Stat4decision
Data science company

CATALOGUE DES FORMATIONS

FORMATIONS DATA & DATA SCIENCE

2024

Qualiopi
processus certifié
RÉPUBLIQUE FRANÇAISE
La certification qualité a été délivrée au titre de la catégorie d'action suivante : actions de formation

python R Power BI data iku git

Nos formations 5
Notre équipe 7
Notre expérience 7
Nos références 8

Formations Python & DATA 9
Python pour la data science 10
Python pour utilisateur de SAS 12
Python pour utilisateur d'Excel 14
Pandas avancé - manipulation des données 16
Spark avec PySpark en Python 18
Deep Learning avec Python 20
Analyse textuelle avec Python 22
Scikit-learn pour le machine learning 24
Data visualisation avec Python 28
Dash pour la construction de dashboard avec python 31

Formations R 33
R pour la data science 34
Logiciel R pour utilisateur de SAS 36
Visualisation avec R 38
Développement d'applications web avec R shiny 39
Data mining et machine learning avec R 41
R pour la Cartographie et les SIG 43
Séries temporelles avec R 45

Formations data 47

Les fondamentaux du machine learning et du deep learning 48
Les fondamentaux du big data et de la data science 50
Intelligence artificielle - les fondamentaux 52
Open Data 54
Data science – bonnes pratiques et outils 56
Culture générale de la donnée 58
Atelier acculturation data 60
Panorama et choix des briques d'exploitation d'un data lake 61

Formations mises en pratique de la data visualisation 63
Datavisualisation 64
Rendre ses présentations claires et engageantes 66

Formations Business Intelligence 68
Microsoft Power BI - initiation 69
Microsoft Power BI - perfectionnement 72
Microsoft Power BI - dataviz & analytics 74
Atelier Découverte Microsoft Power BI 76
Atelier Langage DAX : expertise 77
Utilisation de R dans Power BI 79
Designer Qlik Sense 81
Développeur Qlik Sense 82
Découverte Qlik Sense 84
QlikView – Initiation 85
Tableau Desktop 86

Formations outils data science 88
Git pour versionner votre code 89

SQL - Bases du langage 91
Découverte OpenRefine 93
Dataiku DSS - Fondamentaux 95
Logiciel KNIME – les fondamentaux 97
Logiciel RapidMiner 99
Langage Julia – les fondamentaux 101
Microsoft Azure Machine Learning 103

Formations statistique et analyse de données 105
Statistique et analyse de données 106
Statistique et analyse de données avec XLSTAT 108
Approche PLS (PLS Path Modeling) 110
Les équations structurelles à variables latentes 112
Régression PLS 114
Analyse conjointe (méthodes de trade-off) 115
XLSTAT-R et programmation avec XLSTAT 117

Nos formations

Les formations organisées par Stat4decision répondent à des critères de qualité clairement énoncés et qui sont essentiels à la qualité de nos enseignements :

- **L'adaptation du contenu au public de la formation** : par un échange entre nos équipes et les apprenants (que ce soit en intra ou en inter), nous construisons des groupes homogènes et proposons des formations ajustées à chaque niveau.
- **Des formateurs spécialisés et utilisateurs** : nos formateurs interviennent uniquement sur leur spécialité. De plus, nous exigeons que nos formateurs soient aussi des utilisateurs des outils et concepts enseignés. La sélection des formateurs est un point central du processus qualité de Stat4decision.
- **Des conditions matérielles de qualité** : quel que soit le type d'action de formation, Stat4decision met à la disposition des apprenants tous les outils nécessaires à leur apprentissage et la réutilisation des savoirs appris (supports papier, clé USB, PDF…).
- **L'amélioration continue des formations** : par la prise en compte des avis des apprenants, Stat4decision garantit une amélioration continue de la qualité de ses actions de formation.

Stat4decisiosn est certifié Qualiopi et vous garantit la mise en place de processus qualité de haut niveau. La certification qualité a été délivrée au titre de la catégorie d'action suivante : actions de formation

Qualiopi
processus certifié
RÉPUBLIQUE FRANÇAISE

La certification qualité a été délivrée au titre de la catégorie d'action suivante : actions de formation

Stat4decision propose trois types de formations :

- **Les formations en présentiel** (inter) organisées dans nos locaux à Paris. Ces formations se font en petits groupes avec maximum 7 participants. Si cela est nécessaire, des ordinateurs sont mis à la disposition des apprenants.
- **Les formations en présentiel sur mesure (intra)** organisées dans vos locaux partout en France. Nos formateurs se déplacent directement dans votre structure et adaptent la formation à vos besoins. Le nombre maximal de participants conseillé est de 8.
- **Les formations en salle de classe virtuelle (en ligne)** organisées en ligne et en direct (synchrone) avec un formateur. Ces formations se limitent à 7 participants et vous permettent de vous connecter en utilisant votre navigateur. Les outils utilisés facilitent l'interactivité avec le formateur et les autres apprenants. Ces formations peuvent être mises en place en inter ou en intra.

L'ensemble de nos formations sont proposées dans ces 3 formats et peuvent aussi être proposées en anglais.

Notre équipe

L'équipe de Stat4decision est composée d'experts en traitement et analyse de données et de formateurs expérimentés.

Sous la direction d'Emmanuel Jakobowicz, responsable pédagogique, les équipes d'experts de Stat4decision vous accompagne afin de monter en compétences sur tous les sujets data.

Nos data scientists sont là aussi pour vous accompagner par des actions de conseils, aussi bien pour structurer votre besoin en termes de projets (big data, analytics, infrastructure, méthodologie) que pour l'analyse de vos données.

Notre expérience

Stat4decision forme et aide des centaines d'utilisateurs annuellement lors d'actions de formation et de sessions de conseil et d'accompagnement.

Les chiffres de 2022

1090	360	9.2/10
STAGIAIRES FORMÉS	JOURS DE FORMATION	SATISFACTION MOYENNE*

Nos références

Stat4decision s'adresse à tous les utilisateurs de la data science aujourd'hui.

Nos clients se répartissent dans de nombreux secteurs.

Stat4decision a la capacité de conseiller et former aussi bien des grandes que des petites structures.

FORMATIONS PYTHON

python™

Python pour la data science

Description

Que vous soyez data scientist, utilisateur de R, analyste SAS ou développeur, et que vous désirez découvrir python pour le traitement des données, cette formation est faite pour vous. Trois jours intensifs pour acquérir les bases de l'utilisation du langage python pour des applications en analyse de données, data mining et data science.

Après une introduction au langage python (ou un rappel si vous avez déjà utilisé python), vous apprendrez à maîtriser les bibliothèques spécialisées en data science, data mining et data visualisation. Basée sur de nombreux exercices pratiques, cette formation vous permet d'être immédiatement opérationnel pour le traitement de vos données grâce à python.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-python-pour-la-data-science/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst désirant découvrir python en data science
- Manager désirant comprendre les spécificité de python
- Développeur spécialisé dans d'autres langages

Prérequis :

- Connaissances de base en traitement de données (statistique et tables de données)
- Pas de prérequis en python

Objectifs pédagogiques

- Maîtriser les bases de python
- Etre capable d'analyser des données avec python
- Etre capable d'automatiser le traitement de données avec python
- Connaître les outils pour la data science de python

Programme détaillé

- Introduction au langage python
    - Pour qui ? Pour quoi faire ? Comment ?

- Les consoles et le débogage en python
- Le langage (la syntaxe, les structures (string, listes, dictionnaires…), les opérateurs et les fonctions)
- Construction d'un outil de calcul de statistiques simples
- Les classes, les objets et la création de bibliothèques

- Python pour le traitement des données
  - Le processus de développement d'un outil en python pour la data science
  - L'importation des données (texte, csv, Excel, web scrapping., SQL..)
  - La préparation et le nettoyage des données (jointure, filtre, transformation, traitement des données manquantes avec pandas, numpy et scipy)
- La data visualisation
  - Présentation des principes de la data visualisation
  - Quelques graphiques simples avec matplotlib : scatter plot, box plot, histogrammes…
  - Présentation de visualisation plus avancées
  - Utilisation de Bokeh pour des visualisations interactives
- Introduction au machine learning : présentation des principes de l'apprentissage automatique
  - Préparation des données avec la bibliothèque scikit-learn
- Cas pratiques en machine learning
  - La prévision en marketing
  - L'extraction d'informations pertinentes
  - Les k-means pour la compression d'images
  - Les k-plus proches voisins pour la prévision de scores
  - Le traitement de données textuelles
  - La classification d'image avec le deep learning
- Aller plus loin
  - L'automatisation et la mise en production de traitements avec python (pipeline)
  - Les bibliothèques pour le big data (pyspark, keras)
  - Discussions et démonstrations de fonctionnalités avancées sur les problématiques des participants

Python pour utilisateur de SAS

Description

Vous êtes utilisateur de SAS et habitué à utiliser le langage SAS au quotidien, vous voulez passer à Python pour vos traitements de données ou utiliser Python en supplément de SAS. Notre formation Python pour utilisateur de SAS vous permet de comprendre les points communs, les forces et faiblesses de ces outils.

Par des exemples comparés, nous vous accompagnons dans une transition douce vers Python depuis SAS.

Plus d'informations :
https://www.stat4decision.com/fr/formations/formation-python-utilisateur-sas/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Utilisateurs de SAS.
- Statisticiens
- Analystes

Prérequis :

- Connaissance de base de SAS
- Aucun prérequis sur l'utilisation de Python

Objectifs pédagogiques

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

Programme détaillé

- Introduction
  - Le langage Python
  - Comparaison de Python et de SAS
  - Python, un investissement pour l'avenir
- Python pour le traitement des données
  - Le processus de développement d'un outil en python pour la data science
  - Les IDE

- Les Jupyter Notebook
- Les packages pour la data

- Les bases du langage Python
- La préparation et le nettoyage des données
  - Charger des données, les procédures SAS et les fonctions Python
  - Manipuler des données comparaison étape data de SAS et le package Pandas
- Statistiques et reporting
  - Les statistiques descriptives – comparaison des procédures SAS et des fonctions Python
  - Construire des graphiques / tableaux de bord avec Python
- Analyse de données et machine learning
  - Comparaison de SAS et de statsmodels
  - Aller plus loin avec le machine learning et scikit-learn
- Discussions et démonstrations des différences sur les problématiques des participants

Python pour utilisateur d'Excel

Description

Vous utilisez Excel au quotidien et vous voulez accélérer vos traitements avec python.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Expert Excel
- Expert BI (Business Intelligence)
- Métiers désirant augmenter sa productivité avec python

Prérequis :

- Maîtrise d'Excel
- Pas de prérequis en python

Objectifs pédagogiques

- Maîtriser les bases de python
- Manipuler des données avec python
- Faire communiquer python et excel
- Construire des visualisation python dans Excel
- Coder en python dans Excel

Programme détaillé

- Introduction au langage python
  - Pour qui ? Pour quoi faire ? Comment ?
  - Les différences avec Excel. Comment coder en python pour Excel ?
  - Le langage (la syntaxe, les structures (string, listes, dictionnaires…), les opérateurs et les fonctions)
- Python pour le traitement des données
  - La manipulation de données avec pandas et ses DataFrame (jointure, filtre, transformation, traitement des données manquantes)
- Faire communiquer Python et Excel, les différentes approches :
  - Depuis un Jupyter Notebook
  - Directement dans Excel avec des compléments (pyxll...)
  - Directement dans Excel avec les nouveautés d'Excel
- Manipulation de données avec python dans Excel
  - Lecture d'un fichier Excel
  - Écriture dans un fichier Excel

  - Maîtriser le chargement des formules Excel en python
  - Générer des graphiques dans Excel avec python
- Faire communiquer Excel avec d'autres applications
  - De Excel vers PowerPoint avec python
  - Automatisation de génération de rapport PowerPoint en utilisant des données Excel avec python
- Aller plus loin avec les nouvelles fonctionnalités d'Excel
  - Le partenariat Microsoft-Anaconda pour vous donner accès à Python dans Excel
  - Générer un modèle de machine learning python et des data visualisations sans sortir d'Excel
- Discussions et démonstrations de fonctionnalités avancées sur les problématiques des participants

Organisation de la formation

Moyens pédagogiques et techniques

- Supports sous forme de Jupyter Notebooks
- Nombreux cas pratiques

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Évaluation par des exercices par le formateur
- Quiz en classe

Pandas avancé - manipulation des données

Description

Au fil des années la librairie pandas est devenue incontournable dans les phases de chargement, de compréhension, de préparation et de transformation des données, en amont du Machine Learning. Elle est également très utile pour produire des indicateurs statistiques et des graphiques.

