

Tous documents papier autorisé. Lisez les questions attentivement et répondez de manière argumentée. Essayez d'être concis mais précis. Vous pouvez rédiger en français ou en anglais. En général, il n'y a pas **une bonne réponse unique**. Vous pouvez utiliser des schémas ou des équations si vous le souhaitez.

Vous êtes responsables d'une mission pour la SNCF, et vous devez aider l'entreprise à améliorer ses performances en terme de ponctualité. Dans la suite, nous considérerons toujours que les données sont déjà nettoyées (pas d'outliers/données aberrantes, de données erronées ou manquantes)

1. (4 points) Vous disposez de toutes les données sur les lignes ferroviaires existantes, la fréquence prévu des trains, le nombre de passagers, etc. Vous voudriez commencer par utiliser une approche réseau pour identifier les trajets ou les stations les plus sensibles, c'est à dire celles qui, si jamais elles sont bloquées (par exemple par une alerte colis piégé), affecteront le plus l'ensemble du réseau. Quel réseau proposez-vous de construire, et quelles méthodes pensez-vous pouvoir utiliser pour cette tâche?

2. (4 points) Pour chaque retard ou autre incident, vous disposez d'une fiche la décrivant, contenant plusieurs informations: durée de retard (en minutes), lieu de l'incident (coordonnées géographiques), nombre de passagers affecté, cause du retard (choix parmi 10 raisons possibles), coût pour l'entreprise (en euros). A partir de toutes ces informations, nous voudrions créer une typologie d'incidents, par exemple représenté par 4 ou 5 cas types. Que proposez-vous pour faire cela ? Y compris pré-traitement sur les données. .

3. (2 points) A partir des données de la question précédente, nous avons créé une matrice dont les lignes sont des gares, les colonnes des heures de la semaine (lundi matin 12h=12, mardi matin 12h=36, etc.), et les valeurs correspondent au nombre de retard à ce moment là dans cette gare là. Que proposez-vous de faire pour extraire une information utile de ces données ?