Big Data - Python pour l'analyse de données

Tirer meilleur parti des bases de données grâce à Python

Présentiel

4 jours (28 h) | Prix inter : 3.350,00 € HT | Réf.: BD555

Python est un choix très populaire pour l'analyse de données en raison de sa simplicité d'apprentissage, de son écosystème de bibliothèques robustes (pandas, NumPy, Matplotlib, etc.), de sa polyvalence, de son intégration facile avec d'autres langages, de son statut open source, de sa large adoption industrielle et de sa communauté active.
Ces avantages permettent aux analystes de données d'effectuer efficacement la manipulation, la visualisation et la modélisation de données, ainsi que de mettre en oeuvre des techniques d'apprentissage automatique, ce qui en fait un choix incontournable dans le domaine de l'analyse de données.

A qui s'adresse cette formation ?

- Développeurs en Python
- Développeurs de logiciels
- Programmeurs
- Data analysts
- Data scientists

- Maîtrise de la programmation Python

Programme

Pendant

1 - Positionnement Python

- Besoins des data-scientist : 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, agate, bokeh, scikit-learn, pybrain, tensorflow, keras, mxnet, caffe

2 - 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
- Mise en oeuvre de SciPy/NumPy : manipulation d'images, détection de contours

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

3 - Être capable d'extraire des données d'un fichier

- Pandas : manipulation de tables de données
- Tableaux avec Pandas : indexation, opérations, algèbre relationnelle
- Stockage dans des fichiers : CSV, h5py, netCDF

- Comparaison et performances Pandas / NumPy

  Construction d'ETL de base entre json et csvkagglt.com,

4 - 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

  Mise en oeuvre avec postgresql. Construction d'ETL SQL/json

5 - 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

  Mise en oeuvre de scikit-learn

6 - 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

7 - Sites de références data-sciences

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

8 - 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 pydecodestyle, Pylint, Black
- Analyse et production de code informatique avec une IA
- Génération de code avec OpenAI : démonstrations ChatGPT, apports, bonnes pratiques.

  Utilisation de la génération de code et de snippets Python avec ChatGPT



Les objectifs de la formation

- Savoir utiliser les principaux outils de traitement et d'analyse de données pour Python
- Être capable d'extraire des données d'un fichier et les manipuler
- Apprendre à mettre en place un modèle d'apprentissage simple
- Comprendre les mécanismes d'interconnexion aux bases de données
- Comprendre les principaux outils de traitement et d'analyse de données pour Python



Evaluation

Pendant la formation, le formateur évalue la progression pédagogique des participants via des QCM, des mises en situation et des travaux pratiques. Les participants passent un test de positionnement avant et après la formation pour valider leurs compétences acquises.

Les points forts de la formation

- Une formation très pratique : les participants seront amenés à réaliser de nombreux ateliers qui leur permettront d'acquérir une première expérience pratique de Python pour analyser des données
- Le partage de bonnes pratiques sur l'utilisation de Python
- 83% des participants à cette formation se sont déclarés satisfaits ou très satisfaits au cours des 12 derniers mois.

Evaluation

Pendant la formation, le formateur évalue la progression pédagogique des participants via des QCM, des mises en situation et des travaux pratiques. Les participants passent un test de positionnement avant et après la formation pour valider leurs compétences acquises.

Dates et villes 2027 - Référence BD555

Dernières places disponibles Session garantie

Paris

du 22 févr. au 25 févr.	du 21 juin au 24 juin	du 29 nov. au 2 déc.
du 26 avr. au 29 avr.	du 27 sept. au 30 sept.	