Prédiction des retards de TGV

Le but de ce projet est de prédire le temps de retard des TGV à l’arrivée ainsi que la raison de leur retard le cas échéant. Nous nous focaliserons en particulier sur les trains réservés par la SNCF.

# Obtenir les données

Les données pour réaliser ce projet sont accessibles grâce aux étapes suivantes :

* Rendez-vous sur cette [page](https://www.data.gouv.fr/fr/datasets/regularite-mensuelle-tgv-par-liaisons/).
* Allez tout en bas de la page.
* Téléchargez le fichier au format **csv**.
* Vous pouvez à présent utiliser ces données grâce à la librairie Python **Pandas**. Vous trouverez à ce [lien](https://data36.com/pandas-tutorial-1-basics-reading-data-files-dataframes-data-selection/) un petit tutoriel pour vous aider à manipuler cette librairie.

Afin de vous aider pour la visualisation des données avant de les traiter sous Python, vous pouvez vous rendre sur cette [page](https://ressources.data.sncf.com/explore/dataset/regularite-mensuelle-tgv-aqst/analyze/?sort=date&dataChart=eyJxdWVyaWVzIjpbeyJjaGFydHMiOlt7InR5cGUiOiJzcGxpbmUiLCJmdW5jIjoiQVZHIiwieUF4aXMiOiJwcmN0X2NhdXNlX2V4dGVybmUiLCJzY2llbnRpZmljRGlzcGxheSI6dHJ1ZSwiY29sb3IiOiJyYW5nZS1TZXQxIn1dLCJ4QXhpcyI6ImRhdGUiLCJtYXhwb2ludHMiOiIiLCJ0aW1lc2NhbGUiOiJtb250aCIsInNvcnQiOiIiLCJzZXJpZXNCcmVha2Rvd24iOiJzZXJ2aWNlIiwiY29uZmlnIjp7ImRhdGFzZXQiOiJyZWd1bGFyaXRlLW1lbnN1ZWxsZS10Z3YtYXFzdCIsIm9wdGlvbnMiOnsic29ydCI6ImRhdGUifX19XSwiZGlzcGxheUxlZ2VuZCI6dHJ1ZSwiYWxpZ25Nb250aCI6dHJ1ZSwidGltZXNjYWxlIjoiIn0%3D) et tracer toutes sortes de graphes.

# Objectifs du projet

Les objectifs du projet sont les suivants :

* Appliquer différentes approches vues en cours et en TD sur ce dataset pour prédire le temps de retard des trains à l’arrivée sur les 6 derniers mois (tous les mois de 2023) ainsi que les principales causes de ces retards (c’est-à-dire sur les différents pourcentages relevés dans le dataset). Le dataset de test correspondra donc aux mois de 2023, tandis que celui d’entraînement sera les données sur les années précédentes.
* Comparer les performances de ces différents modèles grâce aux métriques vues en cours.
* Conclure sur l’approche la plus appropriée pour prédire le retard des trains et les différentes causes.
* Écrire un rapport d’environ 5 pages sur le travail effectué et les résultats obtenus.