# Petit ewxo de crunching de data

# 

# Je ne m’attends pas à un résultat parfait ou complet (car l’exercice peut être infini), le but est de tester tes capacités d’apprentissage et ta motivation. 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 des 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 rendu devra être tes scripts et (pas nécessairement) un mini-rapport m’expliquant tes recherches et ta démarche. Cela peut se présenter sous la forme d’un **jupyter notebook** Si tu ne sais pas comment démarrer je te conseille **Python** avec le package **Pandas** (et la méthode *pandas.read\_csv*) Prends le temps qu’il te faut. Si tu as des questions n’hésites pas :-)

# Bon courage !!

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

# ATTENTION : l’objet doit être **DATA@NOM PRENOM**