

# Data Scientist



## Le Campus Numérique in the Alps Formation Data Scientist

### Objectifs de la formation

La formation Data Scientist est un parcours de 10 à 14 mois formant aux métiers de la Data en permettant à des diplômés en reconversion de développer des compétences complémentaires à leur spécialité en analyse de données et en développement de l'intelligence artificielle.

La formation est adaptée aux compétences recherchées par les entreprises en analyse de données, en vue de favoriser leur insertion professionnelle sur des métiers en tension. Nous formons des experts métiers de haut niveau avec des compétences complémentaires à leur cœur de métier en récolte, analyse et exploitation des données.

Le Data Scientist est responsable de l'extraction et l'interprétation de données mises à disposition par les services web et les autres canaux digitaux (big data), dans l'objectif d'aider l'entreprise à prendre des décisions stratégiques.

Les métiers de Data Scientist s'articule autour de 5 activités principales :

- Récupérer et analyser les données pertinentes liées au processus de décision de l'entreprise, à la vente ou encore aux données client ;
- Construire les algorithmes permettant d'améliorer les résultats de recherches et de ciblage ;
- Elaborer les modèles prédictifs permettant d'anticiper l'évolution des données et les tendances relatives à l'activité de l'entreprise ;
- Modéliser les résultats d'analyse des données pour les rendre lisibles et exploitables aux décideurs ;
- Fournir des outils d'aide à la décision et des recommandations stratégiques aux équipes métiers.

### Compétences développées

Le programme aborde les technologies et compétences indispensables pour l'analyse de données : algorithmique, Python, Machine Learning, bases de données, outils de visualisation des données, outils statistiques... et les soft skills nécessaires en entreprise. Les méthodes d'apprentissage « apprendre à apprendre » permettent aux apprenants de se former en autonomie tout au long de leur carrière pour suivre les évolutions technologiques.

Compétences développées à l'issue de la formation :

- Maîtrise des outils statistiques pour l'analyse de données
- Maîtrise des outils analytiques pour l'analyse et l'exploration des données
- Maîtrise des langages de programmation utilisés en analyse de données (Python)
- Maîtrise des méthodes de Machine Learning

- Gestion de données non structurées
- Récolte, analyse et exploitation de grands jeux de données
- Interprétation des données et formulation de recommandations stratégiques

## Méthodes pédagogiques

Le Campus adopte une approche par compétences, les modules de formation étant construits sur la base de compétences à acquérir : savoirs, savoir-faire, savoir-être, outils, méthodologies... Cette approche permet d'assurer une formation professionnalisante avec le développement de compétences opérationnelles correspondant aux exigences des métiers visés.

Le modèle de pédagogie par projet mis en place dans les modules de formation permet de rendre les stagiaires acteurs de la formation et de l'apprentissage. L'autonomie développée dans les modules, la transversalité marquée des enseignements et une pratique pédagogique hautement professionnalisante (développement d'applications dès les premières heures passées au Campus et apprentissage en alternance) sont au cœur du projet pédagogique.

## Modalités d'évaluation et diplôme visé

La formation est non diplômante, elle donne lieu à une attestation de réussite délivrée par Le Campus Numérique in the Alps.

La formation Data Scientist est une formation labellisante par l'Institut MIAI de l'UGA, sous le label « [Applications de l'IA](#) ».

**Modalités d'évaluation :**

- **Validation des compétences** (auto-validation et validation par un formateur professionnel de l'ensemble des compétences associées à chacun des modules de formation)
- **Assiduité**
- **Réalisation de 4 rapports d'activité en entreprise**
- **Réalisation d'un rapport de mission professionnelle**
- **Soutenance devant un jury mixte** (professionnels du secteur numérique et de représentants académiques)

## Durée de la formation

La formation se décompose en deux phases :

- **6 mois de formation intensive au Campus (770 heures) permettant d'acquérir le socle de compétences techniques et de travail en équipe ;**
- **4 à 12 mois en entreprise comprenant 280 heures de formation afin d'élargir et d'approfondir les compétences techniques et les soft skills développées.**

