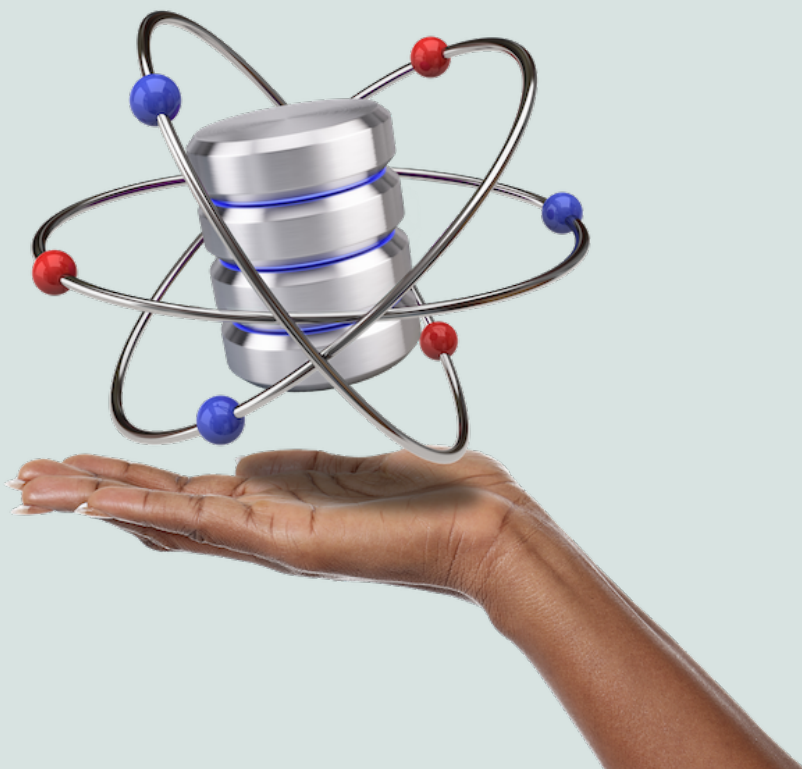


Data



Data Engineer

 Mentor individuel  Diplôme de niveau 7 (bac +5) *

Concevez et gérez les données de demain

 PÉRIODE DE FORMATION

9 mois à temps plein

18 mois en alternance**

 DURÉE DE LA FORMATION

603 heures supervisées

OPENCLASSROOMS

La formation demande un investissement en temps estimé à 1206 heures : 603 heures de formation supervisée (projets encadrés par des mentors) et 603 heures de formation guidée (cours et des ressources pédagogiques). En alternance, la durée totale ne comprend pas le temps passé en entreprise.

La période de formation peut être rallongée en cas de formation à temps partiel. La durée est estimée et dépend du niveau d'entrée en formation, de la disponibilité, du temps alloué par semaine et des capacités et rythmes d'apprentissage de l'étudiant.

*Expert en ingénierie et science des données - 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 31/10/2024

** Vérifiez l'éligibilité du parcours en fonction de votre contrat d'alternance (professionnalisation ou apprentissage).

Devenir Data Engineer : concevoir et gérer les données de demain !

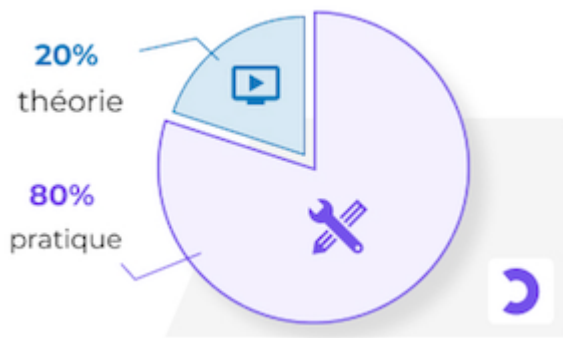
Le Data Engineer est un expert en gestion et transformation de données. Son rôle est crucial pour la mise en place de pipelines de données fiables et efficaces, permettant aux entreprises de tirer pleinement parti de leurs informations.

