



Projet Data Science

Compte rendu - Véhicules Électriques

La répartition des infrastructures de recharge est-elle adaptée en France ?

IG4 - 2022 / 2023

Romain Frezier - Lorenzo Italiano - Tom Sartori

1. Introduction	3
1.1. Présentation du sujet	3
1.2. Découverte des données	3
1.2.1. Données brutes	3
1.2.2. Premiers pas	4
1.3. Lac et entrepôt de données	4
1.3.1. Lac de données AWS S3	4
1.3.2. Entrepôt de données AWS Athena et S3	5
2. Comment le nombre d'infrastructures de recharge évolue-t-il en France ?	7
2.1. Etude du parc d'infrastructure de recharge	7
2.1.1. Total bornes par date	7
2.1.2. Évolution de l'infrastructure au cours du temps	7
2.1.3. Répartition géographique des bornes	8
2.2. Étude du parc de voitures	8
2.2.1. Total voitures par date	8
2.2.2. Total voitures électriques par date	9
2.2.3. Rapport voitures électriques / voitures totales par date	9
2.2.4. Représentation géographique de la quantité de voitures (thermiques, électriques, gaz) par dates et par région ou département	10
2.3. Existe-t-il un lien entre l'évolution du nombre de voitures électriques et l'évolution du nombre d'infrastructures de recharges ?	11
2.3.1. Nombre de bornes par voitures électriques en France	11
2.3.2. Représentation géographique du nombre de bornes par voitures électriques par date et par région/département	11
3. La répartition des infrastructures de recharge sur le territoire correspond-elle à une logique de distribution de la population et de trafic routier ?	12
3.1 Existe-t-il une relation entre trafic, nombre d'habitants et nombre de bornes ?	12
Résultat des tests du chi2	13
Résultat des ACMs	14
3.2 Existe-t-il une relation entre nombre de véhicules électriques et nombre de bornes ?	17
Résultat de l'Anova sur les régions, année 2020	17
Résultat de l'Anova sur les départements, année 2020	17
Résultat de l'Anova sur les régions, année 2021	18
Résultat de l'Anova sur les départements, année 2021	18
Résultat de l'Anova sur les régions, année 2022	18
Résultat de l'Anova sur les départements, année 2022	19
3.3 Existe-t-il une relation entre nombre de véhicules électriques, type de population et nombre de bornes ?	19
3.3.1 Taux de chômage	19
Résultat de l'Anova sur les régions, année 2020	20

Résultat de l'Anova sur les départements, année 2020	21
Résultat de l'Anova sur les régions, année 2021	21
Résultat de l'Anova sur les départements, année 2021	22
Résultat de l'Anova sur les régions, année 2022	23
Résultat de l'Anova sur les départements, année 2022	23
3.3.2 Age	24
Résultat de l'Anova sur les régions, année 1	25
Résultat de l'Anova sur les départements, année 1	25
Résultat de l'Anova sur les régions, année 2	26
Résultat de l'Anova sur les départements, année 2	27
Résultat de l'Anova sur les régions, année 3	27
Résultat de l'Anova sur les départements, année 3	28
3.3.3 Richesse	28
Résultat de l'Anova sur les régions, année 1	29
Résultat de l'Anova sur les départements, année 1	30
Résultat de l'Anova sur les régions, année 2	31
Résultat de l'Anova sur les départements, année 2	31
Résultat de l'Anova sur les régions, année 3	32
Résultat de l'Anova sur les départements, année 3	33
3.3.4 Conclusion de nos analyses	34
4. Conclusion	35
Annexes	36
Code SQL utilisé dans AWS Athena, afin de transformer les données de l'entrepôt.	36

1. Introduction

Ce document existe également en notebook contenant des éléments interactifs et tout le code python des analyses, pour aller sur le google colab, [cliquez ici](#).

1.1. Présentation du sujet

[Véhicules électriques](#) : La répartition des infrastructures de recharge est-elle adaptée ?



Ce sujet traite de l'évolution des bornes de recharges de voitures électriques dans les différents départements et régions de France. L'évolution de ces infrastructures est détaillée sur plusieurs trimestres et années. Le nombre de voitures électriques par zone est également comparé aux caractéristiques des populations dans ces zones.

Nous avons choisi d'étudier ce sujet, car il nous semble pertinent, utile et intéressant. Le choix du rapport sur Google Colab est motivé par Python et la simplicité d'utilisation de cet outil.

1.2. Découverte des données

1.2.1. Données brutes

Les données brutes proviennent toutes du site data.gouv.fr. Nous utilisons principalement les données indiquées dans les recommandations du projet. Cependant, nous avons eu besoin de données complémentaires pour, par exemple, les correspondances entre les codes INSEE des communes, les départements et les régions.

Dès le départ, nous avons eu des problèmes avec certains fichiers csv trop volumineux, qui pouvaient faire jusqu'à 2gb. Ces derniers prenaient beaucoup de temps à s'ouvrir, lorsqu'ils s'ouvraient et ensuite étaient inutilisables par les logiciels de visualisations classiques. Nous avons donc mis en place de nouvelles méthodes pour utiliser ces fichiers, comme indiqué ci-après.

1.2.2. Premiers pas

Dans un premier temps, pour visualiser les données et pallier le problème des fichiers trop volumineux inutilisables, nous avons fait un script bash, permettant d'insérer toutes les données dans une base MongoDB. Nous avons d'abord fait le choix d'une base objet, car on ne connaissait pas la structure des données. Cette méthode et les documentations fournies nous ont permis de visualiser et de naviguer dans les données de manière efficace.

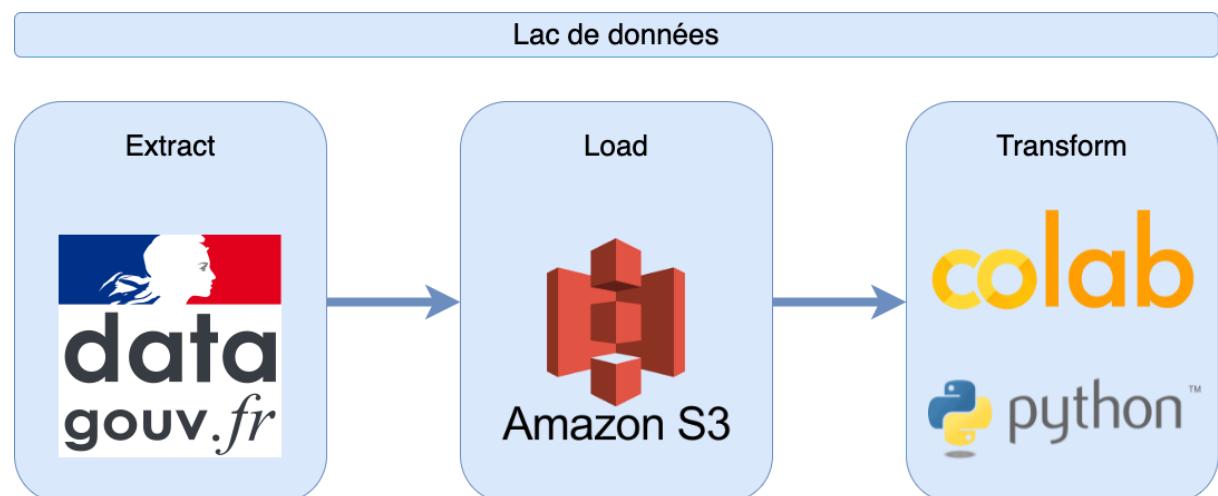
Après cette première approche, nous avons vu que les données étaient globalement très bien structurées. Par la suite, nous avons donc préféré utiliser une base Postgres, car plus à l'aise en sql pour transformer et préparer les données.

1.3. Lac et entrepôt de données

1.3.1. Lac de données AWS S3

Les données extraites de data gouv sont de plusieurs formats, comme le csv, json, geojson ou autre. De plus, Google Colab ne permet pas de stocker indéfiniment des fichiers.

C'est pourquoi, nous avons décidé de créer un lac de données, contenant toutes les données brutes. Pour ce faire, nous utilisons le service Simple Storage Service (S3), de Amazon Web Services (AWS). Nous avons fait le choix de cet outil, car il nous permet de stocker facilement tous types de données, de toutes tailles, à moindre coûts et ensuite, de les récupérer très facilement sur Google Colab.



Une fois sur Colab, les données sont transformées et analysées grâce aux outils Python Dash, Plotly, PySpark et GeoPandas. Ils permettent de faire des analyses statistiques, affichages sur cartes et autres.



plotly | Dash



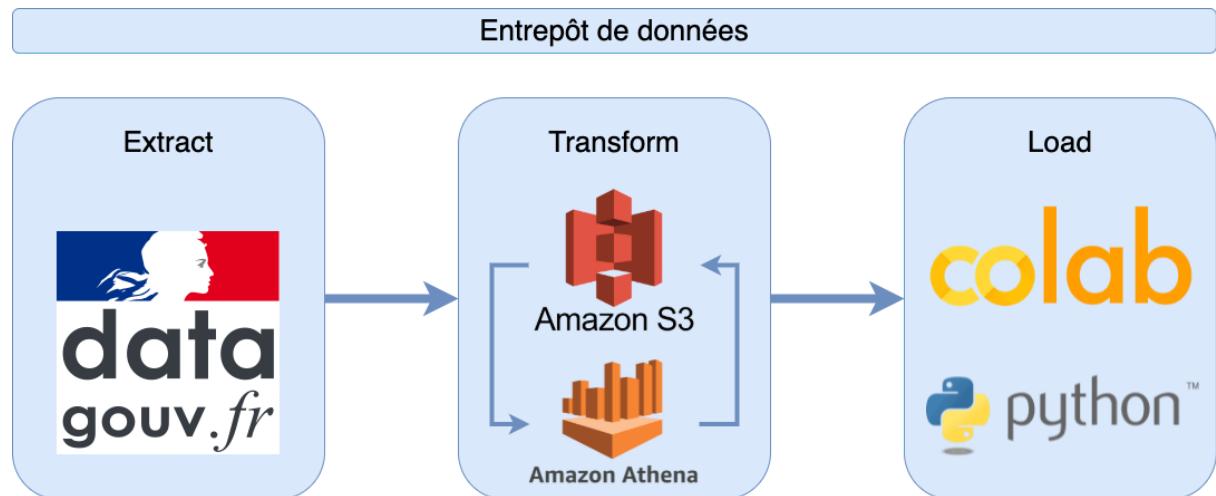
GeoPandas

1.3.2. Entrepôt de données AWS Athena et S3

Lors du téléchargement de fichiers en provenance du lac vers Colab, nous avons rencontré certains problèmes. En effet, des fichiers trop volumineux mettaient beaucoup de temps à télécharger, à cause d'informations pas forcément utiles. C'est pourquoi, nous avons souhaité sélectionner, transformer et réduire les données avant de les télécharger sur Colab. Nous avons donc créé un entrepôt de données à partir du lac précédemment évoqué.

Pour ce faire, nous avons utilisé la technologie Athena, d'Amazon Web Services. Cette dernière permet de récupérer les fichiers contenus dans S3, dans n'importe quel format (dans notre cas csv) ensuite, d'effectuer des requêtes sql dessus et enfin, de mettre les résultats dans une autre compartiment S3.

Pour ce faire, les données sont extraites du lac dans S3, transformées dans Athena, rendues disponibles dans un autre compartiment S3, pour être finalement téléchargées dans Colab.



2. Comment le nombre d'infrastructures de recharge évolue-t-il en France ?

Dans cette partie, nous allons tenter de voir comment les infrastructures de recharge évoluent et également l'évolution du parc de véhicules thermiques et électriques en France. Nous pourrons ainsi voir si l'évolution des infrastructures est proportionnelle ou du moins suffisante à l'évolution des voitures et ainsi avoir un premier élément de réponse à la problématique.

2.1. Etude du parc d'infrastructure de recharge

Premièrement, nous allons étudier le parc d'infrastructures de recharges. Nous allons tout d'abord voir le total des bornes de recharge par date en France. Nous verrons ensuite l'évolution de ce nombre au cours du temps, puis nous verrons une répartition de ces bornes par région et département.

2.1.1. Total bornes par date

Nous allons voir ici le total des bornes pour les années 2020, 2021 et 2022. Pour cela, nous faisons la somme de la colonne "nb_bornes" dans chacun de nos csv concernant les bornes et les dates voulues. Pour cela, nous avons au préalable traité et créé des csv par région et par département pour les bornes afin d'avoir les données triées par région ou département pour chaque date.