L'objectif de cette formation est de présenter certaines caractéristiques avancées de pandas, notamment le style de programmation Method Chaining qui facilite la lecture et la compréhension du code.

D'autres points sont abordés comme l'amélioration de représentation des données et la stylisation des DataFrames.

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data engineer
- Data scientist
- Développeur Python

Prérequis :

- Bases en python et en manipulation de données avec python

Objectifs pédagogiques

- Maîtriser certains concepts avancés de pandas
- Être capable d'écrire un traitement en Method Chaining
- Améliorer la représentation des données
- Apprendre à styliser les dataframes

Programme détaillé

- Rappels sur numpy et pandas
- Rappels sur les Series et les DataFrames
- Method Chaining pour la préparation et le traitement des données avec pandas
- Amélioration de la représentation des données
    - Optimisation des types (numériques, catégories)
    - Valeurs manquantes utilisées par défaut
- Stylisation des DataFrames et utilisation de la librairie sparklines

- Mise en pratique par des exercices variés

Organisation de la formation

Moyens pédagogiques et techniques

- Accueil des apprenants dans une salle dédiée à la formation.
- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation.

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Mises en situation.
- Exercices dans des Jupyter notebooks
- Questions orales ou écrites (QCM).

Spark avec PySpark en Python

Description

L'environnement Apache Spark est aujourd'hui central dans l'approche big data de la donnée. Cette formation spark avec PySpark en python vous permet de maîtriser les principes de l'environnement Apache Spark et l'utilisation du package PySpark pour gérer des données, appliquer des algorithmes de machine learning ou accélérer vos processus.
Cette formation Spark s'adresse à tous ceux qui veulent manipuler Apache Spark en utilisant le langage python.

Trois jours intensifs basés sur des applications réelles pour la préparation, le traitement et l'analyse des données dans l'environnement Apache Spark.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-spark-avec-python/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data scientist désirant monter en compétence sur l'utilisation d'Apache Spark
- Développeur étant amené à automatiser des traitements de données massives

Prérequis :

- Connaissances de base en traitement de données (statistique et tables de données)
- Une connaissance de Python est fortement conseillée.

Objectifs pédagogiques

- Comprendre l'environnement Apache Spark
- Savoir utiliser le package PySpark pour communiquer avec Spark
- Maîtriser l'utilisation de Spark SQL
- Maîtriser l'utilisation de Spark.ml

Programme détaillé

- Rappels sur Python et la manipulation des données
- Introduction à l'environnement Big Data et à Spark
    - Pour qui ? Pour quoi faire ? Comment ?
    - Comment installer Apache Spark

- Pyspark un package Python pour gérer votre environnement Apache Spark
- Quelle infrastructure pour utiliser Spark en entreprise ?
- Les principes de l'environnement : RDD, DataFrame, DataSet…

- Installation de Spark :
  - Sur une infrastructure distribuée
  - En local
  - En cloud (exemples avec Amazon AWS et Microsoft Azure)
- Spark pour la manipulation des données
  - Utilisation de SparkSQL et des DataFrames pour manipuler des données
  - Charger des données depuis Hadoop, depuis des fihiers csv…
  - Transformer des données (création de DataFrames, ajout de colonnes, filtres…)
  - Cas pratiques de chargement et de modifications de données avec Spark et PySpark
- L'utilisation de spark.ml pour le machine learning
  - Apprentissage supervisé : Forêts aléatoires avec Spark
  - Mise en place d'un outil de recommandation
  - Traitement de données textuelles
  - Automatiser vos analyses avec des pipelines
- Introduction et utilisation de Spark Streaming avec PySpark

Deep Learning avec Python

Description

Le Deep Learning est au centre de la révolution de l'Intelligence Artificielle. Ces méthodes d'apprentissage profond vous permettront de résoudre des problèmes jusque-là difficiles à traiter avec des approches classiques de Machine Learning.

Cette formation Deep Learning avec Python vous permet de découvrir et de pratiquer la mise en place de réseaux de neurones profonds. La formation Deep Learning est basé sur des exemples concrets d'utilisation du Deep Learning avec du code en Python.

Si vous avez des bases en Python et que vous voulez aller plus loin avec l'utilisation du Deep Learning, cette formation est faite pour vous.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-deep-learning-avec-python/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- data scientist
- data analyst ayant déjà pratiqué python

Prérequis :

- Avoir suivi la formation Python pour la data science ou avoir de bonnes connaissances en analyse de données et en Python.

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.

Programme détaillé

- A quel problème répond le Deep Learning ?
- Quel est le cadre théorique ?
    - Les réseaux de neurones
    - Le perceptron multicouches
    - Présentation des autres types de réseaux
- Quelle infrastructure et quels outils ?
    - La complexité des modèles

- L'utilisation du GPU
- Les environnements Deep Learning et leurs API Python (TensorFlow, Keras, PyTorch)
- Et le passage en production ?

- Rappel sur la gestion des données avec Python
  - NumPy
  - Pandas
- Un premier réseau de neurones avec Keras
- Les réseaux de neurones à convolutions (CNN)
  - Principes
  - Traitement d'une base d'image
  - Mise en œuvre avec Keras, TensorFlow et PyTorch
- Comment sélectionner mon réseau ?
- Le Transfer Learning
- La génération de features
- Les réseaux de neurones récurrents (RNN)
  - Principes
  - Mise en œuvre pour le traitement de données textuelles
- Passage en production d'un algorithme de Deep Learning
  - La persistance du modèle
  - Création d'une API avec Keras

Analyse textuelle avec Python

Description

Pour les utilisateurs de Python en data science, le traitement automatique du langage (TAL / NLP) est un domaine de premier plan.

Cette formation analyse textuelle avec python se décompose en deux parties : une présentation des principes du NLP et des outils de traitement des chaînes de caractères avec python et une étude plus approfondie des packages spécialisés avec notamment nltk et les packages tels que Spacy.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-analyse-textuelle-python/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data scientist désirant s'initier au traitement des données textuelles

Prérequis :

- Connaissance de Python
- Ou ayant suivi la formation Python pour la Data science

Objectifs pédagogiques

- Maîtrise des 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

Programme détaillé

- Les principes de l'analyse textuelle
  - Qu'est-ce que le traitement du langage naturel (NLP) ?
  - A quelles questions répond le NLP ?
- Les outils de python pour le NLP
  - Le travail sur les chaînes de caractères avec python (string, expressions régulières, transformation vectorisées dans des DataFrame)
  - Les API pour récupérer des données textuelles (Réseaux sociaux, Web scrapping, Open Data)

- La préparation des données textuelles en fonction de la problématique (Données d'enquêtes, Automatisation de processus)
- Les descripteurs en NLP : BOW, TF-IDF, Word Embedding

- Le NLP avec nltk et spaCy
  - Les packages nltk et spaCy : leurs principes et leurs différences
- Combiner analyse textuelle et algorithmes de machine learning
  - L'utilisation de nltk et du machine learning avec scikit-learn pour l'analyse des sentiments
  - Identification de thématiques à partir de textes (topic modeling) en utilisant le package gensim
- L'utilisation du deep learning pour le NLP
  - Les possibilités pour le traitement du langage avec le deep learning
  - Les réseaux de neurones récurrents pour le NLP

Scikit-learn pour le machine learning

Description

Une formation scikit-learn pour le machine learning vous permettant de maîtriser l'utilisation du package scikit-learn, référence mondiale pour le machine learning en python. Par des cas pratiques développés dans des notebooks, vous découvrirez toutes les spécificités de cet outil complet pour la construction de modèles de machine learning.
Cette formation de deux jours convient parfaitement à la formation d'utilisateur de python désirant se perfectionner à l'utilisation de scikit-learn.

Cette formation scikit-learn pour le machine learning vous permet d'approfondir l'utilisation et de maitriser les bonnes pratiques du package scikit-learn.

Formation en petits groupes (maximum 7 participants) pour assurer un échange optimal avec le formateur.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst ou data scientist désirant monter en compétence sur scikit-learn

Prérequis :

- Bases en python et en traitement de données ou avoir suivi la session Python pour la data science

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 machine learning (de la préparation à la validation)

Programme détaillé

- Introduction
  - Les concepts du machine learning
  - Le processus de construction d'un modèle de machine learning
  - L'écosystème de développement autour de scikit-learn (python, numpy, pandas...)
- Utilisation de Scikit-Learn
  - Les principes de scikit-learn pour les modèles de machine learning
  - La notion de pipeline en machine learning

    - Un premier modèle de machine learning avec scikit-learn
- Préparation des données et preprocessing
    - Comment préparer des données pour un modèle de machine learning ?
    - Les preprocessing de scikit-learn
    - Le traitement des données qualitatives
    - Le traitement des données manquantes
    - L'utilisation du ColumnTransformer de scikit-learn
- Sélection de meilleur modèle avec scikit-learn
    - Gérer le sur-apprentissage
    - Validation du modèle (train / test, validation croisée...)
    - Les métriques d'évaluation des modèles de machine learning
- L'ajustement des hyperparamètres avec scikit-learn
    - Gestion automatique
    - Combinaison dans un pipeline
- Focus sur certaines méthodes de machine learning
    - Modèles linéaires
    - Modèles basés sur des arbres (forets aléatoires, GBM...)
    - Autre modèles
- Le cas des modèles non supervisés dans scikit-learn
    - Mise en œuvre de modèle de machine learning non supervisés avec scikitlearn et comparaisons.
- Passage en production d'un modèle de machine learning avec scikit-learn - les différentes approches.

