**Rapport de Travail Pratique : Visualisation des Données de Transport en Commun**

# 1. Introduction

## 1.1. Contexte

Les transports en commun constituent l’un des piliers fondamentaux de la mobilité urbaine et interurbaine, facilitant les déplacements quotidiens de millions de personnes à travers le monde. En Espagne, le réseau de transport public comprend un large éventail de moyens de transport, notamment les bus, les trains, les métros et les tramways. L'efficacité de ces services repose sur une gestion optimale des horaires, des itinéraires et de la fréquentation des lignes, ce qui requiert une analyse approfondie des données collectées.

Avec l’avènement du Big Data et des techniques avancées de visualisation de données, il est possible d’analyser ces flux de transport de manière précise. Les données issues des transports publics permettent d’identifier des tendances et d'optimiser l'allocation des ressources afin de réduire les temps d'attente, minimiser la congestion et améliorer l’expérience utilisateur.

## 1.2. Problématique

Malgré les efforts entrepris pour améliorer les réseaux de transport public, plusieurs problèmes persistent, notamment les retards, la saturation de certaines lignes aux heures de pointe, et une mauvaise répartition des fréquences selon les périodes de la journée. La visualisation des données peut jouer un rôle crucial dans l’identification de ces problèmes et dans la proposition de solutions basées sur des analyses factuelles.

L'objectif de ce travail est donc de réaliser une exploration et une visualisation des données de transport public en Espagne afin de répondre à plusieurs questions clés :

* Quels sont les itinéraires les plus empruntés et ceux qui sont sous-utilisés ?
* Quels sont les horaires de transport les plus fréquentés ?
* Comment évolue la fréquence des trajets en fonction du jour et de l'heure ?
* Existe-t-il des variations saisonnières affectant l’usage des transports en commun ?
* Comment peut-on optimiser l'offre de transport en fonction de ces observations ?

Ce rapport vise à analyser ces questions en exploitant un jeu de données public sur les transports en Espagne, en appliquant diverses techniques de visualisation et en dégageant des tendances pertinentes qui peuvent guider les décideurs.

# 2. Revue de Littérature

De nombreuses recherches ont exploré la visualisation des données de transport en commun. Parmi celles-ci :

* L'utilisation des Systèmes d'Information Géographique (SIG) pour cartographier les itinéraires et analyser les flux de transport.
* L'exploitation des données GPS des bus et trains pour modéliser la ponctualité et l'affluence.
* L'application des algorithmes de clustering pour identifier des modèles de trafic.
* L'usage de la visualisation interactive avec des outils comme Plotly et Matplotlib pour mieux représenter les tendances de transport.

# 3. Questions de Recherche

Voici cinq questions clés explorées dans ce travail :

* **Quels sont les itinéraires les plus empruntés et ceux qui sont sous-utilisés ?**
* **Graphique en lignes** : Utilisé pour représenter la fréquentation des différents itinéraires sur une période donnée.
* **Diagramme de dispersion (scatter plot)** : Pour identifier les itinéraires les plus et les moins empruntés en fonction de différents paramètres (distance, fréquence, etc.).
* **Quels sont les horaires de transport les plus fréquentés ?**
* **Graphique en lignes** : Pour visualiser l'évolution du nombre de trajets par heure de la journée.
* **Diagramme de dispersion** : Identification des pics d’utilisation en fonction des heures.
* **Comment évolue la fréquence des trajets en fonction du jour et de l'heure ?**
* **Graphique en lignes** : Permet de voir l'évolution des trajets sur plusieurs jours et d'identifier des tendances.
* **Heatmap temporelle** : Représentation des jours/heures avec le plus de trajets sous forme de matrice colorée.
* **Existe-t-il des variations saisonnières affectant l’usage des transports en commun ?**
* **Graphique en lignes** : Suivi de la tendance des trajets sur plusieurs mois pour détecter des variations saisonnières.
* **Scatter plot** : Comparaison de la fréquentation des transports selon différentes périodes de l’année.
* **Comment peut-on optimiser l'offre de transport en fonction de ces observations ?**
* **Graphique en lignes** : Analyse de la dispersion des temps d’attente et des retards.
* **Diagramme de dispersion** : Évaluation de la corrélation entre différents facteurs influençant l’optimisation du réseau de transport.

# 4. Méthodologie

Nous avons procédé comme suit :

**Exploration et Prétraitement des Données**

* Chargement du dataset public\_transportation\_data.csv.
* Inspection de la structure des données (types, valeurs manquantes, etc.).
* Conversion des dates en format approprié.
* Filtrage et nettoyage des données incohérentes.

**Visualisation des Données**

* Histogrammes pour représenter la distribution des horaires.
* Cartes interactives pour visualiser les itinéraires les plus fréquents.
* Graphiques temporels pour analyser l'évolution de la fréquence des trajets.
* Heatmaps pour identifier les périodes de forte et faible affluence.