Nombre de bornes de recharge électriques en France en 2020: 9289
Nombre de bornes de recharge électriques en France en 2021: 16691
Nombre de bornes de recharge électriques en France en 2022: 22996

2.1.2. Évolution de l'infrastructure au cours du temps

Maintenant que nous avons la quantité de bornes par date, nous allons pouvoir voir l'évolution de l'infrastructure au cours du temps. Pour cela, nous allons voir l'évolution entre 2020/2021 et entre 2021/2022.

Évolution 2020/2021: 179.69%
Évolution 2021/2022: 137.77%

Nous pouvons voir qu'entre l'année 2020 et l'année 2021 le nombre de bornes a augmenté de 79,69 % et qu'entre 2021 et 2022 celui-ci a augmenté de 37,77 %.

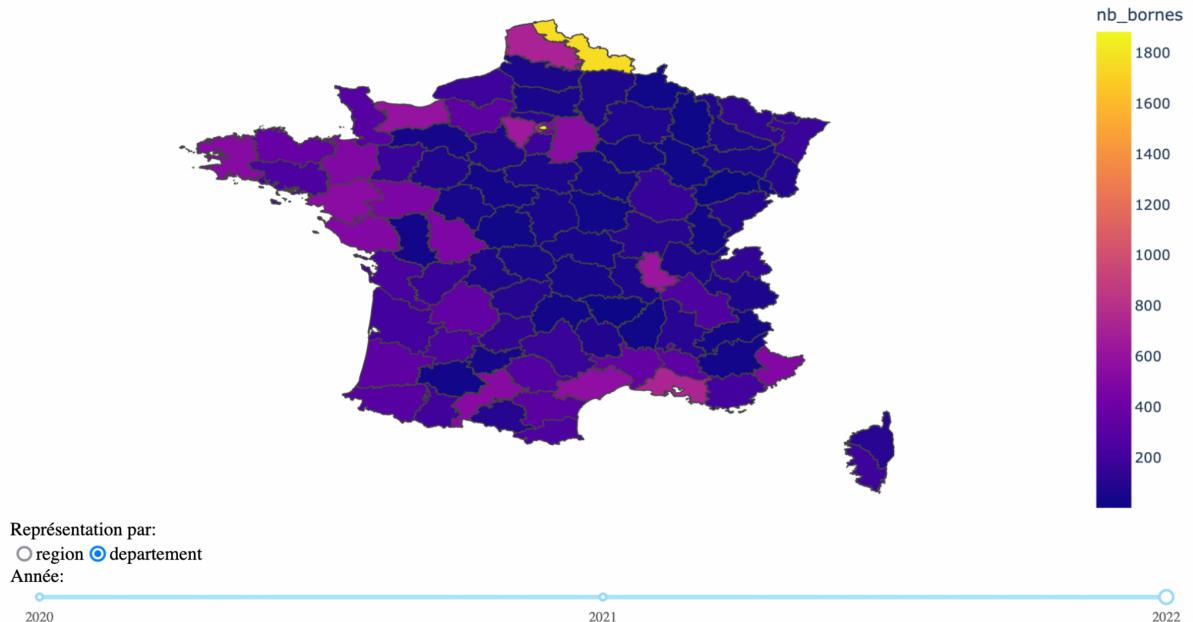
Attention: il est important de noter que nous n'avons pas les données complètes de 2022, nous n'avons que les 3 premiers trimestres, il faut donc prendre en compte que le nombre de bornes a donc augmenté de plus de 37,77 %.

Par conséquent, nous pouvons voir que l'infrastructure s'agrandit de façon rapide.

2.1.3. Répartition géographique des bornes

Nous allons maintenant voir une représentation de ces bornes par département/région et par dates. Ici nous avons une capture d'écran d'une carte interactive, si vous voulez la version interactive vous pouvez vous rendre [ici](#).

Nombre de bornes par département ou région



2.2. Étude du parc de voitures

Dans cette partie, nous allons faire l'état des lieux du parc de voitures français. Nous allons donc d'abord voir le nombre total de voitures en France selon 3 dates:

- 31/12/2020
- 31/12/2021
- 30/09/2022

Puis, nous allons voir l'évolution de ce nombre au cours du temps. Nous allons ensuite voir ces mêmes chiffres pour le parc de voitures électriques en France. Nous verrons finalement le rapport entre voitures thermiques et électriques et son évolution au cours du temps.

2.2.1. Total voitures par date

Nous allons voir dans cette partie l'état du parc de véhicules en France, puis voir son évolution de 2020/2021 et 2021/2022.

Premièrement, nous allons calculer le total des voitures en 2020, 2021 et 2022.

Nombre de voitures en France en 2020: 64618757
Nombre de voitures en France en 2021: 65679590
Nombre de voitures en France en 2022: 66189583

Le parc français de véhicules était ainsi constitué de 64 618 757 véhicules à la fin de l'année 2020, de 65 679 590 véhicules à la fin de l'année 2021 et de 66 189 583 véhicules à la fin du troisième trimestre de 2022.

Nous allons maintenant calculer l'évolution de ce parc de véhicules de 2020 à 2021 et de 2021 à 2022.

Évolution du parc de véhicules entre 2020 et 2021: 101.64167967514447
Évolution du parc de véhicules entre 2021 et 2022: 100.77648627221943

Comme nous pouvons le voir, on recense une augmentation de 1.64 % de véhicules entre l'année 2020 et 2021 et une augmentation de 0.78 % entre 2021 et 2022.

Attention : il est bon de noter que le chiffre de 2022 est daté de la fin du troisième trimestre et non pas de la fin de l'année, ce chiffre est donc logiquement plus grand en réalité que 0.78 %.

2.2.2. Total voitures électriques par date

Nous allons voir dans cette partie le nombre de voitures électriques par dates en France pour les années 2020, 2021 et 2022. Puis nous verrons l'évolution de cette quantité au cours du temps.

Voici l'état du parc électrique en 2020, 2021 et 2022:

Total de voitures électriques en 2020: 409523
Total de voitures électriques en 2021: 715077
Total de voitures électriques en 2022: 944942

Nous avons donc 409 523 voitures électriques en 2020, 715 077 en 2021 et 944 942 en 2022.

Nous allons maintenant voir l'évolution de la quantité des véhicules électriques entre 2020/2021 et 2021/2022.

Évolution entre 2020 et 2021: 174.61%
Évolution entre 2021 et 2022: 132.15%

Nous avons donc une augmentation de 74,61 % entre 2020 et 2021 et de 32,15 % entre 2021 et 2022. Il est à noter qu'il manque le dernier trimestre de 2022 dans nos données, cette augmentation est donc forcément plus grande en réalité. On voit donc que la quantité de véhicules électriques évolue très rapidement en France, le besoin en borne est donc conséquent et grandissant.

2.2.3. Rapport voitures électriques / voitures totales par date

Dans cette partie, nous allons voir le rapport entre les voitures électriques par rapport au total de voitures en France afin de voir la proportion qu'elles représentent par rapport au parc total de voitures en France et également voir son évolution au cours du temps. Pour cela, nous disposons des données pour 2020, 2021 et 2022.

Rapport entre voitures électriques et voitures en 2020: 0.63%
Rapport entre voitures électriques et voitures en 2021: 1.09%
Rapport entre voitures électriques et voitures en 2022: 1.43%

Nous constatons donc que les voitures électriques représentaient 0,63 % des voitures françaises en 2020, 1,09 % en 2021 et 1,43 % en 2022. Il est à noter que le chiffre de 1,43 % est sous-estimé puisqu'il manque les données du dernier trimestre de 2022. Cette donnée nous montre à quel point l'électrique progresse vite dans le parc de véhicules français.

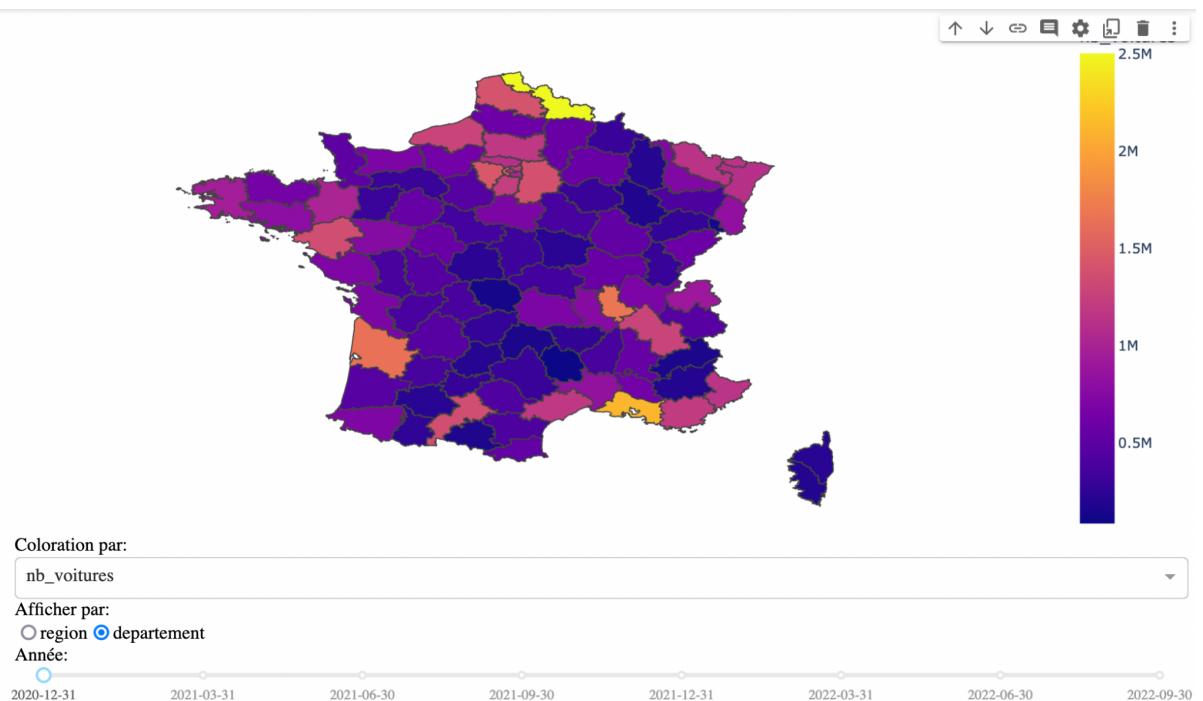
Nous allons maintenant voir l'évolution de ce rapport au cours du temps. Pour cela, nous allons voir les évolutions entre 2020/2021 et 2021/2022.

Évolution du rapport voitures électriques / toutes voitures entre 2020 et 2021: 171.79%
Évolution du rapport voitures électriques / toutes voitures entre 2021 et 2022: 131.13%

Nous pouvons donc constater que nous avons une augmentation de 71,79 % des voitures électriques par rapport aux voitures totales de 2020 à 2021 et une augmentation de 31.13 % entre 2021 et 2022. Il est à noter que nous n'avons pas les données complètes pour 2022, il manque le dernier trimestre de cette année-là.

2.2.4. Représentation géographique de la quantité de voitures (thermiques, électriques, gaz) par dates et par région ou département

Nous allons maintenant voir une représentation des voitures électriques, thermiques et à gaz par département/région et par dates. Ici nous avons une capture d'écran d'une carte interactive, si vous voulez la version interactive vous pouvez vous rendre [ici](#).



2.3. Existe-t-il un lien entre l'évolution du nombre de voitures électriques et l'évolution du nombre d'infrastructures de recharges ?

Dans cette partie, nous allons voir s'il existe un lien entre le nombre de voitures électriques et le nombre de bornes au cours du temps. Pour cela, nous allons calculer le nombre de bornes par voitures électriques en France pour les années 2020, 2021 et 2022.

