



État de l'art

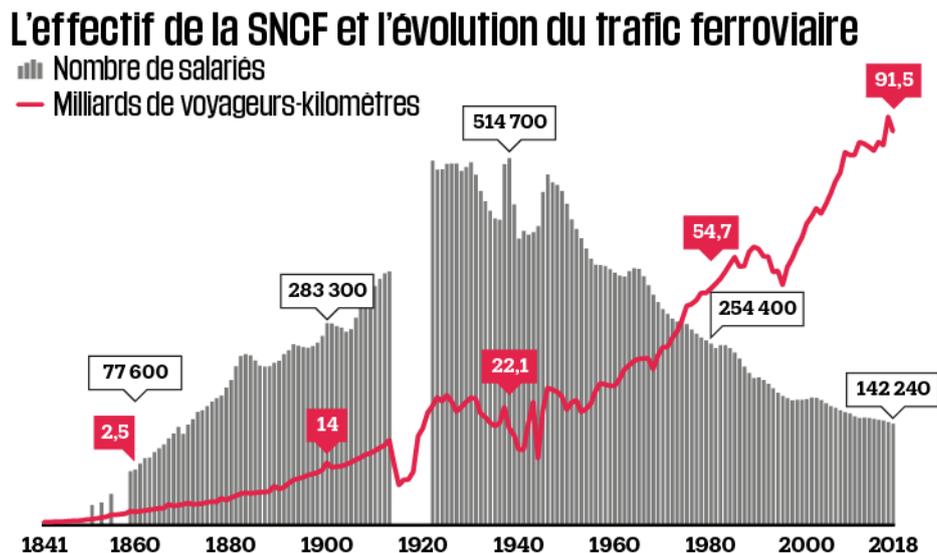
Aspect business :

Aujourd'hui, la SNCF croit en plein essor. Le chiffre d'affaires est d'environ 18 milliards d'euros. La croissance des activités de logistique et de transport de marchandises est soutenue : Geodis (+21%) et le transport ferroviaire de marchandises (+13%) versus le S1 2020. Le transport ferroviaire de voyageurs longue distance a su adapter l'offre à la demande et maintenir un taux d'occupation supérieur à 60%. Une forte demande de mobilité a été enregistrée dès la levée des restrictions en mai grâce à l'adaptation de la politique commerciale. L'activité des métiers conventionnés est bien orientée : Transilien +21%, TER +16% et Keolis +9% versus le chiffre d'affaires du S1 2020. La mise en œuvre d'un plan d'économies ambitieux est poursuivie avec un impact favorable sur la trésorerie du Groupe de près de 830 M€. Aucun renseignement n'a été trouvé concernant l'investissement de la SNCF dans le domaine de la statistique. L'entreprise possède elle-même ses analystes de la donnée.

Nos besoins de mobilités ne cessent de grandir, SNCF Réseau développe le réseau ferré haute performance de demain. À l'heure où le transport de voyageurs s'ouvre à la concurrence, 6,2 milliards d'euros vont ainsi être consacrés en 2020 à la modernisation des quelque 30 000 km de lignes qui composent le réseau français.

À la suite de ces investissements, le but de ce projet est d'avoir une statistique sur les activités principale et connexe de la SNCF (objets trouvés, investissements, fréquentations, temps d'attente dans les points de ventes, nombre de voyageurs, effectifs par métier, ponctualité des trains, répartition du genre et effectif...) et de fournir un tableau de bord récapitulatif des activités qui permettront d'aider à la prise de décision ainsi que s'interroger sur de nouvelles opportunités.

Si l'outil devait être vendu, il serait acheté par la SNCF pour une certaine somme étant donné que ce travail est déjà réalisé par leurs analystes mais pas dans le format prévu et que mon outil est une amélioration d'une analyse déjà faite.

