

SwissTrain Delay

1 Intention, message à transmettre, public

Le but de notre projet est de proposer aux utilisateurs une visualisation des retards des trains en Suisse par le biais d'une carte interactive. Ce projet est destiné aux usagers du train en Suisse et doit leur permettre d'avoir une vision claire des différents retards afin de mieux planifier leur voyage. Les utilisateurs devront pouvoir naviguer librement sur la carte en filtrant les informations qui les intéressent, que ce soit par rapport à une gare ou à une ligne spécifique.

Ils pourront ainsi visualiser de manière interactive des données comme : le nombre de retards sur la journée, le taux de trains en retard ou le temps de retard le plus élevé. Des informations complémentaires aidant les usagers à planifier leur trajet devront également être accessibles (par exemple : le nom des lignes qui passent par une gare ou encore le nom des gares).

Notre projet vise également à être le plus inclusif possible en optant pour un encodage colorimétrique pertinent pour l'ensemble de la population (prise en compte des différents types de daltonisme).

2 Données

Les données du projet ont été recueillies à partir du site web des CFF qui offre une plateforme Open Data. Sur cette dernière, il est possible de trouver plus de 90 jeux de données allant de la répartition des wifis dans les gares, aux horaires des trains en passant par les emplacements des gares.

Pour ce projet, nous avons utilisé les retards des trains en Suisse. Ce jeu de données est mis à jour tous les quelques jours et contient les retards de la veille. Dans notre cas, il s'agit des retards des trains pour le 13 octobre 2024. Les CFF offrent une grande variété de formats de données téléchargeables (CSV, JSON, GEOJSON, KML, PARQUET, ...). Pour notre projet, nous avons utilisé les données sous le format CSV, car il s'agit d'un format très facile à utiliser pour faire du prétraitement en python.

Nous avons également utilisé un autre jeu de données pour notre projet. Il s'agit des données des tronçons de rails ferrées en Suisse. Elles proviennent de la même plateforme Open Data que le premier jeu de données. Ces données sont utilisées pour ajouter des repères visuels sur la carte de la Suisse.

Dans le premier jeu de données, après les étapes de prétraitement, pour chaque arrivée de train en gare, nous avons les informations suivantes :

- Nom de la gare
- Retard max de la gare
- Nombre total de trains s'arrêtant à cette gare
- Nombre de train ayant eu du retard à cette gare
- Toutes les lignes s'arrêtant à cette gare
- Taux de retard pour la gare
- Position longitudinale de la gare
- Position latitude de la gare
- Nom de la ligne du train

- Nombre total de trains qui ont circulé sur cette ligne
- Nombre de trains ayant eu du retard sur cette ligne
- Taux de retard sur la ligne
- Si le train en question a eu du retard
- L'heure d'arrivée du train

Pour le deuxième jeu de donnée, nous n'avons pas appliqué de prétraitement. Cependant, dans les colonnes disponibles, il y a la colonne "Ligne" qui est également présente dans le premier jeu de donnée. Cependant, même si les deux colonnes portent le même nom, l'information qui y est présente est différente. Nous n'avons pas trouvé le moindre lien entre les deux. Par conséquent, nous n'avons pas pu faire de filtre sur les rails des trains.

3 Représentation

3.1 Carte



Le premier graphique est une carte où les différentes gares et rails de trains de la Suisse sont visibles. Les gares sont représentées par des cercles dont la taille dépend du nombre de trains s'arrêtant à cette dernière et dont la couleur dépend du pourcentage de trains en retard. La taille du cercle a pour but de permettre à l'utilisateur d'identifier rapidement les grandes gares de la région qu'il observe et non d'identifier avec précision le nombre de trains s'y arrêtant. La couleur du cercle permet également à l'utilisateur de rapidement identifier les gares avec le plus de retard. Elles ont été choisies pour être intuitives, le vert représente les gares avec le moins de retard et le rose celles avec le plus de retard.

