



INSTITUT DE  
TECHNOLOGIE

UNIVERSITÉ  
Clermont Auvergne

## SAE 2-06 Analyse de données, reporting et datavisualisation

Année : 2025

➤ TOURE Ibrahima

UCA  
Professeur M. BRUNO Lagarde

# Résumé

À partir d'un jeu de données fourni par la SNCF, nous avons défini une problématique essentielle : **"Quel est l'impact du trafic ferroviaire sur les dynamiques régionales, notamment en termes de fréquentation des gares, d'évolution du nombre de voyageurs et du marché immobilier ?"**

Pour y répondre, nous avons mis en place un reporting et une datavisualisation afin d'analyser en profondeur l'influence du trafic ferroviaire sur les territoires. En exploitant **Power BI**, nous avons étudié l'évolution de la fréquentation des gares sur plusieurs années, permettant ainsi d'identifier les tendances et les gares les plus fréquentées.

Une **analyse de corrélation** a été menée entre le nombre de départs de trains et l'affluence des voyageurs afin de mieux comprendre les flux de mobilité. Par ailleurs, nous avons exploré l'impact du réseau ferroviaire sur le marché immobilier, en analysant les variations des prix en fonction de l'accessibilité aux transports.

L'objectif de cette étude est d'illustrer comment le transport ferroviaire façonne les territoires et influence leur attractivité.

# Sommaire

<b>I.</b>	<b>Introduction.....</b>	<b>3</b>
<b>II.</b>	<b>Traitement des Données.....</b>	<b>4</b>
	1. Choix et Analyse des Données	
	2. Nettoyage et préparation des données	
<b>III.</b>	<b>Analyse de la fréquentation des gares.....</b>	<b>5</b>
	1. Les 30 gares les plus fréquentées en France (2015-2022)	
<b>IV.</b>	<b>Évolution des départs de trains (2017-2022) .....</b>	<b>6</b>
	1. Tendances avant la pandémie (2017-2019)	
	2. Impact de la pandémie de 2020	
<b>V.</b>	<b>Évolution du nombre de voyageurs (2015-2022).....</b>	<b>7</b>
	1. Analyse régionale des flux de voyageurs	
	2. Impact de la pandémie et reprise post-Covid	
	3. Relation entre le nombre départ de trains et le nombre de voyageurs	
<b>VI.</b>	<b>Corrélation entre transport ferroviaire et immobilier .....</b>	<b>9</b>
	1. Influence des gares sur les prix immobiliers	
	2. Impact de la densité ferroviaire sur le marché immobilier	
<b>VII.</b>	<b>Conclusion .....</b>	<b>11</b>

## I. Introduction

Ce projet repose sur une base de données fournie par la SNCF, recensant l'ensemble des circulations ferroviaires en France entre 2015 et 2022. Elle regroupe des informations détaillées sur les gares, les types de locomotives, le nombre de voyageurs, des données relatives aux prix de l'immobilier etc.

L'objectif de cette étude est d'explorer les dynamiques du transport ferroviaire et son impact sur le territoire à travers une approche de datavisualisation. **Grâce à Power BI**, nous avons traité et analysé ces données afin d'identifier des tendances clés, telles que l'évolution du nombre de voyageurs, la variation des départs de trains par catégorie et l'influence du réseau ferroviaire sur le marché immobilier.

Nous avons conçu des **visualisations pertinentes** qui permettent d'interpréter ces phénomènes de manière claire et efficace. L'analyse de ces graphiques offre ainsi une meilleure compréhension du rôle du transport ferroviaire dans l'aménagement du territoire et son influence sur l'attractivité des différentes régions.

## II. Traitement des Données

### 1. Choix et Analyse des Données

Dans le cadre de notre analyse, nous avons sélectionné uniquement les colonnes les plus pertinentes pour répondre efficacement à notre problématique. Bien que certaines variables, telles que le nombre de foyers avec un revenu, la population ou encore le nombre de logements, puissent être intéressantes, elles ne contribuaient pas directement à notre étude et ont donc été écartées.

