

# Catalogue des formations AIC Data Services

## Le parcours formation « Maitrise des données »

**Version 2023**

## L'offre

AIC Data Services vous propose une offre de formation centrée autour des métiers de gestion des données, du décisionnel et de la Data Science.

Nous proposons des séminaires sur des sujets de notre expérience de consultant. En complément, nous effectuons une veille permanente sur l'organisation de l'approche des données dans les entreprises, les grandes et les PME. L'attente vis-à-vis des données est en constante évolution et les métiers changent en parallèle. Les bonnes pratiques de Data Management font émerger des activités nouvelles comme l'intendance des données (le **Data Steward**) et des responsabilités nouvelles comme le **CDO, Chief Data Officer** et le **DPO, Data Protection Officer**.

Les nouvelles réglementations forcent la mise en place de ces nouvelles responsabilités et positions, toutefois, celles-ci peuvent et doivent être transformées en gains de performance et en avantage compétitif pour l'entreprise. C'est dans cet esprit que nous avons conçu ces formations.

Cette offre s'appuie sur près de 8 ans d'expérience de séminaires en partenariat avec Orsys. Ce sont des centaines de séminaires qui ont déjà été effectués. Les retours des participants pour nos séminaires mesurés par les organismes de formation sont excellents.

## Les formateurs

Les séminaires AIC Data Services sont animés par des formateurs senior ayant un parcours professionnel dans lequel ils ont mis en œuvre et encadré les sujets couverts.

Les sujets sont abordés sous l'angle théorique et de terrain. Les participants acquièrent un savoir-faire pratique qui peut être mis en application directement.

## L'organisation

Les séminaires se déroulent en général sur 2 ou 3 jours en présentiel ou en classe virtuelle. Il est recommandé de limiter l'effectif à 10 ou 12 personnes maximum pour que les exercices pratiques soient suivis et efficaces.

Nous sommes en mesure d'organiser des séminaires dans vos locaux sur devis. N'hésitez pas à nous contacter.

## Choisir votre parcours Data Management

### Data Management

#### La gouvernance et la gestion des données

- **Comprendre les métiers et maîtriser les activités de Data Management**
  - **DMG** : Stratégie et mise en œuvre la Gouvernance des données (2 j)
  - **DMQ** : Démarche qualité des données (2 j)
  - **DMM** : Le projet Master Data Management (2j)

Ces trois modules permettent d'avoir une vue d'ensemble sur le périmètre complet de la gouvernance des données. En s'appuyant principalement sur la méthodologie de DAMA, le module SGOD permet d'acquérir les bonnes pratiques pour mettre en place une gouvernance des données dans des contextes de grande entreprise ou de PME.