2.3.1. Nombre de bornes par voitures électriques en France

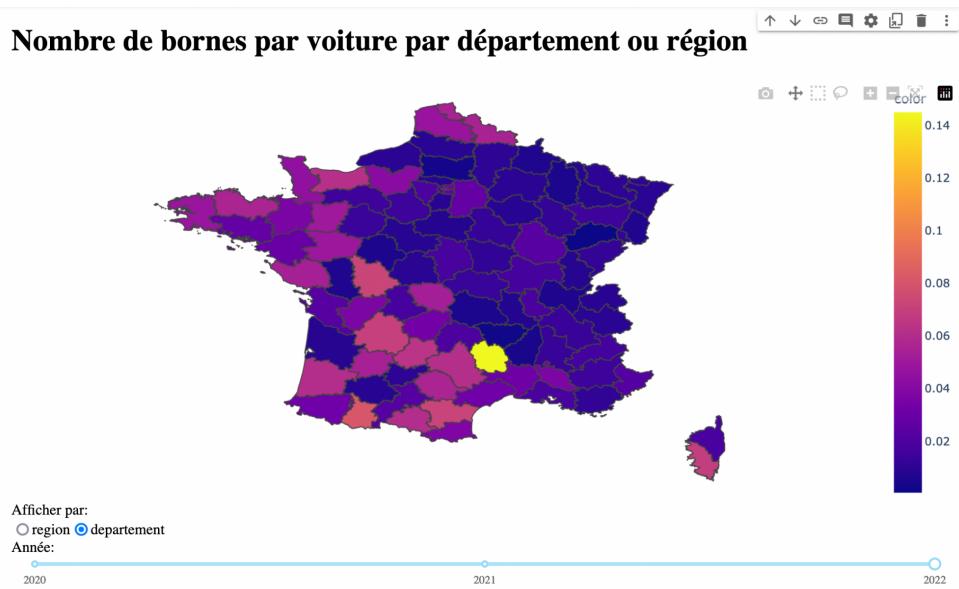
Dans cette partie, nous allons voir le nombre de bornes de recharges par voitures électriques en France sur les années 2020, 2021 et 2022.

En 2020, il y a 0.0227 bornes par voiture électrique
En 2021, il y a 0.0233 bornes par voiture électrique
En 2022, il y a 0.0243 bornes par voiture électrique

Nous pouvons voir que non seulement l'infrastructure maintient sa qualité au cours du temps alors que le nombre de voitures électriques augmente fortement(croissance presque exponentielle), mais en plus, celle-ci s'améliore, en effet le nombre de bornes par voiture augmente au cours du temps. De plus, il est à noter qu'il manque les données du dernier trimestre de 2022, nous pouvons donc aisément supposer que le nombre de bornes par voiture électrique est supérieur au 0,0243 qui est présenté dans les résultats. L'infrastructure a donc maintenu sa qualité au cours du temps, dans la partie suivante nous allons chercher à voir si l'infrastructure est bien répartie sur le territoire français à l'échelle des départements et des régions.

2.3.2. Représentation géographique du nombre de bornes par voitures électriques par date et par région/département

Nous allons maintenant voir une représentation du nombre de bornes par voitures électriques par département/région et par dates. Ici nous avons une capture d'écran d'une carte interactive, si vous voulez la version interactive vous pouvez vous rendre [ici](#).



3. La répartition des infrastructures de recharge sur le territoire correspond-elle à une logique de distribution de la population et de trafic routier ?

3.1 Existe-t-il une relation entre trafic, nombre d'habitants et nombre de bornes ?

Pour cette partie, nous commençons par réaliser des tests du chi2 pour chaque année et échelle géographique. Ces tests permettront de montrer si un lien existe entre les différentes variables concernées, ici les variables sont :

- trafic
- nombre de bornes
- population

Nous commençons par énoncer les hypothèses. Pour chaque test, les hypothèses seront les suivantes :

- H0 (hypothèse nulle) : les deux variables sont indépendantes.
- H1 (hypothèse alternative) : les deux variables ne sont pas indépendantes.

Nous pouvons donc effectuer les tests au seuil de 95 %. C'est-à-dire que si la p-value, la probabilité pour un modèle statistique donnée sous l'hypothèse nulle d'obtenir une valeur au moins aussi extrême que celle observée, est inférieur à 0,05, alors H0 est rejeté.

Il est à noter que les données sur le trafic et la population sont sur les années 2017, 2018, 2019, tandis que les données sur les bornes sont sur les années 2020, 2021, 2022. Il est donc important de se le rappeler lors de l'interprétation des résultats, qui pourrait être légèrement biaisé par cette différence. Nous utiliserons le tableau de correspondance ci-dessous :

Année	1	2	3
Trafic	2017	2018	2019
Population	2017	2018	2019
Bornes	2020	2021	2022

Résultat des tests du chi2

	Échelle régionale					
Année	1		2		3	
	p-value	H0	p-value	H0	p-value	H0
Lien Population-Trafic	0.493	✗	0.493	✗	0.493	✗
Lien Population-Bornes	0.187	✗	0.187	✗	0.187	✗
Lien Trafic-Bornes	0.739	✗	0.534	✗	0.299	✗

	Échelle départementale					
Année	1		2		3	
	p-value	H0	p-value	H0	p-value	H0
Lien Population-Trafic	1.59e-8	✓	3.3e-12	✓	2.7e-11	✓
Lien Population-Bornes	0.008	✓	5.19e-5	✓	2.30e-9	✓
Lien Trafic-Bornes	0.386	✗	0.017	✓	0.001	✓

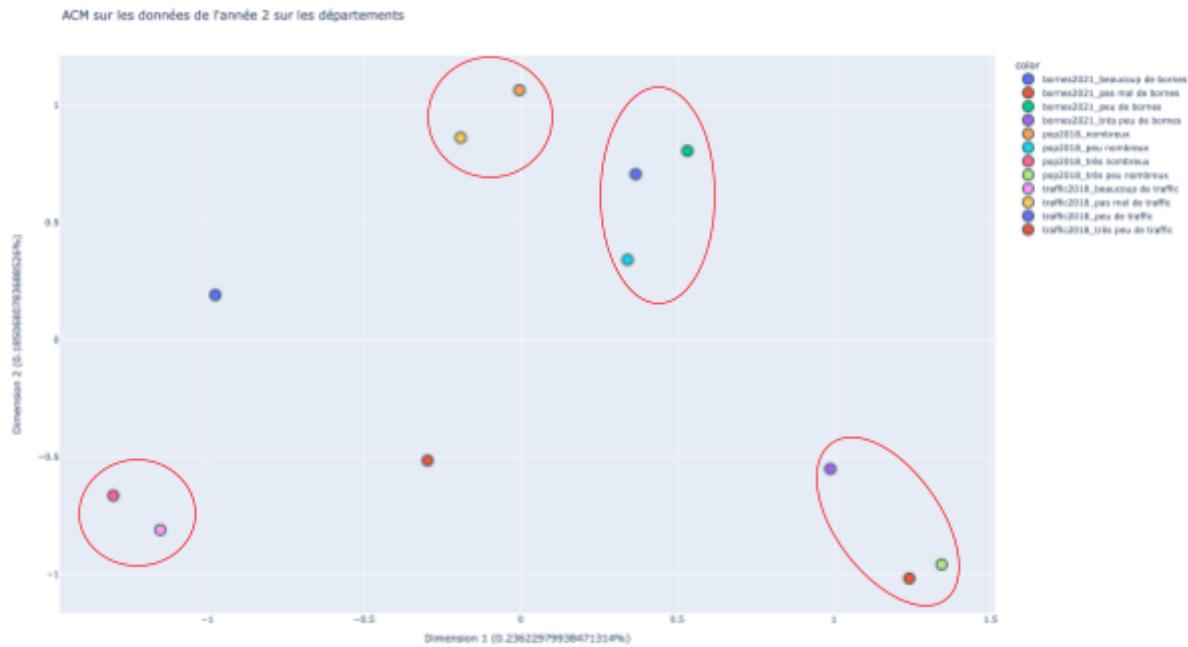
Nous constatons que les variables ne sont pas liées sur les dataframes des régions, ce qui est sûrement dû au fait que nous avons trop peu de valeur (1 par région - 13) et qu'elles ne sont pas assez dispersées (variation de 1 entre 1 et 4). Par ailleurs, les variables sont toutes liées sur les dataframes des départements sauf pour la 1ere année où la variable trafic n'est pas liée à la variable bornes.

Nous pouvons nous poser la question de la pertinence de ces tests sur les dataframes des régions, ainsi que du résultat obtenu sur le dataframe des départements pour la 1ere année. En effet, c'est la seule année où les variables trafic et bornes ne sont pas liées.

Nous allons donc réaliser une ACM chaque année sur les dataframes des départements afin de voir comment les variables sont liées. Les graphes obtenus sont disponible [ici](#) en version interactive.

Résultat des ACMs

Pour l'année 2 nous obtenons :



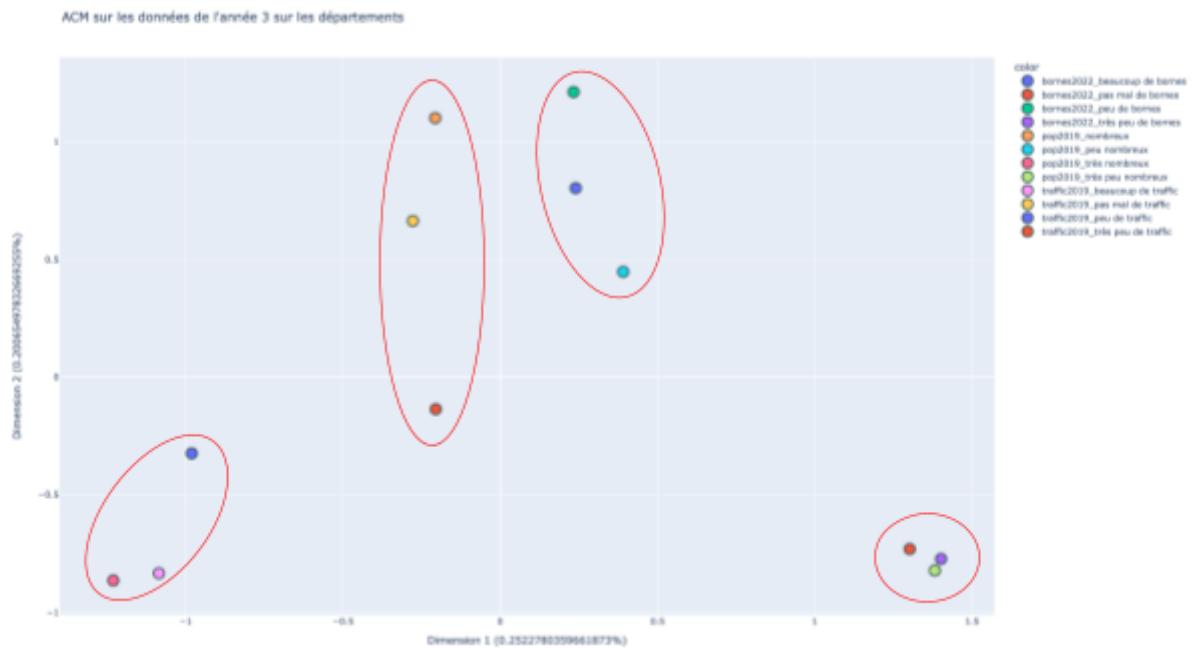
Pour cette première ACM, nous voyons un cluster assez resserré en bas à droite. Il est composé des individus représentants une grande population et un grand nombre de voitures électriques. Pour ce groupe, nous aurions préféré qu'il contienne aussi le point bleu représentant un très grand nombre de bornes de recharges.

Dans la même logique, le groupement en bas à droite indique un lien entre très petite population, très peu de trafic et très peu de bornes. Il y a donc très peu de bornes lorsqu'il y a très peu de besoins. Les bornes inutiles sont donc évitées.

Le groupement en haut à droite est composé des individus moyens bas pour la taille de la population, le nombre de voitures électriques et les bornes. Pour le groupement à sa gauche, ce sont les quantités moyennes hautes, pour les individus population et voitures. Comme pour les départements où il y a une très grande population, nous aurions ici espéré que le point rouge correspondant à "pas mal de bornes", soit avec ce groupe.

En conclusion, cette ACM révèle des liens cohérents pour presque tous les individus. Les deux groupements de droite, avec une plus petite population, sont tout de même plus resserrés que ceux à plus grande population (de gauche). Dans les prochaines années, nous pouvons espérer une évolution qui indiquera un resserrement de ces variables. Les besoins en bornes des départements avec un grand nombre de voitures électriques seront donc plus satisfaits.

