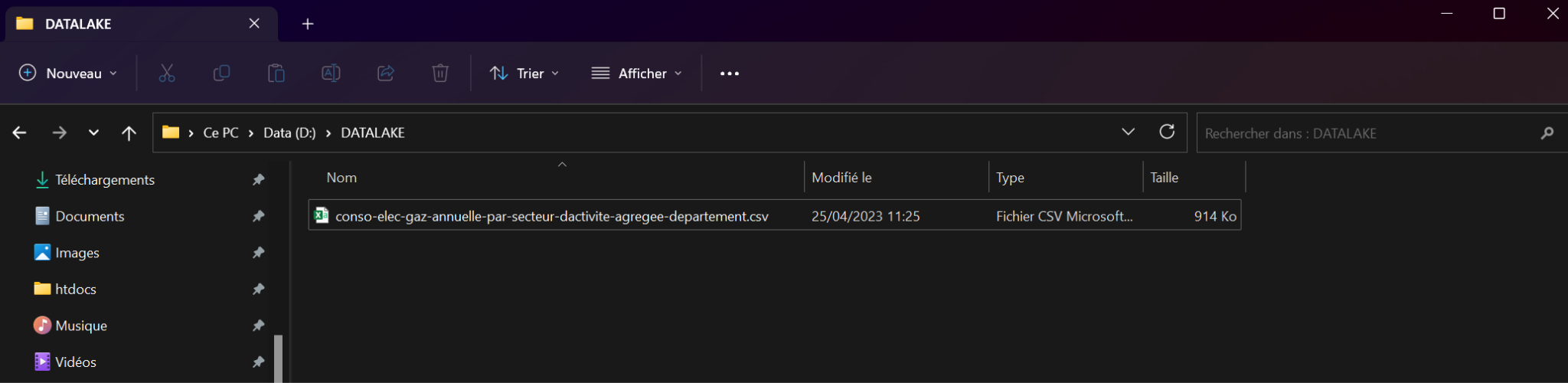
On peut même voir les informations sur un fichier (nombre de blocs, nombre de répliques, taille des blocs…)





A présent, si on veut récupérer un dataset sur la machine en local pour pouvoir l’utiliser, on a juste à le copier dans un dossier sur la machine en local.





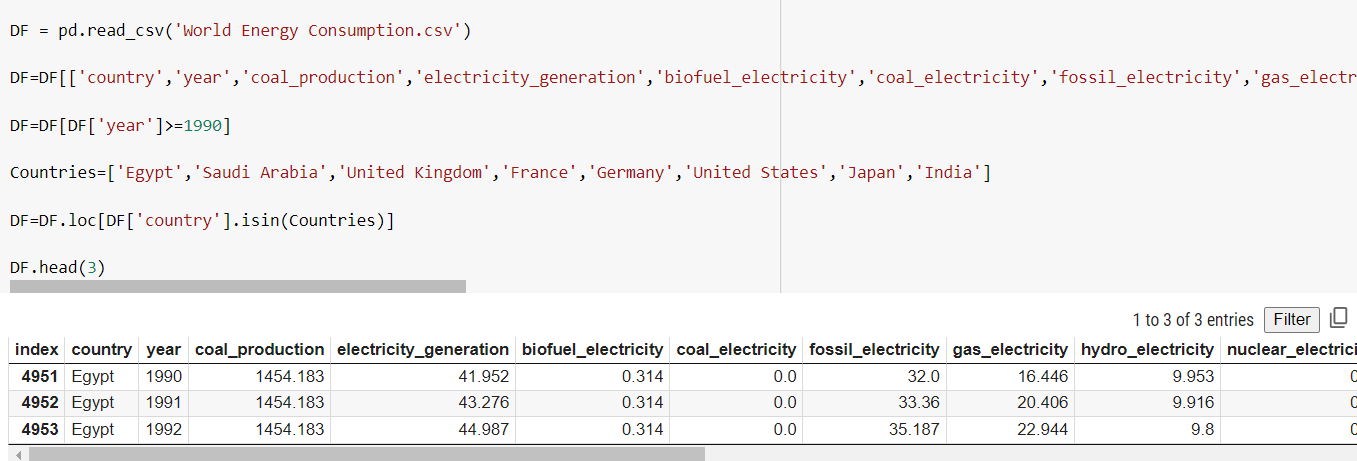
# **III - ANALYSE DESCRIPTIVE ET VISUALISATION DES DONNÉES**

## **1) Consommation d'énergie des différents pays du monde**

Tout d’abord, on importe les données du document **World Energy Consumption.csv.**

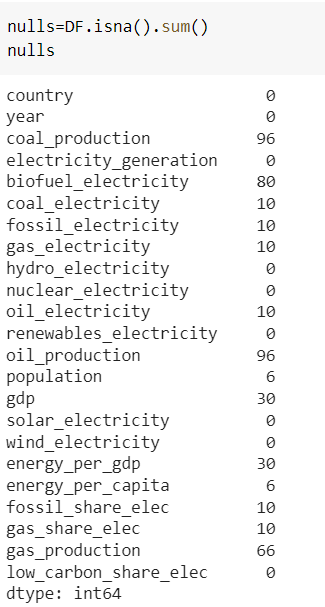
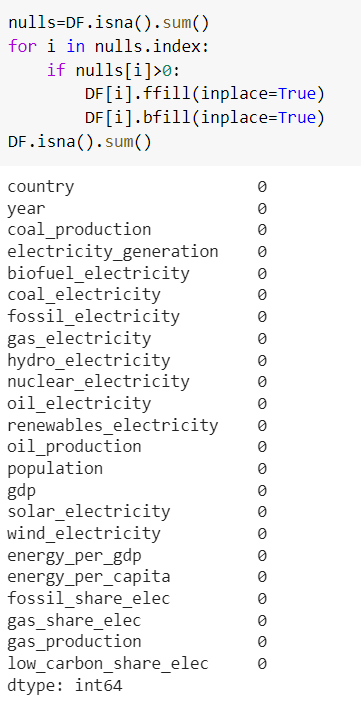
On filtre les colonnes en sélectionnant les données qu’on veut garder (tout ce qui est relié à la consommation d’énergies).

Puis on sélectionne l’année à partir de laquelle les données sont pertinentes et les pays de notre domaine d’étude.

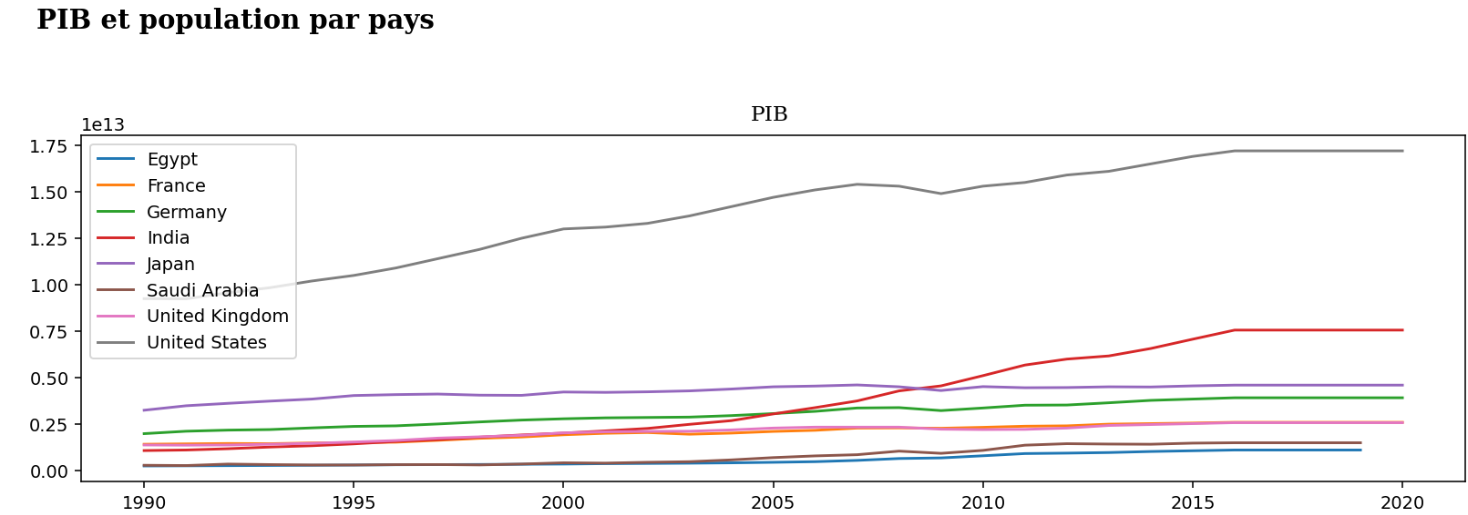


(fig.1 : Récupération des données )

Une fois les données récoltées, on supprime les lignes contenant des données nulles.

