



# Data Analyst

Concepteur développeur en science des données

Titre Professionnel inscrit au Répertoire National des Certifications Professionnelles (RNCP)

BAC +3/4 : Niveau 6

Code(s) NSF :

326 : Informatique, traitement de l'information, réseaux de transmission

114g : Mathématiques de l'informatique, mathématiques financières, statistique de la santé

# DA



## Le métier de Data Analyst

Le Data Analyst a pour mission d'analyser des données brutes collectées en masse (Big Data) sur le web et les réseaux sociaux pour en tirer des conclusions exploitables par les entreprises.

Profil très recherché par les employeurs, l'analyste de données doit ensuite produire des rapports ou reporting, destinés aux différents services de l'entreprise, afin de faciliter la prise de décision des managers.

Le Data Analyst doit savoir analyser et traiter des données grâce à des langages de programmation tels que R, SQL ou HIVE. Il doit également savoir maîtriser des outils comme SAS, Tableau, Microsoft Power BI ou encore Oracle. La maîtrise de ces langages de programmation et de ces outils lui permettront d'effectuer des recherches dans les bases de données brutes. Dans le but de produire du reporting.

Le Data Analyst doit être en mesure d'identifier les problématiques à résoudre grâce aux données recueillies. Il doit être un bon communicant pour pouvoir fournir des informations de façon précise et détaillée aux décideurs.

# Le dispositif de formation

## Public concerné



Toute personne en reconversion professionnelle ou souhaitant monter en compétences. Niveau Bac +2 en informatique (orienté développement web, algorithmie, etc) et une expérience professionnelle en milieu informatique souhaitée.

Toutes nos formations sont accessibles aux personnes en situation de handicap. Pour toute demande d'information, merci de contacter notre référent handicap à [diplome@m2ifformation.fr](mailto:diplome@m2ifformation.fr)

## Métier visés / Passerelles et poursuite d'études



### Métiers accessibles\*

- + Data Quality Manager
- + Cloud Data Engineer
- + Data Manager
- + Chef de Projet CRM
- + Architect Big Data
- + Consultant Big Data



### Passerelles et poursuite d'études possibles\*\* Data Scientist

\* Liste non-exhaustive

\*\* La formation vise l'insertion directe en emploi. Une poursuite de parcours peut néanmoins être envisageable avec des exemples indiqués.

## Prérequis et modalités d'accès



Un bac+2 (niveau 5) dans un domaine scientifique et/ou économique Une pratique (B1-B2) courante de l'anglais Notions Algorithmiques / Python Notions sur les systèmes de programmation informatique.

## Méthodes mobilisées



### Formation en présentiel ou à distance

- + 35 heures/semaine, du lundi au vendredi de 9h00 à 17h00
- + Formation synchrone avec une équipe pédagogique dédiée tout au long du parcours (que cela soit en présentiel en centre ou à distance)
- + Modalités : théorie, pratique, travaux de groupes, individuels, réalisation de projets



### Prérequis techniques fortement conseillés pour suivre cette formation en présentiel à distance

- + Connexion internet « Haut débit », 15 mégabits par seconde minimum
- + Fibre non obligatoire
- + Relier sa box à son ordinateur via un câble réseau
- + Résider en France Métropolitaine
- + Être muni d'un casque audio/micro
- + PC/MAC i7, SSD, 32 Go de RAM
- + Configuration nécessaire pour travailler sur des environnements virtualisés



### Moyens pédagogiques

Un poste informatique/personne en présentiel, les logiciels nécessaires au suivi de la formation, mises en situation, cas pratiques, supports pédagogiques, plateforme multimédia, tutoriels, travaux en sous-groupes, suivi tutoré.

# Le dispositif de formation (suite)

## Modalités d'évaluation

- ☑ Contrôle continu et 2 évaluations en cours de formation (ECF) afin d'évaluer la validation des compétences nécessaires

## Parcours

- 📅 La formation est composée d'une période théorique de 750 heures (en moyenne) et une période pratique de 250 (en moyenne). Le nombre d'heures exact est référencé sur le planning de la session. Compte tenu de l'évolution du référentiel, des compétences métier, des logiciels, les éléments du programme ne sauraient être contractuels.

