

# Python Data Science, manipuler et visualiser les données

certification TOSA® en option

Cours Pratique de 4 jours - 28h

Réf : IYT - Prix 2024 : 2 390€ HT

La science des données est un domaine multidisciplinaire en constante expansion. Elle s'appuie sur des méthodes scientifiques, des algorithmes et des processus que Python a su maîtriser grâce à un écosystème particulièrement riche. Il est devenu aujourd'hui le langage de référence pour l'analyse de données, quels qu'en soient les formats. Notre formation vous permet la prise en main des outils, bibliothèques et modules Python pour obtenir de rapides compétences en data science avec ce langage.

## OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Posséder une vue d'ensemble de l'écosystème scientifique de Python

Connaître les bibliothèques scientifiques incontournables pour la science des données

Être capable de manipuler des données volumineuses avec Python

Comprendre l'intérêt de la datavisualisation

Savoir visualiser des données avec Python

## MÉTHODES PÉDAGOGIQUES

Pédagogie active, retours d'expériences, des démonstrations sont mises en œuvre par le formateur pour une mise en pratique plus rapide par les participants.

## TRAVAUX PRATIQUES

De nombreux exercices sont réalisés pour illustrer les sujets.

## LE PROGRAMME

dernière mise à jour : 10/2023

### 1) Présentation de l'écosystème Python scientifique

- Panorama de l'écosystème scientifique de Python : les bibliothèques incontournables.
- Savoir où trouver de nouvelles bibliothèques et juger de leur pérennité.
- Les principaux outils et logiciels open source pour la data science.
- Pourquoi utiliser une distribution scientifique, Anaconda.
- Comprendre l'intérêt d'un environnement virtuel et savoir l'utiliser.
- L'interpréteur IPython et le serveur Jupyter.
- Les bonnes pratiques pour bien démarrer votre projet de data science avec Python.
- Les formats de fichiers scientifiques et les bibliothèques pour les manipuler.

*Travaux pratiques : Mise en place de l'environnement de développement : installation d'Anaconda, création d'un environnement virtuel, export et duplication d'un environnement, utiliser les notebooks Jupyter.*

### 2) La SciPy Stack

- Le socle de bibliothèques scientifiques incontournables sur lequel sont basées toutes les autres : la SciPy Stack.
- NumPy : calcul numérique et algèbre linéaire (les vecteurs, matrices, images).
- SciPy, basée sur NumPy pour les statistiques, les analyses fonctionnelles, géospatiales, le traitement du signal, etc.
- Pandas : l'analyse de données tabulaires (CSV, Excel, etc.), statistiques, pivots, filtres, recherche...

## PARTICIPANTS

Ingénieur, développeur, chercheur, data scientist, data analyst et toute personne désireuse de se former à l'univers scientifique de Python.

## PRÉREQUIS

Pratique du langage Python.

## COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

## MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

## MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

## ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

- Matplotlib : la librairie de visualisation de données incontournable.

*Travaux pratiques* : Mesurer les performances du NumPy installé par votre Linux et celui d'Anaconda. Traitement d'images avec NumPy. Premiers tracés. Analyses statistiques de fichiers CSV. Premiers éléments de cartographie. Transformées de Fourier.

### 3) Les librairies de visualisation

- Panorama des librairies de visualisation de Python : 2D/3D, desktop/web, statistiques, cartographie, big data...

- Les librairies orientées desktop : Matplotlib, Pandas, Seaborn.

- Les librairies orientées web : Bokeh, Altair, Plotly...

- Les librairies pour la 3D : Plotly, pythreejs, ipyvolumes...

- Les librairies cartographiques : Cartopy, folium, ipyleaflet, Bokeh, cesiumpy...

- Les librairies big data : datashader, Vaex...

*Travaux pratiques* : Réalisation de multiples exercices avec quelques librairies présentées. Visualisation big data, cartographique, 2D et 3D.

### 4) La datavisualisation

- L'intérêt de la datavisualisation

- Utiliser PyViz et l'écosystème HoloViz.

- Présentation des outils SuperSet, Mayavi, Paraview et VisIt.

*Travaux pratiques* : Poursuivre l'utilisation des librairies de visualisation et manipulations des outils.

### 5) Les formats de fichiers scientifiques et la manipulation de données volumineuses

- Panorama des principaux formats de fichiers scientifiques : NetCDF, HDF5, GRIB, JSON, PARQUET, MATLAB, CGNS...

- Manipuler des données volumineuses avec Dask, Vaex et Xarray.

*Travaux pratiques* : Manipulation de données dépassant les Go, lecture et écriture de fichiers NetCDF/HDF5. Visualisation de données climatiques, images satellites, création de vidéos/animations graphiques.

## LES DATES

---

#### CLASSE À DISTANCE

2024 : 11 juin, 17 sept., 26 nov.

#### PARIS

2024 : 04 juin, 10 sept., 19 nov.

#### LYON

2024 : 11 juin, 17 sept., 26 nov.

#### AIX-EN-PROVENCE

2024 : 17 sept., 26 nov.

#### BORDEAUX

2024 : 17 sept., 26 nov.

#### LILLE

2024 : 17 sept., 26 nov.

#### NANTES

2024 : 17 sept., 26 nov.

#### SOPHIA-ANTIPOLIS

2024 : 17 sept., 26 nov.

#### STRASBOURG

2024 : 17 sept., 26 nov.

#### TOULOUSE

2024 : 17 sept., 26 nov.