Ce que vous apprendrez :

- **Effectuer l'analyse prédictive** et le **machine learning** en transformant des données brutes en insights exploitables. Vous comprendrez et anticiperez les tendances.
- **Améliorer les processus décisionnels** et apporter une valeur stratégique ajoutée à votre organisation.
- **Mettre en place des architectures de données** afin de structurer et organiser les flux d'informations. Cela inclut la sélection des technologies appropriées et la création d'architectures évolutives.
- **Développer des pipelines de données robustes** pour garantir l'intégrité et la qualité des données. Cela comprend l'automatisation des processus de collecte et de transformation des données.
- **Optimiser la gestion des bases de données** afin d'assurer des performances optimales. Vous surveillerez les performances des bases de données. Vous apprendrez à résoudre les problèmes de goulot d'étranglement en vous assurant que les systèmes fonctionnent.
- **Implémenter une solution de stockage de données** pour répondre aux besoins spécifiques des projets. Vous devrez choisir entre différentes options de stockage : bases de données relationnelles, NoSQL, data lakes, etc., en fonction des exigences du projet.
- **Collaborer avec les équipes de data science et d'analyse** pour comprendre leurs besoins et fournir des données prêtes à l'emploi.
- **Appliquer des techniques de machine learning** pour analyser les données et développer des modèles prédictifs. Vous utiliserez des algorithmes de machine learning pour résoudre des problèmes complexes et extraire des insights.
- **Gérer des projets de bout en bout** en définissant les objectifs, en planifiant les étapes et en assurant la livraison des solutions de données dans les délais impartis. Vous coordonnerez les équipes et les ressources pour mener à bien les projets tout en respectant les contraintes.

Vous maîtriserez les technologies et les langages de programmation les plus demandés : SQL, Python, Pandas, Sk-Learn, Airbyte, Spark, Kafka, Kestra et beaucoup d'autres.

Nos formations vous permettent d'apprendre par la pratique



Le rôle de Data Engineer est crucial en entreprise : Data Engineer est classé **12ème sur 30** des métiers digitaux qui ont le plus fort besoin de talents en France.*

*(Grande école du numérique, avril 2023)

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



Fayçal, un des mentors sur un parcours Data

« Parfois, les étudiants ont peur du jugement des autres, ou tout simplement d'échouer. Mon rôle de mentor, c'est de leur rappeler qu'ils sont en formation, et qu'ils sont là pour tester le plus de choses possible. Plus ils essaient, plus ils vont avoir des retours sur leur travail. Cette approche permet de nourrir leur curiosité et de les aider à aller plus vite ! » Fayçal

Autonome, mais jamais seul !

Chez OpenClassrooms, vous pourrez compter sur de nombreux soutiens :

- Les mentors, experts du domaine
- Les conseillers pédagogiques dédiés à votre réussite
- Une communauté d'étudiants sur Slack

Vos opportunités de carrière et d'emploi à la suite de ce programme

En suivant notre formation en ligne, vous pourrez exercer les métiers de **Data Engineer** ou **Analytics Engineer**.

Les salaires du marché pour le poste de Data Engineer

Le salaire dépend notamment du niveau d'ancienneté*

- 44000 € : brut annuel pour un profil **débutant**
- 51000 € : brut annuel pour un profil **confirmé**

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

** Glassdoor*

Et si vous suiviez cette formation en alternance ?

La formation de Data Engineer 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 pré-requis pour postuler

Pour accéder directement à la certification, les candidats doivent obligatoirement justifier d'un des **pré-requis** 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.

Langues :

- Un niveau B2 minimum en français
- Un niveau B2 est recommandé en anglais

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.

Pour tous les candidats, la procédure de sélection s'effectue au travers d'un formulaire de candidature validant les pré-requis académiques ou dérogatoires, ainsi que l'adéquation du projet professionnel avec les objectifs de la formation. Seront collectées via ce formulaire les pièces justificatives constituant le dossier d'admission. Si nécessaire, un entretien individuel de motivation est organisé.

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.

Projet 1 - 13 heures

Découvrez votre formation de Data Engineer

Mettez-vous dans les meilleures conditions pour réussir votre parcours : projetez-vous dans votre formation, définissez votre planning et appropriez-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.



Trouvez votre premier emploi en data

 Facile  6 heures

Trouver un emploi dans la data ça se prépare ! Ce cours est là pour vous aider à adopter la bonne approche... et vous aider à le décrocher !



Optimisez votre apprentissage avec l'Intelligence Artificielle



Facile



6 heures

Utiliser l'IA en gardant un esprit critique, pour acquérir plus rapidement des compétences, gagner en productivité et mieux organiser votre planning d'apprentissage.

