DATA SCIENCE AVEC PYTHON

FFCNCERCERXBD13	PRIX : 1 930 €	DURÉE : 2 JOURS

Pauses et déjeuners offerts

AVANCÉ

ATELIER

RÉALISABLE EN ANGLAIS

FAISABLE À DISTANCE

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.)
- 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
- Environnements virtuels et installation de bibliothèques
- Variables, types et structures de contrôle (boucles, conditions, etc.)
- Entrées/sorties
- Fonctions

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

Données tabulaires
- Bibliothèques Numpy et 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

Études 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)

Guillaume WISNIEWSKI
Enseignant-chercheur en informatique à Université Paris Cité. Il est diplômé de Télécom ParisTech (promo 2004) et de l'Université Pierre et Marie Curie (thèse en apprentissage statistique en 2007). Ses travaux de recherche portent sur la linguistique informatique et la compréhension automatique des langues. Depuis 2008, il enseigne la data science, le traitement des langues, l'apprentissage statistique, l'apprentissage profond et les giga-modèles de langue (LLMs).

MODALITÉS PÉDAGOGIQUES

Travaux pratiques et études de cas.