

Catalogue des formations

Data Science & Intelligence Artificielle

2026

80+ formations 9,3/10 de satisfaction (2025) 1200+ apprenants formés

Certifié Qualiopi · **Financement OPCO** · **Inter & intra-entreprise**

BIENVENUE

Formez vos équipes à la data et à l'IA

Plus de 80 formations en présentiel et à distance, du Python à l'IA générative. Une pédagogie concrète, portée par des praticiens qui déploient la data au quotidien — et qui restent à vos côtés une fois la formation terminée.

- **Des formateurs praticiens.** Chaque formateur intervient sur sa seule spécialité et utilise au quotidien les outils qu'il enseigne.
- **De la pratique, surtout.** Exercices et ateliers sur des jeux de données réels pour être opérationnel dès la fin de la formation.
- **De petits groupes.** 7 participants maximum en inter, pour un accompagnement personnalisé et de vrais échanges.
- **Une qualité certifiée.** Organisme certifié Qualiopi au titre des actions de formation, finançable par les OPCO.

TROIS FORMATS, UN MÊME NIVEAU D'EXIGENCE

- **Présentiel inter-entreprises** : dans nos locaux à Paris, en petits groupes de 7 participants maximum.
- **Sur mesure intra-entreprise** : chez vous, partout en France ; le programme est adapté à votre contexte et à vos données.
- **Classe virtuelle à distance** : en direct avec un formateur, depuis votre navigateur, avec des outils pensés pour l'interactivité.

ILS FORMENT LEURS ÉQUIPES AVEC NOUS

EDF · SNCF · Orange · AXA · L'Oréal · Société Générale · Dior · OCDE ·
Thales · ENGIE · Ubisoft · FDJ · Groupama · CNRS · La Banque Postale ·
Ipsos · MGEN · ASNR

SOMMAIRE

Nos formations par domaine

Astuce : pour mettre à jour les numéros de page, faites un clic droit sur le sommaire puis « Mettre à jour les champs ».

Python	6
Formation python pour la data science	7
Formation python pour l'analyse de données	9
Formation manipuler de grandes données avec Polars, DuckDB et Arrow	11
Formation pandas avancé (method chaining et stylisation)	12
Formation data visualisation avec Python	13
Formation analyse textuelle avec python	14
Formation Industrialisation d'un Projet Data en Python	15
Formation python pour utilisateur de SAS	16
Formation python et Excel - Python pour les utilisateurs d'Excel	17
Formation scikit-learn pour le machine learning	18
Formation Deep Learning avec Python	19
Formation Spark avec Python - utilisation de pyspark	20
Logiciel R	21
Formation logiciel R pour la data science	22
Formation R pour utilisateur de SAS	23
Formation data mining et machine learning avec R	24
Formation visualisation avec R	25
Formation séries temporelles avec R	26
Formation R shiny - développement d'applications web	27
Formation R pour la Cartographie et les SIG	28
IA & Machine Learning	29
Formation intelligence artificielle – Les fondamentaux	30
Formation Acculturation aux modèles génératifs	31
Formation IA générative pour la génération de code	32
Formation LLM pour les développeurs - Formation IA	33
Formation exploiter des modèles d'IA préexistants dans des projets data	34
Formation machine learning et deep learning - les fondamentaux	35
Formation Machine Learning distribué avec Spark ML	36
Formation Introduction au Machine Learning avec Scikit-learn	37
Formation Deep Learning avec PyTorch	38
Formation Deep Learning avec TensorFlow	39
Formation Développer un assistant conversationnel avec Rasa et LlamaIndex	40
Formation Applications d'IA générative avec LangChain	41
Formation Gestion du cycle de vie des modèles avec MLflow	42
Formation IA en entreprise : ML et IA générative	43
Data Engineer & Big Data	44
Formation data science - bonnes pratiques et outils	45
Formation big data et data science - les fondamentaux	46
Formation big data et marketing pour décideurs	47
Formation Data Lake - Panorama et choix des briques d'exploitation d'un data lake	48

Formation traitement de données distribuées avec Spark en Scala.....	49
Formation DBT (Data Build Tool).....	50
Formation ElasticSearch et Kibana pour l'analyse de données	51
Formation NoSQL pour les projets data (MongoDB)	52
Formation Git - gestion de versions	53
Formation DevOps pour la Data Science	54
Business Intelligence	55
Formation Power BI initiation : Découvrez Microsoft Power BI Desktop.....	56
Formation Power BI Dataviz et analytics.....	57
Formation DAX expertise - PowerBI	58
Formation découverte Power BI	59
Formation Tableau Desktop	60
Formation QlikView - Initiation	61
Formation Designer Qlik Sense	62
Formation Développeur Qlik Sense	63
Formation Découverte Qlik Sense	64
Formation Superset pour l'analyse et la visualisation de données	65
Formation Power BI Perfectionnement	66
DataViz.....	67
Formation Datavisualisation - L'art de représenter les données	68
Formation rendre ses présentations claires et engageantes.....	69
Statistiques.....	70
Formation statistique et analyse de données	71
Formation XLSTAT - statistique et analyse de données	73
Formation XLSTAT-R et programmation avec XLSTAT	74
Formation approche PLS (PLS Path Modeling).....	75
Formation équations structurelles à variables latentes	76
Formation régression PLS.....	77
Formation analyse conjointe (méthodes de trade-off).....	78
SQL.....	79
Formation SQL - Bases du langage.....	80
Formation SQL pour la construction de datamarts décisionnels	82
Logiciels Data Science.....	83
Formation Dataiku DSS.....	84
Formation Alteryx pour utilisateurs.....	85
Formation Databricks pour Utilisateurs : Maîtriser l'Analyse et le Traitement des Données	86
Formation Microsoft Azure Machine Learning	88
Formation Snowflake pour utilisateurs - Maîtriser la Gestion et l'Analyse des Données Cloud	89
Formation logiciel KNIME	91
Formation OpenRefine initiation	93
Formation langage Julia	94
Formation AutoML avec H2O.ai.....	95
Gouvernance & Qualité.....	96
Formation Gouvernance des données en entreprise.....	97

Formation qualité des données en entreprise	98
Formation RPA pour les équipes data et SI	99
Culture data	100
Formation culture générale de la donnée - De l'open data au big data	101
Formation Open Data	102
Atelier acculturation data	103

DOMAINE

Python

Le langage de référence de la data, de l'initiation à l'industrialisation des modèles.

12 formations

PYTHON · PYT-001

Formation python pour la data science

Maîtrisez Python pour la data science en 4 jours : manipulation, visualisation et machine learning sur cas concrets.

DURÉE 4 jours (28 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · pandas · NumPy · Matplotlib · Seaborn · Plotly · scikit-learn · Jupyter

PRÉSENTATION

La formation Python pour la data science vous donne en 4 jours les bases solides pour exploiter pleinement le langage Python dans vos projets data. Après une introduction au langage (ou un rappel pour les profils déjà initiés), vous montez en compétences sur les bibliothèques de référence : pandas pour la manipulation de tableaux, NumPy pour le calcul scientifique, Matplotlib, Seaborn et Plotly pour la visualisation, et scikit-learn pour le machine learning. L'apprentissage repose sur les Jupyter notebooks et de nombreux exercices pratiques sur des jeux de données réels, pour vous rendre rapidement opérationnels en analyse de données, data mining et data science. Cette formation s'adresse aux data scientists, data analysts et développeurs souhaitant découvrir ou approfondir Python pour la data science.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les bases du langage Python
- Analyser des données avec Python
- Automatiser le traitement de données avec Python
- Connaître les outils de l'écosystème Python pour la data science

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst souhaitant découvrir Python pour la data science
- Data scientist débutant en Python
- Développeur spécialisé dans d'autres langages
- Manager technique souhaitant comprendre les spécificités de Python

Prérequis

- Connaissance des bases de la programmation (variables, boucles, fonctions) dans un langage quelconque
- Notions de base en analyse de données

PROGRAMME DÉTAILLÉ

JOUR 1 Python aujourd'hui : bases et bonnes pratiques

- Introduction au langage Python : pour qui, pour quoi, comment
- Consoles et débogage en Python
- Syntaxe, structures de données (chaînes, listes, dictionnaires), opérateurs et fonctions
- Classes, objets et création de bibliothèques
- Construction d'un outil de calcul de statistiques simples
- Processus de développement d'un outil Python pour la data science

JOUR 2 Manipulation et préparation des données avec pandas

- Importation de données : CSV, Excel, fichiers texte, bases SQL, web (API, scraping léger)
- DataFrame : sélection, filtrage, tri, création et transformation de variables
- Jointures, concaténations et gestion des données manquantes
- Nettoyage, contrôles de qualité et typage des variables
- Travail sur les dates et les données textuelles
- Agrégations (groupby, pivots) et automatisation des traitements
- Atelier fil rouge : préparation complète d'un jeu de données

JOUR 3 Visualisation et analyse exploratoire

- Principes de la data visualisation : choisir le bon graphique
- Visualisation avec Matplotlib et Seaborn (histogrammes, boxplots, scatter plots)
- Visualisations multivariées et comparaison de groupes

- Analyse exploratoire (EDA) : distributions, anomalies, relations entre variables
- Visualisation interactive avec Plotly et Streamlit
- Atelier : exploration visuelle complète d'un jeu de données

JOUR 4 Introduction au machine learning et mise en production

- Comprendre le machine learning : supervisé / non supervisé, régression vs classification
- Sur-apprentissage, biais, variance et importance de la validation
- Préparation des données pour le ML : train/test, standardisation, encodage
- Pipelines avec scikit-learn
- Cas pratiques : régression, classification, clustering (k-means)
- Industrialisation : pipelines complets, sauvegarde de modèles, passage en production

« Très clair, très complet, donnant les moyens de passer de débutant à pleinement opérationnel en un temps record. Excellente mise en contexte pour permettre de comprendre la raison et la logique de chaque élément. »
— **Christophe W.**

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-python-pour-la-data-science/>

PYTHON · PYT-002

Formation python pour l'analyse de données

3 jours pour devenir autonome sur Python en analyse de données : pandas, visualisation et reporting pour data analysts.

DURÉE 3 jours (21 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance ou à Nice
--------------------------------	------------------------------	---	--

Technologies : Python · pandas · NumPy · Matplotlib · Seaborn · Jupyter · JupyterHub

PRÉSENTATION

La formation Python pour l'analyse de données vous donne en 3 jours les compétences nécessaires pour devenir autonome dans vos analyses quotidiennes. Vous découvrez les fondamentaux du langage Python appliqués à l'analyse, puis vous apprenez à importer, nettoyer et transformer des données issues de sources variées (CSV, Excel, bases de données). Vous manipulez et analysez vos jeux de données avec pandas, réalisez des statistiques descriptives et exploratoires, et créez des visualisations pour communiquer vos résultats. La formation vous montre également comment automatiser des tâches répétitives de reporting et construire des analyses reproductibles avec Jupyter Notebook. Idéale pour data analysts et analystes métier souhaitant dépasser les limites d'Excel.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les fondamentaux du langage Python appliqués à l'analyse de données
- Importer, nettoyer et transformer des données provenant de sources variées (CSV, Excel, bases de données)
- Manipuler et analyser des jeux de données avec pandas
- Réaliser des analyses statistiques descriptives et exploratoires
- Créer des visualisations percutantes pour communiquer ses résultats
- Automatiser des tâches répétitives d'analyse et de reporting
- Construire des analyses reproductibles avec Jupyter Notebook

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst souhaitant découvrir Python pour l'analyse de données
- Analyste métier travaillant déjà sur Excel ou SAS
- Développeur spécialisé dans d'autres langages
- Manager technique souhaitant comprendre les spécificités de Python

Prérequis

- Pratique d'un outil d'analyse de données (Excel, SQL, SAS ou équivalent)
- Aucune connaissance préalable de Python n'est requise

PROGRAMME DÉTAILLÉ

JOUR 1 Python aujourd'hui : bases et bonnes pratiques

- Introduction au langage Python : pour qui, pour quoi, comment
- Consoles et débogage en Python
- Syntaxe, structures de données, opérateurs et fonctions
- Classes, objets et création de bibliothèques
- Construction d'un outil de calcul de statistiques simples
- Processus de développement d'un outil Python pour la data

JOUR 2 Manipulation et préparation des données avec pandas

- Importation de données : CSV, Excel, fichiers texte, bases SQL, web (API, scraping léger)
- DataFrame : sélection, filtrage, tri, transformation de variables
- Jointures, concaténations et gestion des données manquantes
- Nettoyage, contrôles de qualité et typage des variables
- Travail sur les dates et les données textuelles
- Agrégations (groupby, pivots) et automatisation des traitements
- Atelier fil rouge : préparation complète d'un jeu de données

JOUR 3 Visualisation et analyse exploratoire

- Principes de la data visualisation : choisir le bon graphique
- Visualisation avec Matplotlib et Seaborn
- Analyse exploratoire (EDA) : distributions, anomalies, relations entre variables
- Introduction aux visualisations interactives
- Construction d'analyses reproductibles avec Jupyter Notebook
- Atelier : exploration visuelle complète d'un jeu de données

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-python-pour-lanalyse-de-donnees/>

PYTHON · PYT-003

Formation manipuler de grandes données avec Polars, DuckDB et Arrow

Aller au-delà de pandas : traitez de grands volumes de données en Python avec Polars, DuckDB et Arrow, sans Spark.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · Polars · DuckDB · Apache Arrow · pandas · Parquet · Jupyter

PRÉSENTATION

La formation Polars, DuckDB et Arrow vous permet en 2 jours d'aller au-delà de pandas pour manipuler de grands volumes de données en Python, sans recourir à une infrastructure Big Data. Vous découvrez les limites structurelles de pandas et trois technologies modernes pour les dépasser. Polars vous offre des transformations rapides et lisibles grâce à son moteur compilé et son API expressive. DuckDB permet d'exécuter des requêtes SQL analytiques performantes directement sur fichiers (Parquet, CSV, JSON). Apache Arrow assure l'interopérabilité haute performance entre ces outils. Au programme : transformations Polars, requêtes DuckDB, pipelines reproductibles, jusqu'à plusieurs dizaines de millions de lignes sur un poste standard. Pour data analysts, data scientists et développeurs Python orientés data.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les limites structurelles de pandas
- Manipuler efficacement de gros volumes de données sans stack Big Data
- Utiliser Polars pour des transformations rapides et lisibles
- Exploiter DuckDB pour des analyses SQL performantes sur fichiers
- Comprendre le rôle clé d'Apache Arrow dans l'écosystème moderne
- Construire des pipelines data performants et reproductibles

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst confronté à de gros volumes de données
- Data scientist cherchant à dépasser les limites de pandas
- Développeur Python orienté data

Prérequis

- Bonne maîtrise de Python pour la manipulation de données
- Expérience pratique de pandas
- Connaissances de base en SQL

PROGRAMME DÉTAILLÉ

JOUR 1 Polars : le DataFrame moderne et performant

- Pourquoi dépasser pandas : limites mémoire et performance, monothread vs multithread
- Fondamentaux de Polars : DataFrame, types de données et schémas
- Lecture de données volumineuses (CSV, Parquet) et expressions Polars
- Manipulation avancée : sélection, filtrage, chaînage d'expressions
- Agrégations, groupby et jointures efficaces
- Lazy execution et optimisations automatiques (pushdown filters et colonnes)
- Atelier : migration d'un pipeline pandas vers Polars et mesure des gains

JOUR 2 DuckDB, Arrow et pipelines data modernes

- Apache Arrow : format columnar mémoire, zéro-copie, interopérabilité
- Rôle d'Arrow dans Polars, DuckDB et pandas
- DuckDB : le SQL analytique embarqué
- Requêtes SQL performantes directement sur fichiers
- Combinaison Polars / DuckDB / Arrow dans un pipeline
- Atelier : construction d'un pipeline data performant et reproductible

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-polars-duckdb-et-arrow/>

PYTHON · PYT-004

Formation pandas avancé (method chaining et stylisation)

Pandas avancé en 1 jour : method chaining, stylisation, performance et code lisible pour vos pipelines de données.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Python · pandas · NumPy · Matplotlib · Jupyter · Anaconda

PRÉSENTATION

La formation pandas avancé s'adresse aux utilisateurs Python déjà familiers de la bibliothèque qui souhaitent gagner en lisibilité, en performance et en élégance dans leur code. En 1 jour, vous approfondissez des concepts avancés de pandas et adoptez les bonnes pratiques de la communauté. Vous apprenez notamment à écrire vos traitements en method chaining pour produire un code plus expressif et maintenable, à styliser vos DataFrames pour mettre en valeur vos résultats, et à optimiser la représentation et la performance de vos pipelines. La formation alterne théorie et ateliers pratiques sur des jeux de données réels. Une étape idéale pour data engineers, data scientists et développeurs Python souhaitant passer un cap technique sur l'écosystème pandas.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser des concepts avancés de pandas
- Écrire un traitement en method chaining
- Améliorer la représentation des données
- Styliser les DataFrames

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data engineer et data scientist souhaitant approfondir leur connaissance de pandas
- Développeur Python travaillant sur des pipelines de données
- Analyste avancé cherchant à industrialiser ses traitements

Prérequis

- Bonne maîtrise de Python
- Pratique régulière de pandas (DataFrames, groupby, merge)

PROGRAMME DÉTAILLÉ

JOUR 1 Pandas avancé : method chaining et stylisation

- Rappels sur NumPy et pandas (Series et DataFrames)
- Method chaining pour la préparation et le traitement des données
- Écriture de traitements lisibles et maintenables
- Amélioration de la représentation des données
- Stylisation des DataFrames pour la restitution
- Bonnes pratiques de performance et optimisation

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-pandas-avance/>

PYTHON · PYT-005

Formation data visualisation avec Python

Visualisez vos données comme un pro avec Python : matplotlib, seaborn, plotly et bonnes pratiques de design graphique.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · Matplotlib · Seaborn · Plotly · Folium · Dash · Streamlit · Jupyter

PRÉSENTATION

La data visualisation est un levier essentiel pour comprendre et communiquer des résultats analytiques. La formation data visualisation avec Python vous apprend en 2 jours à créer des visuels clairs, esthétiques et adaptés à votre message avec les bibliothèques de référence : Matplotlib, Seaborn et Plotly. Vous découvrirez aussi comment construire des applications web interactives de data visualisation avec Dash ou Streamlit, pour partager vos analyses au-delà du notebook. Au-delà des outils, la formation aborde les principes du design graphique appliqués aux données : choix du type de visuel, palette, hiérarchie de l'information, storytelling. Idéale pour data analysts, data scientists et chefs de projet souhaitant valoriser leurs analyses.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les outils pour mettre en valeur graphiquement ses données avec Python
- Créer des data visualisations complexes avec Matplotlib et Seaborn
- Créer des applications web de data visualisation avec Dash ou Streamlit

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et data scientist souhaitant valoriser leurs analyses
- Développeur Python construisant des outils de visualisation
- Chef de projet et responsable data souhaitant produire des rapports impactants

Prérequis

- Bases de Python et pratique de pandas
- Connaissance des principes de manipulation de données

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux et graphiques statiques avec Python

- Bonnes pratiques de la data visualisation : quelle visualisation selon le contexte
- Pièges à éviter pour visualiser des données et des résultats
- Le package Matplotlib : panorama des graphiques et gestion des options
- Nuages de points, bar plots, contour plots, histogrammes, graphiques 3D
- Seaborn pour les graphiques scientifiques : box plot, pair plot, matrices de graphiques
- Ajout d'informations : annotations, légendes, colorbars

JOUR 2 Visualisation avancée, interactive et applications web

- Graphiques animés et génération de vidéos avec Matplotlib
- Affichage de données massives
- Bases de cartographie : cartopy et cartes interactives avec Folium
- Le package Plotly et la création de graphiques web interactifs
- Applications web avec Streamlit et Dash : principes et construction d'un tableau de bord
- Construction d'une application de reporting et passage en production

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-data-visualisation-avec-python/>

PYTHON · PYT-006

Formation analyse textuelle avec python

Initiez-vous au NLP en Python : nettoyage de texte, embeddings, classification de documents et analyse de sentiments.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · spaCy · NLTK · scikit-learn · pandas · Jupyter

PRÉSENTATION

Le traitement automatique du langage naturel (NLP) est devenu un pilier de la data science, avec des applications dans la classification de documents, l'analyse de sentiments, l'extraction d'information ou les chatbots. La formation analyse textuelle avec Python aborde sur 2 jours les techniques essentielles du NLP. Vous maîtrisez les outils pour traiter des données textuelles : prétraitement, tokenisation, vectorisation (TF-IDF, embeddings), classification, clustering. Vous apprenez à identifier la problématique adaptée à chaque type de projet NLP et à choisir les bons outils parmi spaCy, NLTK, scikit-learn et les apports récents des modèles de langage. Tournée pratique, cette formation s'appuie sur des corpus réels pour vous rendre opérationnel sur vos propres projets NLP en Python.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les outils pour traiter des données textuelles avec Python
- Identifier la problématique et les outils à utiliser pour traiter des problèmes de NLP
- Mener à bien un projet de traitement NLP avec Python

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist souhaitant aborder le NLP
- Data analyst exploitant des données textuelles
- Développeur Python intégrant des fonctionnalités d'analyse de texte

Prérequis

- Bonne maîtrise de Python
- Bases du machine learning et de pandas

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du NLP et prétraitement de texte

- Panorama du traitement automatique du langage naturel (NLP) et cas d'usage
- Identification de la problématique et choix des outils
- Prétraitement : nettoyage, tokenisation, normalisation, stop words, lemmatisation
- Manipulation de corpus avec spaCy et NLTK
- Vectorisation : sac de mots, TF-IDF
- Atelier : préparation d'un corpus de documents

JOUR 2 Modélisation et applications NLP

- Représentations vectorielles : word embeddings et embeddings de phrases
- Classification de documents et analyse de sentiments
- Clustering et extraction de thématiques (topic modeling)
- Apports des grands modèles de langage (LLM) pour le NLP
- Évaluation des modèles de traitement du langage
- Atelier : mise en place d'un projet NLP de bout en bout

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-analyse-textuelle-python/>

PYTHON · PYT-007

Formation Industrialisation d'un Projet Data en Python

Passez du notebook à la production : industrialisez vos projets data Python avec packaging, tests, CI/CD et MLOps.

