

# Utiliser Python pour la data science

## Tutorat en option

Formation en ligne - 02h30

Réf : 4GU - Prix 2024 : 95CHF HT

Ce cours en ligne a pour objectif de vous apprendre à utiliser le langage Python dans le cadre de la data science pour être en mesure d'analyser de gros volumes de données. Il s'adresse à des personnes disposant de connaissances en langage Python. La pédagogie s'appuie sur un auto-apprentissage séquencé par actions de l'utilisateur sur l'environnement à maîtriser. Une option de tutorat vient renforcer l'apprentissage.

### OBJECTIFS PÉDAGOGIQUES

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

Connaître NumPy et Pandas, les bibliothèques dédiées à la data science

Préparer et virtualiser des données

Analyser des variables

Estimer des modèles avec la bibliothèque Scikit-Learn

### PÉDAGOGIE ET PRATIQUES

Une évaluation tout au long de la formation grâce à une pédagogie active mixant théorie, exercice, partage de pratique et gamification. Un service technique est dédié au support de l'apprenant. La formation est diffusée au format SCORM (1.2) et accessible en illimité pendant 1 an.

### ACTIVITÉS DIGITALES

Démonstrations, cours enregistrés, partages de bonnes pratiques, quiz, fiches de synthèse.

## LE PROGRAMME

dernière mise à jour : 10/2023

### 1) Python et la data science

- Choisir Python pour la data science.
- La bibliothèque NumPy.
- Le type et la taille des vecteurs NumPy.
- Initialisation et tableau NumPy.
- Accès aux données d'un tableau NumPy à une dimension.
- Accès aux données d'un tableau NumPy à deux dimensions.
- Algèbre linéaire avec NumPy.
- Tableau NumPy vs liste Python.
- Statistique descriptive avec NumPy.

### 2) Visualisation de données

- Installer Anaconda et Jupyter.
- Travailler avec Jupyter.
- Utiliser la bibliothèque Pandas.
- Accéder aux données d'un DataFrame.
- Filtrer les données d'un DataFrame.
- Trier les données d'un DataFrame.
- Gérer les statistiques de base avec un DataFrame.
- Lire des fichiers de grande taille avec Pandas.

### PARTICIPANTS

Développeurs Python et toute personne s'intéressant à la data science.

### PRÉREQUIS

Connaissances solides du langage Python.

### COMPÉTENCES DU FORMATEUR

Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. 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. 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

La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques. Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

### MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels. ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

### 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.

### 3) Statistiques inférentielles avec Python

- Utiliser des méthodes melt et apply.
- Extraire des informations à partir des données existantes.
- Créer de nouvelles variables à partir des données existantes.
- Visualiser les données avec Matplotlib.
- Appréhender les tests d'hypothèses.
- Réaliser des tests statistiques de comparaison de deux moyennes.

### 4) Modélisation multivariée avec Python

- Introduction à la régression linéaire.
- Exemple de modélisation avec la régression linéaire.
- Introduction à l'algorithme Support Vector Machine.
- Exemple de modélisation avec un Support Vector Machine.
- Introduction à l'algorithme de K-means.
- Exemple de l'algorithme K-means.