# Résumé

# Tables des matières

[Résumé 1](#_Toc180070113)

[Tables des matières 2](#_Toc180070114)

[Listes des tableaux 4](#_Toc180070115)

[Listes des figures 5](#_Toc180070116)

[1 Introduction 6](#_Toc180070117)

[1.1 Contexte général du marché du Gazole Non Routier en France 6](#_Toc180070118)

[1.2 Énoncé de la problématique et objectifs de recherche 6](#_Toc180070119)

[1.3 Aperçu de l’approche méthodologique et de la structure du rapport 7](#_Toc180070120)

[2 Revue de littérature 9](#_Toc180070121)

[2.1 Le marché du GNR en France : contexte et facteurs d’influence 9](#_Toc180070122)

[2.1.1 Évolution historique et réglementaire du GNR depuis 2011 9](#_Toc180070123)

[2.1.2 Importance économique et secteurs clés 10](#_Toc180070124)

[2.1.2.1 Part du GNR dans la consommation énergétique française 10](#_Toc180070125)

[2.1.2.2 Analyse sectorielle de l’utilisation du GNR 10](#_Toc180070126)

[2.1.3 Conclusion sur les facteurs économiques et les secteurs clés du GNR 11](#_Toc180070127)

[2.2 Facteurs influençant les prix du GNR 12](#_Toc180070128)

[2.2.1 Facteurs macroéconomiques 12](#_Toc180070129)

[2.2.2 Facteurs géographiques et logistiques 13](#_Toc180070130)

[2.2.3 Facteurs démographiques et économiques régionaux 14](#_Toc180070131)

[2.2.4 Facteurs saisonniers et météorologiques 15](#_Toc180070132)

[2.2.5 Conclusion 16](#_Toc180070133)

[2.3 Méthodes d'analyse des disparités régionales des prix énergétiques 17](#_Toc180070134)

[2.3.1 Analyse statistique descriptive 17](#_Toc180070135)

[2.3.2 Analyse spatiale 18](#_Toc180070136)

[2.3.3 Analyse temporelle 18](#_Toc180070137)

[2.4 Collecte et structuration des données : méthodologies, considérations et outils 18](#_Toc180070138)

[2.4.1 Approche méthodologique de la collecte de données 18](#_Toc180070139)

[2.4.2 Techniquess et architecture de web scraping 21](#_Toc180070140)

[2.4.3 Considérations éthiques et légales du web scraping 25](#_Toc180070141)

[2.4.4 Considérations spécifiques à la recherche sur les prix des carburants 26](#_Toc180070142)

[3 Collecte et structuration des données de l’étude 28](#_Toc180070143)

[3.3 Sources des variables de l’étude 28](#_Toc180070144)

[3.3.1 Sources des données de prix du GNR 28](#_Toc180070145)

[3.3.2 Sources des facteurs macroéconomiques 28](#_Toc180070146)

[3.3.3 Sources des données géographiques et démographiques 28](#_Toc180070147)

[3.4 Couverture géographique des données collectées 28](#_Toc180070148)

[3.4.1 Présentation de la carte des codes postaux scrapés 28](#_Toc180070149)

[3.4.2 Analyse de la répartition géographique des données 28](#_Toc180070150)

[3.4.3 Discussion sur la représentativité de l'échantillon 28](#_Toc180070151)

[3.5 Fréquence et période de collecte des données 28](#_Toc180070152)

[3.6 Structuration de la base de données 28](#_Toc180070153)

[3.7 Organisation des données collectées 28](#_Toc180070154)

[3.8 Description finale de la base de données 28](#_Toc180070155)

[4 Méthodologie d’analyse 29](#_Toc180070156)

[5 Analyse des disparités régionales des prix du GNR 30](#_Toc180070157)

[5.1 Cartographie des prix à l'échelle nationale 30](#_Toc180070158)

[5.2 Identification des facteurs influents par région 30](#_Toc180070159)

[5.3 Patterns géographiques et clusters de prix 30](#_Toc180070160)

[5.4 Cas d'étude : focus sur des régions spécifiques 30](#_Toc180070161)

[Annexes 32](#_Toc180070162)

[Références bibliographiques 33](#_Toc180070163)

# Listes des tableaux

# Listes des figures

# 1 Introduction

## 1.1 Contexte général du marché du Gazole Non Routier en France

Le Gazole Non Routier (GNR) occupe une place centrale dans le paysage énergétique français, jouant un rôle crucial dans des secteurs économiques clés tels que l'agriculture, la sylviculture et les travaux publics. Introduit en 2011 dans le cadre des efforts européens visant à réduire les émissions polluantes, le GNR est rapidement devenu un élément indispensable pour de nombreuses activités économiques (TotalEnergies, 2024).

La dynamique du marché du GNR en France est caractérisée par une complexité croissante, influencée par une multitude de facteurs allant des fluctuations du marché pétrolier mondial aux politiques énergétiques nationales et européennes. L'annulation récente, en janvier 2024, de la hausse progressive de la taxe sur le GNR prévue jusqu'en 2030 illustre la sensibilité de ce marché aux décisions politiques et aux pressions économiques (Ministère de la Transition Écologique, 2022).

Les disparités régionales dans les prix du GNR soulèvent des questions importantes sur l'équité d'accès à cette ressource énergétique et sur ses implications pour la compétitivité des entreprises dans différentes régions de France. Ces variations de prix, parfois significatives, peuvent avoir des répercussions importantes sur les coûts opérationnels des entreprises et, par extension, sur l'économie locale et nationale.

## 1.2 Énoncé de la problématique et objectifs de recherche

Dans ce contexte, notre étude vise à répondre à la question centrale suivante : "Quels sont les facteurs déterminants des disparités régionales des prix du GNR en France, et comment peut-on prédire ces variations de prix de manière fiable ?"

Pour aborder cette problématique, nous nous fixons les objectifs spécifiques suivants :

1. Cartographier et analyser en profondeur les disparités régionales des prix du GNR en France.
2. Identifier et hiérarchiser les facteurs influençant ces variations de prix, qu'ils soient économiques, géographiques, ou politiques.
3. Développer des modèles prédictifs robustes pour anticiper les évolutions futures des prix du GNR à l'échelle régionale.
4. Évaluer l'impact potentiel de ces disparités de prix sur la compétitivité des entreprises et l'économie régionale.
5. Formuler des recommandations basées sur nos résultats pour les acteurs du marché et les décideurs politiques.

## 1.3 Aperçu de l’approche méthodologique et de la structure du rapport

Pour atteindre ces objectifs, notre étude s'appuie sur une approche multidimensionnelle combinant :

* Une analyse approfondie du contexte historique et réglementaire du marché du GNR en France.
* Une collecte de données utilisant des techniques de web scraping pour obtenir des informations détaillées sur les prix du GNR dans plus de 2000 codes postaux.
* L'utilisation de techniques avancées d'analyse spatiale et temporelle pour identifier les patterns et les tendances dans les variations de prix.
* L'application de modèles statistiques et de machine learning pour analyser les facteurs d'influence et développer des capacités prédictives.

Le rapport est structuré comme suit :

* La section 2 présente une analyse détaillée du marché du GNR en France, incluant son évolution historique et les facteurs influençant ses prix.
* La section 3 décrit notre méthodologie de collecte et de structuration des données, mettant en lumière notre approche innovante de web scraping.
* La section 4 expose les méthodologies d'analyse employées dans notre étude.
* Les sections 5 et 6 présentent respectivement nos analyses des disparités régionales et de la dynamique temporelle des prix du GNR.
* La section 7 discute des implications de nos résultats et propose des recommandations.
* Enfin, la conclusion synthétise nos principales découvertes et ouvre des perspectives pour de futures recherches.