DURÉE 3 jours (21 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · Git · Docker · pytest · CI/CD · scikit-learn · FastAPI

PRÉSENTATION

La formation industrialisation d'un projet data en Python aide les équipes data à sortir du notebook et à livrer du code prêt pour la production. Sur 3 jours, vous découvrirez les techniques de software engineering appliquées à la data science : structuration de projet, packaging, écriture de tests, gestion de versions, intégration continue. Vous créez une version packagée et testée d'un modèle simple, puis vous réalisez les différentes étapes d'une mise en production : sérialisation, exposition via API, conteneurisation, supervision. La formation s'appuie sur des ateliers progressifs autour d'un cas concret. Idéale pour data scientists, data engineers ou développeurs souhaitant adopter les bonnes pratiques du développement logiciel appliquées à la data.

OBJECTIFS PÉDAGOGIQUES

- Découvrir les techniques de software engineering appliquées à la data science
- Créer une version packagée et testée d'un modèle simple
- Réaliser les différentes étapes d'une mise en production d'un modèle de data science

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist souhaitant mettre ses modèles en production
- Data engineer responsable du déploiement de projets data
- Développeur Python intervenant sur des projets data

Prérequis

- Bonne maîtrise de Python
- Pratique de pandas et scikit-learn ou équivalent
- Connaissance basique de Git appréciée

PROGRAMME DÉTAILLÉ

JOUR 1 Software engineering pour la data science

- Techniques de software engineering appliquées à la data science
- Structuration d'un projet Python : organisation, modules, environnements
- Gestion de versions avec Git et bonnes pratiques de code
- Qualité de code : linting, formatage, typage

JOUR 2 Packaging, tests et reproductibilité

- Packaging d'un projet Python
- Écriture de tests unitaires et d'intégration
- Création d'une version packagée et testée d'un modèle simple
- Gestion des dépendances et reproductibilité des environnements

JOUR 3 Mise en production d'un modèle de data science

- Étapes d'une mise en production d'un modèle
- Sérialisation et exposition d'un modèle via une API
- Conteneurisation avec Docker
- Intégration continue et déploiement (CI/CD)
- Supervision et bonnes pratiques MLOps

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-industrialisation-data-python/>

PYTHON · PYT-008

Formation python pour utilisateur de SAS

Migrez de SAS vers Python en 2 jours : équivalences pas à pas, étapes data, procédures et bonnes pratiques de transition.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · pandas · NumPy · scikit-learn · statsmodels · Jupyter · Anaconda · SAS

PRÉSENTATION

Conçue pour les utilisateurs SAS souhaitant transposer leurs compétences vers Python, cette formation propose un parcours guidé pas à pas en 2 jours. Vous comprenez les différences et points communs entre SAS et Python, puis vous apprenez à traduire vos codes SAS en codes Python équivalents. Vous découvrez les packages Python de gestion de données et leurs correspondants SAS, en travaillant avec pandas, NumPy et scikit-learn. Au programme : manipulation de données, statistiques descriptives, tests statistiques, modélisation prédictive. La formation est rythmée par des ateliers reprenant des cas typiques rencontrés en environnement SAS, transposés en Python, pour faciliter une migration en douceur. Idéale pour statisticiens, biostatisticiens et analystes en transition vers Python.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les différences et points communs entre SAS et Python
- Traduire des codes SAS en codes Python
- Connaître les packages Python de gestion de données et leur équivalent SAS

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Statisticien et biostatisticien utilisateur de SAS
- Data analyst SAS souhaitant migrer vers Python
- Manager d'équipes en transition technologique SAS vers Python

Prérequis

- Pratique courante de SAS (étape DATA et procédures usuelles)
- Aucune connaissance préalable de Python n'est requise

PROGRAMME DÉTAILLÉ

JOUR 1 De SAS à Python : fondamentaux et manipulation des données

- Le langage Python et sa comparaison avec SAS
- Python, un investissement pour l'avenir
- Environnements de développement : IDE et Jupyter Notebook, packages pour la data
- Les bases du langage : syntaxe, structures de données, opérateurs et fonctions
- Chargement de données : procédures SAS et fonctions Python équivalentes
- Manipulation de données : comparaison étape DATA SAS et package pandas

JOUR 2 Statistiques, analyse et machine learning

- Nettoyage et préparation des données
- Statistiques descriptives : procédures SAS vs fonctions Python
- Construction de graphiques et tableaux de bord avec Python
- Comparaison de SAS et de statsmodels pour la modélisation
- Aller plus loin avec le machine learning et scikit-learn
- Discussions et démonstrations sur les problématiques des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/python-pour-utilisateur-de-sas/>

PYTHON · PYT-009

Formation python et Excel - Python pour les utilisateurs d'Excel

Automatisez vos traitements Excel avec Python : openpyxl, pandas et passerelle Excel/Python pour gagner en productivité.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · pandas · Excel · openpyxl · Jupyter · Anaconda

PRÉSENTATION

La formation Python et Excel s'adresse aux utilisateurs d'Excel souhaitant dépasser les limites du tableur grâce à l'automatisation Python. Sur 2 jours, vous apprenez les bases du langage Python, manipulez des données avec pandas, et faites communiquer Python et Excel via openpyxl, xlwings ou la nouvelle intégration native Python in Excel. Vous construisez des visualisations Python directement dans Excel et apprenez à coder en Python dans Excel. Aucun prérequis Python n'est nécessaire : la formation part des concepts essentiels et vous rend rapidement autonome sur des cas concrets de reporting, consolidation ou nettoyage de données. Idéale pour analystes, contrôleurs de gestion et utilisateurs avancés d'Excel cherchant à industrialiser leurs traitements.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les bases du langage Python
- Manipuler des données avec Python
- Faire communiquer Python et Excel
- Construire des visualisations Python dans Excel
- Coder en Python directement dans Excel

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Utilisateur avancé d'Excel souhaitant gagner en productivité
- Analyste métier et contrôleur de gestion
- Profil data débutant souhaitant industrialiser ses traitements Excel

Prérequis

- Bonne maîtrise d'Excel (formules, tableaux croisés dynamiques)
- Aucune connaissance préalable de Python n'est requise

PROGRAMME DÉTAILLÉ

JOUR 1 Python et traitement des données

- Introduction au langage Python : pour qui, pour quoi, comment
- Les différences avec Excel et comment coder en Python pour Excel
- Syntaxe, structures de données, opérateurs et fonctions
- Manipulation de données avec pandas : jointure, filtre, transformation
- Traitement des données manquantes

JOUR 2 Faire communiquer Python et Excel

- Les différentes approches : Jupyter Notebook, compléments (pyxl), Python dans Excel
- Lecture et écriture de fichiers Excel avec Python
- Maîtrise du chargement des formules Excel en Python
- Génération de graphiques dans Excel avec Python
- Communication Excel vers PowerPoint et automatisation de rapports
- Python dans Excel (partenariat Microsoft-Anaconda) : ML et data visualisation sans sortir d'Excel

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-python-et-excel/>

PYTHON · PYT-010

Formation scikit-learn pour le machine learning

Maîtrisez scikit-learn pour le machine learning : pipelines, modèles supervisés et non supervisés, évaluation et tuning.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · scikit-learn · pandas · NumPy · Jupyter · Anaconda

PRÉSENTATION

La formation scikit-learn vous donne en 2 jours les clés pour construire des modèles de machine learning robustes en Python. Vous découvrez les principaux concepts du machine learning : apprentissage supervisé, non supervisé, séparation entraînement/test, validation croisée, métriques d'évaluation. Vous apprenez à construire un pipeline complet avec scikit-learn, en intégrant prétraitement, sélection de variables, entraînement et évaluation. La formation vous fait maîtriser toutes les étapes de création d'un modèle, de la préparation des données à la validation des performances. Très orientée pratique, elle s'appuie sur des jeux de données réels et des cas métier variés (classification, régression, clustering). Indispensable pour tout data scientist en formation.

OBJECTIFS PÉDAGOGIQUES

- Connaître les principaux concepts du machine learning
- Construire un pipeline de machine learning avec scikit-learn
- Maîtriser les étapes de création d'un modèle, de la préparation à la validation

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist en formation sur le machine learning
- Data analyst souhaitant aborder la modélisation prédictive
- Développeur Python intégrant des modèles ML à ses projets

Prérequis

- Bonne maîtrise de Python et de pandas
- Notions de base en statistique

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux et premiers modèles avec scikit-learn

- Les concepts du machine learning et le processus de construction d'un modèle
- L'écosystème scikit-learn (Python, NumPy, pandas)
- Les principes de scikit-learn et la notion de pipeline
- Un premier modèle de machine learning avec scikit-learn
- Préparation des données et preprocessing : données qualitatives, manquantes, ColumnTransformer

JOUR 2 Sélection, évaluation et méthodes avancées

- Gestion du sur-apprentissage et validation (train/test, validation croisée)
- Les métriques d'évaluation des modèles
- Ajustement des hyperparamètres et combinaison dans un pipeline
- Focus sur les méthodes : modèles linéaires, modèles à base d'arbres (forêts aléatoires, GBM)
- Modèles non supervisés avec scikit-learn
- Passage en production d'un modèle : les différentes approches

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-scikit-learn-pour-le-machine-learning/>

PYTHON · PYT-011

Formation Deep Learning avec Python

Découvrez le deep learning en Python avec Keras et PyTorch : réseaux de neurones, CNN, RNN et applications pratiques.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris et à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · Keras · TensorFlow · PyTorch · NumPy · Jupyter

PRÉSENTATION

La formation Deep Learning avec Python vous accompagne sur 2 jours dans la découverte et la mise en pratique des réseaux de neurones profonds. Vous comprenez les principes du deep learning et les différents types de réseaux : perceptrons multicouches, réseaux convolutifs (CNN) pour l'image, réseaux récurrents (RNN) et transformeurs pour les séquences. Vous maîtrisez les outils pour faire du deep learning avec Python en travaillant avec Keras / TensorFlow et PyTorch, les deux frameworks de référence. La formation vous guide sur la mise en place complète d'un projet de deep learning : préparation des données, choix d'architecture, entraînement, régularisation, évaluation. Idéale pour data scientists et ingénieurs souhaitant intégrer le deep learning à leur boîte à outils.

OBJECTIFS PÉDAGOGIQUES

- Comprendre le deep learning et les différents réseaux de neurones
- Maîtriser les outils pour faire du deep learning avec Python
- Mener à bien un projet de mise en place d'algorithmes de deep learning avec Python

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist souhaitant ajouter le deep learning à sa palette
- Ingénieur ML travaillant sur des projets image, son ou langage
- Développeur Python orienté IA

Prérequis

- Bonne maîtrise de Python et de pandas
- Bases du machine learning (idéalement scikit-learn)
- Notions de calcul matriciel

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du deep learning et premiers réseaux

- Comprendre le deep learning et les différents réseaux de neurones
- Principes : neurones, couches, fonctions d'activation, rétropropagation
- Optimisation, régularisation et bonnes pratiques d'entraînement
- Outils pour le deep learning avec Python : Keras / TensorFlow et PyTorch
- Construction d'un premier réseau de neurones (perceptron multicouche)

JOUR 2 Réseaux avancés et mise en œuvre de projets

- Réseaux convolutifs (CNN) pour le traitement d'images
- Réseaux récurrents (RNN) et transformeurs pour les séquences
- Transfer learning et utilisation de modèles pré-entraînés
- Mener à bien un projet de deep learning de bout en bout
- Atelier : mise en place d'un algorithme de deep learning sur un cas concret

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-deep-learning-avec-python/>

PYTHON · PYT-012

Formation Spark avec Python - utilisation de pyspark

Traitez de grands volumes avec PySpark : DataFrame Spark, SQL distribué et machine learning à l'échelle Big Data.

DURÉE 3 jours (21 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Apache Spark · PySpark · Python · Spark SQL · MLlib · Jupyter

PRÉSENTATION

La formation Spark avec Python vous initie en 3 jours au traitement distribué de grands volumes de données via PySpark. Vous comprenez l'environnement Apache Spark : architecture, exécution distribuée, lazy evaluation, gestion de la mémoire. Vous apprenez à utiliser le package PySpark pour communiquer avec Spark depuis Python, manipulez des DataFrames distribués, et maîtrisez Spark SQL pour exécuter des requêtes à grande échelle. Vous abordez également MLlib pour le machine learning distribué sur de très grands jeux de données. La formation est jalonnée d'ateliers pratiques sur cluster, pour vous rendre opérationnel sur vos projets data engineering ou data science en environnement Big Data.

OBJECTIFS PÉDAGOGIQUES

- Comprendre l'environnement Apache Spark
- Utiliser le package PySpark pour communiquer avec Spark
- Maîtriser Spark SQL
- Maîtriser MLlib

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data engineer et data scientist travaillant sur de grands volumes
- Développeur Python amené à exploiter un cluster Spark
- Architecte data cadrant des projets Big Data

Prérequis

- Bonne maîtrise de Python
- Pratique de pandas
- Connaissances de base en SQL

PROGRAMME DÉTAILLÉ

JOUR 1 L'environnement Spark et PySpark

- Comprendre l'environnement Apache Spark : architecture et exécution distribuée
- Notions de RDD, lazy evaluation et gestion de la mémoire
- Le package PySpark pour communiquer avec Spark depuis Python
- Manipulation de DataFrames distribués

JOUR 2 Spark SQL et traitement de données à grande échelle

- Maîtrise de Spark SQL : requêtes à grande échelle
- Transformations et actions sur de grands volumes
- Lecture et écriture de données (Parquet, CSV, sources distribuées)
- Optimisation des traitements et partitionnement

JOUR 3 Machine learning distribué avec MLlib

- Maîtrise de MLlib pour le machine learning distribué
- Pipelines de machine learning sur Spark
- Entraînement et évaluation de modèles sur de très grands jeux de données
- Atelier : projet complet de data science distribuée sur cluster

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-spark-avec-python/>

DOMAINE

Logiciel R

Le logiciel de référence pour l'analyse, le data mining et les applications Shiny.

7 formations

LOGICIEL R · RDS-001

Formation logiciel R pour la data science

Formez-vous au langage R pour la data science : manipulation tidyverse, visualisation ggplot2 et statistiques appliquées.

DURÉE 3 jours (21 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · tidyverse · dplyr · ggplot2 · data.table

PRÉSENTATION

La formation logiciel R pour la data science vous donne en 3 jours les bases solides pour exploiter le langage R dans vos analyses. Vous maîtrisez l'environnement de travail R avec RStudio, puis manipulez et gérez vos données avec les packages les plus récents : dplyr, tidyr, data.table. Vous découvrez l'application de méthodes de data science : statistiques descriptives, tests, modélisation, visualisation avec ggplot2. La formation aborde également les bases de la programmation avec R (fonctions, structures de contrôle, packages personnalisés). La pédagogie repose sur des cas pratiques issus du quotidien des analystes et data scientists. Une formation R idéale pour toute personne souhaitant adopter un outil open source puissant pour l'analyse statistique.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser l'environnement de travail R avec RStudio
- Manipuler et gérer des données avec R et les packages récents (dplyr, data.table...)
- Maîtriser l'application des méthodes de data science
- Connaître les bases de la programmation avec R

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Statisticien et biostatisticien
- Data analyst découvrant ou approfondissant R
- Chercheur ou doctorant utilisant R pour ses travaux
- Data scientist souhaitant compléter sa boîte à outils

Prérequis

- Aucun prérequis technique sur R
- Bases de la statistique descriptive

PROGRAMME DÉTAILLÉ

JOUR 1 Prise en main de R et de RStudio

- L'environnement de travail R avec RStudio
- Syntaxe de base : objets, vecteurs, fonctions, structures de contrôle
- Types de données et structures (vecteurs, listes, data frames)
- Import et export de données (CSV, Excel, bases de données)
- Premiers traitements et bonnes pratiques

JOUR 2 Manipulation et gestion des données

- Manipulation de données avec le tidyverse : dplyr et tidyr
- Traitement performant avec data.table
- Nettoyage, transformation et recodage des variables
- Jointures, agrégations et données manquantes
- Travail sur les dates et les chaînes de caractères

JOUR 3 Méthodes de data science et programmation

- Statistiques descriptives et premières analyses
- Visualisation des données avec ggplot2
- Application de méthodes de data science
- Bases de la programmation R : fonctions, itérations, packages
- Atelier de synthèse sur un jeu de données réel

« La qualité de la formatrice - Qualité des échanges - Exemples pratiques »
— A. K.

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-logiciel-r/>

LOGICIEL R · RDS-002

Formation R pour utilisateur de SAS

Migrez de SAS à R sans douleur : équivalences SAS/R, étapes data, procédures et bonnes pratiques de transition.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · tidyverse · dplyr · data.table · SAS

PRÉSENTATION

La formation R pour utilisateurs SAS facilite la transition vers un environnement open source. Sur 2 jours, vous comprenez les différences et points communs entre SAS et R, puis apprenez à traduire vos codes SAS en codes R équivalents. Vous découvrez les packages R de gestion de données (tidyverse, data.table) et leur correspondance avec les étapes DATA et procédures SAS. La formation s'appuie sur des ateliers reprenant des cas typiques d'analyse SAS, transposés en R : préparation, statistiques descriptives, tests, régression, reporting. Idéale pour statisticiens, biostatisticiens et analystes souhaitant élargir leur boîte à outils ou préparer une migration vers R.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les différences et points communs entre SAS et R
- Traduire des codes SAS en codes R
- Connaître les packages R de gestion de données et leur équivalent SAS

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Statisticien et biostatisticien utilisateur de SAS
- Data analyst SAS souhaitant adopter R
- Manager d'équipes en transition SAS vers R

Prérequis

- Pratique courante de SAS (étape DATA et procédures usuelles)
- Aucune connaissance préalable de R n'est requise

PROGRAMME DÉTAILLÉ

JOUR 1 De SAS à R : fondamentaux et manipulation

- Différences et points communs entre SAS et R
- Environnement R et RStudio pour un utilisateur SAS
- Équivalents R de l'étape DATA : packages de gestion de données
- Manipulation de données avec dplyr et data.table
- Traduction de codes SAS courants en R

JOUR 2 Statistiques et analyses avancées

- Statistiques descriptives : procédures SAS vs fonctions R
- Tests statistiques et modélisation en R
- Visualisation avec ggplot2
- Reporting et automatisation
- Traduction de cas d'analyse SAS vers R sur les données des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-r-pour-utilisateur-de-sas/>

LOGICIEL R · RDS-003

Formation data mining et machine learning avec R

Faites du data mining et du machine learning avec R : arbres, random forest, clustering et évaluation de modèles.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · caret · tidymodels · randomForest

PRÉSENTATION

La formation data mining et machine learning avec R vous apprend en 2 jours à construire des modèles prédictifs robustes. Vous maîtrisez l'utilisation des packages R dédiés au machine learning, notamment caret et tidymodels, qui standardisent la construction et la comparaison de modèles. Vous comprenez le processus complet de traitement en machine learning : préparation des données, séparation entraînement/test, choix d'algorithme, entraînement, validation, mise en production. La formation couvre les principaux algorithmes (arbres, random forest, boosting, régression logistique, clustering) et vous apprend à interpréter les sorties d'un modèle. Une formation idéale pour analystes, statisticiens et data scientists travaillant en environnement R.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser l'utilisation des packages R pour le machine learning
- Comprendre le processus de traitement en machine learning
- Interpréter les sorties d'un algorithme de machine learning avec R

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et statisticien souhaitant aborder le machine learning
- Data scientist travaillant en environnement R
- Chargé d'étude exploitant des modèles prédictifs

Prérequis

- Bases de R (manipulation de données, fonctions)
- Notions de statistique descriptive

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du machine learning avec R

- Le processus de traitement en machine learning
- Les packages R pour le machine learning : caret et tidymodels
- Préparation des données et feature engineering
- Séparation entraînement / test et validation croisée
- Premiers modèles : régression et classification

JOUR 2 Algorithmes avancés et interprétation

- Arbres de décision, random forest et boosting
- Modèles non supervisés : clustering et réduction de dimension
- Métriques d'évaluation et courbes ROC
- Interprétation des sorties d'un algorithme
- Atelier : projet de data mining de bout en bout

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-data-mining-avec-r/>

LOGICIEL R · RDS-004

Formation visualisation avec R

Créez des visualisations élégantes avec R et ggplot2 : grammaire des graphiques, thèmes, interactivité et storytelling.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · ggplot2 · plotly · leaflet

PRÉSENTATION

La formation visualisation avec R vous donne en 2 jours les compétences pour mettre en valeur graphiquement vos données. Vous maîtrisez les outils dédiés à la visualisation dans l'écosystème R et créez des data visualisations complexes avec ggplot2, la bibliothèque de référence. Vous apprenez la grammaire des graphiques, personnalisez vos visuels avec des thèmes professionnels, et abordez les solutions interactives (plotly, leaflet) pour partager vos analyses. La formation aborde aussi les principes de design et de storytelling appliqués aux données. Idéale pour data analysts, data scientists, chercheurs et toute personne travaillant avec R souhaitant produire des visuels clairs et engageants.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les outils pour mettre en valeur graphiquement ses données avec R
- Créer des data visualisations complexes avec ggplot2

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et data scientist travaillant avec R
- Statisticien et chercheur produisant des rapports analytiques
- Toute personne utilisant R souhaitant produire des visuels impactants

Prérequis

- Bases de R (manipulation de données avec dplyr ou base R)

PROGRAMME DÉTAILLÉ

JOUR 1 Les fondamentaux de ggplot2

- Principes de la data visualisation et choix des graphiques
- La grammaire des graphiques avec ggplot2
- Graphiques de base : points, lignes, barres, histogrammes, boxplots
- Esthétiques, échelles et facettes
- Personnalisation : thèmes, couleurs, annotations

JOUR 2 Visualisations complexes et interactives

- Visualisations multivariées et avancées avec ggplot2
- Extensions de l'écosystème ggplot2
- Graphiques interactifs avec plotly
- Cartographie interactive avec leaflet
- Storytelling et restitution professionnelle

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-visualisation-avec-r/>

LOGICIEL R · RDS-005

Formation séries temporelles avec R

Analysez et prévoyez vos séries temporelles avec R : ARIMA, lissage exponentiel, prophet et modèles de prévision.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · forecast · prophet · lubridate · xts · zoo · ggplot2

PRÉSENTATION

La formation séries temporelles avec R vous apprend en 2 jours à analyser et prévoir des données chronologiques. Vous maîtrisez les principes du traitement des séries temporelles : stationnarité, autocorrélation, saisonnalité, tendance. Vous apprenez à définir et représenter des séries temporelles sous R, à choisir la méthode adaptée à votre problématique parmi les approches classiques (lissage exponentiel, ARIMA, SARIMA) et modernes (prophet, modèles à espace d'état). Vous utilisez R pour appliquer les méthodes de description et de prévision, puis vous savez lire et interpréter les résultats obtenus. Les ateliers pratiques portent sur des cas concrets : ventes, indicateurs économiques, mesures industrielles.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les principes du traitement des séries temporelles
- Définir et représenter des séries temporelles sous R
- Choisir la méthode adaptée à sa problématique
- Utiliser R pour appliquer les méthodes de description et de prévision
- Lire et interpréter les résultats obtenus

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Statisticien et data analyst travaillant sur des données chronologiques
- Économiste, planificateur, prévisionniste
- Data scientist abordant la prévision

Prérequis

- Bases de R (manipulation de données, fonctions)
- Notions de statistique inférentielle

PROGRAMME DÉTAILLÉ

JOUR 1 Comprendre et explorer les séries temporelles

- Introduction aux séries temporelles : définition, intérêt et applications
- Caractéristiques des données temporelles : saisonnalité, tendance, cycles, bruit
- Gestion des dates et formats temporels (lubridate, zoo, xts)
- Création et manipulation d'objets séries temporelles (ts, xts, zoo)
- Analyse descriptive : moyenne, variance, autocorrélation, saisonnalité
- Cas pratique : analyse d'une série de consommation énergétique

JOUR 2 Modélisation et prévision des séries temporelles

- Techniques de filtrage et de lissage (Holt-Winters, moyennes mobiles)
- Modèles ARIMA : identification, estimation et diagnostic
- Séries non stationnaires : différenciation et transformation
- Automatisation avec auto.arima()
- Approches modernes de prévision (prophet)
- Lecture et interprétation des résultats

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-series-temporelles-r/>

LOGICIEL R · RDS-006

Formation R shiny - développement d'applications web

Créez des applications web interactives avec R Shiny : tableaux de bord, dashboards et outils de visualisation en ligne.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · Shiny · shinyWidgets · DT · shinyapps.io

PRÉSENTATION

La formation R Shiny vous apprend en 2 jours à transformer vos analyses R en applications web interactives. Vous construisez des applications web basées sur vos données avec le package Shiny de R, sans nécessiter de compétences en développement web traditionnel. Vous mettez en place l'architecture d'une application Shiny (UI/server, modules, gestion d'état) et apprenez à partager votre application Shiny : déploiement sur shinyapps.io, Posit Connect ou serveur interne. La formation aborde aussi l'intégration de graphiques interactifs, les bonnes pratiques de design et la gestion de la performance. Idéale pour data scientists, analystes et chercheurs souhaitant diffuser leurs résultats sous forme d'outils en ligne.

