

# Data Scientist, parcours de 9 mois en continu tutorat inclus

Formation en ligne - 400h

Réf : 4DB - Prix 2024 : 5 990€ HT

ORSYS et DataScientest, leader dans le domaine de la data science en France, proposent un parcours de formation complet pour exercer le métier de data scientist. Le data scientist traite les données et élabore des modèles prédictifs. Ce parcours en ligne vous apprend les fondamentaux du langage Python et l'usage de ses principales bibliothèques pour la data science. Vous apprenez également toutes les technologies autour de l'analyse de la donnée comme la data visualisation, le machine learning, le big data, le deep learning ainsi que les systèmes complexes et l'intelligence artificielle (IA).

## OBJECTIFS PÉDAGOGIQUES

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

Connaître les fondamentaux du langage Python

Maîtriser les bibliothèques Python dédiées à la data science

Effectuer de la data visualisation

Utiliser les technologies autour de la data science (data visualisation, machine learning, big data, deep learning...)

Comprendre les systèmes complexes et l'intelligence artificielle (IA)

## PÉDAGOGIE ET PRATIQUES

Formation digitale basée sur une pédagogie active et conçue avec des experts en data science. Une combinaison de théorie, de démonstrations, de mises en pratique, de partages d'expériences et de bonnes pratiques. Un test de positionnement, un accompagnement sur mesure dès le début du parcours, un projet fil rouge et des séquences pédagogiques de courte durée permettent de renforcer l'apprentissage et d'évaluer l'apprenant tout au long de sa formation. En cas de besoin technique, une cellule support est disponible en ligne 5 jours sur 7 de 9 h à 18 h 30. À l'issue de son parcours en ligne, l'apprenant reçoit un certificat délivré par Mines ParisTech | PSL Executive Education, ce qui lui permet de bénéficier de la reconnaissance d'un acteur de référence dans le domaine.

## ACTIVITÉS DIGITALES

Test de positionnement sous forme de QCM d'une heure, séance d'introduction à la plateforme à distance, classes collectives, classe de soutien sur mesure, simulation et codage en direct, exercices, fiches de synthèse, projet fil rouge dédié, social learning, échanges avec data scientists.

## PARTICIPANTS

Toutes les personnes ayant une appétence pour la programmation et la manipulation des données.

## PRÉREQUIS

Avoir un niveau bac +3 en mathématiques ou un niveau bac +5 en sciences (ingénieur, mathématiques, statistiques, économétrie...). Des notions de marketing et de communication sont un plus.

## 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.

## LE PROGRAMME

dernière mise à jour : 03/2023

## 1) Fondamentaux de Python

- Découverte des différents types de variables, des listes et des tuples.
- Présentation des divers opérateurs et structures de contrôle.
- Initiation au concept de boucle et ses différents types.
- Introduction aux fonctions et leur documentation.
- Instanciation de classes et utilisation de modules.

## 2) Les bibliothèques NumPy et Pandas

- Création et manipulation d'un tableau NumPy array.
- Présentation des opérations matricielles et manipulation d'un NumPy array.
- Création d'un indicateur statistique et opérations sur un NumPy array.
- Introduction à la bibliothèque Pandas.
- Introduction au nettoyage des données (data cleaning).
- Introduction au data processing.
- En option : Data quality (principes, variables, nettoyage de données, gestion des valeurs manquantes).

## 3) Matplotlib et Seaborn

- Présentation des différents types de graphes.
- Introduction à la personnalisation des graphiques.
- Seaborn : maîtrise de l'analyse des distributions.
- Mise en place de l'analyse statistique.
- Initiation à l'analyse multivariée.
- Bokeh (optionnel).
- Formation à tout type de graphiques interactifs et intégrables sur page web.
- Visualisation de données géographiques.
- Découverte et création de widgets.

## 4) Machine learning

- Introduction à Scikit-learn et présentation de la classification support vector machines (SVM).
- Application de la méthode des k plus proches voisins (k-nearest neighbors ou KNN).
- Sélection de modèles et classification semi-supervisée.
- Présentation des algorithmes des k-moyennes.
- Découvertes des algorithmes de mean shift.
- Introduction à la régression linéaire simple et multiple.
- Introduction à la régression linéaire régularisée.
- Feature selection process.
- Initiation aux analyses en composantes principales.
- Application de l'approche manifold learning.

## 5) Séries temporelles avec Statsmodels et text mining

- Création de nuages de mots (Wordclouds).
- Découverte des modèles de base.
- Découverte des modèles ARIMA.
- Text mining : introduction aux expressions régulières.
- Gestion de données textuelles.

## 6) Machine learning et théorie des graphes avec NetworkX

- Introduction à la théorie des graphes.
- Application des algorithmes fondamentaux : Kruskal et Dijkstra.
- Détection de communautés.
- Application de l'algorithme PageRank au classement de pages web.

## 7) Introduction au data engineering et big data

- Introduction aux 3V du big data (volume, variété, vélocité).

- Présentation des bases de données et des architectures.
- Mise en production de données.
- Data processing et machine learning sur des grandes bases de données.
- Introduction à PySpark.
- Découverte des fonctions de PySpark : data processing, dataframes, régression avec PySpark, ML pipelines, model turning.

#### 8) Deep learning avec la bibliothèque Keras

- Découverte de concepts : Dense neural networks, convolutional neural networks, architecture LeNet, transfer Learning.

#### 9) Introduction à TensorFlow 2.0 et au renforcement learning

- Mise en relation de TensorFlow et Keras.
- Application du word embedding avec Word2vec.
- Présentation du recurrent neural network.
- Présentation du generative adversarial network.
- Définition des éléments de l'apprentissage par renforcement.
- Mise au point sur les fondamentaux mathématiques pour le renforcement learning.
- Présentation de la principale famille d'algorithmes de renforcement learning : Monte-Carlo.
- Application des algorithmes d'apprentissage par différence temporelle : SARSA.
- Application des algorithmes d'apprentissage par différence temporelle : Q-Learning.
- Comparaison entre apprentissage par différence temporelle et Monte-Carlo.

#### 10) Deep reinforcement learning (apprentissage par renforcement profond)

- Application à des business cases.
- Introduction au deep reinforcement learning.
- Compréhension des fondements des méthodes basées sur la valeur.
- Compréhension des problèmes de variance élevée, des risques de surestimation, application des méthodes de résolution.
- Introduction à la méthode de policy gradient.
- Compréhension des fondamentaux de la structure acteur-critique utilisée dans l'apprentissage en deep reinforcement learning.

## NOS POINTS FORTS

---

- Séquences de courte durée
- Activités digitales variées
- Accès illimité pendant 1 an ou pendant la durée du parcours
- Tutorat personnalisé inclus ou en option
- Accès multi-device (smartphone, tablette ou ordinateur)