À travers cette étude, nous visons à apporter un éclairage nouveau sur les dynamiques complexes du marché du GNR en France, offrant ainsi des insights précieux pour les acteurs économiques et les décideurs politiques dans ce domaine crucial de l'économie française.

# 2 Revue de littérature

## 2.1 Le marché du GNR en France : contexte et facteurs d’influence

### 2.1.1 Évolution historique et réglementaire du GNR depuis 2011

L'introduction du Gazole Non Routier (GNR) en France s'inscrit dans le cadre plus large de la politique énergétique européenne. La directive 2009/30/CE du Parlement européen et du Conseil du 23 avril 2009 a posé les bases de cette évolution, visant à réduire l'impact environnemental des carburants utilisés dans les secteurs non routiers (Commission européenne, 2009). Cette directive a marqué un tournant dans la réglementation des carburants en Europe, en imposant des normes plus strictes sur la qualité des carburants et en encourageant l'utilisation de carburants à faible teneur en soufre.

En France, l'implémentation de cette directive s'est traduite par l'introduction du GNR le 1er mai 2011, remplaçant le fioul domestique dans les engins mobiles non routiers, les tracteurs agricoles et forestiers. Cette transition a été encadrée par l'arrêté du 10 décembre 2010, qui a défini les spécifications techniques du GNR et les conditions de son utilisation (Ministère de la Transition écologique, 2010). Les premières années ont été caractérisées par une phase d'apprentissage et d'ajustement, au cours de laquelle les acteurs du marché ont dû s'adapter à ce nouveau carburant et à ses spécificités.

Entre 2014 et 2017, le marché du GNR a connu une période de stabilisation et de croissance régulière. Cette phase a vu une amélioration des réseaux de distribution et une meilleure compréhension des avantages du GNR par les utilisateurs finaux. Cette relative stabilité a cependant été mise à l'épreuve en 2018, lorsque la hausse des cours du pétrole a fortement impacté les prix du GNR, suivie d'une chute marquée qui a souligné la vulnérabilité du marché face aux fluctuations internationales (INSEE, 2018).

La pandémie de COVID-19 en 2020 a introduit un nouveau niveau de complexité, entraînant une volatilité accrue des prix et des perturbations dans les chaînes d'approvisionnement. Cette période a mis en lumière la nécessité d'une plus grande résilience dans le secteur énergétique (Terre-net, 2020).

Les années 2021 à 2023 ont été marquées par un renforcement des normes environnementales et des débats intenses sur l'avenir fiscal du GNR (Assemblée nationale, 2021). Ces discussions reflétaient la tension croissante entre les objectifs de transition écologique et les préoccupations économiques des secteurs dépendant du GNR.

L'année 2024 a débuté avec une décision politique majeure : l'annulation de la hausse progressive de la taxe sur le GNR prévue jusqu'en 2030. Cette décision illustre la complexité des arbitrages entre considérations environnementales, économiques et sociales dans la gestion de ce marché (Inspection générale des finances, 2024).

## 2.1.2 Importance économique et secteurs clés

### 2.1.2.1 Part du GNR dans la consommation énergétique française

Le Gazole Non Routier (GNR) est un élément important du mix énergétique en France, surtout dans des secteurs tels que l’agriculture, les travaux publics et la sylviculture. Selon le Bilan énergétique de la France publié par le Service de la Donnée et des Études Statistiques (SDED) du Ministère de la transition Écologique, le pétrole représente environ 39 % de la consommation finale d'énergie en 2022, dont une part significative est allouée aux carburants pour les secteurs industriels et agricoles (Service de la Donnée et des Études Statistiques, 2023). Cela inclut le GNR, essentiel pour les machines lourdes et les engins non routiers dans ces secteurs.

### 2.1.2.2 Analyse sectorielle de l’utilisation du GNR

Le Gazole Non Routier (GNR) joue un rôle crucial dans plusieurs secteurs clés de l'économie française. Cette section analyse l'utilisation de celui-ci dans les principaux secteurs consommateurs.

#### Secteur agricole :

D'après Audrey Trévisiol, Coordinatrice Agriculture et Changement Climatique à l'Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME), le secteur agricole est le premier consommateur de GNR en France (Trévisiol, 2023). Cette affirmation est confirmée par le Ministère de l’Agriculture et de l'Alimentation, qui a indiqué qu’en 2021, le secteur agricole a utilisé environ 29,4 TWh de Gazole Non Routier, soit 60 % de sa consommation énergétique totale. Ces données soulignent l'importance du GNR pour soutenir les activités agricoles, en en faisant une ressource énergétique clé pour ce secteur​ (SSP, 2024).

#### Secteur des travaux publics et de la construction :

En 2015, le secteur de la construction en France dépend fortement des produits pétroliers, qui représentent 89 % de sa consommation énergétique totale. Le Gazole Non Routier (GNR) en constitue 22 %, étant largement utilisé dans les travaux publics pour les engins lourds et les grands chantiers. Le gazole routier, dominant à 61 %, est plus courant dans le bâtiment pour les petits chantiers et le transport. Dans les travaux publics, 97 % du GNR est dédié aux activités de chantier, tandis que le transport et les locaux administratifs occupent une part secondaire, illustrant ainsi la dépendance majeure de ce secteur aux énergies fossiles (Commissariat général au développement durable, 2018).

#### Secteur forestier :

Selon l'Office National des Forêts (ONF), le secteur forestier, bien que moins consommateur que l'agriculture ou la construction, dépend fortement du GNR pour ses opérations ( Office National des Forêts, 2021). En 2021, la consommation de GNR dans ce secteur était estimée à environ 100 000 mètres cubes par an, représentant une part significative des coûts opérationnels. Une étude menée par l'Institut technologique FCBA (Forêt Cellulose Bois-construction Ameublement) en 2021 a révélé que le GNR constitue en moyenne 12% des coûts totaux d'exploitation forestière, pouvant atteindre 18% pour l'abattage mécanisé et 15% pour le débardage. Ces chiffres soulignent l'importance du GNR dans les activités forestières, particulièrement dans les zones difficiles d'accès où l'électrification des engins n'est pas encore viable (Institut technologique Forêt Cellulose Bois-construction, 2021).

### 2.1.3 Conclusion sur les facteurs économiques et les secteurs clés du GNR

L'analyse de l'importance économique du Gazole Non Routier (GNR) et de son utilisation dans les secteurs clés de l'économie française révèle son rôle crucial dans le paysage énergétique national. Le GNR s'impose comme une ressource énergétique essentielle, particulièrement dans les domaines de l'agriculture, des travaux publics, de la construction et de la sylviculture.

La part significative du GNR dans la consommation énergétique française, notamment dans les secteurs industriels et agricoles, souligne sa place prépondérante dans le mix énergétique national. Cette importance est d'autant plus marquée lorsqu'on examine son utilisation sectorielle.

Le secteur agricole se distingue comme le principal consommateur de GNR, représentant 60% de sa consommation énergétique totale. Cette dépendance illustre le rôle vital du GNR dans le maintien et le développement des activités agricoles françaises.

Dans le domaine des travaux publics et de la construction, le GNR occupe également une place centrale, constituant 22% de la consommation énergétique du secteur. Son utilisation prédominante dans les grands chantiers et pour les engins lourds témoigne de son importance dans la réalisation des infrastructures et des projets de construction à grande échelle.

