



# Crunching de data

## Introduction

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 !

## Objectifs

- 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

Par exemple (0-200km, 201-800km, 800-2000km, 2000+km)

- Le plus d'infos bonus !

Comme par exemple :

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

Langage requis : **Python** + le package **Pandas**

## Les données

stations.csv 516.5KB

providers.csv 15.9KB

cities.csv 634.4KB

ticket\_data.csv 8712.3KB

- **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).

## Le rendu

- Tes scripts
- 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.

### Bonus

- Des points bonus sont attribués à tout exercice faisant plus que le strict minimum.
- Mettre une petite star sur nos repo <https://github.com/tictactrip> ;) (et les lire)

Nous ferons particulièrement attention à la **qualité** et à la **rigueur du code**.

Bon courage !!

Tu veux te faire remarquer ? Et surtout mieux nous connaître ?

- Achète ton prochain trajet chez nous et fait nous tes

retours d'utilisation !

- Et si le service t'a plus laissé nous un avis 5 sur Google ou Truspilot

**Pour ta réponse : envoie un lien vers le repo  
github à [bot+jobs@tictactrip.eu](mailto:bot+jobs@tictactrip.eu)**

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