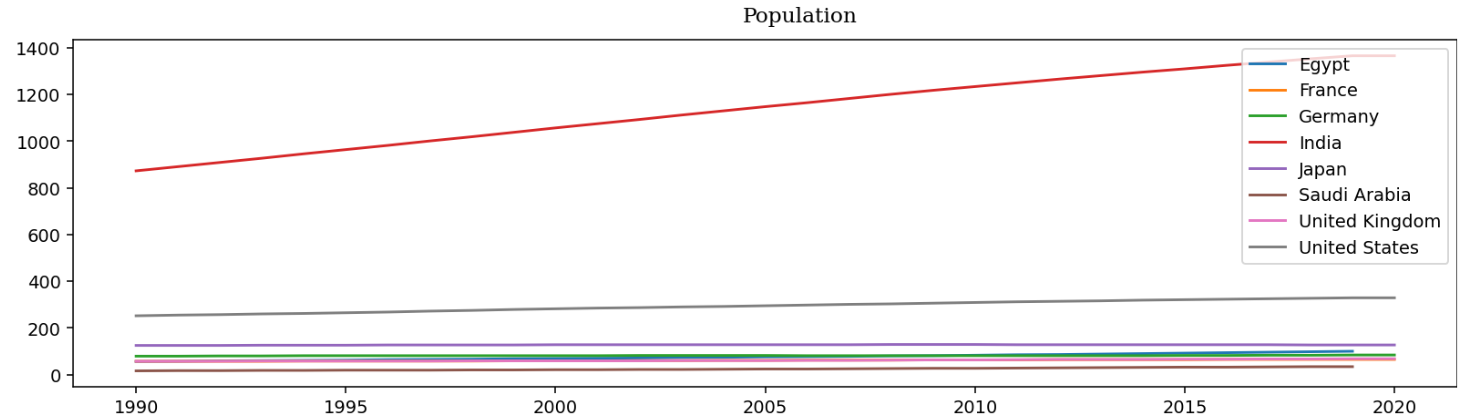
 

(fig.2 : à gauche, les données bruitées, à droite, les données nettoyées)



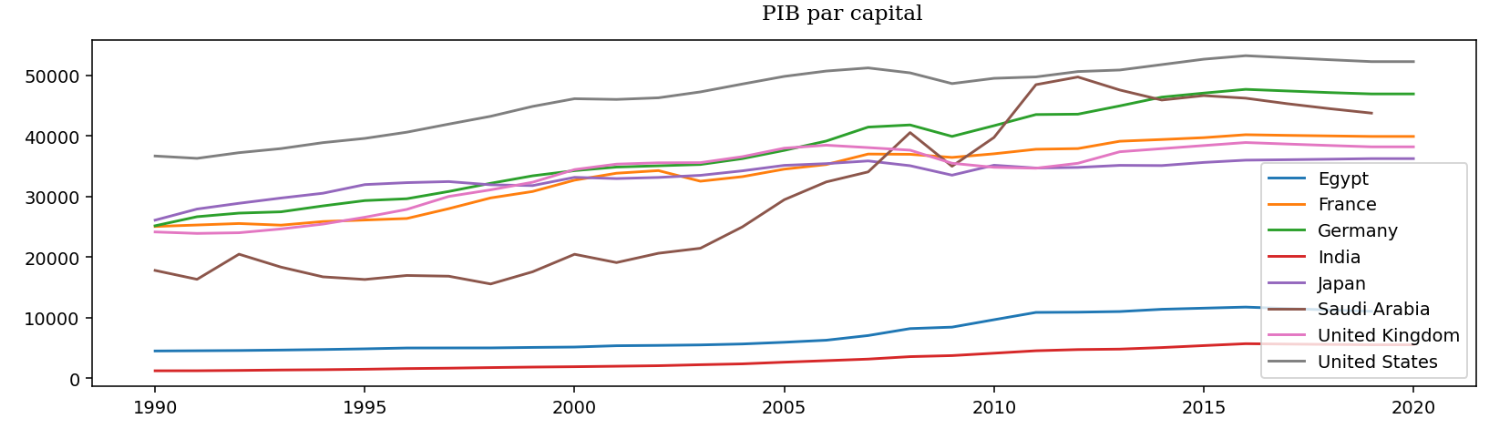
(fig.3 : PIB par pays en fonction de l’année)

Les PIB d’Egypte et d’Inde augmentent linéairement par rapport aux autres pays en fonction du temps.



(fig.4 : Population par pays en fonction de l’année)

Alors que la population des autres pays semble stagner, la population en Inde augmente.



(fig. 5 : PIB par capital en fonction de l’année)

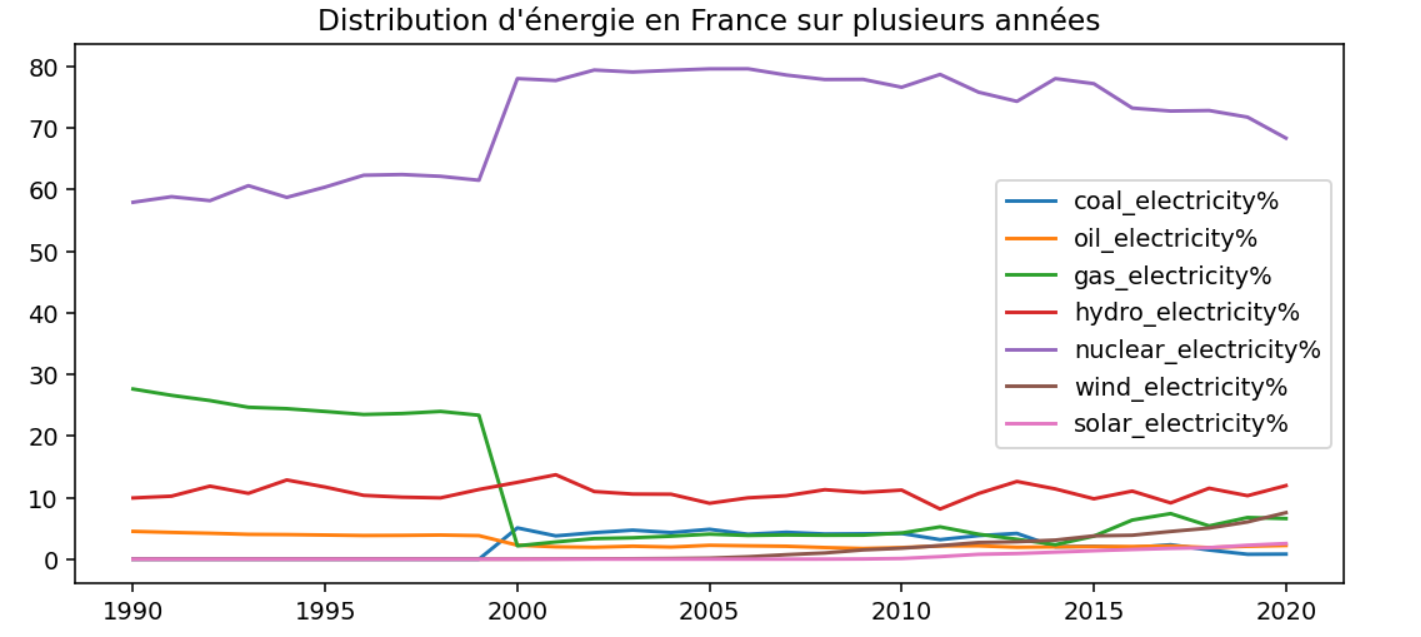
On remarque ici que le PIB par capital est le plus élevé aux États-Unis et en Allemagne.