Bien que moins consommateur en volume, le secteur forestier dépend fortement du GNR pour ses opérations, particulièrement dans les zones difficiles d'accès. La part significative du GNR dans les coûts d'exploitation forestière souligne son impact économique dans ce secteur.

Cette analyse sectorielle met en lumière la dépendance de secteurs clés de l'économie française au GNR, soulignant ainsi l'importance stratégique de ce carburant. Elle révèle également les défis potentiels liés à cette dépendance, notamment en termes de transition énergétique et d'impact environnemental.

La compréhension de ces dynamiques sectorielles est cruciale pour appréhender les enjeux liés aux prix du GNR et leurs répercussions sur l'économie française. Elle fournit un contexte essentiel pour l'analyse des disparités régionales des prix du GNR et souligne l'importance d'une approche nuancée, prenant en compte les spécificités de chaque secteur dans l'élaboration de politiques énergétiques et économiques.

## 2.2 Facteurs influençant les prix du GNR

La formation des prix du GNR en France est le résultat d'une interaction complexe entre divers facteurs, allant des dynamiques macroéconomiques globales aux spécificités locales. Comprendre ces facteurs est essentiel pour appréhender les disparités régionales et anticiper les évolutions futures des prix.

### 2.2.1 Facteurs macroéconomiques

Les facteurs macroéconomiques, qui opèrent à l'échelle mondiale et nationale, jouent un rôle crucial dans la détermination des coûts de production, d'importation et de distribution du GNR. Parmi ces facteurs, deux se distinguent par leur impact significatif et leur nature dynamique sur le marché international : le cours du baril de pétrole et le taux de change USD/EUR.

#### 2.2.1.1 Cours du baril de pétrole

Le cours du baril de pétrole est un des facteurs clés, à la fois visible et volatil, qui influence les prix du Gazole Non Routier (GNR) en France. En tant que composant principal des produits raffinés, une variation du prix du baril de pétrole brut affecte directement les coûts de production du GNR, en plus des coûts de transport et de distribution qui y sont associés.

Les fluctuations du marché pétrolier mondial, qu'elles soient dues à des événements géopolitiques, des décisions de l'OPEP, ou des changements dans l'offre et la demande globales, se répercutent rapidement sur les prix du GNR en France (Fernandois & Medel, 2020). Cette sensibilité aux cours mondiaux du pétrole souligne l'interconnexion du marché français du GNR avec les dynamiques énergétiques globales.

#### 2.2.1.2 Taux de Change USD/EUR

Le taux de change USD/EUR exerce une influence significative sur les prix du GNR. La majeure partie du pétrole brut étant négociée en dollars américains, les variations du taux de change impactent directement les coûts d'importation du pétrole en France. Un euro fort tend à réduire ces coûts, tandis qu'un euro faible les augmente, affectant ainsi les prix finaux du GNR pour les consommateurs français (Beckmann & Czudaj, 2013).

### 2.2.2 Facteurs géographiques et logistiques

Les prix du GNR sont influencés par des facteurs géographiques et logistiques, en particulier la localisation des dépôts de carburant et la qualité de l'infrastructure de transport. Ces éléments jouent un rôle crucial dans la distribution du GNR, impactant directement les coûts de transport et donc les prix finaux dans différentes régions.

#### 2.2.2.1. Localisation des dépôts

La position des dépôts de carburant par rapport aux zones de consommation affecte directement les coûts de distribution. Plus les dépôts sont éloignés des consommateurs finaux (ex. agriculteurs, sites industriels), plus les coûts de transport sont élevés. En effet, les coûts supplémentaires liés aux trajets prolongés et aux besoins en carburant pour les camions-citernes sont répercutés sur les prix du GNR dans les régions isolées.

#### 2.2.2.2. Infrastructure de transport et prix du carburant

Les coûts associés au transport du Gazole Non Routier (GNR) en France reposent sur une combinaison de facteurs logistiques. Tout d’abord, le carburant, représentant une part significative des dépenses, est extrêmement sensible aux variations de prix du pétrole. Ces fluctuations influencent directement le coût des opérations de transport, impactant ainsi les prix finaux pour les consommateurs. Les frais liés à la main-d'œuvre, notamment les salaires et charges sociales des chauffeurs, sont également importants. Toute augmentation dans ces coûts, due par exemple à des ajustements salariaux ou à des pénuries de conducteurs qualifiés, accroît les charges globales du transport​.

Les frais de maintenance et d’entretien des véhicules, qui varient selon leur fréquence d’utilisation et les conditions de transport, ajoutent une autre couche de coûts. Le vieillissement des camions et les normes de sécurité plus strictes contribuent également à ces dépenses. Par ailleurs, l’amortissement des véhicules, calculé en fonction de leur durée de vie et de la fréquence de renouvellement de la flotte, constitue un coût structurel non négligeable. Enfin, les dépenses en assurances et les coûts de gestion administrative associés au transport – incluant la planification logistique et la conformité aux régulations – complètent les charges. Chacun de ces éléments contribue à façonner les coûts logistiques globaux, influençant ainsi le prix du GNR pour les secteurs dépendants (Rodrigue, 2024)​.

### 2.2.3 Facteurs démographiques et économiques régionaux

Pour comprendre comment les facteurs démographiques et économiques régionaux influencent les prix du GNR, on peut effectivement tracer un parallèle avec d’autres produits pétroliers, en s’appuyant sur des études de consommation et de distribution d’énergie en général.

#### 2.2.3.1. Densité de population

Des études sur d’autres carburants, comme l'essence et le diesel, montrent que les zones à haute densité de population tendent à avoir des coûts de distribution plus faibles en raison des économies d'échelle. Ces régions, qui concentrent la demande sur une zone géographique réduite, permettent des trajets de distribution plus courts et donc plus économiques. Dans les zones à faible densité, comme les campagnes, où les distances entre les points de livraison sont plus grandes, les coûts de transport augmentent, ce qui se répercute sur les prix de vente au détail des carburants​. Les produits pétroliers destinés aux zones moins peuplées affichent ainsi une prime logistique, comparable à celle observée pour le GNR dans des régions rurales ou éloignées (Brons, et al., 2008).

#### 2.2.3.2. Activité économique locale

Dans le cadre de l’analyse de l’impact de l’activité économique locale sur les prix du GNR, il est pertinent de considérer les zones où ce carburant est le plus utilisé, comme expliqué dans la section 2.2.2 dédiée à l’utilisation sectorielle du GNR. Les régions dominées par des secteurs consommateurs majeurs – en particulier l’agriculture, les travaux publics et l’industrie – montrent une demande particulièrement forte en GNR. Cette demande soutenue provient des besoins en carburant pour les machines agricoles, les engins de construction et autres équipements énergivores.

Ainsi, dans les zones à forte activité agricole ou industrielle, les volumes de GNR requis sont conséquents, ce qui permet aux acheteurs de négocier des prix plus compétitifs en raison de la demande élevée. En revanche, dans les régions où l’activité économique est plus modérée, les coûts de distribution et de transport augmentent souvent, ce qui se traduit par des prix de GNR plus élevés pour les utilisateurs finaux​.

### 2.2.4 Facteurs saisonniers et météorologiques

L'influence des facteurs saisonniers et météorologiques sur le GNR constitue un aspect intrigant de notre analyse. Bien que leur impact précis reste à quantifier par des études empiriques approfondies, plusieurs hypothèses méritent notre attention.