# Objectifs de formation

Le consultant Data Analyst sera capable de participer à un projet Data de façon agile. Il maîtrisera l'ensemble des étapes de mise en place et d'intégration d'une solution Data (collecte, intégration, transformation, stockage, traitement, analyse et visualisation). Il réalisera dans le cadre de Travaux Pratiques et tout au long du cursus des actions qui permettront de constituer une base de donnée sécurisée et d'appliquer des algorithmes mathématiques de machine. Il apprendra aussi à développer des aptitudes techniques et scientifiques tout au long du cursus.

- + **Bloc n°1** - Construction et alimentation d'une infrastructure de gestion de données
- + **Bloc n°2** - Analyse exploratoire, descriptive et inférentielle de données
- + **Bloc n°3** - Analyse prédictive de données structurées par l'intelligence artificielle
- + **Bloc n°4** - Analyse prédictive de données non-structurées par l'intelligence artificielle
- + **Bloc n°5** - Industrialisation d'un algorithme d'apprentissage automatique et automatisation des processus de décision
- + **Bloc n°6** - Direction de projets de gestion de données



## Les spécificités M2i FORMATION DIPLOMANTE

🔄 Durant l'intégralité du cursus de formation, M2i met à disposition de chaque apprenant :

- + Une licence Microsoft Office 365
- + Un compte sur la plateforme Teams Education
- + Un espace de stockage OneDrive personnel

Cette liste ne saurait être contractuelle. M2i FORMATION DIPLOMANTE, se réserve le droit de faire évoluer son parcours et les technologies enseignées, en fonction des évolutions du référentiel de Jedha.

## Tarif & Financement

€ **Tarif**  
19 € de l'heure.  
**Demandeurs d'emploi ou financement personnel :**  
Tarifs spécifiques à consulter auprès du centre concerné pour un accompagnement personnalisé.

🔄 **Financement**  
Quel que soit votre statut (salarié du secteur privé ou public, demandeur d'emploi...), des [dispositifs de financement](#) vous aident à réaliser votre projet de formation.

Toutes nos formations sont éligibles au CPF.

## Parchemin et livret de certification

- + Un parchemin est attribué au candidat ayant obtenu le titre complet
- + Un livret de certification est remis au candidat en réussite partielle

Ces deux documents sont délivrés par le représentant territorial compétent de Jedha.

Le système de certification du Ministère du Travail du Plein Emploi et de l'Insertion est régi par les textes suivants :

- Code de l'éducation notamment les articles L. 335-5, L 335-6, R 335-7, R . 335-13 et R . 338-1 et suivants ;

- Arrêté du 22 décembre 2015 relatif aux conditions de délivrance du titre professionnel du Ministère du Travail du Plein Emploi et de l'Insertion ;

- Arrêté du 21 juillet 2016 (JO du 28 juillet 2016 modifié par l'arrêté du 15 septembre 2016) portant règlement général des sessions d'examen pour l'obtention du titre professionnel.



## Construction et alimentation d'une infrastructure de gestion de données

- + Concevoir une architecture de données robuste et adaptée en créant des lacs de données (Data Lake en anglais) et des entrepôts de données (Data Warehouse en anglais) afin de répondre aux besoins de stockage, d'utilisation, de sécurité et de protection de l'organisation définie par un cahier des charges.
- + Intégrer la dimension de stockage et de calcul distribuée à l'infrastructure de données via l'utilisation d'outils comme Spark ou AWS Redshift afin de l'adapter à des besoins de gestion de données massives (Big Data en anglais).
- + Collecter des données provenant de différentes sources (Web, Logiciels internes de type Sage/Excel ou externes de type Google Analytics) via des bibliothèques de programmation de type Scrapy ou BeautifulSoup dans le respect des normes de protection des données utilisateurs définies dans le RGPD pour alimenter le Data Lake afin d'affiner le résultat d'analyses futures.
- + Nettoyer et organiser les données dans l'entrepôt de données (Data Warehouse en anglais) en écrivant des processus d'extraction, transformation et chargements (ETL en anglais) afin de rendre ces données disponibles et compréhensibles pour les autres équipes métiers.

## Analyse exploratoire, descriptive et inférentielle de données

- + Traiter des bases de données grâce à des analyses statistiques descriptives et inférentielles via des bibliothèques de programmation comme Numpy ou Pandas, pour les organiser et les nettoyer afin de les normaliser par rapport à la population étudiée.
- + Effectuer des analyses univariées et multivariées sur des bases de données structurées afin de préciser des relations entre plusieurs variables et d'établir des liens statistiques entre elles.
- + Optimiser les analyses statistiques grâce au traitement parallélisé via l'utilisation d'outils comme Spark pour accélérer le temps de calcul d'un ordinateur afin de pouvoir analyser des volumes de données massifs (Big Data)
- + Présenter le résultat d'une analyse statistique de données structurées, massives ou non, grâce à des bibliothèques de programmation comme Plotly ou Matplotlib pour synthétiser ce résultat devant un public profane afin de faciliter la prise de décisions et appuyer leurs déclinaisons opérationnelles.