Organisation de la formation

Moyens pédagogiques et techniques

- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation.
- Utilisation de Notebooks Jupyter et d'un environnement JupyterHub pour les exercices pratiques.

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Mises en situation.
- Questions orales ou écrites (QCM).
- Exercices effectués dans des notebooks

Industrialisation d'un Projet Data en Python

Description

Découvrez comment passer en production un projet data avec cette formation industrialisation de projets data en python.

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst et data scientist désirant acquérir des compétences en industrialisation
- Développeur Python souhaitant accompagner la mise en production de modèles

Prérequis :

- Connaissances de base en traitement de données (statistique et tables de données)
- Connaissances de base du langage Python

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

Programme détaillé

- Qu'est-ce que l'industrialisation ?
    - Quelles sont les grandes thématiques de l'industrialisation d'un projet ?
    - Pourquoi et pour qui industrialiser ?
- Le version control, et les outils qui l'entourent (Gitlab, Github)
- Découverte des principes de clean code
    - Application des différentes règles dans le contexte d'une MR
    - Écriture d'un code en suivant des standards de qualité élevés
    - Application des principes de clean code aux notebooks
- Découverte des outils de vérification statique
    - Utilisation d'un linter
    - Utilisation du typage statique
- Les différents types de tests, et leur application au contexte de la data science
    - Écriture d'un test unitaire, intégration et e2e
    - Application des principes du testing à la thématique ML

    - Écriture d'un pipeline CI / CD lançant plusieurs types de test
- Les outils de documentation et leur principe
    - Documentation automatique du code avec Sphinx
    - La documentation comme substitut à la revue scientifique à partir d'outils collaboratifs tels que Slite
- Découverte du processus de recette d'un algorithme
- Mise en production d'un algorithme
    - Rédaction et application d'un processus de recette d'un algorithme
    - Description du principe d'itération
- Suivi de la qualité et monitoring des performances d'un algorithme
- Mise en production et dockerisation d'un projet

Organisation de la formation

Moyens pédagogiques et techniques

- Supports sous forme de Jupyter Notebooks
- Nombreux cas pratiques

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Évaluation par des exercices par le formateur
- Quiz en classe

Data visualisation avec Python

Description

Vous connaissez déjà l'environnement Python pour la data science et vous désirez vous perfectionner en visualisation de données avec Python ?

Aujourd'hui de nombreux packages sont disponibles, nous en ferons un panorama avec des applications pratiques. Cette formation data visualisation avec Python se décompose en deux parties. Un approfondissement des principes et des représentations avec Matplotlib et Seaborn (graphiques animés, graphiques complexes, graphiques statistiques). La construction de graphiques interactifs et d'applications web de data visualisation avec Plotly, Dash et Streamlit.

Cette formation est organisée en petit groupes (maximum 6 participants) afin de maximiser les échanges avec le formateur et entre participants.

PLUS D'INFORMATIONS :
Formation disponible en intra ou en inter-entreprises
Inscription : https://www.stat4decision.com/fr/formations/formation-data-visualisation-avec-python/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst désirant approfondir son utilisation de la dataviz en python
- Développeur web désirant construire des graphiques interactifs en python

Prérequis :

- Connaissance du langage python et des bases de la manipulation de données avec python (numpy / pandas) ou ayant suivi la formation python pour la data science.

Objectifs pédagogiques

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

Programme détaillé

- Les bonnes pratiques pour la data visualisation
    - Quelle visualisation choisir en fonction du contexte
    - Pièges à éviter pour visualiser des données et des résultats
- Les graphiques avec Python
    - -> Le package Matplotlib pour les graphiques
    - Panorama des graphiques disponibles
    - Gérer les options de vos graphiques
    - Le nuage de points
    - Les graphiques en bâtons (bar plot)
    - Les countour plot
    - Les histogrammes
    - Les graphiques pour représenter des données en 3 dimensions
    - -> Utiliser Seaborn pour obtenir des graphiques scientifiques
    - Le box plot ou la boîte à moustache
    - Le pair plot
    - Les matrices de graphiques
    - -> Ajouter des informations à vos graphiques :
    - Ajout de texte / annotations
    - Les légendes
    - Les colorbar
    - -> Créer des graphiques animés avec Python
    - Générer une vidéo à partir d'un graphique et Matplotlib
    - -> Afficher des données massives sous forme de graphiques
    - -> Quelques bases de cartographie
    - Utilisation de cartopy combiné à Matplotlib
    - Cartes interactives avec Folium
- Les graphiques interactifs et les applications web avec Python :
    - Le package Plotly et ses principes
    - Création de graphiques web interactifs avec Plotly
    - Les applications web avec Stremlit et Dash
    - Principe
    - Rapport au package shiny de R
    - Construction d'un tableau de bord pour visualiser vos données
    - Construction d'une application de reporting avec Streamlit ou Dash
    - Principes pour le passage en production

Organisation de la formation

Moyens pédagogiques et techniques

- Accueil des stagiaires dans une salle dédiée à la formation.
- Documents supports de formation projetés.

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Mises en situation.
- Cas pratiques validés par le formateur.

- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation.
- Outils : Anaconda et python 3 ainsi que toutes les bibliothèques pour la data visualisation
- Évaluation des connaissances (quizz / tests).

Dash pour la construction de dashboard avec python

Description

Une formation pour utilisateur de python désirant construire des dashboard interactifs avec l'outil Dash de Plotly. Des bases de son utilisation à la mise en production d'applications web.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst
- Data scientist

Prérequis :

- Connaissance de base en python

Objectifs pédagogiques

- Construire une application Dash
- Déployer une application Dash

Programme détaillé

- Introduction
  - Préparation de données avec pandas et numpy
  - Présentation de Plotly pour la dataviz
  - Présentation de Dash pour les dashboards
- Les bases de Plotly
  - Principes généraux
  - Les graphiques classiques (nuage de point, bar chart, histogrammes...)
  - Des graphiques plus avancés (DistPlots HeatMaps...)
- Les bases de Dash
  - Les layout d'un dashboard Dash
  - Comment appliquer un style à votre dashboard
  - Du graphique au dashboard
- Les composantes de Dash
  - Les composantes "figées" (HTML, Markdown...)
  - Les composantes interactives (callbacks pour l'interactivité et les graphiques, entrées et sorties multiples
  - Aller plus loin sur les callbacks
  - Aller plus loin sur les interactions avec les données

- Déploiement d'applications Dash
    - Déploiement en local
    - Déploiement avec Dash Entreprise
    - Déploiement avec Flask

Organisation de la formation

Moyens pédagogiques et techniques

- Etude de cas concrets
- Documents supports de formation projetés.
- Exposés théoriques
- Utilisation de VS Code pour le développement

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Evaluations des exercices par le formateur
- Mises en situation.

FORMATIONS R

R pour la data science

Description

R est aujourd'hui le logiciel libre de référence pour traiter des données. Cette formation de trois jours vous ouvre les portes du langage R et de toutes les subtilités de son utilisation pour la science des données.
Vous sortirez de cette formation avec une maîtrise de l'utilisation et des principes de base du logiciel R, référence absolue en analyse des données.

Cette formation s'adresse aussi bien à des utilisateurs d'autres outils (SAS, Matlab, Python) voulant s'initier à R qu'à des utilisateurs métiers désirant découvrir ce logiciel.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-logiciel-r/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Cette formation s'adresse aussi bien à des utilisateurs d'autres outils (SAS, matlab, python) voulant s'initier à R , qu'à des utilisateurs métiers désirant découvrir ce logiciel.
- Data analyst désirant découvrir R
- Manager désirant comprendre les spécificités de R
- Développeur spécialisé dans d'autres langages

Prérequis :

- Aucun prérequis sur l'utilisation de R n'est nécessaire.
- Connaissances de base en traitement de données (statistique et tables de données)

Objectifs pédagogiques

- Maîtriser l'environnement de travail de R avec Rstudio
- Manipuler et gérer des données avec R et les packages de data manipulation les plus récents (dplyr, data.table…)
- Maîtriser l'application de méthodes de data science
- Connaitre les bases de la programmation avec le langage R

Programme détaillé

- Le langage R et son environnement :

- Définition de l'environnement
- Présentation de la notion de packages
- Présentation de RStudio
- Les possibilités de R pour la création de documents (scripts, Rmarkdown, Rnotebook…)
- Présentation des packages importants pour la manipulation de données

- Les structures et les fonctionnalités de base de R :
  - L'importation des données (.txt, .csv, .xls, .sav…)
  - Les fonctionnalités de base pour la gestion des données
- Utilisation avancée de la gestion des données avec R :
  - La grammaire de dplyr
- Statistiques descriptives avec R
  - Les fonctionnalités pour la description statistique des données
  - La gestion des graphiques (histogrammes, box-plot…)
- Le data mining et le machine learning avec R
  - L'analyse descriptive des données multivariées et la réduction de dimensions (ACP) : Présentation et utilisation du package FactoMineR
  - Apprentissage supervisé – cas pratiques avec R (la régression logistique, les arbres de décisions, Les forêts aléatoires et les GBM, Présentation du package caret pour ajuster les hyper-paramètres)
  - L'apprentissage non supervisé – cas pratiques avec R (k-means…)
- Quelques notions de programmation avec R
  - Les standards de la programmation avec R
  - Construire une fonction R pour automatiser des analyses
  - Les structures de contrôle : if, for, while…

Logiciel R pour utilisateur de SAS

Description

Vous êtes utilisateur de SAS et habitué à utiliser le langage SAS au quotidien, vous voulez passer à R pour vos traitements de données ou utiliser R en supplément de SAS. Notre formation logiciel R pour utilisateur de SAS vous permet de comprendre les points communs, les forces et faiblesses de ces outils.

Par des exemples comparés, nous vous accompagnons dans une transition douce vers R depuis SAS.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-r-utilisateur-sas/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Utilisateurs de SAS
- Statisticiens
- Analystes

Prérequis :

- Connaissance de base de SAS
- Aucun prérequis sur l'utilisation de R

Objectifs pédagogiques

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

Programme détaillé

- Introduction
  - R et son écosystème
  - Comparaison des logiciels R et SAS
  - Le logiciel R, un investissement pour l'avenir
- Les outils R
  - RStudio
  - Les packages
  - Le tidyverse
- La syntaxe R vs la syntaxe SAS

- La préparation et le nettoyage des données
    - Charger des données, les procédures SAS et les fonctions R
    - Manipuler des données comparaison étape data de SAS et le tidyverse (dplyr)
- Statistiques et reporting
    - Les statistiques descriptives – comparaison des procédures SAS et des fonctions R
    - Construire des graphiques / tableaux de bord avec R, ggplot2 et shiny
- Analyse de données et machine learning
    - Comparaison de SAS et de R sur des modélisations simples
    - Aller plus loin avec le machine learning et R
- Discussions et démonstrations des différences sur les problématiques des participants

Visualisation avec R

Description

En deux journées, maîtrisez les principes et les outils de la data visualisation avec le langage R. Cette formation visualisation avec R se concentre sur les outils de R et sur le principal package pour la data visualisation : ggplot2. Si vous êtes un utilisateur de R et que vous voulez aller plus loin dans la visualisation des données, cette formation est pour vous !