Pour l'année 3 nous obtenons :



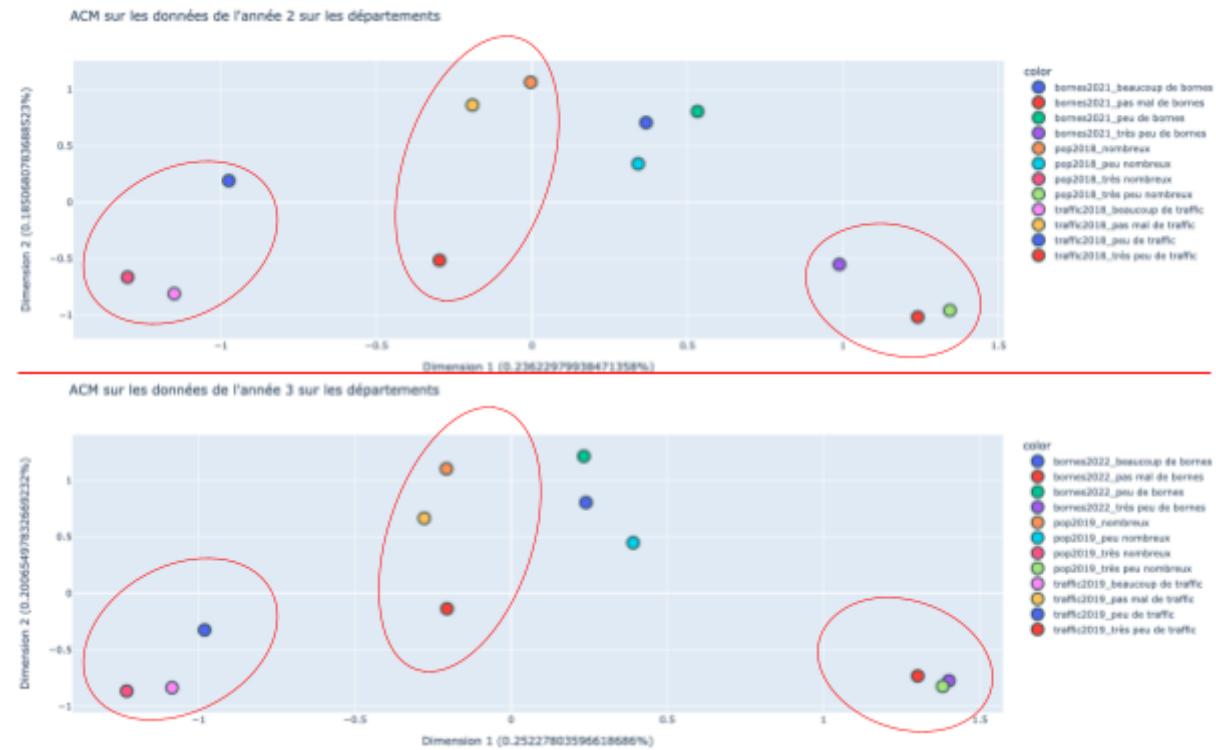
Pour cette ACM, nous distinguons un premier groupe très concentré en bas à droite. Ce dernier est composé des individus "très peu de population", "très peu de trafic" et "très peu de bornes". Ce premier paraît assez logique et cohérent avec ce que nous pourrions attendre en réalité.

Un second groupe un peu plus dispersé, de trois individus, se dégage en bas à droite. Il est formé par les individus correspondants à une population très nombreuse, beaucoup de trafic et beaucoup de bornes de recharges. Dans le cadre de ce projet, ce groupement est révélateur, car il montre que les départements à forte concentration de bornes de recharges sont placés là où il y a le plus de voitures électriques et donc de besoins.

Nous distinguons deux autres groupements de trois individus chacuns, représentés par les cercles plus en haut. Le groupe de gauche semble assez logique, car il représente une population nombreuse, avec pas mal de trafic et pas mal de bornes. Le groupe de droite est aussi très cohérent, car il représente peu de population, peu de trafic et peu de bornes.

Dans l'ensemble, cette ACM est logique et claire. Elle confirme les résultats que nous aurions pu prédire instinctivement au départ.

Comparaison entre l'année 2 et l'année 3 :



Entre l'année 2 (au-dessus) et l'année 3 (au-dessous), nous voyons que les groupes se resserrent. Ces résultats sont encourageants, car ils montrent que de nouvelles bornes sont placées aux endroits où il y a le plus de besoins.

Pour les deux groupes de gauche, les individus correspondant au nombre de bornes se rapprochent progressivement des individus aux quantités similaires de voitures électriques et de population. Au contraire, à droite, là où il y a peu de besoins, il ne doit pas y avoir de nouvelles bornes et cet individu se rapproche donc de son groupe.

Ces résultats sont révélateurs d'une évolution logique et cohérente avec les besoins des français.

3.2 Existe-t-il une relation entre nombre de véhicules électriques et nombre de bornes ?

Nous nous intéressons maintenant à une relation entre nombre de véhicules électriques et nombre de bornes.

La variable quantitative pour notre étude sera le nombre de véhicules électriques et notre variable qualitative sera le nombre de bornes (quartile correspondant).

Pour étudier l'existence d'une telle relation, nous allons réaliser une anova pour chaque année et chaque échelle géographique. Ainsi, nous pourrons voir si le nombre de véhicules électriques est impacté par le nombre de bornes. Pour chaque anova, les hypothèses sont les suivantes :

- H0 (hypothèse nulle) : le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques
- H1 (hypothèse alternative) : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques

Résultat de l'Anova sur les régions, année 2020

	sum_sq	df	F	PR(>F)
bornes2020	2.464084e+09	1.0	5.838372	0.034232
Residual	4.642549e+09	11.0		

La p-value est de 0.034232, cela signifie qu'il y a une probabilité de 3.4% que les différences observées entre les groupes soient dues au hasard. Comme la p-value est inférieure à notre seuil de signification (0.05), nous rejetons l'hypothèse nulle H0 et nous acceptons l'hypothèse alternative **H1 : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques.**

Résultat de l'Anova sur les départements, année 2020

	sum_sq	df	F	PR(>F)
bornes2020	3.660508e+08	1.0	21.574245	0.000014
Residual	1.323428e+09	78.0		

La p-value est de 0.000014, cela signifie qu'il y a une probabilité de 0.001 % que les différences observées entre les groupes soient dues au hasard. Comme la p-value est inférieure à notre seuil de signification (0.05), nous rejetons l'hypothèse nulle H0 et nous acceptons l'hypothèse alternative **H1 : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques.**

Résultat de l'Anova sur les régions, année 2021

	sum_sq	df	F	PR(>F)
bornes2021	7.449633e+09	1.0	6.862653	0.023843
Residual	1.194086e+10	11.0		

La p-value est de 0.023843, cela signifie qu'il y a une probabilité de 2.4 % que les différences observées entre les groupes soient dues au hasard. Comme la p-value est inférieure à notre seuil de signification (0.05), nous rejetons l'hypothèse nulle H0 et nous acceptons l'hypothèse alternative **H1 : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques.**

Résultat de l'Anova sur les départements, année 2021

	sum_sq	df	F	PR(>F)
bornes2021	1.459012e+09	1.0	36.994851	2.603065e-08
Residual	3.667756e+09	93.0		

La p-value est de 2.603065e-08, cela signifie qu'il y a une probabilité d'environ 0 % que les différences observées entre les groupes soient dues au hasard. Comme la p-value est inférieure à notre seuil de signification (0.05), nous rejetons l'hypothèse nulle H0 et nous acceptons l'hypothèse alternative **H1 : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques.**

Résultat de l'Anova sur les régions, année 2022

	sum_sq	df	F	PR(>F)
bornes2022	7.729047e+09	1.0	3.559	0.085884
Residual	2.388860e+10	11.0		

La p-value est de 0.085884, cela signifie qu'il y a une probabilité de 8.6 % que les différences observées entre les groupes soient dues au hasard. Comme la p-value est supérieur à notre seuil de signification (0.05), nous rejetons l'hypothèse alternative H1 et nous acceptons l'hypothèse nulle **H0 : le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.**

Résultat de l'Anova sur les départements, année 2022

	sum_sq	df	F	PR(>F)
bornes2022	3.196612e+09	1.0	55.869859	3.977556e-11
Residual	5.378240e+09	94.0		

La p-value est de 3.977556e-11, cela signifie qu'il y a une probabilité d'environ 0 % que les différences observées entre les groupes soient dues au hasard. Comme la p-value est supérieur à notre seuil de signification (0.05), nous rejetons l'hypothèse alternative H1 et nous acceptons l'hypothèse nulle **H0 : le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.**

Pour conclure, les résultats ont montré que lorsque que nous étudions les données sur l'échelle de la région, la p-value est proche du seuil de signification de 5 %. De plus, sur l'année 2022, nous rejetons l'hypothèse H1. Nous pouvons alors se demander si l'échelle géographique des régions n'est pas un peu trop grande. Par ailleurs, les données de 2022 s'arrêtent à l'avant dernier trimestre (septembre), et cela peut expliquer cette différence par rapport aux autres anova.

Lorsque nous nous penchons sur l'échelle des départements, nous observons dans tous les cas une p-value largement inférieure au seuil de 5 %.

Nous pouvons alors affirmer, d'après notre étude, que **le nombre de bornes dans les départements impact le nombre de véhicules électriques**, et ce au fil des années.

3.3 Existe-t-il une relation entre nombre de véhicules électriques, type de population et nombre de bornes ?

Nous souhaitons à présent étudier l'impact, ou non, du nombre de bornes et du type population sur le nombre de véhicules électriques. Pour cela, nous avons choisi trois facteurs différents pour catégoriser la population, l'âge, le taux de chômage et la richesse.

3.3.1 Taux de chômage

Nous nous intéressons maintenant à une relation entre le nombre de véhicules électriques, le nombre de bornes et le taux de chômage.

La variable quantitative pour notre étude sera le nombre de véhicules électriques et nos variables qualitatives seront le nombre de bornes (quartile correspondant) et le taux de chômage de la population (quartile correspondant).

Pour étudier l'existence d'une telle relation, nous allons réaliser une anova à deux facteurs pour chaque année et chaque échelle géographique. Ainsi, nous pourrons voir si le nombre de véhicules électriques est impacté par le nombre de bornes, par le taux de chômage, ou les deux. Pour chaque anova, les hypothèses sont les suivante :

- H01 (hypothèse nulle) : le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques

- H11 (hypothèse alternative) : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques
- H02 (hypothèse nulle) : le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques
- H12 (hypothèse alternative) : le facteur du taux de chômage a un impact sur le nombre de véhicules électriques
- H03 (hypothèse nulle) : le facteur de l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques
- H13 (hypothèse alternative) : le facteur de l'interaction entre taux de chômage et nombre de bornes a un impact sur le nombre de véhicules électriques

Résultat de l'Anova sur les régions, année 2020

	sum_sq	df	F	PR(>F)
bornes2020	2.650568e+09	1.0	7.023117	0.026471
taux_chomage_2020	3.185662e+08	1.0	0.844094	0.382191
Interaction	9.273273e+08	1.0	2.457107	0.151438
Residual	3.396655e+09	9.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.026471, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur du taux de chômage, la p-value est de 0.382191, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre taux de chômage et nombre de bornes, la p-value est de 0.151438, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les départements, année 2020

	sum_sq	df	F	PR(>F)
bornes2020	3.653722e+08	1.0	21.895640	0.000012
taux_chomage_2020	3.434974e+07	1.0	2.058475	0.155464
Interaction	2.086750e+07	1.0	1.250525	0.266974
Residual	1.268211e+09	76.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.000012, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur du taux de chômage, la p-value est de 0.155464, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre taux de chômage et nombre de bornes, la p-value est de 0.266974, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les régions, année 2021

	sum_sq	df	F	PR(>F)
bornes2021	8.645882e+09	1.0	9.329314	0.013695
taux_chomage_2021	1.349569e+09	1.0	1.456249	0.258288
Interaction	2.250596e+09	1.0	2.428499	0.153575
Residual	8.340692e+09	9.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.013695, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur du taux de chômage, la p-value est de 0.258288, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse

nulle (H02) et que le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre taux de chômage et nombre de bornes, la p-value est de 0.153575, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les départements, année 2021

	sum_sq	df	F	PR(>F)
bornes2021	1.467387e+09	1.0	37.304480	2.455313e-08
taux_chomage_2021	3.584257e+07	1.0	0.911204	3.423241e-01
Interaction	5.239052e+07	1.0	1.331892	2.514909e-01
Residual	3.579523e+09	91.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 2.455313e-08, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur du taux de chômage, la p-value est de 0.3423241, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre taux de chômage et nombre de bornes, la p-value est de 0.2514909, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les régions, année 2022

	sum_sq	df	F	PR(>F)
bornes2022	4.157856e+09	1.0	1.824277	0.209783
taux_chomage_2022	2.931664e+09	1.0	1.286280	0.286039
Interaction	4.443080e+08	1.0	0.194942	0.669252
Residual	2.051263e+10	9.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.209783, ce qui est supérieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse alternative (H11) et accepter l'hypothèse nulle (H01), c'est-à-dire que le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente et éventuellement un problème sur les données de 2022 (manque d'un trimestre).

