

# Python avancé pour data-scientists

## INFORMATIONS GÉNÉRALES

**Type de formation :** Formation continue

**Éligible au CPF :** Non

**Domaine :** IA, Big Data et Bases de données

**Action collective :** Non

**Filière :** Big Data

**Rubrique :** Fondamentaux

**Code de formation :** BD060

## € Tarifs

**Prix public :** 2190 €

### Tarif & financement :

Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :

- **Le plan de développement des compétences** de votre entreprise : rapprochez-vous de votre service RH.
- **Le dispositif FNE-Formation.**
- **L'OPCO** (opérateurs de compétences) de votre entreprise.
- **Pôle Emploi** sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.
- **CPF -MonCompteFormation**

Contactez nous pour plus d'information

## PRÉSENTATION

### Objectifs & compétences

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.

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

### Public visé

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

### Pré-requis

Maîtrise de la programmation Python.

## 📍 Lieux & Horaires

**Campus :** Ensemble des sites

**Durée :** 21 heures

### Délai d'accès :

Jusqu'à 8 jours avant le début de la formation

**Distanciel possible :** Oui

## 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

### 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 csv

### Comprendre les mécanismes d'interconnexion aux bases de données

Définitions : pilotes, connexions, curseurs, CRUD, transactions

## 📅 Prochaines sessions

Cliquez sur la date choisie pour vous inscrire :

### ■ 18 / 11 / 2024

📍 : Ensemble des sites

✓ : Distanciel possible

🕒 : 21 heures

📅 : 3 jours

### ■ 16 / 12 / 2024

📍 : Ensemble des sites

✓ : Distanciel possible

🕒 : 21 heures

📅 : 3 jours

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 et génération de jeux de données.

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

Présentation de pyspark

Machine learning et deep learning

TensorFlow : principe de fonctionnement, plateformes supportées, distribution,

## **MODALITÉS**

### **Modalités**

**Modalités :** en présentiel, distanciel ou mixte – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise

**Pédagogie :** essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques.

**Ressources techniques et pédagogiques :** Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom

**Pendant la formation :** mises en situation, autodiagnostic, travail individuel ou en sous-groupe sur des cas réels

### **Méthode**

**Fin de formation :** entretien individuel

**Satisfaction des participants :** questionnaire de satisfaction réalisé en fin de formation

**Assiduité :** certificat de réalisation (validation des acquis)