



Data Scientist

Analysez des données pour identifier des tendances et faire des prédictions...
Maîtrisez la Data Science !



Durée de la formation :
600 heures



Emploi garanti**



Durée de la formation en alternance :
18 mois



Temps d'études conseillé à temps plein :
9 mois



Master's-level diploma*

*Data scientist - NSF 326 - Master's-level diploma - enregistré au RNCP par arrêté du 3/27/20 publié au JO du 3/20/20

**Conditions disponibles sur la page <https://openclassrooms.com/fr/job-guarantee>

Financez ce parcours grâce à vos crédits CPF directement depuis la plateforme Mon Compte Formation. [Plus d'infos](#) (Temps complet - 9 mois). [Plus d'infos](#) (Temps partiel - 18 mois).

Quel est le rôle d'un Data Scientist ?

Les entreprises produisent une quantité astronomique de données. **Être capable de les analyser et 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 les équipes métiers, vous mènerez à bien un projet data, de la collecte des données à la mise en production de vos algorithmes.

Ce que vous saurez faire

- Collecter et préparer les données en vue de l'analyse
- Programmer des algorithmes de Machine Learning à l'aide du langage Python
- Déployer des algorithmes dans le cloud avec les outils du Big Data
- Communiquer les résultats à des spécialistes ou des néophytes



CentraleSupélec

Parcours en partenariat avec CentraleSupélec

Votre rémunération

Rémunérations moyennes pour le métier de Data Scientist :

- Débutant : 35 000 € à 45 000 € annuels bruts
- Expérimenté : 45 000 € à 65 000 € annuels bruts

(Source : <http://datarecrutement.fr/etude-salaire/>)

Ces profils sont très recherchés, majoritairement en CDI. En freelance, les Data Scientist facturent jusqu'à 1000€ par jour de travail.

Retrouvez sur [cette page](#) les indicateurs de performance des formations OpenClassrooms.

Prérequis

Niveau de langue : Pour les apprenants étrangers, un niveau de français B1-B2 (utilisateur indépendant) est conseillé pour la réussite de la formation.

Matériel : Accès à un ordinateur (PC ou Mac), muni d'un microphone, une webcam et 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](#).

Ce parcours nécessite également 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.

Niveau requis : Prépa scientifique ou Bac + 2 en mathématiques.

Prérequis techniques :

- Maths (analyse réelle, algèbre, probabilités, statistiques).
- Notions d'informatique (algorithmique, base de données, terminal).

Si vous n'avez pas réalisé une prépa scientifique ou validé un bac+2 en mathématiques:

- réalisez notre [test de positionnement](#) pour évaluer votre niveau.
- remplissez ce [formulaire d'admission](#) et précisez votre score au test de positionnement dans la case où vous renseignez vos motivations. Notez que ce test ne constitue pas une validation.

Votre orientation

Ce parcours donne accès aux métiers suivants :

- Data Scientist
- Data Analyst, Business Analyst, BI Analyst

Quel parcours Data est fait pour vous ?

[Data Analyst](#) : Vous débuterez dans la data en analysant des données et en réalisant des reportings et des dashboards.

Data Scientist : Vous avez un bagage mathématique, et vous souhaitez réaliser des analyses poussées à l'aide d'algorithmes.

[Ingénieur Machine Learning](#) : Vous avez un solide bagage mathématique et vous souhaitez développer des algorithmes de machine learning avancés.

[Data Architect](#) : Vous avez un bagage informatique et vous souhaitez mettre en place l'architecture et les outils de traitement des données.

Si vous ne possédez pas le niveau de prérequis attendu et/ou que vous êtes déjà en activité, la durée de votre formation sera allongée.

Projet 1 - 10 heures

Définissez votre stratégie d'apprentissage

Vous embarquez sur un grand parcours d'apprentissage ! Équipez-vous des outils et des bonnes pratiques dont vous aurez besoin tout au long de vos cours et de vos projets.