Pour le facteur du taux de chômage, la p-value est de 0.286039, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre taux de chômage et nombre de bornes, la p-value est de 0.669252, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre qu'**aucun facteur n'a d'impact sur le nombre de véhicules électriques**.

Résultat de l'Anova sur les départements, année 2022

	sum_sq	df	F	PR(>F)
bornes2022	3.186640e+09	1.0	54.913853	5.949502e-11
taux_chomage_2022	2.225900e+07	1.0	0.383579	5.372249e-01
Interaction	1.723954e+07	1.0	0.297081	5.870372e-01
Residual	5.338741e+09	92.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 5.949502e-11, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur du taux de chômage, la p-value est de 0.5372249, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter

l'hypothèse nulle (H02) et que le facteur du taux de chômage n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre taux de chômage et nombre de bornes, la p-value est de 0.5870372, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre taux de chômage et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

3.3.2 Age

Nous nous intéressons maintenant à une relation entre le nombre de véhicules électriques, le nombre de bornes et l'âge de la population.

La variable quantitative pour notre étude sera le nombre de véhicules électriques et nos variables qualitatives seront le nombre de bornes (quartile correspondant) et l'âge de la population (quartile correspondant).

Il est à noter que les données sur la population sont sur les années 2017, 2018, 2019, tandis que les données sur les bornes et sur les voitures sont sur les années 2020, 2021, 2022. Il est donc important de se le rappeler lors de l'interprétation des résultats, qui pourrait être légèrement biaisé par cette différence. Nous utiliserons le tableau de correspondance ci-dessous :

Année	1	2	3
Population	2017	2018	2019
Voitures	2020	2021	2022
Bornes	2020	2021	2022

Pour étudier l'existence d'une telle relation, nous allons réaliser une anova à deux facteurs pour chaque année et chaque échelle géographique. Ainsi, nous pourrons voir si le nombre de véhicules électriques est impacté par le nombre de bornes, par l'âge de la population, ou les deux. Pour chaque anova, les hypothèses sont les suivante :

- H01 (hypothèse nulle) : le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques
- H11 (hypothèse alternative) : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques
- H02 (hypothèse nulle) : le facteur de l'âge n'a pas d'impact sur le nombre de véhicules électriques
- H12 (hypothèse alternative) : le facteur de l'âge a un impact sur le nombre de véhicules électriques
- H03 (hypothèse nulle) : le facteur de l'interaction entre âge et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques

- H13 (hypothèse alternative) : le facteur de l'interaction entre âge et nombre de bornes a un impact sur le nombre de véhicules électriques

Résultat de l'Anova sur les régions, année 1

	sum_sq	df	F	PR(>F)
bornes2020	6.166050e+09	1.0	16.640467	0.002217
age_group_2017	6.927026e+08	1.0	1.869413	0.201487
Interaction	4.459681e+08	1.0	1.203545	0.298329
Residual	3.705455e+09	10.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.002217, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de l'âge, la p-value est de 0.201487, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur de l'âge n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre âge et nombre de bornes, la p-value est de 0.298329, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre âge de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les départements, année 1

	sum_sq	df	F	PR(>F)
bornes2020	1.473699e+08	1.0	8.800120	0.004172
age_group_2017	1.418043e+08	1.0	8.467776	0.004902
Interaction	5.028190e+06	1.0	0.300256	0.585543
Residual	1.122005e+09	67.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.004172, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de l'âge, la p-value est de 0.004902, ce qui est aussi inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse alternative (H12) et que le facteur de l'âge a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre âge et nombre de bornes, la p-value est de 0.585543, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre âge de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de l'âge, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

Résultat de l'Anova sur les régions, année 2

	sum_sq	df	F	PR(>F)
bornes2021	1.808550e+10	1.0	18.703876	0.001502
age_group_2018	1.855053e+09	1.0	1.918481	0.196142
Interaction	1.013812e+09	1.0	1.048476	0.329990
Residual	9.669387e+09	10.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.001502, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de l'âge, la p-value est de 0.196142, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur de l'âge n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre âge et nombre de bornes, la p-value est de 0.329990, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre âge de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les départements, année 2

	sum_sq	df	F	PR(>F)
bornes2021	6.919566e+08	1.0	18.573587	0.000045
age_group_2018	4.451448e+08	1.0	11.948633	0.000869
Interaction	5.550365e+07	1.0	1.489836	0.225740
Residual	3.054900e+09	82.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.000045, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de l'âge, la p-value est de 0.000869, ce qui est aussi inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse alternative (H12) et que le facteur de l'âge a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre âge et nombre de bornes, la p-value est de 0.225740, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre âge de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de l'âge, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

Résultat de l'Anova sur les régions, année 3

	sum_sq	df	F	PR(>F)
bornes2022	2.513003e+10	1.0	13.617352	0.004176
age_group_2019	1.543858e+09	1.0	0.836579	0.381902
Interaction	3.943661e+09	1.0	2.136974	0.174487
Residual	1.845442e+10	10.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.004176, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de l'âge, la p-value est de 0.381902, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur de l'âge n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre âge et nombre de bornes, la p-value est de 0.174487, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre âge de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **seul le facteur du nombre de bornes a un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs n'ont pas d'impact significatif.

Résultat de l'Anova sur les départements, année 3

	sum_sq	df	F	PR(>F)
bornes2022	1.672769e+09	1.0	29.578052	5.301661e-07
age_group_2019	5.557152e+08	1.0	9.826208	2.378887e-03
Interaction	4.168838e+07	1.0	0.737138	3.930520e-01
Residual	4.694015e+09	83.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 5.301661e-07, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de l'âge, la p-value est de 2.378887e-03, ce qui est aussi inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse alternative (H12) et que le facteur de l'âge a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre âge et nombre de bornes, la p-value est de 0.393052, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre âge de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de l'âge, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

3.3.3 Richesse

Nous nous intéressons maintenant à une relation entre le nombre de véhicules électriques, le nombre de bornes et la richesse de la population.

La variable quantitative pour notre étude sera le nombre de véhicules électriques et nos variables qualitatives seront le nombre de bornes (quartile correspondant) et la richesse de la population (quartile correspondant).

Il est à noter que les données sur la richesse de la population sont sur les années 2017, 2018, 2019, tandis que les données sur les bornes et sur les voitures sont sur les années 2020, 2021, 2022. Il est donc important de se le rappeler lors de l'interprétation des

résultats, qui pourrait être légèrement biaisé par cette différence. Nous utiliserons le tableau de correspondance ci-dessous :

Année	1	2	3
Richesse	2017	2018	2019
Voitures	2020	2021	2022
Bornes	2020	2021	2022

Pour étudier l'existence d'une telle relation, nous allons réaliser une anova à deux facteurs pour chaque année et chaque échelle géographique. Ainsi, nous pourrons voir si le nombre de véhicules électriques est impacté par le nombre de bornes, par la richesse de la population, ou les deux. Pour chaque anova, les hypothèses sont les suivante :

- H01 (hypothèse nulle) : le facteur du nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques
- H11 (hypothèse alternative) : le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques
- H02 (hypothèse nulle) : le facteur de la richesse n'a pas d'impact sur le nombre de véhicules électriques
- H12 (hypothèse alternative) : le facteur de la richesse a un impact sur le nombre de véhicules électriques
- H03 (hypothèse nulle) : le facteur de l'interaction entre richesse et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques
- H13 (hypothèse alternative) : le facteur de l'interaction entre richesse et nombre de bornes a un impact sur le nombre de véhicules électriques

Résultat de l'Anova sur les régions, année 1

	sum_sq	df	F	PR(>F)
bornes2020	3.459146e+09	1.0	13.731524	0.004878
richesse_2017	2.104926e+09	1.0	8.355775	0.017866
Interaction	2.704076e+08	1.0	1.073418	0.327209
Residual	2.267215e+09	9.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.004878, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de la richesse, la p-value est de 0.017866, ce qui est inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse alternative (H12) et que le facteur de richesse a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre richesse et nombre de bornes, la p-value est de 0.327209, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre richesse de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de la richesse, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

Résultat de l'Anova sur les départements, année 1

	sum_sq	df	F	PR(>F)
bornes2020	2.979884e+08	1.0	22.378575	0.000012
richesse_2017	3.498331e+08	1.0	26.272046	0.000003
Interaction	2.684669e+07	1.0	2.016155	0.160269
Residual	8.921580e+08	67.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.004172, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de la richesse, la p-value est de 0.004902, ce qui est aussi inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur de la richesse a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre richesse et nombre de bornes, la p-value est de 0.298329, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre richesse de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de la richesse, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

Résultat de l'Anova sur les régions, année 2

	sum_sq	df	F	PR(>F)
bornes2021	1.122386e+10	1.0	22.329074	0.001081
richesse_2018	6.208178e+09	1.0	12.350728	0.006573
Interaction	1.208768e+00	1.0	2.404757	0.155380
Residual	4.523912e+09	9.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.001081, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H_01) et accepter l'hypothèse alternative (H_11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de la richesse, la p-value est de 0.006573, ce qui est inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse alternative (H_12) et que le facteur de la richesse a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre richesse et nombre de bornes, la p-value est de 0.155380, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H_03) et que l'interaction entre richesse de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de la richesse, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

Résultat de l'Anova sur les départements, année 2

	sum_sq	df	F	PR(>F)
bornes2021	9.885992e+08	1.0	32.375354	1.906586e-07
richesse_2018	9.808864e+08	1.0	32.122769	2.092921e-07
Interaction	7.074681e+07	1.0	2.316867	1.318247e-01
Residual	2.503915e+09	82.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 1.906586e-07, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H_01) et accepter l'hypothèse alternative (H_11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de la richesse, la p-value est de 2.092921e-07, ce qui est aussi inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H_02) et que le facteur de la richesse a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre richesse et nombre de bornes, la p-value est de 0.131825, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre richesse de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, et celui de la richesse, ont un impact significatif sur le nombre de véhicules électriques**, tandis que le facteur de l'interaction n'a pas d'impact significatif.

Résultat de l'Anova sur les régions, année 3

	sum_sq	df	F	PR(>F)
bornes2022	1.060764e+10	1.0	5.358663	0.045869
richesse_2019	5.796342e+09	1.0	2.928139	0.121209
Interaction	2.764764e+08	1.0	0.139668	0.717263
Residual	1.781578e+10	9.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 0.045869, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de la richesse, la p-value est de 0.121209, ce qui est supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse alternative (H12) et que le facteur de la richesse n'a pas d'impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre richesse et nombre de bornes, la p-value est de 0.717263, ce qui est également supérieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H03) et que l'interaction entre richesse de la population et nombre de bornes n'a pas d'impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes est le seul à avoir un impact significatif sur le nombre de véhicules électriques**, tandis que les autres facteurs de l'interaction n'ont pas d'impact significatif.

Résultat de l'Anova sur les départements, année 3

	sum_sq	df	F	PR(>F)
bornes2022	1.762554e+09	1.0	38.936076	1.322558e-08
richesse_2019	1.002435e+09	1.0	22.144517	8.888012e-06
Interaction	2.111592e+08	1.0	4.664659	3.338953e-02
Residual	4.164645e+09	92.0		

Nous pouvons voir que la p-value pour le facteur du nombre de bornes est de 1.322558e-08, ce qui est inférieur à notre seuil de significativité de 0.05. Cela signifie que nous pouvons rejeter l'hypothèse nulle (H01) et accepter l'hypothèse alternative (H11), c'est-à-dire que le facteur du nombre de bornes a un impact sur le nombre de véhicules électriques. Ceci confirme l'étude faite dans la partie précédente.

