



Data Scientist

Réalisez des analyses de données et des prédictions métiers poussées grâce à la data science.



Durée de la formation :
600 heures



Embauché ou remboursé**



Durée de la formation en alternance :
18 mois



Temps d'études conseillé à temps plein :
9 mois



Diplôme de niveau 7 (bac +5) *

*Expert en data science - code NSF 326 - Diplôme de niveau 7 (bac +5) - certification professionnelle enregistrée au Répertoire National des Certifications Professionnelles (RNCP) par décision de France compétences publiée le 27/03/2023

**Conditions disponibles sur la page <https://openclassrooms.com/fr/terms-of-sale>

OPENCLASSROOMS

Devenez data scientist

Réalisez des analyses de données et des prédictions métiers poussées grâce à la data science

Les entreprises produisent une quantité astronomique de données. Être capable de les analyser et de les valoriser représente un enjeu crucial et un avantage compétitif indéniable.

En tant que data scientist, votre rôle sera de **traduire un besoin métier en une problématique de data science, puis de la résoudre grâce à vos algorithmes**. Vous réaliserez par exemple des moteurs de recommandations, des prédictions pour améliorer les ventes de votre entreprise ou encore des intelligences artificielles pour des applications mobiles. Au contact avec des équipes métiers, vous mènerez à bien un projet data, de la collecte des données à la mise en production de vos algorithmes.



CentraleSupélec

Parcours en partenariat avec CentraleSupélec

En tant que data scientist, quelles sont vos missions ?

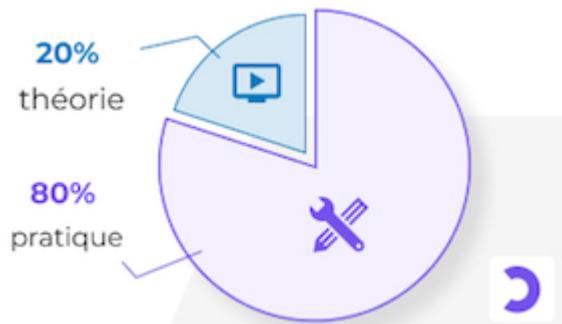
1. **Collecter et préparer les données** avec Python en vue de l'analyse.
2. **Explorer et analyser les données** pour comprendre les tendances et les performances d'une entreprise ou d'un produit.
3. **Élaborer des modèles prédictifs** pour identifier les nouvelles tendances et opportunités.
4. **Déployer des modèles de données** (sur le cloud) pour automatiser les processus d'entreprise et améliorer leur efficacité.
5. **Communiquer les résultats** à des spécialistes ou des néophytes grâce à des visualisations.
6. **Participer à l'organisation et au pilotage d'un projet** de data science / IA.

Vous maîtriserez les technologies et les langages techniques les plus demandés :

Python (Jupyter Notebook), Git, Github, Streamlit, MLFlow

Technologies : NLP, Computer Vision, Deep Learning, MLOps, Cloud (AWS, Azure, Heroku, Google Cloud), Databricks, API, Hadoop/PySpark

Nos formations vous permettent d'apprendre par la pratique



Réalisez des projets concrets basés sur des mises en situation professionnelles. Savoir c'est bien, savoir faire c'est mieux !

Vous êtes accompagné par un mentor, professionnel du métier, tout au long de votre formation.



Alison Patou, data scientist et mentor OpenClassrooms sur les parcours data depuis plus de 4 ans

« Accompagner les étudiants tout au long de leur formation en leur partageant mon expérience, c'est là tout le sens de mon rôle de mentor.
» Alison

Autonomes, mais jamais seuls !

Chez OpenClassrooms, vous pourrez compter aussi sur de nombreux soutiens :

- les professeurs, experts du domaine ;
- les conseillers pédagogiques dédiés à votre réussite ;
- une communauté d'étudiants sur Slack.

Quelles sont vos opportunités de carrières et perspectives d'emploi à la suite de cette formation diplômante ?

En suivant notre formation en ligne, vous pourrez exercer de nombreux métiers dont celui de **data scientist**, de **data architect**, ou aussi de **data engineer**.

Les salaires du marché pour le poste de data scientist :

- **35 000 € à 45 000 €** brut annuel pour un profil **débutant**
- **45 000 € à 65 000 €** brut annuel pour un profil **confirmé**

Cette fourchette peut toutefois varier selon la taille de l'entreprise et la région.