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-visualisation-avec-r/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst / statisticien désirant découvrir la data visualisation avec R

Prérequis :

- Public ayant des bases pour l'utilisation du logiciel R ou ayant suivi une formation R pour la data science

Objectifs pédagogiques

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

Programme détaillé

- La data visualisation
  - Principes de base pour une présentation efficace
  - Différents types de graphiques pour représenter différentes variables
  - Efficacité d'un graphique
- Le langage R pour représenter des données
  - Le package ggplot2 : Principes, Exporter un graphique, sauvegarde et recyclage
  - La fonction qplot : Syntaxe de base, Diagramme en bâtons, histogramme, boxplot, nuage de points, courbe
  - La fonction ggplot et ses compléments : Les graphiques, Les données, Traitement statistique
- Aller plus loin pour des représentations avancées avec R

Développement d'applications web avec R shiny

Description

Vous voulez développer à partir de vos données des applications web interactives et attractives sans être un développeur web ? Vos données méritent mieux que des représentations figées ?
La formation développement d'applications web avec R shiny est faite pour vous.

Shiny est un package de R qui vous simplifie la vie et vous permettre de représenter vos données sous forme d'applications web simples à mettre en œuvre.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst
- Data scientist
- Statisticien

Prérequis :

- Connaissance de base en R ou avoir suivi une formation R pour la data science

Objectifs pédagogiques

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

Programme détaillé

- Introduction
  - Découverte et installation
  - Dépendances
  - Quelques exemples
- Principes de base
  - Les fichiers (ui.R / server.R)
  - Lancement d'applications
  - Le fichier global.R
  - Mise en pratique
- La structure des fichiers de votre application
  - ui.R
  - server.R
  - global.R

- Apparence de votre application
  - Les panels (wellPanel, ConditionalPanel., tab panel..) avec mise en pratique
  - Les layouts (fluidRow, flowLayout, FluidPage,...) avec mise en pratique
- Les widgets
  - Les différents widgets de shiny (choix simples ou multiples, valeurs numériques, textes, déclencheurs...) avec mise en pratique
  - Aller plus loin avec shinyWidgets
- Les sorties (outputs)
  - Utilisation de la fonction render... (renderPlot, renderUI...) avec mise en pratique
  - Le cas renderUI en particulier
  - Le cas renderImage en particulier
  - L'utilisation du package DT pour l'affichage des tables
- La réactivité de votre application
  - Concepts
  - Les fonctions observe() et observeEvent() et leur mise en pratique
  - Les fonctions reactive(), eventReactive(), isolate()...
- Aller plus loin dans vos applications shiny
  - Personnalisation (html, css)
  - Utilisation de shinythemes
  - D'autres packages pour personnaliser vos applications
- Partager votre application shiny
  - Déploiement avec shinyapps.io
  - Créer un serveur shiny
  - Autres modes de déploiement
- Discussion et mise en pratique sur les données des participants
  - Apportez vos données et construisez votre application shiny

Data mining et machine learning avec R

Description

Vous êtes analyste, vous voulez compléter vos compétences afin de devenir data scientist, vous vous intéressez au langage R ?
Cette formation data mining et machine learning avec R de 2 jours est faite pour vous ! Maîtrisez l'utilisation du langage R et les principales méthodes d'analyse de données par des applications pratiques.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-data-mining-avec-r/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Analyste
- Statisticien

Prérequis :

- Utilisateurs occasionnels de R
- Connaissances de base en analyse de données

Objectifs pédagogiques

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

Programme détaillé

- Le data mining et le machine learning – qu'est-ce que c'est ? (Panorama des méthodes et des applications)
- Le langage R pour le data mining et la machine learning
- Le langage R pour le data mining
- L'analyse descriptive des données multivariées et la réduction de dimensions :
    - ACP/ACM/AFC
    - Présentation et utilisation du package FactoMineR
- Apprentissage supervisé – cas pratiques avec R
    - La régression logistique
    - Les arbres de décisions
    - Les forets aléatoires et les GBM

- Utilisation du package caret pour ajuster les hyper-paramètres
- L'apprentissage non supervisé – cas pratiques avec R (k-means…)
- Discussions et évaluation de la formation

R pour la Cartographie et les SIG

Description

Une formation R pour la Cartographie vous permet de maîtriser les outils modernes de manipulation, de traitement et de restitution graphiques de données spatiales dans le logiciel R.

Vous saurez importer, manipuler et traiter des données géographiques au format vectoriel, réaliser différentes opérations de géomatique.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-cartographie-sig-r/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst
- Statisticiens
- Cartographes

Prérequis :

- Avoir suivi une initiation au logiciel R ou de savoir manipuler les différents objets R avec les outils du {tidyverse}.

Objectifs pédagogiques

- Importer, manipuler et traiter des données géographiques au format vectoriel
- Réaliser différentes opérations de géomatique telles que les jointures spatiales, les zones tampons ou le calcul d'indices topographiques
- Produire des cartes légendées, annotées et interactives

Programme détaillé

- INTRODUCTION A LA CARTOGRAPHIE
  - Présentation des SIG
  - Présentation des formats des objets géographiques dans R
  - Importation et affichage de cartes existantes (shapefiles, raster)
  - Projections de cartes
- MANIPULATION D'OBJETS VECTORIELS
  - Importation de tableau de données de points spatialisés
  - Jointure spatiale points – polygones
  - Manipulation des tables attributaires

- Sélection de données
- MANIPULATION AVANCÉE DES OBJETS VECTORIELS
  - Création d'objets spatiaux (polygones, polylignes, points)
  - Création de zones tampon
  - Calculs de distances
  - Connexion à une base de données (PostGIS, GeoJSON)
- VISUALISATION STATIQUE ET DYNAMIQUE
  - Mise en forme de cartes pour la recherche reproductible
  - Cartes dynamiques avec leaflet

Séries temporelles avec R

Description

Deux jours pour maîtriser les séries temporelles ou séries chronologiques avec le langage R.

Au cours de cette formation, les notions de base liées aux séries temporelles sont abordées et les méthodes d'analyse de données temporelles sont appliquées sur des données réelles.

L'objectif étant la maîtrise du langage R pour le traitement, l'analyse et la prévision sur des données temporelles.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-series-temporelles-r/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Statisticien
- Analyste
- Economiste

Prérequis :

- Connaissances de base en R (avoir suivi la formation R pour la data science ou niveau équivalent)
- Connaissances de base en traitement de données (statistique et tables de données)

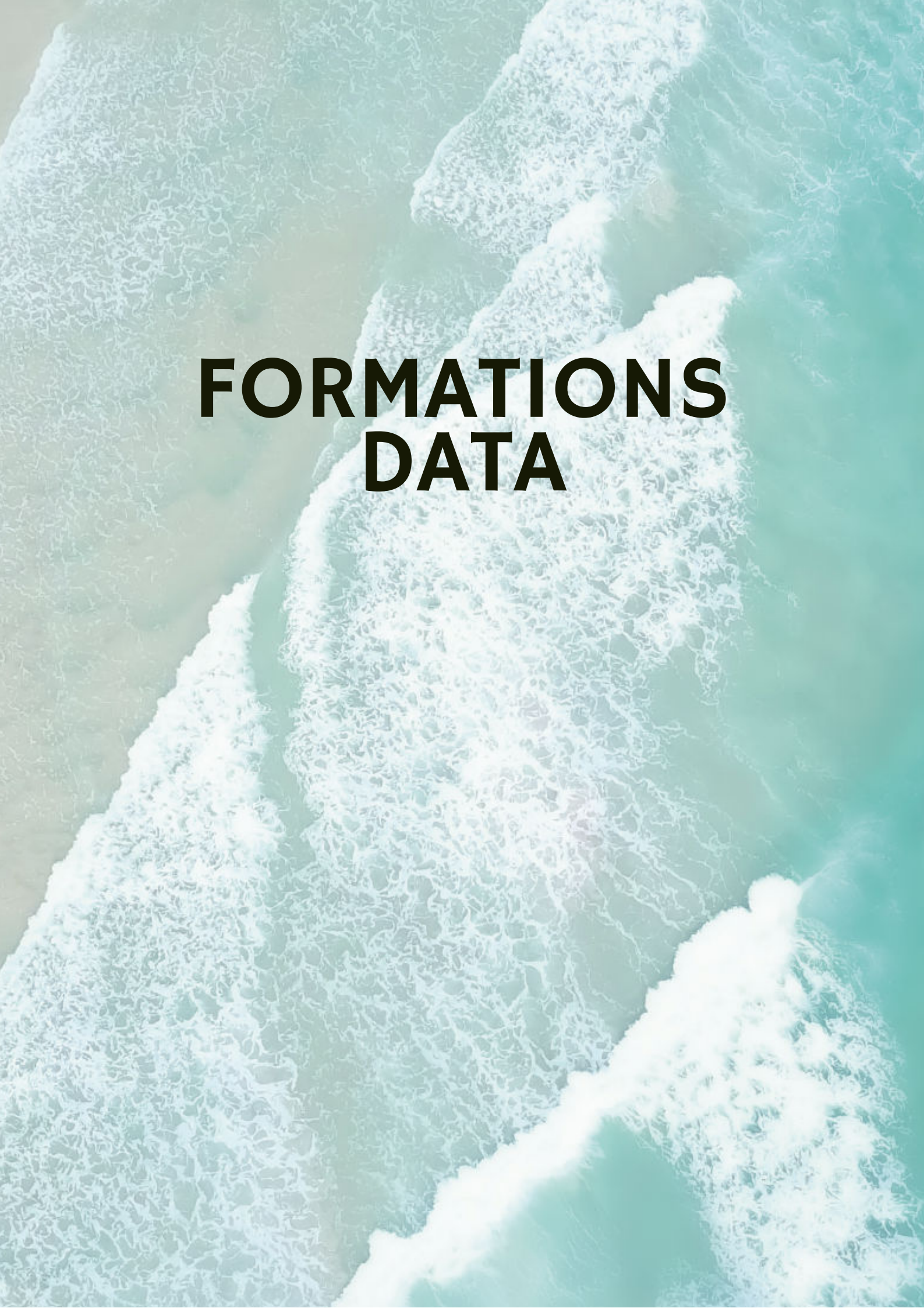
Objectifs pédagogiques

- Maîtriser les principes du traitement des séries temporelles
- Définir et représenter des séries temporelles sous R
- Savoir quelle méthode utiliser en fonction de votre problématique
- Utiliser R pour appliquer les méthodes de description et de prévision de séries temporelles
- Lire et interpréter les résultats obtenus

Programme détaillé

- Introduction aux séries temporelles
- Les formats dates et time dans R
- Les objets séries temporelles dans R et leurs packages associés
- La description d'une série temporelle
- Les méthodes de filtrage et de lissage

- De la description à la prévision : les modèles ARIMA
- Applications sur des données réelles
    - Les participants sont invités à apporter leurs données s'ils en ont.



FORMATIONS DATA

Les fondamentaux du machine learning et du deep learning

Description

Que vous soyez analyste BI, chef de projet, développeur ou juste intéressé par le traitement des données, cette formation les fondamentaux du machine learning et du deep learning est faite pour vous.

