

## Formation **Le rôle du Data Scientist**

Familiarisez vous avec le rôle et les responsabilités d'un Data Scientist avec notre formation d'une journée !

Cette formation d'une journée est destinée aux professionnels souhaitant se familiariser avec le rôle et les responsabilités d'un Data Scientist. Les apprenants découvriront les concepts fondamentaux de la science des données, les rôles stratégiques et défis du métier, ainsi que les outils et technologies couramment utilisés par un Data Scientist.

### **Durée**

1 jour

### **Objectifs pédagogiques**

- ◆ Identifier les principales tâches et responsabilités d'un Data Scientist
- ◆ Comprendre les concepts fondamentaux de la science des données (exploration, préparation des données, modèles d'apprentissage automatique)
- ◆ Connaître les outils et technologies couramment utilisés par un Data Scientist

### **Public**

Tous

### **Prérequis**

Des connaissances de base en statistiques et en bases de données sont un plus.

## Programme de formation

Introduction à la formation d'acculturation au rôle de Data Scientist  
Présentation du rôle stratégique du Data Scientist  
Évolution du domaine des sciences de données  
Principaux défis et opportunités actuels  
Différences entre Data Engineer, Data Scientist et Data Analyst  
Vue d'ensemble des compétences requises

Concepts fondamentaux de la Science des Données  
Exploration et visualisation des données  
Techniques de nettoyage et de préparation des données  
Concepts statistiques essentiels  
Introduction aux modèles d'apprentissage automatique : types de modèles, sélection et ajustement, validation croisée et hyperparamétrage

Tâches Typiques du Data Scientist

Collecte et exploration de données venant de différentes sources  
Nettoyage et préparation des données pour garantir leur qualité  
Sélection de modèles adaptés aux problématiques spécifiques  
Évaluation des performances des modèles et optimisation  
Communication des résultats aux non-spécialistes

Panorama des outils et technologies couramment utilisés  
Environnements de Développement Intégrés (IDE) : Jupyter Notebooks, Anaconda  
Langages de Programmation : Python, R, langages spécialisés (SQL, Scala)  
Bibliothèques de Traitement de Données : Pandas, Numpy, Apache Spark  
Bibliothèques d'Apprentissage Automatique : Scikit-Learn, TensorFlow, PyTorch, XGBoost et LightGBM  
Outils de Visualisation : Matplotlib, Seaborn, Power BI, D3.js

## Moyens et méthodes pédagogiques

- ◆ La formation alterne entre présentations des concepts théoriques et mises en application à travers d'ateliers et exercices pratiques (hors formation de type séminaire).
- ◆ Les participants bénéficient des retours d'expérience terrains du formateur ou de la formatrice
- ◆ Un support de cours numérique est fourni aux stagiaires

## Modalités d'évaluation

- ◆ **En amont de la session de formation**, un questionnaire d'auto-positionnement est remis aux participants, afin qu'ils situent leurs connaissances et compétences déjà acquises par rapport au thème de la formation.
- ◆ **En cours de formation**, l'évaluation se fait sous forme d'ateliers, exercices et travaux pratiques de validation, de retour d'observation et/ou de partage d'expérience, en cohérence avec les objectifs pédagogiques visés.
- ◆ **En fin de session**, le formateur évalue les compétences et connaissances acquises par les apprenants grâce à un questionnaire reprenant les mêmes éléments que l'auto-positionnement, permettant ainsi une analyse détaillée de leur progression.