### Décisionnel

#### Le décisionnel

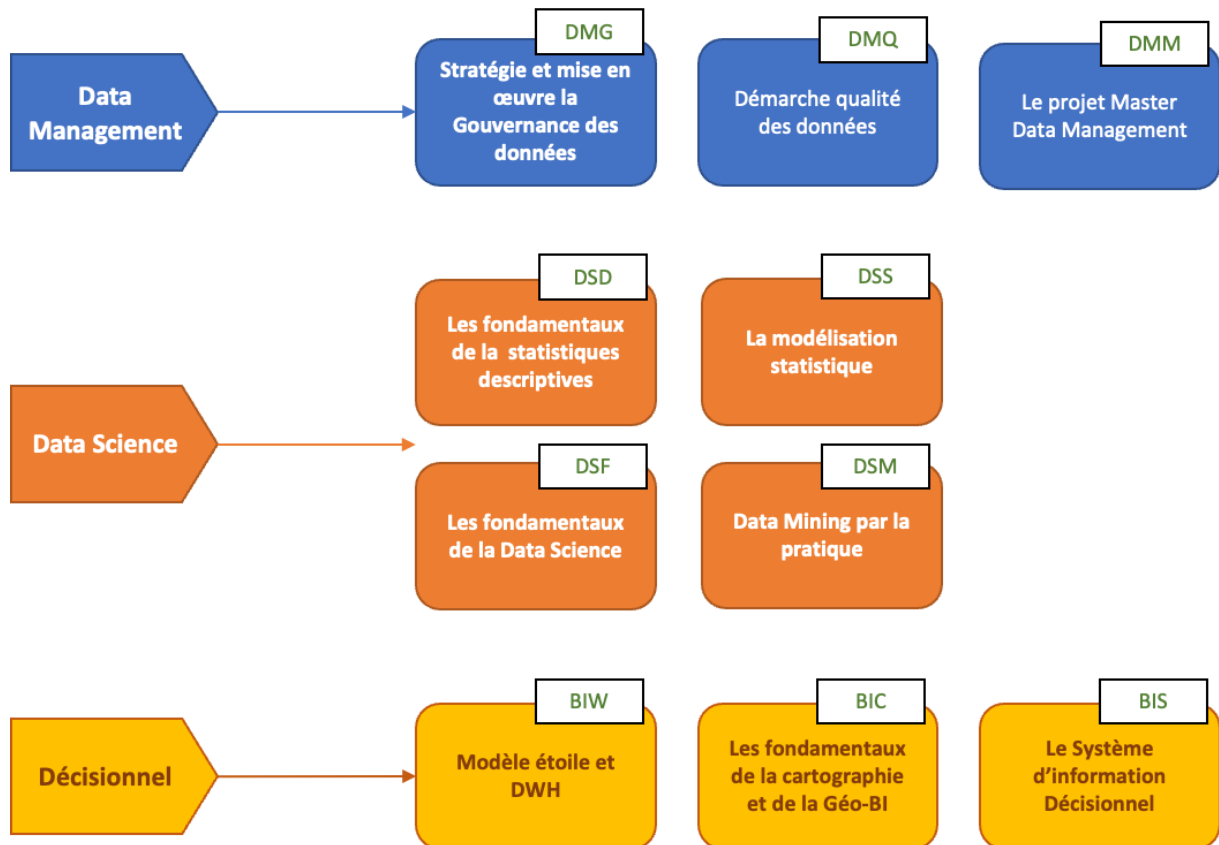
- **Concevoir le cœur du système d'analyse des données pour les statistiques**
  - **BIW** : Modélisation étoile et conception d'un entrepôt des données (3 j)
  - **BIC** : Les fondamentaux de la cartographie et la GeoBI (2 j)
  - **BIS** : Le système d'information décisionnel (3 j)

### Data Science

#### La Data Science

- **Parcours complet de formation aux statistiques avancées depuis les fondamentaux**
  - **DSF** : Les fondamentaux de la Data Science (2 j)
  - **DSD** : Les fondamentaux de la statistiques descriptive (2 j)
  - **DSS** : La modélisation statistique (2 j)
  - **DSM** : Le Data Mining par la pratique (3 j)

## Choisir votre parcours Data Management



**Data Management**  
**DMG****2 Jours**

## Stratégie et mise en œuvre de la Gouvernance des données

### Objectifs pédagogiques

La gouvernance des données est une démarche et une organisation transverse à l'entreprise ayant pour objectif mise en place de bonnes pratiques en matière de bonne gestion des données. Disposer d'un ensemble de données cohérent et fiable au niveau le plus transverse permet de servir et superviser l'ensemble des processus métier. Ce séminaire a pour objectif de vous aider à positionner la démarche sur les pôles les plus critiques et apporteurs de valeur. Il s'agit également de couvrir les aspects réglementaires et les risques sur les données.

La démarche gouvernance couvre l'organisation, les acteurs clés, la présentation des méthodologies recommandées, en particulier la méthodologie de DAMA. Le reporting et les plans d'actions pour la gouvernance des données sont présentés sous forme d'exercices afin de disposer des outils de la création de valeur autour des données.