OBJECTIFS PÉDAGOGIQUES

- Construire des applications web basées sur ses données avec le package Shiny de R
- Mettre en place l'architecture d'une application Shiny
- Partager son application Shiny

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist et data analyst souhaitant partager leurs analyses
- Chercheur diffusant des outils interactifs
- Statisticien construisant des applications métier

Prérequis

- Bonne maîtrise de R
- Notions de base en HTML/CSS appréciées mais non indispensables

PROGRAMME DÉTAILLÉ

JOUR 1 Principes et structure d'une application Shiny

- Découverte, installation et dépendances de Shiny
- Principes de base : ui.R, server.R, global.R
- Lancement et structure d'une application
- Apparence : panels (wellPanel, conditionalPanel, tabPanel)
- Layouts (fluidRow, flowLayout, fluidPage)
- Les widgets de Shiny et shinyWidgets

JOUR 2 Réactivité, personnalisation et déploiement

- Les sorties (outputs) : renderPlot, renderUI, renderImage
- Affichage de tables avec le package DT
- La réactivité : observe, observeEvent, reactive, eventReactive, isolate
- Personnalisation avancée (HTML, CSS, shinythemes)
- Partage et déploiement : shinyapps.io, serveur Shiny
- Atelier : construction d'une application sur les données des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-developpement-dapplications-web-avec-r-shiny/>

LOGICIEL R · RDS-007

Formation R pour la Cartographie et les SIG

Faites de la cartographie et du SIG avec R : données géospatiales, sf, leaflet et visualisations cartographiques avancées.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : R · RStudio · sf · terra · tmap · ggplot2 · leaflet

PRÉSENTATION

La formation R pour la cartographie et les SIG vous initie en 2 jours à l'analyse géospatiale avec le langage R. Vous apprenez à importer, manipuler et traiter des données géographiques au format vectoriel avec le package sf, devenu standard de l'écosystème R. Vous réalisez différentes opérations de géomatique : jointures spatiales, zones tampons, calcul d'indices topographiques, intersections, agrégations spatiales. Vous produisez des cartes légendées, annotées et interactives avec ggplot2, tmap et leaflet. La formation aborde aussi les fonds de carte, les projections et les jeux de données géographiques de référence. Idéale pour analystes, géomaticiens, chercheurs et data scientists travaillant avec des données localisées.

OBJECTIFS PÉDAGOGIQUES

- Importer, manipuler et traiter des données géographiques au format vectoriel
- Réaliser des opérations de géomatique : jointures spatiales, zones tampons, indices topographiques
- Produire des cartes légendées, annotées et interactives

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Géomaticien et géographe souhaitant utiliser R
- Data analyst travaillant sur des données localisées
- Chercheur exploitant des données spatiales

Prérequis

- Bases de R (manipulation de données, packages)
- Notions de cartographie ou de SIG appréciées

PROGRAMME DÉTAILLÉ

JOUR 1 Données géographiques et manipulation spatiale avec R

- Introduction aux données géospatiales : vecteur et raster
- Le package sf : importer et manipuler des données vectorielles
- Systèmes de coordonnées et projections
- Opérations de géomatique : jointures spatiales, zones tampons, intersections
- Calcul d'indices topographiques

JOUR 2 Cartographie et visualisation

- Cartes statiques avec ggplot2 et tmap
- Cartes légendées et annotées
- Cartes interactives avec leaflet
- Fonds de carte et données de référence
- Atelier : production d'une carte thématique complète

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/r-pour-la-cartographie-sig/>

DOMAINE

IA & Machine Learning

Du machine learning au deep learning et à l'IA générative : RAG, agents, LLM.

14 formations

IA & MACHINE LEARNING · IAM-001

Formation intelligence artificielle – Les fondamentaux

Comprenez les fondamentaux de l'IA en 1 journée : concepts clés, applications, limites et enjeux pour votre organisation.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : ChatGPT · Hugging Face · Mistral · scikit-learn

PRÉSENTATION

L'intelligence artificielle est aujourd'hui au cœur des stratégies d'innovation et de transformation des entreprises, portée par des avancées majeures comme les modèles de langage de type ChatGPT. La formation intelligence artificielle vous offre en une journée une vision claire et accessible des concepts clés de l'IA, sans nécessiter de compétences techniques préalables. Vous comprenez les principes fondamentaux de l'intelligence artificielle, apprenez à reconnaître ses utilisations et les besoins associés, et découvrez les différents cas d'usage dans l'industrie, les services, la santé ou le secteur public. La formation aborde également les limites, les enjeux éthiques et les facteurs clés de succès d'un projet IA. Idéale pour dirigeants, managers et collaborateurs souhaitant acculturer leurs équipes ou cadrer leurs initiatives autour de l'IA.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes de l'intelligence artificielle
- Reconnaître l'utilisation d'IA et les besoins associés
- Connaître les différents cas d'usage de l'intelligence artificielle

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Dirigeant, manager et chef de projet
- Responsable métier souhaitant comprendre l'IA
- Collaborateur impliqué dans des initiatives d'IA

Prérequis

- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Les fondamentaux de l'intelligence artificielle

- Comprendre les principes de l'intelligence artificielle
- Histoire et grandes approches : du symbolique à l'apprentissage profond
- Machine learning, deep learning et IA générative : de quoi parle-t-on ?
- Reconnaître l'utilisation d'IA et identifier les besoins
- Cas d'usage de l'IA par secteur (industrie, services, santé, public)
- Limites, enjeux éthiques et facteurs clés de succès d'un projet IA

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-intelligence-artificielle/>

IA & MACHINE LEARNING · IAM-002

Formation Acculturation aux modèles génératifs

Découvrez les modèles génératifs et l'IA générative : ChatGPT, LLM, applications métier et bonnes pratiques.

DURÉE 1 jour (7 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : ChatGPT · Hugging Face · LLM · Python

PRÉSENTATION

La formation acculturation aux modèles génératifs vous donne en une journée les clés pour comprendre l'IA générative et son impact sur les entreprises. Vous saisissez les enjeux et l'importance de l'IA générative au sein de la roadmap de transformation data des organisations. Vous découvrez le fonctionnement des grands modèles de langage (LLM) comme ChatGPT, Claude ou Gemini, et identifiez les étapes clés d'un projet de mise en place d'IA générative ainsi que les principaux acteurs impliqués (fournisseurs de modèles, intégrateurs, équipes data internes). À l'issue de la formation, vous êtes capable de faciliter la réalisation de projets d'IA générative dans votre métier ou département. Idéale pour dirigeants, managers et chefs de projet en quête d'acculturation rapide.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les enjeux et l'importance de l'IA générative dans la roadmap data transformation des entreprises
- Connaître les étapes d'un projet de mise en place d'IA générative et les principaux acteurs impliqués
- Faciliter la réalisation de projets d'IA générative dans son métier ou département

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Dirigeant et manager souhaitant cadrer une stratégie IA générative
- Chef de projet métier ou data
- Collaborateur curieux des apports de l'IA générative

Prérequis

- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Acculturation aux modèles génératifs

- Module 1 - Introduction au NLP, au Deep Learning et à la Gen AI
- Principes de l'IA générative et cas d'usage réels (image, son, texte, vidéo)
- Nouvelles compétences et acteurs de l'IA générative
- Module 2 - Algorithmes et outils : prompt engineering
- Tokenisation, mécanisme d'attention et embeddings
- Enjeux de mise en place de l'IA générative dans l'entreprise

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-acculturation-ia-aux-modeles-generatifs/>

IA & MACHINE LEARNING · IAM-003

Formation IA générative pour la génération de code

Boostez votre productivité dev avec l'IA générative : Copilot, ChatGPT, Cursor et bonnes pratiques de génération de code.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : GitHub Copilot · ChatGPT · Cursor · VS Code · MCP · Git

PRÉSENTATION

La formation IA générative pour la génération de code accompagne développeurs et équipes techniques dans l'adoption efficace des assistants IA. Sur une journée, vous apprenez à utiliser les capacités avancées des IA génératives pour optimiser la production et la qualité du code. Vous maîtrisez les techniques de génération automatique de code, de tests et de documentation avec des outils comme GitHub Copilot, ChatGPT, Claude ou Cursor. Vous découvrez comment identifier et corriger efficacement bugs et vulnérabilités grâce à l'IA, et appliquez les bonnes pratiques d'intégration sécurisée et responsable de l'IA dans les environnements de développement. La formation insiste sur l'importance de garder un regard critique sur le code produit. Idéale pour développeurs, lead techs et CTO.

OBJECTIFS PÉDAGOGIQUES

- Utiliser les capacités avancées des IA génératives pour optimiser la production et la qualité du code
- Maîtriser les techniques de génération automatique de code, de tests et de documentation
- Identifier et corriger efficacement bugs et vulnérabilités via l'IA
- Appliquer les bonnes pratiques d'intégration sécurisée et responsable de l'IA dans les environnements de développement
- Garder un regard critique sur le code produit

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Développeur souhaitant intégrer l'IA à ses pratiques
- Lead tech et tech lead encadrant des équipes de développement
- CTO et responsable d'ingénierie

Prérequis

- Expérience pratique du développement logiciel
- Familiarité avec un IDE et avec Git

PROGRAMME DÉTAILLÉ

JOUR 1 Introduction et pratique de l'IA générative pour le code

- Principes de fonctionnement des IA génératives appliquées au code
- Limites, risques et apports réels pour le développement
- Panorama des principaux outils IA pour le développement
- Coding agentique : la boucle plan → edit → run → fix → commit
- Démo agent loop (VS Code Copilot Agent Mode)
- Atelier guidé : feature + tests + documentation, du brief à la code review

JOUR 2 Optimisation, MCP et bonnes pratiques

- Les MCP en pratique : configuration de serveurs et outils (GitHub, scripts, docs, tickets)
- Du single agent au multi-agent : spécialisation, coordination, parallélisation
- Personnalisation avancée des environnements (MCP, extensions, skills)
- Garde-fous et LLMOps : politiques d'accès, gestion des secrets, validation par tests
- Sandboxing d'agent et règles de confidentialité
- Atelier : création d'un MCP sur une ressource interne

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-ia-generative-generation-code/>

IA & MACHINE LEARNING · IAM-004

Formation LLM pour les développeurs - Formation IA

Intégrez les LLM dans vos applications : API, prompt engineering, RAG, fine-tuning et patterns d'architecture IA.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · Hugging Face Transformers · OpenAI API · LangChain · bases vectorielles

PRÉSENTATION

La formation LLM pour les développeurs aborde sur 2 jours l'intégration des grands modèles de langage dans des applications réelles. Vous acquérez une vision claire des étapes d'un projet de mise en place d'IA générative et des principaux acteurs impliqués. Vous identifiez les différents outils et la complexité de chaque étape d'un projet d'IA générative textuelle : choix de modèle, prompt engineering, architectures RAG (retrieval augmented generation), bases vectorielles, fine-tuning, évaluation et déploiement. À l'issue, vous êtes capable de faciliter la réalisation de projets d'IA générative dans votre métier ou département. Une formation indispensable pour développeurs, architectes et lead techs souhaitant industrialiser l'IA générative dans leur SI.

OBJECTIFS PÉDAGOGIQUES

- Connaître les étapes d'un projet de mise en place d'IA générative et les principaux acteurs impliqués
- Identifier les différents outils et la complexité de chaque étape pour la réalisation d'un projet d'IA générative textuelle
- Faciliter la réalisation de projets d'IA générative dans son métier ou département

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Développeur intégrant des LLM dans des applications
- Architecte logiciel concevant des solutions IA
- Data scientist se spécialisant sur les LLM

Prérequis

- Bonne maîtrise d'un langage de programmation (Python recommandé)
- Notions sur les API REST

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux des LLM et des Transformers

- Module 1 - Introduction au NLP, au Deep Learning et à la Gen AI
- Cas d'utilisation de l'IA générative dans le monde réel
- Module 2 - Les bases des Transformers et des modèles de langage
- Prompt engineering : concevoir des prompts efficaces
- Tokenisation, mécanisme d'attention et embeddings
- Structure et comportement des modèles de langage

JOUR 2 Intégration des LLM dans les applications

- Appels d'API de modèles (OpenAI, Anthropic, modèles open source)
- Architectures RAG (retrieval augmented generation) et bases vectorielles
- Notions de fine-tuning et d'adaptation de modèles
- Conception d'agents et d'assistants métier
- Évaluation, coûts et bonnes pratiques de mise en production
- Atelier : construction d'un cas d'usage LLM de bout en bout

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-llms-pour-les-developpeurs/>

IA & MACHINE LEARNING · IAM-005

Formation exploiter des modèles d'IA préexistants dans des projets data

Exploitez les modèles d'IA préexistants dans vos projets : Hugging Face, APIs et intégration de modèles sur étagère.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · scikit-learn · Hugging Face · PyTorch · pandas · NumPy · APIs IA

PRÉSENTATION

La formation exploiter des modèles d'IA préexistants vous apprend en 2 jours à intégrer des modèles existants plutôt que de tout réinventer. Vous comprenez les différentes familles de modèles d'IA disponibles (vision, langage, audio, multimodaux) et identifiez les cas d'usage adaptés à l'exploitation de modèles préexistants. Vous découvrez l'écosystème Hugging Face, les principales APIs commerciales (OpenAI, Anthropic, Mistral, Google), et apprenez à intégrer des modèles d'IA dans un projet data. Vous évaluez la qualité, les limites et les biais des modèles utilisés, puis exploitez les résultats de manière fiable et responsable. La formation aborde aussi les questions de coût, de performance et de souveraineté. Idéale pour data scientists et développeurs cherchant à accélérer leurs projets IA.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les différentes familles de modèles d'IA existants
- Identifier des cas d'usage adaptés à l'exploitation de modèles préexistants
- Intégrer des modèles d'IA dans un projet data
- Évaluer la qualité, les limites et les biais des modèles utilisés
- Exploiter les résultats de manière fiable et responsable

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist intégrant des modèles existants
- Développeur souhaitant ajouter des fonctionnalités IA à ses produits
- Chef de projet data cadrant des solutions à base d'IA

Prérequis

- Bonne maîtrise de Python
- Notions de base en machine learning

PROGRAMME DÉTAILLÉ

JOUR 1 Panorama des modèles d'IA et intégration technique

- Rappels sur l'IA, le machine learning et le deep learning
- Modèles statistiques, ML et modèles préentraînés
- Cas d'usage métiers et data, critères de choix d'un modèle existant
- Bibliothèques Python : transformers, PyTorch, Hugging Face
- Appels à des API de services d'IA générative (OpenAI, etc.)
- Intégration des modèles dans des pipelines data

JOUR 2 Qualité, évaluation et exploitation responsable

- Gestion des performances et des coûts
- Mesure de la performance des modèles
- Analyse des biais et des limites, validation des résultats
- Bonnes pratiques de tests et de contrôle
- Exploitation des résultats dans des outils métiers
- Documentation, traçabilité, surveillance et maintenance des modèles

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-exploiter-modeles-ia/>

IA & MACHINE LEARNING · IAM-006

Formation machine learning et deep learning - les fondamentaux

Machine learning et deep learning : fondamentaux, algorithmes clés, évaluation et bonnes pratiques sur cas concrets.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · scikit-learn · Keras · TensorFlow · pandas · Jupyter · Anaconda

PRÉSENTATION

La formation machine learning et deep learning vous donne en 2 jours une vision complète des fondamentaux de l'apprentissage automatique. Vous comprenez les principes du machine learning, du deep learning et de l'intelligence artificielle, et leurs différences. Vous découvrez les principales méthodes du machine learning (régression, classification, clustering, ensembles) et du deep learning (réseaux denses, CNN, RNN, transformeurs). Vous comprenez les limites des algorithmes : biais, surapprentissage, exigences en données, interprétabilité. La formation vous apprend surtout à traduire un problème métier en algorithme pour le résoudre concrètement. Idéale pour data analysts et profils techniques souhaitant prendre une vraie hauteur sur le machine learning avant d'attaquer la pratique.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes du machine learning, du deep learning et de l'intelligence artificielle
- Connaître les méthodes du machine learning et du deep learning
- Comprendre les limites des algorithmes
- Traduire un problème métier en algorithme pour le résoudre

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst souhaitant aborder le machine learning
- Data scientist en formation initiale
- Profil technique souhaitant comprendre les principes du ML et du DL

Prérequis

- Bases de la statistique descriptive et de l'algèbre
- Familiarité avec un outil de traitement de données (Python, R, SAS ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du machine learning

- Principes du machine learning, du deep learning et de l'IA
- Apprentissage supervisé et non supervisé
- Traduire un problème métier en problème d'apprentissage
- Préparation des données et feature engineering
- Évaluation des modèles et gestion du sur-apprentissage
- Principaux algorithmes : régression, classification, clustering

JOUR 2 Méthodes avancées et deep learning

- Méthodes ensemblistes : forêts aléatoires, boosting
- Introduction au deep learning : réseaux de neurones, CNN, RNN
- Connaître les méthodes adaptées à chaque type de problème
- Comprendre les limites des algorithmes
- Bonnes pratiques de mise en œuvre
- Atelier : résolution d'un problème métier par le machine learning

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-machine-learning/>

IA & MACHINE LEARNING · IAM-007

Formation Machine Learning distribué avec Spark ML

Développez et déployez des modèles de machine learning distribués à grande échelle avec Apache Spark ML.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Apache Spark · Spark MLlib · Python · Scala

PRÉSENTATION

La formation machine learning distribué avec spark ml vous permet, en 3 jour(s) (21 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : avancé.

OBJECTIFS PÉDAGOGIQUES

- Développer des modèles ML distribués
- Optimiser les performances de calcul
- Préparer et transformer des données pour le ML
- Évaluer et déployer les modèles

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data engineer
- Data scientist confronté à de grands volumes de données

Prérequis

- Connaissances en Python ou Scala
- Notions de Spark et de machine learning

PROGRAMME DÉTAILLÉ

JOUR 1 Spark ML : architecture et préparation des données

- Architecture distribuée de Spark et de MLlib
- Pipelines de préparation et transformation de données à grande échelle
- Manipulation des DataFrames Spark pour le ML

JOUR 2 Modèles distribués et optimisation

- Algorithmes de ML distribués : classification, régression, clustering
- Optimisation des performances de calcul et gestion des ressources cluster
- Validation croisée et tuning des hyperparamètres

JOUR 3 Évaluation et industrialisation

- Évaluation et comparaison des modèles
- Déploiement et industrialisation des modèles distribués
- Atelier de synthèse sur cas concret

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-machine-learning-distribue-spark-ml/>

IA & MACHINE LEARNING · IAM-008

Formation Introduction au Machine Learning avec Scikit-learn

Créez et évaluez vos premiers modèles de machine learning avec Scikit-learn, la bibliothèque de référence.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Python · scikit-learn · pandas · NumPy

PRÉSENTATION

La formation introduction au machine learning avec scikit-learn vous permet, en 2 jour(s) (14 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : intermédiaire.

OBJECTIFS PÉDAGOGIQUES

- Construire des modèles supervisés et non supervisés
- Évaluer et comparer les modèles
- Préparer les données pour le ML

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Débutant en machine learning
- Data scientist junior
- Analyste souhaitant s'initier au ML

Prérequis

- Connaissances en Python
- Notions de statistiques de base

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du ML et préparation des données

- Panorama du machine learning et de l'écosystème Scikit-learn
- Préparation des données : nettoyage, encodage, normalisation
- Premiers modèles supervisés : classification et régression

JOUR 2 Modèles, évaluation et bonnes pratiques

- Apprentissage non supervisé : clustering et réduction de dimension
- Évaluation, métriques et comparaison de modèles
- Bonnes pratiques et premiers pas vers la mise en production

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-introduction-machine-learning-scikit-learn/>

IA & MACHINE LEARNING · IAM-009

Formation Deep Learning avec PyTorch

Concevez et entraînez des réseaux de neurones profonds avec PyTorch, du tenseur au déploiement.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : PyTorch · Python · CUDA · Deep Learning

PRÉSENTATION

La formation deep learning avec pytorch vous permet, en 3 jour(s) (21 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : avancé.