On choisit alors de comparer la distribution d’énergies entre la France, les Etats-Unis et l’Inde.

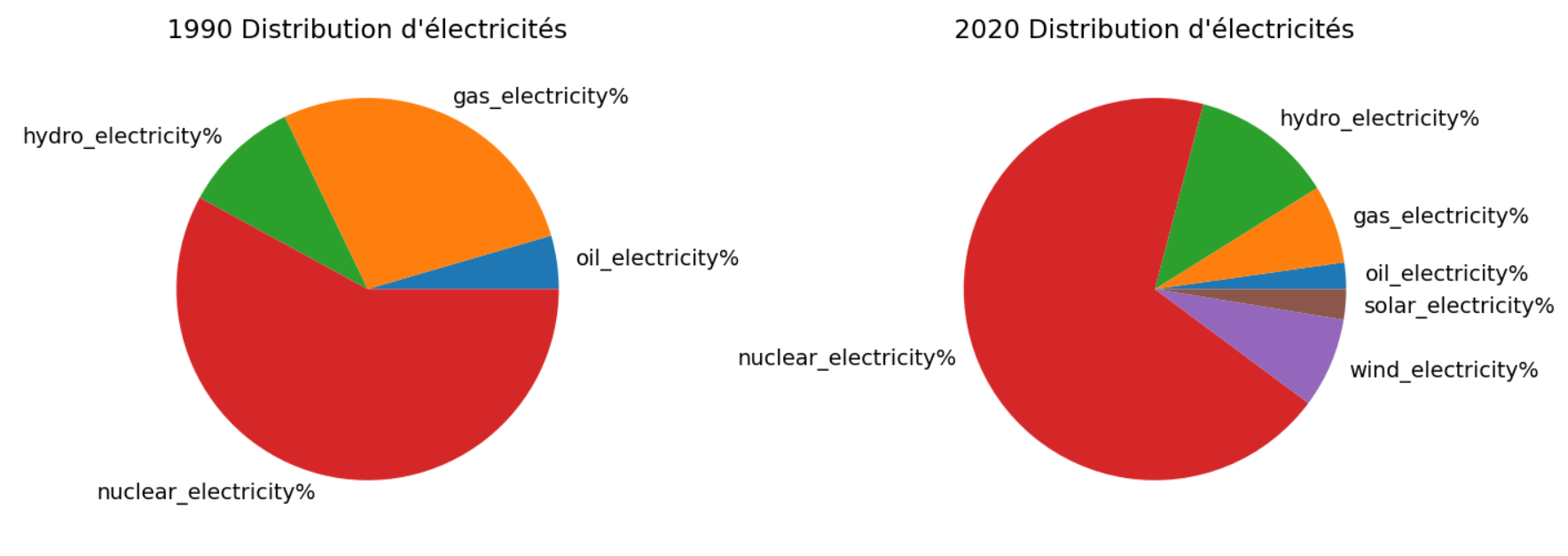
On sélectionne comme source d'électricité le charbon, l’huile, le gaz, l'hydraulique, le nucléaire, l’éolien et le solaire comme sources d’énergies.

L'étude des distributions d'énergies permet de mieux comprendre la situation énergétique du pays et d'anticiper les risques éventuels de pénuries ou de surconsommation.

* En France



(fig. 6 : Distribution d’énergies en France entre 1990 et 2020)



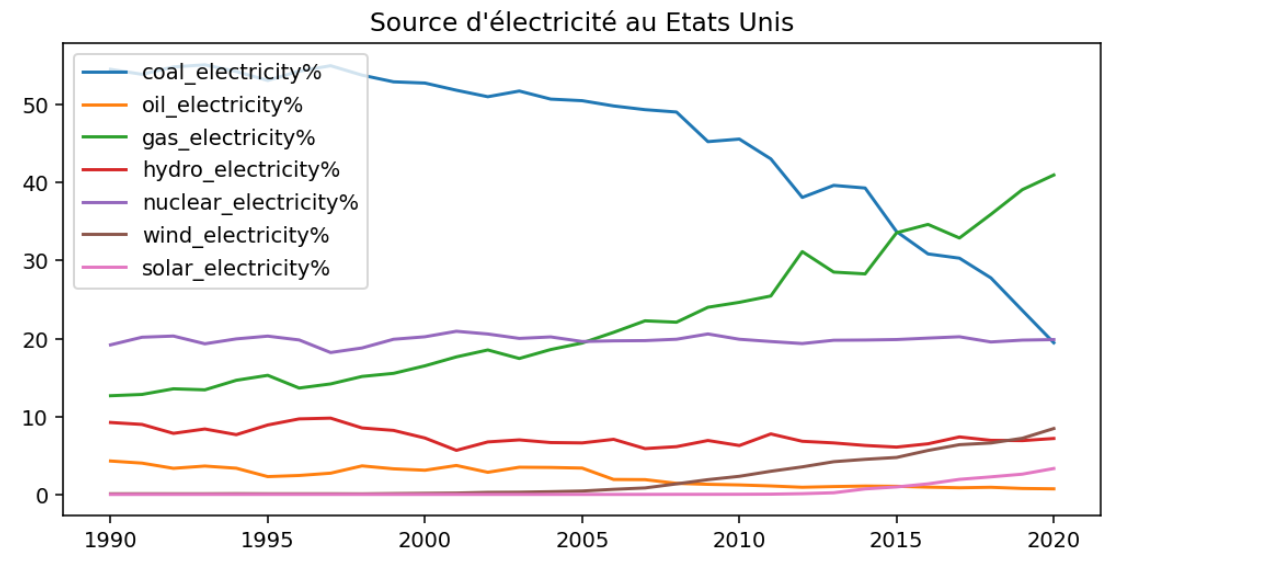
(fig. 7 : Piechart de la distribution d’électricités en France)

D’après la **fig. 6**, à la fin des années 1990, on remarque une baisse dans l’utilisation du gaz comme source d’électricité. Tandis que l’utilisation du nucléaire augmente. Vers les années 2020, on remarque une légère variation notamment concernant l’utilisation du nucléaire qui baisse et l’énergie éolienne qui augmente.

De manière générale, on remarque **fig. 7** que l’énergie nucléaire (nuclear\_electricity) reste une source importante d’électricité en France au fil des années.

De plus, les énergies renouvelables, telles que l’énergie solaire (solar\_electricity) et éolienne (wind\_electricity), sont en augmentation mais représentent encore une part relativement faible de la consommation d’énergies en France.

