

[Formation] Data Science avec Python

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, Jupyter notebook
- 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 ?

DATES ET LIEUX

Du 24/09/2026 au 25/09/2026 à Paris

PUBLIC / PREREQUIS

Ingénieurs, développeurs devant traiter des données en Python.

Des connaissances de base en programmation sont nécessaires afin de tirer pleinement profit de la formation.

COORDINATEURS

Guillaume WISNEWSKI

Docteur en informatique à Université Paris Dauphine (2000) et de l'Université Pierre et Marie Curie (2004 en apprentissage statistique en 2007). Ses travaux de recherche portent sur la

MODALITES PEDAGOGIQUES

Travaux pratiques et études de cas.

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 interfacer Sklearn et Pandas ?

Étude de cas

Analyse des données du Covid

- Intégration et consolidation des données provenant de plusieurs sources
- Représentation graphique
- Analyse des données et prédition de l'évolution

Synthèse et conclusion

Appelez le 01 75 31 95 90
International : +33 (0)1 75 31 95 90

contact.exed@telecom-paris.fr / executive-education.telecom-paris.fr