Les couleurs choisies sont pensées pour ne pas poser de problèmes aux daltoniens. Ce tableau montre comment sont perçues les couleurs choisies par chaque type de daltonisme :

Normal	Achromatopsie	Monochromatisme à cônes bleus
Protanomalie	Deutéranomalie	Tritanomalie

		
Protanopie	Deutranopie	Tritanopie
		

Le seul type de daltonisme qui ne peut pas correctement différencier les couleurs utilisées est l'achromatopsie qui est définie par une absence totale de vision des couleurs. Il aurait fallu utiliser des intensités différentes pour leur permettre de différencier les différents pourcentages, cependant, il n'est pas possible de réaliser cette modification sur Carto.

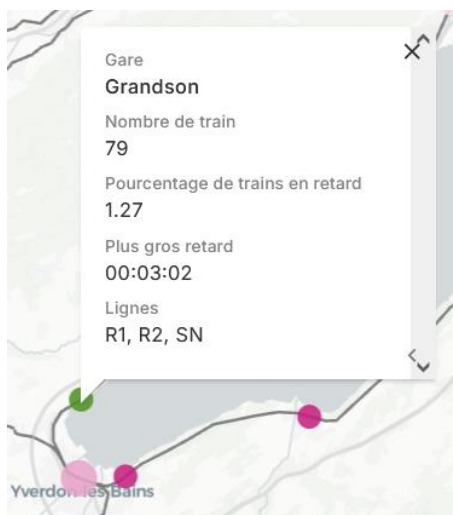
3.1.1 Interactions disponibles sur la carte

Passer la souris sur une gare :



Passer la souris sur une gare permet d'afficher le nom de la gare. Cela permet à l'utilisateur de rapidement identifier une gare.

Cliquer sur une gare :



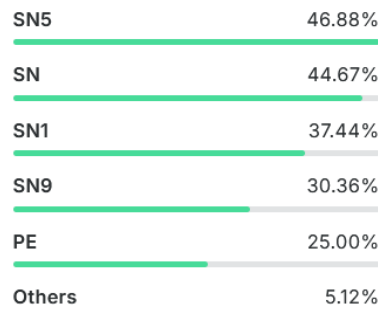
Cliquer sur une gare fournit des informations détaillées sur cette dernière. L'utilisateur peut directement y voir le nom de la gare et son pourcentage de trains en retard. Un autre point important est l'indication de toutes les lignes s'arrêtant à cette gare. En effet, un usager ne connaît pas forcément le nom de la ligne qu'il emprunte. Cependant, il connaît généralement le nom de la gare où il doit monter ou descendre, cette fonctionnalité lui permet donc d'identifier facilement le nom de la ligne. S'il ne connaît pas la localisation de la gare, il peut utiliser la barre de recherche pour la retrouver (fonctionnalité par défaut proposée par Carto).

3.2 Diagrammes à barres horizontales

Il y a deux graphiques en barres :

Pourcentage de trains en retard
par ligne

All

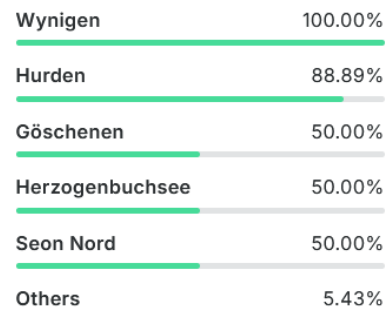


[Search in 94 elements](#)

Un train est considéré en retard s'il arrive avec au moins 3 minutes de retard

Pourcentage de trains en retard
par gare

All



[Search in 581 elements](#)

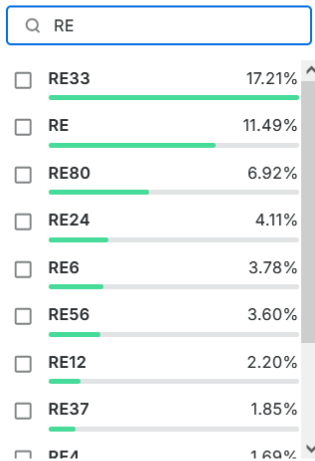
Un train est considéré en retard s'il arrive avec au moins 3 minutes de retard

Les deux graphiques sont très similaires et sont donc traités en même temps. Ces graphiques affichent le pourcentage de retard par ligne (respectivement par gare) trié par ordre décroissant de la valeur. Ils permettent à l'utilisateur d'identifier les lignes/gares où il y a le plus haut taux de retards.