* Aux Etats-Unis



(fig. 8 : Distribution d’énergies en France entre 1990 et 2020)



(fig. 9 : Piechart de la distribution d’électricités aux Etats-Unis)

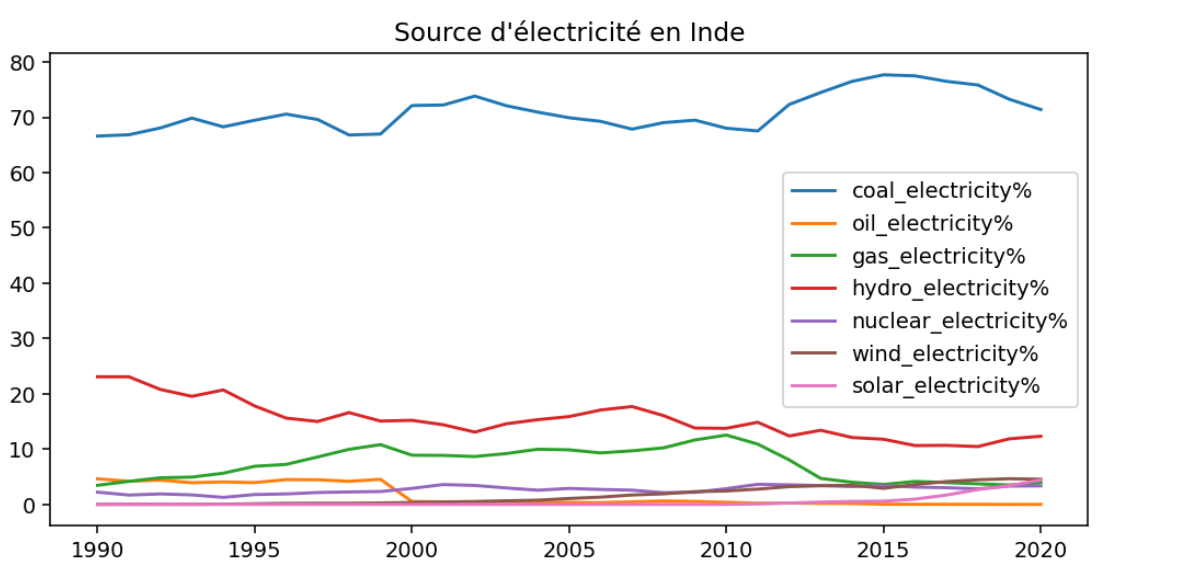
D’après la **fig. 8**, jusqu’à 2015, la source d’électricité la plus importante aux Etats-Unis est le charbon. Cependant, depuis 2010, elle est en déclin et dès 2015 est remplacée par le gaz.

Vers les années 2020, on remarque une légère variation concernant l’utilisation de l’énergie éolienne qui augmente.

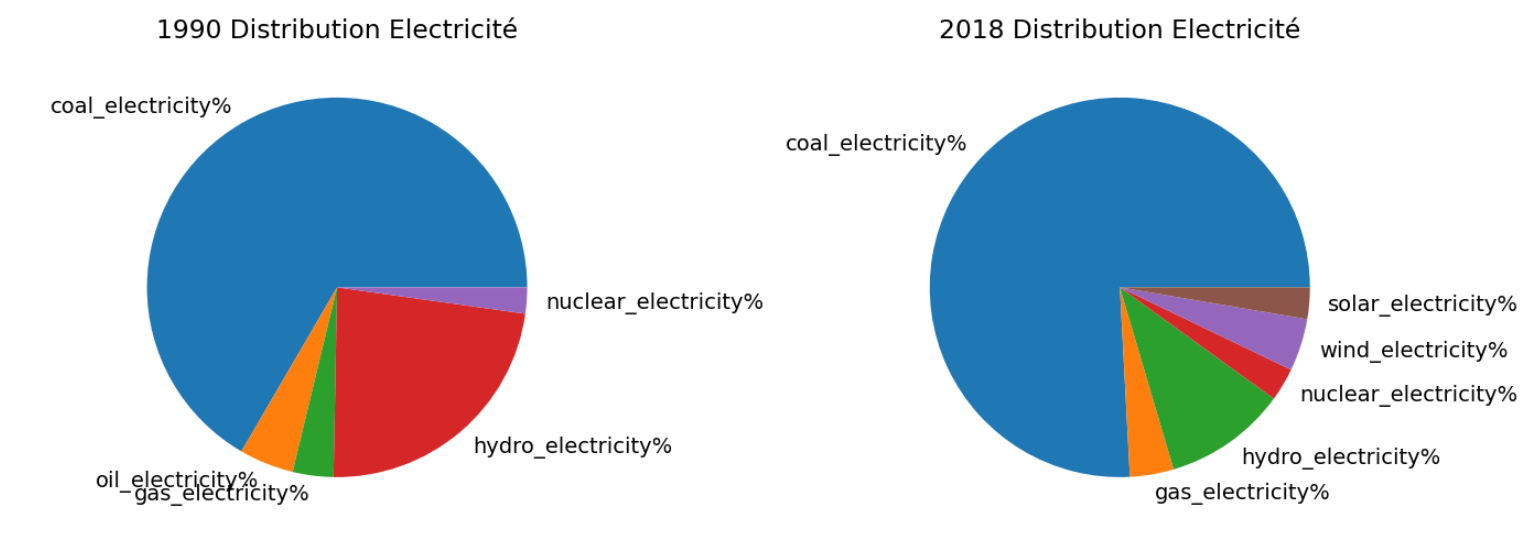
De manière générale, on remarque **fig. 9** que l'électricité provenant de charbon (coal\_electricity) est une source importante aux Etats-Unis en 1990 qui diminue drastiquement en 2020. Tandis que le gaz (gaz\_electricity) devient la source d’électricité de prédilection des Etats-Unis.

De plus, à partir de 2020, on constate l’apparition de l’énergie solaire (solar\_electricity).

* En Inde



(fig. 10 : Distribution d’énergies en Inde entre 1990 et 2020)



(fig. 11 : Piechart de la distribution d’électricités aux Etats-Unis)

D’après la **fig. 10**, la source principale d’électricité en Inde est le charbon.

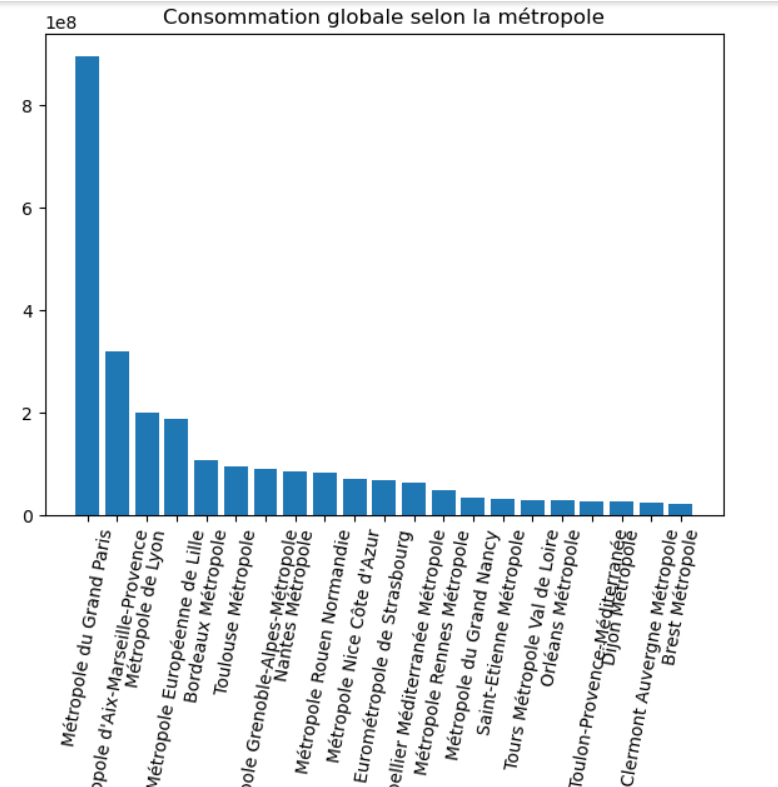
Il n’y a pas d’évolution particulière malgré la **fig. 4** qui montre une population grandissante.

De manière générale, on remarque **fig. 11** l’apparition de l’énergie solaire (solar\_electricity).