Source : étude de salaire, Data Recrutement

Les profils de data scientist sont très recherchés, majoritairement en CDI. En freelance, les data scientists facturent jusqu'à 1000€ par jour de travail.

Découvrez les regards croisés de professionnels sur les métiers de la Data grâce à notre podcast



Apprenez en plus sur le métier de data scientist à travers l'épisode dédié de notre podcast, le Joboscope

Et si vous suiviez cette formation en alternance ?

La formation de data scientist est disponible en contrat de professionnalisation et en apprentissage.

Avec OpenClassrooms vous pouvez démarrer votre alternance à n'importe quel moment de l'année. Votre formation est financée par l'entreprise et vous touchez un salaire mensuel, calculé selon votre situation. En plus des projets concrets à valider dans la formation, vous pouvez mettre en pratique directement vos compétences acquises durant votre parcours.

Pour vous soutenir dans cette démarche, les équipes d'OpenClassrooms vous aident à trouver une entreprise prête à vous embaucher dès le début de vos études !

Découvrez-en plus sur l'[alternance](#) en ligne chez OpenClassrooms.

Nos alumni racontent

Les prérequis pour postuler

Pour accéder directement à la certification, les candidats doivent obligatoirement justifier d'un des **prérequis** suivants :

- être titulaire d'une certification professionnelle de **niveau 6** (bac+3 ou équivalent) en mathématiques, informatique, technologie, économie, finance ou analyse de données ;

ou

- être titulaire d'une certification professionnelle de **niveau 5** (ou équivalent) avec une expérience professionnelle d'au moins 2 ans en mathématiques, informatique, technologie, économie, finance ou analyse de données ;

Tout autre profil fera l'objet d'une **procédure dérogatoire**, à savoir :

- être titulaire d'une certification professionnelle de niveau 6 (bac+3 ou équivalent) tous domaines avec une expérience professionnelle tous secteurs d'au moins 1 an ;
- et effectuer un test de positionnement. Ce test vous sera transmis lors du processus d'admission.

- **Langue** : un niveau B2 minimum en français.

- **Matériel et logiciels** : l'accès à un ordinateur (PC ou Mac) muni d'un casque son, d'une webcam et d'une bonne connexion Internet (3,2 Mbps en envoi et 1,8 Mbps en réception de données). Pour tester la qualité de votre connexion, cliquez sur ce [lien](#). De plus, il faut être administrateur de votre ordinateur, afin de pouvoir installer des programmes complémentaires. Ce parcours nécessite d'être équipé d'un ordinateur avec les spécifications suivantes :

- core i3 ou AMD Ryzen 3 minimum (idéalement core i7 ou AMD Ryzen 7) ;
- mémoire vive (RAM) de 8 Go minimum (16 Go idéalement, mais possibilité d'utiliser Google Colaboratory pour les fichiers de données les plus importants du parcours, avec un compte gratuit) ;
- espace disque disponible 256 Go, SSD idéalement.

Vous ne répondez pas aux critères d'admissibilité ?

Nous vous invitons à déposer votre candidature, un conseiller en formation étudiera votre dossier et vous recontactera.

Projet 1 - 10 heures

Démarrez votre formation de Data Scientist

Mettez-vous dans les meilleures conditions pour réussir votre parcours : projetez-vous dans votre formation, définissez votre planning et apprenez-vous les outils essentiels pour apprendre.

Compétences cibles

- Définir le cadre de votre formation

Cours associés



Engagez-vous dans votre formation OpenClassrooms

Facile 2 heures

Prenez en main votre parcours OpenClassrooms et réalisez votre premier projet en suivant ce cours conçu pour vous accompagner dans ces premières étapes de formation.



Objectif IA : initiez-vous à l'intelligence artificielle

Facile 6 heures

Vous découvrirez ce que recouvre vraiment l'intelligence artificielle ! Ses enjeux, ses possibilités, son fonctionnement scientifique ainsi que ses sous-disciplines, comme le Machine Learning et le Deep Learning.

Projet 2 - 60 heures

Analysez des données de systèmes éducatifs

L'entreprise "academy" cherche à s'étendre à l'international. Dans ce premier projet, vous ferez des recommandations stratégiques à partir de données de systèmes éducatifs.