Analysez les données de systèmes éducatifs

Renforcez vos fondamentaux de Python pour la Data Science en analysant les données de systèmes éducatifs

Compétences cibles


- Appliquer des analyses statistiques descriptives et naviguer visuellement dans les données
- Configurer l'environnement de travail nécessaire à l'exploitation des données
- Corriger les anomalies manuellement et à l'aide d'outils adaptés

Cours associés

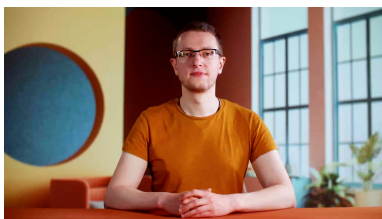


Initiez-vous à Python pour l'analyse de données

 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.

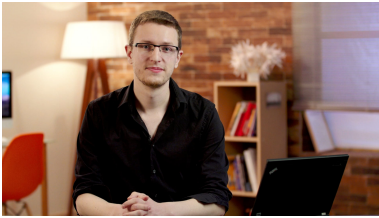


Découvrez les bibliothèques Python pour la Data Science

 Moyenne

 8 heures

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

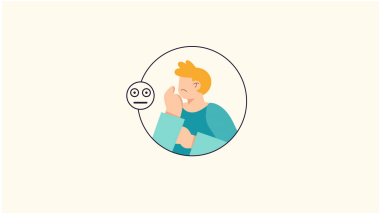


Nettoyez et analysez votre jeu de données

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.



Mettez en place votre environnement sous Poetry

Facile

2 heures

À l'aide de ce cours Poetry, mettez en place votre environnement virtualisé pour Python dans le cadre de vos projets Data et maîtrisez les fonctionnalités avancées.

Entraînez-vous avec SQL et créez votre BDD

Renforcez vos compétences en SQL pour réussir votre parcours de Data Engineer



Compétences cibles

- Créer des bases de données relationnelles afin de contenir les données
- Structurer les données et leurs relations en cohérence avec leurs caractéristiques

Cours associés



Modélisez vos bases de données

 Moyenne  8 heures

Apprenez à modéliser vos bases de données avec des diagrammes de classe UML et à passer du modèle conceptuel de données au modèle relationnel.



Requêtez une base de données avec SQL

 Moyenne  12 heures

Initiez-vous à la modélisation relationnelle et construisez des requêtes SQL avec des fonctions pertinentes pour alimenter vos data analyses avec les bonnes données.

Projet 4 - 40 heures

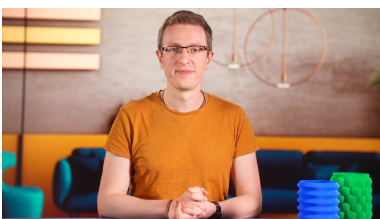
Auditez un environnement de données

Faites du retro engineering sur une infrastructure de données et présentez votre rapport

Compétences cibles

- Analyser la donnée afin d'en déterminer l'adéquation au regard des besoins identifiés
- Effectuer un prototype de la solution afin d'en confirmer la faisabilité technique.
- Identifier et évaluer les risques de la solution data
- Identifier une solution technique afin de répondre aux besoins
- Présenter le projet data et expliquer ses choix afin d'en démontrer la pertinence

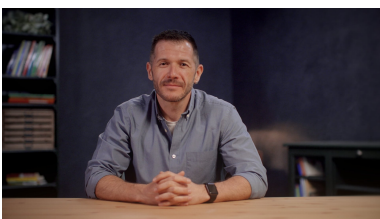
Cours associés



Requêtez une base de données avec SQL

 Moyenne  12 heures

Initiez-vous à la modélisation relationnelle et construisez des requêtes SQL avec des fonctions pertinentes pour alimenter vos data analyses avec les bonnes données.





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.



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 5 - 40 heures

Maintenez et documentez un système de stockage des données sécurisé et performant

Aidez une entreprise à maintenir et à documenter son système de stockage des données sécurisé pour le rendre performant et optimisé sur le plan énergétique

Compétences cibles

- Définir et formaliser les processus de traitement et de stockage des données
- Mettre en place un système d'authentification afin de garantir la sécurité des données
- Configurer l'environnement de travail

Cours associés



Optimisez votre déploiement en créant des conteneurs avec Docker

 Moyenne  8 heures

Installez Docker, lancez des conteneurs avec des images stockées sur Docker Hub, écrivez un Dockerfile et utilisez Docker-Compose pour orchestrer vos conteneurs !

