



ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

Formation Python pour la science des données *Python avancé pour data-scientists*

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Ce cours de 3 jours apprend aux participants à programmer en Python pour la data science, depuis les bases du langage jusqu'aux bibliothèques essentielles pour la data science.

A travers de nombreux exercices pratiques ils prendront en main les environnements de développement Python, la distribution Anaconda, l'environnement de travail collaboratif JupyterLab ainsi que les bibliothèques Python majeures telles que NumPy, Pandas, Matplotlib, Seaborn et Scikit-Learn pour faire de la data science.

Référence	PYDATA
Durée	3 jours (21h)
Tarif	1 800 €HT
Certification	90 €HT

Objectifs

- | utiliser Python en standalone et à travers JupyterLab.
- | acquérir des données à partir des formats de fichiers standards (ex: CSV) ou par web scraping.
- | mettre en oeuvre des analyses de données exploratoires quantitatives avec Numpy et Pandas.
- | mettre en oeuvre des analyses de données exploratoires visuelles avec Matplotlib, Pandas, Seaborn, Bokeh,...
- | résoudre des problèmes réels d'analyse de données à grande échelle.

Public

- | Consultants, développeurs, chefs de projet, data scientists, data engineers

Prérequis

- | Connaissances de base en programmation (logique, structures de données) et mathématiques (fonctions, vecteurs, matrices)
- | Des bases en Python seront un plus mais ne sont pas requises, les concepts fondamentaux du langage étant présentés la première journée

Programme de la formation

Introduction à l'analyse de données avec Python

- | Pourquoi Python pour l'analyse de données ?
- | Introduction au langage Python: bref historique, versions, outils.
- | Les bibliothèques Python essentielles pour l'analyse de données.
- | Notebooks interactifs collaboratifs: JupyterLab, Kaggle Kernels, Google Colab.
- | Ateliers: Installer un environnement Python complet pour l'analyse des données, type Anaconda.
- | Prendre en main de l'environnement de développement JupyterLab.

Les bases de Python pour la manipulation de données

- | Premier programme Python.
- | Organiser et écrire son code, PEPs.
- | Expressions, variables et types.
- | Chaînes de caractère et texte.
- | Interagir avec les utilisateurs: inputs et affichage formaté.
- | Structures de données: listes, ensembles, dictionnaires, tuples.
- | Logique et boucles.

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 17 au 19 novembre 2025

[VOIR TOUTES LES DATES](#)

- | Fonctions.
- | Travailler avec des fichiers.
- | Itérateurs et Générateurs.
- | Atelier: Explorer des jeux de données dans des fichiers tenant ou non en mémoire.

Au delà du prototypage dans les notebooks

- | Bases de POO en Python (classes, objets).
- | Modulariser son code.
- | IDE et debuggers.
- | Ateliers: Refactoring du code d'exploration de données avec une approche modulaire et POO.
- | Développer une application web basique avec Flask permettant de partager les résultats de l'exploration des données.

Les bases de Numpy: tableaux et calculs vectoriels

- | Représentation des tableaux avec Numpy.
- | Fonction universelles et accès rapide aux éléments.
- | Programmation orientée tableaux.
- | Tableaux et fichiers.
- | Algèbre linéaire.
- | Génération de nombres pseudo aléatoire.
- | SciPy.
- | Ateliers: Implémenter des fonctions avec Numpy et Traiter des données avec Numpy et SciPy.

Les bases de Pandas

- | Les structures de données Pandas.
- | Fonctionnalités essentielles.
- | Statistiques descriptives.
- | Atelier: Mettre en oeuvre des analyses de données exploratoires quantitatives.

Gestion des fichiers de données sous Pandas

- | Les différents formats de données.
- | Bonnes pratiques en terme de manipulation de gros fichiers.
- | Stratégies pour analyser les données à l'échelle.
- | Atelier: Explorer des jeux de données tenant ou non en mémoire.

Visualisation des données en Python

- | Visualisation avec Matplotlib.
- | Visualisation avec Pandas et Seaborn.
- | Autres outils de visualisation: bokeh, plotly,...
- | Atelier: Mettre en oeuvre des analyses de données exploratoires visuelles.

Pandas avancé

- | Nettoyer et préparer les données.
- | Joindre, combiner et reshaper des données.
- | Agréger et grouper les données.
- | Séries de temps.
- | Atelier: Manipulation avancée des fichiers .CSV avec Pandas.

Web scraping

- | Rappels sur le fonctionnement du web et des sites web.
- | Qu'est-ce que le web crawling et le web scraping ? Est-ce légal, faut-il se conformer ? Quels sont les outils communément utilisés ?
- | Télécharger des pages web et effectuer des requêtes HTTP.
- | Web scraping avec BeautifulSoup.
- | Atelier: Scraping web avec Request et BeautifulSoup.

Introduction à Scikit-learn

- | Pré-traitement des données: chargement et transformations.
- | Analyse et prédiction sur les données avec les algorithmes de machine learning: régression, classification, clustering, réduction de dimensionnalité.
- | Atelier: Prise en main de Scikit-learn: pré-processing et premier modèle ML.

Certification

Cette formation prépare au passage de la certification suivante.
N'hésitez pas à nous contacter pour toute information complémentaire.

TOSA Python

Programmer et automatiser des tâches avec Python (Tosa) L'examen TOSA se passe en centre agréé ou en ligne. Votre score (sur 1000) vous permettra de communiquer sur votre niveau de connaissances du langage.

- | Durée : 1h30
- | 35 questions QCM et cas pratiques
- | Score minimum : 551
- | Version 3.10
- | Date d'enregistrement de la certification : 19/12/2024

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

Accessibilité



Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.