## Analyse prédictive de données structurées par l'intelligence artificielle

- + Traiter des données structurées en créant un pipeline de traitement grâce à des bibliothèques de programmation comme Scikit-Learn pour encoder, normaliser et découper des données afin de les rendre interprétables par un algorithme d'apprentissage automatique (Machine Learning en anglais).
- + Effectuer des analyses prédictives sur un jeu de données structurées grâce à des algorithmes d'apprentissage automatique supervisés adaptés afin d'automatiser des tâches liées aux résultats des prédictions de ces algorithmes.
- + Élaborer un algorithme d'apprentissage automatique non-supervisé pour segmenter une base de données en différents groupes homogènes ou réduire la dimension de cette dernière afin de pouvoir comprendre des observations de manière granulaire et de permettre leur visualisation.
- + Évaluer la performance prédictive des algorithmes d'apprentissage automatique en déterminant l'influence des différentes variables pour pouvoir l'améliorer afin de démontrer son utilité aux directions métiers, par rapport aux processus déjà établis dans l'organisation.

## Analyse prédictive de données non-structurées par l'intelligence artificielle

- + Traiter des données non-structurées (image, texte, audio) par la création de fonction de traitements via l'utilisation de bibliothèques de programmation comme TensorFlow ou Numpy pour les transformer en matrices afin de les rendre interprétables par un algorithme d'apprentissage automatique profond (Deep learning en anglais).
- + Élaborer des réseaux de neurones adaptés (classiques, convolutifs ou récurrents) en superposant des couches neuronales via des bibliothèques de programmation comme TensorFlow pour analyser des données non-structurées afin de détecter des signaux sur ces dernières.
- + Créer un algorithme robuste et précis en configurant un réseau de neurones pré-entraîné profond afin de répondre à des problématiques de prédiction sur des volumes de données massifs.
- + Créer des données non-structurées en élaborant des réseaux de neurones adverses afin de construire de nouvelles bases d'entraînement pour des applications d'intelligence artificielle.
- + Évaluer la performance d'un algorithme d'apprentissage automatique profond en évaluant des indicateurs sur des données d'entraînement et de validation afin d'industrialiser son utilisation.

## Industrialisation d'un algorithme d'apprentissage automatique et automatisation des processus de décision

- + Standardiser la construction et l'environnement informatique d'un algorithme d'apprentissage automatique grâce des outils de production comme MLflow et Docker afin de faciliter la mise en production de projets d'intelligence artificielle sur tous types de plateformes.
- + Créer une interface de programmation applicative grâce à des outil comme AWS sagemaker afin de donner un accès à échelle aux prédictions des algorithmes d'apprentissage automatique à l'ensemble des équipes métiers concernées.
- + Déployer une application web intégrant des algorithmes de statistiques prédictives (Machine Learning et Deep Learning) grâce à des outils comme Flask, Heroku ou AWS sagemaker pour les rendre utilisables par l'ensemble des équipes métiers afin d'automatiser leurs processus de décision.

