Contactez-nous Tél. : +33 (0)1 83 35 34 40
Email inscription@lepont-learning.com
Site web www.lepont-learning.com

PARCOURS DATA

Data Avant Garde

REF : DDC-DAG

DUREE : 30h

Présentiel Classe virtuelle

PUBLIC

Cette formation s'adresse aux analystes qui souhaitent travailler sur des questions métiers requérant l'analyse de jeux de données volumineux et complexes.

Modalités et délais d'accés : les inscriptions sont fernées 24h avant la 1ére journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Pour suivre cette formation il est nécessaire de maîtriser un outil de reporting et connaître SQL et des techniques avancées de capture, préparation, analyse et visualisation.

MODALITES PEDAGOGIQUES

1 poste par stagiaire

Remise d'un support pédagogique (numérique ou papier) pendant la formation

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

****

Formation distancielle, synchrone, avec le formateur en visioconférence

Partage d'écran formateur et stagiaire pour les présentations, démonstrations et vérification des mises en application

Un support technique accompagne la préparation et le suivi du parcours

OBJECTIFS PEDAGOGIQUES

Maîtriser les techniques de traitement de données avancé.
Connaître les clés du langage Python et du module Pandas appliqués à la data et les méthodes de préparation efficace.
Savoir comment appliquer algorithmes et visualisation avec pertinence.
Acquérir confiance, autonomie et performance pour être en mesure de s'attaquer sans hésitation et avec pragmatisme aux challenges métiers complexes.
Développer des processus analytiques avancés efficaces, élégants et durables et créer des passerelles avec la data science.

PROGRAMME

1. La Capture de la donnée
Accéder à tout type de sources
• Les fichiers plats
• Les bases de données
• Les données non structurées, les pages web et documents
Focus sur le "data scraping"
• Récupérer des données sur les sites d'information
• Les techniques avancées de scraping de documents PDF
2. Le Profilage des données
Les étapes du profilage
• L'examen des types et de la structure des données
• L'analyse de la cohérence des information
• Les rapports de synthèse
Les outils du profilage
• Les astuces pythons
• Les commande Pandas
• Les conseils de Data Officers
3. La Préparation des données
La préparation des données
• Eliminer le superflux ou l'inutilisable
• Traiter les doublons ou les manquants
• Ajuster les formats et les types de données
• Nettoyer les chaines de caractères
• Inférer les valeurs manquantes
4. Faire parler la data
• Filtrer vos données
• Hierarchiser
• Grouper et Calculer
• Joindre
5. La visualisation des données
Les techniques de visualisation
• Pourquoi visualiser ?
• Programmer simplement ses visualisations
Les grandes familles de visualisation
• Les visualisations simples: avantages et limites
• Les visualisations avancées
• La visualisation des données géographiques

d'apprentissage

Séquençage pédagogique construit et adapté au mode d'organisation distanciel

Remise d'une documentation pédagogique numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

****

Évaluation du formateur au fil des séquences pédagogiques.

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

Les bonnes pratiques

6. Les règles d'Or du Data Wrangling

- Le code vient en dernier
- Toujours s'inspirer des bibliothèques existantes
- Coder le moins possible
- Commenter et respecter le formalisme
- Suivre le process de la donnée: charger-transformer-visualiser
- Ecrire du code que l'on peut comprendre

7. Pour aller plus loin

Des concepts complémentaires

- Listes
- Tuples
- Sets
- Dictionnaires
- Lire et écrire du texte

Les librairies pythons

8. L'application du Machine Learning

Les grandes familles d'algorithmes

- Régressions
- Classifications
- Clustering

L'utilisation des librairies

9. Les Responsabilités de l'Avant garde

La collaboration

- Data Science
- Informatique
- Métiers

La transparence

- Documenter
- Expliquer
- Former

Version du : 02/12/2021