

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 bibliothèque externe

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`

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

Pandas

- Contraction de Panel Data
- Intègre des fonctionnalités de numpy
- Permet de gérer les données labellisées (dites tabulaires)
 - Super-excel en programmation
- Intégrée par défaut dans Anaconda / Google Colab → pas besoin de pip

Pratiquons ! - pandas

Pré-requis :

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

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

Questions ?

modele	marque	gamme	vitesse_max	nbre_sieges	longueur
AMG	BMW	luxe	210	7	3000
XC60	Volvo	haut de gamme	160	5	2750
Ka	Ford	compacte	120	4	2000
Cactus	Citroen	citadine	130	5	2500