OBJECTIFS PÉDAGOGIQUES

- Construire des réseaux de neurones
- Entraîner et évaluer des modèles
- Optimiser et déboguer les architectures
- Déployer des modèles de deep learning

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist
- AI engineer

Prérequis

- Connaissances en Python et machine learning
- Notions de mathématiques (algèbre linéaire, dérivées)

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux et construction de réseaux

- Fondamentaux du deep learning et tenseurs PyTorch
- Construction de réseaux de neurones (autograd, modules)
- Premiers entraînements

JOUR 2 Entraînement et architectures avancées

- Fonctions de perte, optimiseurs, régularisation
- Architectures avancées (CNN, RNN) et transfert d'apprentissage
- Suivi et évaluation des modèles

JOUR 3 Optimisation et déploiement

- Optimisation, débogage et accélération GPU
- Déploiement et sauvegarde des modèles
- Atelier de synthèse

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-deep-learning-pytorch/>

IA & MACHINE LEARNING · IAM-010

Formation Deep Learning avec TensorFlow

Développez, optimisez et déployez des modèles de deep learning avec TensorFlow et Keras.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : TensorFlow · Keras · Python · Deep Learning

PRÉSENTATION

La formation deep learning avec tensorflow vous permet, en 3 jour(s) (21 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : avancé.

OBJECTIFS PÉDAGOGIQUES

- Créer des réseaux de neurones
- Entraîner et évaluer des modèles
- Utiliser les APIs TensorFlow avancées
- Déployer et monitorer des modèles de deep learning

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist
- AI engineer

Prérequis

- Connaissances en Python et machine learning
- Notions de mathématiques (algèbre linéaire, dérivées)

PROGRAMME DÉTAILLÉ

JOUR 1 Écosystème TensorFlow et Keras

- Écosystème TensorFlow et Keras
- Construction de modèles avec l'API Keras
- Premiers entraînements et suivi des métriques

JOUR 2 Entraînement et APIs avancées

- Callbacks, régularisation et suivi des métriques
- APIs avancées et personnalisation (couches, boucles d'entraînement)
- Évaluation des modèles

JOUR 3 Optimisation et déploiement

- Optimisation des performances et accélération matérielle
- Déploiement et monitoring (TensorFlow Serving)
- Atelier de synthèse

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-deep-learning-tensorflow/>

IA & MACHINE LEARNING · IAM-011

Formation Développer un assistant conversationnel avec Rasa et LlamaIndex

Concevez, entraînez et déployez un assistant conversationnel intelligent avec Rasa et LlamaIndex.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Rasa · LlamaIndex · Python · NLP · RAG

PRÉSENTATION

La formation développer un assistant conversationnel avec rasa et llamaindex vous permet, en 3 jour(s) (21 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : avancé.

OBJECTIFS PÉDAGOGIQUES

- Concevoir un chatbot intelligent
- Gérer le traitement du langage naturel
- Intégrer et déployer le chatbot
- Optimiser la compréhension et les réponses

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Développeur AI
- NLP engineer

Prérequis

- Connaissances en Python
- Notions de machine learning

PROGRAMME DÉTAILLÉ

JOUR 1 Architecture et NLU

- Architecture d'un assistant conversationnel (NLU, dialogue)
- Mise en œuvre de Rasa : intents, entités, stories
- Premiers dialogues

JOUR 2 RAG et intégration

- Intégration de LlamaIndex et approche RAG
- Connexion aux sources de données et aux canaux
- Gestion du contexte conversationnel

JOUR 3 Optimisation et déploiement

- Optimisation de la compréhension et des réponses
- Déploiement et supervision de l'assistant
- Atelier de synthèse

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-assistant-conversationnel-rasa-llamaindex/>

IA & MACHINE LEARNING · IAM-012

Formation Applications d'IA générative avec LangChain

Développez des applications d'IA générative robustes avec LangChain : chaînes, agents et RAG.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : LangChain · LLM · Python · RAG · IA générative

PRÉSENTATION

La formation applications d'ia générative avec langchain vous permet, en 3 jour(s) (21 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : avancé.

OBJECTIFS PÉDAGOGIQUES

- Créer des applications basées sur l'IA générative
- Intégrer des modèles de langage
- Automatiser des tâches génératives
- Déployer des solutions d'IA

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Développeur AI
- Data engineer

Prérequis

- Connaissances en Python
- Notions sur les LLM

PROGRAMME DÉTAILLÉ

JOUR 1 LLM et fondamentaux LangChain

- Panorama des LLM et du framework LangChain
- Chaînes, prompts et gestion du contexte
- Premières applications

JOUR 2 Agents et RAG

- Intégration de modèles de langage et d'agents
- Approche RAG et connexion aux bases de connaissances
- Automatisation de tâches génératives

JOUR 3 Déploiement et bonnes pratiques

- Déploiement des applications d'IA générative
- Bonnes pratiques : coûts, sécurité, évaluation
- Atelier de synthèse

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-ia-generative-langchain/>

IA & MACHINE LEARNING · IAM-013

Formation Gestion du cycle de vie des modèles avec MLflow

Industrialisez vos modèles de machine learning avec MLflow : tracking, registry et déploiement.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : MLflow · Python · MLOps · CI/CD

PRÉSENTATION

La formation gestion du cycle de vie des modèles avec mlflow vous permet, en 2 jour(s) (14 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : avancé.

OBJECTIFS PÉDAGOGIQUES

- Suivre et versionner les modèles ML
- Déployer et monitorer les modèles
- Automatiser les expériences ML

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Data scientist
- ML engineer

Prérequis

- Connaissances en machine learning et Python

PROGRAMME DÉTAILLÉ

JOUR 1 MLOps et tracking

- Enjeux du MLOps et présentation de MLflow
- Tracking : suivi des expériences et des métriques
- Packaging et reproductibilité (MLflow Projects, Models)

JOUR 2 Registry, déploiement et automatisation

- Model Registry : versionnement et gouvernance des modèles
- Déploiement et monitoring des modèles en production
- Automatisation et intégration CI/CD

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-mlops-mlflow/>

IA & MACHINE LEARNING · IAM-014

Formation IA en entreprise : ML et IA générative

Comprenez les usages, opportunités et enjeux du machine learning et de l'IA générative en entreprise, sans prérequis technique.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Machine Learning · IA générative · LLM

PRÉSENTATION

La formation ia en entreprise : ml et ia générative vous permet, en 1 jour(s) (7 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : débutant.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les usages du ML et de l'IA générative
- Identifier les opportunités et enjeux pour l'entreprise

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Manager
- Décideur
- Business analyst

Prérequis

- Aucune notion particulière

PROGRAMME DÉTAILLÉ

JOUR 1 Comprendre et exploiter l'IA en entreprise

- Comprendre l'IA : ML, deep learning, IA générative (sans prérequis technique)
- Panorama des usages concrets en entreprise par fonction
- Opportunités, création de valeur et cas d'usage sectoriels
- Enjeux : données, éthique, RGPD, conduite du changement
- Comment lancer et cadrer un projet IA
- Tendances et perspectives

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-ia-en-entreprise-ml-ia-generative/>

DOMAINE

Data Engineer & Big Data

Pipelines, traitement distribué et écosystème big data pour passer à l'échelle.

10 formations

DATA ENGINEER & BIG DATA · DEB-001

Formation data science - bonnes pratiques et outils

Bonnes pratiques et outils de la data science : structuration de projet, méthodologie et outils incontournables.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Python · Git · Docker · MLflow · Jupyter

PRÉSENTATION

La formation data science - bonnes pratiques et outils vous donne en 1 jour une vision transverse des méthodes et outils du métier. Vous apprenez à éviter les pièges courants des projets data science : qualité des données, sur-ajustement, mauvais cadrage métier, biais, fuite de données. Vous comprenez les bonnes pratiques du domaine : structuration de projet (CRISP-DM), gestion de versions, reproductibilité, documentation, tests, collaboration en équipe. Vous connaissez l'offre d'outils pour la data science : environnements de développement, plateformes intégrées, outils de MLOps, services cloud. Idéale pour data scientists et data analysts souhaitant gagner en maturité et fiabiliser leurs livrables, ainsi que pour managers cherchant à cadrer la pratique de leurs équipes data.

OBJECTIFS PÉDAGOGIQUES

- Éviter les pièges courants des projets data science
- Comprendre les bonnes pratiques du métier
- Connaître l'offre d'outils pour la data science

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Data scientist et data analyst en activité
- Chef d'équipe data souhaitant cadrer les pratiques
- Architecte ou responsable data

Prérequis

- Expérience pratique en data science ou analyse de données
- Pratique d'un langage data (Python, R ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Bonnes pratiques et outils de la data science

- Éviter les pièges courants des projets data science
- Méthodologie de projet (CRISP-DM)
- Bonnes pratiques : reproductibilité, gestion de versions, documentation
- Organisation du code et collaboration en équipe
- Panorama de l'offre d'outils (environnements, plateformes, MLOps, cloud)
- Critères de choix selon le contexte

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-data-science-bonnes-pratiques-et-outils/>

DATA ENGINEER & BIG DATA · DEB-002

Formation big data et data science - les fondamentaux

Big Data et data science - les fondamentaux : panorama des technos, architectures et cas d'usage incontournables.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Hadoop · Apache Spark · HDFS · Kafka · cloud

PRÉSENTATION

La formation big data et data science - les fondamentaux vous offre en 2 jours un panorama complet de l'écosystème Big Data et de la data science. Vous comprenez les principes du big data : volumes, variétés, vitesses, valeur, véracité. Vous maîtrisez le vocabulaire lié au big data et à la data science : lake, warehouse, lakehouse, batch, streaming, ETL, ELT, ML, IA, MLOps. Vous comprenez les enjeux pour mettre en œuvre une stratégie big data dans votre entreprise : cas d'usage, gouvernance, architectures, organisation, ROI. Vous connaissez les outils et leurs usages (Hadoop, Spark, Kafka, services cloud) et savez choisir ceux adaptés à votre contexte. Idéale pour décideurs, chefs de projet et profils techniques souhaitant cadrer leurs initiatives data.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes du big data
- Maîtriser le vocabulaire lié au big data et à la data science
- Identifier les enjeux pour mettre en œuvre une stratégie big data dans son entreprise
- Connaître les outils et leurs usages

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Décideur, manager et chef de projet
- Profil technique abordant le Big Data
- Responsable métier exploitant des données

Prérequis

- Aucun prérequis technique
- Culture générale data ou SI appréciée

PROGRAMME DÉTAILLÉ

JOUR 1 Comprendre le Big Data

- Le big data : définitions et usages
- Business cases de projets big data
- Les métiers de la donnée : data scientist, data analyst, data engineer, architecte
- Les données non structurées
- Vocabulaire clé du big data et de la data science

JOUR 2 Architectures, outils et stratégie

- Le passage au big data : quels changements pour l'organisation
- Le data lake et les formats de stockage (HDFS, in memory)
- Choisir sa solution de stockage et de traitement
- Panorama des outils et de leurs usages
- Enjeux d'une stratégie big data en entreprise

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-big-data-fondamentaux/>

DATA ENGINEER & BIG DATA · DEB-003

Formation big data et marketing pour décideurs

Big Data et marketing pour décideurs : leviers data, cas d'usage CRM, personnalisation et ROI des projets marketing data.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Hadoop · HDFS · MapReduce · machine learning

PRÉSENTATION

La formation big data et marketing pour décideurs explore en 2 jours les apports concrets de la data au marketing. Vous découvrez comment le big data transforme les pratiques marketing : personnalisation, attribution, scoring, recommandation, prédiction du churn. Vous identifiez les principaux cas d'usage data appliqués au marketing : segmentation comportementale, CRM analytique, attribution multi-touch, optimisation média, expérience client unifiée. Vous découvrez les architectures de référence (CDP, datalake marketing, activation cross-canal) et les enjeux RGPD associés. Vous comprenez les facteurs clés de succès des projets data marketing : alignement métier/data, qualité des données, mesure de la valeur, conduite du changement. Idéale pour directions marketing, CDO et chefs de projet.

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Direction marketing et CMO
- Chef de projet marketing et CRM
- Chief Data Officer en relation avec le marketing

Prérequis

- Aucun prérequis technique
- Connaissance des problématiques marketing

PROGRAMME DÉTAILLÉ

JOUR 1 L'économie du Big Data et les données

- La data pour réinventer son business model (étude de cas Lapeyre)
- Big Data, Small Data, Smart Data : les différences et les 5V
- Vocabulaire et notions de base (data science, data lake, machine learning)
- Les sources de données
- Les compétences du Big Data et le rôle du data scientist
- Constituer une équipe pour un projet Big Data

JOUR 2 Traitement, analyse et technologies

- Traitement des données non structurées (messages, documents, semi-structuré)
- Analyse sémantique : sentiment, univers, corrélation
- Analyse d'images pour le machine learning
- Analyse prédictive : catégories, tendances, regroupements
- Technologies du Big Data : HDFS, Hadoop, MapReduce
- Panorama des outils pour traiter de grandes masses de données

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-big-data-et-marketing/>

DATA ENGINEER & BIG DATA · DEB-004

Formation Data Lake - Panorama et choix des briques d'exploitation d'un data lake

Data Lake : panorama des architectures, choix technologiques et briques essentielles pour exploiter un lac de données.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : HDFS · Amazon S3 · Azure Data Lake Storage · Google Cloud Storage · Apache Spark · Apache Flink · Hadoop

PRÉSENTATION

La formation Data Lake vous donne en 2 jours une vision complète des architectures et choix technologiques pour exploiter un lac de données. Vous réalisez un panorama des différents outils et thématiques de data engineering : ETL et ELT, datalake et datawarehouse, bases SQL et NoSQL, moteurs SQL distribués, catalogues de données. Vous faites une revue des différentes méthodologies de choix d'une technologie selon le contexte : volumétrie, latence, coût, compétences disponibles, écosystème existant. Vous savez proposer et décrire une architecture répondant à un cas d'usage : ingestion, stockage, transformation, consommation, gouvernance. La formation s'appuie sur des cas réels d'entreprise. Idéale pour architectes data, chefs de projet et data engineers cadrant leur plateforme data.

OBJECTIFS PÉDAGOGIQUES

- Réaliser un panorama des différents outils et thématiques de data engineering (ETL, datalake, SQL, NoSQL)
- Faire une revue des méthodologies de choix d'une technologie
- Proposer et décrire une architecture répondant à un cas d'usage

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Architecte data et data engineer
- Chef de projet data cadrant une plateforme
- Responsable SI ou décideur technique

Prérequis

- Culture générale data ou SI
- Notions de base sur les bases de données

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du data engineering et stockage

- Le data engineering et sa fonction dans l'entreprise
- Qu'est-ce qu'un data lake et quelles sont ses briques
- Manières de stocker la donnée : SQL vs NoSQL
- Technologies de stockage : HDFS, GCS, S3, Azure Data Lake Storage
- Orchestration et ingestion avec un ETL
- Data processing avec Spark, Flink et Hadoop

JOUR 2 Gouvernance, sécurité et architecture

- Data lineage et data management
- Gestion des accès et de la sécurité
- Analyser sa donnée avec des outils de data analytics
- Proposer et documenter une architecture (description visuelle, RFC)
- Monitorer et gérer les performances du data lake
- Mise en place de principes de data gouvernance

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-data-lake/>

DATA ENGINEER & BIG DATA · DEB-005

Formation traitement de données distribuées avec Spark en Scala

Apache Spark avec Scala : traitement distribué, DataFrame, SQL et machine learning à l'échelle Big Data en Scala.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Apache Spark · Scala · Spark SQL · RDD · DataFrames

PRÉSENTATION

La formation traitement de données distribuées avec Spark en Scala vous apprend en 3 jours à exploiter Spark dans son langage natif. Vous comprenez l'architecture et le fonctionnement interne de Spark : driver, executors, partitions, DAG, shuffle, gestion mémoire. Vous développez des traitements distribués en Scala en utilisant l'écosystème Spark de manière idiomatique. Vous utilisez efficacement RDD, DataFrames et Spark SQL selon les besoins de performance et de typage. Vous optimisez les performances et la consommation de ressources : choix de partitionnement, gestion du cache, broadcast, plans d'exécution. Vous analysez et corrigez des problématiques en production : skew, OOM, lenteur, debugging avec Spark UI. Idéale pour data engineers et développeurs Scala traitant de grands volumes.

OBJECTIFS PÉDAGOGIQUES

- Comprendre l'architecture et le fonctionnement interne de Spark
- Développer des traitements distribués en Scala
- Utiliser efficacement RDD, DataFrames et Spark SQL
- Optimiser les performances et la consommation de ressources
- Analyser et corriger des problématiques en production

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data engineer travaillant sur Spark
- Développeur Scala intervenant sur des projets data
- Architecte Big Data

Prérequis

- Bonne maîtrise de Scala
- Notions de base en SQL et en systèmes distribués

PROGRAMME DÉTAILLÉ

JOUR 1 Architecture de Spark et programmation Scala

- Architecture et fonctionnement interne de Spark
- Driver, executors, partitions, DAG, shuffle
- Rappels de Scala pour Spark
- Développement de traitements distribués en Scala
- Premiers jobs Spark

JOUR 2 RDD, DataFrames et Spark SQL

- Utilisation des RDD
- DataFrames et Datasets
- Spark SQL et requêtes structurées
- Lecture et écriture de données
- Travaux pratiques

JOUR 3 Optimisation et production

- Optimisation des performances et de la consommation de ressources
- Partitionnement, cache et broadcast
- Analyse des plans d'exécution avec Spark UI
- Diagnostic et correction des problèmes en production (skew, OOM)
- Atelier de synthèse sur cluster

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-spark-avec-scala/>

DATA ENGINEER & BIG DATA · DEB-006

Formation DBT (Data Build Tool)

DBT (Data Build Tool) : transformations SQL versionnées, tests, documentation et bonnes pratiques de modern data stack.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : dbt · dbt Cloud · SQL · Snowflake · BigQuery · PostgreSQL · Git

PRÉSENTATION

La formation DBT (Data Build Tool) vous initie en 2 jours à l'outil de transformation SQL au cœur de la modern data stack. Vous comprenez l'architecture et les concepts fondamentaux de dbt : modèles, sources, tests, snapshots, macros, packages. Vous savez structurer et exécuter un projet dbt de bout en bout, depuis la configuration jusqu'au déploiement sur un data warehouse cloud (Snowflake, BigQuery, Redshift, Databricks). Vous créez des modèles SQL dynamiques et maintenables grâce au templating Jinja et aux macros. Vous mettez en place des tests automatisés, de la documentation et des snapshots pour garantir la qualité de votre couche de transformation. Vous intégrez dbt dans un pipeline de données avec de bonnes pratiques : organisation des modèles, gestion d'environnements, CI/CD. Idéale pour analytics engineers et data engineers.

OBJECTIFS PÉDAGOGIQUES

- Comprendre l'architecture et les concepts fondamentaux de dbt
- Structurer et exécuter un projet dbt de bout en bout
- Créer des modèles SQL dynamiques et maintenables
- Mettre en place des tests, de la documentation et des snapshots
- Intégrer dbt dans un pipeline de données avec de bonnes pratiques

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et analytics engineer
- Data engineer travaillant sur la couche de transformation
- Responsable data en modernisation de stack

Prérequis

- Bonne maîtrise de SQL
- Accès à un data warehouse cloud (Snowflake, BigQuery, Redshift, Databricks)
- Notions de Git appréciables

PROGRAMME DÉTAILLÉ

JOUR 1 De la découverte à la modélisation maîtrisée

- Fondamentaux de dbt et mise en route
- Architecture et concepts : modèles, sources, références
- Structuration d'un projet dbt
- Création de modèles SQL et organisation par couches
- Connexion à un data warehouse cloud (Snowflake, BigQuery, Redshift)

JOUR 2 Qualité, documentation et industrialisation

- Modèles dynamiques avec Jinja et macros
- Tests automatisés et qualité des données
- Documentation et snapshots
- Intégration dans un pipeline (orchestration, CI/CD)
- Bonnes pratiques et gestion d'environnements

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-dbt/>

DATA ENGINEER & BIG DATA · DEB-007

Formation Elasticsearch et Kibana pour l'analyse de données

ElasticSearch et Kibana : indexation, recherche full-text, agrégations et tableaux de bord analytiques en temps réel.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : ElasticSearch · Kibana · Elastic Stack · JSON

PRÉSENTATION

La formation Elasticsearch et Kibana vous donne en 2 jours les clés pour exploiter la stack Elastic à des fins analytiques. Vous comprenez le rôle d'ElasticSearch et de Kibana dans un écosystème data : moteur de recherche distribué, indexation full-text, analytique en temps réel. Vous savez structurer et indexer des données exploitables : mappings, types de champs, analyseurs, gestion des index. Vous formulez des requêtes pertinentes pour l'analyse via le DSL Query d'Elasticsearch (term, match, bool, agrégations). Vous construisez des visualisations et tableaux de bord interactifs avec Kibana : Discover, Visualize, Dashboards, Canvas. Vous identifiez les cas d'usage data et SI adaptés : analyse de logs, recherche métier, observabilité, dashboarding temps réel. Idéale pour data analysts et ingénieurs SI.

