



Traitement et analyse de données dans Python

Référence et durée

PYTHDM

2 jours

Objectif

Savoir utiliser Python pour accéder, traiter, analyser des données.

Public

Des statisticiens utilisant un langage tel que SAS ou R pour préparer et analyser les données. Des informaticiens maîtrisant des langages de script et désirant s'initier au traitement de la donnée.

Prérequis

Connaissances de base en programmation informatique.

Programme

- Utiliser Python comme un logiciel de data mining et data science
 - Python 2 ou Python 3 ?
 - Les différentes distributions, Focus sur la distribution Anaconda
 - Environnement de travail Spyder et IPython Notebook
 - Gestion des librairies
 - Principales librairies pour la data science (Scipy, Numpy, Pandas, Scikit-Learn, StatsModel...)
- La programmation
 - Le langage Python (syntaxe, types, opérateurs)
 - Les structures de données dans Python (dict, list, tuple, json, ...)
 - Les structures de contrôle (if then else, for, while ...)
 - Création de fonctions, Entrées / Sorties
 - Exercices orientés programmation
- La manipulation de données avec Scipy, Numpy et Pandas
 - Accès aux données (fichiers, bases de données ...), export
 - Manipulation de données avec pandas (via les objets Series et DataFrame)
 - Librairies d'Analyse et d'exploration : Matplotlib
 - Expressions régulières pour traiter les chaînes de caractères
 - Exercices : lire des données externes, connexion à une base SQL, préparation de données (création de variables, jointures, filtres, union ...), paramétrage et automatisation