### Programme

#### L'enjeu données et système d'information

- L'enjeu des données dans l'entreprise
- Les types de données
- Valeur métier et risques attachés aux données. (Exercices sur cas métier)
- Les enjeux réglementaires
- Les enjeux technologiques : Opérationnels, IoT, Décisionnel, Data Lake, Big Data

#### Acteurs et méthodologie pour la maîtrise des données

- La démarche Gouvernance et approche méthodologique
- Les acteurs et les métiers liés au Data Management
- La qualité des données comme outil de mesure et de communication
- Les données de référence transverses et entités transverses (MDM)
- L'acculturation orientation données

#### *Étude pratique : identification des besoins de gouvernance des données*

#### Déployer la gouvernance des données

- Choisir la bonne approche pour la Gouvernance selon l'entreprise
- Le métier de Chief Data Officer, garant de la stratégie orienté données
- Le rôle clé de l'intendance des données
- Élaboration des principes de Data Management et grille de maturité
- Suivi d'un agenda de gouvernance – Piloter et Mesurer les gains

#### *Cas pratique de mise en place d'une grille de maturité et agenda de gouvernance*

#### Transformer les données en stratégie de la connaissance

- Le capital informationnel
- Les tendances technologiques
- Développer la maîtrise des données
- Les opportunités de la BI et Data Science

### Synthèse

- Organisation et feuille de route
- Les outils du marché

### Participants

Ce séminaire s'adresse au *Chief Data Officer*, au *Directeur des Systèmes d'Information*, au *chef de projet* ainsi qu'aux *Maîtrises d'ouvrages* impliquées dans un projet mise en place d'une gouvernance

Data Management  
**DMM**

**2 Jours**

## Le projet Master Data Management

### Objectifs pédagogiques

Le projet de refonte des référentiels transverses de l'entreprise, encore appelé Master Data Management est une application opérationnelle de la gouvernance des données. Le projet doit être porté par une organisation et suivre des lignes guides présentées dans ce séminaire de façon très pratique. La définition du cadre référentiel est détaillée à partir de cas concrets. Les priorités et le phasage est abordé. Les aspects techniques sont évoqués car ils peuvent déterminer la faisabilité de l'approche attendue.

### Programme

#### Introduction

- Rôle stratégique des données pour l'entreprise.
- Différence entre données et information.
- Principe d'organisation des données de référence
- Enjeux d'organisation et avantage compétitif

#### Mise en place de la méthode projet

- MDM et gouvernance : cadre méthodologique basé sur DAMA
- Créer la valeur métier avec les données transverses
- Mesurer la maturité de gestion des données
- Identifier les acteurs et l'organisation

#### ***Atelier de mise en place d'une mesure de maturité***

#### Concevoir les données Master

- Identifier les sources de référence et les acteurs
- Construire des modèles de référentiels transverses à partir de cas pratiques

#### ***Atelier de conception des données partagées, réconciliation des points de vue métier***

#### Les modèles d'architecture

- Principes d'architecture MDM
- Les modèles d'architecture
- Les solutions éditeurs

#### ***Étude comparative de solutions techniques, avantages et inconvénients***

#### Maintenir les référentiels dans le temps

- Les outils disponibles
- Organiser le cycle de vie des données
- Les glossaires et dictionnaires de données

#### Synthèse de la démarche et tendances

### Participants

Ce séminaire s'adresse aux intendants ou référents données, aux chefs de projet Data ainsi qu'aux MOA impliquées dans une dans un projet MDM ou la gouvernance des données

## Démarche qualité des données

### Objectifs pédagogiques

La mesure de la qualité est une problématique complexe car de nombreuses données concourent à produire des informations métier. Le séminaire Démarche Qualité des Données présente comment faire le lien entre les informations métier critiques et les données impliquées. Il couvre la démarche méthodologique de mise en place des indicateurs, l'organisation. Il détail également les méthodes statistiques à utiliser pour superviser de façon efficace la qualité. Le séminaire présente des cas d'usage et les solutions qui peuvent être utilisées pour l'acquisition des données, le calcul des indicateurs et les exemples de restitution sous forme de tableaux de bord et de rapports.

