

# DATA SCIENCE – LES FONDAMENTAUX

Durée : 2 jours (14 heures)

## CONNAISSANCES PREALABLES

---

- Maîtriser l'utilisation courante d'un ordinateur
- Connaissances générales en statistiques descriptives recommandées
- Connaissance de base des données et d'Excel appréciée
- Aucune compétence en programmation avancée n'est requise.

## PROFIL DES STAGIAIRES

---

- Chefs de projet Data et BI
- Data Analysts
- Consultants décisionnels
- Responsables métiers souhaitant exploiter la donnée
- Managers impliqués dans des projets Data ou IA
- Développeurs et ingénieurs souhaitant découvrir la Data Science
- Toute personne désirant comprendre les principes, méthodes et usages de la Data Science.

## OBJECTIFS

---

À l'issue de cette formation, les participants seront capables de :

- Comprendre les concepts fondamentaux de la Data Science
- Identifier les étapes d'un projet Data Science
- Préparer et explorer des données
- Comprendre les principes du Machine Learning
- Différencier les principaux algorithmes supervisés et non supervisés
- Évaluer les performances d'un modèle prédictif
- Identifier les cas d'usage pertinents en entreprise
- Comprendre les limites, risques et enjeux éthiques de la Data Science.

## CERTIFICATION PREPAREE

---

Aucune

## METHODES PEDAGOGIQUES

---

- Apports théoriques illustrés par de nombreux exemples concrets
- Études de cas issues de projets réels
- Démonstrations d'outils Data Science
- Travaux pratiques guidés
- Exercices d'interprétation de résultats
- Échanges d'expériences et retours terrain.

## FORMATEUR

---

Consultant Data Scientist expérimenté, spécialiste de l'analyse de données, du Machine Learning et de l'accompagnement des projets Data et IA en entreprise..

## METHODE D'EVALUATION DES ACQUIS

---

- Quiz de validation des connaissances
- Exercices pratiques individuels et collectifs
- Études de cas
- Évaluation continue durant les ateliers
- Validation des acquis en fin de formation.

## CONTENU DU COURS

---

### Module 1 – Introduction à la Data Science (2h)

#### Objectifs

- Comprendre ce qu'est la Data Science et son positionnement dans l'écosystème Data
- Identifier les différents métiers de la Data
- Découvrir les principaux cas d'usage en entreprise

#### Contenu

- Définition de la Data Science
- Différences entre Data Science, Data Analytics, Business Intelligence et Intelligence Artificielle
- Les métiers de la Data :
  - Data Analyst
  - Data Scientist
  - Data Engineer
  - Machine Learning Engineer
- Les composantes d'un projet Data
- Panorama des usages :
  - Préviation
  - Segmentation client
  - Maintenance prédictive
  - Détection de fraude
  - Recommandation

#### Mise en pratique

- Tour de table sur les usages de la donnée dans les organisations des participants
- Brainstorming collectif : identifier des cas d'usage Data Science applicables à son activité
- Analyse commentée de plusieurs projets Data réels

### Module 2 – Le cycle de vie d'un projet Data Science (1h30)

#### Objectifs

- Comprendre les différentes phases d'un projet Data Science
- Identifier les facteurs de succès et les points de vigilance

#### Contenu

- Les étapes d'un projet Data Science :
  - Compréhension du besoin métier
  - Collecte des données

- Préparation des données
- Modélisation
- Évaluation
- Déploiement
- Industrialisation
- Présentation de la méthodologie CRISP-DM
- Organisation des équipes Data
- Gouvernance et qualité des données

#### **Mise en pratique**

- Étude de cas : analyse d'un besoin métier et définition des étapes du projet
- Atelier collaboratif de construction d'un cycle projet Data

### **Module 3 – Comprendre, préparer et explorer les données (3h30)**

#### **Objectifs**

- Identifier les différents types de données
- Comprendre les enjeux de qualité des données
- Réaliser une première analyse exploratoire