Pour le facteur de la richesse, la p-value est de 8.888012e-06, ce qui est aussi inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H02) et que le facteur de la richesse a un impact sur le nombre de véhicules électriques.

Enfin, pour le facteur de l'interaction entre richesse et nombre de bornes, la p-value est de 3.338953e-02, ce qui est également inférieur à notre seuil de significativité. Cela signifie que nous ne pouvons pas rejeter l'hypothèse nulle (H13) et que l'interaction entre richesse de la population et nombre de bornes a un impact sur le nombre de véhicules électriques.

En résumé, notre analyse montre que **le facteur du nombre de bornes, celui de la richesse et l'interaction entre les deux, ont un impact significatif sur le nombre de véhicules électriques.**

3.3.4 Conclusion de nos analyses

Nous avons donc analysé d'éventuels impacts du nombre de bornes et du type de population (taux de chômage, richesse, âge) sur le nombre de véhicules électriques.

Voici un tableau récapitulatif de nos résultats:

Année		1		2		3	
Échelle géographique		Dep	Reg	Dep	Reg	Dep	Reg
Chômage	Bornes	✓	✓	✓	✓	✓	✓
	Chômage	✗	✗	✗	✗	✗	✗
	Bornes & Chômage	✗	✗	✗	✗	✗	✗
Âge	Bornes	✓	✓	✓	✓	✓	✓
	Âge	✓	✗	✓	✗	✓	✗
	Bornes & Âge	✗	✗	✗	✗	✗	✗
Richesse	Bornes	✓	✓	✓	✓	✓	✓
	Richesse	✓	✓	✓	✓	✓	✗
	Bornes & Richesse	✗	✗	✗	✗	✓	✗

D'après nos résultats, il semble assez clair que le nombre de bornes à toujours un impact sur le nombre de véhicules électriques. Ceci valide les résultats obtenus dans la partie [3.2](#).

Par ailleurs, nous remarquons que le taux de chômage n'a aucun impact sur le nombre de véhicules électriques. Au contraire, la richesse de la population (basée sur le salaire médian dans la zone géographique) à un impact sur le nombre de véhicules électriques. Pour l'âge, c'est un peu plus mitigé. Nous remarquons un impact seulement dans l'échelle géographique des départements. Ceci peut encore une fois être dû à une échelle trop englobante pour les régions.

Nos résultats laissent donc penser que pour pouvoir rouler en électrique, il faudrait plutôt avoir un salaire élevé, et ainsi ne pas être trop jeune (petit salaire), ni trop vieux (petite retraite). Ainsi nous comprenons que certains types de population sont plus aptes à acquérir des véhicules électriques et donc que l'augmentation du nombre de bornes peut être ciblé en fonction de la population locale.

4. Conclusion

Au fil de notre étude, nous avons remarqué que le parc de bornes de recharges s'agrandit de façon rapide, et qu'il en est de même pour le parc de voitures électriques. De plus, le nombre de bornes à suivi l'augmentation du nombre de voitures électriques, en gardant le rapport (nombre de bornes) / (nombre de voitures) constant au fil des années (voir [2.3.1](#)).

Néanmoins, l'augmentation du nombre de bornes s'est adapté au besoin des français. En effet, nous avons observé une augmentation ciblée du parc de bornes de recharges. Premièrement, cette augmentation est fortement liée au nombre d'habitants, mais également au trafic routier. Dans la partie [3.1](#), nous avons pu observer une clusterisation de l'augmentation du nombre de bornes, adapté au trafic et au nombre d'habitants.

De plus, nous avons constaté que le nombre de bornes avait un fort impact sur le nombre de véhicules électriques. Mais le type de la population est aussi un facteur impactant le nombre de véhicules électriques. En effet, d'après les résultats présentés [ici](#), nous remarquons que le taux de chômage n'a aucun impact sur le nombre de véhicules électriques. Au contraire, la richesse de la population et l'âge ont un impact.

Il faut toutefois noter que notre étude a pu être impactée par un manque de données, ou encore par une différence entre les années de mesures de nos données (3 ans maximum).

De plus, nous avons remarqué que les données sur une échelle géographique trop large (régions) pouvait être imprécise. Il serait donc intéressant de se plonger dans une étude sur échelle plus petite (les communes par exemple) afin de vérifier les résultats obtenus pour les départements.

Pour rappel, notre problématique était de savoir si la répartition des infrastructures de recharges est adaptée en France. Nous pouvons donc dire, d'après notre étude, que la répartition des infrastructures de recharges est adaptée aux besoins des français en termes de population et de trafic routier.

De plus, nous avons remarqué que le nombre de véhicules électriques est impacté par le nombre de bornes ou le type de population. En vue de la loi de l'Union Européenne mettant fin à la vente de véhicules thermiques en 2035, nous pouvons nous demander si une augmentation du nombre de bornes, permettant de dépasser les besoins des français en infrastructure de recharges, ne pourrait pas entraîner une augmentation du nombre de véhicules électriques en France.

Annexes

Code SQL utilisé dans AWS Athena, afin de transformer les données de l'entrepôt.

```
-- Code insee vers code nouvelle région
CREATE OR REPLACE VIEW insee_nouvelle_region as
SELECT "Code INSEE", "Code Postal", "Commune",
      "Code Département", "Département",
      "Anciens Code" as "Ancien Code Région", "Anciens Nom" as "Ancienne
Région",
      "Nouveau Code" as "Code Région", "Nouveau Nom" as "Région"
FROM correspondance_code_insee_code_postal
JOIN anciennes_nouvelles_regions ON
correspondance_code_insee_code_postal."Code Région" =
anciennes_nouvelles_regions."Anciens Code";


-- Global
CREATE materialized view borne_region_global as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_global_shape as
SELECT borne_region_global."Code Région", borne_region_global."Région",
date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_global
JOIN shape_region ON borne_region_global."Code Région" =
CAST(shape_region."Code_region" AS int);


-- 2022-09-30
CREATE materialized view borne_region_2022_09_30 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
```

```

= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2022-09-30'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2022_09_30_shape as
SELECT borne_region_2022_09_30."Code Région",
borne_region_2022_09_30."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2022_09_30
JOIN shape_region ON borne_region_2022_09_30."Code Région" =
CAST(shape_region."Code_region" AS int);

-- 2022-06-30
CREATE materialized view borne_region_2022_06_30 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_comunes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_comunes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2022-06-30'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2022_06_30_shape as
SELECT borne_region_2022_06_30."Code Région",
borne_region_2022_06_30."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2022_06_30
JOIN shape_region ON borne_region_2022_06_30."Code Région" =
CAST(shape_region."Code_region" AS int);

-- 2022-03-31
CREATE materialized view borne_region_2022_03_31 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_comunes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_comunes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2022-03-31'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2022_03_31_shape as

```

```

SELECT borne_region_2022_03_31."Code Région",
borne_region_2022_03_31."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2022_03_31
JOIN shape_region ON borne_region_2022_03_31."Code Région" =
CAST(shape_region."Code_region" AS int);

-- 2021-12-31
CREATE materialized view borne_region_2021_12_31 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2021-12-31'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2021_12_31_shape as
SELECT borne_region_2021_12_31."Code Région",
borne_region_2021_12_31."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2021_12_31
JOIN shape_region ON borne_region_2021_12_31."Code Région" =
CAST(shape_region."Code_region" AS int);

-- 2021-09-30
CREATE materialized view borne_region_2021_09_30 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2021-09-30'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2021_09_30_shape as
SELECT borne_region_2021_09_30."Code Région",
borne_region_2021_09_30."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2021_09_30
JOIN shape_region ON borne_region_2021_09_30."Code Région" =
CAST(shape_region."Code_region" AS int);

```

```

-- 2021-06-30
CREATE materialized view borne_region_2021_06_30 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2021-06-30'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2021_06_30_shape as
SELECT borne_region_2021_06_30."Code Région",
borne_region_2021_06_30."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2021_06_30
JOIN shape_region ON borne_region_2021_06_30."Code Région" =
CAST(shape_region."Code_region" AS int);

-- 2021-03-31
CREATE materialized view borne_region_2021_03_31 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,
sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2021-03-31'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2021_03_31_shape as
SELECT borne_region_2021_03_31."Code Région",
borne_region_2021_03_31."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2021_03_31
JOIN shape_region ON borne_region_2021_03_31."Code Région" =
CAST(shape_region."Code_region" AS int);

-- 2020-12-31
CREATE materialized view borne_region_2020_12_31 as
SELECT "Code Région", "Région", date_arrete, sum(nb_vp) as nb_voitures,
sum(nb_vp_rechargeables_el) as nb_voitures_elec,

```

```

sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region ON voitures_par_communes_par_energie.codgeo
= insee_nouvelle_region."Code INSEE"
WHERE date_arrete = '2020-12-31'
group by "Code Région", "Région", date_arrete
order by "Région", date_arrete desc;

CREATE OR REPLACE view borne_region_2020_12_31_shape as
SELECT borne_region_2020_12_31."Code Région",
borne_region_2020_12_31."Région", date_arrete, nb_voitures,
nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_region_2020_12_31
JOIN shape_region ON borne_region_2020_12_31."Code Région" =
CAST(shape_region."Code_region" AS int);
```

```sql
-- Code insee vers code nouvelle region
CREATE OR REPLACE VIEW insee_nouvelle_region_1 as
SELECT "Code INSEE", "Code Postal", "Commune",
      "Code Département", "Département",
      "Anciens Code" as "Ancien Code Région", "Anciens Nom" as "Ancienne
Région",
      "Nouveau Code" as "Code Région", "Nouveau Nom" as "Région"
FROM correspondance_code_insee_code_postal_1
JOIN anciennes_nouvelles_regions ON
correspondance_code_insee_code_postal_1."Code Région" =
anciennes_nouvelles_regions."Anciens Code";

-- global
CREATE materialized view borne_departement_global as
SELECT
      "Code Région", "Région",
      "Code Département", "Département",
      date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codgeo = insee_nouvelle_region_1."Code
INSEE"
group by
      "Code Région", "Région",
      "Code Département", "Département", date_arrete
order by
      "Région",

```

```

date_arrete desc;

CREATE OR REPLACE view borne_departement_global_shape as
SELECT
    borne_departement_global."Code Région",
    borne_departement_global."Région",
    borne_departement_global."Code Département",
    borne_departement_global."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_global
JOIN shape_departement ON borne_departement_global."Code Département" =
shape_departement."Code Département";

-- 2022-09-30
CREATE materialized view borne_departement_2022_09_30 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
    date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2022-09-30'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2022_09_30_shape as
SELECT
    borne_departement_2022_09_30."Code Région",
    borne_departement_2022_09_30."Région",
    borne_departement_2022_09_30."Code Département",
    borne_departement_2022_09_30."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2022_09_30
JOIN shape_departement ON borne_departement_2022_09_30."Code Département" =
shape_departement."Code Département";

-- 2022-06-30
CREATE materialized view borne_departement_2022_06_30 as

```

```

SELECT
    "Code Région", "Région",
    "Code Département", "Département",
    date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2022-06-30'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2022_06_30_shape as
SELECT
    borne_departement_2022_06_30."Code Région",
    borne_departement_2022_06_30."Région",
    borne_departement_2022_06_30."Code Département",
    borne_departement_2022_06_30."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2022_06_30
JOIN shape_departement ON borne_departement_2022_06_30."Code Département" =
shape_departement."Code Département";

-- 2022-03-31
CREATE materialized view borne_departement_2022_03_31 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
    date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2022-03-31'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2022_03_31_shape as

```

```

SELECT
    borne_departement_2022_03_31."Code Région",
    borne_departement_2022_03_31."Région",
    borne_departement_2022_03_31."Code Département",
    borne_departement_2022_03_31."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2022_03_31
JOIN shape_departement ON borne_departement_2022_03_31."Code Département" =
shape_departement."Code Département";

-- 2021-12-31
CREATE materialized view borne_departement_2021_12_31 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
    date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2021-12-31'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2021_12_31_shape as
SELECT
    borne_departement_2021_12_31."Code Région",
    borne_departement_2021_12_31."Région",
    borne_departement_2021_12_31."Code Département",
    borne_departement_2021_12_31."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2021_12_31
JOIN shape_departement ON borne_departement_2021_12_31."Code Département" =
shape_departement."Code Département";

-- 2021-09-30
CREATE materialized view borne_departement_2021_09_30 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
    date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as