### Programme

#### Introduction

- Rôle stratégique des données pour l'entreprise.
- Différence entre données et information.
- Les différentes sources de données de l'entreprise.
- Les différentes formes d'exploitation de données.
- Architectures : relationnelle, NoSQL ou Big Data.

#### Définition de la qualité et cadre méthodologique

- Définition et mesure de la qualité des données.
- Les méthodologies de la qualité des données.
- Comparaison entre TDQM/DWQ/AIMQ/ORME Data Quality.
- Les principes d'évaluation de la qualité des données des entreprises.
- Synthèse sur les méthodologies qualité.
- Dimensions : définitions et mesures.

***Simulation d'un atelier d'identification des valeurs métiers et des dimensions qualités relatives. Construction des indicateurs qualité***

#### Démarche gestion de la qualité des données

- La place de la qualité dans la démarche de gouvernance.
- Les acteurs et l'organisation. La méthodologie DAMA.
- Mise en œuvre de la démarche Projet.
- Le coût de la non-qualité.
- Périmètre de la démarche qualité.
- Niveau d'approche et de granularité.

***Simulation d'un atelier de mise en place de l'organisation et de la démarche***

#### Contrôle qualité et bonnes pratiques

- Audit de la qualité.
- Identification des données de faible qualité.
- Collecte et stockage des mesures de la qualité.
- Une approche centralisée de la qualité des données.
- Types de contrôles et outils statistiques.



- Exploitation et évaluation des mesures de la qualité.

***Exercices sur la mesure de la qualité. Formalisation des indicateurs.***

### **Suivi opérationnel de la qualité des données**

- Tableaux de bord qualité.
- Cycle d'amélioration continue.
- Quel rôle pour la gouvernance ?

### **Hausser le niveau de la qualité**

- Intervenir en amont de la chaîne.
- Identification des cas atypiques.
- Gestion ciblée des cas de données de faible qualité.
- La réconciliation des données.

### **Synthèse**

## **Participants**

Ce séminaire s'adresse aux intendants ou référents données, aux chefs de projet Data ainsi qu'aux MOA impliquées dans une démarche qualité, dans un projet MDM ou la gouvernance des données

Aucune connaissance technique n'est requise pour ce séminaire, mais ce séminaire s'adresse plutôt à des personnes déjà impliquées dans ces problématiques.

**Décisionnel  
BIW****3 Jours**

## La conception du Data Warehouse

### Objectifs pédagogiques

Concevoir un entrepôt de données dans les années 2020 s'appuie toujours sur des principes fondateurs des années 90, mais les volumétries, les performances, les outils et surtout les attentes ont beaucoup évolués. Ce séminaire vise à former aux fondamentaux du schéma en étoile, des dimensions à évolution, mais également de prendre en compte les données du Big Data et lacs de données. Les datamart et la « self BI » sont abordés avec les problématiques de sécurité qu'ils engendrent. De nombreux exercices pratiques amènent les participants à trouver les solutions à partir d'exemples commentés.

### Programme

#### Le Data Warehouse, finalité et principes

- Les enjeux stratégiques d'un SI décisionnel.
- Les raisons techniques et culturelles qui conduisent au Data Warehouse.
- Définition du Data Warehouse selon Bill Inmon.
- Les solutions apportées par l'architecture technique et fonctionnelle du Data Warehouse.
- Caractéristiques des données du SI décisionnel.
- L'Infocentre et le SI décisionnel.
- Présentation des différentes approches Data Warehouse et Infocentre, leurs avantages et inconvénients.

#### L'architecture d'un SI décisionnel d'entreprise

- Les différentes couches de l'organisation d'un Data Warehouse.
- La collecte et l'intégration des données.
- L'Operational Data Store et la Data Staging area.
- La couche de présentation, le portail décisionnel.
- Les moteurs d'analyse OLAP : (MOLAP) et/ou relationnel OLAP (ROLAP).
- Analyse des données : méthodes descriptives et prédictives
- Croissance du volume et de la nature des données, les enjeux du Big Data.
- La documentation des données du DW : notions de référentiel de données.
- Comment le DW fiabilise la gestion du référentiel de données (MDM)
- La gestion des flux : la capture des données sources, les règles de transformation.