Compétences cibles

- Effectuer une analyse univariée et des représentations graphiques avec une librairie
- Maîtriser les opérations fondamentales du langage Python pour la Data Science
- Manipuler des données avec des librairies Python spécialisées
- Mettre en place un environnement Python
- Utiliser un notebook Jupyter pour faciliter la rédaction du code et la collaboration

Cours associés



Initiez-vous à Python pour l'analyse de données

Facile 6 heures

Dans ce cours, vous apprendrez un langage de programmation indispensable pour l'analyse de données : Python. Nous aborderons ensemble les notions fondamentales de la programmation Python, à l'aide d'exemples simples et d'exercices pratiques.



Découvrez les librairies Python pour la Data Science

Moyenne 8 heures

Découvrez les librairies Python spécialisées pour la Data Science. Maîtrisez NumPy, Matplotlib, Pandas et Seaborn pour créer des Data Frames et des Data Visualisations.



Apprenez à utiliser la ligne de commande dans un terminal

Facile 6 heures

Apprenez à écrire des lignes de commandes qui vous permettront de communiquer avec votre ordinateur.



Nettoyez et analysez votre jeu de données

Facile 10 heures

Prêt à entrer dans l'univers de la statistique descriptive ? Après ce cours, vous serez capable de nettoyer et décrire un jeu de données.

Projet 3 - 70 heures

Préparez des données pour un organisme de santé publique

L'agence "Santé publique France" a lancé un appel à projet autour des problématiques alimentaires. Vous proposerez une application basée sur des données nutritionnelles.

Compétences cibles

- Déterminer les objectifs du nettoyage des données et de la démarche de mise en œuvre
- Effectuer des analyses statistiques univariées et multivariées
- Effectuer des opérations de nettoyage sur des données structurées
- Représenter des données grâce à des graphiques afin justifier les analyses réalisées

Cours associés



Initiez-vous au Machine Learning

Moyenne 10 heures

Découvrez le Machine Learning et ses différentes techniques (régression linéaire, classification non supervisée...). Vous verrez comment un algorithme apprend pour résoudre un problème de Data Science, et vous entrainerez votre premier modèle !



Réalisez une analyse exploratoire de données

Moyenne 15 heures

Comprenez les tendances de votre jeu de données avec la puissante ACP (Analyse en Composantes Principales) et d'autres méthodes de classification automatique comme le k-means.



Gérez votre temps efficacement

 Facile

 12 heures

La gestion du temps est une compétence qui peut s'apprendre. Découvrez une série de lois du temps et de bonnes habitudes à prendre, et devenez plus performant et efficace dans votre travail.

Projet 4 - 90 heures

Anticipez les besoins en consommation de bâtiments

Pour atteindre son objectif de ville neutre en émissions de carbone en 2050, la ville de Seattle a besoin de vous. Votre mission ? Prédire la consommation électrique des bâtiments municipaux.

Compétences cibles

- Créer les variables pertinentes d'un modèle d'apprentissage supervisé ou non-supervisé
- Sélectionner, entraîner et évaluer des modèles d'apprentissage supervisé

Cours associés



Evaluez les performances d'un modèle de machine learning

Moyenne 10 heures

Apprenez à évaluer un algorithme de machine learning, évitez le sur-apprentissage, et choisissez le meilleur modèle pour votre problème, à l'aide de la validation croisée et la grid-search.



Entraînez un modèle prédictif linéaire

Moyenne 10 heures

Découvrez les algorithmes d'apprentissage supervisés. Appliquez une régression linéaire ou logistique et appröhendez les méthodes à large marge (SVM).



Utilisez des modèles supervisés non linéaires

Moyenne 12 heures

Etendons les méthodes linéaires à la modélisation de relations non linéaires entre les données, notamment à l'aide du SVM et du perceptron. Vous découvrirez aussi une famille d'algorithme très populaire... les réseaux de neurones !



Modélez vos données avec les méthodes ensemblistes

Moyenne 15 heures

Découpez la robustesse et l'efficacité de vos algorithmes à l'aide des méthodes ensemblistes, le bagging et le boosting. Vous découvrirez aussi les forêts aléatoires et le très prisé XGBoost.

