

# Data analyst, parcours intensif de 9 semaines tutorat inclus

Formation en ligne - 300h

Réf : 4ID - Prix 2024 : 5 590€ HT

ORSYS et DataScientest proposent un parcours de formation complet pour le métier de data analyst. Le data analyst ou business analyst traite et analyse les données dans une optique de prise de décision business, organisationnelle ou projet. Ce parcours vous apprend les fondamentaux du langage Python et ses bibliothèques principales pour la data science. Vous apprenez également toutes les technologies autour de l'analyse des données comme la data visualisation, le machine learning, l'extraction et la gestion de données texte, le big data ainsi que la business intelligence (BI).

## OBJECTIFS PÉDAGOGIQUES

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

Connaître les fondamentaux de Python

Maîtriser les bibliothèques Python dédiées à la data science

Effectuer de la data visualisation

Utiliser les technologies autour de l'analyse données : machine Learning, big data, Business Intelligence (BI)...

Comprendre l'extraction et la gestion de données texte

## 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

Personnes ayant une appétence pour la programmation et la manipulation des données.

## PRÉREQUIS

Avoir un niveau bac+2 à composante business ou sciences est souhaité. Des notions de marketing et de statistiques sont nécessaires.

## 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 : 10/2023

## 1) Fondamentaux de Python

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

## 2) NumPy et Pandas

- Création et manipulation d'un tableau NumPy array.
- Présentation des opérations matricielles.
- Création d'un indicateur statistique et application.
- Chargement et première exploration d'un jeu de données.
- Introduction au data cleaning.
- Introduction au data processing.

## 3) Data quality (optionnel)

- Présentation des principes essentiels à la data quality.
- Traitement des variables numériques et textuelles.
- Introduction au nettoyage de données.
- Initiation à la gestion des valeurs manquantes.

## 4) Matplotlib

- Présentation de différents types de graphes.
- Introduction à la personnalisation de graphique.
- En option : présentation de graphiques avancés avec Matplotlib et création de figures avec classes et objets.

## 5) Seaborn

- Initiation à l'analyse de distribution.
- Découverte de l'analyse statistique.
- Initiation à l'analyse multivariée.
- En option : Bokeh.
- Utiliser des graphiques interactifs et intégrables sur une page web.

## 6) Algorithme et méthodologie de classification avec Scikit-learn

- Présentation des algorithmes de classification (régression logistique, KNN, arbre de décision, forêt aléatoire, SVM...).
- Initiation aux algorithmes de boosting et de bagging.
- Découverte des différents modèles et sélections.
- Classification des données déséquilibrées.

## 7) Méthode de réduction de dimension

- Découverte du clustering avec l'algorithme des k-means.
- Introduction à l'analyse en composantes principales.
- Présentation de l'algorithme T-SNE.
- Introduction à l'analyse discriminante linéaire.

## 8) Méthode de régression et text mining

- Introduction à la régression linéaire simple.
- Introduction aux expressions régulières.
- Mise au point à la gestion de données textuelles.
- Création de nuages de mots.
- Présentation de l'analyse de sentiments.

## 9) Web scraping et introduction aux data engineering et big data

- Introduction au langage web (HTML, CSS).

- Extraction de contenu web avec Beautiful Soup.
- Application du scraping sur Google.
- Introduction aux 3V du big data (volume, variété, vitesse).
- Présentation des bases de données et architectures.
- Mise en production de données.

#### 10) Data processing et machine learning sur de grandes bases de données

- Introduction à PySpark.
- Découverte de fonctions PySpark (Data processing, data frames, régression avec PySpark, ML pipelines, model turning...).

## 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)