## 

## **2) Consommation d'énergie dans les métropoles françaises**

**Histogramme de la consommation globale selon la métropole entre 2017 et 2023**

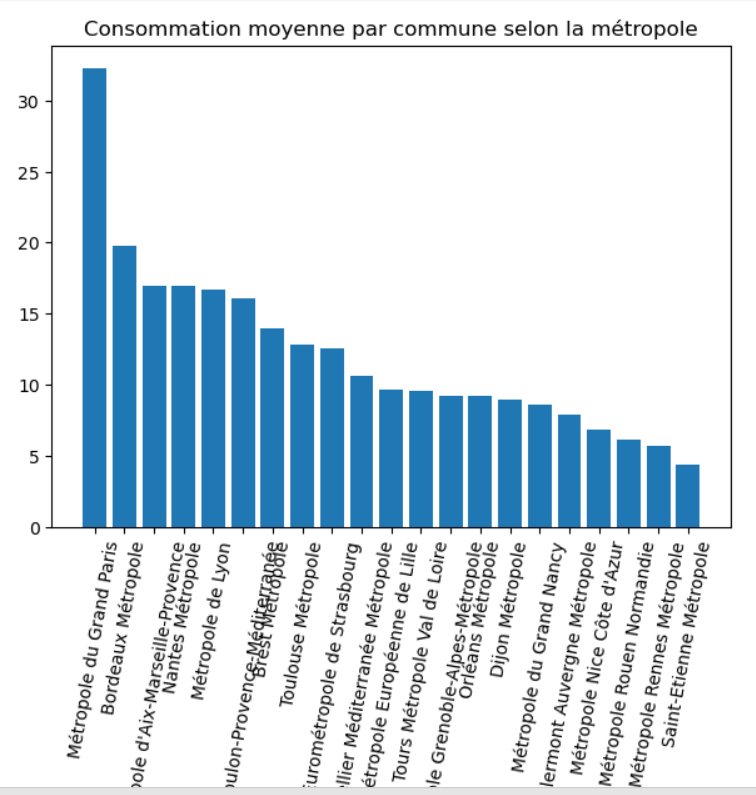


Le graphe en question affiche les données relatives à la consommation totale d'énergie dans les métropoles importantes de France.

D'après les données présentées ci-dessus, on constate que la métropole ayant de très loin la consommation d'énergie la plus élevée est celle du Grand Paris, et que cette dernière consomme significativement plus d'énergie que les autres métropoles.

De plus, il est à noter que les quatre plus grandes métropoles de France, à savoir Paris, Marseille, Lyon et Lille, sont également les quatre métropoles les plus énergivores.

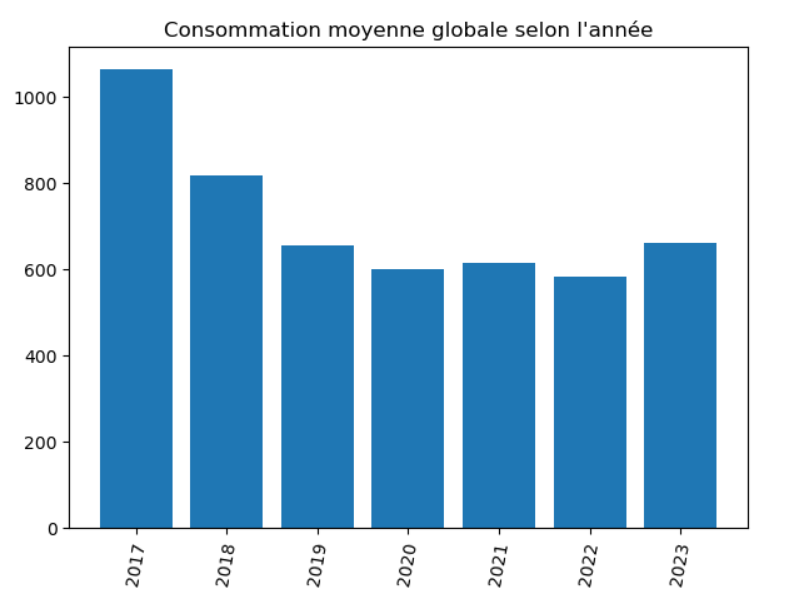
**Histogramme de la consommation moyenne par commune selon la métropole**



Le graphique en question illustre les données relatives à la consommation énergétique moyenne par commune dans chaque métropole de France. L'objectif de ce graphique est de permettre une comparaison facile de la consommation énergétique moyenne par commune entre différentes métropoles en France, en fournissant une représentation visuelle claire et concise de ces données.

Cela signale donc que bien que la consommation globale d'une zone (métropole) soit plus élevée qu'une autre, il est possible que la moyenne de la consommation par entité individuelle (commune) soit plus élevée dans la zone ayant une consommation globale inférieure. En d'autres termes, la zone ayant une consommation globale inférieure peut avoir des entités individuelles qui consomment en moyenne plus que celles de la zone ayant une consommation globale supérieure.

**Histogramme de la consommation moyenne globale selon l'année**



Le graphique en question affiche les données relatives à la consommation énergétique moyenne globale sur une période donnée, en prenant en compte toutes les sources d'énergie utilisées pendant cette période et dans l'ensemble des villes concernées. L'objectif de ce graphique est de permettre une comparaison facile de la consommation énergétique moyenne globale entre différentes années, en fournissant une représentation visuelle claire et concise de ces données.

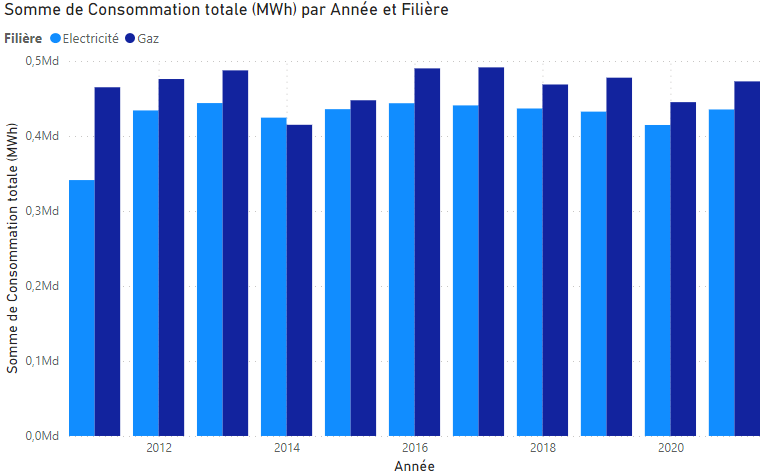
D'après les données présentées, on peut constater une variation de la consommation d'énergie entre 2017 et 2023. Plus précisément, l'année 2017 est celle où la consommation d'énergie a été la plus élevée, suivie d'une baisse de la consommation jusqu'en 2021. Ensuite, on observe une légère augmentation de la consommation d'énergie en 2021, suivie d'une nouvelle hausse en 2023. Cette observation suggère une tendance à la baisse de la consommation d'énergie jusqu'en 2021, qui a été suivie d'une légère augmentation. Ces données peuvent avoir des implications importantes pour la planification énergétique et la politique environnementale.

**Nuage de points de la consommation moyenne de la population selon la métropole**



Ce graphique illustre une corrélation entre la population et la consommation d'énergie, suggérant que plus la population est importante, plus la consommation d'énergie est élevée.