Projet 5 - 70 heures

Segmentez des clients d'un site e-commerce

Vous êtes consultant pour Olist, un site e-commerce brésilien. Les équipes marketing ont besoin de segmenter leurs clients pour optimiser les campagnes de communication.

Compétences cibles

- Évaluer les performances des modèles d'apprentissage non supervisé
- Sélectionner et entraîner des modèles d'apprentissage non-supervisé

Cours associés



Explorez vos données avec des algorithmes non supervisés

Difficile 15 heures

Comment faire parler vos données sans les étiqueter ? Apprenez à mettre en œuvre le clustering (k-means, DBSCAN, clustering hiérarchique) et la réduction dimensionnelle (ACP, MDS, t-SNE).



Améliorez l'impact de vos présentations

Moyenne 8 heures

Découvrez le design de présentation, avec des techniques de facilitation visuelle. Faites passer un message à impact émotionnel positif grâce à des diaporamas créatifs et modernes.

Projet 6 - 100 heures

Classifiez automatiquement des biens de consommation

Votre entreprise cherche à lancer une place de marché e-commerce. Vous devrez tester la faisabilité d'un moteur de classification de biens de consommation.

Compétences cibles

- Définir la stratégie d'élaboration d'un modèle d'apprentissage profond
- Évaluer la performance des modèles d'apprentissage profond
- Prétraiter des données non structurées de type image
- Prétraiter des données non structurées de type texte
- Réaliser la collecte des données répondant à des critères définis via une API
- Réduire la dimension de données de grande dimension
- Utiliser des techniques appropriées de réduction en deux dimensions

Cours associés



Modélez des sujets avec des méthodes non supervisées

Analysez vos données textuelles

Moyenne 8 heures

Les données textuelles, non structurées, sont omniprésentes dans vos fils d'actualité, ou encore sur les réseaux sociaux. Transformez et modélez vos données textes grâce aux bag of words, aux word embedding et même aux réseaux de neurones !



Classez et segmentez des données visuelles

Difficile 15 heures

Enrichissez votre palette de Data Scientist en classant des données visuelles. Dans ce cours, vous allez prétraiter des images et les modéliser grâce au SIFT et au Deep Learning (CNN).



Initiez-vous au Deep Learning

Moyenne 8 heures

Nous allons étudier les principes de l'intelligence artificielle à travers des architectures de type réseaux de neurones. Après avoir présenté les réseaux classiques (MLP, CNN, GAN), nous aborderons les modèles récurrents dédiés aux données séquentielles (RNN, LSTM).

Projet 7 - 80 heures

Implémentez un modèle de scoring

Au sein d'une société financière, vous allez développer et implémenter un modèle de scoring pour aider les équipes métiers à accorder un crédit à un client.

Compétences cibles

- Concevoir un déploiement continu d'un moteur d'inférence sur une plateforme Cloud
- Définir et mettre en œuvre un pipeline d'entraînement des modèles
- Définir la stratégie d'élaboration d'un modèle d'apprentissage supervisé
- Évaluer les performances des modèles d'apprentissage supervisé
- Mettre en œuvre un logiciel de version de code
- Suivre la performance d'un modèle en production et en assurer la maintenance

Cours associés



Gérez du code avec Git et GitHub

Facile 6 heures

Installez Git et GitHub et apprenez les commandes de base de Git pour gérer votre code et déployer vos projets de développement.

Projet 8 - 40 heures

Réalisez un dashboard et assurez une veille technique

Toujours au sein d'une société financière, vous allez développer un dashboard interactif pour que des chargés de relation client puissent expliquer de façon transparente les décisions d'octroi de crédit. Vous ferez aussi une veille sur l'état de l'art.

Compétences cibles

- Réaliser la présentation orale d'une démarche de modélisation à un client interne/externe
- Réaliser une veille sur les outils et tendances en data science et IA
- Réaliser un tableau de bord afin de présenter son travail de modélisation à un public
- Rédiger une note méthodologique afin de communiquer sa démarche de modélisation

Cours associés



Prenez la parole en public

Facile 6 heures

Mobilisez des techniques de communication en public. Créez votre contenu avec le storytelling et le mind mapping, et entraînez-vous pour maîtriser votre prestation orale.