#### **Contenu**

- Les différents types de données :
  - Structurées
  - Semi-structurées
  - Non structurées
- Les sources de données internes et externes
- Les problématiques de qualité des données :
  - Données manquantes
  - Doublons
  - Valeurs aberrantes
  - Incohérences
- Préparation et transformation des données
- Introduction à l'analyse exploratoire (EDA)
- Statistiques descriptives essentielles
- Identification des tendances et corrélations

#### **Mise en pratique**

- Étude d'un jeu de données métier
- Détection des anomalies et problèmes de qualité
- Réalisation d'une analyse exploratoire guidée
- Interprétation des résultats obtenus

### **Module 4 – Introduction au Machine Learning (2h)**

#### **Objectifs**

- Comprendre les principes fondamentaux du Machine Learning
- Différencier les principales familles d'algorithmes

#### **Contenu**

- Intelligence Artificielle, Machine Learning et Deep Learning
- Fonctionnement général d'un modèle prédictif
- Notion d'apprentissage à partir des données
- Jeux d'entraînement, de validation et de test
- Présentation des grandes familles :
  - Apprentissage supervisé
  - Apprentissage non supervisé

- Apprentissage par renforcement
- Cas d'utilisation associés

#### **Mise en pratique**

- Quiz interactif sur les concepts fondamentaux
- Identification du type d'apprentissage adapté à différents cas métiers
- Analyse commentée de modèles existants

### **Module 5 – Les principaux algorithmes de Machine Learning (2h)**

#### **Objectifs**

- Comprendre le fonctionnement général des algorithmes les plus utilisés
- Identifier leurs domaines d'application

#### **Contenu**

##### **Algorithmes supervisés**

- Régression linéaire
- Régression logistique
- Arbres de décision
- Random Forest
- Gradient Boosting

##### **Algorithmes non supervisés**

- K-Means
- Clustering hiérarchique
- Détection d'anomalies
- Analyse en composantes principales (ACP)

#### **Choisir un algorithme selon son besoin**

- Prédiction
- Classification
- Segmentation
- Détection d'événements inhabituels

#### **Mise en pratique**

- Étude comparative de plusieurs algorithmes sur un même cas métier
- Atelier de choix d'algorithmes selon différents scénarios d'entreprise

### **Module 6 – Évaluer la performance d'un modèle (1h30)**

#### **Objectifs**

- Comprendre comment mesurer la qualité d'un modèle
- Identifier les risques liés à une mauvaise évaluation

#### **Contenu**

- Surapprentissage (Overfitting)
- Sous-apprentissage (Underfitting)
- Validation croisée
- Matrice de confusion
- Mesures de performance :
  - Accuracy
  - Precision
  - Recall
  - F1-Score
- Introduction aux courbes ROC et AUC

#### **Mise en pratique**

- Analyse d'indicateurs de performance réels
- Interprétation de résultats de modèles

- Atelier de diagnostic de modèles prédictifs

### **Module 7 – Valoriser les résultats et communiquer efficacement (1h)**

#### **Objectifs**

- Savoir restituer les résultats d'une analyse Data
- Faciliter la prise de décision grâce à la donnée

#### **Contenu**

- Principes de Data Visualisation
- Les bonnes pratiques du Data Storytelling
- Construction d'indicateurs pertinents
- Présentation des résultats aux décideurs
- Erreurs fréquentes de communication

#### **Mise en pratique**

- Analyse critique de tableaux de bord
- Atelier de restitution des résultats d'une étude Data

### **Module 8 – Éthique, IA responsable et étude de cas de synthèse (0h30)**

#### **Objectifs**

- Comprendre les enjeux éthiques liés à la Data Science
- Identifier les bonnes pratiques d'une démarche responsable

#### **Contenu**

- Les biais algorithmiques
- Qualité et représentativité des données
- Transparence et explicabilité des modèles
- Protection des données personnelles
- Introduction aux principes de l'IA responsable

#### **Mise en pratique**

- Étude de cas de synthèse reprenant l'ensemble des concepts abordés
- Réflexion collective sur les risques et impacts d'un projet Data Science
- Débriefing et échanges de bonnes pratiques

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:01.71.19.70.30) ou par mail à <mailto:referent.handicap@edugroupe.com> pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.