

# Python Data Science

Nicolas Rousset

Formation Data Science, day 1

# Qu'est-ce que Python?

La langue la plus populaire selon l'index Tiobe !



# Comment est-ce utilisé ?

- Machine learning / AI
- Tensorflow - pytorch
- Analyse / visualisation des données
- Web (pas un acteur principal)
- Scripts

# Caractéristiques de python

- Créé en 1991 par Guido Van Rossum
- Interpreted langage using compiled library
- Langage interprété utilisant des librairies compilées
- Typage dynamique
- Open source
- multi-paradigme

# Avantage comparatifs de python

Par rapport à Excel, SAS, BI, MATLAB et d'autres outils couramment utilisés pour l'analyse et les rapports des données, Python a plusieurs avantages : - gratuit (parfois surfait) - Langage de programmation de niveau industriel - générique

D'un autre côté, il sera souvent battu par d'autres outils sur leurs fonctionnalités spécialisées

Le programme n'est pas directement exécuté par le système d'exploitation, mais par un interpréteur

```
python hello_world.py
```

Et pas :

```
./hello_world.py
```

# Python est extraordinairement permissif

Vous pouvez tout modifier!

```
import math  
math.pi = "bonjour"
```

Python essaiera toujours de faire ce que vous lui dites de faire, pas de ce que vous voulez. Comme le dit l'adage, cette machine n'a pas de cerveau, utilisez le vôtre

# Quand vous allez en production, le TDD est obligatoire

TDD => Test Driven Development

Les tests sont très importants en Python (et autre langage interprétés)



# Python est multi-paradigme

POO (programmation orientée objet) n'est pas le paradigme de programmation le plus courant dans Python. Les objets ne sont pas les plus faciles à tester, et la POO va de pair avec la compilation

OOP dans Python en un mot => Pourquoi n'utilisez pas C #, Java ou C ++?

À des fins d'analyse des données, le moyen le plus simple est d'utiliser une distribution, comme Anaconda:

<https://www.anaconda.com/products/distribution>

La distribution contient Python et toutes les bibliothèques scientifiques les plus couramment utilisées

# Lancement de Jupyter notebook

Après avoir installé Anaconda, la façon de lancer le serveur Jupyter est notebook :

```
jupyter notebook
```

Il ouvrira un navigateur avec les fichiers locaux

# Spécification de modules Python

L'outil utilisé pour gérer le package dans Python est pip (python install package) :

```
pip install tensorflow  
pip install -r requirements.txt
```

Lorsque vous avez plusieurs dépendances de modules, vous devez les mettre dans un fichier **requirements.txt**, qui ressemble à ceci:

```
setuptools==45  
asgiref==3.2.10  
attrs==19.3.0  
autobahn==20.12.3  
Automat==0.8.0  
bokeh==2.0.2
```

# Spécification de modules Python (2)

La commande pour fixer (*freezer*) la liste des modules est :

```
pip freeze > requirements.txt
```

Cela est essentiel pour la reproductibilité