Par des applications et des présentations de use cases, vous découvrez les modèles utilisés en machine learning (apprentissage automatique) ainsi qu'en deep learning pour l'intelligence artificielle.

Que ce soit pour de la classification, du clustering, du traitement d'images, de données textuelles ou d'autres applications, cette formation machine learning et deep learning vous fournit tout le bagage nécessaire à la compréhension de ce nouveau domaine passionnant.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-machine-learning/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Développeur
- Chef de projet
- Analyste BI
- Chargé d'études

Prérequis :

- Pas de pré-requis (le langage python sera utilisé avec des librairies open-source exclusivement. Aucune connaissance préalable de python n'est demandée.)

Objectifs pédagogiques

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

Programme détaillé

- Introduction au Machine Learning
    - Lien entre Big Data et Machine Learning
    - Du machine learning au deep learning pour des applications spécifiques
    - Découverte de l'éco-système Python en cloud
- Les problématiques de machine learning
    - D' un problème métier à un problème de machine learning
- Feature engineering
    - Apprendre à réduire la complexité d'un problème pour le résoudre
    - Data augmentation : création de nouvelles variables pour aider à résoudre le problème
- Classification par catégories
    - Arbre de décisions et forêts aléatoires
    - Régression logistique
    - Analyse de données non structurées : classification de textes
    - Analyse de sentiments
    - Mesures de performance d'une classification
- Régression : détection de tendances et prévisions
- Clustering : regroupement automatique par familles
- Le deep learning
    - Les réseaux de neurones
    - Les principes
    - Traitement des images avec le deep learning : les réseaux de neurones à convolution
    - Traitement des données textuelles avec le deep learning : les réseaux de neurones récurrents
- Déployer un modèle en production
    - Mise en pratique

Les fondamentaux du big data et de la data science

Description

Cette formation comprendre les fondamentaux du big data et de la data science est basée sur des cas pratiques afin de vous familiariser avec les concepts du big data et de la data science.
Vous apprendrez par des exemples à comprendre l'environnement du big data (Hadoop, Spark…) et ses applications (open data, internet des objets…).

Cette formation big data vise un public d'analystes, de chargés d'études voulant comprendre les enjeux liés au big data et ne demande aucun prérequis techniques.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-introduction-au-big-data/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Analystes
- Chargés d'études
- Tout public intéressé par la compréhension des fondamentaux du big data et de la data science

Prérequis :

- Tout public

Objectifs pédagogiques

- Comprendre les principes du big data
- Maîtriser le vocabulaire lié au big data et à la data science
- Maîtriser les enjeux pour mettre en œuvre une stratégie big data dans votre entreprise

Programme détaillé

- Le big data :
  - Le big data – définitions et usages
  - Quelques cas d'usages de projets big data
  - Qui pour s'occuper des données dans votre structure : le data scientist, le data analyst, le data engineer, le big data architect…
- Les données et leur gestion :
  - Les données non structurées

- Le passage au big data : quels changements pour votre organisation ?
- Le data lake et les formats de stockage (HDFS, in memory…), quelle solution choisir ?
- Hadoop et son environnement (Hive, Pig, MapReduce, Kafka…)

- Les étapes d'un projet big data :
  - Les données
  - Le stockage
  - Les outils liés au big data : Hadoop, Spark, Kafka…
- La data analytics (predictive analytics) et le machine learning pour l'intelligence artificielle :
  - Différence entre le traitement des données classiques et la data analytics pour le big data
  - Le machine learning : panorama des méthodes et des utilisations (supervisé, non supervisé, renforcement, deep learning)
  - Présentation d'outils pour le machine learning dans le cadre du big data (MLlib de Spark)
- Les objectifs d'une stratégie big data :
  - Comment définir une stratégie big data dans votre entreprise.
  - Présentation de cas réels d'applications big data
  - Comment éviter les pièges liés à un projet big data

Intelligence artificielle - les fondamentaux

Description

L'intelligence artificielle est au cœur de tous les sujets.
Elle représente aujourd'hui un point critique dans toutes les entreprises.
Cette formation sur une journée vous permet de comprendre les tenants et les aboutissants de l'intelligence artificielle dans le cadre de l'entreprise.

Cette formation intelligence artificielle vous permet de construire votre regard critique sur l'intelligence artificielle.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-intelligence-artificielle/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Manager
- Responsable data
- Chef de projet
- Analyste BI
- Chef de produit
- Chargé d'études

Prérequis :

- Aucun prérequis technique

Objectifs pédagogiques

- Comprendre les principes de l'intelligence artificielle
- Être capable de reconnaître l'utilisation d'intelligence artificielle
- Connaître les différents cas d'usage de l'intelligence artificielle

Programme détaillé

- Introduction à l'intelligence artificielle
  - Définitions et principes de l'intelligence artificielle
  - Les concepts de l'intelligence artificielle : Big data, machine learning et deep learning
  - Les outils utilisés en intelligence artificielle
- L'intelligence artificielle par des cas d'usage

- Les 3 principales applications de l'intelligence artificielle : la vision par ordinateur, le traitement automatique du langage et les "robots intelligents"
- L'intelligence artificielle en entreprise (bot, assistants, maintenance prédictive…)
- L'intelligence artificielle dans votre quotidien
- L'internet des objets et l'intelligence artificielle
- Les LLM et l'intelligence artificielle

- Construction d'un projet d'intelligence artificielle
  - Répartition des projets et construction d'un projet de mise en œuvre d'une intelligence artificielle (étapes, planning, ressources nécessaires, objectifs, retours sur investissement)
  - Présentation et compte-rendu des projets des participants
- Les enjeux et le futur de l'intelligence artificielle
  - Les biais des algorithmes et des données par des exemples
  - Les intelligences artificielles de demain
  - Discussions sur les opportunités liés à l'intelligence artificielle

Open Data

Description

Cette formation vous permet d'appréhender le contexte et les opportunités d'une politique Open Data.
La formation open data vous permet d'appréhender les outils pour la mise à disposition des données et leur réutilisation.
Des cas pratiques seront présentés avec les différents outils du marché permettant de publier vos données.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-open-data/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Directeur des Systèmes d'Information
- Chargé(e) de mission Open Data
- Chef de projet en charge de l'Open Data
- DSI ; responsable SI
- Toute personne impliquée dans un projet d'ouverture de données ou d'utilisation de données ouvertes.

Prérequis :

- Pas de pré-requis

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

Programme détaillé

- Comprendre l'open data
  - Origines historiques de l'ouverture des données
  - Les grands principes de l'ouverture des données
  - Aperçu des principaux acteurs politiques, économiques et associatifs de l'open data
  - Cadre juridique de l'ouverture des données
  - Open government et Etat plateforme : l'ouverture des données dans la modernisation de l'Etat

- Les formats de données
- Tour d'horizon des portails d'open data (CKAN, Open Data Soft, data.gouv.fr…)
- Les success stories de l'open data

- Ouvrir des données
  - La conception d'une stratégie d'open data : principaux leviers politiques, organisationnels, économiques
  - L'identification des données : de la tournée des bureaux à l'inventaire systématique
  - Open Data Census : exercice d'identification de données à ouvrir
  - Négociations et frictions dans l'ouverture des données : comment éviter les obstacles de l'open data ?
  - Extraction des données et évolution des systèmes d'information
  - Conversion et mise en qualité des données
  - Les standards émergents de l'open data
  - Publication des données et production de métadonnées
  - Animation de l'open data et incitations à la réutilisation de données
  - Gouvernance des données et gestion de la relation avec les réutilisateurs

Data science – bonnes pratiques et outils

Description

Une journée pour échanger sur les bonnes pratiques et les outils de la data science. La data science est aujourd'hui centrale dans les entreprises digitalisées.
Cette formation data science d'un jour est destinée à des décideurs ou à des data scientists désirant connaître les bonnes pratiques ainsi que les outils à utiliser en data science.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-data-science-bonnes-pratiques-et-outils/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Responsable data science
- Chief data officer
- Data scientist

Prérequis :

- Pas de pré-requis

Objectifs pédagogiques

- Eviter les pièges
- Comprendre les bonnes pratiques
- Connaître l'offre d'outils pour la data science.

Programme détaillé

- Introduction à la data science
- La data gouvernance et ses bonnes pratiques
- Les projets data science
    - Déroulé d'un projet data science
    - Les erreurs à éviter
    - Le passage en production
- Les outils pour la data science
    - L'open source : R, python, julia…
    - Les plateformes et leurs possibilités : KNIME, Rapidminer, DataIku…
- Des cas concrets

- Internet des objets
- Maintenance prédictive
- Détection de fraude
- Moteur de recommandation

Culture générale de la donnée

Description

Rares sont aujourd'hui les entreprises ou les institutions qui n'affichent pas des objectifs ambitieux de transformation de l'organisation par les données, ne tentent pas de se saisir de la « révolution du big data » ou n'essaient pas d'extraire de la valeur du « nouveau pétrole des données ».

Pour enfin comprendre de quoi on parle lorsqu'on évoque le « big data », la « data science », le « machine learning », le « deep learning » ou « l'open data », cette formation offre un panorama large des concepts clés et des bonnes pratiques de l'exploitation et de la mise en circulation des données.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-culture-generale-donnees/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Public qui souhaite approfondir ses connaissances des notions et des bonnes pratiques d'exploitation et de mise en circulation des données.
- Membre d'une équipe : Innovation Digital
- Membre d'une équipe : DSI
- Membre d'une équipe : Stratégie
- Membre d'une équipe : Marketing
- Membre d'une équipe : Ressources humaines

Prérequis :

- Pas de pré-requis

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

Programme détaillé

- Comprendre l'écosystème des données
    - Workshop 1 — Qu'est-ce qu'une donnée ? Exercice d'identification, définition et étymologie du terme

- Données, données… quelles données ? Quantitatives, qualitatives, structurées, non structurées, capturées, échappées, dérivées…
- Petite histoire sociale des données et de leur exploitation
- Datafication : les promesses de la mise en données du monde
- Workshop 2 — regards critiques sur la révolution des données – Limites et critiques académiques, politiques, sociales, techniques et organisationnelles pour mieux appréhender la question dans votre organisation
- Open data, big data, linked data, data API : les habits neufs de la donnée
- Cadre juridique et protection de la vie privée
- Modèles économiques de l'open data et du big data
- Workshop 3 — analyse du modèle économique d'entreprises de la data

- Ouvrir la boite noire de la data science
  - Qu'est-ce que la data science ?
  - Workshop 4 – quelles compétences recruter dans un projet de data science ?
  - L'écosystème matériel, logiciel, informationnel
  - Le data pipeline : acquérir les données, les vérifier, les nettoyer, les analyser, les présenter
  - Workshop 5 – data visualisation : quel format et quel outil choisir pour quel fonction ?
  - Modéliser pour analyser : introduction au machine learning et deep learning
  - Workshop 6 – décortiquons quelques algorithmes notoires