Rédigez des écrits professionnels

Facile 4 heures

La communication écrite est primordiale dans le monde professionnel. Quel que soit votre métier, si vous voulez maîtriser la rédaction d'un écrit de la structuration à la relecture, ce cours vous sera utile !



Mettez en place un système de veille informationnelle

Facile 4 heures

Apprenez à faire de la veille pour vous adapter aux évolutions de votre métier et à utiliser des outils de veille informationnelle.

Projet 9 - 40 heures

Réalisez un traitement dans un environnement Big Data sur le Cloud

Votre startup AgriTech souhaite développer une application mobile permettant de détecter des fruits sur une photo. À vous d'industrialiser le modèle à grande échelle grâce aux outils du big data.

Compétences cibles

- Modéliser des données dans un environnement Big Data et en utilisant les outils du Cloud
- Réaliser des calculs distribués sur des données massives en utilisant les outils adaptés
- Sélectionner les outils du Cloud permettant de traiter et stocker des données Big Data

Cours associés



Découvrez le cloud avec Amazon Web Services

Moyenne 12 heures

Grâce à ce cours, prenez en main le cloud Amazon Web Services (AWS), montez un serveur web EC2, installez une base de données sur RDS et stockez des fichiers sur S3.



Concevez des architectures Big Data

Moyenne 6 heures

Nous sommes à l'âge d'or du Big Data et les Data Architects disposent de tous les outils dont ils ont besoin pour gérer des données massives. Mais comment les assembler ? Familiarisez-vous avec une vision d'ensemble pour la conception d'architectures Big Data complètes.



Réalisez des calculs distribués sur des données massives

Difficile 20 heures

Dans ce cours, vous apprendrez à réaliser des analyses de données massives sur des centaines de machines dans le cloud grâce à Hadoop MapReduce, Spark et Amazon Web Services.

Projet 10 - 40 heures

Réalisez le cadrage d'un projet IA

Dans quelques jours va se tenir le COMEX qui validera la roadmap produit de votre entreprise. Vous allez cadrer le projet afin de convaincre les dirigeants que votre produit possède un fort potentiel.

Compétences cibles

- Analyser les risques inhérents au projet et à la gestion des données personnelles
- Identifier des besoins métiers d'un projet et les formaliser sous forme de user stories
- Identifier les ressources humaines, techniques et financières d'un projet de data science
- Intégrer le respect des contraintes légales et éthiques relatives à la collecte de données
- Piloter ou participer au pilotage du projet en planifiant les différents sprints

Cours associés



Analysez les risques de votre projet

Moyenne 10 heures

Assurez le succès de votre projet en anticipant ses risques éventuels. Apprenez à analyser les risques de tous types de projets et à élaborer des plans de prévention pour y faire face.



Initiez-vous à la gouvernance des données

Moyenne 6 heures

Apprenez à mettre en place un bon plan d'action de gouvernance des données dans votre entreprise.



Travaillez efficacement en équipe

 Moyenne  10 heures

Le travail en équipe est une compétence essentielle... Et cela s'apprend ! Mobilisez des techniques comportementales pour profiter d'un levier formidable de l'équipe : l'intelligence collective.



alternance

**OpenClassrooms, 1^{ère} école en ligne
à proposer des formations diplômantes**

L'alternance en ligne permet à l'entreprise et à l'étudiant de trouver le **meilleur rythme** adapté à leurs besoins respectifs.

Côté étudiant :

une formation diplômante et 100% financée

L'alternance est une modalité de formation qui mêle acquisition des connaissances (en ligne) et des compétences (en entreprise).

Comment se déroule un parcours en alternance OpenClassrooms ?

- Après avoir choisi le parcours de formation qui lui convient, **l'étudiant bénéficie d'un accompagnement gratuit de 2 mois pour trouver son entreprise.**
- La formation peut se réaliser en contrat de professionnalisation ou d'apprentissage. Rendez-vous à la page suivante pour savoir lequel choisir !

Est ce que la formation en alternance a un coût pour l'étudiant ?

- La formation en alternance est gratuite ! Une rémunération est également prévue : jusqu'à 100% du SMIC en fonction de votre âge et de votre année de formation.
- Pour se lancer, une seule adresse : openclassrooms.com/fr/apprenticeship

Côté entreprise :

I'accès à des profils qualifiés toute l'année, en toute simplicité !