#### *Exercice sur un projet d'analyse décisionnelle divers.*

#### Les principes de la modélisation en étoile

- Les modèles relationnels opérationnels et dénormalisés.
- Les modèles hybrides.
- Les modèles génériques.
- Comprendre le modèle en étoile, sa finalité.
- Comprendre les notions de faits et d'analyse. Les hiérarchies d'axe d'analyse.
- Le modèle en flocon.
- La problématique des dimensions à évolution.
- Gestion des agrégats et de la stabilité du périmètre fonctionnel.

- Quelle approche favorise les informations de détail ou les agrégats ? Bonnes pratiques, questions à poser au métier.

***Exercice de construction et d'enrichissement d'un modèle de données en étoile, suivant plusieurs études de cas données.***

### **La démarche de construction d'un Data Warehouse**

- Identifier le périmètre fonctionnel candidat. Déterminer l'objectif et les événements de gestion à suivre.
- Estimer la volumétrie du périmètre.
- Analyse fonctionnelle, recueil des besoins utilisateurs.
- Conception de l'architecture technique détaillée.
- Etablir une démarche générique de mise en œuvre.
- Les apports d'une démarche itérative, le contenu d'une itération.
- Première itération ou projet pilote, bien le choisir. Rôle du sponsor, de la MOA, de la MOE, impact sur l'organisation.
- L'administration et le suivi de la solution opérationnelle.

***Mise en pratique d'une démarche de conception fonctionnelle du Data Warehouse.***

### **Organisation du projet, les acteurs et les livrables**

- Le pilotage du projet
- Le rôle de l'équipe fonctionnelle, du groupe de projet utilisateurs : valider le design de l'environnement utilisateur.
- Le transfert de compétences vers les utilisateurs finaux par l'équipe fonctionnelle : formation et documentation.
- L'équipe technique, les architectes.
- Les principaux livrables d'un projet décisionnel.

***Présentation des livrables et de leur responsable suivant les étapes de la démarche.***

### **Les outils dans le domaine du décisionnel**

- Les dernières évolutions techniques des SGBDR dans le domaine du décisionnel.
- Panorama et typologie des solutions BI du marché. Les offres en mode SaaS.
- Les solutions de reporting : SSRS, IBM Cognos, SAS, BusinessObjects... La mise en œuvre des outils de requête.
- Les outils d'analyse OLAP côté serveur et côté client : utilisation, évolutivité, approche DataMart, temps de réponse.
- Les solutions d'analyse data mining : SAS Enterprise Miner, IBM, OBI data mining. Exigences et points forts.
- Les solutions ETL : IBM, Informatica, Oracle, SAP, Talend...
- Les outils de modélisation relationnelle : possibilités et limites.

***Présentation des possibilités de divers outils BI.***

### **Synthèse**

#### **Participants**

Ce séminaire s'adresse aux chefs de projet, analystes décisionnel ou développeurs sur projets décisionnels souhaitant acquérir les méthodes de conception d'un entrepôt de données.

Une connaissance des principes décisionnels est recommandée

Décisionnel  
BIC

3 Jours

## Les fondamentaux de la cartographie pour la GéoBI

### Objectifs pédagogiques

Comprendre et maîtriser les principes et modes de représentation cartographique, et les appliquer à l'analyse de données

Apprendre à installer un serveur cartographique et apprendre le développement à partir des bibliothèques OpenLayers et Leaflet

### Programme

#### Premier jour - Concepts de base - Fonctions des SIG - Contexte :

- Introduction aux SIG : Comprendre ce qu'est un SIG et comment s'articulent ses différentes composantes (logiciel, données, ressources matérielles, humaines...).
- Logiciels SIG : Découvrir l'offre actuelle en logiciels et comprendre leurs différentes fonctionnalités.
- Contexte institutionnel des SIG : Identifier les grands acteurs publics et privés du domaine. Comprendre les démarches d'accès aux données et de diffusion.
- Les fondamentaux de coordonnées et de projection (Lambert, WGS84...)