#### 2.2.4.1. Variations saisonnières de la demande

Les fluctuations saisonnières de la demande en GNR pourraient jouer un rôle significatif dans la dynamique des prix. L'hiver semble particulièrement crucial, avec une possible augmentation de la demande liée à deux facteurs principaux : d'une part, l'utilisation accrue du GNR pour le chauffage dans certains secteurs, et d'autre part, la consommation potentiellement plus élevée des engins agricoles et de chantier en conditions froides. À l'inverse, l'été pourrait connaître une baisse de la demande, résultant de la diminution des besoins en chauffage et du ralentissement saisonnier de certaines activités industrielles ou agricoles.

Au-delà de cette dichotomie été-hiver, certains secteurs utilisant le GNR, notamment l'agriculture, pourraient connaître des pics d'activité saisonniers lors des périodes de semis ou de récoltes, influençant ainsi la demande de manière ponctuelle mais potentiellement significative. Ces variations cycliques, si elles se confirment, pourraient se traduire par des fluctuations prévisibles des prix du GNR au fil des saisons.

#### 2.2.4.2. Impact des conditions météorologiques

Au-delà des variations saisonnières régulières, les phénomènes météorologiques ponctuels et parfois extrêmes peuvent exercer une influence significative sur le marché du GNR. Ces événements climatiques intenses affectent à la fois l'offre et la demande du carburant. Du côté de l'offre, les intempéries peuvent perturber la chaîne logistique, entraînant des pénuries localisées et des variations de prix à court terme. Les conditions extrêmes peuvent également impacter la production, avec le froid affectant le fonctionnement des raffineries et les canicules pouvant réduire l'activité de certaines installations industrielles.

Côté demande, les variations climatiques influencent directement la consommation de GNR. Par temps froid, l'efficacité réduite des moteurs augmente la consommation. Le secteur agricole, particulièrement sensible au climat, voit sa demande en GNR fluctuer selon les conditions météorologiques, que ce soit pour l'irrigation en période de sécheresse ou pour s'adapter à des saisons de culture modifiées par le climat.

Ces impacts s'inscrivent dans un contexte énergétique plus large, où des phénomènes météorologiques affectant d'autres sources d'énergie peuvent indirectement influencer la demande en GNR. Pour quantifier précisément ces effets, une analyse croisée des données météorologiques, de production, de distribution et des indicateurs économiques sectoriels est nécessaire, permettant ainsi de distinguer l'impact spécifique des conditions météorologiques sur le marché du GNR.

### 2.2.5 Conclusion

L'analyse des facteurs influençant les prix du Gazole Non Routier (GNR) en France révèle un réseau complexe d'interactions entre variables macroéconomiques, géographiques, logistiques, démographiques, économiques et météorologiques. Cette multiplicité de facteurs explique en grande partie les disparités régionales observées dans les prix du GNR, soulignant la complexité du marché énergétique français.

Les facteurs macroéconomiques, tels que le cours du baril de pétrole et le taux de change USD/EUR, jouent un rôle fondamental dans la détermination des prix de base du GNR. Cependant, ce sont les facteurs géographiques et logistiques qui contribuent significativement aux variations régionales des prix. La localisation des dépôts et la qualité des infrastructures de transport influencent directement les coûts de distribution, créant ainsi des disparités entre les régions urbaines et rurales.

Les aspects démographiques et économiques régionaux ajoutent une couche supplémentaire de complexité. La densité de population et l'activité économique locale, notamment dans les secteurs fortement consommateurs de GNR comme l'agriculture et l'industrie, influencent la demande et, par conséquent, les prix. Ces facteurs peuvent expliquer pourquoi certaines régions bénéficient de prix plus compétitifs malgré leur éloignement géographique des sources d'approvisionnement.

Enfin, l'impact des facteurs saisonniers et météorologiques, bien que moins étudié, semble jouer un rôle non négligeable dans les fluctuations temporelles des prix du GNR. Les variations saisonnières de la demande et les événements météorologiques extrêmes peuvent entraîner des perturbations à court terme dans l'offre et la demande, influençant ainsi les prix de manière ponctuelle mais potentiellement significative.

Cette analyse exhaustive des facteurs influençant les prix du GNR souligne la nécessité d'une approche multidimensionnelle pour comprendre et prédire les disparités régionales. Elle met en évidence l'importance de prendre en compte non seulement les tendances macroéconomiques globales, mais aussi les spécificités locales et temporelles dans toute analyse ou modélisation des prix du GNR en France. Cette compréhension approfondie servira de base solide pour les analyses quantitatives et les modélisations prédictives qui suivront dans cette étude.

## 2.3 Méthodes d'analyse des disparités régionales des prix énergétiques

L'analyse des disparités régionales des prix énergétiques, et plus particulièrement du GNR, nécessite une approche multidimensionnelle combinant diverses techniques statistiques et analytiques. Cette section présente les principales méthodes utilisées dans la littérature récente, qui seront appliquées dans notre étude.

### 2.3.1 Analyse statistique descriptive

Les analyses statistiques descriptives constituent la base de toute étude sur les disparités de prix. Elles permettent de caractériser la distribution des prix et d'identifier les tendances générales. Asche et al. (2013) ont utilisé ces techniques pour analyser les prix du gaz naturel en Europe, mettant en évidence l'importance des mesures de tendance centrale et de dispersion pour comprendre les dynamiques de marché. Dans notre étude, nous emploierons des histogrammes et des boîtes à moustaches pour visualiser la distribution des prix du GNR, suivant l'approche de Mirza et Bergland (2012) dans leur analyse des prix de l'électricité dans les pays nordiques.

### 2.3.2 Analyse spatiale

Les méthodes d'analyse spatiale sont cruciales pour comprendre les disparités géographiques des prix. La cartographie thématique, notamment les cartes choroplèthes, offre une visualisation intuitive des variations spatiales. Ramos et al. (2016) ont utilisé cette approche pour étudier les prix de l'essence au Portugal, révélant des patterns géographiques significatifs. Dans notre étude, nous adopterons une approche similaire pour le GNR en France.

### 2.3.3 Analyse temporelle

L'aspect temporel est essentiel dans l'analyse des prix énergétiques. La décomposition des séries temporelles, comme l'ont fait Zhao et al. (2018) pour les prix du charbon en Chine, permet d'isoler les tendances, les saisonnalités et les composantes irrégulières. Nous appliquerons des techniques similaires pour comprendre l'évolution des prix du GNR au fil du temps.

Les modèles ARIMA (Autoregressive Integrated Moving Average), largement utilisés dans la prévision des prix énergétiques (Nogales et Conejo, 2006), seront également considérés pour modéliser et prévoir les tendances futures des prix du GNR.

## 2.4 Collecte et structuration des données : méthodologies, considérations et outils

### 2.4.1 Approche méthodologique de la collecte de données

#### 2.4.1.1. Revue des méthodes de collecte de données dans les études de marchés énergétiques

Le web scraping est devenu une technique essentielle pour collecter des données sur les prix, soutenant ainsi les statistiques officielles et l’analyse économique. Les instituts de statistique explorent de plus en plus cette méthode pour moderniser leurs processus de collecte, en particulier pour les informations sur les prix (Virgillito & Polidoro, 2019). Comparé à la collecte manuelle, le web scraping présente des avantages importants, notamment une efficacité accrue et la production de statistiques de prix plus précises.