Anticipez les besoins en consommation de bâtiments

Aidez une grande ville à atteindre son objectif de neutralité en émissions de carbone

Compétences cibles

- Entraîner un modèle d'apprentissage
- Évaluer le modèle d'apprentissage
- Exposer les résultats aux directions (via une API) en vue de leur exploitation
- Identifier ou créer une API compatible et l'intégrer pour permettre l'accès aux résultats
- Préparer et transformer des données afin de les adapter au modèle d'apprentissage.
- Présenter ses résultats

Cours associés



Initiez-vous au Machine Learning

 Moyenne  10 heures

Découvrez dans ce cours les techniques incontournables du Machine Learning. Leurs applications sont variées et très prometteuses !



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.



Devenez un expert de Git et GitHub

■ ■ Moyenne ⌚ 6 heures

Devenez un expert de Git et GitHub : utiliser une méthode et des commandes pour gérer les conflits, mettez en place des workflows, des hooks et utiliser un client GUI.



Maîtrisez l'apprentissage supervisé

■ ■ Moyenne ⌚ 10 heures

Maîtrisez les concepts avancés de l'apprentissage supervisé ainsi que les bases du MLOps.

Concevez et analysez une base de données NoSQL

Combinez vos compétences en Python et en bases de données NoSQL (MongoDB) pour manipuler, requêter et analyser de grands volumes de données.

Compétences cibles

- Construire et manipuler des bases de données
- Élaborer un processus
- Schématiser une architecture physique on premise
- Schématiser une solution de collecte et de stockage des données

Construisez et testez une infrastructure de données

Construisez votre premier système complexe de gestion des données de la modélisation au test de fonctionnalité

Compétences cibles

- Définir et formaliser les processus de traitement et de stockage des données
- Établir et exécuter un processus de test du SGDB
- Installer et paramétrer un système de gestion de base de données et un outil d'extraction
- Installer et tester l'infrastructure de gestion des données
- Modéliser une infrastructure compatible avec le SI
- Structurer l'architecture des données et concevoir des BDD

Cours associés



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.



Initiez-vous au logiciel Airbyte pour intégrer vos données

 Facile  6 heures

Installez Airbyte sur votre ordinateur et utilisez les principales fonctionnalités de cet outil pour extraire vos données.

Modélisez une infrastructure dans le cloud

Modélisez votre infrastructure dans le cloud

Compétences cibles

- Identifier et sélectionner les composants nécessaires à une infrastructure de données
- Documenter son travail
- Représenter visuellement une infrastructure de gestion des données
- Transformer des données afin de les adapter à leur utilisation finale.
- Charger des données afin de les stocker dans un emplacement adapté
- Évaluer la compatibilité des composants avec l'environnement SI de l'organisation.
- Extraire des données issues de toutes sources confondues pour les traiter ou les déplacer

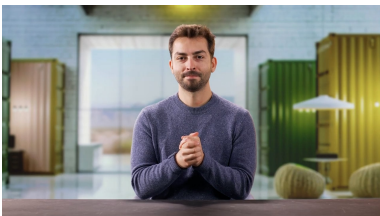
Mettez en place un pipeline d'orchestration des flux

Construisez le pipeline des flux de données d'une entreprise dans le secteur technologique

Compétences cibles

- Mettre en place l'ordonnancement des flux des données pour programmer leur déclenchement
- Créer des processus de test afin de valider la mise en production des pipelines

Cours associés



Orchestrez vos workflows avec Kestra

 Moyenne  6 heures

Et si vous pouviez simplifier vos workflows en quelques clics ? Maîtrisez Kestra et découvrez comment automatiser et optimiser vos flux de données facilement !

Concevez et déployez un système RAG

Réalisez un système Retrieval-Augmented Generation fonctionnel basé sur LangChain et le modèle Mistral, soutenu par une base vectorielle Faiss.

Compétences cibles

- Identifier ou créer un modèle d'apprentissage adapté aux contraintes et besoins métiers
- Mettre en place un processus de nettoyage afin d'améliorer la qualité des données
- Configurer l'environnement de travail nécessaire à l'exploitation des données

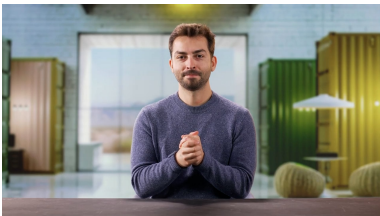
Gérez un projet d'infrastructure

Aidez une plateforme d'entraînement en course à pied à automatiser leur infrastructure de données

Compétences cibles

- Ajuster les requêtes et les paramétrages des pipelines
- Concevoir un système de suivi des flux de données pour valider leur intégration en continu
- Mettre en place un pipeline de données automatisé et adapté au besoin

Cours associés



Orchestrez vos workflows avec Kestra

 Moyenne  6 heures

Et si vous pouviez simplifier vos workflows en quelques clics ? Maîtrisez Kestra et découvrez comment automatiser et optimiser vos flux de données facilement !