Compétences cibles

- Construire pas à pas son projet professionnel

Cours associés



Apprenez à apprendre

 Facile  6 heures

Être capable d'apprendre vite et bien est une compétence clé dans n'importe quel domaine. Suivez ce cours pour améliorer votre capacité d'apprentissage !



Comprendre le Web

 Facile  6 heures

Le monde du Web vous intéresse, mais vous n'y connaissez rien ? Vous décrochez quand vous entendez du vocabulaire geek ? Imprégnez-vous de la culture du Web avec ce cours !



Développez vos soft skills

 Facile  6 heures

Les soft skills, comme l'adaptabilité, la collaboration, la résolution de problèmes, sont des compétences de plus en plus recherchées. Développez ces compétences transversales et comportementales.



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 2 - 60 heures

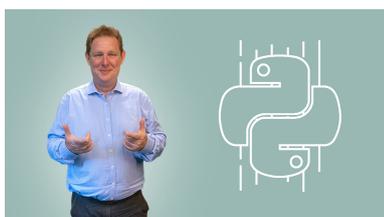
Analysez des données de systèmes éducatifs

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

- Utiliser un notebook Jupyter pour faciliter la rédaction du code et la collaboration
- Manipuler des données avec des bibliothèques Python spécialisées
- Mettre en place un environnement Python
- Effectuer une représentation graphique à l'aide d'une bibliothèque Python adaptée
- Maîtriser les opérations fondamentales du langage Python pour la Data Science

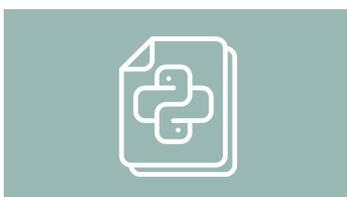
Cours associés



Initiez-vous à Python pour l'analyse de données

 Facile  12 heures

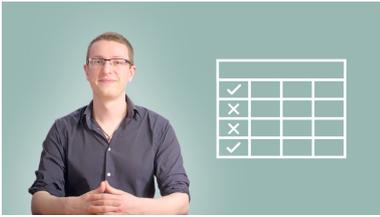
Dans ce cours, vous apprendrez un langage de programmation indispensable pour l'analyse de données : Python. Nous aborderons ensemble les notions fondamentales de la programmation Python, à l'aide d'exemples simples et d'exercices pratiques.



Découvrez les bibliothèques Python pour la Data Science

 Moyenne  10 heures

Python s'est imposé comme le langage incontournable pour la Data Science et le Machine Learning, avec de nombreuses bibliothèques spécialisées. Découvrez les notebooks Jupyter et les bibliothèques incontournables que sont Numpy, Matplotlib et Pandas.



Décrivez et nettoyez votre jeu de données

■ Moyenne ⌚ 15 heures

Prêt à entrer dans l'univers de la statistique descriptive ? Avec ce cours, vous découvrirez comment se compose un jeu de données. Vous serez capable de le nettoyer et le décrire en vue de l'analyser.



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.

Projet 3 - 70 heures

Concevez une application au service de la 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

- Communiquer ses résultats à l'aide de représentations graphiques lisibles et pertinentes
- Effectuer des opérations de nettoyage sur des données structurées
- Effectuer une analyse statistique multivariée
- Effectuer une analyse statistique univariée

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 entraînerez votre premier modèle !

Anticipez les besoins en consommation électrique 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

- Adapter les hyperparamètres d'un algorithme d'apprentissage supervisé afin de l'améliorer
- Évaluer les performances d'un modèle d'apprentissage supervisé
- Mettre en place le modèle d'apprentissage supervisé adapté au problème métier
- Transformer les variables pertinentes d'un modèle d'apprentissage supervisé

Cours associés



Évaluez 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 appréhendez 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'algorithmes très populaire... les réseaux de neurones !



Modélisez vos données avec les méthodes ensemblistes

■ Moyenne ⌚ 15 heures

Décuplez 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

- Mettre en place le modèle d'apprentissage non supervisé adapté au problème métier
- Transformer les variables pertinentes d'un modèle d'apprentissage non supervisé
- Adapter les hyperparamètres d'un algorithme non supervisé afin de l'améliorer
- Évaluer les performances d'un modèle 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).