Pour une extraction optimale des informations, nous avons privilégié les colonnes suivantes :

- **Code géographique de la commune**
- **Code du département**
- **Code unique de la gare**
- **Code postal de la commune**
- **Longitude et latitude**
- **Nombre de voyageurs (2015-2022)**
- **Nombre de départs selon le type de locomotive (TER, TGV, Intercités, Transilien)**
- **Prix médians au mètre carré**

Afin d'interpréter correctement les données, nous avons utilisé un **dictionnaire des colonnes**, permettant d'associer chaque intitulé à sa véritable signification.

Par ailleurs, pour enrichir notre analyse, nous avons intégré une **nouvelle base de données** contenant des informations sur les régions, les codes départementaux et d'autres variables géographiques. Celle-ci sera exploitée ultérieurement via des jointures, afin de déterminer à quelle région appartient chaque gare.

Enfin, des codes DAX ont été créés pour générer des tables permettant de structurer et d'organiser les données de manière efficace. Ains Chaque **table** est conçue pour apporter une information essentielle, permettant ainsi d'améliorer la qualité et la pertinence des visualisations.

### 2. Nettoyage et préparation des données

En premier lieu, nous avons ouvert le jeu de données "DonneesSNCF.csv" pour comprendre son format. Lors de cette ouverture, nous avons constaté qu'il y avait un problème avec les données. Pour mieux comprendre leur structure, nous avons essayé d'ouvrir le fichier sur Power BI. Cependant, Power BI nous a signalé que certaines transformations n'étaient pas appliquées correctement à toutes les colonnes.

En examinant cela de plus près, nous avons remarqué que certaines colonnes, censées être des valeurs numériques, étaient traitées comme des chaînes de texte. Lors de la tentative de transformation de ces colonnes en entiers, nous avons perdu certaines données existantes. Ce problème était principalement lié à des espaces entre les nombres et à l'utilisation de points au lieu de virgules dans les valeurs numériques. Pour répondre efficacement à notre problématique,

nous avons mis en place un script Python permettant d'extraire et de nettoyer les données issues du fichier **DonneesSNCF.csv**.

Ce script commence par charger les données tout en s'assurant que l'encodage est compatible. Ensuite, il supprime les espaces superflus dans les noms de colonnes et extrait uniquement les colonnes pertinentes, incluant des informations sur les gares (nom, code, géolocalisation), le nombre de voyageurs (2015-2022), le prix médian de l'immobilier et les départs de trains (TER, TGV, Intercités, Transilien). Pour garantir la qualité des données, le script nettoie également les noms des gares, remplaçant les séparateurs incohérents et supprimant les espaces inutiles. Les colonnes numériques sont corrigées en supprimant les espaces et en convertissant les valeurs en entiers, tout en gérant les erreurs de conversion. Une **vérification** est effectuée pour détecter d'éventuelles différences avec les données initiales. Enfin, les données nettoyées sont sauvegardées dans un nouveau fichier "**Problematique\_liees\_a\_la\_gare.csv**", prêtes à être exploitées pour nos analyses.

### **III. Analyse de la fréquentation des gares**

#### ***1. Les 30 gares les plus fréquentées en France (2015-2022)***

L'analyse de la fréquentation des gares françaises de 2015 à 2022 repose sur le jeu de données proposé par la SNCF comprenant 2761 gares. L'objectif est d'identifier celles qui accueillent le plus de voyageurs.

Les résultats montrent une forte concentration en Île-de-France. Parmi les trente 30 gares les plus fréquentées, vingt-quatre 24 se situent à Paris. Les plus importantes sont Paris Saint-Lazare et Paris Gare de Lyon. Seules six 6 gares appartiennent à d'autres villes françaises.