#### *Exercices de manipulation*

#### Deuxième jour - Représentation de l'information - Données :

- Modes de représentation : Comprendre comment les données sont structurées dans un SIG, ce qu'apporte la dimension géographique par rapport aux données alphanumériques classiques. Connaître les modes d'acquisition spécifiques aux données spatiales.
- Bases de données géographiques disponibles : Analyser le contenu des bases de données les plus courantes, afin d'identifier leur niveau d'utilisation et leur complémentarité.
- Les usages des données géographiques : Manipuler des échantillons de bases de données et construire une grille d'analyse des contenus et des usages.
- Bases de la cartographie : Connaître les bases de la sémiologie cartographique.
- Exploitation des données cartographiques pour produire des cartes
- Créer des cartes statistiques, des cartes thématiques et un atlas en appliquant les règles de sémiologie cartographique.

#### *Exercices de manipulation de données et développement de cartes*

#### Troisième jour - Cartographie et traitement des données :

- Traitement des données et requêtes : Effectuer des calculs sur les données : calculs de distance, de surface, d'indicateurs. Sélectionner des objets géographiques à partir d'un ou plusieurs critères.
- Analyse spatiale : Savoir ce que sont les outils de géotraitement, les chaînes de géotraitement.
- SGBD spatiaux : Appréhender les différentes méthodes de stockage des données géographiques et les SGBD (Systèmes de Gestion de Bases de Données) spatiaux.
- Serveurs géographiques : Savoir créer des services web cartographiques sous GeoServer.

- Développement d'applications pour clients légers : Découvrir les clients légers (web) permettant de construire des applications cartographiques en ligne (OpenLayers, LeafLet).
- Développement d'applications pour clients lourds : Connaître les environnements de développement des principaux SIG du marché (QGIS, ArcGis, MapInfo, GeoMedia)

### **Participants**

Ce séminaire s'adresse aux consultants, aux développeurs et aux Data Scientists souhaitant acquérir les bases de la représentation cartographique des données.

Aucune connaissance technique avancée n'est requise pour ce séminaire. Il est recommandé d'avoir quelques bases de programmation

Data Science  
DSD

2 Jours

## Les fondamentaux de la statistique descriptive

### Objectifs pédagogiques

Ce séminaire est formation et/ou une remise à niveau sur les techniques de la statistique afin d'aborder les séminaires avancés du Data Mining., de l'intelligence artificielle et du prédictif

Les notions sont présentées de façon pratiques, toutes les notions sont illustrées par des exemples et complétées par des exercices.

### Programme

#### La formalisation mathématique pour les statistiques

- Indexation de 1 à n. La valeur absolue.
- Le Symbole Sigma pour l'écriture de sommes.
- Effectif, fréquence, quartile, centile : explication et représentations graphiques.
- Calcul d'intervalles : le traitement des données continues.

*Application de chaque notion présentée sur des exercices.*

#### Traitement statistique des données à une dimension

- Type de données : qualitative ou quantitative.
- Données avec effectif : calcul de fréquences et interprétation.
- Tri et traitement des données : mise en forme statistique de différents exemples de données brutes.
- Représentations graphiques.
- Paramètres de position : moyenne, mode, médiane.
- Les paramètres de dispersion : étendue, quantiles, décile, variance.
- La variance : une moyenne "d'écarts".

*Transformation, tri et représentation des données. Mesure de leur dispersion.*

#### Variables aléatoires

- Définition et catégories de variables.
- Exemples et examen de variables aléatoires.
- Courbes de distribution.
- Explications des intervalles de confiance.
- La loi normale.

*Utilisation d'une table de loi normale.*

#### Statistique descriptive à deux dimensions : les tableaux de contingence

- Les données.
- Représentations graphiques.
- La covariance.
- Le coefficient de corrélation linéaire.