Projet 6 - 100 heures

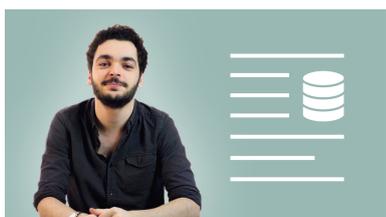
Classifiez automatiquement des biens de consommation

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

- Prétraiter des données image pour obtenir un jeu de données exploitable
- Prétraiter des données texte pour obtenir un jeu de données exploitable
- Représenter graphiquement des données à grandes dimensions
- Mettre en œuvre des techniques de réduction de dimension

Cours associés



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 modélisez 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).

Projet 7 - 120 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

- Déployer un modèle via une API dans le Web
- Réaliser un dashboard pour présenter son travail de modélisation
- Rédiger une note méthodologique afin de communiquer sa démarche de modélisation
- Utiliser un logiciel de version de code pour assurer l'intégration du modèle
- Présenter son travail de modélisation à l'oral

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 - 70 heures

Déployez un modèle dans le cloud

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

Compétences cibles

- Utiliser les outils du cloud pour manipuler des données dans un environnement Big Data
- Identifier les outils du cloud permettant de mettre en place un environnement Big Data
- Paralléliser des opérations de calcul avec Pyspark

Cours associés



Découvrez le cloud avec Amazon Web Services

 Moyenne  20 heures

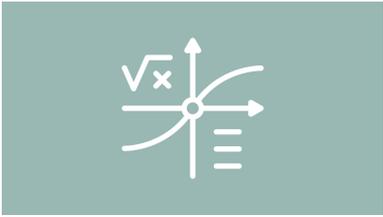
Vous avez entendu parler du cloud et notamment d'Amazon Web Services (AWS), le leader du cloud ? Venez découvrir comment l'utiliser dans ce cours d'introduction !



Concevez des architectures Big Data

 Moyenne  6 heures

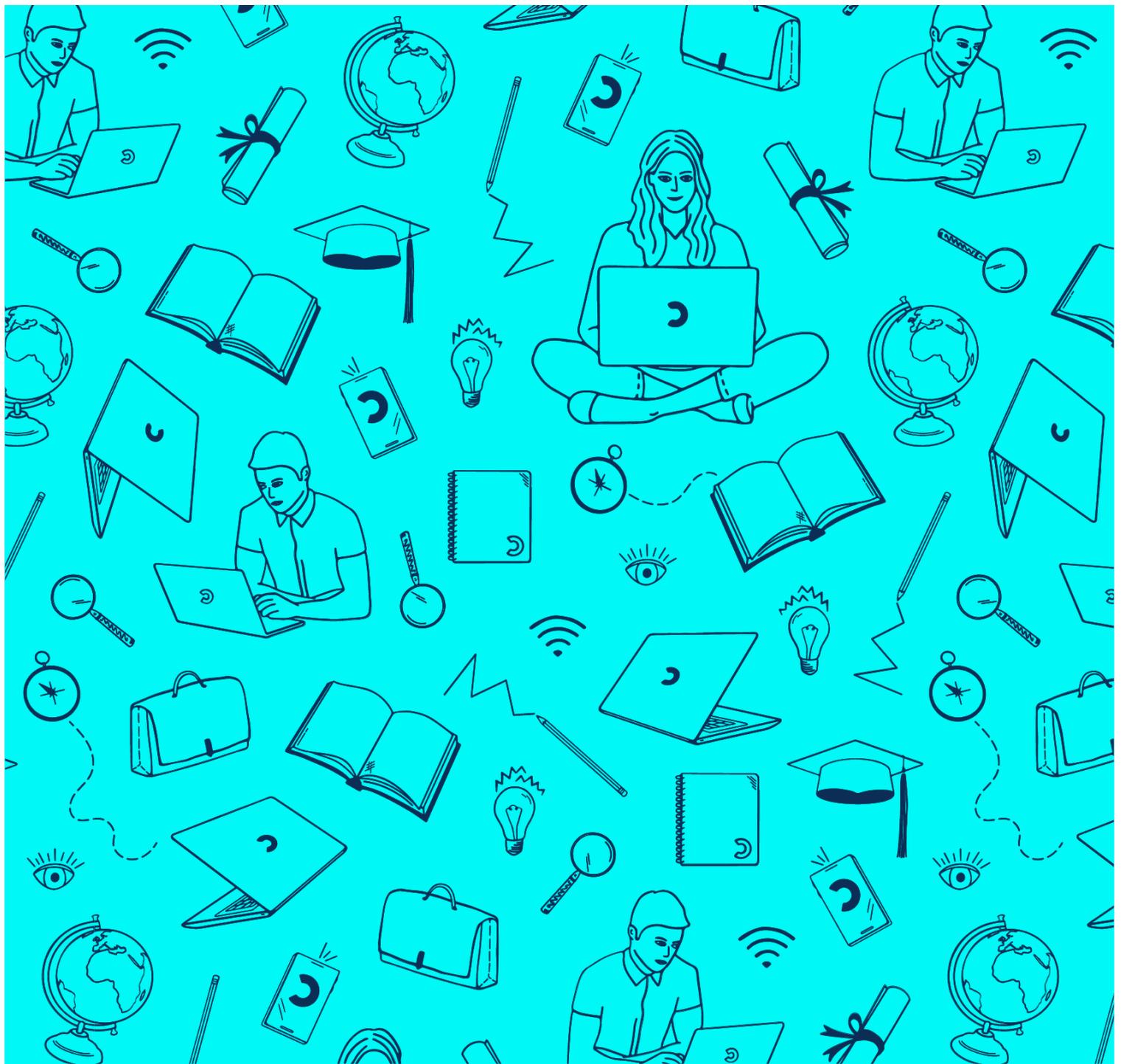
Nous sommes à l'âge d'or du Big Data et les Data Architects disposent de tous les outils dont ils ont besoin pour gérer des données massives. Mais comment les assembler ? Familiarisez-vous avec une vision d'ensemble pour la conception d'architectures Big Data complètes.