OBJECTIFS PÉDAGOGIQUES

- Comprendre le rôle d'ElasticSearch et de Kibana dans un écosystème data
- Structurer et indexer des données exploitables
- Formuler des requêtes pertinentes pour l'analyse
- Construire des visualisations et tableaux de bord
- Identifier des cas d'usage data et SI adaptés

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et ingénieur exploitant des logs
- Développeur intégrant un moteur de recherche
- Profil SI travaillant sur l'observabilité

Prérequis

- Notions de base sur le format JSON et les API REST
- Pratique d'un langage de programmation appréciée

PROGRAMME DÉTAILLÉ

JOUR 1 Elasticsearch : indexation et recherche

- Rôle d'ElasticSearch et de Kibana dans un écosystème data
- Architecture distribuée et concepts clés
- Indexation : mappings, types de champs, analyseurs
- Structurer des données exploitables
- Requêtes avec le DSL Query (term, match, bool)

JOUR 2 Agrégations, visualisations et cas d'usage

- Agrégations pour l'analyse
- Construction de visualisations avec Kibana
- Tableaux de bord interactifs (Discover, Dashboards, Canvas)
- Cas d'usage : analyse de logs, observabilité, recherche métier
- Bonnes pratiques d'exploitation

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-elasticsearch-et-kibana/>

DATA ENGINEER & BIG DATA · DEB-008

Formation NoSQL pour les projets data (MongoDB)

NoSQL avec MongoDB : modélisation orientée document, requêtes, agrégations et intégration dans vos projets data.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : MongoDB · NoSQL · MongoDB Compass · MongoDB Atlas

PRÉSENTATION

La formation NoSQL pour les projets data (MongoDB) vous apprend en 2 jours à exploiter la base orientée document leader du marché. Vous comprenez les principes fondamentaux des bases NoSQL et leur positionnement face aux bases relationnelles : flexibilité du schéma, scalabilité horizontale, modèles de données. Vous identifiez les cas d'usage pertinents pour MongoDB : catalogues, contenus, IoT, profils utilisateurs, agrégation de logs. Vous apprenez à modéliser efficacement des données orientées documents : embedding vs referencing, design patterns NoSQL, anti-patterns à éviter. Vous interrogez et agrégez des données avec MongoDB (find, framework d'agrégation, index). Vous intégrez MongoDB dans un écosystème data existant : connecteurs Spark, exports, synchronisation, BI. Idéale pour développeurs, data analysts et data engineers.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes fondamentaux des bases NoSQL
- Identifier les cas d'usage pertinents pour MongoDB
- Modéliser efficacement des données orientées documents
- Interroger et agréger des données avec MongoDB
- Intégrer MongoDB dans un écosystème data existant

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Développeur et data engineer
- Architecte data évaluant des solutions NoSQL
- Data analyst confronté à des données semi-structurées

Prérequis

- Bases en bases de données relationnelles ou SQL
- Pratique d'un langage de programmation appréciée

PROGRAMME DÉTAILLÉ

JOUR 1 Comprendre le NoSQL et modéliser avec MongoDB

- Panorama des bases NoSQL et principes fondamentaux
- Limites des bases relationnelles dans les projets data
- Positionnement de MongoDB dans une architecture data
- Cas d'usage concrets en environnement SI et data
- Modélisation orientée documents : documents imbriqués et références
- Travaux pratiques de modélisation

JOUR 2 Requêtes, agrégations et intégration

- Langage de requêtes MongoDB
- Indexation et performance
- Framework d'agrégation
- Analyse et restitution de données
- MongoDB dans un pipeline data et interopérabilité avec les ETL
- Bonnes pratiques d'exploitation et de maintenance

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-nosql-mongodb/>

DATA ENGINEER & BIG DATA · DEB-009

Formation Git - gestion de versions

Git pour profils data : versioning, branches, fusion, GitHub/GitLab et workflow collaboratif pour le code analytique.

DURÉE 1 jour (7 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU à distance
------------------------------	------------------------------	---	---------------------------

Technologies : Git · GitHub · GitLab · Git Bash · VS Code

PRÉSENTATION

La formation Git - gestion de versions est pensée pour les profils data (analystes, data scientists, data engineers) souhaitant adopter les bonnes pratiques de versioning. Sur 1 jour, vous mettez en place un projet versionné avec Git et configurez votre environnement. Vous connaissez les différents concepts clés : commit, push, pull, branch, merge, rebase, tag, stash, HEAD, index. Vous appliquez les bonnes pratiques d'utilisation de Git : messages de commit clairs, granularité, gestion de fichiers binaires, .gitignore adapté aux projets data. Vous maîtrisez les workflows collaboratifs (feature branch, GitFlow, GitHub flow) et la collaboration via GitHub, GitLab ou Bitbucket : pull requests, revues de code, gestion de conflits. Indispensable pour fiabiliser et industrialiser vos projets data.

OBJECTIFS PÉDAGOGIQUES

- Mettre en place un projet versionné avec Git
- Connaître les différents concepts (commit, push, etc.)
- Appliquer les bonnes pratiques d'utilisation de Git
- Maîtriser les workflows

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist et data analyst codant en Python ou R
- Data engineer travaillant en équipe
- Tout profil data souhaitant collaborer sur du code

Prérequis

- Pratique régulière d'un langage de programmation (Python, R, SQL ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Git : versionner et collaborer

- Principes de Git et son aspect décentralisé
- Les différentes interfaces et le lien avec GitHub / GitLab
- Prise en main : installation, premier dépôt, gestion des fichiers
- La notion de commit et les bonnes pratiques
- Les branches : création, gestion, résolution de conflits
- Dépôt distant (push / pull / clone) et workflows Git

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-git/>

DATA ENGINEER & BIG DATA · DEB-010

Formation DevOps pour la Data Science

DevOps pour la data science : conteneurs, CI/CD, infra as code et industrialisation des projets data et ML.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Git · GitHub · GitLab · Docker · Apache Airflow · CI/CD · cloud

PRÉSENTATION

La formation DevOps pour la data science vous apprend en 2 jours à appliquer les principes DevOps à vos projets data et machine learning. Vous comprenez les principes fondamentaux du DevOps appliqués aux projets data : automatisation, reproductibilité, collaboration, observabilité. Vous maîtrisez Git et GitHub/GitLab pour le versioning et la collaboration. Vous mettez en place des workflows d'intégration et déploiement continu (CI/CD) pour les scripts analytiques et modèles ML. Vous automatisez le traitement et l'orchestration des pipelines de données avec Apache Airflow. Vous déployez des modèles et analyses dans un environnement cloud (conteneurs Docker, Kubernetes, services managés). Vous appliquez les bonnes pratiques de surveillance et de monitoring des analyses et modèles en production. Idéale pour data scientists et data engineers visant la production.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes fondamentaux du DevOps appliqués aux projets data
- Maîtriser Git et GitHub/GitLab pour le versioning et la collaboration
- Mettre en place des workflows d'intégration et déploiement continu (CI/CD) pour les scripts analytiques et modèles ML
- Automatiser le traitement et l'orchestration des pipelines de données avec Apache Airflow
- Déployer des modèles et analyses dans un environnement cloud
- Appliquer les bonnes pratiques de surveillance et de monitoring des analyses et modèles en production

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist visant la mise en production
- Data engineer industrialisant des pipelines
- Ingénieur ML et ML engineer

Prérequis

- Bonne maîtrise de Python
- Bases de Git et de la ligne de commande Linux
- Notions de cloud appréciables

PROGRAMME DÉTAILLÉ

JOUR 1 Principes du DevOps, automatisation et CI/CD

- Introduction au DevOps appliqué à la data science
- Différences entre DevOps classique et MLOps / DataOps
- Git et GitHub/GitLab pour le versioning et la collaboration
- Conteneurisation avec Docker
- Mise en place de workflows d'intégration et déploiement continu (CI/CD)
- CI/CD pour les scripts analytiques et les modèles ML

JOUR 2 Orchestration, déploiement cloud et monitoring

- Orchestration des pipelines avec Apache Airflow
- Gestion des environnements et des dépendances
- Déploiement de modèles et d'analyses dans le cloud
- Bonnes pratiques de mise en production
- Surveillance et monitoring des modèles en production
- Atelier : pipeline complet de bout en bout

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-devops/>

DOMAINE

Business Intelligence

Concevoir des tableaux de bord clairs et efficaces qui parlent à vos décideurs.

11 formations

BUSINESS INTELLIGENCE · BIN-001

Formation Power BI initiation : Découvrez Microsoft Power BI Desktop

Démarrez Power BI Desktop en 2 jours : Power Query, modélisation, DAX et création de rapports interactifs.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou Lille ou à distance
--------------------------------	------------------------------	---	---

Technologies : Microsoft Power BI Desktop · Power Query · DAX · Power BI Service

PRÉSENTATION

La formation Power BI initiation vous permet d'acquérir en 2 jours les compétences essentielles pour analyser, transformer et visualiser vos données avec Microsoft Power BI Desktop. Vous apprenez à créer des tableaux de bord interactifs et des rapports dynamiques adaptés à vos besoins professionnels. Au programme : acquérir une méthodologie pour l'exploitation des données, connecter et transformer vos données avec Power Query, modéliser et créer des indicateurs avec le langage DAX, concevoir des rapports et tableaux de bord interactifs, publier vos rapports vers le service Power BI et construire des espaces de travail collaboratifs. Idéale pour data analysts, contrôleurs de gestion, chefs de projet décisionnel et utilisateurs Excel souhaitant monter en compétence sur Power BI.

OBJECTIFS PÉDAGOGIQUES

- Acquérir une méthodologie pour l'exploitation des données
- Développer des rapports avec Power BI Desktop
- Publier des rapports vers le service Power BI et construire des tableaux de bord

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Professionnel BI et data analyst
- Contrôleur de gestion
- Manager et chef de projet décisionnel
- Utilisateur d'Excel, SAS ou Tableau souhaitant découvrir Power BI

Prérequis

- Bonne connaissance d'Excel (fonctions de recherche, tableaux croisés dynamiques, graphiques)
- Aucune connaissance préalable de Power BI n'est nécessaire

PROGRAMME DÉTAILLÉ

JOUR 1 Méthodologie, préparation des données et modélisation

- Acquérir une méthodologie pour l'exploitation des données
- Découverte de l'interface de Power BI Desktop
- Connexion à des sources de données variées
- Transformation et nettoyage des données avec Power Query
- Construction d'un modèle de données (relations, schéma en étoile)
- Introduction au langage DAX : colonnes calculées et mesures

JOUR 2 Indicateurs, rapports et publication

- Fonctions DAX essentielles et contexte de filtre
- Time intelligence et création d'indicateurs métiers
- Conception de rapports et de tableaux de bord interactifs
- Choix des visuels, filtres, segments et interactions
- Publication vers le service Power BI
- Partage sécurisé et espaces de travail collaboratifs

« La pédagogie et l'écoute du formateur. Un support et des exercices bien pensés qui permettent de visualiser concrètement les notions abordées. »

— Maxime B.

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-microsoft-power-bi-desktop/>

BUSINESS INTELLIGENCE · BIN-002

Formation Power BI Dataviz et analytics

Allez plus loin avec Power BI : datavisualisation avancée, analytics, storytelling et rapports à fort impact métier.

DURÉE 1 jour (7 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Microsoft Power BI Desktop · DAX · Power Query · Power BI Service

PRÉSENTATION

La formation Power BI Dataviz et Analytics vous permet en 1 jour d'aller au-delà des bases de Power BI pour produire des rapports véritablement impactants. Vous améliorez les restitutions visuelles dans Power BI en explorant les visuels avancés, les bonnes pratiques de design et de storytelling. Vous construisez une progression dans l'approche des données : du vue d'ensemble au détail, du KPI à l'analyse multi-dimensionnelle. Vous exploitez la puissance analytique de Power BI grâce aux fonctionnalités intégrées : prévisions, analyse d'influences, regroupements, segmentation visuelle. Une formation idéale pour analystes Power BI souhaitant produire des dashboards plus efficaces, plus lisibles et plus convaincants pour leurs métiers.

OBJECTIFS PÉDAGOGIQUES

- Améliorer les restitutions visuelles dans Power BI
- Construire une progression dans l'approche des données
- Exploiter la puissance analytique de Power BI

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Data analyst Power BI souhaitant améliorer ses rapports
- Développeur BI travaillant sur Power BI
- Contrôleur de gestion et analyste métier

Prérequis

- Pratique régulière de Power BI Desktop
- Connaissance de base du langage DAX

PROGRAMME DÉTAILLÉ

JOUR 1 Data visualisation et analytique avancée avec Power BI

- Comprendre la valeur ajoutée de la data visualisation
- Évaluer l'efficacité d'un visuel (data-ink ratio, parcours de l'œil)
- Choisir le bon graphique selon le message et les familles de graphiques
- Exploiter des visuels avancés (flux, Kiviat, treemap, heat map, Marimekko)
- Cartes géographiques et data storytelling
- Filtres drill through, signets, mise en surbrillance et interactions
- Thème personnalisé JSON, affichage mobile et visuels personnalisés

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-power-bi-dataviz-et-analytics/>

BUSINESS INTELLIGENCE · BIN-003

Formation DAX expertise - PowerBI

Devenez expert DAX : contextes d'évaluation, time intelligence, mesures avancées et optimisation des modèles Power BI.

DURÉE 1 jour (7 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Microsoft Power BI Desktop · DAX · DAX Studio · Tabular Editor

PRÉSENTATION

La formation DAX expertise Power BI s'adresse aux utilisateurs avancés souhaitant maîtriser à fond le langage DAX, au cœur de Power BI et des modèles tabulaires Microsoft. Sur 1 jour, vous apprenez à travailler avec des tables calculées, à exploiter les fonctions de génération dynamique et à structurer vos modèles. Vous maîtrisez les contextes d'évaluation (ligne et filtre), CALCULATE, les fonctions de time intelligence, ainsi que les patterns DAX avancés utilisés par les experts. Vous optimisez les requêtes DAX en comprenant le moteur de stockage (VertiPaq), l'évaluation par moteur de formule, et les techniques de profilage. Indispensable pour passer un cap technique sur Power BI et pour les développeurs BI ambitieux.

OBJECTIFS PÉDAGOGIQUES

- Travailler avec des tables calculées
- Maîtriser les contextes d'évaluation (ligne et filtre)
- Optimiser les requêtes DAX

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Développeur BI confirmé sur Power BI
- Data analyst expert souhaitant approfondir DAX
- Architecte BI travaillant sur des modèles tabulaires

Prérequis

- Pratique approfondie de Power BI Desktop
- Maîtrise de base du langage DAX (CALCULATE, mesures simples, contexte de filtre)

PROGRAMME DÉTAILLÉ

JOUR 1 DAX expertise : optimisation et fonctions avancées

- Environnement développeur : DAX Studio, SQL Server Management Studio, profilage des requêtes
- Création de variables DAX (VAR... RETURN)
- Tables calculées et reverse linked table (CALENDARAUTO)
- Forcer le contexte d'évaluation avec CALCULATE et filtrage bidirectionnel
- Calendriers personnalisés (type 4/4/5)
- Fonctions itératives (SUMX, AVERAGEX) et de filtre (FILTER, EARLIER)
- Fonctions d'agrégation avancées (SUMMARIZE, SUMMARIZECOLUMNS, ADDCOLUMNS)
- Fonctions de jointure (CROSSJOIN, UNION, INTERSECT) et Row-Level Security

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-power-bi-dax-expertise/>

BUSINESS INTELLIGENCE · BIN-004

Formation découverte Power BI

Découvrez Power BI en 1 journée : panorama de l'outil, premiers rapports et bases pour cadrer un projet décisionnel.

DURÉE ½ journée (3,5 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
-----------------------------------	------------------------------	---	------------------------------------

Technologies : Microsoft Power BI Service · Power BI Desktop

PRÉSENTATION

La formation découverte Power BI vous offre en une demi-journée un tour d'horizon de la solution Microsoft Power BI Service, côté consommation des rapports. Vous apprenez à manipuler des tableaux de bord avec Microsoft Power BI Service : navigation, filtres, drill-down, mise en favoris, abonnements. Vous découvrez comment modifier des tableaux de bord (filtres, paramètres, signets) sans toucher au modèle de données, pour adapter rapidement les rapports à vos besoins. La formation aborde aussi le partage, les espaces de travail et l'organisation des contenus. Idéale pour utilisateurs métiers, décideurs et chefs de projet souhaitant exploiter pleinement les rapports Power BI mis à leur disposition, sans rentrer dans le développement avec Power BI Desktop.

OBJECTIFS PÉDAGOGIQUES

- Manipuler des tableaux de bord avec Microsoft Power BI Service
- Modifier des tableaux de bord (filtres...) avec Microsoft Power BI Service

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Utilisateur métier consommateur de rapports Power BI
- Décideur et manager
- Chef de projet cadrant un déploiement Power BI

Prérequis

- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte de Power BI Service (½ journée)

- Panorama de Power BI et de son écosystème
- Manipuler des tableaux de bord avec Power BI Service
- Navigation, filtres et drill-down
- Modifier des tableaux de bord (filtres, signets)
- Partage et organisation des contenus

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/decouverte-power-bi/>

BUSINESS INTELLIGENCE · BIN-005

Formation Tableau Desktop

Maîtrisez Tableau Desktop : connexion aux données, visualisations interactives, calculs et tableaux de bord pro.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Tableau Desktop · Tableau Cloud · Tableau Server

PRÉSENTATION

La formation Tableau Desktop vous donne en 2 jours les compétences clés pour créer des analyses visuelles puissantes. Vous apprenez à charger des sources de données multiples (Excel, bases de données, cloud, fichiers plats) pour la création de tableaux de bord interactifs. Vous construisez les indicateurs et visualisations de votre tableau de bord, en exploitant la richesse graphique de Tableau et ses fonctionnalités analytiques (calculs, paramètres, ensembles, niveau de détail). Vous utilisez le cloud Tableau (Tableau Cloud, Tableau Server) comme espace d'hébergement et de partage de vos tableaux de bord avec vos équipes. Idéale pour data analysts, marketeurs, contrôleurs de gestion et chefs de projet souhaitant maîtriser Tableau de bout en bout.

OBJECTIFS PÉDAGOGIQUES

- Charger des sources de données multiples pour la création d'un tableau de bord
- Construire les indicateurs et les visualisations du tableau de bord
- Utiliser le cloud comme espace d'hébergement et de partage du tableau de bord

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et marketeur exploitant des données
- Contrôleur de gestion et chef de projet décisionnel
- Utilisateur d'Excel souhaitant passer à un outil dédié à la dataviz

Prérequis

- Bonne maîtrise d'Excel
- Aucune connaissance préalable de Tableau n'est nécessaire

PROGRAMME DÉTAILLÉ

JOUR 1 Prise en main et visualisations avec Tableau

- Découverte de l'interface de Tableau Desktop
- Connexion à des sources de données multiples (Excel, bases, cloud)
- Dimensions, mesures et types de données
- Construction de visualisations : barres, lignes, nuages, cartes
- Tris, filtres et hiérarchies

JOUR 2 Calculs, tableaux de bord et partage

- Champs calculés et paramètres
- Fonctions analytiques : ensembles, niveaux de détail (LOD)
- Construction d'indicateurs et de KPI
- Assemblage de tableaux de bord interactifs
- Publication et partage sur Tableau Cloud / Server

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-tableau-desktop/>

BUSINESS INTELLIGENCE · BIN-006

Formation QlikView - Initiation

Initiez-vous à QlikView : script de chargement, modèle associatif, expressions et conception d'applications analytiques.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU —
--------------------------------	------------------------------	---	------------------

Technologies : QlikView · QlikView Desktop

PRÉSENTATION

La formation QlikView initiation vous donne les bases pour construire des applications analytiques avec QlikView. Vous apprenez à charger des sources de données multiples pour la création d'un tableau de bord et à construire le modèle associatif au cœur de la plateforme Qlik. Vous construisez les indicateurs et visualisations de votre tableau de bord : graphiques, tableaux, listes de sélection, expressions et set analysis. Vous découvrez comment utiliser le cloud comme espace d'hébergement et de partage de vos applications QlikView avec vos équipes. La formation est rythmée d'ateliers pratiques pour ancrer les acquis. Idéale pour analystes, consultants, contrôleurs de gestion et chargés d'étude souhaitant prendre en main QlikView.

OBJECTIFS PÉDAGOGIQUES

- Charger des sources de données multiples pour la création d'un tableau de bord
- Construire les indicateurs et les visualisations du tableau de bord
- Utiliser le cloud comme espace d'hébergement et de partage du tableau de bord

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Analyste et consultant BI
- Contrôleur de gestion
- Chargé d'étude

Prérequis

- Aucun prérequis technique
- Bonne maîtrise d'Excel appréciée

PROGRAMME DÉTAILLÉ

JOUR 1 Prise en main et chargement de données QlikView

- Présentation de QlikView et du modèle associatif
- Découverte de l'interface et des principes de la BI
- Chargement de sources de données multiples
- Édition et mise en forme de documents QlikView
- Création d'un premier document

JOUR 2 Visualisations et partage

- Construction des indicateurs et des visualisations
- Représentation graphique avec QlikView
- Introduction à la modélisation des données
- Expressions et bonnes pratiques
- Utilisation du cloud pour l'hébergement et le partage

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-qlikview/>

BUSINESS INTELLIGENCE · BIN-007

Formation Designer Qlik Sense

Devenez designer Qlik Sense : conception d'applications, visualisations, storytelling et bonnes pratiques de dashboarding.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Qlik Sense · Qlik Sense Cloud

PRÉSENTATION

La formation Designer Qlik Sense vous apprend en 1 jour à concevoir et faire évoluer des applications analytiques sur Qlik Sense. Vous apprenez à modifier des applications Qlik Sense existantes, en travaillant sur les feuilles, les visualisations et la structure de l'application. Vous réalisez des analyses avancées avec Qlik Sense, en exploitant set analysis, expressions complexes et fonctionnalités d'investigation. Vous savez utiliser les bonnes visualisations selon le message à transmettre et appliquez les principes de design propres à Qlik Sense. Enfin, vous partagez votre application modifiée avec vos collaborateurs via Qlik Sense Cloud ou Enterprise. Idéale pour utilisateurs et analystes Qlik Sense souhaitant produire des applications à fort impact métier.

OBJECTIFS PÉDAGOGIQUES

- Modifier des applications Qlik Sense existantes
- Réaliser des analyses avancées avec Qlik Sense
- Utiliser les bonnes visualisations
- Partager son application modifiée

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Utilisateur Qlik Sense débutant ou confirmé
- Data analyst produisant des applications Qlik Sense
- Designer ou chef de projet BI travaillant sur Qlik Sense

Prérequis

- Première expérience sur Qlik Sense ou bonne culture des outils de BI

PROGRAMME DÉTAILLÉ

JOUR 1 Designer Qlik Sense : conception et partage d'applications

- Prendre en main une application existante et découvrir l'interface Designer
- Modifier les éléments d'une page d'analyse : formules, graphiques, tableaux
- Utilisation des master items
- Analyses avancées : choisir les bonnes visualisations selon les problématiques
- Création d'applications responsive
- Storytelling pour présenter ses analyses et partage avec d'autres utilisateurs

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-designer-qlik-sense/>

BUSINESS INTELLIGENCE · BIN-008

Formation Développeur Qlik Sense

Côté développeur Qlik Sense : script de chargement, modélisation, expressions set analysis et optimisation des applis.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Qlik Sense · Qlik Sense Cloud · QVD · Set Analysis

PRÉSENTATION

La formation Développeur Qlik Sense couvre en 2 jours ce qu'il faut savoir pour construire des applications robustes côté développeur. Vous apprenez à effectuer des analyses avancées avec Qlik Sense en maîtrisant le langage d'expression et set analysis. Vous créez des modèles de données Qlik Sense optimisés et savez résoudre les clés synthétiques et les références circulaires, problématiques classiques du modèle associatif. Vous maîtrisez les fonctions avancées de l'éditeur de scripts de chargement de données : variables, sous-routines, chargements incrémentaux, fichiers QVD. Vous transformez les données pour les rendre prêtes à l'analyse. Idéale pour développeurs Qlik et data engineers livrant des applications industrielles et performantes.