Son accessibilité et son faible coût en font également une option attrayante pour les chercheurs et les entreprises, avec des applications variées, allant des calculs d’indices de prix à l’analyse des tendances du marché (Śpiewanowski, et al., 2021). Malgré ses nombreux atouts, le web scraping présente certains défis, notamment en matière de réglementation, de risques technologiques et de besoins de formation (Boettcher, 2015). Toutefois, il reste un outil précieux pour les chercheurs et la société civile, permettant une collecte et une interprétation des données socioéconomiques plus rapides et approfondies (Uriarte, et al., 2020).

#### 2.4.1.2. Critères de sélection des sources de données

Lors de l'élaboration d'un projet visant à collecter des données à partir de sources ouvertes, d'APIs et de techniques de web scraping, en particulier pour l’étude analytique du marché du GNR, il est crucial de prendre en compte plusieurs critères clés. Ces critères garantissent la qualité, la fiabilité et la pertinence des données collectées. La littérature scientifique offre des perspectives précieuses sur ces considérations, notamment dans le contexte du big data et de la science des données.

**- La qualité des données :** La qualité des données constitue le fondement d'une analyse efficace. Des données de haute qualité se caractérisent par leur exactitude, leur exhaustivité, leur cohérence et leur actualité. Cai et Zhu (2015) soulignent les défis de l'évaluation de la qualité des données à l'ère du big data, notant que les méthodes traditionnelles d'évaluation de la qualité peuvent s'avérer insuffisantes en raison du volume et de la variété des données disponibles. Ils proposent un cadre complet pour évaluer la qualité des données, incluant des dimensions telles que l'exactitude, l'exhaustivité, la cohérence et l'actualité, essentielles pour garantir la fiabilité des données utilisées dans les analyses. Ce cadre est particulièrement pertinent pour les projets impliquant des ensembles de données diversifiés, tels que les prix du GNR et les données démographiques, où assurer l'intégrité des données est crucial pour une analyse précise (Cai & Zhu, 2015).

- **Crédibilité des sources :** La crédibilité des sources est vitale pour garantir l'intégrité de l'analyse. Landers et al. (2016) discutent de l'importance d'évaluer la crédibilité des sources de données, particulièrement dans le contexte du web scraping pour la recherche en psychologie. Ils soulignent que les ressources requises pour le web scraping nécessitent une planification minutieuse et une prise en compte des implications techniques et éthiques impliquées. Cela est particulièrement pertinent lors de la collecte de données à partir de diverses plateformes en ligne, car la crédibilité de la source peut avoir un impact significatif sur la validité des résultats. Les chercheurs doivent évaluer de manière critique la fiabilité des sources à partir desquelles ils extraient des données, en particulier lorsqu'il s'agit d'informations sensibles comme les données démographiques (Landers, et al., 2016).

- **Fréquence de mise-à-jour des sources :** La fréquence de mise à jour des données est un autre critère important. Des mises à jour régulières garantissent que les données restent pertinentes et reflètent les conditions actuelles. Khder (2021) discute du rôle de l'informatique en nuage (Cloud Computing) dans l'amélioration de la fiabilité des outils de web scraping, permettant une collecte et des mises à jour de données plus fréquentes. Cela est crucial pour les projets qui s'appuient sur des données en temps réel, telles que les relevés de température et les fluctuations de prix, car des mises à jour opportunes peuvent avoir un impact significatif sur la précision des analyses (Khder, 2021).

**- La couverture géographique :** La couverture géographique des données est cruciale, en particulier pour les analyses qui dépendent d'informations spécifiques à la localisation. La capacité à collecter et analyser des données à travers différentes régions géographiques peut fournir des informations précieuses sur les variations de prix et les tendances démographiques. Dogucu et Çetinkaya-Rundel (2020) soulignent la dépendance croissante des agences statistiques nationales au web scraping comme méthode de collecte de données, ce qui permet une couverture géographique plus complète. Cela est particulièrement pertinent pour les projets analysant les prix du GNR et les données démographiques, car la compréhension des variations régionales peut éclairer des analyses plus ciblées (Dogucu & Çetinkaya-Rundel, 2020).

**- La fiabilité des données :** La fiabilité des sources de données est cruciale pour garantir que les résultats dérivés des données sont valides. Juszczak (2023) souligne l'importance d'utiliser des données obtenues par web scraping pour analyser la dynamique des prix, notant que cette technique permet des réductions de coûts significatives et une amélioration de la précision dans la recherche sur les prix. Cela met en évidence la nécessité d'utiliser des données fiables pour soutenir une analyse efficace et une prise de décision, en particulier dans les études économiques où la précision des prix est primordiale (Juszczak, 2023).

**- L’accessibilité via les données ouvertes, les API ou le web scraping :** L'accessibilité est un facteur critique lors de la considération des sources de données. Les initiatives de données ouvertes fournissent un accès gratuit aux ensembles de données, améliorant la transparence et facilitant la recherche. Les APIs servent de points d'accès structurés aux données, permettant aux chercheurs de récupérer efficacement les informations. Virgillito et Polidoro (2019) discutent de la manière dont les offices statistiques explorent de plus en plus l'utilisation d'Internet comme source de données pour moderniser leurs processus de collecte, notamment par le biais de techniques de web scraping. Cette accessibilité est vitale pour les chercheurs, car elle leur permet de rassembler les données nécessaires à leurs analyses sans obstacles significatifs (Virgillito & Polidoro, 2019).

En conclusion, la collecte de données provenant de sources diverses ou par le biais du web scraping nécessite une attention particulière à plusieurs critères cruciaux : la qualité des données, la crédibilité des sources, les considérations éthiques, la couverture géographique, la fréquence de mise à jour, la fiabilité globale et l'accessibilité via les données ouvertes, les API ou le web scraping. L'adhésion rigoureuse à ces critères non seulement renforce la fiabilité des données collectées, mais garantit également que les résultats qui en découlent sont valides, pertinents et exploitables dans un contexte de recherche ou de prise de décision.

### 2.4.2 Techniquess et architecture de web scraping

#### 2.4.2.1. Revue des techniques de web scraping

Le web scraping s'est imposé comme une technique fondamentale dans le domaine de l'extraction de données à partir de sources web, englobant un éventail de méthodologies et d'outils adaptés à la diversité des structures web, qu'elles soient statiques ou dynamiques. La sélection judicieuse des outils d'extraction est intrinsèquement liée aux spécificités de la tâche d'extraction, notamment la nature du contenu web, la complexité de la structure HTML, le volume de données à extraire, et le format de sortie requis pour l'analyse ultérieure.

Dans l'écosystème Python, plusieurs bibliothèques se distinguent par leur efficacité et leur polyvalence dans le domaine du web scraping. Beautiful Soup, Requests, Scrapy et Selenium constituent le quartet prédominant, chacune offrant des fonctionnalités distinctes et complémentaires (Hajba, 2018; García, et al., 2020).

**Beautiful Soup** et **Requests** forment un tandem particulièrement efficace pour l'extraction de données à partir de pages web statiques. Beautiful Soup excelle dans l'analyse syntaxique des documents HTML et XML, offrant une interface intuitive pour la navigation et la recherche au sein de la structure arborescente, tandis que Requests simplifie l’envoi de requêtes HTTP (Landers, et al., 2016). Cette combinaison est particulièrement efficace pour traiter des documents HTML mal formés, comme l’indique Almarzouq et al. (2020), qui soulignent l’importance de l’extraction de données dans des environnements web hétérogènes. Requests, quant à elle, simplifie considérablement le processus d'envoi de requêtes HTTP, facilitant ainsi la récupération du contenu web (AlMarzouq, et al., 2020).