Pourcentage de trains en retard
par ligne

All

[Apply](#)



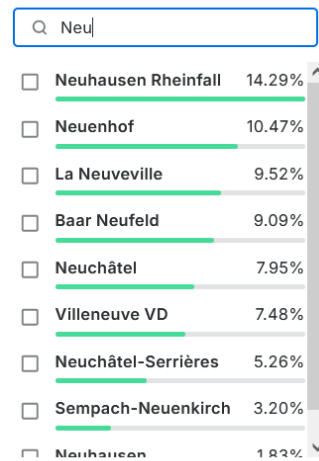
[Cancel](#)

Un train est considéré en retard s'il arrive avec au moins 3 minutes de retard

Pourcentage de trains en retard
par gare

All

[Apply](#)



[Cancel](#)

Un train est considéré en retard s'il arrive avec au moins 3 minutes de retard

Une option « Search » permet de chercher un élément spécifique.

3.3 Graphique en courbes



Ce graphique permet de comparer le nombre de trains en retard et le nombre de train à l'heure en fonction de l'heure de la journée. Sur l'abscisse, la valeur 14 indique le changement de jour (les données allant du 13 octobre au début du 14 octobre). L'objectif est de permettre à l'utilisateur de pouvoir déterminer les heures à éviter et les heures où les trains circulent le mieux.

Passer sa souris sur la courbe permet de voir le nombre de trains exact sur une plage de temps précise.

4 Présentation et interaction

Les différents graphiques, séparément, ne permettent pas à un utilisateur d'obtenir des informations spécifiques sur le trajet qu'il prend. Cependant, ces différents graphiques, utilisés ensemble permettent de visualiser des informations très précises.

Les deux graphiques en barres servent de filtre. Il est possible de sélectionner une ligne/gare spécifique pour que le graphique en courbe affiche le nombre de trains en retard et le nombre de trains à l'heure pour cette sélection spécifique. Également sur la carte, seuls les éléments sélectionnés sont affichés.

Par exemple, si un utilisateur est intéressé par la ligne RE33 (Annemasse – Genève – Martigny), il peut sélectionner la ligne via le widget « Pourcentage de trains en retard par ligne »



Cette sélection provoque les modifications suivantes :

- Sur la carte, seuls les gares où la ligne passe sont affichées. Cela permet à l'utilisateur de facilement visualiser le parcours réalisé par la ligne sélectionnée.
- Sur le graphique en barres « Pourcentage de trains en retard par gare », seuls les gares présentent sur la ligne R33 sont affichées.
- Sur le graphique en courbes, seuls les trains de la ligne R33 sont comptés pour le nombre de trains en retard et le nombre de trains à l'heure. Dans cet exemple, nous voyons que durant la soirée et il y a eu des moments où il y avait plus de trains en retard que de trains à l'heure. L'utilisateur peut donc identifier cette période comme une période à éviter au maximum.