Ce phénomène s'explique en partie par le rôle de Paris, qui est à la fois la capitale et l'une des villes les plus visitées au monde. Cependant, l'organisation du réseau ferroviaire influence aussi cette tendance. L'évolution des offres de train varie selon leur nature. Les trains du quotidien, comme les TER en région et le Transilien en Île-de-France, répondent aux besoins des trajets courts et fréquents. Les trains longue distance, tels que les TGV et les Intercités, assurent des liaisons entre les grandes villes sur des distances plus importantes.



#### IV. Évolution des départs de trains (2017-2022)

L'analyse se concentre sur l'évolution des offres ferroviaires en mesurant le nombre de départs effectués par chaque type de train.

##### 2. Tendances avant la pandémie (2017-2019)

Avant l'arrivée de la pandémie de COVID-19, les données indiquent une diminution du nombre de départs pour les Intercités, les TGV et les Transilien. Cette tendance pourrait résulter de divers facteurs, tels que des ajustements dans l'offre ferroviaire, une baisse de la demande ou une réorientation vers d'autres modes de transport.

##### 3. Impact de la pandémie en 2020

L'année 2020 a été marquée par une baisse significative du nombre de départs pour tous les types de trains, probablement en raison des restrictions de déplacement liées à la crise sanitaire. Les pourcentages de diminution par rapport à 2019 sont les suivants :

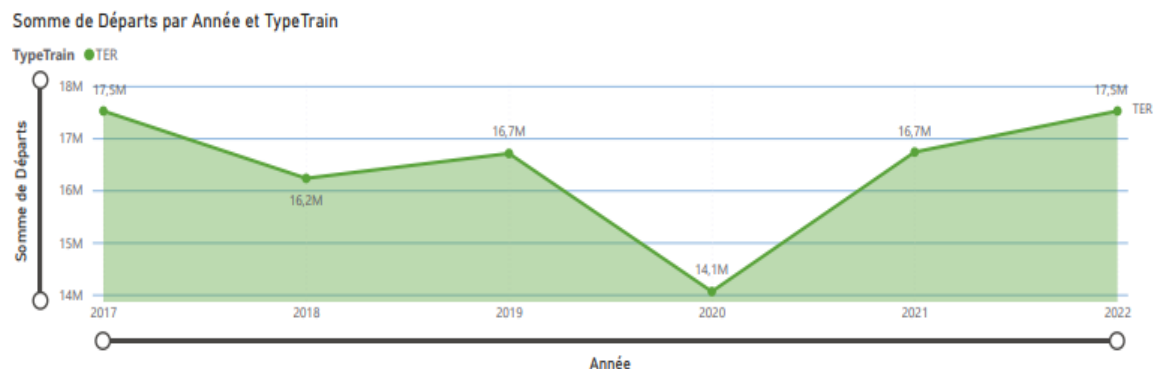
- **Intercités** : une diminution de 66,67%.
- **TER** : une diminution de 15,57%.
- **TGV** : une diminution de 31,33%.
- **Transilien (TN)** : une diminution de 18,45%.

##### 4. Reprise après la pandémie (2021-2022)

Après la levée progressive des restrictions, une reprise du nombre de départs a été observée pour la plupart des types de trains. Cependant, les Intercités ont connu une reprise plus lente, suggérant un impact durable de la crise sur ce service. Cette situation pourrait s'expliquer par une demande moins forte ou par des ajustements opérés par la SNCF dans l'offre de trains longue distance classiques.

**Remarque** : Il est important de noter que le nombre de départs de trains et le nombre de voyageurs transportés ne sont pas nécessairement corrélés. En effet, une réduction du nombre de départs n'implique pas forcément une diminution proportionnelle du nombre de voyageurs,

car d'autres facteurs, tels que la capacité des trains et la demande des passagers, peuvent influencer cette relation. Dans ce projet, nous nous efforcerons de mieux illustrer cette distinction dans les prochaines visualisations.



#### IV. Évolution du nombre de voyageurs (2015-2022)

##### 1. Analyse régionale des flux de voyageurs

Pour analyser l'évolution du nombre de voyageurs entre 2015 et 2022, nous avons effectué une étude par région. En téléchargeant une base de données complémentaire contenant des informations sur les régions, nous avons pu faire une jointure avec notre base de données existante. Cela nous a permis de répartir les gares par région et d'analyser les flux de voyageurs de manière plus détaillée.