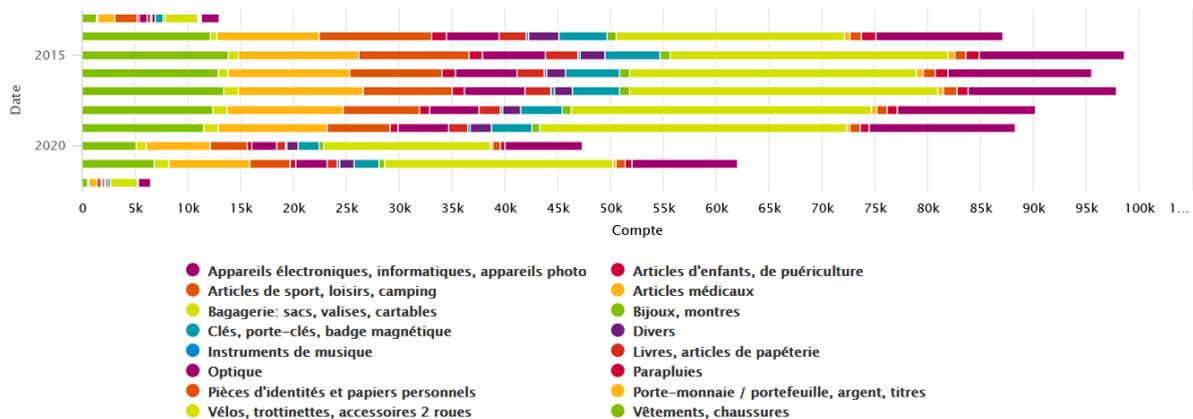


Source : SNCF (avant 1937, effectif agrégé des différentes compagnies ferroviaires)

Aspect de l'art technique :

Le projet a déjà été réalisé par la SNCF (<https://ressources.data.sncf.com>) en prenant toutes les activités. Mon but est de reprendre ses données, m'inspirer des analyses afin d'améliorer leurs visualisations en choisissant des variables pertinentes.

L'outil utilisé par la SNCF est QLIK et Artique MAP. Les graphiques majoritairement présents sont les graphiques en barre ou en carte proportionnelle permettant de mettre en évidence les types d'objets retrouvés par catégorie (Bagagerie, bijoux...) et le nombre de gares par régions.



Le problème de ses graphiques est l'interprétation, on ne peut pas visualiser correctement les objets trouvés ou les régions ayant le plus de gares. Des logiciels permettant de retrouver les objets perdus ont été créés mais mon objectif est d'analyser les statistiques. De mon point de vue, il faudrait fournir de meilleures analyses en gardant le top 5-10 des variables et/ou choisir une visualisation plus facile à comprendre.

Afin de répondre au besoin, je me suis posé plusieurs questions qui me permettront d'établir un rapport sur quelques activités de la SNCF :

- **Quelles sont les gares les plus fréquentées ?**
- **Comment évolue l'investissement des 5 dernières années ?**
- **Quel est l'effectif par métier et par sexe ?**
- **Temps moyen d'attente dans les points de vente ?**
- **Quels sont les objets les plus perdus en gare ?**
- **Quels sont les trains les plus ponctuels ?**
- **Quel est la répartition des gares ?**
- **Combien y'a-t-il de voyageurs transportés ?**

Pour répondre à ses questions, j'ai regardé les données présentes sur le site [open data sncf](#). À la suite de l'analyse de leurs données, j'ai également ajouté d'autres données qui me pourront être utile pour mon rapport. Les données présentes sur leur site sont téléchargeables en plusieurs versions : csv, Json et Excel.

Mon travail réalisé a été de :

- Collecter les données pertinentes
- Nettoyer les données (supprimer les colonnes inutiles, lignes manquantes, ...) sur Excel
- Importer les données sur Power BI
- Réaliser des visualisations pertinentes (choisir les bonnes visualisations)
- Faire des calculs de mesure (moyenne, somme, pourcentage) en codant avec le langage DAX.
- Créer une interface de l'outil.