*Calcul de covariances et de coefficients de corrélation. Analyse.*

### **Cas pratique : utilisation des données des participants**

- Mise en évidence de la problématique statistique.
- Mise en forme des données.
- Calcul des statistiques de base et représentations graphiques.
- Recherche de la méthode appropriée au problème

### **Participants**

Ce séminaire s'adresse aux personnes qui souhaitent aborder les statistiques avec un bagage solide, une bonne maîtrise des notions avant des séminaires plus complexes sur la modélisation statistique.

Aucune connaissance technique n'est requise pour ce séminaire

Data Science  
DSF

2 Jours

## Les fondamentaux de la Data Science

### Objectifs pédagogiques

La Data Science : être capable d'en donner une définition, connaître les principes de base et l'organisation de la démarche. Appréhender son application afin de résoudre des questions et les limites.

Développer sa capacité d'analyse et d'interprétation des chiffres par la représentation graphique, la maîtrise du vocabulaire, et connaître les domaines d'utilisation.

Être en capacité de décision pour mettre en œuvre et utiliser les modèles et les outils de la Data Science dans un contexte professionnel. La Data Science est au service de la stratégie de développement commerciale et de la gestion de la relation client sur le moyen et long terme

La Data Science est un enjeu pour l'exploitation des données dans un contexte concurrentiel et d'amélioration continue de l'offre et des services

Appréhender l'organisation, les moyens et l'infrastructure pour les services et pour les projets de Data Science dans l'organisation

### Programme

#### La Data Science pour un système d'information de la connaissance

- Comprendre et démystifier l'importance de la Data Science : Analyse descriptive, prédictive et prescriptive
- Définition des concepts : Big data, le Data Lake, Machine Learning, Deep learning,...
- Évolution de la place de la Data Science dans les organisations
- Gouvernance et contraintes réglementaires
- Architecture des systèmes d'information, comment organiser les
- Les nouveaux défis de la Data Science : multiplication des sources de données, hétérogénéités des données et explosion des volumes de données à prendre en compte.
- Les métiers de la Data Science : vocabulaire, compétences, savoir-faire et savoir-être

**Cas pratique : Organiser les compétences et les outils pour déployer l'étude des ventes en flux tendu**

#### Méthodes et modèles

- Les langages de la statistique et du ML : R et Python et leurs environnements
- Les méthodologies
- Les données et la qualité
- Définition des métriques
- Analyse et modélisation d'un problème de Data Science
- Les principales classes d'algorithmes supervisés : Arbres, KNN, Régressions, Naive Bayes
- Les principales classes d'algorithmes non-supervisé : Clustering, ACP, CAH
- Les réseaux de neurones
- Interprétation des résultats
- Le text-mining et IA générative (Chat-GPT)

**Programmer des exemples en Python ou R**



### **Analyse et modélisation d'un problème de Data Science**

- La data visualisation
- Le story telling
- Fondamentaux d'un projet Data Science : les étapes, les rôles, les outils,
- Identifier les besoins et les objectifs métiers,
- La Data Science au service de la prospective (Forecast/Foresight),
- Intégrer la Data Science à la production informatique
- Concevoir un projet Data Science ouvert et évolutif
- Difficultés, craintes et obstacles spécifiques à la Data Science

*Exercice sur un cas métier*

### **Participants**

Ce séminaire s'adresse aux :

- Directeur/Responsable des SI, responsable de projets en lien avec l'analyse de données, responsable d'études statistiques
- Analystes décisionnels souhaitant évoluer vers la Data Science
- Data Steward souhaitant comprendre les enjeux de la Data Science

Data Science  
DSS

2 Jours

## La modélisation statistique

### Objectifs pédagogiques

Ce séminaire permet de maîtriser l'essentiel des techniques de la modélisation statistique. C'est le prolongement des outils du décisionnel pour aller chercher les informations non visibles. Il permet aux participants de découvrir par la pratique les outils de programmation pour mettre en œuvre ces fonctions. C'est un prolongement du séminaire FSTD (Statistiques Descriptives)

### Programme

#### Rappel des fondamentaux de la statistique

- Les éléments de la statistique descriptive.
- L'échantillonnage et l'analyse d'une population.
- Le traitement des variables qualitatives et quantitatives.
- Les effectifs : Fréquences, cumuls
- Représentation graphique des variables.