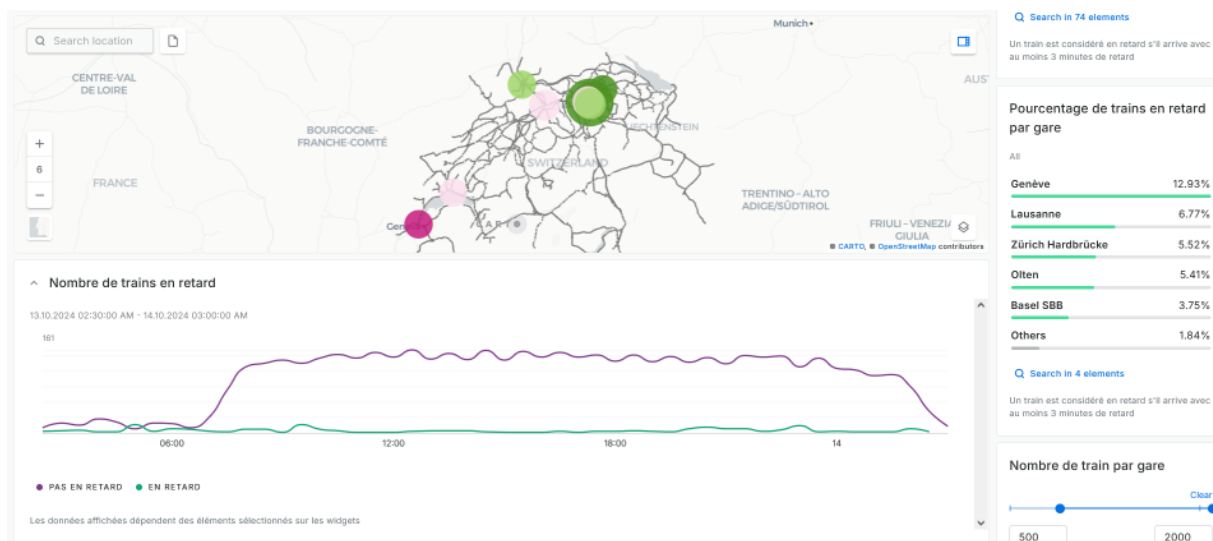
Si c'est une gare qui est sélectionnée, sur la carte, il ne restera que la gare en question affichée. Le graphe en courbe affichera les retards uniquement pour cette gare. Également, il est possible de sélectionner plusieurs lignes et plusieurs gares à la fois, les éléments sélectionnés seront alors affichés sur les différents graphiques.

Nombre de train par gare



Une autre option disponible est le filtre des gares par le nombre de trains. Cela permet à l'utilisateur de n'afficher qu'une partie des gares en fonction du nombre de trains y passent.

Voici un exemple où seules les gares avec plus de 500 trains par jour sont affichées.



Le comportement est similaire à celui où une ligne est sélectionnée :

- Sur la carte, seul les gares avec au moins 500 trains sont affichés.
- Sur le graphique en barres « Pourcentage de trains en retard par gare », seuls les gares respectant ce critère sont affichées.
- Sur le graphique en barres « Pourcentage de trains en retard par ligne », seuls les lignes passant par au moins une gare respectant ce critère sont affichées.

- Sur le graphique en courbes, seuls les trains passant par une gare où au moins 500 trains passent sont comptés pour le nombre de trains en retard et le nombre de trains à l'heure.

Les différents graphiques s'adaptent en fonction des éléments sélectionnés sur les autres graphiques.

Il y a d'autres éléments avec lesquels il est possible d'interagir, mais il s'agit de fonctionnalités de base de Carto, comme la possibilité de changer de carte de fond (mettre un fond noir, passer en image satellite ou encore afficher le terrain) ou télécharger un rapport de l'état de la carte (l'état des divers graphiques en fonction d'une sélection).

5 Outils

5.1 Python

Les données exploitées pour ce projet ont dû être prétraitées en Python avant de pouvoir être utilisées par le logiciel Carto. En effet, pour pouvoir pleinement exploiter Carto, des modifications du jeu de données sont nécessaires. Cela nous permet de sélectionner et d'ajouter des informations que nous souhaitons afficher dans cette plateforme. Par exemple, les données fournies par les CFF sont relatives à chaque train, et nous avons dû créer un script qui regroupe et traite les informations par gare et par ligne. Permettant ainsi d'avoir pour chaque gare ou ligne le nombre de retard, le taux de retard, le nombre de trains, etc.

5.2 Carto

Le principal outil utilisé dans ce projet est Carto. Il s'agit d'une plateforme de cartographie basée sur le cloud. Il permet d'analyser, de visualiser et de développer en utilisant des données spatiales. L'outil supporte diverses sources de données allant du simple CSV à la base de données externes. Il permet par conséquent la création de carte interactive et leur déploiement.