## Modalités d'accès et prérequis

Tous les candidats titulaires d'un diplôme de niveau Master ou Doctorat (mathématiques, informatique, ingénierie, physique, chimie, statistiques, matériaux, sciences de la terre, économie, génie civil, génie industriel, mécanique, nanotechnologies, traitement du signal, astrophysique, biomédical, sciences humaines et sociales...), pratiquant au moins un langage de programmation et

étant demandeurs d'emploi peuvent s'inscrire à la formation. Les critères de sélection sont basés sur :

- Motivation, cohérence et clarté du projet professionnel : vouloir développer des compétences complémentaires au cœur de métier en analyse de données
- Aptitudes scientifiques (mathématiques, probabilités, apprentissage de la programmation)
- Aptitudes métier, qualités professionnelles et aptitudes au travail en équipe

Processus de sélection :

- Envoi d'un CV, d'une lettre de motivation et d'une vidéo pitch
- Immersion dans un travail collaboratif non numérique (piscine de 2 jours)

Délais d'accès : la formation débute en moyenne deux mois après l'ouverture des candidatures.

## Prix de la formation

Gratuit grâce aux partenaires financeurs de la formation et les fonds de financement de la formation.

## Lieu de formation

La formation Data Scientist est réalisée :

### Campus de Grenoble

Le Totem

16 boulevard Maréchal Lyautey

38000 Grenoble

07 50 66 51 51 / [audrey.graffagnino@le-campus-numerique.fr](mailto:audrey.graffagnino@le-campus-numerique.fr)

## Accessibilité aux personnes en situation de handicap

La formation est accessible aux personnes en situation de handicap, avec une adaptation possible des modalités pédagogiques et d'évaluation. Le Campus Numérique est engagé dans la [démarche H+ Formation](#) de la Région Auvergne Rhône-Alpes.

Référents handicap : Audrey Graffagnino ([audrey.graffagnino@le-campus-numerique.fr](mailto:audrey.graffagnino@le-campus-numerique.fr)) / Pauline Lacour ([pauline.lacour@le-campus-numerique.fr](mailto:pauline.lacour@le-campus-numerique.fr))

## Suites de parcours

A visée professionnalisante, la formation vous permet d'exercer le métier de Data Scientist.

## Programme Pédagogique (prévisionnel)

Socle		
Module	Durée	Objectifs pédagogiques
Positionnement	4 heures	Bilan, diagnostic et définition du parcours individualisé de formation
Outils numériques pour formation à distance	3 heures	Mise en place de l'environnement de formation à distance, formation sur les outils numériques utilisés à distance (installation de logiciels, vérification de connexion internet, tests techniques)
TRE	21 heures	Maîtriser les techniques de recherche d'emploi, de rédaction du CV et de la lettre de motivation. Connaître les techniques de création d'un profil professionnel numérique (LinkedIn notamment). Connaître les techniques et stratégies nécessaires pour s'engager dans un processus actif de recherche d'emploi. Maîtriser les techniques de valorisation des profils et de présence numérique.
Développement Python & Algorithmie	91 heures	Maîtriser les concepts algorithmiques de la programmation (variables, opérateurs, conditions, boucles et tableaux). Connaître la méthodologie de conception d'une solution informatique. Analyser la complexité d'un algorithme. Maîtriser les bases de la programmation par la pratique à l'aide du langage Python (variables, instructions, conditions, boucles, fonctions, structures de données, programmation orientée objet). Maîtrise de Jupyter Notebooks et de la bibliothèque Python Panda.
Machine Learning	119 heures	Implémentation et apprentissage sur des jeux de données réels pour appréhender les principaux outils de Machine Learning (classification, décomposition en composantes principales, K-Mean, K-Folds, Descentes de gradients, réseaux de neurones).
Outils statistiques pour la donnée	77 heures	Maîtrise des outils mathématiques pour l'analyse de données (statistiques descriptives, tests d'hypothèses, nettoyage de données, régressions).
Bonnes pratiques du développement logiciel en équipe	21 heures	Utilisation des méthodes de gestion de projet Agile et des outils de travail collaboratifs dans une équipe de développement informatique.
Système d'exploitation et bases de données	42 heures	Maîtrise du système d'exploitation open source Linux et construction et manipulation de bases de données.

