

Big Data Open Data

MMI 2 – TP#5 S4

Danielo **JEAN-LOUIS**
Développeur front-end

Numpy

- Bibliothèque externe à Python
- Optimisé pour les tableaux jusqu'à que votre machine ne puisse
- ~50x plus rapide que Python pour la gestion des tableaux
- Intégrée par défaut dans Anaconda / Google Colab → pas besoin de pip

pip

- Signifie Pip Installs Python
- Gestionnaire de bibliothèques externes

Numpy

- Rajoute de nouvelles fonctionnalités aux tableaux en Python
- Possibilité d'utiliser certaines syntaxes de Python sur les tableaux numpy
- Spécialisé pour gérer les données non labellisées. De préférence des chiffres.

Pratiquons ! - Numpy

Pré-requis :

- Avoir le notebook `ressources/numpy/numpy.ipynb`

A télécharger ici :

<https://downgit.github.io/#/home?url=https://github.com/DanYellow/cours/tree/main/big-data-s4/travaux-pratiques/numero-5/ressources/numpy>

Pandas

- Bibliothèque Python
- Contraction de Panel Data
 - ou Python Data Analysis
- Intègre des fonctionnalités de numpy

Pandas

- Permet de gérer les données labellisées (dites tabulaires)
 - Super-excel en programmation
- Gère sans problème de très gros fichiers
 - Plus d'un million de lignes
- Intégrée par défaut dans Anaconda / Google Colab → pas besoin de pip

Pratiquons ! - Pandas

Pré-requis :

- Avoir le notebook `ressources/pandas/initiation-pandas-partie-1.ipynb`

A télécharger ici :

<https://downgit.github.io/#/home?url=https://github.com/DanYellow/cours/tree/main/big-data-s4/travaux-pratiques/numero-5/ressources/pandas/pratique>

Questions ?