# 5. Résultats et Analyses

Les analyses réalisées ont permis d'obtenir plusieurs observations :

* Certains itinéraires sont significativement plus utilisés que d'autres.
* La distribution des horaires met en évidence des pics de fréquentation aux heures de pointe.
* La fréquence des trajets varie selon les jours de la semaine, avec une baisse le week-end.
* Des tendances saisonnières sont perceptibles, notamment une baisse en période estivale.

# 6. Discussion et Commentaire

L'analyse des données du réseau de transport en commun en Espagne nous a permis de dégager des tendances significatives et de mieux comprendre les flux de mobilité urbaine. Toutefois, plusieurs aspects méritent d'être discutés pour mieux interpréter les résultats obtenus.

Les visualisations ont révélé plusieurs tendances clés. Par exemple, l'analyse des stations d'origine et de destination a montré que certaines gares sont nettement plus utilisées que d'autres, comme illustré par les diagrammes en barres obtenus via sns.countplot(). Ces résultats indiquent des points de concentration du trafic, nécessitant potentiellement une meilleure répartition des flux.

D'autre part, la répartition des types de trains montre que certains modèles sont préférés par les usagers, probablement en raison de leur rapidité ou de leur confort. Le countplot des types de train a permis d’identifier les catégories les plus utilisées, tandis qu’un lineplot a mis en évidence une relation entre le type de train et le prix moyen des billets.

En analysant les tendances temporelles, nous avons observé une fluctuation du trafic en fonction des jours et des heures. Un lineplot des prix des trajets par heure de la journée a révélé que les coûts sont souvent plus élevés aux heures de pointe, ce qui peut être un indicateur d’une tarification dynamique en réponse à la demande accrue.

Un autre point intéressant concerne la fréquence d'utilisation des transports publics. Une heatmap illustrant la répartition des trajets en fonction de l'heure et du jour a permis d’identifier les périodes les plus chargées et celles où le réseau est sous-utilisé. Ces informations sont précieuses pour ajuster les horaires et optimiser l’offre de transport.

Enfin, nous avons également utilisé des boxplots pour évaluer la variabilité des prix en fonction de divers facteurs, tels que le type de train et la distance parcourue. Ces visualisations permettent de mieux comprendre les écarts tarifaires et d’identifier des opportunités d’optimisation.

Les résultats obtenus soulignent l'importance d’une gestion optimisée des transports en commun. Une meilleure allocation des ressources pourrait être envisagée, notamment en adaptant la fréquence des trajets aux périodes de forte demande et en améliorant les infrastructures dans les stations les plus fréquentées.

Par ailleurs, l’analyse des tarifs pourrait être approfondie afin d’explorer l’impact d’une tarification dynamique sur la répartition du trafic et l’accessibilité des services. En combinant ces données avec des analyses de satisfaction des usagers, il serait possible d’affiner davantage les recommandations pour améliorer l'expérience de voyage.

# 7. Conclusion

L’analyse et la visualisation des données de transport en commun en Espagne ont permis de dégager plusieurs tendances intéressantes. En examinant les itinéraires les plus fréquemment empruntés, il a été possible d'identifier des zones de forte affluence ainsi que des lignes moins utilisées. De plus, l’étude des horaires a révélé des pics de fréquentation aux heures de pointe, confirmant la nécessité d’adapter la fréquence des services en conséquence.

L’analyse temporelle des trajets a mis en évidence des variations selon les jours de la semaine et les saisons, permettant ainsi d'anticiper les besoins des usagers en fonction des périodes de l'année. Par exemple, une baisse de fréquentation a été observée pendant certaines périodes de vacances scolaires, tandis que les jours ouvrables présentent une congestion accrue le matin et en fin d’après-midi.

L’utilisation de techniques de visualisation interactives a été un atout précieux pour mieux interpréter ces données. En permettant d’explorer dynamiquement les tendances, elles offrent un outil puissant aux gestionnaires de transport public pour ajuster les services en fonction des besoins réels.

En somme, la visualisation des données de transport constitue un levier essentiel pour comprendre, analyser et améliorer les réseaux de transport public. Cette étude a permis de poser les bases d'une analyse approfondie, et ouvre la voie à de nombreuses améliorations possibles pour une meilleure gestion de la mobilité urbaine.

Table des matières

**[Rapport de Travail Pratique : Visualisation des Données de Transport en Commun](#_Toc189958017)** [1](#_Toc189958017)

[1. Introduction 1](#_Toc189958018)

[1.1. Contexte 1](#_Toc189958019)

[1.2. Problématique 1](#_Toc189958020)

[2. Revue de Littérature 2](#_Toc189958021)

[3. Questions de Recherche 2](#_Toc189958022)

[4. Méthodologie 3](#_Toc189958023)

[**Exploration et Prétraitement des Données** 3](#_Toc189958024)

[**Visualisation des Données** 3](#_Toc189958025)

[5. Résultats et Analyses 3](#_Toc189958026)

[6. Discussion et Commentaire 4](#_Toc189958027)

[7. Conclusion 5](#_Toc189958028)