L'Île-de-France se distingue comme la région ayant accueilli le plus grand nombre de voyageurs, dépassant les 10 milliards de passagers entre 2015 et 2021. Toutefois, il est important de noter que les données de 2022 ne sont pas disponibles pour cette région. Après l'Île-de-France, la région Auvergne-Rhône-Alpes est en deuxième position, bien qu'aucune autre région n'ait atteint le seuil d'un milliard de voyageurs pendant la période de 2015 à 2022.

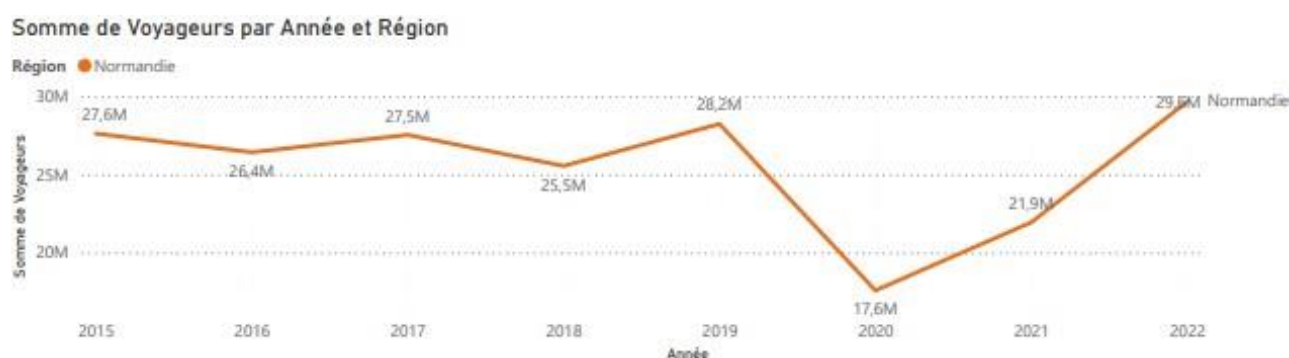


##### 2. Impact de la pandémie et reprise post-Covid



L'impact de la pandémie sur les flux de voyageurs est particulièrement significatif, avec une forte réduction des déplacements en 2020 dans toutes les régions. Par exemple, l'Île-de-France a connu une diminution de 50,28 %, passant de 1.79 milliard de voyageurs en 2019 à 0.89 milliard en 2020. Les régions telles que la Bretagne, la Bourgogne-Franche-Comté, et les Hauts-de-France ont également observé des baisses importantes, avec des diminutions allant de 28,82 % à 37,59 %.

Les données de 2020 montrent des baisses très marquées, suivies d'une reprise partielle en 2021. Cette tendance pourrait être due à la levée des restrictions sanitaires et à une reprise progressive de l'activité économique.



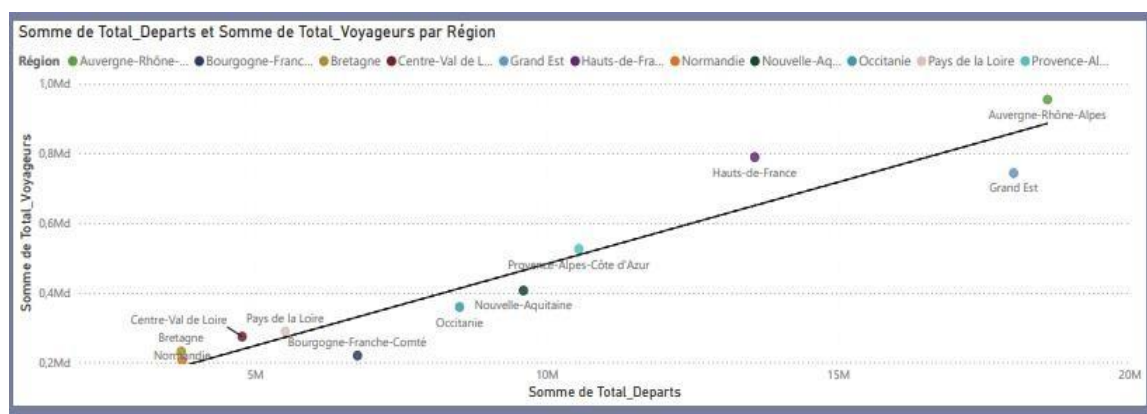
### 3. Relation entre le nombre départ de trains et le nombre de voyageurs Comme mentionné

