

Exercice de crunching de data

Page 1 sur 3

Nous ne nous attendons pas à un résultat parfait car l'exercice peut être infini, le but est de tester tes capacités d'apprentissage et ta motivation.

Plus tu vas loin, mieux ce sera !

Tu trouveras dans le dossier *data*, 4 fichiers CSV

- ***ticket_data.csv*** : Contenant un historique de ticket (une ligne => une proposition de ticket sur tictactrip)

- ***cities.csv*** les villes desservies par tictactrip (lien grâce aux colonnes o_city (origin_city), d_city (destination_city) de ticket_data)

- ***stations.csv*** les stations desservies par tictactrip (lien via o_station, d_station de ticket_data)

- ***providers.csv*** infos sur les différents providers (lien via company de ticket_data)

Un provider est une "sous-compagnie". Par exemple TGV et TER sont deux providers de VSC (voyages-sncf).

Ta mission (si tu l'acceptes) :

- extraire les infos intéressantes type prix min, moyen et max, durée min/max/moyenne par trajet
 - différence de prix moyen et durée selon le train, le bus et le covoit selon la distance du trajet (0-200km, 201-800km, 800-2000km, 2000+km)
 - le plus d'infos bonus
-

Le rendu devra être :

- Tes scripts
- et (pas nécessairement) un mini-rapport expliquant tes recherches et ta démarche.

Cela peut se présenter sous la forme d'un **jupyter notebook** publié sur Github / Gitlab ou une autre plateforme en ligne.

Langage requis : **Python**, fortement conseillé de l'accompagner du package **Pandas**
Prends le temps qu'il te faut. Si tu as des questions n'hésites pas :-)

NB : Nous ferons particulièrement attention à la qualité et à la rigueur du code.
Des points bonus sont attribués à tout exercice faisant **plus que le strict minimum**.

Comme par exemple :

Graphes, prédictions de prix, rapport des soucis relevés dans les données, visualisation interactive, sourcing & utilisation de données externes pertinentes

Bon courage !!

Pour ta réponse : envoie un lien vers le repo github à **jobs@tictactrip.fr**

ATTENTION : l'objet doit être **DATA@NOM
PRENOM**