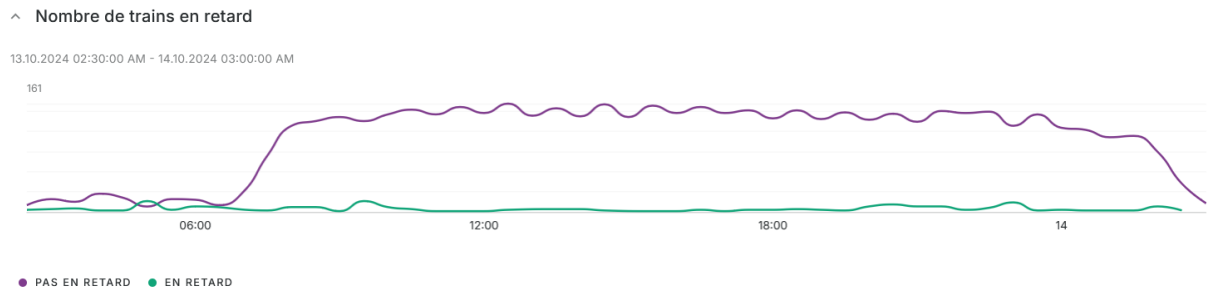
Son utilisation est normalement payante, cependant avec GitHub Student, il est possible de l'utiliser gratuitement. Toutes les fonctionnalités ne sont pas disponibles avec le statut d'étudiant, mais celles pertinentes pour le développement de ce projet sont utilisables.

Il a également été nécessaire de réaliser un gros pré-traitement des données pour qu'elles soient utilisées comme voulu sur la carte. Carto dispose d'un outil workflow pour réaliser le prétraitement, mais nous avons préféré réaliser ce dernier en python, car nous y sommes plus à l'aise.

6 Limitations liées aux choix techniques

6.1 Limitations liées à Carto

Carto est un outil très puissant, mais nous avons rencontré quelques limitations.



Sur ce graphique, il ne nous a pas été possible de modifier les couleurs des 2 courbes. Il serait plus logique que les trains « Pas en retard » soient de couleur verte et les trains « En retard » soient de couleur rouge. Un autre problème est les labels sur l'axe des x, il y a un « 14 » vers la droite, il s'agit du moment où l'on passe du 13 au 14 octobre. Cependant, dans notre cas, cela n'est pas intuitif et il serait plus pertinent d'afficher l'heure « 00:00 ».

Q Search in 74 elements

Les différents éléments textuels comme l'option de recherche provenant de Carto sont obligatoirement en anglais, il n'est pas possible de les mettre en français.

Nous voulions également réaliser une interaction lorsque l'utilisateur clique sur un élément de la carte. C'est-à-dire que nous souhaitons que lorsque l'utilisateur appuie sur une gare, les différents graphiques se mettent à jour en conséquent. Cependant, aucune option dans Carto ne permet de réaliser cela.

Il n'est également pas possible d'ajouter un widget contenant uniquement du texte. Nous souhaitons ajouter un widget avec des informations diverses sur les données. En raison de cette impossibilité, les détails sont présents dans ce rapport et dans le readme.

6.2 Limitations liées aux jeux de données.

Le dernier point que nous n'avons pas pu réaliser est lié aux jeux de données. Nous voulions que quand l'utilisateur sélectionne une ligne, seuls les rails où passe cette ligne soit affichées. Cependant, comme nous n'avons pas pu faire de correspondance entre le jeu de données des retards et celui des rails, cette interaction est impossible. Pour la même raison, nous n'avons pas pu faire en sorte que quand l'utilisateur clique sur un rail, il puisse voir les lignes qui y passent.

7 Conclusion

Ce projet permet à un utilisateur de pouvoir visualiser efficacement et rapidement les retards de trains en Suisse. Pour cela, nous avons créé une carte interactive à l'aide de la plateforme Carto en utilisant les données fournies par les CFF que nous avons prétraitées avec Python.

L'utilisation de ses divers outils a amené à quelques contraintes qui nous ont empêchés de réaliser tout ce que nous souhaitons faire. Cependant, les éléments principaux que nous voulions implémenter ont pu être réalisés.

Pour conclure, malgré les diverses contraintes, nous sommes satisfaits du résultat, car nous avons atteint notre objectif de fournir une carte pertinente pour l'analyse des retards des trains. Nous sommes également contents d'avoir pu découvrir un nouvel outil : Carto.

8 Références

Color blind simulator : [Color blindness](#)

Carto : [Carto](#)

Données retard CFF : [CFF OpenData - Retard Train](#)

Données du réseau ferroviaire des CFF : [CFF OpenData - Réseau ferroviaire](#)