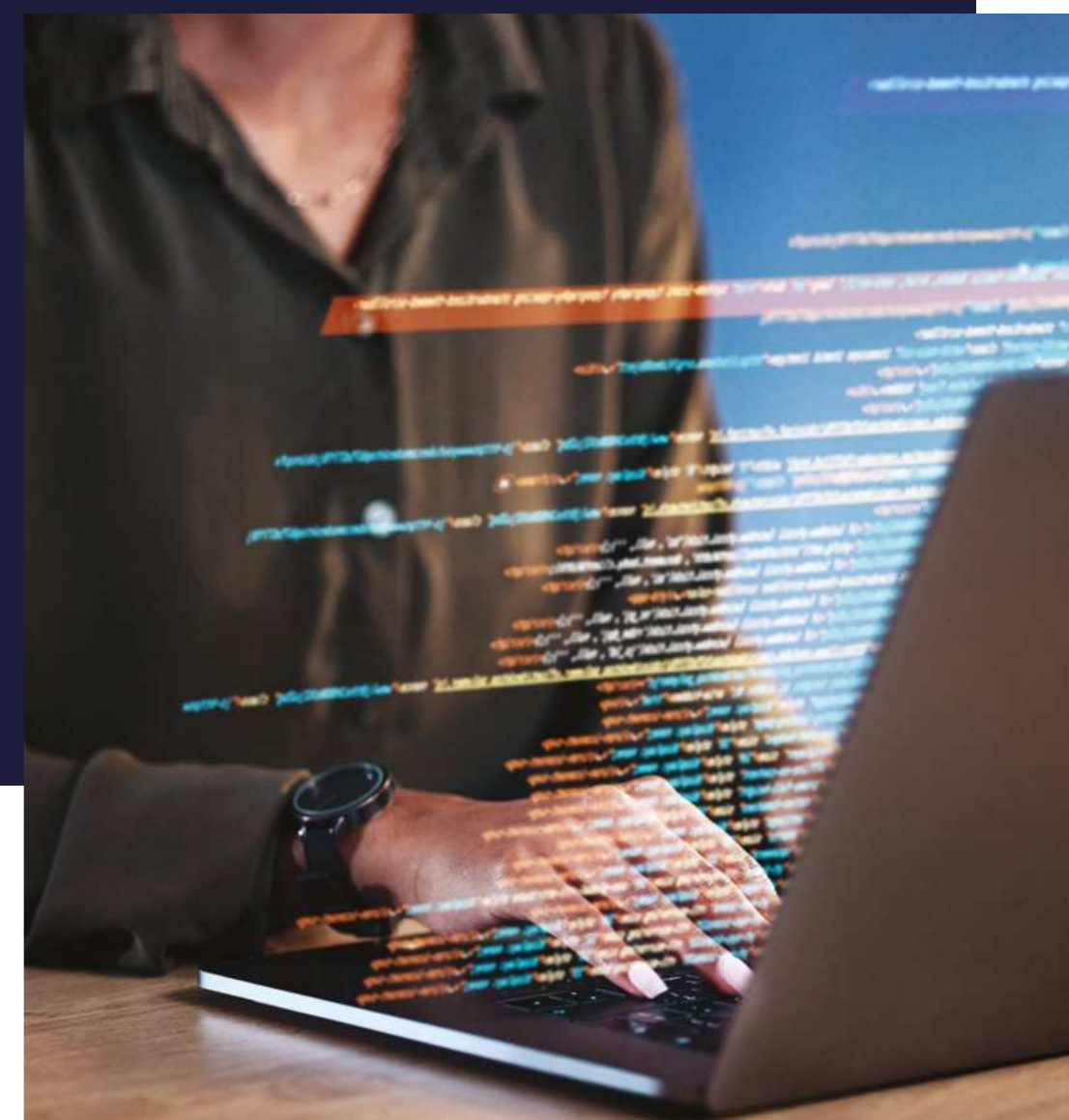
## Direction de projets de gestion de données

- + Traduire les enjeux métiers en problématiques mathématiques/ data grâce à une compréhension des besoins propres à chaque projet data afin de pouvoir répondre aux objectifs de l'organisation.
- + Maîtriser les technologies les plus récentes et adaptées du marché grâce à de la veille technologique et de la pratique constante pour développer une expertise afin d'être à même de proposer aux directions métiers les solutions les plus adaptées actuellement à une problématique et l'amélioration constante des process de gestion de données déjà en place.
- + Définir un cahier des charges, un retroplanning et un budget afin de défendre et détailler aux directions métier un projet data répondant aux besoins de l'organisation.
- + Gérer un projet d'analyse et de gestion de données (analyse statistique descriptive, Machine Learning, Deep Learning, Big Data ou non) grâce à l'élaboration d'indicateurs adaptés et de tableaux de bords, afin de faire le suivi et le bilan de l'action, ainsi que de la déclinaison opérationnelle de ses résultats, le tout dans le respect des normes de protection des données utilisateurs définies dans le RGPD.
- + Transmettre aux directions-métiers le process d'extraction d'informations et d'analyse de données en le vulgarisant afin de soutenir la mise en place d'une stratégie et d'actions futures.
- + Diriger un projet de gestion de données, allant de sa conception à la mise en place de solutions, afin de le mener jusqu'à son terme, d'être la personne clé disposant de toutes les informations sur le projet à tout moment, et d'accompagner d'autres services de l'organisation dans l'ensemble des activités relatives à celui-ci.