		Maîtrise des concepts et l'utilisation d'un système de gestion de base de données (SGBD).
<b>Outils de visualisation</b>	14 heures	Cartographier et adapter les outils de visualisation aux jeux de données pour créer une représentation simplifiée.
<b>Compétences douces</b>	7 heures	Maîtrise des outils et techniques de communication Vulgarisation et présentation de résultats scientifiques.

<b>Machine Learning avancé</b>	21 heures	Savoir reconnaître et résoudre des problèmes d'optimisation.
<b>Gestion de projet Data</b>	21 heures	Maîtrise des méthodologies et des outils de gestion de projet appliquées à l'analyse de données (méthodologies agiles, projets cycle en V).
<b>Optimisation</b>	14 heures	Modélisation des modèles d'optimisation par contrainte. Résolution des contraintes Optimisation.
<b>Mathématiques pour l'informatique</b>	21 heures	Outils permettant d'implémenter les articles de la littérature Intégration & dérivation numérique multi-dimensionnelle Calcul matriciel
<b>Visualisation et UI</b>	14 heures	Outils et méthodes de visualisation de grands jeux de données Vulgarisation et représentation simplifiée Création d'un Dashboard interactif de visualisation de données
<b>Advanced computer vision (ACV)</b>	42 heures	Traitement de l'image Deep learning (augmentation de données, GAN) Deep learning vidéo
<b>Apprentissage par renforcement</b>	21 heures	Création d'algorithme de prise de décision sans jeux de données à priori Prise de décision itérative Optimisation sous contrainte
<b>Natural Language Processing</b>	21 heures	Traitement naturel du langage : compréhension du langage humain grâce à l'intelligence artificielle
<b>Ethique de la donnée</b>	17.5 heures	Compréhension des problématiques de données personnelles et services numériques, d'éthique et de gouvernance des données et d'anonymisation des données personnelles.
<b>Open Source</b>	7 heures	Maîtrise des problématiques liées aux aspects communautaire du libre, aux différents types de licences et aux aspects légaux des projets d'analyse de données.
<b>Big Data</b>	35 heures	Application de gestion de projet / Machine Learning & Deep Learning sur un grand jeu de données
<b>Dev Ops</b>	14 heures	Technologies et outils du DevOps : - GIT et GIT flow

		<ul style="list-style-type: none"> <li>- Tests et déploiement</li> <li>- Conteneurisation</li> <li>- Linux avancé</li> <li>- Docker &amp; containers</li> </ul>
<b>Conception Objet</b>	14 heures	Réalisation d'un projet Orienté Objet avancé : Conception UML.
<b>Support Projet</b>	28 heures	Accompagnement sur les problématiques entreprises par un expert en Machine Learning
<b>Préparation Soutenance</b>	7 heures	Préparation des rapports professionnels Préparation soutenances et présentations professionnelles.
<b>Green IT</b>	14 heures	Sensibilisation à l'éco-conception Elaboration d'un guide des bonnes pratiques environnementale Elaboration de la fresque du numérique
<b>Culture mathématiques et informatique</b>	21 heures	Choix des structures de données Fonctionnement des machines Algorithmique avancée Langage C++ Bases de données NoSQL Statistiques avancées Epistémologie Scrapping/ API
<b>Laïcité, citoyenneté et valeurs de la République</b>	3.5 heures	Maîtrise des principes et valeurs de laïcité et des notions liées aux valeurs de la République Française (devise, laïcité, citoyenneté).
<b>Anglais</b>	7 heures	Utilisation de l'anglais technique pour évoluer dans un contexte professionnel anglophone et utiliser la documentation spécifique en anglais.
<b>RGPD</b>	7 heures	Compréhension du RGPD, les différents principes à respecter (minimisation de la collecte, finalité, consentement, conservation limitée des données, sécurité et confidentialité des données), les enjeux du RGPD pour les développeurs web et l'analyse de données.
<b>Cloud</b>	21 heures	Maîtrise des technologies Cloud (Azure, GCP, AWS) pour développer, mettre en place et maintenir les outils et infrastructures des applications Data. Maîtrise des concepts généraux du cloud computing (différents providers, avantages...)

*Ce projet est cofinancé par l'Union Européenne dans le cadre du Fonds Social Européen (FSE)*



Cofinancé par  
l'Union européenne

*En partenariat avec*



**Hewlett Packard  
Enterprise**