Réalisez des calculs distribués sur des données massives

 Moyenne  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.



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 :

l'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. Une formation gratuite pour l'étudiant et financée pour l'entreprise

- La formation est gratuite.
- Dans le cadre d'un **contrat de professionnalisation, elle est 100% financée** par votre OPCO.
- Dans le cadre d'un **contrat d'apprentissage**, elle est financée par la taxe d'apprentissage que paient les entreprises privées.

3. Une formation de qualité, conçue pour une application directe au sein de l'entreprise.

- Les formations OpenClassrooms sont **professionnalisantes**. Elles sont conçues par et pour l'entreprise, sur des métiers recherchés.
- Les formations OpenClassrooms sont de qualité, avec des **diplômes inscrits au RNCP** (Répertoire National des Certifications Professionnelles) et dont les niveaux sont reconnus par l'État.
- Une plateforme collaborative favorise les échanges entre apprentis et aide à la recherche d'emploi.
- Des espaces de coworking sont proposés dans toute la France.

2. Une solution flexible, adaptée aux besoins de l'étudiant et de l'entreprise

- Le contrat peut démarrer **à tout moment de l'année**.
- Il est possible de faire soit un contrat de professionnalisation, soit un contrat d'apprentissage.
- Le(s) **jour(s)** dédié(s) à la **formation** est (sont) **flexible(s)** en fonction de l'organisation de l'étudiant et de l'entreprise.

4. Un accompagnement dédié à chaque étape, pour l'étudiant comme pour l'entreprise

- Des **profils d'étudiants disponibles et pré-sélectionnés par nos experts pour répondre à vos besoins en recrutement sur toute la France**.
- Un accompagnement par nos conseillers pédagogiques **à chaque étape de l'alternance**.
- Un point de suivi hebdomadaire par un **mentor individuel** pour **motiver l'étudiant et l'accompagner** pendant toute sa formation.



Tout savoir sur les contrats d'alternance

Une question ? Un projet ?

Contactez : 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.