L'extraction de données à partir de pages web dynamiques, dont le contenu est généré ou modifié par JavaScript, présente des défis spécifiques que les méthodes traditionnelles peinent à surmonter. Dans ce contexte, **Selenium** s'impose comme un outil indispensable. En permettant l'automatisation des navigateurs web, Selenium facilite l'interaction avec le contenu dynamique d'une manière analogue à celle d'un utilisateur humain. Cette fonctionnalité est cruciale pour accéder aux données qui ne sont pas immédiatement disponibles dans le code source HTML initial, mais qui sont générées ou modifiées par des scripts côté client Néanmoins, l'utilisation de Selenium peut induire une latence supérieure à celle observée avec des bibliothèques comme Requests, en raison de la surcharge computationnelle liée à l'automatisation complète du navigateur (García, et al., 2020).

**Scrapy** se positionne comme un framework robuste, spécifiquement conçu pour les projets de web scraping à grande échelle. Son architecture modulaire intègre des fonctionnalités avancées de gestion des requêtes, de pipelines de données et d'analyse des réponses. La capacité de Scrapy à effectuer un traitement asynchrone représente un atout majeur, permettant une optimisation significative des performances lors de l'extraction simultanée de données à partir de multiples sources. Cette caractéristique le rend particulièrement adapté aux projets nécessitant une collecte de données massive et efficiente.

L'intégration du Cloud Computing aux méthodologies de web scraping représente une évolution significative, offrant des perspectives prometteuses en termes de scalabilité et de performance. Khder (2021) met en lumière le potentiel des plateformes cloud pour fournir un environnement d'exécution plus robuste et flexible pour les tâches de scraping.

La gestion efficace des défis liés au web scraping, tels que la détection et le contournement des mécanismes anti-bot, nécessite une approche sophistiquée. L'implémentation de techniques de rotation des adresses IP, de gestion des délais entre les requêtes, et l'utilisation de user-agents variés sont des stratégies couramment employées pour minimiser le risque de blocage. Ces techniques doivent être mises en œuvre avec discernement, en respectant les politiques d'utilisation des sites web ciblés et les considérations éthiques liées à l'extraction de données.

En conclusion, le choix des outils et techniques de web scraping est dicté par une analyse approfondie des caractéristiques du contenu web ciblé et des objectifs spécifiques du projet d'extraction. Beautiful Soup et Requests s'avèrent particulièrement efficaces pour l'extraction de données à partir de pages statiques, tandis que Selenium est indispensable pour l'interaction avec le contenu dynamique. Scrapy se distingue par sa capacité à gérer efficacement des projets d'extraction à grande échelle, offrant une plateforme robuste pour le développement de spiders complexes. L'intégration judicieuse de ces outils, couplée à l'utilisation stratégique des technologies de Cloud Computing, permet d'optimiser significativement l'efficacité, la scalabilité et la performance des opérations de web scraping.

#### 2.4.2.2. Architecture du système de scraping

L'intégration du cloud computing dans l'architecture des systèmes de web scraping a marqué un tournant significatif dans la collecte de données à grande échelle. Cette évolution répond aux besoins croissants de flexibilité, de scalabilité et d'efficacité face à des volumes de données en constante augmentation. Cette revue examine l'architecture cloud des systèmes de web scraping à travers cinq modules essentiels : planification des tâches, extraction des données, traitement et nettoyage, stockage, et suivi. En analysant les récentes avancées dans ce domaine, nous mettrons en lumière les innovations et les défis à l'intersection du web scraping et du cloud computing, offrant ainsi un aperçu des tendances actuelles et futures de cette technologie en pleine expansion.

**- Module de planification des tâches :**

Le module de planification des tâches est essentiel pour orchestrer l'exécution des tâches de scraping dans un environnement cloud. Aslam (2023) souligne les avantages du cloud computing dans la gestion du big data, notamment dans la planification des tâches, où l'allocation dynamique des ressources est cruciale (Aslam, 2023). Les architectures serverless proposées par les principales plateformes cloud permettent une exécution efficiente des tâches de scraping, minimisant la latence et maximisant le débit.

Des outils avancés de planification, tels qu'Apache Airflow et AWS Step Functions et Google Cloud Job Scheduler peuvent être intégrés dans les environnements cloud pour gérer des flux de travail complexes. Ces outils permettent l'orchestration de tâches multiples, facilitant la gestion des dépendances, des réessais et de l'exécution parallèle des tâches de scraping.

**- Module d'extraction des données :**

Le module d'extraction des données est responsable de la récupération des données à partir de diverses sources web. Mosa et Sakellariou (2019) mettent en évidence l'importance du placement dynamique des machines virtuelles dans les environnements cloud, optimisant ainsi les ressources utilisées pour l'extraction de données (Mosa & Sakellariou, 2019). Cette approche distribuée améliore non seulement la vitesse du processus d'extraction mais renforce également la tolérance aux pannes.

L'extraction basée sur les événements, rendue possible par des services cloud permet une collecte de données en temps réel. Ce modèle améliore la réactivité des opérations de scraping, permettant aux organisations de rester à jour avec les informations les plus récentes.

**- Module de traitement et de nettoyage des données :**

Zhang et al. (2013) soulignent l'efficacité des algorithmes de nettoyage de données spécifiquement conçus pour les environnements cloud. Ces algorithmes peuvent être déployés sur des plateformes cloud pour traiter efficacement de grands volumes de données. La nature distribuée du cloud computing permet un traitement parallèle des tâches de nettoyage des données, accélérant significativement la transformation des données brutes en formats structurés (Zhang, et al., 2013).

L'intégration d'algorithmes d'apprentissage automatique dans les solutions de traitement de données basées sur le cloud peut améliorer la qualité des données. L'application de techniques telles que la détection d'anomalies peut améliorer la précision et la pertinence des données collectées.

**- Module de stockage des données :**

Latifian (2022) discute du rôle significatif du cloud computing dans la gestion du big data, notamment des diverses solutions de stockage adaptées à différents types de données et modèles d'accès. La flexibilité des solutions de stockage cloud permet aux organisations de choisir le type de stockage le plus approprié en fonction de leurs besoins spécifiques, assurant une récupération et une gestion efficaces des données (Latifian, 2022).

Les solutions de stockage cloud offrent également des fonctionnalités de gestion du cycle de vie des données, permettant aux organisations d'automatiser le déplacement des données entre différentes classes de stockage en fonction des modèles d'accès.

**- Module de suivi et de contrôle :**

Dautov et al. (2014) soulignent l'importance des capacités autonomes dans les plateformes cloud, qui peuvent aider à automatiser les processus de surveillance et d'adaptation. L'utilisation d'outils de surveillance natifs du cloud permet aux organisations d'obtenir des informations sur les performances de leurs tâches de scraping, de détecter les anomalies et de mettre en œuvre des réponses automatisées aux problèmes qui surviennent (Dautov, et al., 2014).

Les capacités d'analyse intégrées aux plateformes cloud permettent aux organisations de visualiser les performances de leurs tâches de scraping dans le temps, facilitant l'identification des domaines d'amélioration et l'optimisation de leurs stratégies de scraping.

En conclusion, l'exploitation des solutions cloud pour l'architecture des systèmes de web scraping améliore la fonctionnalité et l'efficacité de chaque module. L'intégration des technologies cloud permet des opérations de web scraping évolutives, flexibles et robustes, capables de s'adapter aux demandes croissantes de collecte de données. Chaque module bénéficie des capacités uniques du cloud computing, permettant aux organisations d'optimiser efficacement leurs efforts de web scraping.

