

## Data Science – Les fondamentaux

2 jours (14 heures)

Initiation

Présentiel ou classe à distance

à partir de

**1 570 € HT** par personne (TVA en vigueur) Repas inclus en présentiel

Contenu mis à jour le 25/09/2026. Document téléchargé le 06/10/2026.

### Objectifs

A l'issue de cette formation, vous serez capable de :

- Identifier les enjeux de la Data Science et de l'IA
- Schématiser le cycle d'un projet Data Science
- Appliquer les meilleures pratiques en matière de nettoyage et de préparation de vos données avant l'analyse
- Modéliser un problème de Data Science.

Les compétences visées par cette formation font l'objet d'une évaluation.

### À qui s'adresse cette formation ?

Développeurs de logiciels, programmeurs, Data Analysts et/ou statisticiens.

#### Prérequis

Avoir des connaissances de base en statistiques et de l'expérience avec le langage Python.

### Programme

**1** Pendant votre formation

## Data Science – Les fondamentaux

Jour 1

### Introduction à la Data Science

- Qu'est-ce que la Data Science ?
- Différence entre statistiques, Data Science et IA
- Champs d'application de la Data Science
- Outils et algorithmes pour faire la Data Science
- Les cycles d'un projet Data Science

### L'apprentissage automatique

- L'apprentissage
  - Automatique dans le passé
  - Supervisé
  - Non supervisé
  - Renforcé

### Rappel Python

- Introduction à Python
- Les structures / propriétés / fonctions spéciales
- La programmation orientée objet
- Manipuler les librairies :
  - NumPy
  - SciPy
  - Pandas
  - Matplotlib

### Les algorithmes de l'apprentissage automatique

- Réduction de la dimensionnalité avec l'Analyse en Composantes Principales (ACP)
- Analyse de régression linéaire, multilinéaire, polynomiale, logistique et régularisation
- Classification supervisée
- Classification non supervisée et régression avec arbres de décision
- KNN (K-Nearest Neighbors)
- Le K-Means clustering
- Le perceptron et le support des machines vectorielles (SVM)
- Réseaux de neurones artificiels

### Exemples de travaux pratiques (à titre indicatif)

- Réalisation d'un projet Data Science de bout en bout
  - Collecte des besoins et de l'objectif du projet
  - Exploration des données avec Python

- Visualisation avec Matplotlib
  - Nettoyage des données
  - Imputation des valeurs manquantes (avec moyenne, médiane, most frequent, KNN)
  - Réduction de la dimension avec une ACP
- 
- Entraînement de modèles de Machine Learning selon différents algorithmes et comparaison des performances selon différentes métriques

Jour 2

## Les indispensables de la Data Science

- Cross-validation
- Les métriques d'évaluation
- Overfitting ou surapprentissage
- Biais vs variance
- Etapes de préparation des données (du nettoyage des données à la compréhension des données)
- Feature Engineering
- Data Visualisation
- La malédiction de la dimension
- Bonnes pratiques

*Exemples de travaux pratiques (à titre indicatif)*

- Reprise du projet précédent, ingénierie des attributs, amélioration de la qualité des données, et réapprentissage

## Data Science et le Big Data

- Initiation aux traitements parallèles avec Spark MLlib

# Financement

Plusieurs dispositifs de financement sont accessibles via les OPCO (Opérateurs de Compétences), organismes agréés par le ministère du Travail dont le rôle est d'accompagner, collecter et gérer les contributions des entreprises au titre du financement de la formation professionnelle.

Pour plus d'information, une équipe de gestionnaires SKOLAE Formation spécialisée vous accompagne dans le choix de vos formations et la gestion administrative.

[Nous contacter pour en savoir plus →](#)

## Modalités pédagogiques

Formation délivrée en présentiel ou distanciel\* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode\*\* démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2I, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

\* nous consulter pour la faisabilité en distanciel

\*\* ratio variable selon le cours suivi

### Accessibilité aux personnes en situation de handicap

## Détails de la session

Du 7 décembre 2026 au 8 décembre 2026 (2 jours / 14 heures)

Mulhouse

Du 7 décembre 2026 au 8 décembre 2026 **Mulhouse**

- 7 décembre 2026 ----- 9h00 – 16h30
- 8 décembre 2026 ----- 9h00 – 16h30