Précédemment, notre objectif était d'analyser si une augmentation du nombre de départs de trains entraîne systématiquement une hausse du nombre de voyageurs.

Dans le cas de l'Île-de-France, cette relation ne semble pas s'appliquer directement. En effet, malgré un nombre impressionnant de plus de **10 milliards de voyageurs**, cette région demeure paradoxalement celle ayant le **moins de départs de trains**.

Cependant, hormis l'Île-de-France qui représente une modalité atypique dans notre analyse nous avons décidé de l'exclure de notre visualisation afin de mieux évaluer la tendance générale. Une fois cette exclusion faite, nous avons observé une **corrélation positive** entre le **nombre de départs de trains** et le **nombre de voyageurs** dans les autres régions. Autrement dit, plus une région bénéficie d'un grand nombre de départs, plus elle enregistre un nombre élevé de voyageurs.

Cela confirme qu'en dehors du cas particulier de l'Île-de-France, l'offre de trains et la demande en transport ferroviaire évoluent généralement dans le même sens.

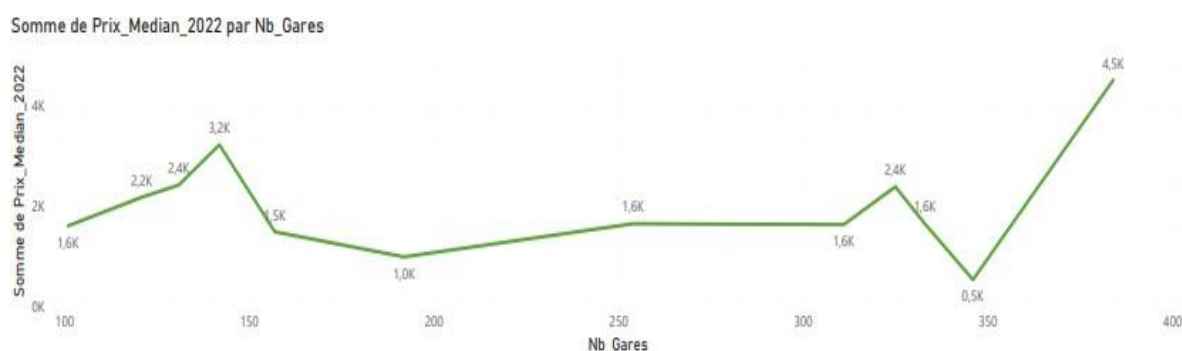


## V. Corrélation entre transport ferroviaire et immobilier

### 1. Influence des gares sur les prix immobiliers

Nous avons analysé l'impact des gares sur les prix médians de l'immobilier au mètre carré en étudiant le nombre de gares par région. Il est important de noter que notre étude des prix immobiliers porte uniquement sur le premier semestre de l'année 2022. La courbe des tendances montre que les indications seraient similaires même si l'on considère les années précédentes. Certaines régions comptent de nombreuses gares avec un grand nombre de voyageurs. Par exemple, l'Île-de-France possède 384 gares et a enregistré plus de 10 milliards de voyageurs entre 2015 et 2021. Toutefois, en raison de sa spécificité, nous avons choisi de l'exclure de certaines visualisations pour éviter que sa taille et son nombre élevé de voyageurs ne faussent ou n'altèrent la représentation des données des autres régions.

