

DATA ANALYST

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 : NE142

€ Tarifs

Prix public : 5100 €

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

Connaître les fondamentaux de Python
Maîtriser les bibliothèques Python dédiées à la data science
Effectuer de la data visualisation
Utiliser les technologies autour de l'analyse données : machine Learning, big data, Business Intelligence (BI)..
Comprendre l'extraction et la gestion de données texte

Public visé

Personnes ayant une appétence pour la programmation et la manipulation des données.

Pré-requis

Avoir un niveau bac+2 à composante business ou sciences est souhaité. Des notions de marketing et de statistiques sont nécessaires.

Lieux & Horaires

Durée : 300 heures

Délai d'accès :

Jusqu'à 8 jours avant le début de la formation

Informations :

NULL

PROGRAMME

Fondamentaux de Python

Découverte des différentes variables, listes et tuples.
Présentation des divers opérateurs et structures de contrôle.
Initiation au concept de boucle et ses différents types.
Introduction aux fonctions et leurs documentations.
Instanciation de classes et utilisation de modules.

NumPy et Pandas

Création et manipulation d'un tableau NumPy array.
Présentation des opérations matricielles.
Création d'un indicateur statistique et application.
Chargement et première exploration d'un jeu de données.
Introduction au data cleaning.
Introduction au data processing.

Data quality (module optionnel)

Présentation des principes essentiels à la data quality.
Traitement des variables numériques et textuelles.
Introduction au nettoyage de données.
Initiation à la gestion des valeurs manquantes.

Matplotlib

Présentation de différents types de graphes.
Introduction à la personnalisation de graphique.
En option : présentation de graphiques avancés avec Matplotlib et création de figures avec classes et objets.

Seaborn

Initiation à l'analyse de distribution.
Découverte de l'analyse statistique.
Initiation à l'analyse multivariée.
En option : Bokeh.
Utiliser des graphiques interactifs et intégrables sur une page web.

Prochaines sessions

Consultez-nous pour les prochaines sessions.

Algorithme et méthodologie de classification avec Scikit-learn

Présentation des algorithmes de classification (régression logistique, KNN, arbre de décision, forêt aléatoire, SVM...).

Initiation aux algorithmes de boosting et de bagging.

Découverte des différents modèles et sélections.

Classification des données déséquilibrées.

Méthode de réduction de dimension

Introduction à l'analyse en composantes principales.

Présentation de l'algorithme T-SNE.

Introduction à l'analyse discriminante linéaire.

Découverte du clustering avec l'algorithme des K-means.

Méthode de régression et text mining

Introduction à la régression linéaire simple.

Découverte des expressions régulières.

Mise au point à la gestion de données textuelles.

Création de Wordclouds (nuages de mots-clés).

Présentation de l'analyse de sentiments.

Web scraping et introduction aux data engineering et big data

Introduction au langage web (HTML, CSS).

Extraction de contenu web avec BeautifulSoup.

Application du scraping sur Google.

Introduction aux 3V du big data (volume, variété, vélocité).

Présentation des bases de données et architectures.

Mise en production de données.

Data processing et machine learning sur de grandes bases de données

Introduction à PySpark.

Découverte de fonctions Pyspark (Data processing, data frames, régression avec Pyspark, ML pipelines, model turning...).

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, autodiagnosics, 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)