Après avoir terminé les tâches, j'ai préparé mon rapport ainsi que ma présentation orale.

Modélisation

Expliquer le processus qui va des données décrites dans le cahier des charges et des questions posées et la manière dont vous y répondez :

- J'ai choisi d'extraire les informations qui me seront utile parmi mes 9 fichiers Excel. La difficulté était de réaliser une jointure entre chaque table ayant un lien commun qui était la date. Cette date était différente parmi les différents fichiers (parfois en AAAA, MM/AAAA), j'ai créé un calendrier par un code DAX qui m'a permis de faire une jointure entre chaque fichier.

Énumérer et argumenter tous les indicateurs que vous allez utiliser pour répondre aux questions :

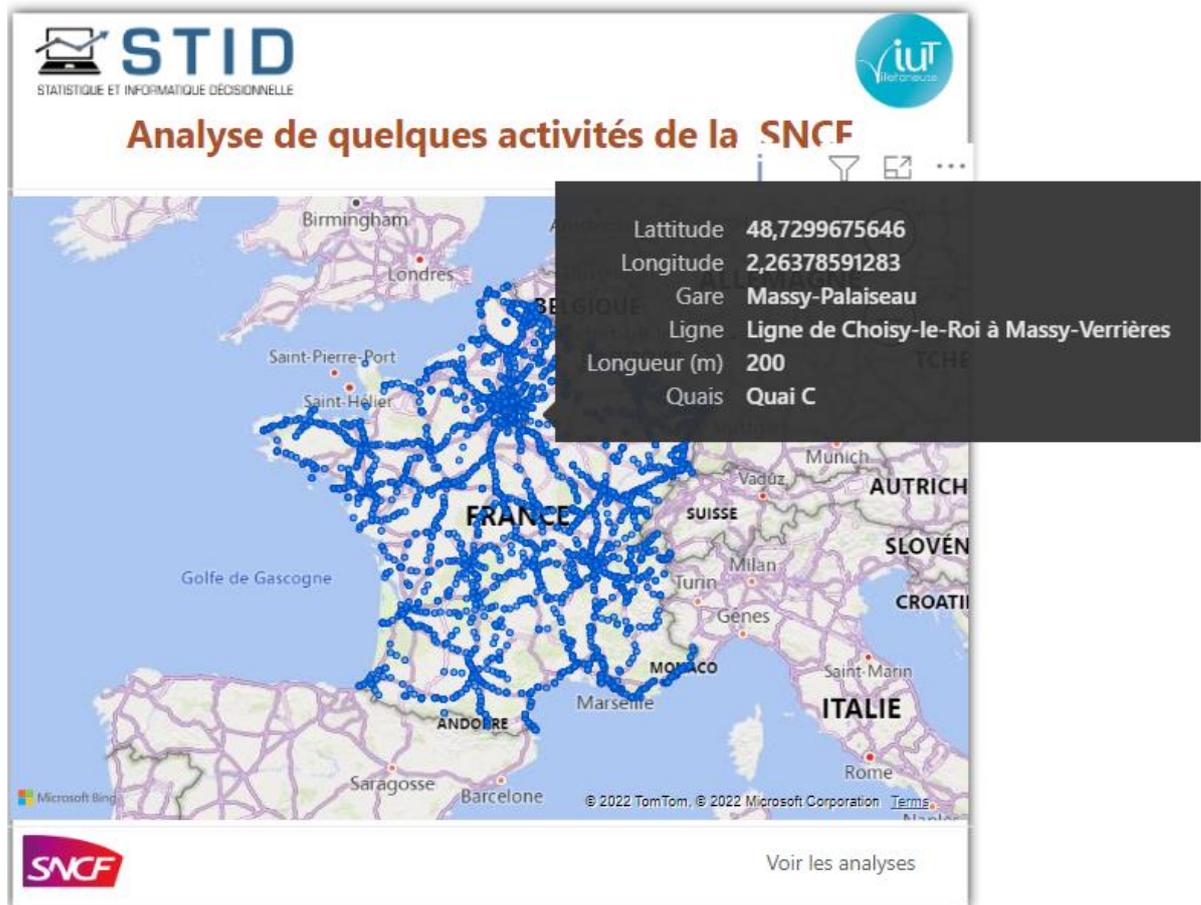
- Les indicateurs utilisés sont : la moyenne, somme ainsi que les statistiques liées aux gares, temps d'attente, objets perdus et l'effectif de la SNCF qui ont été représenté sous formes de visuels qui permettront l'aide à la prise de décision.