Atelier acculturation data

Description

Un atelier d'une journée destiné à un public large afin d'appréhender les spécificités de la data.
Notre atelier acculturation data est basé sur des échanges en petits groupes – mini-projets, quiz, échanges de retour d'expériences – et permet d'appréhender les défis liés aux données de manière simple et ludique.
Nous pouvons vous proposer ces ateliers dans votre structure afin d'offrir une acculturation data au plus grand nombre. Dans ce cas, le format est adaptable (en durée et en contenu). Nos ateliers acculturation data sont déployables à grande échelle dans votre structure si vous avez de nombreux participants à former.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/atelier-acculturation-data/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Public ayant peu de connaissances sur les données

Prérequis :

- Aucun prérequis

Objectifs pédagogiques

- Comprendre ce qu'est une donnée
- Etre capable d'identifier des données pour un projet
- Connaître la terminologie liée aux données

Programme détaillé

- Workshop 1 — Qu'est-ce qu'une donnée ? Exercice d'identification, définition et étymologie du terme
- Données, données… quelles données ? Quantitatives, qualitatives, structurées, non structurées, capturées, échappées, dérivées
- Open data, big data, linked data, data API : les habits neufs de la donnée
- Workshop 2 — analyse du modèle économique d'entreprises de la data
- Les nouveaux métiers de la donnée
- Workshop 3 - Comment utiliser les données pour améliorer votre organisation
- Echange sur les pratiques liées aux données

Panorama et choix des briques d'exploitation d'un data lake

Description

Une formation pour comprendre les concepts et les choix existants pour la construction d'un data lake.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Chef de projet DSI désirant faciliter l'exploitation de données déjà présentes dans un data lake
- Manager souhaitant accompagner une maîtrise d'ouvrage technique

Prérequis :

- Une première utilisation de données provenant d'un data lake
- Connaissances de base en traitement de données (statistique et tables de données)

Objectifs pédagogiques

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

Programme détaillé

- Qu'est-ce que le data engineering?
    - Quelle est sa fonction à l'intérieure de l'entreprise
    - Qu'est-ce qu'un data lake
    - Retour sur le rôle des différentes briques qui composent un data lake
- Comprendre les différentes manières de stocker sa donnée, les bases des technologies et les différences d'approches
    - SQL et noSQL
    - HDFS
    - GCS , S3, Azure Data Lake Storage
- Orchestration des transformations et ingestion de la donnée avec un ETL

- Data Processing avec Spark, Flink et Hadoop
- 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 d'une architecture
    - Utilisation d'une RFC
- Monitorer et manager les performances du data lake
- Mise en places de principes de data gouvernance

Organisation de la formation

Moyens pédagogiques et techniques

- Support PDF avec de nombreuses références et cas pratiques
- Utilisation d'une plateforme de test des outils du data lake

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Cas pratiques validés par le formateur
- Quiz en classe

FORMATIONS MISES EN PRATIQUE DE LA DATA VISUALISATION

Datavisualisation

Description

La datavisualisation est aujourd'hui centrale dans la plupart des métiers. Cette formation datavisualisation vous permet de maîtriser la construction de visualisations efficaces et lisibles en appliquant les bonnes pratiques du domaine.
Notre formateur expert vous propose de manipuler et de réfléchir aux bonnes pratiques de data visualisation pour mettre en valeur vos données et l'histoire qu'elles racontent.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Toute personne ayant la nécessité de diffuser des données et souhaitant les communiquer de façon engageante et lisible

Prérequis :

- Niveau initiation, pas de prérequis

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 afin de toucher un public plus large

Programme détaillé

- Introduction, les enjeux de la datavisualisation
    - Définition de la visualisation de données
    - La différence entre l'infographie et la datavisualisation
    - Panorama des différentes utilisations métiers possibles
    - Les ingrédients nécessaires pour bien réussir sa visualisation
    - L'équilibre à respecter entre l'information et l'esthétique
- L'aspect éditorial
    - Identifier son audience
    - Les clés pour s'adapter au public cible
    - Méthodologie et étapes à respecter
    - La théorie de la conception de l'information
    - Démonstration par l'exemple
    - Définir un angle et choisir les informations pertinentes
    - Hiérarchiser et contextualiser l'information
    - Les différents niveaux de lecture

- L'écriture infographique
- Optimiser la lecture pour ne pas perdre le lecteur
  - La sémiologie graphique
  - Mesurer les limites de l'interprétation
  - L'importance du choix des couleurs dans un graphique
  - Connaître les bonnes pratiques à déployer
  - Repérer les erreurs typiques dans un graphique
  - Les spécificités des supports (print, digital)
  - L'utilité d'une charte graphique pour obtenir des graphiques cohérents et homogènes
- Capter l'attention du lecteur
  - Les différents principes de mise en avant de l'information
  - Présenter ses graphiques de façons pertinentes
  - Créer un rythme graphique
  - L'utilisation de gabarits adaptés aux données - Exercice, pas-à-pas
  - Le datastorytelling
  - Les déclinaisons pour toucher encore plus de monde
- Savoir choisir la bonne forme graphique pour les données
  - Les grandes familles de représentations graphiques
  - Les différentes possibilités de mises en forme - Exercice, pas-à-pas
  - Identifier les points forts et trouver des alternatives aux points faibles - Démonstration par l'exemple
  - Tester la lisibilité et la compréhension - Argumenter son choix
  - Discussion collective

Rendre ses présentations claires et engageantes

Description

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Apprenant étant amené à construire des présentations

Prérequis :

- Connaissance de base pour la construction de contenu (type présentation)

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 des données

Programme détaillé

- Intégrer les bonnes pratiques pour présenter clairement ses données
    - contextualiser le data design
    - Appréhender les règles de la sémiologie graphique
    - Méthodologie : - Six principes de mise en avant - Six principes de composition
    - Établir le rapport idéal entre texte et visuel pour être lisible
    - Définir le point d'équilibre entre information et esthétique
    - Faciliter la mémorisation des informations
    - Utiliser correctement les pictogrammes et les illustrations pour contextualiser un sujet
    - Varier les formes des graphes pour donner du rythme
    - Comparer les mises en forme possibles pour « faire parler » les chiffres clés
    - Examiner les principaux gabarits de mise en page adaptés à la mise en scène des données
    - Études de cas
- Prendre en compte les aspects éditoriaux
    - Panorama des outils de mise en page et de présentation
    - Définir un angle
    - Hiérarchiser et contextualiser l'information : - La théorie de la conception de l'information - Les différents niveaux de lecture - Le sens de lecture - L'écriture infographique
    - Analyse de projets remarquables

- Quiz n°1

- Capter l'attention du lecteur
  - Délivrer des présentations cohérentes et homogènes en jouant sur : - la grille (modulaire/multi-colonne) - les lignes de force - les marges comme principe graphique - les couleurs - les typos (importance du choix et de son utilisation) - l'élément perturbateur
  - Exercice : tester et expérimenter des principes de constructions pour un slide
  - Évaluer l'utilité d'une charte graphique
  - Mesurer les limites de l'interprétation
  - Initier des bonnes pratiques : - tester des méthodes pour mettre en avant une information - présenter ses graphiques de façon pertinente et créer du rythme - utiliser des gabarits adaptés aux données
  - Éviter les erreurs récurrentes et anticiper les difficultés
  - Veiller à l'harmonie des éléments textuels
  - Examiner les spécificités selon les supports (print, numérique)
- Choisir la meilleure forme graphique pour présenter des données
  - Expérimenter les différentes possibilités de mise en forme
  - Identifier les points forts et trouver des alternatives aux points faibles
  - Tester la lisibilité et la compréhension
  - Argumenter ses choix
  - Exercice : tester et expérimenter des mises en forme pour représenter les données
  - Quiz n°2

Organisation de la formation

Moyens pédagogiques et techniques

- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation.

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Quiz en salle
- Etude de cas concrets
- Mises en situation.

FORMATIONS BUSINESS INTELLIGENCE

Power BI

Microsoft Power BI - initiation

Description

Que vous soyez analyste, consultant, contrôleur de gestion, chargé d'étude ou simplement intéressé par la business intelligence, cette formation vous permet de devenir autonome pour l'analyse de vos données avec Microsoft Power BI Desktop.

Microsoft propose désormais un logiciel regroupant les compléments Power BI initialement développés sous Excel. Ce logiciel s'adresse aussi bien à des développeurs qu'à des utilisateurs métier (« end-user BI »). La diffusion des tableaux de bord se fait alors principalement au travers d'un compte en ligne (« cloud-based BI »).

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-microsoft-power-bi-desktop/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Professionnel BI
- Data analyste
- Contrôleur de gestion
- Manager

Prérequis :

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

Objectifs pédagogiques

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

Programme détaillé

- Introduction
  - Présentation des outils décisionnels Microsoft et de leurs capacités
  - Présentation des principes de bases pour la visualisation et la business intelligence
- Sources de données

- Intégrer des données de type fichiers plats (texte, CSV), classeurs Excel simples ou multi-onglets, bases de données Access, page Web
- Choisir le bon encodage de fichier

- Transformations
  - Lancer l'éditeur de requête
  - Appliquer les transformations disponibles dans l'éditeur
  - Exploiter le fonctionnement des étapes de transformation
  - Définir les types de données
  - Choisir les lignes et les colonnes à conserver, supprimer les doublons
  - Ajouter une colonne personnalisée, sur la base d'exemples
  - Ajouter des requêtes entre elles, sur la base de colonnes identiques
- Modélisation
  - Désactiver le chargement d'une requête
  - Définir les relations entre les tables : un à plusieurs, un à un, active ou inactive
  - Ajouter une table de dates
  - Trier une colonne par une autre colonne
  - Créer une hiérarchie, exploiter la hiérarchie de dates
  - Typer les champs géographiques pour une meilleure géolocalisation
- Formules de calcul
  - Maîtriser la syntaxe globale des formules DAX
  - Choisir entre colonne calculée et mesure
  - Faire la différence entre mesure implicite et mesure explicite
  - Exploiter rapidement les fonctions DAX équivalentes aux fonctions Excel
  - Générer automatiquement des mesures rapides
  - Déplacer et formater une mesure explicite
  - Maîtriser les premières fonctions indispensables (DISTINCTCOUNT, COUNTROWS, RELATED, CALENDAR, CALENDARAUTO, FORMAT, USERLATIONSHIP)
  - Première découverte de la fonction CALCULATE
  - Utiliser correctement les fonctions de Time-Intelligence indispensables (SAMEPERIODLASTYEAR, DATESYTD...)
- Visualisations
  - Créer et paramétrer des visuels (tableaux et matrices, Graphiques simples, Cartes géographiques, Segments, listes déroulantes, chronologie)
  - Exploiter les filtres basiques et avancés au niveau des visuels, pages et rapport
  - Insérer une image, une zone de texte
  - Afficher les interactions
  - Ajouter un visuel personnalisé (« custom viz »)
  - Créer une info-bulle sur mesure
- Publication et partage
  - Se connecter au service Power BI
  - Publier un rapport sur le service Power BI
  - Exporter un rapport en PDF ou au format Power Point
  - Créer et partager un tableau de bord
  - S'abonner à un tableau de bord

- Ajouter d'autres vignettes (image, URL)
- Créer une alerte sur une vignette
- Appliquer un rôle de sécurité simple

Microsoft Power BI - perfectionnement

Description

Que vous soyez analyste, consultant, contrôleur de gestion, chargé d'étude ou simplement intéressé par la business intelligence, ces formations vous permettent de devenir autonome pour l'analyse de vos données avec Microsoft Power BI Desktop.
Microsoft propose désormais un logiciel regroupant les compléments Power BI initialement développés sous Excel. Ce logiciel s'adresse aussi bien à des développeurs qu'à des utilisateurs métier (« end-user BI »).