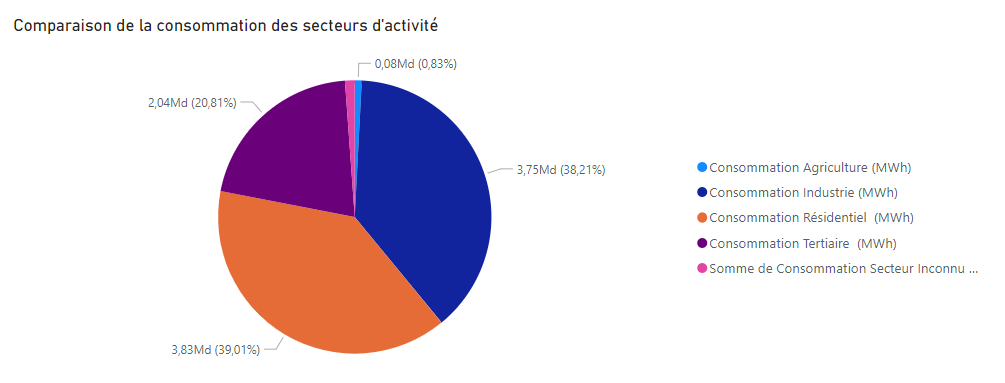
## **3) Consommation d'électricité et de gaz en France par département et par secteur d'activité**



(fig: Histogramme sur la consommation globale d'énergie en france entre 2011 et 2021)

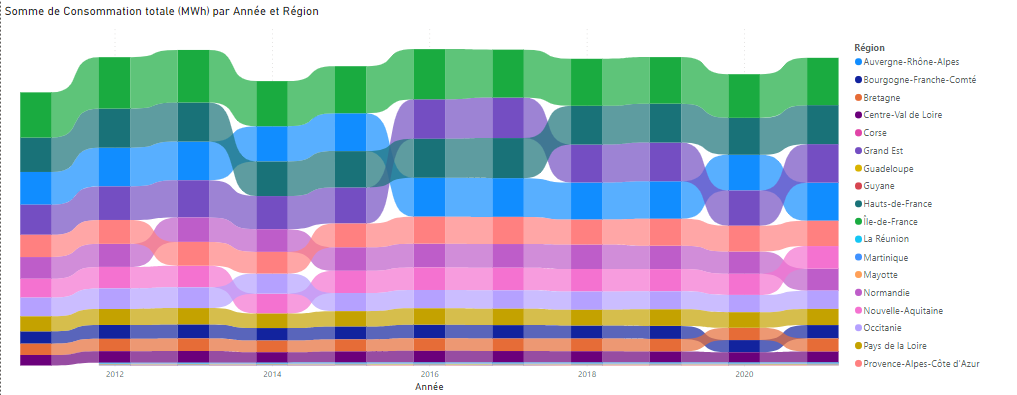
On constate ici que la consommation globale en gaz de 2011 à 2021 est toujours supérieure à la consommation globale en électricité sauf en 2014 où l'électricité est plus consommée que le gaz.

Cette observation suggère que généralement, le gaz est plus consommé que l'électricité en France.



(fig: comparaison de la consommation globale d'énergie en france par secteur d'activité)

On constate ici que 39% de l'énergie consommée en France se fait en secteur résidentiel, suivi de très près par 38% pour le secteur industriel.

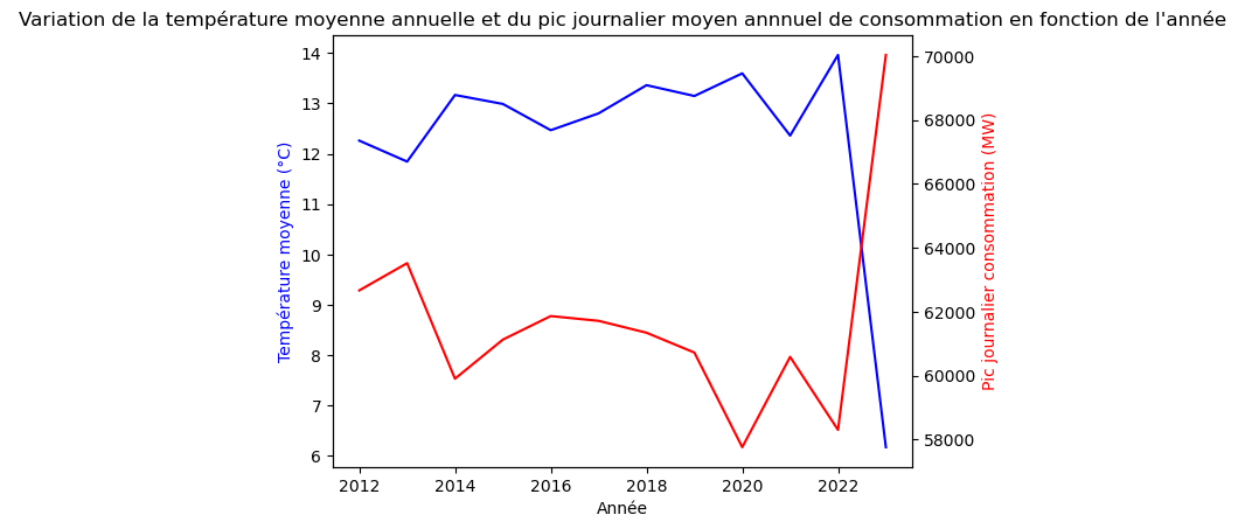
****

(fig : comparaison de la consommation globale d'énergie en france par région)

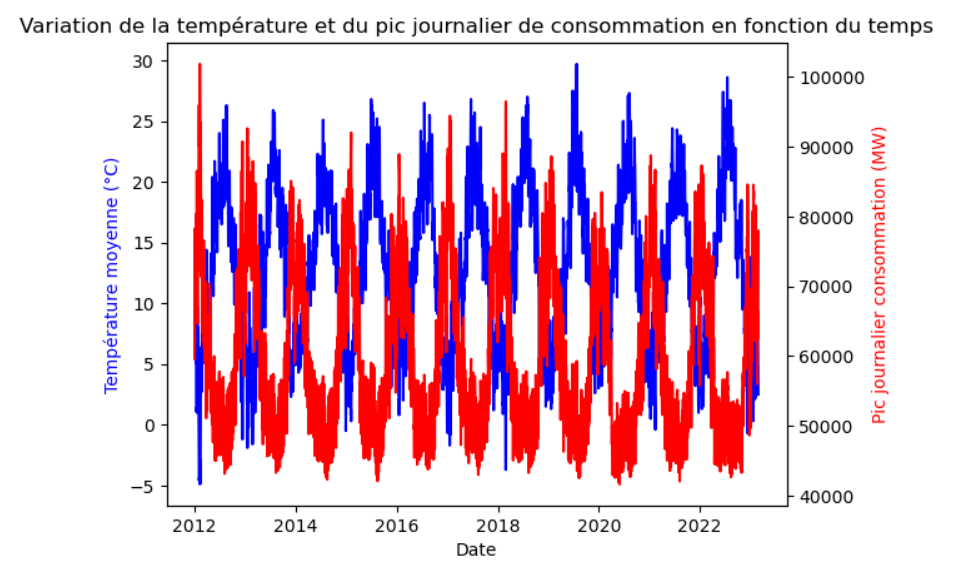
On constate ici que la consommation d'énergie en France est plus élevée dans la région de l’Ile de France, suivie par les régions Auvergne-Rhône-Alpes, Centre-Val de Loire, Hauts-de-France.

## **4) Consommation d'électricité (pic journalier de consommation brute) avec les données de température et de dates**

**Variation de la température moyenne annuelle et du pic journalier moyen annnuel de consommation en fonction de l'année**

****

**Courbe de la Variation de la température et du pic journalier de consommation en fonction du temps**

****

Ce graphique illustre le propos selon lequel la consommation d’électricité est principalement impactée par la météorologie et les éléments du calendrier.Le graphe suggère que la consommation d'électricité augmente lorsque la température diminue, et diminue lorsque la température augmente.

Ceci peut s'expliquer par le fait que lorsque les températures sont basses, nous avons tendance à utiliser des équipements électriques tels que le chauffage, qui ont une consommation énergétique élevée.