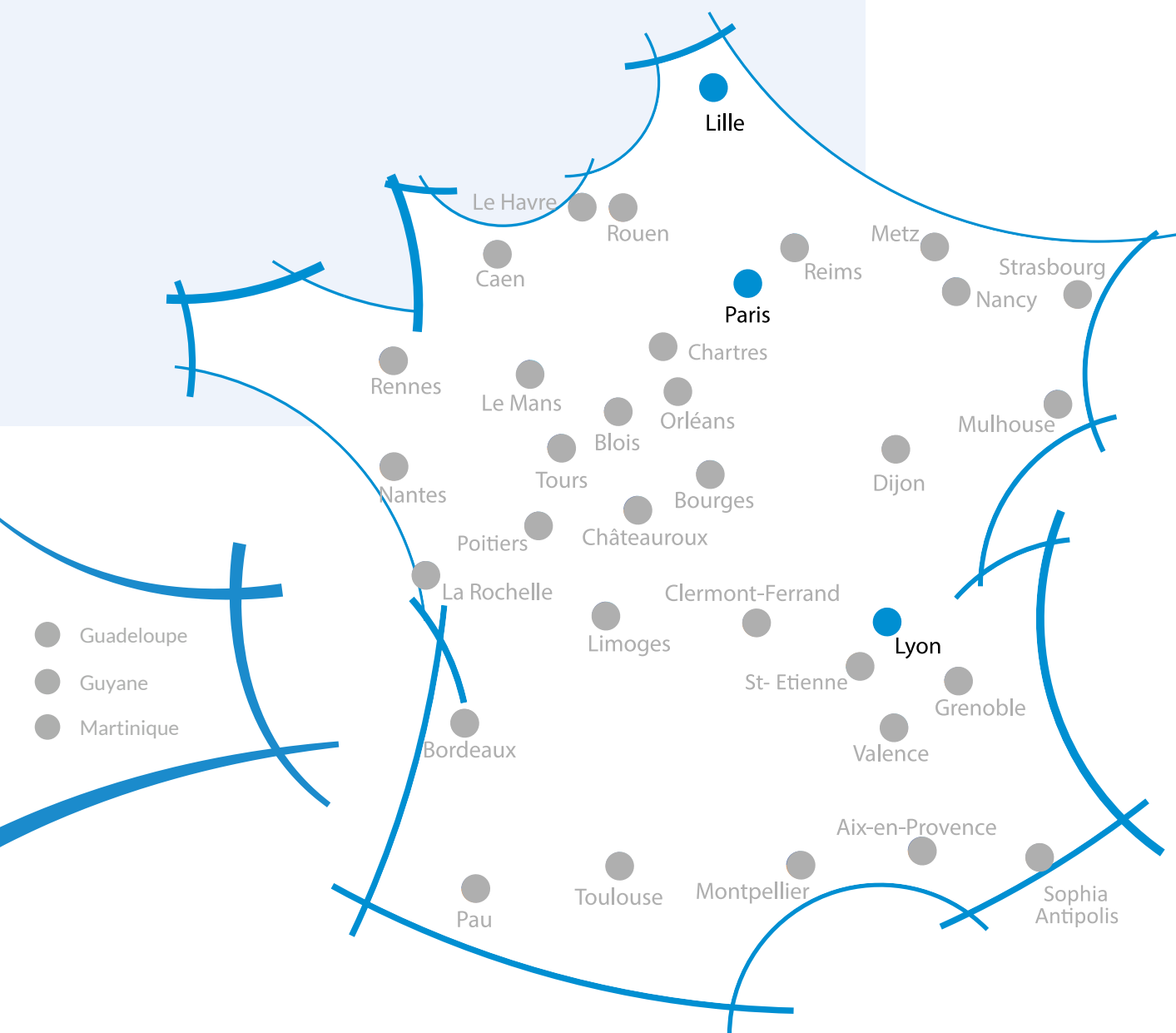
## Modalités d'obtention du titre

Évaluation par le jury de certification sur la base de mises en situation professionnelle :

- + Études de cas réels pour les blocs 1 à 5.
- + Conception et réalisation d'un projet data pour le bloc 6.



# M2i, LEADER DE LA FORMATION IT, DIGITAL ET MANAGEMENT EN FRANCE



Retrouvez tous les détails de notre offre  
sur [diplome.m2iformation.fr](https://diplome.m2iformation.fr)

Tél. Paris : 01 44 53 36 30 - Tél. Lyon : 04 78 02 38 90 - Tél. Lille : 03 20 19 07 19