Réalisation

L'outil comportera plusieurs sections d'analyse :

- Général : Quelques informations générales tel que l'investissement annuel, répartition des genres et effectif.
- Gare : La répartition d'usage des trains par tranches horaire, le taux de ponctualité des trains d'Île-de-France ainsi qu'une moyenne de ponctualité par mois de 2013 à 2021.
- Temps d'attente : Temps d'attente au guichet, les gares ayant le plus long temps d'attente ainsi qu'une moyenne de temps d'attente par mois de 2019 à 2022.
- Objets perdus : Répartition des objets perdus, gares ayant le plus d'objets perdus ainsi que l'évolution des objets perdus par mois.
- Actions : Cette page réunira quelques mots clés utile à la prise de décision ainsi que quelques solutions apportées pour résoudre les problèmes.

Dès l'ouverture de l'outil, vous retrouverez la localisation des 7000 gares de France ainsi que quelques informations liées à celles-ci.



Dans la rubrique « Général », vous trouverez plusieurs analyses notamment sur l'investissement de la SNCF entre 1997 et 2015. Ce graphique en courbe est le plus pertinent et choisis généralement pour représenter une évolution temporelle.

Pour le réaliser, j'ai sélectionné les données sur l'investissement et choisi le graphique en courbe.

Investissements (milliards d'€) par années

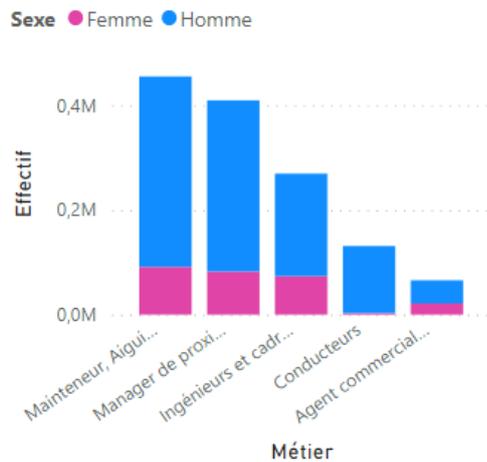


Dans notre cas, on peut voir que l'investissement de l'entreprise connaît quelques fluctuations mais reste en moyenne à 2,45 milliards par an et ne cesse de croître pour améliorer le confort des voyageurs en aménageant les gares et trains.

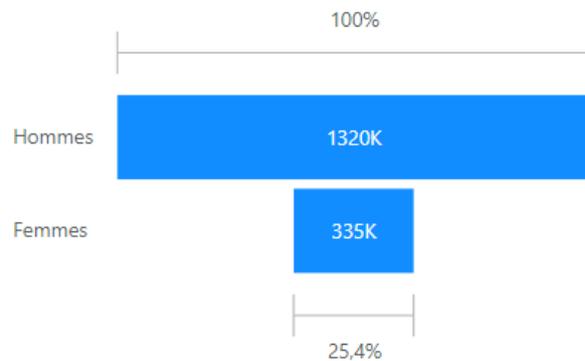
La connaissance du nombre d'employés et sa répartition est aussi importante d'autant plus qu'on observe une minorité de femmes qui représentent seulement 23% parmi tous les métiers et très peu

en tant que conductrice. Cette minorité tente d'être amélioré par la SNCF qui s'engage concrètement pour renforcer la mixité dans tous nos métiers. Par ailleurs, la SNCF a lancé une campagne de recrutement réservée aux femmes. Le choix de l'histogramme empilé et le diagramme en entonnoir m'ont paru pertinent afin de voir rapidement la part qu'occupe chaque sexe.