# **IV - ANALYSE PRÉDICTIVE**

## **Consommation d'électricité en France (pic journalier de consommation brute) avec les données de température et de date**

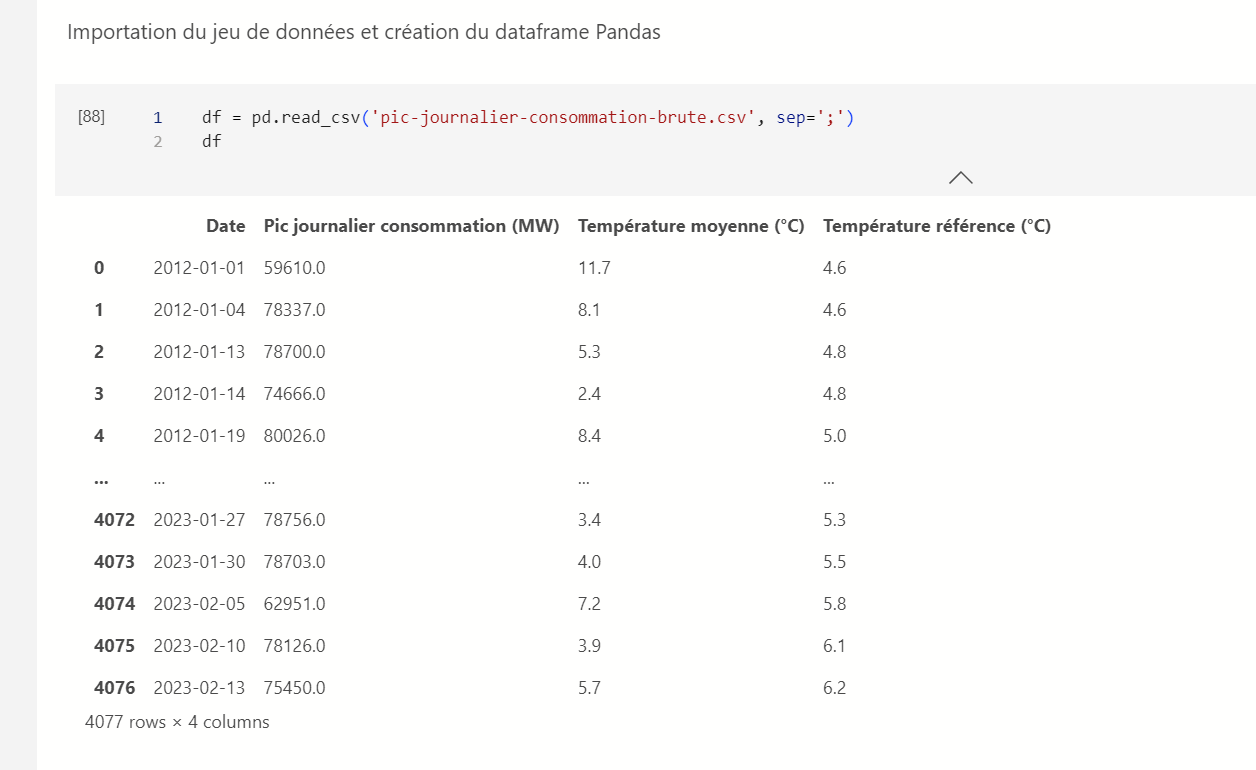
Le dataset ***pic-journalier-consommation-brute.csv*** contient les données de consommation d'électricité (pic de consommation journalier) en France ainsi que les données journalières de température.

Le but de cette analyse prédictive est de prédire la consommation d'électricité (pic de consommation journalier) entre 2021 et 2023 grâce à un modèle de machine learning qu'on aura au préalable entraîné sur les données allant de 2012 à 2020.

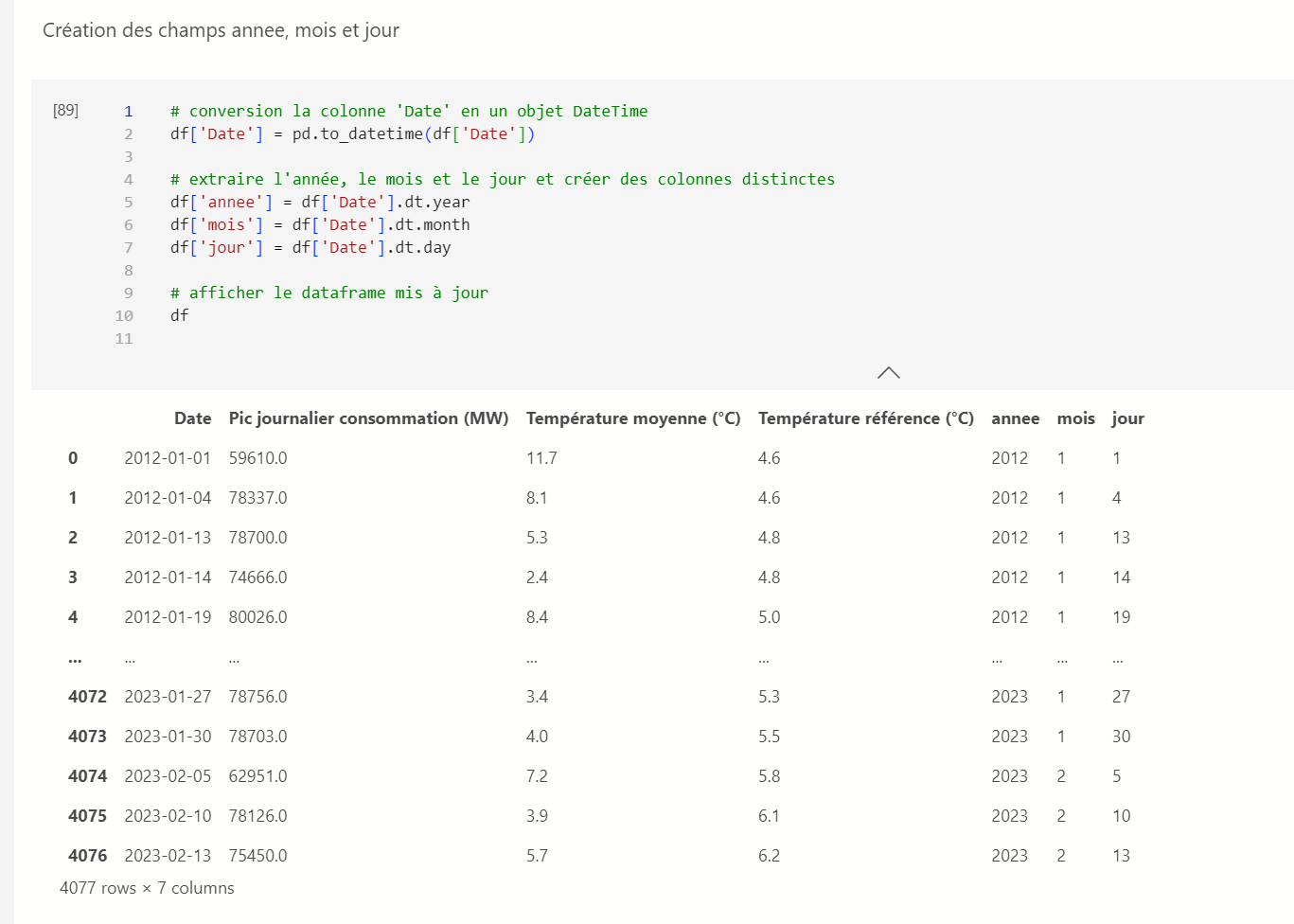
Nous allons utiliser un modèle RandomForest (Forêts aléatoires).

Le principe d'une forêt aléatoire est tout simplement de générer un grand nombre d'arbres de décisions qui seront ensuite combinés (en utilisant des moyennes ou des règles de "majorité") à la fin du processus.

Grâce à cela, si par exemple à un instant t en 2023, on a les prédictions météo (température) sur une certaine plage temporelle, on pourra utiliser ce modèle pour prédire la consommation d'électricité sur cette plage temporelle.



