



Python avancé pour data-scientists

LY010

Durée: 4 jours

3 230 €

28 au 31 janvier

14 au 17 avril

30 juin au 3 juillet

29 septembre au 2 octobre

15 au 18 décembre

Public :

développeurs en Python, Développeurs de logiciels, programmeurs, Data analysts, Data scientists.

Objectifs :

Savoir utiliser les principaux outils de traitement et d'analyse de données pour Python, savoir extraire des données d'un fichier et les manipuler, mettre en place un modèle d'apprentissage simple

Connaissances préalables nécessaires :

Bonne connaissance de la programmation Python.

Objectifs pédagogiques :

Savoir utiliser les principaux outils de traitement et d'analyse de données pour Python

Savoir appliquer les pratiques optimales en matière de nettoyage et de préparation des données avant l'analyse

Être capable d'extraire des données d'un fichier

Comprendre les mécanismes d'interconnexion aux bases de données

Comprendre les principaux outils de traitement et d'analyse de données pour Python

Programme :

Positionnement Python dans l'analyse de données

Besoins des data-scientists : calculs, analyse d'images, machine learning, interface avec les bases de données

Apports de python : grande variété d'outils, expertise dans le domaine du calcul scientifique

Tour d'horizon des outils:

pandas, pyarrow, agate, bokeh, scikit-learn, pybrain, tensorflow, keras, mxnet, caffe, Pytorch

Calculs et graphiques

NumPy : Base du calcul sur des tableaux

SciPy : Scientific Tools for Python, couche scientifique

Manipulation de tableaux, fonctions mathématiques.

Représentation graphique avec basemap et matplotlib.

Atelier : Mise en oeuvre de SciPy/NumPy : manipulation d'images, détection de contours



Être capable d'extraire des données d'un fichier

Pandas : manipulation de tables de données. Notion de dataframe.
Manipulation de données relationnelles
Tableaux avec Pandas: indexation, opérations, algèbre relationnelle
Stockage dans des fichiers: CSV, JSON
Comparaison et performances Pandas / pyarrow / NumPy

Atelier : construction d'ETL de base entre json et csvkaggl.com,

Comprendre les mécanismes d'interconnexion aux bases de données

Définitions : pilotes, connexions, curseurs, CRUD, transactions
Les pilotes : postgresql, mysql, mariadb, ... Présentation de sql-alchemy
Opérations : gestion du curseur, chargement de données, insertion et modification d'enregistrements

Atelier : mise en oeuvre avec postgresql. Construction d'ETL SQL/json

Comprendre les principaux outils de traitement et d'analyse de données pour Python

Présentation des outils d'apprentissage Python : scikit-learn, pybrain, TensorFlow/keras, mxnet, caffe

Atelier : mise en oeuvre de scikit-learn

Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances

Présentation de pyspark
Machine learning et deep learning : les solutions Python,
TensorFlow : principe de fonctionnement, plateformes supportées, distribution

Sites de références data-sciences

Ressources d'apprentissage, datasets, modèles de données pré-entraînés, etc ..
Présentation de : kaggle.com, data-puzzles.com, huggingface.co

Optimisation des développements

Tour d'horizon des outils actuels et futurs:
Jupyter notebook, Aide à la vérification de code, respect des recommandations PEP8 :
exemples avec pydocstyle, Pylint, Black
Analyse et production de code informatique avec une IA.
Génération de code avec OpenAI : démonstrations ChatGPT, apports, bonnes pratiques.

Atelier : utilisation de la génération de code et de snippets Python avec ChatGPT