Effectif par métier et genre

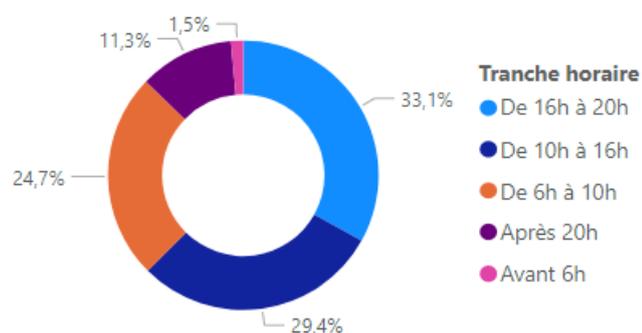


Répartition des genres



Dans la rubrique « Gare », vous retrouverez différentes analyses liées à la gare. La fréquentation par tranche horaire sous forme de graphique en anneau permet de visualiser plusieurs éléments ainsi que leur part par rapport au total. Cette information est importante afin de pouvoir prévoir davantage de trains lors des heures de pointes.

Fréquentation par tranche horaire

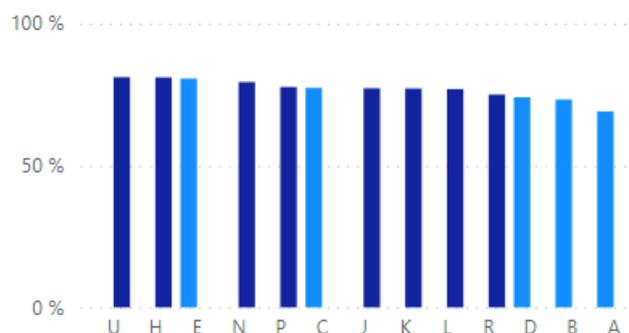


On peut voir que la fréquentation des trains est très faible avant 6h puisqu'il y a peu de trains qui circulent, homogène de 6h à 20h et faible après 20h car les usagers sont majoritairement terminés leur journée.

Un trafic dense nécessite une analyse de la ponctualité des trains afin de pouvoir trouver des solutions.

Taux de ponctualité par ligne

Service ● RER ● Transilien



On observe rapidement grâce à l'histogramme les trains du plus ponctuel au moins ponctuel. Les transiliens sont plus ponctuels que les RER ce qui peut s'expliquer par différentes raisons : en effet, les trains traversant la capitale sont les RER (D, B et A). Ce sont les trains qui passent parmi les gares très fréquentées comme Saint-Denis et Gare du Nord ou bien des routes d'embranchement (Plusieurs terminus). De plus, la population qui l'occupe n'est pas la plus « civilisée » car on retrouve fréquemment des personnes sur les rails, des colis suspects ainsi qu'une difficulté à rentrer dans le train. Ce sont tous ces facteurs qui entraînent le retard et un faible taux de ponctualité pour les RER avec un taux de 69% pour le RER A.

Ainsi, il est nécessaire de connaître les fluctuations du taux de ponctualité par mois.

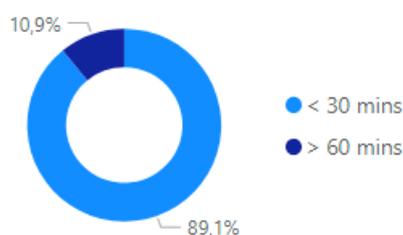
Évolution de la ponctualité par mois



On observe certaines tendances de baisse en début-fin d'année ainsi qu'en milieu d'année notamment liés aux fêtes, grèves ou autres. Mise à part cela, le taux de ponctualité est assez constant à plus de 80% en moyenne.