OpenClassrooms est la 1ère école en ligne à proposer des **formations en alternance, aux titres reconnus par l'État** de niveau Bac+2 à Bac+5, sur des **métiers en tension, très recherchés par les recruteurs**. Les formations en alternance peuvent être financées, soit par la **taxe d'apprentissage**, soit par les **OPCO (Opérateurs de Compétences)**.

Comment lancer des parcours d'alternance au sein de l'entreprise ?

- Pour chercher des profils à recruter en alternance ou lancer le recrutement d'un étudiant déjà identifié, n'hésitez pas à contacter :

job.placement@openclassrooms.com

- Les équipes d'OpenClassrooms vous accompagnent à chaque étape d'une démarche d'alternance, du dossier de financement au suivi de vos alternants.

Quand peut-on lancer une démarche d'alternance au sein de l'entreprise ?

- À tout moment : le contrat peut démarrer tout au long de l'année, il n'y a pas de date de rentrée imposée.



Pourquoi l'alternance en ligne ?

Les avantages de l'alternance OpenClassrooms sont nombreux : **date de début flexible, formations créées par des experts métiers, accompagnement personnalisé, formation financée...**

1. Gratuite pour l'étudiant et financée pour l'entreprise

La formation en alternance est gratuite pour l'étudiant et financée par l'OPCO de son entreprise pour les contrats éligibles. Dans certains cas, l'entreprise complète le financement si besoin. En parallèle l'alternant perçoit aussi son salaire.

2. Flexible et adaptée aux besoins de l'étudiant et de l'entreprise

Le contrat peut démarrer à tout moment de l'année et les jours dédiés à la formation sont flexibles en fonction de l'organisation de l'étudiant et de l'entreprise. Le rythme d'alternance prévoit 3 ou 4 jours en entreprise par semaine.

3. Une formation conçue pour une application directe en entreprise

Nos formations sont conçues par et pour l'entreprise, pour des métiers recherchés. Elles sont également certifiantes, avec des certifications inscrites au RNCP (Répertoire national des certifications professionnelles), reconnues par l'État.

4. Un accompagnement dédié à chaque étape

Pour les entreprises, des profils d'étudiants disponibles et présélectionnés par nos experts pour répondre à vos besoins en recrutement dans toute la France.

Pour l'étudiant, un accompagnement par nos conseillers pédagogiques à chaque étape de l'alternance et un point de suivi hebdomadaire par un mentor individuel.



Tout savoir sur les contrats d'alternance

Une question ? Un projet ?

Contacter : job.placement@openclassrooms.com

Contrat de professionnalisation

1 jour en formation / 4 jours en entreprise.

- L'entreprise embauche l'étudiant en CDD sur 12 ou 24 mois (selon le parcours de formation).
- La formation est financée par un OPCO. OpenClassrooms est référencé dans les principaux OPCO grâce à ses titres certifiés et sa certification Datadock.
- L'entreprise fait la démarche de demande de prise en charge auprès de son OPCO. Nos équipes sont présentes à chaque étape pour l'accompagner.
- L'étudiant est rémunéré **sur une base qui va de 65% à 100% du SMIC** (pour un étudiant de plus de 26 ans).
- Si l'étudiant a plus de 26 ans et est demandeur d'emploi, Pôle Emploi octroie une aide à l'emploi à l'entreprise.

Contrat d'apprentissage

2 jours en formation / 3 jours en entreprise.

- L'étudiant est embauché sur 12 ou 24 mois selon sa formation.
- Le contrat d'apprentissage concerne les personnes de 16 à 29 ans révolus (sans limite d'âge pour les publics RQTH, mais aussi les créateurs d'entreprise, les sportifs de haut niveau et l'encadrement de haut niveau).
- Le contrat d'apprentissage peut être signé par les employeurs dans le secteur public et les entreprises privées.
- La formation est **100% financée grâce à la taxe d'apprentissage** pour les entreprises du secteur privé. Pour le secteur public, une convention financière est établie entre OpenClassrooms et l'employeur.
- Une rémunération de l'apprenti est prévue : jusqu'à 100% du SMIC en fonction de son âge et de l'année de sa formation.
- Notre CFA vous accompagne dans toutes les démarches administratives concernant la mise en place de contrat d'apprentissage.