



# LE CHIFFRE

## Data Analyst

Alternance • 18 mois • 630 heures

### Les modules de formation

#### 1 – Prérequis :

- Test d'entrée : UML (classes, objets, relations, ...), SQL, HTML, CSS, JS, fonctions mathématiques de base, tableur excel (notion de données sous fichier), notions sur les données...
- Entretien

#### 2 – Fondamentaux (Modélisation, Langages et Systèmes):

- UML
- LANGAGE SQL et Interrogation d'une Bdd
- SGDB du marché
- Notions de statistiques et probabilités
- Python
- Notions de langage R

#### 3 – Fondamentaux de l'analyse de données :

*(Identifier et collecter des données et s'assurer de leur qualité et de leur alignement sur les besoins et usages métiers de l'entreprise)*

- Collecte des données : typologie des sources, les données non structurées, ETL / ELT, typologie 3V des sources, API de réseaux sociaux
- Hadoop : HDFS et écosystème distribué.
- Formes de stockage : bases de données NoSQL, Graphes, Objets,

Document

- Préparation et nettoyage des données. Qualifications
- Analytics
- Apprentissage : Choix d'un algorithme, modèles automatiques, librairies standards
- SPARK : gestion des cluster, ML dans SPARK, variables prédictives, classification et régression
- Traitement en flux Streams, STORM
- Analyser et modéliser des séries temporelles
- Python : Manipulation avec Pandas, Numpy, et Ipython

#### 4 – Analyse et Machin Learning :

*(Maîtriser le Mining et le Machine Learning pour explorer de très importants volumes de données et construire des modèles répondant aux problèmes des entreprises impossibles de résoudre par les méthodes classiques)*

- Apprentissage supervisé vs Apprentissage non supervisé et méta-apprentissage
- Ingénierie de la décision : chaîne de Markove discrète, jeu d'entraînement et jeu de test
- Echantillonnages
- Régression logistique binaire, regroupements, arbres aléatoires, réseaux de neurones
- Text Mining : analyse statistique de corpus
- Big Data : gestion de gros volumes de données

#### 5 – Mise en œuvre de l'analyse de données :

*(Savoir mettre en œuvre une solution complète en environnement Hadoop pour l'analyse des données)*

- Collecter les données : sources, outils de collectes et de traitement, données semi-structurées et non-structurées
- Stocker les données : les formes, les technologies, écosystème Hadoop, infrastructure matérielle et logicielle,
- Analyser les données : Requêter, analyse de données non structurées, analyse statistique et prédictive, tendances, programmes d'automatisation des analyses, Machine Learning, Deep Learning
- Transformer les données en décisions : besoins utilisateurs, traduction en requêtes, vérification de la qualité des données,

#### 6 – Data visualisation pour la restitution de données :

*(Concevoir des modèles de documents adaptés aux besoins métiers de l'entreprise)*

- Graphique, Tableau Software, formalisation
- Python Matplotlib
- Rapports statistiques



@lechiffrelille



@EcoleLeChiffre



Le Chiffre



## Le métier

Le Data Analyst – Développeur big data prend en charge la conception et le développement d'applications informatiques.

Le volume des données et leurs variétés étant en perpétuelle évolution dans les entreprises, il conçoit, modélise et actualise la structure des bases de données selon les orientations et besoins de l'entreprise.

Il deviendra donc un collaborateur incontournable de par sa capacité à fournir aux décideurs à tout moment les éléments déterminants nécessaires aux prises de décisions stratégiques pour faire évoluer l'entreprise.

Le Data Analyst doit donc avoir une compréhension fine des activités de son entreprise, et être un excellent communicant. Doté de solides compétences en statistiques, à l'aise avec la visualisation de données, le Data Analyst doit également maîtriser les outils de Business Intelligence et le requêtage de bases de données.

## Le titre professionnel

*Eligible au CQP2 "base de données"*

*du Titre Professionnel du Ministre du Travail de niveau 6*