OBJECTIFS PÉDAGOGIQUES

- Effectuer des analyses avancées avec Qlik Sense
- Créer des modèles de données Qlik Sense
- Résoudre les clés synthétiques et les références circulaires
- Maîtriser les fonctions avancées de l'éditeur de scripts de chargement
- Transformer les données

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Développeur Qlik Sense
- Data engineer construisant des modèles pour Qlik
- Consultant BI sur la plateforme Qlik

Prérequis

- Pratique de Qlik Sense côté utilisateur ou designer
- Notions de SQL et de modélisation de données

PROGRAMME DÉTAILLÉ

JOUR 1 Interface, script et modèles de données

- Découvrir l'interface de développement (Data Manager, Data Load Editor, Data Model Viewer)
- Le langage de script : structuration, contrôles, préfixes, expressions, variables
- Fonctions de base pour les scripts
- Chargement de données depuis fichiers, éditeur et autres sources
- Modèles de données : principes, modèles en étoile
- Gérer les clés synthétiques et les références circulaires

JOUR 2 Visualisations avancées et déploiement

- Types de visualisations, mesures et dimensions
- Utilisation de fonctions dans les graphiques
- Utilisation avancée du Set Analysis
- Travailler avec des master items
- Calculs avancés et bonnes pratiques
- Déploiement et sécurité

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-developpeur-qlik-sense/>

BUSINESS INTELLIGENCE · BIN-009

Formation Découverte Qlik Sense

Découvrez Qlik Sense en 1 journée : panorama des fonctionnalités, premières applications et cadrage de projet BI.

DURÉE ½ journée (3,5 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
-----------------------------------	------------------------------	---	------------------------------------

Technologies : Qlik Sense · Qlik Sense Cloud

PRÉSENTATION

La formation découverte Qlik Sense vous offre en une demi-journée une prise en main pratique de la plateforme analytique de Qlik. Vous apprenez à manipuler une application avec Qlik Sense : navigation, sélections associatives, filtres, drill-down, signets. Vous découvrez comment modifier et partager une application Qlik Sense pour l'adapter à vos besoins et la diffuser à vos collaborateurs. Vous apprenez aussi à modifier des visualisations dans une application Qlik Sense : ajout, suppression, changement de type, paramétrage des dimensions et mesures. La formation s'adresse à des utilisateurs métiers, chefs de projet ou consommateurs de rapports souhaitant gagner en autonomie sur la plateforme Qlik Sense sans se former au développement complet.

OBJECTIFS PÉDAGOGIQUES

- Manipuler une application avec Qlik Sense
- Modifier et partager une application Qlik Sense
- Modifier des visualisations dans une application Qlik Sense

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Utilisateur métier consommateur d'applications Qlik Sense
- Décideur et manager
- Chef de projet cadrant un déploiement Qlik Sense

Prérequis

- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte de Qlik Sense (½ journée)

- Panorama de Qlik Sense et du modèle associatif
- Manipuler une application : navigation, sélections, filtres
- Modifier des visualisations dans une application
- Modifier et partager une application
- Bonnes pratiques d'utilisation

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/decouverte-qlik-sense/>

BUSINESS INTELLIGENCE · BIN-010

Formation Superset pour l'analyse et la visualisation de données

Apache Superset pour la data viz open source : connexion SQL, dashboards interactifs et alternative cloud-native.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Apache Superset · SQL · PostgreSQL

PRÉSENTATION

La formation Apache Superset vous initie en 2 jours à la solution open source de data visualisation et de business intelligence. Vous comprenez le positionnement d'Apache Superset dans un SI décisionnel moderne, face aux solutions propriétaires comme Power BI ou Tableau. Vous apprenez à vous connecter à des sources de données SQL variées (PostgreSQL, MySQL, Snowflake, BigQuery, etc.) et à créer des visualisations pertinentes adaptées à votre message. Vous construisez et partagez des tableaux de bord interactifs et appliquez des bonnes pratiques de gouvernance et de performance : gestion des accès, mise en cache, contrôle de la consommation. Idéale pour équipes data et SI souhaitant déployer une solution BI open source mature.

OBJECTIFS PÉDAGOGIQUES

- Comprendre le positionnement d'Apache Superset dans un SI décisionnel
- Se connecter à des sources de données SQL
- Créer des visualisations pertinentes
- Construire et partager des tableaux de bord
- Appliquer des bonnes pratiques de gouvernance et de performance

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst travaillant en environnement open source
- Data engineer déployant une solution BI
- Responsable data dans une démarche de souveraineté

Prérequis

- Bonne maîtrise de SQL
- Notions sur les bases de données relationnelles

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte de Superset et connexion aux données

- Positionnement d'Apache Superset dans un SI décisionnel
- Installation et architecture de Superset
- Connexion à des sources de données SQL
- Exploration des données et SQL Lab
- Création de datasets

JOUR 2 Visualisations, dashboards et gouvernance

- Création de visualisations pertinentes
- Construction de tableaux de bord interactifs
- Filtres, croisements et interactions
- Partage des tableaux de bord
- Bonnes pratiques de gouvernance, sécurité et performance

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-superset/>

BUSINESS INTELLIGENCE · BIN-011

Formation Power BI Perfectionnement

Perfectionnez-vous sur Power BI : DAX avancé, modélisation, performance et techniques de pros pour des rapports d'élite.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris, Lille ou à distance
--------------------------------	------------------------------	---	---

Technologies : Microsoft Power BI Desktop · DAX · DAX Studio · Power Query · Performance Analyzer

PRÉSENTATION

La formation Power BI Perfectionnement s'adresse aux utilisateurs ayant déjà une bonne maîtrise de Power BI Desktop et souhaitant franchir un cap. Sur 2 jours, vous approfondissez les techniques avancées qui font la différence dans les projets décisionnels exigeants. Vous renforcez votre maîtrise de la modélisation et du langage DAX (CALCULATE, contextes, time intelligence, patterns avancés). Vous optimisez les performances de vos modèles tabulaires et apprenez à analyser les requêtes avec les outils dédiés (Performance Analyzer, DAX Studio). Vous explorez aussi les techniques avancées de visualisation et de mise en forme conditionnelle, pour produire des rapports vraiment impactants. Idéale pour analystes et développeurs BI confirmés.

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst Power BI confirmé
- Développeur BI souhaitant aller plus loin sur Power BI
- Architecte décisionnel travaillant sur des modèles complexes

Prérequis

- Bonne maîtrise de Power BI Desktop (Power Query, modélisation, premières mesures DAX)
- Expérience de plusieurs projets Power BI

PROGRAMME DÉTAILLÉ

JOUR 1 Modélisation avancée et DAX

- Rappels et bonnes pratiques de modélisation tabulaire
- DAX avancé : contextes d'évaluation, CALCULATE, variables
- Time intelligence et calendriers personnalisés
- Patterns DAX courants en entreprise
- Gestion des relations complexes (plusieurs à plusieurs, rôles)

JOUR 2 Performance, sécurité et visualisation avancée

- Optimisation des performances et du moteur VertiPaq
- Analyse des requêtes avec Performance Analyzer et DAX Studio
- Row-Level Security et gestion des droits
- Techniques avancées de visualisation et mise en forme conditionnelle
- Atelier : refonte d'un rapport sur les données des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-power-bi-perfectionnement/>

DOMAINE

DataViz

Représenter la donnée et rendre vos analyses claires, lisibles et engageantes.

2 formations

DATAVIZ · DVZ-001

Formation Datavisualisation - L'art de représenter les données

L'art de représenter les données : sémiologie graphique, choix de visuels, palette et storytelling visuel efficace.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Excel · Power BI · Tableau · Python · R

PRÉSENTATION

La datavisualisation est devenue un levier incontournable pour comprendre, analyser et communiquer efficacement des résultats. La formation datavisualisation vous apprend en 2 jours à créer des visuels clairs, impactants et adaptés à votre audience. Vous maîtrisez les bonnes pratiques pour communiquer vos données de façon pertinente : choix du type de graphique, principes de perception visuelle, hiérarchie de l'information, palette de couleurs. Vous apprenez à créer de l'engagement et à favoriser la compréhension de l'information grâce au data storytelling. Vous appliquez une méthodologie pour toucher un public plus large, qu'il soit expert ou non. La formation est indépendante des outils et bénéficie à tout utilisateur d'Excel, Power BI, Tableau, Python ou R. Idéale pour profils data et métiers diffusant des analyses.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les bonnes pratiques pour communiquer ses données de façon pertinente
- Créer de l'engagement et favoriser la compréhension de l'information
- Appliquer une méthodologie pour toucher un public plus large

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Toute personne ayant la nécessité de diffuser des données
- Data analyst, contrôleur de gestion, chargé d'étude
- Communicant, journaliste de données, chef de projet

Prérequis

- Aucun prérequis technique
- Pratique d'un outil de traitement de données (Excel, Power BI, Tableau ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux de la datavisualisation

- Bonnes pratiques pour communiquer ses données de façon pertinente
- Principes de perception visuelle
- Choisir le bon graphique selon les données et le message
- Pièges et mauvaises pratiques à éviter
- Hiérarchie de l'information et usage de la couleur

JOUR 2 Data storytelling et mise en pratique

- Créer de l'engagement et favoriser la compréhension
- Méthodologie pour toucher un public large
- Construction d'un récit avec les données (data storytelling)
- Manipulation d'outils de datavisualisation
- Atelier : refonte de visuels sur vos propres jeux de données

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-datavisualisation/>

DATAVIZ · DVZ-002

Formation rendre ses présentations claires et engageantes

Des présentations claires et engageantes : structure du message, slides percutants et prise de parole impactante.

DURÉE 2 jours (14 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : PowerPoint · Canva

PRÉSENTATION

La formation rendre ses présentations claires et engageantes vous apprend en 2 jours à concevoir des supports de présentation à fort impact, particulièrement pour partager des résultats data. Vous maîtrisez les fondamentaux du design graphique appliqués aux supports : typographie, couleurs, contrastes, mise en page. Vous connaissez les principes de construction d'un slide : format, grille, ligne de force, hiérarchie visuelle, équilibre. Vous savez choisir la meilleure forme graphique pour présenter vos données selon le message à transmettre et le public visé. La formation alterne théorie, analyse de cas et exercices pratiques sur vos propres supports. Idéale pour profils data, chefs de projet et managers qui présentent régulièrement leurs résultats à des audiences variées.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les fondamentaux du design graphique
- Connaître les principes de construction (format, grille, ligne de force)
- Choisir la meilleure forme graphique pour présenter ses données

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et data scientist amenés à présenter leurs résultats
- Chef de projet et manager
- Consultant et formateur souhaitant améliorer ses supports

Prérequis

- Aucun prérequis technique
- Pratique d'un outil de présentation (PowerPoint, Google Slides, Keynote)

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux du design pour la présentation

- Les fondamentaux du design graphique
- Principes de construction : format, grille, ligne de force
- Typographie, couleurs et contrastes
- Hiérarchie visuelle et équilibre
- Analyse de bons et mauvais exemples

JOUR 2 Structurer et délivrer un message percutant

- Structurer son message et son fil narratif
- Choisir la meilleure forme graphique pour présenter ses données
- Concevoir des slides épurés et mémorables
- Techniques de prise de parole
- Atelier : refonte d'une présentation sur vos propres supports

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-presentations-claires-et-engageantes/>

DOMAINE

Statistiques

Méthodes statistiques, XLSTAT, équations structurelles et analyse de données appliquée.

7 formations

STATISTIQUES · STA-001

Formation statistique et analyse de données

Maîtrisez les statistiques appliquées en 3 jours : descriptives, tests, modélisation et interprétation rigoureuse.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Excel · R · XLSTAT

PRÉSENTATION

La formation statistique et analyse de données vous donne en 3 jours les clés pour analyser vos données avec rigueur statistique. Vous comprenez les principes de base de la statistique : population, échantillon, distributions, lois de probabilité, intervalles de confiance. Vous apprenez à décrire des données en utilisant des statistiques descriptives (tendance centrale, dispersion, formes, relations) et à choisir les bons indicateurs selon la nature de vos variables. Vous comprenez les bases de la modélisation et de l'analyse de données : tests statistiques (Student, Chi², ANOVA), corrélation, régression. La pédagogie privilégie l'interprétation des résultats sur des cas concrets, plutôt que le formalisme mathématique. Idéale pour analystes, chercheurs et profils métier data.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes de base de la statistique
- Décrire des données en utilisant des statistiques descriptives
- Comprendre les bases de la modélisation et de l'analyse de données

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Analyste et chargé d'étude
- Profil métier exploitant des données quantitatives
- Data analyst débutant souhaitant consolider ses bases statistiques
- Chercheur et doctorant

Prérequis

- Aucun prérequis technique
- Aisance avec les chiffres et les pourcentages

PROGRAMME DÉTAILLÉ

JOUR 1 Statistiques descriptives

- Principes de base de la statistique : population, échantillon, variables
- Décrire des données : tendance centrale et dispersion
- Distributions et représentations graphiques
- Statistiques bivariées : tableaux croisés et corrélations
- Interprétation et restitution des résultats

JOUR 2 Statistique inférentielle et tests

- Lois de probabilité et théorème central limite
- Estimation et intervalles de confiance
- Principe des tests d'hypothèses
- Tests usuels : Student, Chi², ANOVA
- Lecture et interprétation des résultats de tests

JOUR 3 Introduction à la modélisation

- Bases de la modélisation et de l'analyse de données
- Régression linéaire simple et multiple
- Qualité d'ajustement et diagnostic
- Notions d'analyse factorielle
- Atelier de synthèse sur un jeu de données réel

« Le formateur a sur-délivré. Le meilleur formateur que j'ai rencontré de loin. Il s'exprime clairement, se met à la place de l'autre, adapte son langage en fonction de son interlocuteur, une excellente connaissance du métier et des différents métiers. Il a su complètement s'adapter à nos domaines respectifs (marketing et pour l'autre participant, les tests cliniques). Un niveau de compétences très très avancé. Meilleur formateur. »

| — G.L.

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-statistique-analyse-de-donnees/>

STATISTIQUES · STA-002

Formation XLSTAT - statistique et analyse de données

Faites des stats avancées dans Excel avec XLSTAT : ACP, ACM, régression, tests, segmentation et plans d'expériences.

DURÉE 3 jours (21 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Excel · XLSTAT

PRÉSENTATION

La formation XLSTAT vous apprend à exploiter pleinement la puissance statistique du complément XLSTAT dans Excel. Sur 3 jours, vous maîtrisez l'utilisation du logiciel XLSTAT pour analyser des données et réaliser un large éventail d'analyses sans quitter Excel. Vous comprenez quelle méthode utiliser dans quel cas : statistiques descriptives, tests d'hypothèses, ACP, ACM, AFC, régressions, segmentation, plans d'expériences, séries temporelles. Vous apprenez à interpréter les résultats d'une analyse avec XLSTAT, à lire les sorties et à les restituer à des décideurs ou à des publications. Idéale pour statisticiens, qualitatifs, chargés d'études, marketeurs et chercheurs ayant besoin d'outils statistiques avancés directement dans Excel.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser l'utilisation du logiciel XLSTAT pour analyser des données
- Comprendre quelle méthode utiliser dans quel cas
- Interpréter les résultats d'une analyse avec XLSTAT

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Statisticien et chargé d'étude
- Qualitatif et ingénieur méthode
- Marketeur et chargé d'études consommateurs
- Chercheur en sciences expérimentales ou sociales

Prérequis

- Bonne maîtrise d'Excel
- Bases en statistique descriptive

PROGRAMME DÉTAILLÉ

JOUR 1 Prise en main et statistiques descriptives avec XLSTAT

- Présentation du logiciel XLSTAT et de son intégration à Excel
- Préparation et description des données
- Statistiques descriptives et représentations graphiques
- Tests de normalité et premières analyses
- Comprendre quelle méthode utiliser selon le cas

JOUR 2 Tests statistiques et analyses factorielles

- Tests d'hypothèses : tests paramétriques et non paramétriques
- ANOVA et comparaisons de moyennes
- Analyse en composantes principales (ACP)
- Analyse des correspondances (AFC, ACM)
- Interprétation des résultats

JOUR 3 Modélisation et méthodes avancées

- Régression linéaire et logistique
- Segmentation et classification
- Introduction aux plans d'expériences
- Séries temporelles avec XLSTAT
- Atelier sur les données des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-xlstat/>

STATISTIQUES · STA-003

Formation XLSTAT-R et programmation avec XLSTAT

Automatisez vos analyses XLSTAT : programmation, scripts R via XLSTAT-R et industrialisation de vos analyses récurrentes.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Excel · XLSTAT · XLSTAT-R · R · VBA

PRÉSENTATION

La formation XLSTAT-R et programmation avec XLSTAT vous apprend en 2 jours à automatiser et étendre les capacités du logiciel. Vous apprenez à utiliser XLSTAT pour automatiser des traitements en VBA, et ainsi industrialiser vos analyses récurrentes et vos rapports. Vous découvrez comment utiliser XLSTAT-R pour appliquer des codes R directement dans Excel, en combinant la puissance statistique de R avec l'ergonomie d'XLSTAT. Vous créez vos propres traitements R pour les intégrer dans XLSTAT, et ainsi ajouter des méthodes non couvertes nativement. Une formation parfaite pour utilisateurs avancés d'XLSTAT cherchant à industrialiser leurs analyses récurrentes ou à connecter XLSTAT à des méthodes R personnalisées.

OBJECTIFS PÉDAGOGIQUES

- Utiliser XLSTAT pour automatiser des traitements en VBA
- Utiliser XLSTAT-R pour appliquer des codes R dans Excel
- Créer ses propres traitements R pour les intégrer dans XLSTAT

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Utilisateur avancé d'XLSTAT
- Statisticien souhaitant industrialiser ses analyses
- Développeur amené à étendre XLSTAT

Prérequis

- Bonne maîtrise d'XLSTAT
- Notions de programmation (VBA ou R) appréciées

PROGRAMME DÉTAILLÉ

JOUR 1 XLSTAT-R : connecter R et Excel

- Présentation d'XLSTAT-R et de l'intégration R dans Excel
- Installation et configuration de l'environnement
- Appliquer des codes R directement depuis XLSTAT
- Créer ses propres traitements R intégrés à XLSTAT
- Cas pratiques de méthodes R non couvertes nativement

JOUR 2 Automatisation et programmation VBA

- Automatiser des traitements XLSTAT en VBA
- Industrialisation des analyses récurrentes
- Génération automatique de rapports
- Bonnes pratiques de programmation et de maintenance
- Atelier : automatisation d'une chaîne d'analyse complète

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-programmation-avec-xlstat/>

STATISTIQUES · STA-004

Formation approche PLS (PLS Path Modeling)

Approche PLS et PLS Path Modeling : modèles à variables latentes pour analyser des structures causales complexes.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : XLSTAT · SmartPLS · R

PRÉSENTATION

La formation approche PLS (PLS Path Modeling) vous initie en 2 jours à la modélisation par équations structurelles selon la méthode des moindres carrés partiels. Vous comprenez les principes de l'approche PLS pour analyser des modèles à variables latentes, et ses spécificités par rapport aux SEM covariance-based. Vous savez utiliser les outils permettant d'appliquer l'approche PLS (XLSTAT-PLS, SmartPLS, packages R), construire votre modèle de mesure et structurel, et évaluer la qualité de l'ajustement. À l'issue, vous êtes capable d'interpréter les résultats d'une analyse par l'approche PLS : coefficients, R^2 , t-statistiques, comparaisons de groupes. Idéale pour chercheurs, marketeurs et analystes étudiant des relations causales complexes.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes de l'approche PLS pour analyser des modèles à variables latentes
- Utiliser les outils permettant d'appliquer l'approche PLS
- Interpréter les résultats d'une analyse par l'approche PLS

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Chercheur et doctorant en sciences sociales ou marketing
- Chargé d'étude et statisticien
- Data analyst travaillant sur des modèles à variables latentes

Prérequis

- Bases solides en statistique (régression, ACP)
- Pratique d'un outil statistique (R, XLSTAT, SmartPLS ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Principes et concepts des modèles à variables latentes

- L'approche PLS (PLS Path Modeling) : principes
- Algorithme, validation et interprétation
- Introduction aux outils pour l'approche PLS
- Cas pratique : application à l'analyse de la satisfaction
- Applications aux sciences sociales et autres domaines

JOUR 2 Utilisation avancée de l'approche PLS

- Effets médiateurs
- Effets modérateurs
- Comparaisons multi-groupes
- Segmentation
- Discussions et applications sur les données des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-approche-pls/>

STATISTIQUES · STA-005

Formation équations structurelles à variables latentes

Équations structurelles à variables latentes (SEM) : modélisation causale, validation et interprétation pour la recherche.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : SPSS AMOS · R · lavaan

PRÉSENTATION

La formation équations structurelles à variables latentes (SEM) vous apprend en 2 jours à modéliser des relations causales complexes. Vous comprenez les principes théoriques des modèles à équations structurelles, leur logique et leurs hypothèses. Vous savez vous servir des outils permettant d'analyser un modèle à équations structurelles (LISREL, AMOS, lavaan en R), depuis la spécification du modèle jusqu'à son estimation. Vous apprenez à interpréter les résultats d'un modèle à variables latentes : indices d'ajustement, coefficients standardisés, modifications du modèle, comparaison de groupes. La formation s'appuie sur des cas concrets issus du marketing, de la psychologie et des sciences sociales. Idéale pour chercheurs, doctorants et analystes utilisant des questionnaires.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes théoriques des modèles à équations structurelles
- Utiliser les outils permettant d'analyser un modèle à équations structurelles
- Interpréter les résultats d'un modèle à variables latentes

