

DATA SCIENCE AVEC PYTHON

FFCNCERCERXBD13

RIX : 1 910 €

DURÉE : 2 JOURS

Pauses et déjeuners offerts



AVANCÉ



ATELIER



FAISABLE À
DISTANCE



RÉALISABLE
EN ANGLAIS

PRÉSENTATION

Le langage Python est reconnu comme l'un des langages majeurs en informatique. C'est l'un des plus utiles pour les utilisateurs scientifiques et techniques, particulièrement dans le domaine des sciences des données.

Pour les entreprises, utiliser Python signifie non seulement accéder à une large gamme de ressources et d'outils, mais aussi bénéficier d'une plateforme solide pour développer des applications intelligentes capables de transformer des volumes de données en informations exploitables, améliorant ainsi la prise de décision et l'efficacité opérationnelle.

OBJECTIFS

- Lire les données depuis différentes sources (fichiers HTML, fichiers Excel, bases de données, etc.)
- Nettoyer et fusionner les données
- Réaliser des analyses simples (statistiques descriptives, visualisations graphiques, régression, etc.)
- Préparer les données pour leur utilisation dans des pipelines d'apprentissage statistique

PROGRAMME

Introduction

Prise en main de Python

- Python : environnements virtuels, installation de bibliothèques
- Variables, types et structures de contrôle (boucles, conditions, etc.)

Entrées/sorties en Python

- Lire et écrire des données depuis un fichier
- Rappel sur la notion de chemin d'accès
- Erreurs et exceptions : comment Python signale-t-il les erreurs ? Comment gérer les erreurs ?

Fonctions

- Pourquoi et comment définir une fonction en Python ?
- Différentes manières d'appeler une fonction en Python
- Utilisation de modules et de bibliothèques tierces

Structures de données

- Listes, ensembles et dictionnaires
- Comment utiliser des classes existantes en Python

Données tabulaires

- Bibliothèque Numpy
- Bibliothèque Pandas
- Chargement de données depuis des fichiers Excel, des fichiers HTML et des bases de données
- Opérations basiques (filtres, opérations sur les colonnes, etc.)

Représentations graphiques

- Découverte de Matplotlib et Seaborn
- Personnalisation des graphes

Manipulation des DataFrames

- Groupby
- Jointures
- Nettoyage de données (gestion des valeurs manquantes, validation, etc.)

Manipulation de séries temporelles

- Index et filtrage
- Opérations sur des fenêtres glissantes

Classification et régression

- Aperçu de la bibliothèque Sklearn
- Mise en place d'un pipeline d'apprentissage statistique
- Comment interfacier Sklearn et Pandas ?

Étude de cas

- Analyse des données du Covid
- Intégration et consolidation de données
- Représentation graphique
- Analyse des données et prédiction de l'évolution

Synthèse et conclusion

PUBLIC/PRÉREQUIS

Ingénieurs, développeurs devant traiter des données en Python.

Des connaissances de base en programmation sont nécessaires pour tirer pleinement profit de cette formation.

RESPONSABLE(S)

Yann BALGOBIN

Yann Balgobin est titulaire d'un doctorat de Télécom Paris. Son expertise porte sur l'économie numérique et ses enjeux : protection des données personnelles, économie des plateformes, enjeux socioéconomiques du numérique. Il est également responsable pédagogique des domaines « Management de la transformation numérique » et « Intelligence Artificielle et Science des données » à Télécom Paris Executive Education.

MODALITÉS PÉDAGOGIQUES

Travaux pratiques et études de cas.