```

```

nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2021-09-30'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2021_09_30_shape as
SELECT
    borne_departement_2021_09_30."Code Région",
    borne_departement_2021_09_30."Région",
    borne_departement_2021_09_30."Code Département",
    borne_departement_2021_09_30."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2021_09_30
JOIN shape_departement ON borne_departement_2021_09_30."Code Département" =
shape_departement."Code Département";

-- 2021-06-30
CREATE materialized view borne_departement_2021_06_30 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
    date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2021-06-30'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2021_06_30_shape as
SELECT
    borne_departement_2021_06_30."Code Région",
    borne_departement_2021_06_30."Région",

```

```

    borne_departement_2021_06_30."Code Département",
    borne_departement_2021_06_30."Département",
        date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2021_06_30
JOIN shape_departement ON borne_departement_2021_06_30."Code Département" =
shape_departement."Code Département";

-- 2021-03-31
CREATE materialized view borne_departement_2021_03_31 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
        date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON
voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code
INSEE"
WHERE date_arrete = '2021-03-31'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2021_03_31_shape as
SELECT
    borne_departement_2021_03_31."Code Région",
    borne_departement_2021_03_31."Région",
    borne_departement_2021_03_31."Code Département",
    borne_departement_2021_03_31."Département",
        date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2021_03_31
JOIN shape_departement ON borne_departement_2021_03_31."Code Département" =
shape_departement."Code Département";

-- 2020-12-31
CREATE materialized view borne_departement_2020_12_31 as
SELECT
    "Code Région", "Région",
    "Code Département", "Département",
        date_arrete, sum(nb_vp) as nb_voitures, sum(nb_vp_rechargeables_el) as
nb_voitures_elec, sum(nb_vp_rechargeables_gaz) as nb_voitures_gaz
FROM voitures_par_communes_par_energie
LEFT JOIN insee_nouvelle_region_1 ON

```

```

voitures_par_communes_par_energie.codegeo = insee_nouvelle_region_1."Code INSEE"
WHERE date_arrete = '2020-12-31'
group by
    "Code Région", "Région",
    "Code Département", "Département", date_arrete
order by
    "Région",
    date_arrete desc;

CREATE OR REPLACE view borne_departement_2020_12_31_shape as
SELECT
    borne_departement_2020_12_31."Code Région",
    borne_departement_2020_12_31."Région",
    borne_departement_2020_12_31."Code Département",
    borne_departement_2020_12_31."Département",
    date_arrete, nb_voitures, nb_voitures_elec, nb_voitures_gaz, wkt
FROM borne_departement_2020_12_31
JOIN shape_departement ON borne_departement_2020_12_31."Code Département" =
shape_departement."Code Département";

```

```

/* ----- BORNES AU 31 DÉCEMBRE 2020 ----- */
/* Detail de chaque bornes en 2020 */
SELECT * FROM bornes WHERE date_mise_en_service <= '2020-12-31';

/* Nombre de bornes par département et shape département */
/* Nombre de bornes par département */
CREATE OR REPLACE VIEW nb_bornes_par_departement_2020 AS
    SELECT code_geo."Code Département", code_geo."Département", count(*) AS NB_BORNES FROM bornes
        JOIN code_geo ON code_geo."Code INSEE"=bornes.code_insee_commune
        WHERE date_mise_en_service <= '2020-12-31'
        GROUP BY code_geo."Département", code_geo."Code Département";
/* Jointure pour obtenir la Shape département en 2020 */
SELECT "Département", NB_BORNES, wkt FROM nb_bornes_par_departement_2020
        JOIN shape_departement ON shape_departement."Code Département"=nb_bornes_par_departement_2020."Code Département";

/* Nombre de bornes par département normalisé en 2020 */
SELECT "Code Département",
    "Département",
    NB_BORNES,
    NTILE(4) OVER (ORDER BY NB_BORNES) AS Quartile
FROM nb_bornes_par_departement_2020;

/* Nombre de bornes par région et shape région */

```

```

/* Nombre de bornes par région */
CREATE OR REPLACE VIEW nb_bornes_par_region_2020 AS
    SELECT inr."Code Région", inr."Région", count(*) AS NB_BORNES FROM bornes
        JOIN code_geo ON code_geo."Code INSEE"=bornes.code_insee_commune
        JOIN insee_nouvelle_region inr ON code_geo."Code INSEE" = inr."Code INSEE"
        WHERE date_mise_en_service <= '2020-12-31'
        GROUP BY inr."Région", inr."Code Région";
/* Jointure pour obtenir la Shape région en 2020 */
SELECT NBPR."Région", NB_BORNES, wkt FROM nb_bornes_par_region_2020 NBPR
    JOIN shape_region ON shape_region."Code_region"=NBPR."Code Région";

/* Nombre de bornes par région normalisé en 2020 */
SELECT "Code Région",
    "Région",
    NB_BORNES,
    NTILE(4) OVER (ORDER BY NB_BORNES) AS Quartile
FROM nb_bornes_par_region_2020;

/* ----- BORNES AU 31 DÉCEMBRE 2021 ----- */
/* Detail de chaque bornes en 2021 */
SELECT * FROM bornes WHERE date_mise_en_service <= '2021-12-31';

/* Nombre de bornes par département et shape département */
/* Nombre de bornes par département */
CREATE OR REPLACE VIEW nb_bornes_par_departement_2021 AS
    SELECT code_geo."Code Departement", code_geo."Departement", count(*) AS
NB_BORNES FROM bornes
        JOIN code_geo ON code_geo."Code INSEE"=bornes.code_insee_commune
        WHERE date_mise_en_service <= '2021-12-31'
        GROUP BY code_geo."Departement", code_geo."Code Departement";
/* Jointure pour obtenir la Shape département en 2021 */
SELECT "Departement", NB_BORNES, wkt FROM nb_bornes_par_departement_2021
    JOIN shape_departement ON shape_departement."Code
Département"=nb_bornes_par_departement_2021."Code Departement";

/* Nombre de bornes par département normalisé en 2021 */
SELECT "Code Departement",
    "Departement",
    NB_BORNES,
    NTILE(4) OVER (ORDER BY NB_BORNES) AS Quartile
FROM nb_bornes_par_departement_2021;

/* Nombre de bornes par région et shape région */

```

```

/* Nombre de bornes par région */
CREATE OR REPLACE VIEW nb_bornes_par_region_2021 AS
    SELECT inr."Code Région", inr."Région", count(*) AS NB_BORNES FROM bornes
        JOIN code_geo ON code_geo."Code INSEE"=bornes.code_insee_commune
        JOIN insee_nouvelle_region inr ON code_geo."Code INSEE" = inr."Code INSEE"
        WHERE date_mise_en_service <= '2021-12-31'
        GROUP BY inr."Région", inr."Code Région";
/* Jointure pour obtenir la Shape région en 2021 */
SELECT NBPR."Région", NB_BORNES, wkt FROM nb_bornes_par_region_2021 NBPR
    JOIN shape_region ON shape_region."Code_région"=NBPR."Code Région";

/* Nombre de bornes par région normalisé en 2021 */
SELECT "Code Région",
    "Région",
    NB_BORNES,
    NTILE(4) OVER (ORDER BY NB_BORNES) AS Quartile
FROM nb_bornes_par_region_2021;

/* ----- BORNES AU 31 DÉCEMBRE 2022 ----- */
/* Detail de chaque bornes en 2022 */
SELECT * FROM bornes;

/* Nombre de bornes par département et shape département */
/* Nombre de bornes par département */
CREATE OR REPLACE VIEW nb_bornes_par_departement_2022 AS
    SELECT code_geo."Code Département", code_geo."Département", count(*) AS
NB_BORNES FROM bornes
        JOIN code_geo ON code_geo."Code INSEE"=bornes.code_insee_commune
        GROUP BY code_geo."Département", code_geo."Code Département";
/* Jointure pour obtenir la Shape département en 2022 */
SELECT "Département", NB_BORNES, wkt FROM nb_bornes_par_departement_2022
    JOIN shape_departement ON shape_departement."Code
Département"=nb_bornes_par_departement_2022."Code Département";

/* Nombre de bornes par département normalisé en 2022 */
SELECT "Code Département",
    "Département",
    NB_BORNES,
    NTILE(4) OVER (ORDER BY NB_BORNES) AS Quartile
FROM nb_bornes_par_departement_2022;

/* Nombre de bornes par région et shape région */
/* Nombre de bornes par région */
CREATE OR REPLACE VIEW nb_bornes_par_region_2022 AS

```

```

SELECT inr."Code Région", inr."Région", count(*) AS NB_BORNES FROM bornes
JOIN code_geo ON code_geo."Code INSEE"=bornes.code_insee_commune
JOIN insee_nouvelle_region inr on code_geo."Code INSEE" = inr."Code INSEE"
GROUP BY inr."Région", inr."Code Région";
/* Jointure pour obtenir la Shape région en 2022 */
SELECT NBPR."Région", NB_BORNES, wkt FROM nb_bornes_par_region_2022 NBPR
    JOIN shape_region ON shape_region."Code_region"=NBPR."Code Région";

/* Nombre de bornes par region normalisé en 2022 */
SELECT "Code Région",
    "Région",
    NB_BORNES,
    NTILE(4) OVER (ORDER BY NB_BORNES) AS Quartile
FROM nb_bornes_par_region_2022;

```

```

/* ----- CHOMAGE 2020 ----- */
/* Taux de chomage par département */
CREATE OR REPLACE VIEW taux_chomage_dep_2020 AS
    SELECT "Code",
        "Libellé" ,
        CAST(replace(avg_2020, ',', '.') AS DOUBLE PRECISION) AS moy_2020
    FROM tx_chomage;

/* Taux de chomage par département normalisé en 2020 */
SELECT "Code",
    "Libellé",
    moy_2020,
    NTILE(4) OVER (ORDER BY moy_2020) AS Quartile
FROM taux_chomage_dep_2020;

/* Taux de chomage par région */
CREATE OR REPLACE VIEW tx_chomage_reg_2020 AS
    SELECT inr."Code Région",
        inr."Région", avg(CAST(replace(avg_2020, ',', '.') AS DOUBLE PRECISION))
AS moy_2020
    FROM tx_chomage
    JOIN insee_nouvelle_region inr on tx_chomage."Code" = inr."Code Département"
    GROUP BY inr."Région",
        inr."Code Région";

/* Taux de chomage par region normalisé en 2020 */
SELECT "Code Région",
    "Région",
    moy_2020,
    NTILE(4) OVER (ORDER BY moy_2020) AS Quartile

```

```

FROM tx_chomage_reg_2020;

/* ----- CHOMAGE 2021 ----- */
/* Taux de chomage par département */
CREATE OR REPLACE VIEW taux_chomage_dep_2021 AS
    SELECT "Code",
           "Libellé",
           CAST(replace(avg_2021, ',', '.') AS DOUBLE PRECISION) AS moy_2021
    FROM tx_chomage;

/* Taux de chomage par département normalisé en 2021 */
SELECT "Code",
       "Libellé",
       moy_2021,
       NTILE(4) OVER (ORDER BY moy_2021) AS Quartile
FROM taux_chomage_dep_2021;

/* Taux de chomage par région */
CREATE OR REPLACE VIEW tx_chomage_reg_2021 AS
    SELECT inr."Code Région",
           inr."Région", avg(CAST(replace(avg_2021, ',', '.') AS DOUBLE PRECISION))
AS moy_2021
    FROM tx_chomage
    JOIN insee_nouvelle_region inr ON tx_chomage."Code" = inr."Code Département"
    GROUP BY inr."Région",
             inr."Code Région";

/* Taux de chomage par region normalisé en 2021 */
SELECT "Code Région",
       "Région",
       moy_2021,
       NTILE(4) OVER (ORDER BY moy_2021) AS Quartile
FROM tx_chomage_reg_2021;

/* ----- CHOMAGE 2022 ----- */
/* Taux de chomage par département */
CREATE OR REPLACE VIEW taux_chomage_dep_2022 AS
    SELECT "Code",
           "Libellé",
           CAST(replace(avg_2021, ',', '.') AS DOUBLE PRECISION) AS moy_2022
    FROM tx_chomage;