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Chercheur et doctorant en sciences sociales, marketing ou psychologie
- Statisticien travaillant sur des questionnaires
- Chargé d'étude et data analyst

Prérequis

- Bases solides en statistique (régression, ACP)
- Pratique d'un outil statistique (R, SPSS, AMOS, LISREL ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Fondements et estimation des modèles SEM

- Principes théoriques des modèles à équations structurelles (SEM)
- Variables latentes et variables manifestes
- Modèle de mesure et modèle structurel
- Présentation des outils (lavaan, AMOS, LISREL)
- Spécification et estimation d'un premier modèle
- Analyse factorielle confirmatoire

JOUR 2 Évaluation, modèles avancés et applications

- Indices d'ajustement et qualité du modèle
- Modification et amélioration du modèle
- Comparaison de groupes et effets médiateurs / modérateurs
- Interprétation des résultats
- Bonnes pratiques de publication
- Atelier sur les données des participants

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-equations-structurelles/>

STATISTIQUES · STA-006

Formation régression PLS

Maîtrisez la régression PLS : prédiction sur données multicolinéaires, peu d'individus et applications industrielles.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : XLSTAT · R

PRÉSENTATION

La formation régression PLS vous donne en 1 jour les compétences pour appliquer cette méthode adaptée aux jeux de données complexes : nombreuses variables explicatives, fortes multicolinéarités, peu d'observations. Vous maîtrisez les principes de la régression PLS et son positionnement face à la régression linéaire classique et à la régression sur composantes principales. Vous comprenez les algorithmes PLS1 (une variable à expliquer) et PLS2 (plusieurs variables à expliquer simultanément), leur logique itérative et leurs propriétés. Vous savez appliquer une régression PLS sur vos données et pouvez interpréter les résultats : composantes, coefficients, VIP, qualité prédictive. Idéale pour statisticiens et qualitatifs en chimie, agroalimentaire ou industrie.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les principes de la régression PLS
- Comprendre les algorithmes PLS1 et PLS2
- Appliquer une régression PLS et interpréter les résultats

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Statisticien et qualitatif en industrie
- Chercheur en chimie, agroalimentaire ou pharmacie
- Data analyst confronté à des jeux de données très larges

Prérequis

- Bases solides en statistique (régression linéaire, ACP)
- Pratique d'un outil statistique (XLSTAT, R, SAS ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Régression PLS : principes et applications

- Introduction à la problématique de la régression PLS
- Rappels des notions de base en régression linéaire
- La régression sur composantes principales
- La régression PLS1 et la régression PLS2
- L'analyse discriminante PLS
- Applications sur des données réelles

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-regression-pls/>

STATISTIQUES · STA-007

Formation analyse conjointe (méthodes de trade-off)

Analyse conjointe et méthodes de trade-off : mesurez les préférences clients et optimisez vos offres produit.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : XLSTAT · Excel

PRÉSENTATION

La formation analyse conjointe vous initie aux méthodes de trade-off, incontournables en études marketing pour mesurer rigoureusement les préférences consommateurs. Sur 2 jours, vous comprenez les principes de l'analyse conjointe : modélisation des préférences, utilités partielles, simulation de parts de marché. Vous savez mettre en place toute la méthodologie de l'analyse conjointe : définition des attributs et niveaux, construction du plan d'expériences, administration du questionnaire, estimation et analyse. Vous maîtrisez l'utilisation d'XLSTAT pour construire une analyse conjointe (CBC, ACBC, MaxDiff). Vous comprenez les méthodes utilisées et savez lire et interpréter les résultats obtenus. Idéale pour chargés d'études, marketeurs et data analysts en études consommateurs.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes de l'analyse conjointe
- Mettre en place toute la méthodologie de l'analyse conjointe
- Maîtriser l'utilisation d'XLSTAT pour construire une analyse conjointe
- Comprendre les méthodes utilisées et interpréter les résultats obtenus

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Chargé d'études consommateurs
- Marketeur et chef de produit
- Data analyst en études marketing
- Chercheur en sciences du comportement

Prérequis

- Bases en statistique descriptive
- Pratique d'Excel et idéalement de XLSTAT

PROGRAMME DÉTAILLÉ

JOUR 1 Principes et conception d'une analyse conjointe

- Principes de l'analyse conjointe et des méthodes de trade-off
- Modélisation des préférences et utilités partielles
- Définition des attributs et des niveaux
- Construction du plan d'expériences
- Conception du questionnaire (CBC, ACBC, MaxDiff)

JOUR 2 Analyse et exploitation des résultats avec XLSTAT

- Estimation des utilités avec XLSTAT
- Lecture et interprétation des résultats
- Simulation de parts de marché
- Segmentation des préférences
- Atelier : étude conjointe complète de bout en bout

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-analyse-conjointe/>

DOMAINE

SQL

Interroger et structurer vos données relationnelles et vos datamarts décisionnels.

2 formations

SQL · SQL-001

Formation SQL - Bases du langage

Apprenez SQL en 3 jours : SELECT, jointures, agrégations, sous-requêtes et bonnes pratiques pour interroger toute base.

DURÉE 3 jours (21 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : SQL · PostgreSQL · MySQL · Microsoft SQL Server · Oracle

PRÉSENTATION

La formation SQL - Bases du langage est incontournable pour toute personne travaillant avec des bases de données relationnelles. Le langage SQL permet d'interroger, manipuler et structurer efficacement vos données, de la simple extraction à la gestion complète d'une base. Sur 3 jours, vous maîtrisez les bases du SQL : SELECT, WHERE, ORDER BY, jointures, agrégations, sous-requêtes. Vous savez différencier les principaux dialectes SQL (PostgreSQL, MySQL, SQLite, Microsoft SQL Server, Oracle, MariaDB) et leurs spécificités. Vous construisez des requêtes SQL adaptées à vos besoins métier et manipulez des données : INSERT, UPDATE, DELETE, transactions. La formation s'appuie sur des exemples pratiques et des exercices concrets, avec la possibilité de travailler sur vos propres jeux de données. Idéale pour data analysts débutants, profils métier et utilisateurs souhaitant gagner en autonomie.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser les bases du SQL
- Différencier les dialectes SQL
- Construire une requête SQL
- Manipuler des données avec SQL

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et analyste métier débutant
- Utilisateur potentiel de bases de données
- Expert BI souhaitant consolider ses bases SQL
- Profil non technique amené à interroger des données

Prérequis

- Aucun prérequis technique
- Aisance avec un outil de traitement de données (Excel ou équivalent)

PROGRAMME DÉTAILLÉ

JOUR 1 Bases de données relationnelles et premières requêtes

- Qu'est-ce qu'une base de données relationnelle et son schéma
- Les différents SGBD (Oracle, MySQL, PostgreSQL, SQL Server)
- Structure d'une base : base, schéma, tables
- Types d'instructions SQL : DDL, DML, DCL
- Mise à jour des données : INSERT, UPDATE, DELETE
- Extraire les données : SELECT, WHERE, NULL, DISTINCT, alias

JOUR 2 Agrégation et jointures

- Fonctions d'agrégats : MIN, MAX, AVG, SUM
- Clause GROUP BY et calcul d'agrégats relatifs
- Filtrer les agrégats avec HAVING
- Principe des jointures : interne et externe
- Opérateurs ensemblistes (UNION, INTERSECT)
- Les vues : principe et intérêt

JOUR 3 Fonctions avancées et optimisation

- Conversion de types et opérateur CASE
- Tri des données avec ORDER BY
- Opérations sur les chaînes de caractères et les dates
- Sous-requêtes
- Bonnes pratiques d'écriture et d'optimisation des requêtes

- Atelier sur vos propres jeux de données
-

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-sql-bases-du-langage/>

SQL · SQL-002

Formation SQL pour la construction de datamarts décisionnels

SQL pour datamarts décisionnels : modélisation en étoile, ETL, fonctions analytiques et bonnes pratiques de reporting.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : SQL · PostgreSQL · CTE · window functions

PRÉSENTATION

La formation SQL pour la construction de datamarts décisionnels vous apprend en 2 jours à exploiter SQL au-delà de l'extraction simple. Vous savez concevoir des datamarts décisionnels cohérents en appliquant les principes de la modélisation dimensionnelle : modèle en étoile, modèle en flocon, slowly changing dimensions. Vous identifiez les indicateurs et la granularité adaptés aux besoins métiers, en lien avec les utilisateurs et les analyses cibles. Vous construisez tables de faits et dimensions en SQL, en exploitant les fonctionnalités avancées du langage : window functions, CTE, MERGE, partitionnement. Vous mettez en place des contrôles de qualité des données : tests de cohérence, contraintes, alertes, audits. Idéale pour data engineers, développeurs BI et data analysts confirmés concevant des couches d'analyse robustes.

OBJECTIFS PÉDAGOGIQUES

- Concevoir des datamarts décisionnels cohérents
- Identifier indicateurs et granularité adaptés aux besoins métiers
- Construire tables de faits et dimensions en SQL
- Mettre en place des contrôles de qualité des données

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data engineer et développeur BI
- Data analyst confirmé concevant des couches d'analyse
- Architecte décisionnel

Prérequis

- Bonne maîtrise de SQL (jointures, agrégations, sous-requêtes)
- Notions de modélisation de données

PROGRAMME DÉTAILLÉ

JOUR 1 Du besoin métier au modèle décisionnel

- Comprendre le rôle d'un datamart dans un SI décisionnel
- Identifier indicateurs, axes d'analyse et granularité
- Analyse collective d'un besoin métier réel
- Modélisation faits / dimensions
- Construction des requêtes SQL analytiques
- Première version du datamart

JOUR 2 Construction, qualité et exploitation

- Agrégations, jointures et CTE
- Calcul et validation des indicateurs
- Tests unitaires simples sur les résultats
- Contrôles de complétude et de cohérence
- Détection d'anomalies
- Préparation à l'exploitation BI

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-sql-datamarts-decisionnels/>

DOMAINE

Logiciels Data Science

Dataiku, Databricks, Snowflake, KNIME : maîtriser les grandes plateformes data.

9 formations

LOGICIELS DATA SCIENCE · LDS-001

Formation Dataiku DSS

Prenez en main Dataiku DSS : préparation visuelle, machine learning low-code et industrialisation de projets data.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Dataiku DSS · Python · SQL · Spark

PRÉSENTATION

La formation Dataiku DSS vous apprend à exploiter la plateforme leader du data science low-code. Sur 2 jours, vous apprenez à utiliser l'interface de DSS pour naviguer, créer des datasets et organiser vos projets. Vous comprenez les principes des installations et systèmes de licensing de Dataiku, pour mieux dialoguer avec les administrateurs de la plateforme. Vous construisez des projets et des flux de manipulation de données avec les recettes visuelles, SQL et Spark. Vous mettez en place un premier modèle de machine learning à l'aide du visual ML et comprenez la mise en production des modèles et des flux : scoring, API, déploiement automatisé. Idéale pour data analysts, data scientists et chefs de projet souhaitant accélérer leurs livraisons grâce à Dataiku.

OBJECTIFS PÉDAGOGIQUES

- Utiliser l'interface de DSS
- Comprendre les principes des installations et systèmes de licensing
- Construire des projets et des flux de manipulation de données
- Mettre en place un premier modèle de machine learning
- Comprendre la mise en production des modèles et des flux

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst souhaitant accélérer ses livraisons
- Data scientist intégrant une équipe utilisant Dataiku
- Chef de projet et architecte data sur Dataiku

Prérequis

- Notions de manipulation de données (SQL ou Excel avancé)
- Bases en statistique ou en machine learning appréciées

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte de Dataiku DSS et manipulation des données

- Introduction à Data Science Studio : usages, utilisateurs, positionnement
- Mode de licensing et d'utilisation
- L'interface DSS : notions de projet, de flux et de Lab
- Travailler en équipe : partage et versionnement
- Préparation des données : import (CSV, XLS, SQL), recettes, transformations
- Gestion des types, données textuelles et jointures de fichiers

JOUR 2 Visualisation, machine learning et mise en production

- La data visualisation avec DSS
- Construction de projets et de flux complets
- Mise en place d'un premier modèle de machine learning (Visual ML)
- Évaluation et comparaison de modèles
- Mise en production des modèles et des flux : scoring, API, automatisation
- Bonnes pratiques et déploiement

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-dataiku/>

LOGICIELS DATA SCIENCE · LDS-002

Formation Alteryx pour utilisateurs

Maîtrisez Alteryx Designer : workflows visuels de préparation, jointures, géocodage et analytique self-service.

DURÉE 2 jours (14 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Alteryx Designer · Power BI · Tableau · Qlik Sense

PRÉSENTATION

La formation Alteryx vous donne en 2 jours les compétences pour construire des workflows analytiques sans coder. Vous découvrez l'environnement Alteryx Designer et ses fonctionnalités principales : interface, palette d'outils, exécution. Vous apprenez à connecter, manipuler et transformer des données issues de sources variées (Excel, SQL, API, CRM, etc.). Vous automatisez la préparation des données et améliorez leur qualité grâce aux outils de nettoyage, déduplication, géocodage. Vous créez des workflows optimisés pour gagner du temps et limiter les erreurs, utilisez les outils analytiques pour explorer et visualiser les résultats, puis exportez et partagez vos données vers Power BI, Tableau, Qlik Sense ou autres outils BI. Idéale pour analystes et contrôleurs cherchant à industrialiser leurs préparations de données.

OBJECTIFS PÉDAGOGIQUES

- Découvrir l'environnement Alteryx Designer et ses fonctionnalités principales
- Connecter, manipuler et transformer des données issues de sources variées (Excel, SQL, API, CRM, etc.)
- Automatiser la préparation des données et améliorer leur qualité
- Créer des workflows optimisés pour gagner du temps et limiter les erreurs
- Utiliser les outils analytiques pour explorer et visualiser les résultats
- Exporter et partager les données vers Power BI, Tableau, Qlik Sense ou autres outils BI

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst souhaitant automatiser ses préparations de données
- Contrôleur de gestion et analyste métier
- Utilisateur d'Excel cherchant un outil plus puissant

Prérequis

- Bonne maîtrise d'Excel
- Aucune connaissance préalable d'Alteryx n'est nécessaire

PROGRAMME DÉTAILLÉ

JOUR 1 Prise en main et manipulation des données

- Introduction à Alteryx et à l'environnement de travail
- Découverte de Designer : interface, palette d'outils, exécution
- Connexion à des sources variées (Excel, SQL, API, CRM)
- Manipulation et transformation des données
- Nettoyage, déduplication et amélioration de la qualité

JOUR 2 Workflows analytiques et partage

- Création de workflows optimisés
- Outils analytiques : exploration et visualisation des résultats
- Analyses spatiales et géocodage
- Automatisation et industrialisation des traitements
- Export et partage vers Power BI, Tableau, Qlik Sense

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-alteryx/>

LOGICIELS DATA SCIENCE · LDS-003

Formation Databricks pour Utilisateurs : Maîtriser l'Analyse et le Traitement des Données

Databricks pour utilisateurs : notebooks, Spark, Delta Lake et analyse de données distribuée sur la lakehouse platform.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Databricks · Apache Spark · PySpark · Delta Lake · SQL · Python

PRÉSENTATION

La formation Databricks pour utilisateurs vous initie en 3 jours à la plateforme lakehouse leader du marché. Vous comprenez les fondamentaux de Databricks et son positionnement dans l'écosystème Big Data, face aux solutions traditionnelles d'entrepôt et de data lake. Vous explorez l'interface Databricks et manipulez des notebooks collaboratifs en Python, SQL ou Scala. Vous ingérez, transformez et analysez des jeux de données avec PySpark et Spark SQL. Vous maîtrisez l'architecture des clusters Databricks et optimisez les performances de vos traitements. Vous construisez des pipelines de données et orchestrez des workflows avec Databricks Workflows ou Delta Live Tables. Vous appliquez des cas d'usage concrets autour de la business intelligence, de la data science et du machine learning.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les fondamentaux de Databricks et son positionnement dans l'écosystème Big Data
- Explorer l'interface Databricks et manipuler des notebooks collaboratifs
- Ingérer, transformer et analyser des jeux de données avec PySpark et SQL
- Maîtriser l'architecture des clusters Databricks et optimiser les performances
- Construire des pipelines de données et orchestrer des workflows
- Appliquer des cas d'usage concrets autour de la BI, de la data science et du machine learning

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst, data scientist et data engineer utilisant Databricks
- Architecte data sur des plateformes lakehouse
- Profil tech accompagnant un déploiement Databricks

Prérequis

- Connaissances de base en SQL
- Notions de Python ou Spark appréciées

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte de Databricks et premiers traitements

- Fondamentaux de Databricks et positionnement dans l'écosystème Big Data
- Architecture lakehouse et concepts clés
- Exploration de l'interface et des notebooks collaboratifs
- Premiers traitements en Python et SQL
- Gestion des clusters

JOUR 2 Traitement de données avec Spark et Delta Lake

- Ingestion, transformation et analyse avec PySpark et Spark SQL
- Delta Lake : fiabilité et performance des données
- Optimisation des performances et architecture des clusters
- Gestion des tables et du catalogue
- Travaux pratiques sur de grands jeux de données

JOUR 3 Pipelines, orchestration et cas d'usage

- Construction de pipelines de données
- Orchestration de workflows (Workflows, Delta Live Tables)
- Cas d'usage business intelligence
- Cas d'usage data science et machine learning

- Atelier de synthèse

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-databricks/>

LOGICIELS DATA SCIENCE · LDS-004

Formation Microsoft Azure Machine Learning

Azure Machine Learning : entraînement, déploiement et MLOps sur la plateforme cloud Microsoft pour la data science.

DURÉE 3 jours (21 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Microsoft Azure · Azure Machine Learning Studio · Python · scikit-learn

PRÉSENTATION

La formation Microsoft Azure Machine Learning vous apprend en 3 jours à exploiter la plateforme cloud de Microsoft pour vos projets de data science. Vous prenez en main l'interface d'Azure Machine Learning Studio, ses workspaces, datasets, expérimentations et environnements. Vous apprenez à choisir le bon algorithme selon la problématique métier et à identifier le meilleur modèle parmi plusieurs algorithmes équivalents, en utilisant les outils intégrés d'AutoML et les modules d'évaluation. Vous exploitez une expérience au travers d'un web service : déploiement de modèles, exposition d'endpoints en temps réel ou batch, intégration avec d'autres services Azure. Idéale pour data scientists et ingénieurs souhaitant industrialiser leurs modèles dans un environnement cloud Microsoft.

OBJECTIFS PÉDAGOGIQUES

- Prendre en main l'interface d'Azure Machine Learning Studio
- Choisir le bon algorithme selon la problématique et le meilleur parmi plusieurs algorithmes équivalents
- Exploiter une expérience au travers d'un web service

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data scientist souhaitant industrialiser ses modèles sur Azure
- Data engineer travaillant sur l'écosystème Microsoft
- Ingénieur ML opérant en environnement cloud

Prérequis

- Bases du machine learning
- Pratique de Python et de scikit-learn
- Notions de cloud Azure appréciées

PROGRAMME DÉTAILLÉ

JOUR 1 Introduction et prise en main d'Azure ML

- Introduction à Azure Machine Learning et à l'écosystème Azure
- Prise en main de l'interface Azure ML Studio
- Workspaces, datasets et environnements
- Premières expérimentations
- Notions de cloud appliquées au machine learning

JOUR 2 Entraînement et sélection de modèles

- Choisir le bon algorithme selon la problématique
- Entraînement de modèles et comparaison de plusieurs algorithmes
- AutoML et réglage des hyperparamètres
- Modules d'évaluation et métriques
- Sélection du meilleur modèle

JOUR 3 Déploiement et MLOps

- Exploitation d'une expérience via un web service
- Déploiement d'endpoints (temps réel et batch)
- Pipelines Azure ML
- Notions de MLOps : versionnement, monitoring
- Intégration avec les autres services Azure

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-azure-machine-learning/>

LOGICIELS DATA SCIENCE · LDS-005

Formation Snowflake pour utilisateurs - Maîtriser la Gestion et l'Analyse des Données Cloud

Snowflake pour utilisateurs : architecture cloud, SQL, performance et bonnes pratiques de la data platform leader.

DURÉE 3 jours (21 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Snowflake · SQL · Power BI · Tableau

PRÉSENTATION

La formation Snowflake pour utilisateurs vous donne en 3 jours les clés pour exploiter pleinement la data platform cloud Snowflake. Vous comprenez l'architecture Snowflake et son fonctionnement : séparation compute/storage, virtual warehouses, micro-partitions, services cloud. Vous prenez en main l'interface utilisateur pour explorer et interroger des bases de données, et exécutez des requêtes SQL avancées sur Snowflake. Vous apprenez à importer et exporter des données depuis et vers Snowflake (COPY, Snowpipe, stages internes et externes). Vous gérez les rôles et permissions pour sécuriser l'accès aux données via le modèle RBAC. Vous utilisez Snowflake pour le reporting et la business intelligence, et optimisez l'exécution des requêtes et la gestion des ressources pour maîtriser les coûts. Idéale pour analystes et data engineers.

OBJECTIFS PÉDAGOGIQUES

- Comprendre l'architecture Snowflake et son fonctionnement
- Prendre en main l'interface utilisateur pour explorer et interroger des bases de données
- Exécuter des requêtes SQL avancées sur Snowflake
- Importer et exporter des données depuis et vers Snowflake
- Gérer les rôles et permissions pour sécuriser l'accès aux données
- Utiliser Snowflake pour le reporting et la business intelligence
- Optimiser l'exécution des requêtes et la gestion des ressources

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et data scientist utilisant Snowflake
- Data engineer construisant des pipelines vers Snowflake
- Architecte data sur des plateformes cloud

Prérequis

- Bonne maîtrise de SQL
- Notions de modélisation de données

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte de Snowflake et exploration des données

- Comprendre Snowflake et son écosystème
- Architecture : séparation compute/storage, virtual warehouses, micro-partitions
- Prise en main de l'interface utilisateur
- Exploration et interrogation de bases de données
- Premières requêtes SQL

JOUR 2 Requêtes avancées et chargement de données

- Requêtes SQL avancées sur Snowflake
- Importer et exporter des données (COPY, Snowpipe, stages)
- Gestion des formats et des données semi-structurées
- Partage de données (data sharing)
- Travaux pratiques

JOUR 3 Sécurité, BI et optimisation

- Gestion des rôles et permissions (RBAC)
- Sécurisation de l'accès aux données
- Utilisation de Snowflake pour le reporting et la BI

- Optimisation des requêtes et gestion des ressources
 - Maîtrise des coûts
-

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-snowflake/>

LOGICIELS DATA SCIENCE · LDS-006

Formation logiciel KNIME

KNIME Analytics Platform : workflows visuels, préparation, machine learning et analytique open source sans code.