La diffusion des tableaux de bord se fait alors principalement au travers d'un compte en ligne (« cloud-based BI »).

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-microsoft-power-bi-desktop/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Expert BI
- Analyste
- Contrôleur de gestion
- Chargé d'étude

Prérequis :

- Avoir suivi la formation initiation ou avoir une bonne maîtrise des bases de Power BI

Objectifs pédagogiques

- Construire un projet Power BI robuste en appliquant les meilleures pratiques
- Maîtriser les syntaxes de formules DAX avancées
- Exploiter les fonctionnalités collaboratives du service Power BI

Programme détaillé

- Sources de données
  - Importer des fichiers identiques depuis un dossier
  - Se connecter à une base de données (import ou direct query), un cube (import ou live connection)

- Exporter un fichier Power BI en modèle (.pbit)

- Transformations
  - Fusionner des requêtes sur le principe des jointures SQL
  - Gérer les erreurs
  - Modifier directement le langage M
  - Exploiter les objets listes
  - Utiliser les paramètres
  - Mettre en place des fonctions M
  - Optimiser les requêtes
  - Utiliser les diagnostiques de requêtes
  - Comprendre les outils profilage et la qualité des colonnes
- Modélisation
  - Construire une modélisation en étoile
  - Gestion des relations de principe « Many to Many »
  - Appliquer une direction de filtre croisée
- Formules de calcul (langage DAX)
  - Comprendre la notion de contexte de ligne, contexte de filtre
  - Exploiter les familles de fonctions DAX :
    - Fonctions Time Intelligence : DATEADD, DATESYTD, DATESBETWEEN…
    - Fonctions itératives : SUMX, AVERAGEX, MAXX, MINX…
    - Fonctions de filtre : CALCULATE, ALL, FILTER, RELATEDTABLE
    - Fonctions de table : SUMMARIZE, CALCULATETABLE, ADDCOLUMN
    - Autres fonctions : HASONEVALUE, VALUES, HASONEFILTER
  - Utiliser DAX Studio pour développer des syntaxes DAX avancées
- Visualisations
  - Créer un segment pilotant le choix des mesures d'un visuel
  - Utiliser un thème augmentant la taille de toutes les polices
  - Afficher des images dans un segment
- Publication et partage
  - Créer un espace de travail d'application en lecture ou en écriture
  - Appliquer une stratégie de sécurité à la ligne (RLS) avancée
- Power Platform
  - Introduction à Power Automate :
    - Mise en pratique d'un cas concret
    - Rafraîchir un jeu de données
    - Mettre un place un flux d'approbation
    - Interaction avec Teams

Microsoft Power BI - dataviz & analytics

Description

Que vous soyez analyste, consultant, contrôleur de gestion, chargé d'étude ou simplement intéressé par la business intelligence, ces formations permettent de devenir autonome pour l'analyse de vos données avec Microsoft Power BI Desktop. Microsoft propose désormais un logiciel regroupant les compléments Power BI initialement développés sous Excel, et s'adresse aussi bien à des développeurs qu'à des utilisateurs métier (« end-user BI »). La diffusion des tableaux de bord se fait alors principalement au travers d'un compte en ligne (« cloud-based BI »).Formation en modules de 2 jours. Suivant votre niveau, besoin, vous sélectionnez les modules souhaités.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-microsoft-power-bi-desktop/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Expert BI
- Analyste
- Contrôleur de gestion, Chargé d'étude

Prérequis :

- Avoir suivi la formation Power BI Initiation ou Perfectionnement
- ou avoir une bonne maîtrise des bases de Power BI

Objectifs pédagogiques

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

Programme détaillé

- Introduction à la data visualisation
  - Comprendre la valeur ajoutée de la data visualisation
  - Evaluer l'efficacité d'un visuel (« data ink ratio », parcours de l'œil)
  - Eviter les pièges et les mauvaises pratiques
- Choisir le bon graphique en fonction du message
  - Distinguer les différentes familles de graphiques
  - Organiser les données pour faciliter la création des visuels

- Utiliser un « chart chooser »
- Exploiter des visuels originaux et mieux adaptés (diagramme de flux, diagramme de Kiviat, graphes de relation, Treemap, Marimekko, heat map…)
- Réaliser des cartes géographiques, éviter les doublons géographiques

- Aller vers le data story telling
  - Réflexions sur l'organisation des pages de rapport et du tableau de bord
  - Proposer des filtres d'extraction (« drill through »)
  - Exploiter les signets (« bookmarks »)
  - Sélectionner les visuels et mettre en surbrillance
- Exploiter les fonctionnalités de Power BI pour la data visualisation
  - Utiliser le quadrillage et l'alignement des objets
  - Définir un fichier de thème personnalisé au format JSON
  - Adapter les visuels pour un affichage mobile
  - Paramétrer les interactions et explorer un visuel grâce aux hiérarchies
- Exploiter et interpréter les visuels pour l'analyse
  - Utiliser les visuels personnalisés (« custom viz ») d'analyse avancée : Correlation plot, Box and Whisker chart, Decision Tree chart, Time series decomposition chart
  - Découvrir les visuels de Microsoft Research : Sand Dance, Timeline Storyteller
- Adopter une démarche méthodologique pour l'analyse des données
  - Notions de variables explicatives et variable à expliquer
  - Afficher les valeurs en pourcentage du total général, lignes ou colonnes
  - Exploiter les mesures rapides (« quick measures ») pour réaliser des calculs statistiques
- Exploiter les options analytiques des visuels
  - Animer un nuage de bulles dans le temps
  - Découvrir une segmentation grâce au nuage de points (« clustering »)
  - Afficher une trame de symétrie ou une ligne de ratio
  - Ajouter une moyenne, médiane ou tendance sur un visuel
  - Prolonger une courbe dans le temps (« forecast »)
  - Expliquer les évolutions d'une courbe (« explain the increase / decrease »)
  - Ajouter un facteur explicatif dans un graphique en cascade (« waterfall »)
- Utiliser les scénarios paramétrés (« what if »)
  - Définir un nouveau paramètre
  - Exploiter un paramètre dans une mesure (SELECTEDVALUE)
  - Croiser différents paramètres dans des scénarios complexes (GENERATESERIES)

Atelier Découverte Microsoft Power BI

Description

Une formation sous forme d'ateliers pour découvrir Microsoft Power BI pour des utilisateurs métiers non développeurs. Les exemples sont adaptés à vos problématiques métiers afin que les participants s'adaptent à l'utilisation de tableaux de bord avec Microsoft Power BI. Cette formation découverte vous permet d'acculturer vos équipes à l'utilisation d'un outil comme Microsoft Power BI.

Durée

0.50 jours (3.50 heures)

Profils des stagiaires et prérequis

Profils :

- Consommateurs de tableaux de bord avec Microsoft Power BI

Prérequis :

- Aucun pré-requis

Objectifs pédagogiques

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

Programme détaillé

- Introduction
  - Présentation des outils Microsoft Power BI : Desktop et Service
  - Présentation des principes de bases pour la visualisation et la business intelligence
- Utilisation de Microsoft Power BI Service
  - Comment se connecter à Microsoft Power BI
  - Comment naviguer dans Microsoft Power BI Service
  - Que puis-je faire avec Microsoft Power BI Service ?
- Aller plus loin avec Microsoft Power BI Service
  - Comment partager un tableau de bord avec Microsoft Power BI
  - Que puis-je créer avec Microsoft Power BI service ?
- Séance de questions / réponses sur l'utilisation de Microsoft Power BI

Atelier Langage DAX : expertise

Description

Une journée intensive pour des experts Microsoft Power BI et maîtriser le langage DAX.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/consulting-data-science/accompagnement-et-developpement-power-bi/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- BI professionnal
- Analystes

Prérequis :

- Avoir suivi la formation perfectionnement ou avoir un niveau équivalent en Power BI

Objectifs pédagogiques

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

Programme détaillé

- Environnement développeur (outils de monitoring et d'optimisation)
  - Reconnaissance du DAX dans Notepad++
  - SQL SERVER Management Studio : profiler les requêtes DAX
  - DAX Studio : tester les requêtes tables
    - EVALUATE
    - DEFINE MEASURE
    - ORDER BY
    - START AT
  - PowerPivot Utilities (add-in Excel)
  - OLAP PivotTable Extensions : obtenir le langage MDX d'une requête vers un cube - Création de variables DAX (VAR… RETURN)
- Tables calculées ou « reverse linked table »
  - Créer une table calculée dans Power BI Desktop (CALENDARAUTO)

- Relation plusieurs à plusieurs
    - Forcer le contexte d'évaluation (CALCULATE)
    - Utiliser le filtrage bidirectionnel
- Travailler avec des calendriers personnalisés
    - Calendrier de type 4/4/5
- Fonctions DAX avancées
    - Rappels sur les fonctions itératives - SUMX, AVERAGEX, …
    - Rappels sur les fonctions de filtre - EARLIER, FILTER
    - Fonctions d'agrégation
        - SUMMARIZE et options ROLLUP, ROLLUPGROUP
        - SUMMARIZECOLUMNS et options IGNORE, ROLLUPADDISSUBTOTAL, ROLLUPGROUP
        - ADDCOLUMNS
        - ADDMISSINGITEMS
        - GENERATE
    - Fonctions de jointure - CROSSJOIN, NATURALINNERJOIN, NATURALLEFTOUTERJOIN, UNION, INTERSECT
    - Gestion avancée des droits (« Row Line Security ») - CONTAINS, CUSTOMDATA

Utilisation de R dans Power BI

Description

Une journée pour découvrir l'utilisation du langage R dans Microsoft Power BI pour créer des modèles de data science et des visualisations avancées

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/consulting-data-science/accompagnement-et-developpement-power-bi/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- BI professionnal
- Data analyst
- Data scientist

Prérequis :

- Avoir suivi la formation perfectionnement ou avoir un niveau équivalent en Power BI
- Utilisateur occasionnel de R

Objectifs pédagogiques

- Connaître les différents niveaux d'interaction de R avec Power BI
- Appliquer les méthodes statistiques de R au travers de Power BI
- Développer les visualisations grâce aux librairies graphiques de R

Programme détaillé

- Installation de R et RStudio (IDE)
  - Distribution Microsoft de R (ROpen)
  - Options dans Power BI Desktop
  - Paramètres de Power BI Service
- Import / export de données
  - Charger des données (intérêt et limites)
  - Reconnaître les valeurs manquantes
  - Joindre des jeux de données (rbind, cbind, merge)
  - Exemple de web scrapping avec R
  - Exporter le résultat d'une requête
- Transformations des données

- Remplacer les valeurs manquantes
- Repérer les valeurs aberrantes (boxplot)
- Recoder une variable numérique en plages (cut)
- Recoder les niveaux d'un facteur (levels)
- Créer une fonction personnalisée (UDF)

- Algorithmes de base
  - Segmentation non supervisée par clustering (KMeans)
  - Réduction des dimensions par Analyse en Composantes Principales
  - Scoring par régression logistique
  - Prévision par séries chronologiques
  - Analyse de données textuelles
- Visualisations
  - Galerie des visualisations R (R-powered visuals)
  - Utiliser l'élément visuel R (principes et limites)
  - Utiliser les librairies graphiques (ggplot2)
  - Exporter un graphique

Designer Qlik Sense

Description

Une formation pour débuter en tant que designer Qlik Sense. Sur une journée, découverte de l'interface designer pour modifier des applications existantes.