### 2.4.3 Considérations éthiques et légales du web scraping

Le web scraping s'est imposé comme un outil incontournable pour la collecte de données dans la recherche sur les prix alimentaires. Cependant, son utilisation soulève des questions éthiques et légales significatives qui méritent une attention particulière. Cette section examine les principales considérations à prendre en compte lors de l'utilisation du web scraping dans ce contexte spécifique, en s'appuyant sur la littérature académique et les meilleures pratiques du domaine.

#### 2.4.3.1. Cadre juridique et éthique

Le cadre juridique du web scraping est complexe et en constante évolution. Une relation "symbiotique" existe souvent entre les scrapers de données et les propriétaires de sites web, suggérant une acceptation pragmatique de cette pratique dans certains contextes (Hirschey, 2014).

Dans ce paysage juridique, plusieurs défis légaux et stratégies pour la comparaison des prix et la réutilisation des données ont été identifiés. Le respect des conditions d'utilisation des sites web, la prévention de la surcharge des serveurs, et la protection de la propriété intellectuelle des propriétaires de sites sont des aspects cruciaux à considérer (Zhu & Madnick, 2010). Ces considérations sont particulièrement importantes dans la recherche sur les prix alimentaires, où les données sont souvent sensibles et peuvent avoir des implications concurrentielles significatives.

#### 2.4.3.2. Mise en œuvre pratique

La mise en œuvre éthique du web scraping repose sur l'adoption de bonnes pratiques techniques. L'identification appropriée des bots de scraping, le respect des fichiers robots.txt, et la limitation de la fréquence des requêtes sont essentiels pour minimiser l'impact sur les serveurs (Hemenway & Calishain, 2004). Ces pratiques, bien que techniques, ont des implications éthiques importantes en assurant que le processus de collecte de données reste non intrusif et respectueux des ressources des sites web ciblés.

La transparence joue également un rôle crucial dans l'éthique du web scraping. Informer les propriétaires de sites web de l'intention de collecter des données, obtenir si possible un consentement explicite, et offrir une possibilité de désengagement du processus de collecte sont des pratiques recommandées (Vanden Broucke & Baesens, 2018).Cette approche transparente atténue les risques légaux et favorise une relation de collaboration bénéfique avec les sources de données, essentielle pour la pérennité de la recherche.

### 2.4.4 Considérations spécifiques à la recherche sur les prix des carburants

Dans le contexte de la recherche sur les prix du carburant, la protection des données sensibles revêt une importance capitale. L'anonymisation des données collectées pour protéger les identités des stations-service et des distributeurs est cruciale. L'agrégation des informations de manière à ne pas révéler de données commerciales sensibles, telles que les marges ou les stratégies de tarification, est également essentielle. Le respect des réglementations sur la protection des données, comme le RGPD en Europe, est une obligation légale qui influence directement la collecte, le stockage et l'utilisation des données de prix du carburant (Mondschein & Monda, 2018).

L'équité concurrentielle est particulièrement sensible dans le secteur du carburant. La collecte et l'utilisation des données de prix ne doivent en aucun cas créer d'avantages concurrentiels déloyaux ou faciliter des pratiques anticoncurrentielles. Le partage des résultats agrégés de la recherche pour bénéficier à l'ensemble du secteur est encouragé, tout en veillant scrupuleusement à éviter toute utilisation des données qui pourrait être perçue comme une tentative de manipuler les marchés ou les prix (Graef, et al., 2019).

# 3 Collecte et structuration des données de l’étude

## 3.3 Sources des variables de l’étude

Cette étude s'appuie sur un ensemble diversifié de sources de données, sélectionnées pour leur pertinence dans l'analyse des disparités régionales des prix du Gazole Non Routier (GNR) en France. La méthodologie employée vise à conjuguer des techniques contemporaines de collecte de données avec l'utilisation de sources officielles établies, dans le but d'obtenir une représentation multidimensionnelle du marché du GNR.

### 3.3.1 Sources des données de prix du GNR

Les données relatives aux prix du GNR constituent le fondement empirique de cette recherche. Un système de collecte automatisée de données (web scraping) a été mis en place, ciblant plusieurs sites de vente en ligne de GNR. Cette méthode permet d'obtenir des données de prix quotidiennes depuis avril 2024, couvrant plus de 2000 codes postaux uniques. Bien que cette couverture géographique soit étendue, il convient de noter qu'elle n'englobe pas la totalité des départements français, ce qui pourrait limiter la généralisation des résultats à l'échelle nationale.

L'anonymisation des sources a été rigoureusement observée, conformément aux principes éthiques de la recherche académique. En complément, des informations confidentielles concernant les dépôts et les régions ont été obtenues auprès de Carrefour, apportant un éclairage supplémentaire sur la structure logistique du marché.

### 3.3.2 Sources des facteurs macroéconomiques

L'étude intègre plusieurs indicateurs macroéconomiques susceptibles d'influencer les prix du GNR. Les fluctuations du taux de change de l’euro au dollar, EUR/USD sont analysées à partir de deux sources : les données annuelles de *Statista* et les données historiques détaillées de *Exchange-rates.org* et *Investing.com*. Les données relatives au cours du baril de pétrole, issues de l'Institut National de la Statistique et des Études Économiques (INSEE), sont également intégrées à l'analyse. L'indice du Comité National Routier (CNR), accessible via le site officiel de l'organisme, fournit également des informations sur l'évolution des coûts du transport routier en France.

### 3.3.3 Sources des données géographiques et démographiques

Afin de contextualiser l'analyse, diverses données géographiques et démographiques ont été incorporées. Les données de température quotidienne départementale, obtenues via Data.gouv.fr, permettent d'examiner l'impact potentiel des conditions météorologiques. Il convient toutefois de noter que ces données pourraient présenter des lacunes pour certains codes postaux ou périodes.

La liste des communes, comprenant leurs coordonnées géographiques et leur population, également issue de Data.gouv.fr, sert de base à l'analyse spatiale. Les données sectorielles, notamment le nombre d'exploitations agricoles et la répartition des entreprises par secteur, proviennent de l'Observatoire des territoires.

## 3.4 Analyse de la couverture géographique et de la représentativité de l'échantillon

### 3.4.1 Cartographie des points de collecte de données

Notre étude sur les prix du Gazole Non Routier (GNR) en France métropolitaine s'appuie sur un vaste échantillon de plus de 2000 codes postaux uniques. Cette couverture géographique étendue englobe une diversité de territoires, allant des zones urbaines densément peuplées aux régions rurales plus isolées. La collecte de données a été effectuée quotidiennement depuis avril 2024, fournissant ainsi un ensemble de données riche et actualisé.

La figure 1 illustre la distribution géographique des points de collecte de données sur les prix du GNR

Une image contenant carte, texte, atlas

Description générée automatiquement

Figure 1Distribution géographique des points de collecte de données sur les prix du GNR en France

### 3.4.2 Analyse de la répartition spatiale des données

### 3.4.3 Évaluation de la représentativité de l'échantillon

## 3.5 Fréquence et période de collecte des données

## 3.6 Structuration de la base de données

## 3.7 Organisation des données collectées

## 3.8 Description finale de la base de données

# 4 Méthodologie d’analyse

Approche analytique globale

Techniques d'analyse spatiale

Analyse des séries temporelles

# 5 Analyse des disparités régionales des prix du GNR

## 5.1 Cartographie des prix à l'échelle nationale

## 5.2 Identification des facteurs influents par région

## 5.3 Patterns géographiques et clusters de prix

## 5.4 Cas d'étude : focus sur des régions spécifiques

# Annexes

# Références bibliographiques

Office National des Forêts, 2021. *Rapport annuel 2021,* Paris: s.n.