Dans la rubrique « Temps d'attente », vous trouverez les informations liées à l'attente dans les guichets. Ces informations sont importantes afin d'améliorer le confort de transport des voyageurs et d'améliorer les performances de vente.

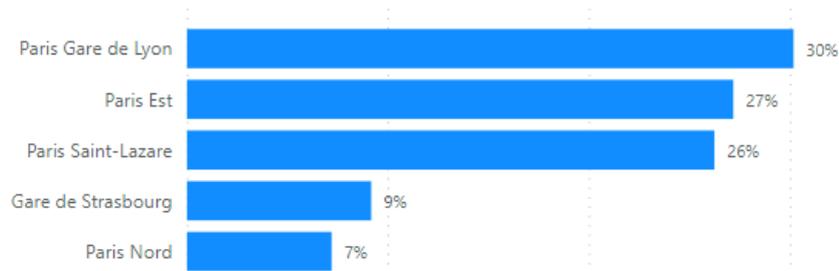
Temps d'attente au guichet



Près de 90% des voyageurs attendent moins de 30 minutes dans un guichet, ce qui est très positif.

On pourrait se demander quelles sont les gares ayant un plus long temps d'attente.

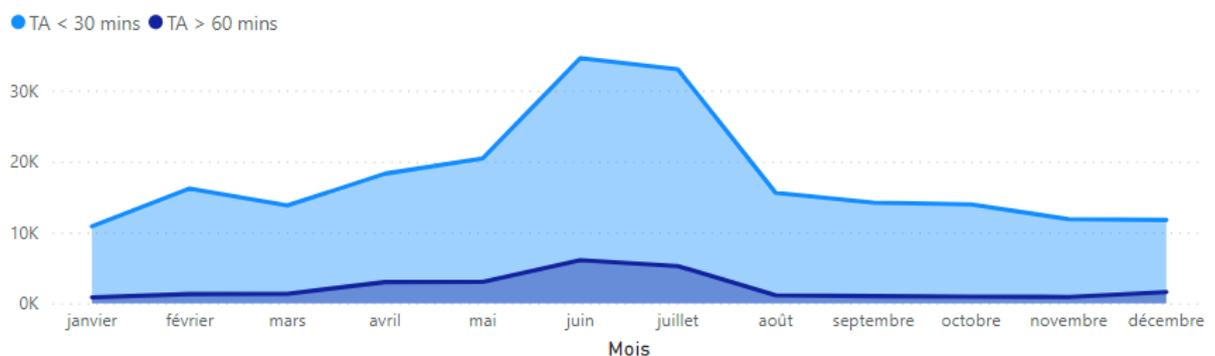
Top 5 gares ayant le plus long temps d'attente



Voici un classement filtré pour avoir les 5 premières gares qui ont le plus long temps d'attente. Comme on pourrait se douter, ce sont les plus grandes gares de France qui ont le temps d'attente le plus long. Grâce à ce graphique, la SNCF peut s'interroger à chercher des solutions pour diminuer ce temps d'attente en créant de nouveaux guichets et plus de personnel.

On peut voir que le temps d'attente est plus important en période estivale, à partir de cela, nous pouvons trouver des solutions pour limiter le nombre de voyageurs en attente en promouvant davantage les applications SNCF pour la réservation de billets.