```

```

/* Taux de chomage par département normalisé en 2022 */
SELECT "Code",
       "Libellé",
       moy_2022,
       NTILE(4) OVER (ORDER BY moy_2022) AS Quartile
FROM taux_chomage_dep_2022;

/* Taux de chomage par région */
CREATE OR REPLACE VIEW tx_chomage_reg_2022 AS
SELECT inr."Code Région",
       inr."Région", avg(CAST(replace(avg_2022, ',', '.') AS DOUBLE PRECISION))
AS moy_2022
FROM tx_chomage
JOIN insee_nouvelle_region inr ON tx_chomage."Code" = inr."Code Département"
GROUP BY inr."Région",
         inr."Code Région";

/* Taux de chomage par region normalisé en 2022 */
SELECT "Code Région",
       "Région",
       moy_2022,
       NTILE(4) OVER (ORDER BY moy_2022) AS Quartile
FROM tx_chomage_reg_2022;

```

```

-- FONCTION POUR NORMALISER LES DONNEES SUR LES POPULATIONS
CREATE OR REPLACE FUNCTION classify_department(total BIGINT,
                                                nb_0_19 BIGINT,
                                                nb_20_39 BIGINT,
                                                nb_40_59 BIGINT,
                                                nb_60_74 BIGINT,
                                                nb_75_plus BIGINT)
RETURNS INT AS $$

DECLARE
    age_group_0_19_percent NUMERIC;
    age_group_20_39_percent NUMERIC;
    age_group_40_59_percent NUMERIC;
    age_group_60_74_percent NUMERIC;
    age_group_75_plus_percent NUMERIC;
    max_percent NUMERIC;

BEGIN
    -- Calculer les pourcentages de la population pour chaque tranche d'âge
    age_group_0_19_percent := (nb_0_19 / CAST(total AS NUMERIC)) * 100;
    age_group_20_39_percent := (nb_20_39 / CAST(total AS NUMERIC)) * 100;
    age_group_40_59_percent := (nb_40_59 / CAST(total AS NUMERIC)) * 100;
    age_group_60_74_percent := (nb_60_74 / CAST(total AS NUMERIC)) * 100;

```

```

age_group_75_plus_percent := (nb_75_plus / CAST(total AS NUMERIC)) * 100;

-- Trouver la tranche d'âge la plus représentée
max_percent := GREATEST(age_group_0_19_percent, age_group_20_39_percent,
age_group_40_59_percent, age_group_60_74_percent, age_group_75_plus_percent);

-- Classer le département en fonction de la tranche d'âge la plus
représentée
IF max_percent = age_group_0_19_percent THEN
    RETURN 1;
ELSIF max_percent = age_group_20_39_percent THEN
    RETURN 2;
ELSIF max_percent = age_group_40_59_percent THEN
    RETURN 3;
ELSIF max_percent = age_group_60_74_percent THEN
    RETURN 4;
ELSE
    RETURN 5;
END IF;
END;

$$ LANGUAGE plpgsql;

/*
----- POP 2017 -----
/* POP par département normalisé en 2017 */
SELECT code_dep,
       nom_dep,
       "Total",
       classify_department(
           "Total",
           "0 à 19 ans",
           "20 à 39 ans",
           "40 à 59 ans",
           "60 à 74 ans",
           "75 ans et plus"
       ) AS age_group,
       NTILE(4) OVER (ORDER BY "Total") AS "pop_group"
FROM "pop-2017";

/*
POP par région normalisé en 2017 */
SELECT "Code Région",
       "Région",
       SUM("Total") AS "Total",
       classify_department(
           SUM("Total"),
           SUM("0 à 19 ans"),

```

```

        SUM("20 à 39 ans"),
        SUM("40 à 59 ans"),
        SUM("60 à 74 ans"),
        SUM("75 ans et plus")
    ) AS age_group,
    NTILE(4) OVER (ORDER BY SUM("Total")) AS "pop_group"
FROM "pop-2017"
JOIN insee_nouvelle_region inr ON inr."Code Département" = "pop-2017".code_dep
GROUP BY "Code Région", "Région";

/* ----- POP 2018 ----- */
/* POP par département normalisé en 2018 */
SELECT code_dep,
    nom_dep,
    "Total",
    classify_department(
        "Total",
        "0 à 19 ans",
        "20 à 39 ans",
        "40 à 59 ans",
        "60 à 74 ans",
        "75 ans et plus"
    ) AS age_group,
    NTILE(4) OVER (ORDER BY "Total") AS "pop_group"
FROM "pop-2018";

/* POP par région normalisé en 2018 */
SELECT "Code Région",
    "Région",
    SUM("Total") AS "Total",
    classify_department(
        SUM("Total"),
        SUM("0 à 19 ans"),
        SUM("20 à 39 ans"),
        SUM("40 à 59 ans"),
        SUM("60 à 74 ans"),
        SUM("75 ans et plus")
    ) AS age_group,
    NTILE(4) OVER (ORDER BY SUM("Total")) AS "pop_group"
FROM "pop-2018"
JOIN insee_nouvelle_region inr ON inr."Code Département" = "pop-2018".code_dep
GROUP BY "Code Région", "Région";

```

```

/* ----- POP 2019 ----- */
/* POP par département normalisé en 2019 */
SELECT code_dep,
       nom_dep,
       "Total",
       classify_department(
           "Total",
           "0 à 19 ans",
           "20 à 39 ans",
           "40 à 59 ans",
           "60 à 74 ans",
           "75 ans et plus"
       ) AS age_group,
       NTILE(4) OVER (ORDER BY "Total") AS "pop_group"
FROM "pop-2019";

/* POP par région normalisé en 2019 */
SELECT "Code Région",
       "Région",
       SUM("Total") AS "Total",
       classify_department(
           SUM("Total"),
           SUM("0 à 19 ans"),
           SUM("20 à 39 ans"),
           SUM("40 à 59 ans"),
           SUM("60 à 74 ans"),
           SUM("75 ans et plus")
       ) AS age_group,
       NTILE(4) OVER (ORDER BY SUM("Total")) AS "pop_group"
FROM "pop-2019"
JOIN insee_nouvelle_region inr ON inr."Code Département" = "pop-2019".code_dep
GROUP BY "Code Région", "Région";

```

```

/* ----- RICHESSE 2017 ----- */
/* Richesse par département en 2017 */
SELECT "Code Département",
       "Département",
       med17 as "Mediane revenus mensuels en 2017",
       NTILE(4) OVER (ORDER BY med17) AS Quartile
FROM richesse_2017
JOIN insee_nouvelle_region inr ON richesse_2017.codgeo = inr."Code Département"
GROUP BY "Code Département",

```

```

    "Département",
    "Mediane revenus mensuels en 2017";

/* Richesse par région en 2017 */
SELECT "Code Région",
    "Région",
    avg(med17) as "Mediane revenus mensuels en 2017",
    NTILE(4) OVER (ORDER BY avg(med17)) AS Quartile
FROM richesse_2017
JOIN insee_nouvelle_region inr ON richesse_2017.codgeo = inr."Code Département"
GROUP BY "Code Région",
    "Région";

/* ----- RICHESSE 2018 ----- */
/* Richesse par département en 2018 */
SELECT "Code Département",
    "Département",
    med18 as "Mediane revenus mensuels en 2018",
    NTILE(4) OVER (ORDER BY med18) AS Quartile
FROM richesse_2018
JOIN insee_nouvelle_region inr ON richesse_2018.codgeo = inr."Code Département"
GROUP BY "Code Département",
    "Département",
    "Mediane revenus mensuels en 2018";

/* Richesse par région en 2018 */
SELECT "Code Région",
    "Région",
    avg(med18) as "Mediane revenus mensuels en 2018",
    NTILE(4) OVER (ORDER BY avg(med18)) AS Quartile
FROM richesse_2018
JOIN insee_nouvelle_region inr ON richesse_2018.codgeo = inr."Code Département"
GROUP BY "Code Région",
    "Région";

/* ----- RICHESSE 2019 ----- */
/* Richesse par département en 2019 */
SELECT "Code Département",
    "Département",
    med19 as "Mediane revenus mensuels en 2019",
    NTILE(4) OVER (ORDER BY med19) AS Quartile

```

```

FROM richesse_2019
JOIN insee_nouvelle_region inr ON richesse_2019.codgeo = inr."Code Département"
GROUP BY "Code Département",
        "Département",
        "Mediane revenus mensuels en 2019";

/* Richesse par région en 2019 */
SELECT "Code Région",
       "Région",
       avg(med19) AS "Mediane revenus mensuels en 2019",
       NTILE(4) OVER (ORDER BY avg(med19)) AS Quartile
FROM richesse_2019
JOIN insee_nouvelle_region inr ON richesse_2019.codgeo = inr."Code Département"
GROUP BY "Code Région",
        "Région";

```

```

/* ----- TMJA 2017 ----- */
/* TMJA par département */
CREATE OR REPLACE VIEW tmja_dep_2017 AS
SELECT "depPrD" AS "Code département",
       inr."Département",
       avg(tmja) AS tmja
FROM "tmja-2017"
JOIN insee_nouvelle_region inr ON CAST("depPrD" AS TEXT) = inr."Code
Département"
WHERE "tmja-2017".tmja IS NOT NULL
GROUP BY "Département", "Code département";

/* TMJA par département normalisé en 2017 */
SELECT "Code département",
       "Département",
       tmja,
       NTILE(4) OVER (ORDER BY tmja) AS Quartile
FROM tmja_dep_2017;

/* TMJA par région */
CREATE OR REPLACE VIEW tmja_reg_2017 AS
SELECT inr."Code Région",
       inr."Région",
       avg(tmja) AS tmja
FROM tmja_dep_2017
JOIN insee_nouvelle_region inr ON inr."Code Département" =
CAST(tmja_dep_2017."Code département" AS TEXT)
GROUP BY inr."Région",
        inr."Code Région";

```

```

/* TMJA par région normalisé en 2017 */
SELECT "Code Région",
       "Région",
       tmja,
       NTILE(4) OVER (ORDER BY tmja) AS Quartile
FROM tmja_reg_2017;

/* ----- TMJA 2018 ----- */
/* TMJA par département */
CREATE OR REPLACE VIEW tmja_dep_2018 AS
  SELECT "depPrD" AS "Code département",
         inr."Département",
         avg(tmja) AS tmja
    FROM "tmja-2018"
   JOIN insee_nouvelle_region inr ON CAST("depPrD" AS TEXT) = inr."Code
Département"
   WHERE "tmja-2018".tmja IS NOT NULL
  GROUP BY "Département", "Code département";

/* TMJA par département normalisé en 2018 */
SELECT "Code département",
       "Département",
       tmja,
       NTILE(4) OVER (ORDER BY tmja) AS Quartile
FROM tmja_dep_2018;

/* TMJA par région */
CREATE OR REPLACE VIEW tmja_reg_2018 AS
  SELECT inr."Code Région",
         inr."Région",
         avg(tmja) AS tmja
    FROM tmja_dep_2018
   JOIN insee_nouvelle_region inr ON inr."Code Département" =
CAST(tmja_dep_2018."Code département" AS TEXT)
  GROUP BY inr."Région",
          inr."Code Région";

/* TMJA par region normalisé en 2018 */
SELECT "Code Région",
       "Région",
       tmja,
       NTILE(4) OVER (ORDER BY tmja) AS Quartile

```

```

FROM tmja_reg_2018;

/*
----- TMJA 2019 -----
/* TMJA par département */
CREATE OR REPLACE VIEW tmja_dep_2019 AS
SELECT "depPrD" AS "Code département",
       inr."Département",
       avg(tmja) AS tmja
  FROM "tmja-2019"
    JOIN insee_nouvelle_region inr ON CAST("depPrD" AS TEXT) = inr."Code
Département"
 WHERE "tmja-2019".tmja IS NOT NULL
 GROUP BY "Département", "Code département";

/* TMJA par département normalisé en 2019 */
SELECT "Code département",
       "Département",
       tmja,
       NTILE(4) OVER (ORDER BY tmja) AS Quartile
  FROM tmja_dep_2019;

/* TMJA par région */
CREATE OR REPLACE VIEW tmja_reg_2019 AS
SELECT inr."Code Région",
       inr."Région",
       avg(tmja) AS tmja
  FROM tmja_dep_2019
    JOIN insee_nouvelle_region inr ON inr."Code Département" =
CAST(tmja_dep_2019."Code département" AS TEXT)
 GROUP BY inr."Région",
          inr."Code Région";

/* TMJA par region normalisé en 2019 */
SELECT "Code Région",
       "Région",
       tmja,
       NTILE(4) OVER (ORDER BY tmja) AS Quartile
  FROM tmja_reg_2019;

```