**INTRODUCTION AU LANGAGE DE PROGRAMMATION PYTHON POUR LA SCIENCE DES DONNEES**

1. **POUR COMMENCER**
   1. **A propos de python**

Python est un langage de programmation interprété, multi-paradigme et multiplateformes créé par le programmeur Guido van Rossum en 1989. Il est conçu pour optimiser la productivité des programmeurs en offrant des outils de haut niveau et une syntaxe simple à utiliser.

En Octobre 2020, ce langage est à sa version 3.9 qui sera supporté jusqu’en 2025.

Le présent document vous apprend les bases de ce langage, partant de zéro jusqu’à atteindre un niveau adéquate pour aborder la science des données.

* 1. **Structure du document**

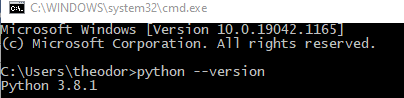
Ce document se subdivise en dix chapitres : les huit premiers sont consacrés aux notions de base de Python alors que les deux derniers chapitres vont vous introduire dans le monde de la science des données.

* 1. **Les outils de développement sous Python**
* **Python**

Python est l’interpréteur du programme écrit en Python ; il est installé dans tout ordinateur destiné à exécuter un programme Python ; de ce fait, il est l’intermédiaire d’une part entre le programmeur et sa machine lors de la conception du programme, et d’autre part, il permet à l’utilisation du programme sur la machine de l’utilisateur.

Vérifier si Python est installé à l’aide de la commande ci-dessous :

python –version



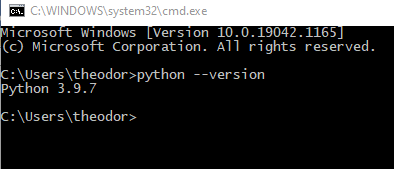
Comme vous pouvez le remarquer, la commande affiche Python 3.8.1 : Nous allons donc installer la version 3.9.7

Vous pouvez télécharger Python sur le site suivant :

* **Pour Windows** : <https://www.python.org/ftp/python/3.9.7/python-3.9.7-amd64.exe>
* **Pour Linux** : <https://www.python.org/ftp/python/3.9.7/Python-3.9.7.tgz>
* **Pour macOs** : <https://www.python.org/ftp/python/3.9.7/python-3.9.7-macosx10.9.pkg>

Plus d’information sur le site internet[**https://www.python.org**](https://www.python.org)

Après installation, nous allons vérifier à l’aide de la même commande que ci-haut :



Cette fois-ci la version 3.9.7 de Python est belle est bien installée.

Si vous n’arrivez pas à voir le message ci-dessus, c’est que quelque chose n’a pas bien marché. Peut-être qu’il vous faudra reprendre dès le début en suivant chacune des étapes de manière minutieuse ; au cas où vous n’arrivez toujours pas à voir un message du genre, vous devriez peut-être chercher de l’aide sur internet.

* **Visual Studio Code**

Visual Studio Code est un éditeur de code très extensible développé par Microsoft pour Windows, Linux et macOS. Ce logiciel nous permettra d’écrire les programme Python.

Vous pouvez télécharger Visual Studio Code sur les liens suivants :

* **Pour windows**

<https://code.visualstudio.com/sha/download?build=stable&os=win32-x64-user>

* **Pour Linux**

<https://az764295.vo.msecnd.net/insider/8d30d4d922dda917d22f05a81f2ab4fad7349779/code-insiders_1.61.0-1632475234_amd64.deb>

* **Pour macOS**

<https://code.visualstudio.com/sha/download?build=insider&os=darwin-universal>

Plus d’information sur le site internet<https://code.visualstudio.com>

* **Jupiter Notebooks (Anaconda)**

Jupyter Notebook propose une interface de programmation interactive permettant la combinaison des sections écrites en langage naturel et des sections écrites en langage informatique.

Pour installer Jupiter Notebooks, il faut d’abord installer Python afin de procéder via la commande ci-après :

pip install jupyterlab

* **Les extensions python pour visual studio code**

Voici quelques extensions utiles à installer afin de mieux pratiquer Python avec Visual Studio Code :

* **Extension Python :** <https://marketplace.visualstudio.com/items?itemName=ms-python.python>
* **Extension Jupyter :** <https://marketplace.visualstudio.com/items?itemName=ms-toolsai.jupyter>

Maintenant que les outils sont installés, nous pouvons écrire notre premier programme en Python.

1. **Premiers pas avec Python**

* **Hello world**
* **Les commentaires**
* **Les variables**
* **Capture de saisie utilisateur**

1. **Types des variables**

* **Types de base**
* **Conversion des types**
* **Le type dictionnaire**
* **Les listes, les intervalles, les tuples, et les ensembles**

1. **Les opérateurs arithmétiques et logiques**

* **Les opérateurs arithmétiques**
* **Priorités des opérateurs**
* **Les opérateurs logiques**
* **Les comparaisons**

1. **Les fonctions**

* **Définition d’une fonction Python**
* **L’instruction return**
* **Les paramètres d’une fonction**
* **Les paramètres avec valeurs initiales**
* **Les paramètres magiques**
* **Les fonctions mathématiques**
* **Les fonctions des chaines de caractères**
* **Les fonctions des listes et tuples**
* **Les fonctions des dictionnaires**

1. **Les chaines de caractères**

* **Introduction**
* **Construction des chaines des caractères complexes**
* **Quelques fonctions utiles aux chaines de caractères**

1. **Les boucles et tests décisionnels**

* **IF**
* **La boucle FOR**
* **La boucle WHILE**
* **Les instructions break et continue**

**Chap 8. Programmation orientée objets**

* **Les classes et les objets**
* **Les méthodes**
* **Les paramètres**
* **L’héritage**
* **L’encapsulation**

1. **Autres notions importantes**

* **La gestion des exceptions**
* **Les modules**
* **Les packages**

1. **Bibliothèques importantes en science des données**

* **Numpy**

Numpy est une bibliothèque python très populaire, qui fournit des nombreuses fonctions pour manipuler les listes

* **Pandas**

Pandas est une bibliothèque qui permet la manipulation aisée des données grâce au concept de Data Frame.

* **MathPlotLib**
* **Scikit-Learn**

1. **Machine Learning**

* **Importation des données**
* **Nettoyage des données**
* **Séparation des données d’entrainement et données des tests**
* **Création du modèle**
* **Entrainer le modèle**
* **Prédictions**
* **Evaluation des prédictions**

Le but de l’évaluation est d’améliorer le score