Évolution du temps d'attente par mois



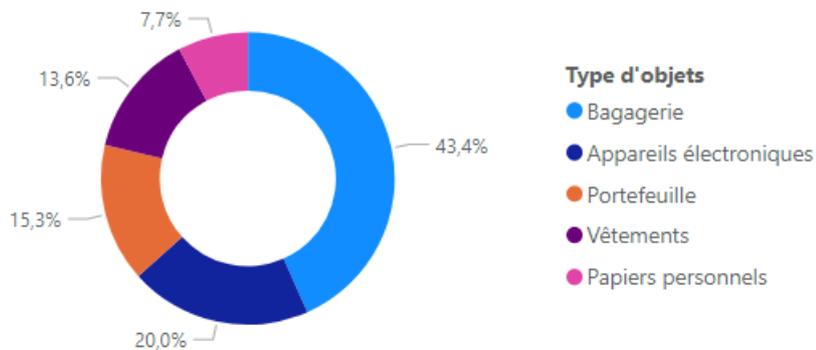
Dans la rubrique « Objets trouvés », vous trouverez l'évolution des objets perdus qui croît en également en période estivale avec une moyenne de 5 000 objets perdus par mois.

Nombre d'objets trouvés par mois



Parmi ces objets, voici une répartition des 5 objets les plus perdus :

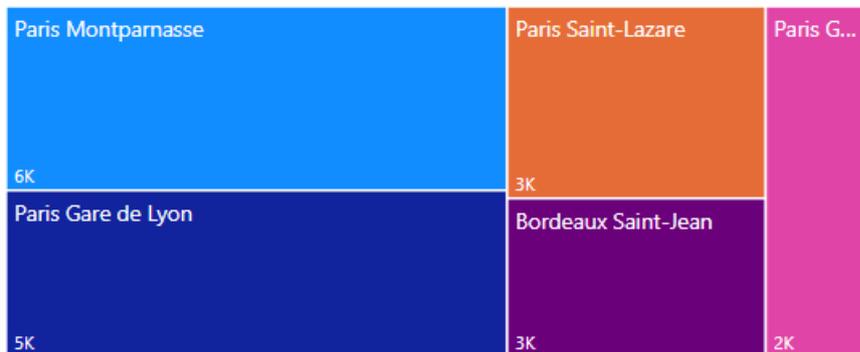
Répartition des types d'objets trouvés



Majoritairement les bagages qui sont considérés comme colis suspect entraînent de grosses perturbations du trafic ferroviaire. La SNCF sensibilise les voyageurs à vérifier derrière eux à la sortie des trains, rappeler cela à chaque avant chaque arrêt de train peut être une solution qui diminuera les pertes.

On retrouve les plus grandes gares avec un fort taux d'oublis.

Top 5 gares ayant le plus d'objets trouvés



Enfin, dans la rubrique « Actions », vous trouverez un récapitulatif des points essentiels de la SNCF ainsi que quelques idées de solutions.



L'effectif majoritairement composé d'hommes dans chaque métier. Cependant le recrutement des femmes est en hausse.



Les RER sont les trains les moins ponctuels notamment les lignes A,B et D. Cela peut s'expliquer au vu des stations déservies et leur population. L'augmentation de la fréquence des trains reste en pleine croissance avec le passage d'un train chaque 5 minutes en moyenne.



Afin de lutter contre la perte d'objets, il faudrait sensibiliser les voyageurs à chaque arrêt par le message " Avant de descendre, assurez-vous de n'avoir rien oublié". Mettre en place des agents qualifiés en gare pouvant résoudre le problème sans que la police intervienne.

Métier	Femme	Homme	Total
Agent commercial à bord des trains	21456	44031	65487
Cadre permanent	19506	40457	59963
Contractuels	1950	3574	5524
Conducteurs	2937	128124	131061
Cadre permanent	2687	122618	125305
Contractuels	250	5506	5756
Ingénieurs et cadres	73526	195776	269302
Cadre permanent	59201	172290	231491
Contractuels	14325	23486	37811
Mainteneur, Aiguilleur, Agent commercial, gestionnaire	91314	363658	454972
Manager de proximité, maîtrise	82377	326936	409313
Total	271610	1058525	1330135

Ligne	Service	Taux ponctualité
U	Transilien	81,14 %
H	Transilien	81,05 %
E	RER	80,64 %
N	Transilien	79,41 %
P	Transilien	77,67 %
C	RER	77,35 %
J	Transilien	77,22 %
K	Transilien	77,14 %
L	Transilien	76,97 %
R	Transilien	75,01 %
D	RER	74,06 %

Codage en DAX

Le langage DAX est un langage de programmation propre à Power BI, il permet de réaliser plusieurs mesures dont une somme, moyenne ou autre.

