

# DATA SCIENTIST 1/2

## DESCRIPTION

Le rôle d'un **Data Scientist** est de travailler avec la data, afin de lui donner un sens humain. Il s'agit d'une part de séparer le signal du bruit, de « dépolluer » la data. Rôle de préparation, de structuration. La difficulté est de faire en sorte que la data, qui peut provenir de façon massive et hétérogène, soit organisée, et, dans une certaine mesure, homogénéisée.

Le Data Scientist plonge les mains là où la Data se trouve, et la prépare. Il lit la Data, et l'interprète, grâce à des traitements statistiques. Il essaie de lui donner une forme, et de la reconnaître, grâce à des algorithmes de Machine Learning.

Le Data Scientist met au point des programmes au brouillon, des prototypes, qui, lorsqu'ils sont branchés en production par des développeurs, permettent de traiter la Data « at scale », c'est-à-dire à grande échelle, et, lorsque c'est nécessaire, en temps réel - auquel cas la légèreté et la rapidité des algorithmes est aussi un enjeu pour le travail du Data Scientist.

Le Data Scientist est capable non seulement d'analyser les informations, mais aussi de faire preuve de créativité et d'un talent pour générer du profit à partir du Big Data.

## MISSIONS

- Collecter et convertir de larges quantités de données
- Détecter des tendances dans les ensembles de données
- Résoudre les problèmes de l'entreprises grâce aux données
- Communiquer avec les différents responsables de l'entreprise
- Rédiger des rapports actionnables pour la direction

## COMPÉTENCES

- Compétences en informatique, mathématiques et algorithmie
- Expertise en analyse de données
- Maîtrise des langages de programmation Python, R
- Maîtrise des technologies Big Data (Hadoop, MapReduce...)
- Connaissance en Dataviz
- Compréhension Business
- Curiosité naturelle

De solides connaissances en mathématiques, algèbre linéaire, et en statistiques sont nécessaires pour réussir dans ce métier. Une formation approfondie en Machine Learning permettra au Data Scientist de travailler avec des outils modernes, et d'aller plus loin dans l'interprétation de la Data.

Un Data Scientist ayant de l'ambition s'intéressera de près au Deep Learning, et effectuera une veille importante sur ces sujets.

Il peut se spécialiser dans différents domaines : computer vision, reconnaissance d'images, traitement automatique du langage (NLP)...

Bien sûr, un Data Scientist doit avoir une connaissance suffisamment bonne de la programmation (C++, Python, Java, R...) afin de ne pas être freiné dans la construction de ses expériences. Apprendre de nouveaux langages, ou se familiariser avec un nouvel environnement, ne doit pas lui poser de problèmes.

# DATA SCIENTIST 2/2

## FORMATION

**Bac+4-5** en informatique, management, statistiques ou marketing  
Formation en mathématique appliquée, en statistique ou ingénierie informatique  
Formation Big Data et machine learning.  
Expérience professionnelle (Data Analyst)

## DÉBOUCHÉS

Ce poste peut évoluer vers des fonction de management, soit un rôle de lead Data Scientist. Un Data Scientist peut aussi créer des solutions sur mesure pour des clients, et se rapprocher du business, de la vente.

Un Data Scientist peut théoriquement travailler pour n'importe quelle société qui possède des données, et souhaitant les exploiter, pour améliorer sa connaissance client, son marketing, pour optimiser son budget publicité, etc. Les compétences d'un Data Scientist peuvent donc s'appliquer à un vaste champ de domaines professionnels.

## TESTIMONIAUX

5 ans / 3 ans / 3 ans

**Valeur ajoutée de la fonction** : examiner des data multi sources en profondeur au service du business

**Une journée type** : des maths, des statistiques de la modélisation

**Le prochain job** : ne sais pas, j'adore ce que je fais

10 ans / 4 ans / 2 ans

**Valeur ajoutée de la fonction** : faire le bridge entre la data et le développement pour construire des solutions d'analyse poussées; apporter une vision scientifique dans l'analyse des données non standardisée; prédiction et insights actionnables

**Une journée type** : processus d'extraction de contrôle de qualité et d'analyse primaire; définition d'algorithmes d'analyse en fonction des objectifs; définition d'architectures des données et de solutions technologiques; production, tests, automatisation & publication

**Le prochain job** : avoir les mêmes missions mais en tant que manager d'une équipe data science avec des différents profils de compétences