DURÉE 3 jours (21 h)	PARTICIPANTS 6 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : KNIME Analytics Platform · Python · R · Spark

PRÉSENTATION

La formation KNIME Analytics Platform vous initie en 3 jours au logiciel open source de data science low-code. Vous maîtrisez l'utilisation de l'interface du logiciel KNIME, ses nœuds, ses workflows et son organisation des projets. Vous savez explorer des données avec KNIME et construire un flux de traitement data science complet : ingestion, préparation, transformation, modélisation, évaluation. Vous maîtrisez la programmation avec KNIME et les intégrations possibles. Vous allez plus loin sur l'importation et le reporting avec KNIME (génération de PDF, dashboards, BIRT). Vous comprenez l'utilisation des technologies Big Data dans KNIME (Spark, Hive) et êtes capable d'utiliser d'autres outils de data science (Python, R) dans KNIME. Idéale pour analystes, statisticiens et data scientists cherchant un outil flexible et gratuit.

OBJECTIFS PÉDAGOGIQUES

- Maîtriser l'utilisation de l'interface du logiciel KNIME
- Explorer des données avec KNIME
- Construire un flux de traitement data science
- Maîtriser la programmation avec KNIME
- Aller plus loin sur l'importation et le reporting avec KNIME
- Comprendre l'utilisation des technologies Big Data dans KNIME
- Utiliser d'autres outils de data science dans KNIME

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst et statisticien souhaitant un outil low-code
- Profil métier exploitant des données avec KNIME
- Data scientist cherchant un outil flexible et gratuit

Prérequis

- Notions de manipulation de données
- Aucune connaissance préalable de KNIME n'est nécessaire

PROGRAMME DÉTAILLÉ

JOUR 1 Prise en main de KNIME et exploration des données

- Présentation de KNIME Analytics Platform
- L'interface : nœuds, workflows, organisation des projets
- Import et exploration des données
- Premières transformations
- Bonnes pratiques de construction de workflows

JOUR 2 Flux de traitement et data science

- Construction d'un flux de traitement data science complet
- Préparation et transformation avancée des données
- Premiers modèles de machine learning
- Évaluation des modèles
- Programmation et logique conditionnelle dans KNIME

JOUR 3 Big Data, reporting et intégrations

- Importation avancée et reporting (PDF, BIRT)
- Technologies Big Data dans KNIME (Spark, Hive)
- Intégration d'autres outils (Python, R)
- Automatisation et déploiement
- Atelier de synthèse

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-logiciel-knime/>

LOGICIELS DATA SCIENCE · LDS-007

Formation OpenRefine initiation

OpenRefine pour le nettoyage de données : déduplication, normalisation, GREL et préparation rapide de gros fichiers.

DURÉE 1 jour (7 h)	PARTICIPANTS 7 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : OpenRefine · GREL

PRÉSENTATION

La formation OpenRefine initiation vous apprend en une journée à exploiter le logiciel libre de référence pour le nettoyage de données. Vous apprenez à importer des données dans OpenRefine depuis différents formats (CSV, Excel, JSON, XML, bases de données). Vous savez préparer et nettoyer des données dans OpenRefine grâce aux facettes, aux transformations en masse, à la déduplication et à la réconciliation avec des bases externes comme Wikidata. Vous savez manipuler des tables en utilisant OpenRefine : regroupements, scissions, transformations conditionnelles, langage d'expression GREL. La formation s'appuie sur des cas pratiques courants : nettoyage de bases clients, normalisation d'adresses, harmonisation de catalogues, préparation de jeux de données ouvertes. Idéale pour profils data, archivistes et documentalistes.

OBJECTIFS PÉDAGOGIQUES

- Importer des données dans OpenRefine
- Préparer et nettoyer des données dans OpenRefine
- Manipuler des tables en utilisant OpenRefine

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Data analyst, data steward et profil data
- Archiviste et documentaliste
- Chargé de données ouvertes en collectivité ou institution

Prérequis

- Aucun prérequis technique
- Pratique d'Excel ou d'un tableur

PROGRAMME DÉTAILLÉ

JOUR 1 Nettoyer et préparer ses données avec OpenRefine

- Présentation d'OpenRefine et de l'interface
- Importer des données (CSV, Excel, JSON, XML, bases)
- Explorer avec les facettes et les filtres
- Préparer et nettoyer : transformations en masse, déduplication
- Manipuler des tables : regroupements, scissions, langage GREL
- Réconciliation avec des bases externes et export des données

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-logiciel-openrefine-initiation/>

LOGICIELS DATA SCIENCE · LDS-008

Formation langage Julia

Découvrez Julia : un langage moderne, rapide et expressif pour le calcul scientifique et la data science haute performance.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : Julia · DataFrames.jl · Plots.jl · Flux.jl · Pluto.jl

PRÉSENTATION

La formation langage Julia vous initie en 2 jours à un langage moderne combinant la lisibilité de Python et la performance du C. Vous savez lancer des projets basés sur le langage Julia : installation, environnements, gestion des packages avec Pkg. Vous maîtrisez les forces et les faiblesses du langage : performance native, multiple dispatch, typage progressif, écosystème scientifique en croissance, communauté plus restreinte. Vous identifiez les outils nécessaires et les principes de codage en Julia : types, fonctions, modules, métaprogrammation. La formation aborde les principaux packages scientifiques (DataFrames.jl, Plots.jl, Flux.jl, DifferentialEquations.jl) et s'appuie sur des cas concrets de calcul scientifique et de data science. Idéale pour chercheurs, ingénieurs et data scientists confrontés à des contraintes de performance.

OBJECTIFS PÉDAGOGIQUES

- Lancer des projets basés sur le langage Julia
- Maîtriser les forces et les faiblesses du langage
- Identifier les outils nécessaires et les principes de codage en Julia

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Chercheur et ingénieur en calcul scientifique
- Data scientist confronté à des contraintes de performance
- Développeur curieux des langages modernes

Prérequis

- Bonne maîtrise d'un langage de programmation (Python, R, Matlab ou équivalent)
- Notions de calcul scientifique appréciées

PROGRAMME DÉTAILLÉ

JOUR 1 Découverte du langage Julia et programmation

- Présentation de Julia : forces et faiblesses
- Installation, environnements et gestion des packages (Pkg)
- Syntaxe de base : types, variables, fonctions
- Structures de contrôle et de données
- Le multiple dispatch et le système de types
- Modules et organisation du code

JOUR 2 Data science et calcul scientifique

- Manipulation de données avec DataFrames.jl
- Visualisation avec Plots.jl
- Calcul numérique et algèbre linéaire
- Machine learning avec Flux.jl
- Interopérabilité avec Python et R
- Atelier : cas concret de calcul scientifique

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-au-langage-julia/>

LOGICIELS DATA SCIENCE · LDS-009

Formation AutoML avec H2O.ai

Automatisez la création de modèles de machine learning performants avec la plateforme H2O.ai.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : H2O.ai · Python · R · AutoML

PRÉSENTATION

La formation automl avec h2o.ai vous permet, en 2 jour(s) (14 heures), de monter en compétences sur ce sujet clé de la data et de l'intelligence artificielle. Animée par des experts praticiens, elle alterne apports, démonstrations et travaux pratiques sur des jeux de données concrets, avec un effectif limité pour favoriser les échanges. Disponible en présentiel et en classe virtuelle. Niveau : intermédiaire.

OBJECTIFS PÉDAGOGIQUES

- Créer des modèles ML automatisés
- Comparer les performances des modèles
- Déployer des modèles H2O.ai

PROFILS DES STAGIAIRES & PRÉREQUIS

Profils

- Data scientist
- Analyste avancé souhaitant accélérer la modélisation

Prérequis

- Connaissances en machine learning de base
- Pratique de Python ou R

PROGRAMME DÉTAILLÉ

JOUR 1 AutoML et génération de modèles

- Principes de l'AutoML et présentation de la plateforme H2O.ai
- Préparation des données et configuration des expériences
- Génération automatisée et comparaison de modèles (leaderboard)

JOUR 2 Interprétabilité et déploiement

- Interprétabilité et explicabilité des modèles
- Déploiement et export des modèles H2O
- Atelier : pipeline AutoML complet sur cas métier

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-automl-h2o-ai/>

DOMAINE

Gouvernance & Qualité

Gouvernance, qualité des données et automatisation au service de la valeur data.

3 formations

GOUVERNANCE & QUALITÉ · GQA-001

Formation Gouvernance des données en entreprise

Gouvernance des données en entreprise : rôles, processus, data office et démarche pour structurer votre patrimoine data.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris et à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Collibra · Power BI · Qlik · DAMA-DMBOK

PRÉSENTATION

La formation gouvernance des données en entreprise vous donne en 1 jour les clés pour structurer une démarche pérenne. Vous comprenez les principes fondamentaux de la gouvernance des données et leurs enjeux stratégiques pour l'entreprise. Vous identifiez les rôles et responsabilités clés (Data Owner, Data Steward, Data Protection Officer, CDO) et savez comment structurer une gouvernance efficace, adaptée à la taille et à la maturité de l'organisation. Vous connaissez les principales approches, méthodes et frameworks (COBIT, DAMA-DMBOK, EDM Council) pour mettre en place une gouvernance des données. Vous êtes capable de définir et piloter des politiques de gestion, de contrôle et de sécurité des données en adéquation avec la stratégie de l'entreprise. Vous savez évaluer la maturité de la gouvernance et mettre en œuvre un plan d'amélioration continue. Idéale pour CDO, responsables data et chefs de projet.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes fondamentaux de la gouvernance des données et leurs enjeux stratégiques pour l'entreprise
- Identifier les rôles et responsabilités clés (Data Owner, Data Steward, Data Protection Officer, etc.) et structurer une gouvernance efficace
- Connaître les principales approches, méthodes et frameworks (COBIT, DAMA-DMBOK, etc.) pour mettre en place une gouvernance des données
- Définir et piloter des politiques de gestion, de contrôle et de sécurité des données en adéquation avec la stratégie de l'entreprise
- Évaluer la maturité de la gouvernance des données et mettre en œuvre un plan d'amélioration continue

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Chief Data Officer et responsable data
- Chef de projet data ou SI
- Data steward et data owner
- Responsable conformité et DPO

Prérequis

- Culture générale data ou SI
- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Gouvernance des données en entreprise

- Introduction : définition et enjeux (stratégiques, réglementaires, opérationnels)
- Composantes clés : politiques, processus, rôles, technologies
- Rôles et responsabilités (Data Owner, Data Steward, DPO, CDO)
- Frameworks de référence (COBIT, DAMA-DMBOK)
- Définir et piloter des politiques de gestion et de sécurité des données
- Évaluer la maturité et bâtir un plan d'amélioration continue

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-gouvernance/>

GOUVERNANCE & QUALITÉ · GQA-002

Formation qualité des données en entreprise

Qualité des données : dimensions, indicateurs, processus de remédiation et outils pour fiabiliser votre patrimoine data.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : Collibra · Talend · Power BI · Excel

PRÉSENTATION

La formation qualité des données en entreprise vous apprend en 1 jour à structurer une démarche data quality efficace. Vous comprenez les bases de la qualité des données : définitions, enjeux, indicateurs de mesure, coûts de la non-qualité. Vous identifiez les facteurs clés qui influencent la qualité des données : complétude, cohérence, exactitude, fraîcheur, unicité, conformité. Vous découvrez les bonnes pratiques et méthodes d'amélioration de la qualité : profilage, règles métier, processus de remédiation, gouvernance, outils dédiés. Vous mettez en œuvre des ateliers pratiques pour analyser et améliorer la qualité de vos données. Vous établissez une feuille de route pour la mise en place d'une démarche qualité dans votre entreprise. Idéale pour data stewards, responsables qualité, CDO et chefs de projet menant des initiatives de fiabilisation.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les bases de la qualité des données : définitions, enjeux, indicateurs de mesure
- Identifier les facteurs clés qui influencent la qualité des données (complétude, cohérence, exactitude, etc.)
- Découvrir les bonnes pratiques et méthodes d'amélioration de la qualité
- Mettre en œuvre des ateliers pratiques pour analyser et améliorer la qualité de ses données
- Établir une feuille de route pour la mise en place d'une démarche qualité dans son entreprise

PROFILS DES STAGIAIRES & PRÉREQUIS**Profil**

- Data steward et data owner
- Responsable qualité et conformité
- Chief Data Officer et chef de projet data
- Profil métier impliqué dans des projets de fiabilisation

Prérequis

- Culture générale data ou SI
- Aucun prérequis technique

PROGRAMME DÉTAILLÉ**JOUR 1 Qualité des données : concepts et démarche**

- Introduction : définitions et concepts clés de la qualité des données
- Les grandes dimensions : complétude, cohérence, exactitude, fraîcheur
- Enjeux et impacts sur la performance de l'entreprise
- Études de cas de mauvaise qualité et leurs conséquences
- Lien avec la gouvernance et la conformité RGPD
- Facteurs influençant la qualité et indicateurs de suivi (KPIs)

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-qualite-des-donnees/>

GOUVERNANCE & QUALITÉ · GQA-003

Formation RPA pour les équipes data et SI

RPA pour équipes data et SI : automatisation de processus, robots logiciels et cas d'usage pour gagner en productivité.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : UiPath · Power Automate · Blue Prism

PRÉSENTATION

La formation RPA pour les équipes data et SI vous initie en 2 jours aux principes et apports de la Robotic Process Automation. Vous comprenez les principes et limites du RPA : automatisation à l'écran, robots logiciels, attended vs unattended, RPA vs API, RPA vs scripts. Vous savez identifier des processus automatisables grâce à une grille d'évaluation : volumes, règles claires, stabilité, ROI. Vous positionnez le RPA dans une architecture data et SI, en lien avec les autres outils d'automatisation et de transformation : ETL, ESB, iPaaS, IA. Vous contribuez au cadrage d'un projet RPA : choix d'outils, gouvernance, périmètre, conduite du changement. Idéale pour directions SI, responsables data et chefs de projet intégrant la RPA dans une stratégie d'automatisation.

OBJECTIFS PÉDAGOGIQUES

- Comprendre les principes et limites du RPA
- Identifier des processus automatisables
- Positionner le RPA dans une architecture data et SI
- Contribuer au cadrage d'un projet RPA

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Responsable SI et architecte d'entreprise
- Responsable data en réflexion sur l'automatisation
- Chef de projet cadrant des initiatives RPA

Prérequis

- Culture générale SI ou data
- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Fondamentaux et identification des processus

- Principes du RPA et cas d'usage
- Différences RPA / scripts / traitements data
- Critères d'éligibilité à l'automatisation
- Études de cas data et SI

JOUR 2 Intégration et gouvernance

- Intégration du RPA dans le SI
- Gestion des erreurs et sécurité
- Bonnes pratiques de maintenance
- Synthèse et recommandations

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-rpa-data-et-si/>

DOMAINE

Culture data

Diffuser la culture de la donnée, de l'open data au big data, dans toute l'entreprise.

3 formations

CULTURE DATA · CDA-001

Formation culture générale de la donnée - De l'open data au big data

Culture générale de la donnée : open data, big data, IA et data science expliqués clairement aux non-spécialistes.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : open data · big data · machine learning

PRÉSENTATION

La formation culture générale de la donnée - de l'open data au big data vous offre en 2 jours un panorama complet des concepts clés et des bonnes pratiques de l'exploitation des données. Les données sont au cœur des stratégies des entreprises et institutions : big data, data science, machine learning, deep learning, open data... Ces termes sont omniprésents mais souvent flous. La formation vous aide à acquérir une vision d'ensemble des nouvelles pratiques de diffusion et d'exploitation de données. Vous disposez d'un panorama des outils et des meilleures pratiques du domaine. Vous obtenez les clés de lecture pour mettre en place une stratégie data dans votre organisation. Vous savez guider un prestataire ou un collaborateur dans la mise en œuvre d'un projet data. Idéale pour décideurs, managers et responsables métiers souhaitant développer leur culture data.

OBJECTIFS PÉDAGOGIQUES

- Avoir une vision d'ensemble des nouvelles pratiques de diffusion et d'exploitation de données
- Disposer d'un panorama des outils et des meilleures pratiques
- Avoir les clés de lecture pour mettre en place une stratégie data
- Pouvoir guider un prestataire dans la mise en œuvre d'un projet data

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Décideur, manager et responsable métier
- Membre des équipes Innovation, Digital, DSI, Stratégie, Marketing, RH
- Toute personne souhaitant développer sa culture data

Prérequis

- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Comprendre l'écosystème des données

- Workshop 1 - Qu'est-ce qu'une donnée ? Définition et typologies
- Données quantitatives, qualitatives, structurées, non structurées
- Petite histoire sociale des données et datafication
- Workshop 2 - Regards critiques sur la révolution des données
- Open data, big data, linked data, data API
- Cadre juridique, vie privée et modèles économiques

JOUR 2 Ouvrir la boîte noire de la data science

- Qu'est-ce que la data science ?
- Workshop 4 - Quelles compétences recruter dans un projet data science ?
- L'écosystème matériel, logiciel et informationnel
- Le data pipeline : acquérir, vérifier, nettoyer, analyser, présenter
- Workshop 5 - Datavisualisation : quel format et quel outil choisir ?
- Introduction au machine learning et au deep learning

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-culture-generale-donnees/>

CULTURE DATA · CDA-002

Formation Open Data

Open Data : enjeux, écosystème, exploitation des jeux de données publics et cas d'usage pour valoriser les données ouvertes.

DURÉE 2 jours (14 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
--------------------------------	------------------------------	---	------------------------------------

Technologies : data.gouv.fr · CKAN · OpenDataSoft · Eurostat

PRÉSENTATION

La formation Open Data vous donne en 2 jours une vision complète de l'écosystème des données ouvertes. Vous comprenez le contexte et les opportunités liés à la mise en œuvre de l'Open Data : enjeux de transparence, valorisation économique, innovation, démocratie. Vous découvrez le cadre réglementaire français et européen (loi pour une République numérique, directive PSI, Data Governance Act), les principales plateformes (data.gouv.fr, Eurostat, OpenStreetMap, banque mondiale) et les standards d'ouverture (cinq étoiles, formats, licences). Vous connaissez les outils pour l'ouverture des données et leur réutilisation : portails, API, formats pivots, qualité des métadonnées. Idéale pour collectivités, administrations, chercheurs, journalistes data et entreprises souhaitant intégrer l'open data à leurs démarches d'innovation ou de transparence.

OBJECTIFS PÉDAGOGIQUES

- Comprendre le contexte et les opportunités liés à la mise en œuvre de l'Open Data
- Connaître les outils pour l'ouverture des données et leur réutilisation

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Agent de collectivité ou d'administration
- Chercheur et journaliste de données
- Entreprise souhaitant valoriser l'open data
- Responsable data en démarche de transparence

Prérequis

- Aucun prérequis technique

PROGRAMME DÉTAILLÉ

JOUR 1 Comprendre l'open data

- Origines historiques de l'ouverture des données
- Les grands principes de l'ouverture des données
- Acteurs politiques, économiques et associatifs de l'open data
- Cadre juridique et open government
- Les formats de données et les portails (CKAN, OpenDataSoft, data.gouv.fr)
- Les success stories de l'open data

JOUR 2 Ouvrir des données

- Concevoir une stratégie d'open data (leviers politiques, organisationnels, économiques)
- Identification des données et Open Data Census
- Négociations et frictions dans l'ouverture
- Extraction, conversion et mise en qualité des données
- Publication, métadonnées et standards émergents
- Animation de l'open data et relation avec les réutilisateurs

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/formation-open-data/>

CULTURE DATA · CDA-003

Atelier acculturation data

Atelier d'acculturation data : format court et interactif pour sensibiliser vos équipes aux enjeux et opportunités data.

DURÉE 1 jour (7 h)	PARTICIPANTS 8 max	FORMATS Présentiel · À distance	LIEU Paris ou à distance
------------------------------	------------------------------	---	------------------------------------

Technologies : open data · big data · data API

PRÉSENTATION

L'atelier acculturation data propose un format court et participatif pour sensibiliser un public large aux enjeux de la donnée. Vous apprenez à comprendre ce qu'est une donnée : nature, typologies (quantitative, qualitative, structurée, non structurée), métadonnées, cycle de vie. Vous êtes capable d'identifier des données pour un projet : sources internes et externes, jeux d'open data, données partenaires. Vous connaissez la terminologie liée aux données : big data, open data, linked data, data API, smart data. L'atelier est basé sur des échanges en petits groupes (mini-projets, quiz, retours d'expériences) pour appréhender les défis liés aux données de manière simple et ludique. Le format est adaptable en durée et en contenu, et nous pouvons le déployer en intra-entreprise pour acculturer un grand nombre de collaborateurs.

OBJECTIFS PÉDAGOGIQUES

- Comprendre ce qu'est une donnée
- Identifier des données pour un projet
- Connaître la terminologie liée aux données

PROFILS DES STAGIAIRES & PRÉREQUIS

Profil

- Tout collaborateur souhaitant s'acculturer aux enjeux data
- Équipe métier impliquée dans un projet data
- Public large dans le cadre d'un déploiement intra-entreprise

Prérequis

- Aucun prérequis

PROGRAMME DÉTAILLÉ

JOUR 1 Atelier acculturation data

- Workshop 1 - Qu'est-ce qu'une donnée ? Identification, définition, étymologie
- Typologies : quantitatives, qualitatives, structurées, non structurées
- Open data, big data, linked data, data API
- Workshop 2 - Analyse du modèle économique d'entreprises de la data
- Les nouveaux métiers et enjeux de la donnée
- Workshop 3 - Utiliser les données pour améliorer son organisation

En savoir plus & sessions à venir : <https://www.stat4decision.com/fr/formations/atelier-acculturation-data/>

Parlons de votre projet de formation

Inter, intra ou sur mesure : nous construisons le parcours adapté à vos équipes et à vos données.

Téléphone +33 (0)1 72 25 40 82

E-mail info@stat4decision.com

Site www.stat4decision.com

Adresse 37-39 avenue Ledru Rollin, 75012 Paris

SIRET 810 489 856 00015 · Déclaration d'activité 11755352275 (préfet de région Île-de-France) · Organisme de formation certifié Qualiopi au titre des actions de formation.