AlMarzouq, M., AlZaidan, A. & Dallal, J. A., 2020. Mining github for research and education: challenges and opportunities. *International Journal of Web Information Systems,* pp. 451-473.

Aslam, F., 2023. Unleashing the power of cloud computing for big data management: advantages, challenges, and future prospects. *Asian Journal of Research in Computer Science,* pp. 290-295.

Assemblée nationale, 2021. *LOI n° 2021-1104 du 22 août 2021 portant lutte contre le dérèglement climatique et renforcement de la résilience face à ses effets,* Paris: Légifrance.

Beckmann, J. & Czudaj, R., 2013. Oil Prices and Effective Dollar Exchange Rates. *International Review of Economics & Finance,* pp. 21-31.

Boettcher, I., 2015. *Automatic data collection on the Internet (web scraping).* s.l.:s.n.

Brons, M., Nijkamp, P., Pels, E. & Rietveld, P., 2008. *A Meta-Analysis of the Price Elasticity of Gasoline Demand. A System of Equations Approach.* 2105-2122: Energy Economics.

Cai, L. & Zhu, Y., 2015. *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era.* s.l.:Data Science Journal, 14, 2..

Commissariat général au développement durable, 2018. *Consommation d'énergie dans le secteur de la construction : Les produits pétroliers, principale énergie dans le secteur de la construction,* s.l.: Ministère de la Transition écologique. URL : https://www.statistiques.developpement-durable.gouv.fr/media/573/download.

Commission européenne, 2009. *Directive 2009/30/CE du Parlement européen et du Conseil du 23 avril 2009,* Bruxelles: Union européenne.

Dautov, R., Paraskakis, I. & Stannett, M., 2014. Utilising stream reasoning techniques to underpin an autonomous framework for cloud application platforms. *Journal of Cloud Computing.*

Dogucu, M. & Çetinkaya-Rundel, M., 2020. Web scraping in the statistics and data science curriculum: challenges and opportunities. *Journal of Statistics and Data Science Education, ,* pp. 112-122.

Fernandois, A. & Medel, C., 2020. Geopolitical tensions, OPEC news, and the oil price: A granger causality analysis.. *Economic Analysis Review,* pp. 57-90.

García, B., Gallego, M., Gortázar, F. & Muñoz-Organero, M., 2020. A survey of the selenium ecosystem. *Electronics, 9(7), 1067.*

Graef, I., Tombal, T. & Streel, A., 2019. Limits and Enablers of Data Sharing. An Analytical Framework for EU Competition, Data Protection and Consumer Law. Dans: s.l.:Tilburg Law & Economics Center (TILEC).

Hajba, G. L., 2018. Website Scraping with Python: Using BeautifulSoup and Scrapy. Dans: s.l.:s.n., p. 41–96.

Hemenway, K. & Calishain, T., 2004. Spidering Hacks. Dans: s.l.:O'Reilly Publishing, Sebastopol, CA.

Hirschey, J. K., 2014. Symbiotic relationships : pragmatic acceptance of data scraping. Dans: s.l.:Berkeley Technology Law Journal. Vol. 29 No. 4, pp. 979-927.

INSEE, 2018. *En décembre 2018, le prix du pétrole chute de nouveau,* Paris: INSEE.

Inspection générale des finances, 2024. *Adaptation de la fiscalité aux exigences de la transition écologique,* Paris: Inspection générale des finances.

Institut technologique Forêt Cellulose Bois-construction, 2021. *Analyse des coûts d'exploitation forestière et impact des prix énergétiques,* s.l.: s.n.

Juszczak, A., 2023. The use of web-scraped data to analyse the dynamics of clothing and footwear prices. *The Polish Statistician,* pp. 15-33.

Khder, M. A., 2021. *Web scraping or web crawling: state of art, techniques, approaches and application.* s.l.:International Journal of Advances in Soft Computing and Its Applications, 145-168.

Landers, R. N., Brusso, R. C., Cavanaugh, K. J. & Collmus, A. B., 2016. *A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research..* s.l.:Psychological Methods, 21, 475-492.

Latifian, A., 2022. How does cloud computing help businesses to manage big data issues. *Kybernetes,* pp. 1917-1948.

Ministère de la Transition écologique, 2010. *Arrêté du 10 décembre 2010 définissant les spécifications techniques du GNR.* Paris: Ministère de la Transition écologique.

Ministère de la Transition Écologique, 2022. *Évolution des taxes et subventions sur les carburants non routiers,* Paris: Ministère de la Transition Écologique.

Mondschein, C. & Monda, C., 2018. The EU’s General Data Protection Regulation (GDPR) in a Research Context. Dans: s.l.:Fundamentals of Clinical Data Science, pp. 55-71.

Mosa, A. & Sakellariou, R., 2019. *Dynamic virtual machine placement considering cpu and memory resource requirements.* s.l., s.n.

Rodrigue, J.-P., 2024. *The spatial organization of transportation and mobility.* [En ligne]   
Available at: https://transportgeography.org/contents/chapter3/transport-costs/transport-costs-conditions/

Service de la Donnée et des Études Statistiques, 2023. *Chiffres clés de l'énergie - Édition 2023,* Paris: Ministère de la transition Écologique.

Śpiewanowski, P., Talavera, O. & Vi, L., 2021. *Applications of Web Scraping in Economics and Finance.* s.l.:Oxford Research Encyclopedia of Economics and Finance.

SSP, I. &., 2024. *Transformations de l'agriculture et des consommations alimentaires,* s.l.: Ministère de l’Agriculture et de la Souveraineté Alimentaire.

Terre-net, 2020. *Comment a évolué le prix du GNR en 2020.* [En ligne]   
Available at: https://www.terre-net.fr/carburants/article/175351/prix-du-gnr-quelle-evolution-en-2020-

TotalEnergies, 2024. *Gazole Non Routier : le carburant pour les engins non routiers.* [En ligne]   
Available at: https://services.totalenergies.fr/professionnels/energies-mobilite/carburants/produits/diesel/gazole-non-routier

Trévisiol, A., 2023. *« Le gazole non routier (GNR) reste la principale dépense énergétique des exploitations agricoles »* [Interview] (5 août 2023).

Uriarte, J. I., Toro, G. R. R. M. d. & Larrosa, J., 2020. *Web scraping based online consumer price index: The “IPC Online” case.* s.l.:s.n.

Vanden Broucke, S. & Baesens, B., 2018. Practical Web Scraping For Data Science. Dans: New York: Apress.

Virgillito, A. & Polidoro, F., 2019. *Big Data Techniques for Supporting Official Statistics : The Use of Web Scraping for Collecting Price Data.* s.l.:Web Services. (pp.728-744) .

Zhang, F. et al., 2013. Big data cleaning algorithms in cloud computing. International Journal of Online and Biomedical Engineering. *International Journal of Online and Biomedical Engineering (iJOE),,* p. 7.

Zhu, H. & Madnick, S. E., 2010. Legal challenges and strategies for comparison shopping and data reuse. Dans: s.l.:Journal of Electronic Commerce Research, Vol. 11 No.3, pp. 231-239.