Par rapport à mes données, programmer sur python ou R n'était pas utile.

Voici quelques code DAX réalisés :

- Un affichage de la période analysée

DateSelected =

```
VAR MinDate = MIN('Fonction appelée'[Date])
```

```
VAR MaxDate = MAX('Fonction appelée'[Date])
```

```
RETURN
```

```
FORMAT (Mi
```

Date d'analyse :



2013 - 2021

- Moyenne de l'investissement annuel

```
Moy_Invest_Années = SUM(Investissement[Investissements SNCF Mobilités (fonds propres)])/COUNT(Investissement[Date])
```

Résultat :

Investissements moyen
(milliards d'€) par an

2,45

- Somme de l'effectif cumulé

```
Effectif = SUM('Genre-effectif'[Nombre d'hommes]) + SUM('Genre-effectif'[Nombre de femmes])
```

Résultat :

Effectif total

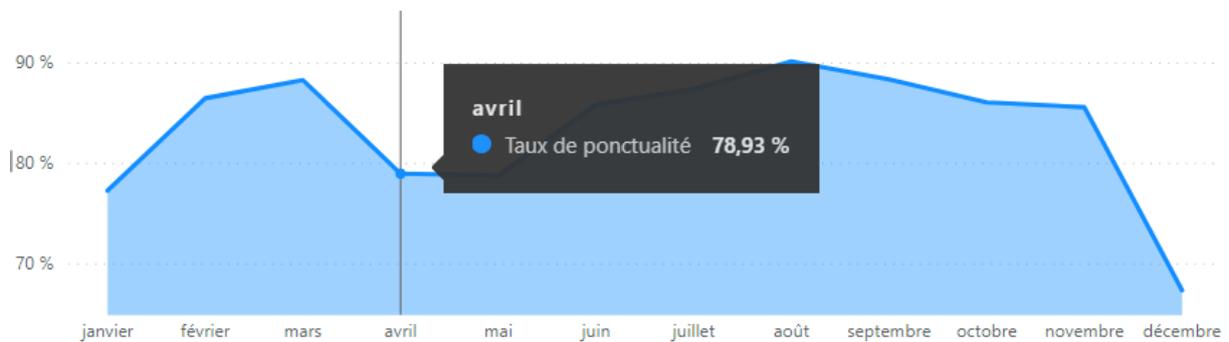
1,65M

- Taux de ponctualité mensuel

Taux ponctualité = ('Ponctualité_Train'[Taux de ponctualité] / SUM('Ponctualité_Train'[Taux de ponctualité])) * 10

Résultat :

Évolution de la ponctualité par mois



Difficultés rencontrées

Comme dit précédemment, la plus grande difficulté était de trouver des données pertinentes et fournir un outil n'ayant pas trop d'informations mais les informations jugées essentielles à une analyse générale. Les tables n'avaient pas de lien entre-elles et j'ai dû quelques fois coder en DAX pour réaliser des calculs plus ou moins complexes.

Conclusion

Selon moi, ce projet a été très intéressant car il nous a permis de mettre en œuvre nos connaissances théoriques et pratiques. L'avantage est que l'on pouvait choisir notre sujet ce qui nous permet d'être davantage investi dans notre travail et le choix de l'outil était libre. Cela m'a permis d'augmenter en compétence dans la collecte/nettoyage de données, l'interprétation et la restitution des résultats oraux et écrits.

Merci d'avoir pris le temps de me lire,

Joyce NANA NJEUTCHOU