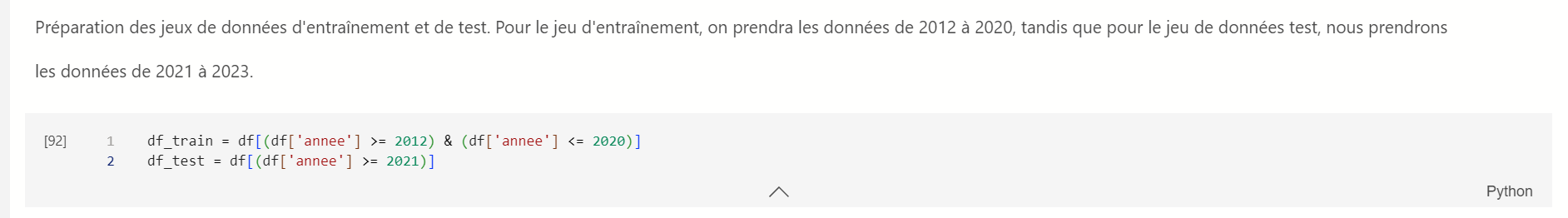
Nous allons tout d’abord rajouter les champs annee, mois et jour. Ainsi, nous pourrons nous débarrasser du champ Date.



Ensuite nous allons créer le jeu de données test et le jeu de données d’entraînement.

Pour le jeu d'entraînement, on prendra les données de 2012 à 2020, tandis que pour le jeu de données test, nous prendrons

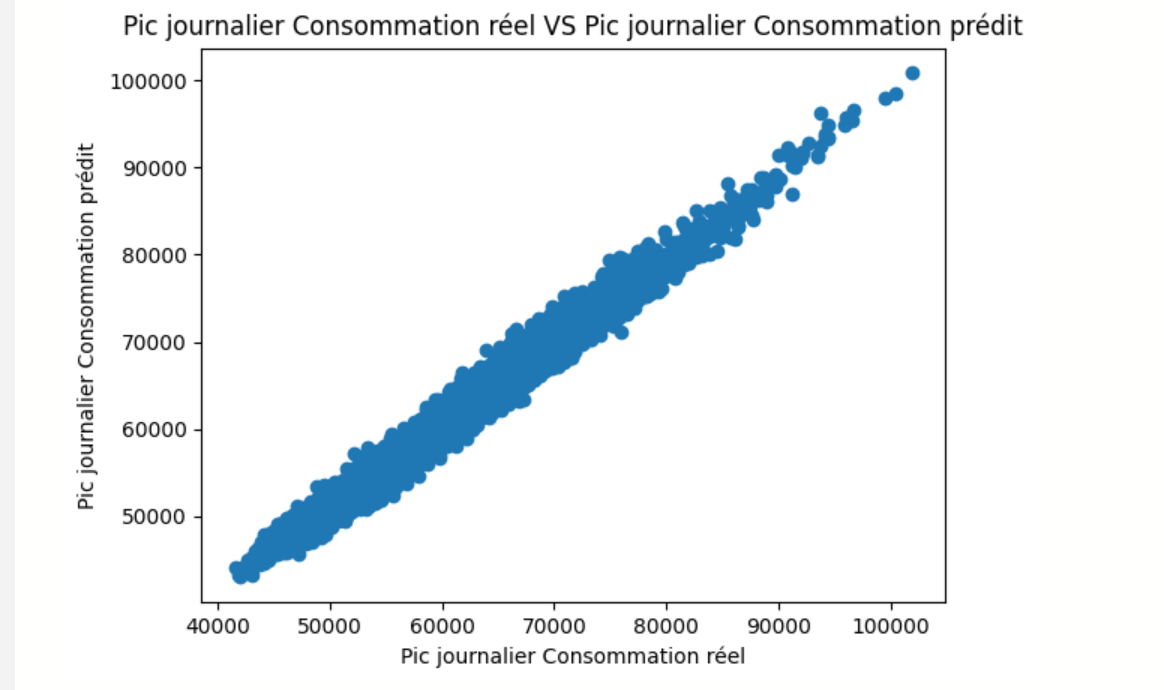
les données de 2021 à 2023.



Maintenant, nous allons initialiser le modèle et l’entraîner sur les données du jeu d’entraînement.

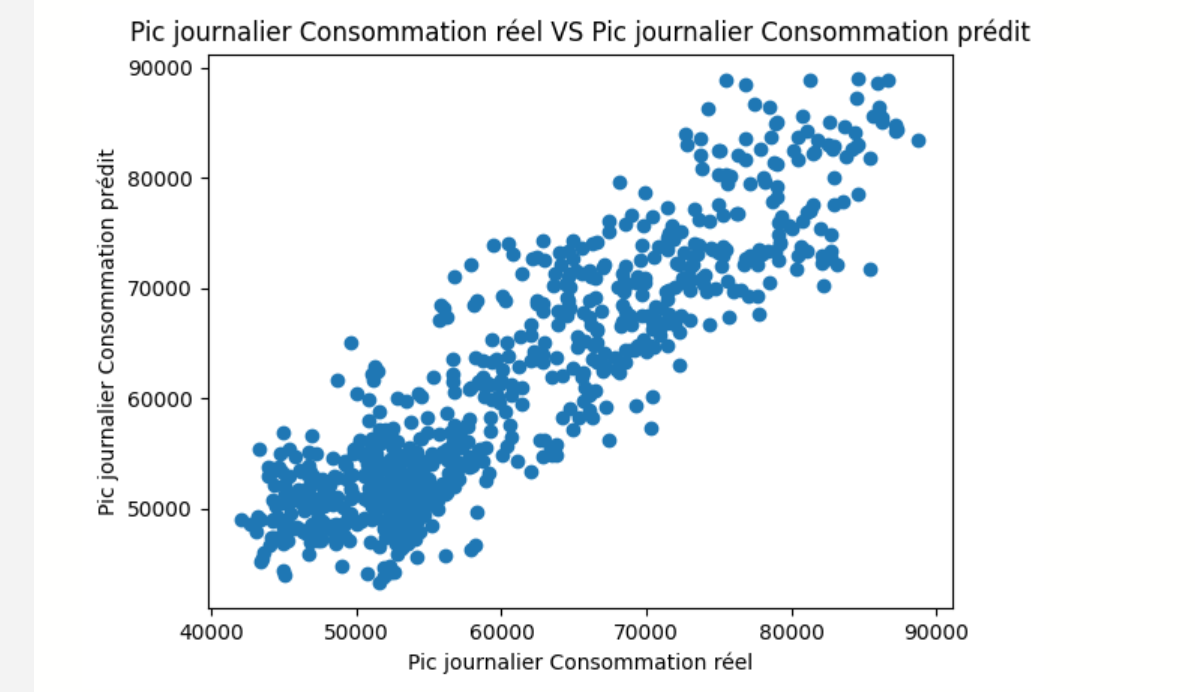


Une fois cela fait, nous pouvons faire prédire la consommation au modèle sur les données du jeu d’entraînement.



le R squared = 0.9817973872262719 est très élevé (proche de 1), donc la régression est très bonne.

A présent nous allons faire prédire à notre modèle la consommation d’électricité sur les données du jeu de test. Nous précisons qu’il n’a pas connaissance des résultats de ces données de test. C’est la première fois qu’il les voit.



le R squared = 0.8022536070525198 est élevé (proche de 1), donc la régression est très bonne. On a donc réussi à générer un modèle capable de prédire

# la consommation d'électricité entre 2023 avec de bonnes performances.

# **CONCLUSION**

A travers notre étude comparative de la consommation d’énergies entre les différents pays, nous prenons conscience de leurs différences et similitudes, ainsi que des facteurs influençant la consommation d’énergie de chaque pays.

En comparant les pays entre eux, il apparaît comme nécessaire pour la France de changer de dynamique énergétique et de réduire sa consommation globale d’énergies. La France privilégie d’ailleurs la consommation de gaz à la consommation d’électricité.