

Programme de remise à niveau sur le thème « Statistiques sous Python ».

Objectifs : Savoir charger et manipuler les données sous Python (package Pandas). Initiation aux principales commandes pour réaliser des statistiques descriptives simples.

Distrib. Python : Installez la distribution Anaconda Python (<https://www.anaconda.com/download/> ; **version 3.x**).

Outil : Nous utiliserons JUPYTER NOTEBOOK qui est automatiquement installé avec ANACONDA. Sous Windows, vous le lancerez en insérant la commande JUPYTER NOTEBOOK dans la zone de recherche en bas à gauche de la barre des tâches (ou à partir d'une invite de commande MS-DOS - <https://jupyter.readthedocs.io/en/latest/running.html>) (sous Mac, je ne sais pas mais ça ne doit pas être bien compliqué).

1. Supports à lire

- Distribution Anaconda

<http://tutoriels-data-mining.blogspot.fr/2015/08/python-la-distribution-anaconda.html>

- Programmation sous Python (*pour ceux qui veulent aller plus loin*)

Cours en L3 (Licence) : https://eric.univ-lyon2.fr/~ricco/cours/cours_programmation_python.html

Page Python sur Developpez.com : <https://python.developpez.com/>

2. Mode de travail

Pour chaque exercice, créez un notebook (NEW / PYTHON 3 ; fichier **.ipynb**) que vous renommez de manière à l'identifier clairement. Alternez les zones (CELLS) de code et de résultats, y compris graphiques. Numérotez les questions pour que je puisse me repérer facilement. Rajouter des zones spécifiques pour les éventuels commentaires (CELL TYPE « MARKDOWN »).

A l'issue de l'exercice, le plus simple est d'imprimer en PDF votre travail. Ne le verrouillez pas pour que je puisse l'éditer et y insérer mes commentaires. Vous devez me rendre les fichiers **.ipynb** et **.pdf**

3. Exercices à réaliser

Fichiers à traiter :

1. « Produits.xlsx »
2. « Census.xlsx »
3. « Iris.txt »