

Python pour la Data Science

Formation Informatique / Bureautique / Sécurité



La Data Science nécessite des compétences spécifiques afin de prévoir le comportement des clients ou d'anticiper des tendances. Cette formation vous permettra de découvrir la Data Science via Python, les différentes étapes qui interviennent lors de l'élaboration d'un modèle statistique, ainsi que l'analyse textuelle et la Data Visualisation.

OBJECTIFS

- Savoir modéliser un problème de Data Science en Python
- Découvrir les principales bibliothèques de calcul numérique dont Numpy, Pandas et Matplotlib

PUBLIC

Développeurs, Data Analystes, Data Scientists, et toute personne souhaitant acquérir des connaissances dans le domaine de la Data Science et Python

PRE-REQUIS

- Une première expérience en programmation Python Avoir des connaissances de base en mathématiques et statistiques

PROGRAMME

Introduction à la Data Science

- Qu'est-ce que la Data Science ?
- Définition
- Domaines d'application de la Data Science
- Outils et algorithmes pour la Data Science
- De l'analyse statistique au Machine Learning
- Enjeux de la Data Science

Le langage Python

- Qu'est-ce que Python ?
- Shell, IDE, NoteBook
- Les bases de Python
- Structures de données
- Modules & fonctions
- Gestion des fichiers

Collecte et préparation des données avec Numpy et Pandas

- Importation et stockage de données
- Nettoyage et filtrage des données (valeurs manquantes, erronées...)
- Transformation des données
- Jointure & combinaison
- Agrégation & Grouping

Exercices

Data Analysis – Modeling avec Scikit-Learn

- Introduction à Scikit-Learn
- Étapes d'élaboration d'un modèle
- Use case
- Apprentissage supervisé vs apprentissage non supervisé
- Evaluation du modèle

Exercices

Data Visualisation avec Matplotlib

- Principe de la Data Visualisation
- Matplotlib : Scatter Plot, Box Plot, histogrammes
- Visualisation interactive

Exercices

Introduction à l'utilisation de Spark avec Python

- Présentation de Spark
- Présentation de la bibliothèque PySpark
- Les différents types d'analyses avec PySpark (SQL, Streaming, Machine Learning, Graphix)

Horaires, Planning et Déroulement de cette formation



A retenir

Durée : **2 jours** soit 14h.
Réf. **GKBDPYT**

☎ 01 42 93 52 72

Dates des sessions

Cette formation est également proposée en formule **INTRA-ENTREPRISE.**



Inclus dans cette formation



Coaching Après-COURS

Pendant 30 jours, votre formateur sera disponible pour vous aider. CERTyou s'engage dans la réalisation de vos objectifs.

100%
SATISFACTION
GARANTIE

Votre garantie 100% SATISFACTION

Notre engagement 100% satisfaction vous garantit la plus grande qualité de formation.

Horaires

- Formation de 9h00 (9h30 le premier jour) à 17h30.
- Deux pauses de 15 minutes le matin et l'après-midi.
- 1 heure de pause déjeuner

DEROULEMENT

- Les horaires de fin de journée sont adaptés en fonction des horaires des trains ou des avions des différents participants.
- Une attestation de suivi de formation vous sera remise en fin de formation.
- Cette formation est organisée pour un maximum de 14 participants.

PROCHAINES FORMATIONS

[Réussir la Certification Gestion de Projet PMP du PMI](#)

[Réussir la Certification PRINCE2 Foundation](#)

[Réussir les Certifications PRINCE2 Foundation et PRINCE2 Practitioner](#)

[Réussir la Certification ITIL Foundation](#)

[Réussir la Certification Agile certifié SCRUM Master](#)

[Réussir les Certifications TOGAF Certified et TOGAF Foundation](#)

Retrouvez cette formation sur notre site :

[Python pour la Data Science](#)