#### Démarche et modélisation d'une analyse statistique

- Les principes de modélisation.
- Les phases d'apprentissage.
- Les méthodes pour estimer et anticiper.
- Comment modéliser un phénomène ; Exercices pratiques

#### Comprendre et déterminer les paramètres de position et de dispersion

- Mode, valeur modale, valeur la plus probable.
- Moyenne d'une population.
- Médiane, partager une série numérique.
- Étendue, différence entre valeurs extrêmes.
- Utiliser les quantiles.
- Écart-Type, calculer la dispersion d'un ensemble de données.
- Calcul de la variance et de la covariance.

#### Tests et intervalle de confiance

- Lois statistiques et intervalle de confiance.
- Tests statistiques courants : Test de Student, Analyse de variances et  $\text{Khi}^2$
- Valider la précision d'une estimation. Amplitude de l'intervalle.

### Participants

Ce séminaire s'adresse aux personnes qui souhaitent maîtriser ou se remettre à niveau sur les outils avancés des sciences des données.

## Le Data Mining par la Pratique

### Objectifs pédagogiques

Ce séminaire a pour objectif d'apprendre à conduire de bout en bout un projet d'analyse statistique en partant d'une problématique. Il aborde chaque étape de l'analyse avec des exercices, de la programmation sous R et Python. Le formateur encadre la mise en pratique des principales méthodes pour garantir leur acquisition. Les méthodes les plus avancées comme les réseaux de neurones sont également abordées.

### Programme

#### Le projet de Data Mining

- Data Scientist : Sa démarche pour aller de la donnée à l'information.
- Le vocabulaire et les concepts.
- Étape d'exploration descriptive du jeu de données.
- Concevoir et stocker les métadonnées pour le suivi du projet de Data Mining.
- Introduction au logiciel R et à Python

#### Les techniques pour aborder un projet de Data Mining

- Méthode basée sur la classification : identification de groupes d'individus statistiques.
- Méthode par association : mise en évidence d'une cause et d'une conséquence.
- Méthode de l'estimation : complément d'un effectif ou d'une fréquence d'un jeu de données.
- Apports du Data Mining pour le traitement des grandes volumétries de données.
- Méthode de segmentation : définition de critères, extension de la méthode de classification et principe des k-means.
- Méthode de prévision : importance de la temporalité et des hypothèses.

#### Les outils statistiques

- Méthodes descriptives : corrélation, classification, réseaux de Kohonen, règles d'association.
- Méthodes prédictives : régression, arbres de décision, réseaux de neurones, les K plus proches voisins.
- Mise en œuvre de classification par k-means et CAH (Classification Ascendante Hiérarchique).
- Principe des méthodes supervisées.

#### La visualisation des données

- Les objectifs de la visualisation des données.
- Les différents types de représentations de données quantitatives.
- Concevoir des tableaux de bord.

### **Analyse de données qualitatives et textuelles**

- Spécificités de la problématique et les alternatives (Analyse Factorielle des Correspondances, table de contingence).
- Présentation d'instanciation, de pattern, de vecteur et d'heuristique.
- Comment utiliser un espace de vecteurs, d'indexation, de scoring.
- Différents types de transformations et traitement d'un document textuel.

### **Participants**

Ce séminaire s'adresse aux personnes qui souhaitent acquérir les méthodes pour mettre en œuvre un projet de Data Mining et la connaissance des méthodes principales.

Le pré-requis de connaissance des statistiques descriptives est indispensable pour ne pas perdre le fil et tirer pleinement partie de l'ensemble du séminaire.