Notre analyse montre que le nombre total de gares dans une région n'influence pas directement les prix de l'immobilier au mètre carré. Certaines régions avec de nombreuses gares n'ont pas connu de variations significatives des prix immobiliers, malgré un nombre élevé de gares 311 en Nouvelle-Aquitaine et 346 dans le Grand Est, ces régions affichent des prix immobiliers médians relativement bas.

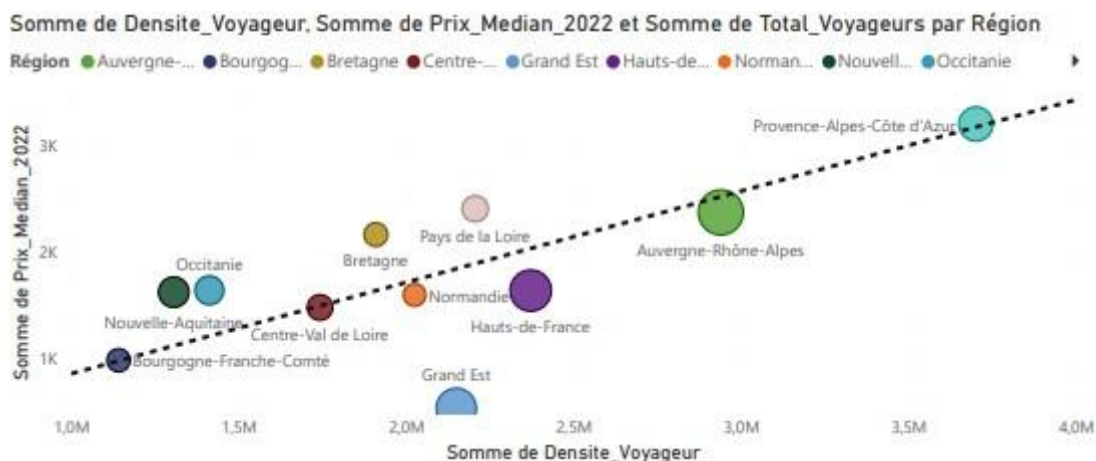


## 2. Impact de la densité ferroviaire sur le marché immobilier

En approfondissant, nous avons examiné la densité de voyageurs par gare et sa relation avec les prix immobiliers. Les résultats indiquent que plus la densité de voyageurs est élevée, plus les prix de l'immobilier ont tendance à augmenter. Cette corrélation peut s'expliquer par :

- **Attractivité accrue** : Les zones avec une forte densité de voyageurs sont souvent mieux desservies, les rendant plus attractives pour les résidents et les entreprises.
- **Développement économique** : Une fréquentation élevée peut stimuler le développement commercial et des infrastructures, augmentant ainsi la demande immobilière.

Ces observations suggèrent que la densité de voyageurs peut être un indicateur pertinent pour anticiper les tendances du marché immobilier dans les zones desservies par le réseau ferroviaire.



## **VI. Conclusion**

Cette étude nous a permis d'explorer l'évolution du transport ferroviaire en France entre 2015 et 2022 et d'évaluer son impact sur le territoire. Nous avons constaté une forte variation du nombre de voyageurs au fil des années, avec une baisse marquée en 2020 due à la pandémie, suivie d'une reprise progressive. L'analyse des flux régionaux a montré que l'Île-de-France domine largement en termes de fréquentation, tandis que d'autres régions, bien que possédant de nombreuses gares, enregistrent un trafic plus modéré. Concernant le lien entre transport ferroviaire et immobilier, nos résultats indiquent que ce n'est pas le nombre de gares qui influence directement les prix, mais plutôt la densité de voyageurs dans une région. En conclusion, le transport ferroviaire joue un rôle structurant dans l'organisation des territoires, mais son impact sur le marché immobilier varie selon la fréquentation des gares plutôt que leur simple présence. Cette étude met en lumière des tendances essentielles et pourrait servir de base à de futures analyses approfondies.