**Analyse critique de la pertinence et des limites d'une étude sur les disparités régionales des prix du GNR en France : Évaluation des données collectées et du potentiel analytique**

Points positifs :

1. Richesse des données :
   * Vous disposez d'une large gamme de variables (32 au total), couvrant des aspects géographiques, démographiques, économiques, et météorologiques.
   * La variable 'prix' vous permet d'analyser directement les disparités de prix du GNR.
2. Couverture temporelle :
   * Avec un scraping quotidien depuis avril 2024, vous avez environ 6 mois de données, ce qui est suffisant pour détecter des tendances saisonnières et des patterns à court terme.
3. Granularité géographique :
   * Avec plus de 1900 codes postaux uniques, vous avez une granularité fine qui permet une analyse détaillée des variations locales.
4. Variables explicatives diversifiées :
   * Des variables comme 'indice\_CNR', 'USD\_to\_EUR', 'cours\_baril\_en\_USD' permettent d'analyser l'impact des facteurs macroéconomiques.
   * Les variables démographiques et économiques locales (population, densité, nombre d'entreprises par secteur) permettent d'étudier les facteurs locaux influençant les prix.
5. Potentiel pour des analyses multidimensionnelles :
   * La combinaison de données géographiques, temporelles, et économiques permet des analyses croisées riches (par exemple, l'évolution des prix dans le temps selon la densité de population).

Points négatifs :

1. Couverture géographique incomplète :
   * Le fait de ne pas couvrir tous les départements peut limiter la généralisation des résultats à l'échelle nationale.
2. Période d'observation relativement courte :
   * 6 mois de données peuvent être insuffisants pour capturer des tendances à long terme ou des cycles économiques complets.
3. Potentielles données manquantes :
   * Certaines variables comme 'tmoy' (température moyenne) pourraient avoir des données manquantes pour certains codes postaux ou certaines périodes.
4. Absence de certaines variables potentiellement pertinentes :
   * Par exemple, des données sur les politiques locales ou les infrastructures de transport pourraient manquer.

Discussion et recul :

1. Pertinence de l'étude :
   * Malgré les limitations, cette étude offre une analyse inédite et détaillée des prix du GNR en France, avec une granularité géographique et temporelle rare.
   * Elle permet d'explorer les facteurs influençant les prix du GNR à différentes échelles (locale, régionale, nationale).
2. Approches analytiques possibles :
   * Analyse des séries temporelles par code postal pour identifier les tendances et saisonnalités locales.
   * Modélisation géospatiale (comme la régression géographiquement pondérée) pour cartographier les variations de prix et leurs déterminants.
   * Clustering des codes postaux basé sur les caractéristiques économiques et démographiques pour identifier des "zones de prix" similaires.
3. Potentiel prédictif :
   * Les variables macroéconomiques et locales permettent de développer des modèles prédictifs robustes, en utilisant des techniques de machine learning comme les forêts aléatoires ou les réseaux de neurones.
4. Limites et perspectives :
   * L'étude pourrait être enrichie à l'avenir en incluant plus de départements et en prolongeant la période d'observation.
   * Une analyse comparative avec d'autres pays européens pourrait offrir des perspectives intéressantes sur les spécificités du marché français du GNR.

En conclusion, malgré certaines limitations, votre étude offre une base solide pour une analyse approfondie et pertinente des disparités régionales des prix du GNR en France. La richesse et la diversité des variables, combinées à une granularité géographique fine, permettent des analyses multidimensionnelles qui peuvent apporter des insights précieux pour les acteurs du secteur et les décideurs politiques