Passez votre système IA du POC au MVP et réalisez votre portfolio de Data Engineer

Organisez le déploiement du RAG d'une entreprise et démontrez vos compétences de Data Engineer dans un portfolio

Compétences cibles

- Apporter un appui stratégique et méthodologique pour faciliter la prise de décision
- Collecter les besoins métiers et analyser le contexte de l'organisation
- Conduire les actions et les échanges entre les parties prenantes du projet data
- Contrôler et analyser le projet data en termes de délais, coûts, livrables et performance
- Définir les modalités de réalisation et de suivi du projet data et le planifier
- Gérer plusieurs priorités et s'adapter à un environnement en constante évolution.
- Identifier de nouvelles opportunités, solutions ou pratiques dans le champ de la data
- S'exprimer à l'oral en respect des bonnes pratiques du métier

Cours associés



Trouvez votre premier emploi en data



Facile



6 heures

Trouver un emploi dans la data ça se prépare ! Ce cours est là pour vous aider à adopter la bonne approche... et vous aider à le décrocher !



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.



L'alternance par OpenClassrooms. Pratique et pratique.

Apprenez où que vous soyez

Pas besoin de déménager pour se former : choisissez une entreprise près de chez vous et formez-vous en ligne

Côté étudiant :

L'alternance à tout âge avec OpenClassrooms

Démarrez une nouvelle carrière avec nos formations en alternance 100% en ligne ! Véritables accélérateurs de carrière, elles sont diplômantes, rémunérées et financées par des entreprises partout en France. Débutez où vous voulez, pendant toute l'année.

La pédagogie et l'expérience OpenClassrooms, les clés de votre réussite

- **Apprenez où que vous soyez**

Pas besoin de déménager pour se former : choisissez une entreprise près de chez vous et formez-vous en ligne.

- **Travaillez sur des projets
professionnalisants**

Réalisez des projets concrets, issus de scénarios métiers, directement applicables en entreprise.

- **Un mentor pour vous accompagner**

Bénéficiez chaque semaine des conseils d'un expert du métier qui vous aide à progresser tout au long de votre formation.

- **Un salaire et aucun frais**

L'entreprise paie votre formation et vous verse un salaire mensuel, calculé selon votre situation personnelle.

Côté employeur :

Recrutez et formez les talents de demain avec l'alternance

Recrutez parmi notre base de candidats et formez vos futurs talents sur les métiers en tension grâce à l'alternance.

Accédez gratuitement aux alternants OpenClassrooms

- **Découvrez des profils motivés et de qualité**

Trouvez des candidats qui correspondent réellement à vos besoins.

- **Recrutez rapidement grâce à notre base d'alternants.**

Dénichez vos futurs talents via votre espace recruteur.

- **Réalisez vos démarches administratives facilement**

Finie la paperasse : nos équipes s'occupent aussi de l'administratif.



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. Un salaire et une formation financée par l'entreprise, qui dit mieux ?

La formation en alternance, c'est 0 frais pour l'étudiant car financée par l'entreprise. Et comme pour toute alternance, vous percevrez un salaire durant votre formation.

2. Un rythme d'alternance flexible : pratique pour vous et pour votre entreprise

Votre contrat peut démarrer à tout moment de l'année, plus besoin d'attendre septembre ou janvier ! Le rythme d'alternance prévoit 3 ou 4 jours en entreprise par semaine et les jours de formation peuvent être adaptés.

3. Une formation au plus proche de la réalité du métier

Nos formations sont conçues avec des experts reconnus dans leur domaine, pour répondre aux besoins des entreprises. Elles sont inscrites au Répertoire national des certifications professionnelles (RNCP) et sont reconnues par l'Etat.

4. En ligne, mais toujours bien accompagné

Parce que se former n'est jamais facile, vous êtes suivi individuellement par un mentor, qui vous aide à progresser. Notre équipe de conseillers pédagogiques est aussi là pour vous accompagner à chaque étape de votre parcours.



Tout savoir sur les contrats d'alternance

Une question ? Un projet ?

Contacter : job.placement@openclassrooms.com

Contrat de professionnalisation

4 jours par semaine (jours au choix) avec présence réduite à 3 jours 1 à 2 fois par mois.

- 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, France Travail octroie une aide à l'emploi à l'entreprise.