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Utilisateurs métiers de Qlik Sense
- Utilisateurs avec des besoins de manipulations de tableaux de bord

Prérequis :

- Avoir déjà manipuler un tableau de bord avec Qlik Sense ou avoir suivi la formation "Découverte Qlik Sense"

Objectifs pédagogiques

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

Programme détaillé

- Prendre en main une application existante avec Qlik Sense
  - Découverte de l'interface Designer
  - Modifier une application existante avec l'interface Designer
  - Modifier les éléments d'une page d'analyse, les formules, les graphiques et les tableaux
  - Utilisation des masters items
- Effectuer des analyses avancées avec Qlik Sense
  - Connaître les bonnes visualisations suivant vos problématiques
  - Choisir les bonnes analyses suivant vos données
- Partage d'applications avec Qlik Sense
  - Partager vos applications avec d'autres utilisateurs
  - Créer des applications "responsive"
  - Utiliser le « Story Telling » pour présenter vos analyses

Développeur Qlik Sense

Description

Une formation sur 2 jours pour des utilisateurs orientés data qui désirent utiliser les capacités de Qlik Sense pour construire des applications complètes. Une formation basée sur des cas pratiques et de la manipulation pour découvrir les possibilités de Qlik Sense.

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Utilisateurs data de Qlik Sense
- Data analysts
- Data scientists

Prérequis :

- Avoir construit des applications Qlik Sense ou avoir suivi la formation Designer Qlik Sense

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 de données
- Transformer les données

Programme détaillé

- L'interface de Qlik Sense
  - Découvrir l'interface de développement
  - Data Manager, Data Load Editor et Data Model Viewer
  - Naviguer et interagir avec l'interface utilisateur
- Le langage de script de Qlik Sense
  - Structuration du langage
  - Contrôles / préfixes / expressions / variables
  - Fonctions de bases pour les scripts
- Les modèles de données avec Qlik Sense
  - Bonnes pratiques pour les modèles de données avec Qlik Sense
  - Charger des données : depuis un fichier / depuis l'éditeur de chargement / avec d'autres sources de données

- Manipulation de données
- Le modèle de données Qlik Sense : principes
- Créer des modèles en étoile
- Gérer des clés synthétiques et des références circulaires

- Création d'applications et visualisations
  - Connaître les différents types de visualisations de Qlik Sense
  - Mesures et dimensions
  - Utiliser des fonctions dans des graphiques
  - Utilisation avancée du Set Analysis
  - Travailler avec des master
- Allez plus loin avec Qlik Sense
  - Déploiement
  - Calculs avancés
  - Sécurité

Découverte Qlik Sense

Description

Une formation sous forme d'ateliers pour découvrir Qlik Sense pour des utilisateurs métiers non développeurs. Les exemples sont adaptés à vos problématiques métiers afin que les participants s'adaptent à l'utilisation de tableaux de bord avec Qlik Sense. Cette formation découverte vous permet d'acculturer vos équipes à l'utilisation d'un outil comme Qlik Sense.

Durée

0.50 jours (3.50 heures)

Profils des stagiaires et prérequis

Profils :

- Tous utilisateurs métiers débutants ou futurs utilisateurs de Qlik Sense

Prérequis :

- Pas de pré-requis

Objectifs pédagogiques

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

Programme détaillé

- Introduction
  - Présentation de Qlik Sense
  - Présentation des types d'utilisateurs Qlik Sense
  - Qu'est-ce qu'une application Qlik Sense ?
- Utilisation de Qlik Sense
  - Comment se connecter à Qlik Sense
  - Comment naviguer dans une application Qlik Sense
  - Que puis-je faire avec Qlik Sense ?
- Aller plus loin avec Qlik Sense
  - Comment partage des visualisations Qlik Sense
  - Que puis-je créer avec Qlik Sense en tant qu'utilisateur métier ?
- Séance de questions / réponses sur l'utilisation de Qlik Sense

QlikView – Initiation

Description

Que vous soyez analyste, consultant, contrôleur de gestion, chargé d'étude ou simplement intéressé par la business intelligence, cette formation vous permet de devenir autonome pour l'analyse de vos données et la construction d'applications avec QlikView.
Ce logiciel s'adresse aussi bien à des développeurs qu'à des utilisateurs métier (« end-user BI »).

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-qlikview/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Tout public
- Analyste
- Consultant
- Contrôleur de gestion
- Chargé d'étude

Prérequis :

- Pas de pré-requis : accessible à tout public

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

Programme détaillé

- Présentation des outils et de leurs capacités
- Présentation des principes de base pour la visualisation et la business intelligence
- Chargement des données et création de document Qlikview
- Edition et mise en forme de documents QlikView
- Représentation graphique avec QlikView
- Introduction à la modélisation avec QlikView

Tableau Desktop

Description

Que vous soyez analyste, consultant, contrôleur de gestion, chargé d'étude ou simplement intéressé par la business intelligence, cette formation Tableau Desktop vous permet de devenir autonome pour l'analyse de vos données avec Tableau Desktop.
Ce logiciel s'adresse aussi bien à des développeurs qu'à des utilisateurs métier (« end-user BI »).

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-tableau-desktop/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Analyste
- Consultant
- Contrôleur de gestion
- Chargé d'étude
- Utilisateur habitué à Windows et à l'utilisation de tableurs.

Prérequis :

- Utilisateur habitué à Windows et à l'utilisation de tableurs.

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

Programme détaillé

- Présentation des outils et de leurs capacités
  - L'outil Tableau Desktop
  - Définitions et terminologies
- Présentation des principes de bases pour la visualisation et la business intelligence
  - Quelques bases pour passer d'un tableur classique à Tableau Desktop
  - Bien présenter des visualisations de données
  - Illustrations avec Tableau Desktop

- Connexion aux données (Excel, base de données)
    - Créer des connexions
    - Rafraîchir des données
    - Panorama des options de connections de Tableau desktop
- Création de visualisations de données avec Tableau Desktop
    - Initialiser une visualisation
    - Filtrer et trier des données
    - Organisation et exploration des données (Groupes, hierarchies, dates…)
    - Utilisation de mesures multiples
    - D'autres utilisations : Corrélations, gestion de données cartographiques, histogrammes, tableaux croisés…
- Introduction à des analyses avancées
    - Créer des calculs dans Tableau Desktop
    - Créer des graphiques animés
- Visualisation et partage des tableaux de bord avec Tableau Desktop
    - Construire un tableau de bord
    - Les possibilités de partage de tableaux de bord

FORMATIONS OUTILS DATA SCIENCE

SQL

OPEN Refine

git

Open for Innovation
KNIME

data iku

Git pour versionner votre code

Description

Prêt à maîtriser Git comme un pro ? Une formation pour devenir un expert de la gestion de versions et de la collaboration avec Git

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst et data scientist désirant utiliser Git au quotidien
- Développeur souhaitant utiliser Git pour ses projets

Prérequis :

- Aucun prérequis

Objectifs pédagogiques

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

Programme détaillé

- Principes de Git
  - L'aspect décentralisé de Git
  - Utilisation de Git au quotidien
  - Les différentes interfaces pour utiliser Git
  - Le lien avec GitHub et GitLab
- Prise en main de Git
  - Installation de Git
  - La création d'un premier dépôt avec Git
  - La gestion des fichiers dans le dépôt
  - La notion de commit et les bonnes pratiques pour tagger ses commits
- Les branches
  - Création et principes des branches
  - Bonne gestion des branches
  - Les conflits liés aux fusions de branches et leur résolution
- Bonne utilisation d'un dépôt disntant
  - Différences entre push / pull / clone...

- Gestion de la synchronisation avec votre dépôt distant
- Les workflows git
  - Principes des workflows git
  - Mise en pratique

Organisation de la formation

Moyens pédagogiques et techniques

- Support avec nombreux cas d'usage
- Mise à disposition d'un environnement de test pour lancer des commandes Git sur GitHub / GitLab

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Cas pratiques validés par le formateur
- Quiz en ligne

SQL - Bases du langage

Description

Découvrez le langage SQL pour manipuler vos données dans des bases de données relationnelles.

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst
- Utilisateurs potentiels de bases de données

Prérequis :

- Aucun prérequis sur le langage SQL

Objectifs pédagogiques

- Comprendre la structuration des bases de données relationnelles
- Savoir écrire des requêtes SQL
- Savoir utiliser les fonctions standards du langage SQL

Programme détaillé

- Base de données relationnelles
  - Qu'est-ce qu'une base de données relationnelle ?
  - Savoir identifier le schéma d'une base de données
  - Les différents systèmes de bases de données (Oracle, MySQL, PostgreSQL, Microsoft SQL...)
  - Outil pour interroger une base
  - Structure générale d'une base de données (Base, Schéma, Tables)
  - Types d'instructions SQL : DDL, DML, DCL
  - Éléments de la syntaxe SQL
- La mise à jour des données
  - L'insertion de données (insert)
  - La mise à jour (update)
  - La suppression d'informations (delete)
- Extraire les données d'une table
  - La clause WHERE pour filtrer les données
  - L'absence de valeur (marqueur NULL)
  - Retourner des lignes sans doublon (DISTINCT)

  - Alias de colonnes et alias de tables
- Fonctions d'agrégats
  - Clause GROUP BY
  - Clause HAVING
- Interroger les données de plusieurs tables
  - Principe des jointures
  - Jointure interne et jointure externe
  - Opérateurs ensemblistes (UNION, INTERSECT...)
  - Les vues : principe et intérêt
- Statistiques
  - Trouver des valeurs agrégées (MIN, MAX, AVG, SUM...)
  - Calculer des agrégats relatifs avec GROUP BY
  - Filtrer les valeurs agrégées avec HAVING
  - Ordonnancement des résultats avec RANK, ROW_NUMBER
- Présenter et trier les données
  - Conversion d'un type à un autre
  - Effectuer des choix à l'aide de l'opérateur CASE
  - Trier les données avec ORDER BY
  - Opérations sur les chaînes de caractères, sur les dates

Organisation de la formation

Moyens pédagogiques et techniques

- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation.

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Questions orales ou écrites (QCM).
- Mises en situation.

Découverte OpenRefine

Description

Une formation de découverte pour prendre en main l'outil openRefine pour la manipulation de données. Formation basée sur de la pratique.

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Analystes
- Utilisateurs d'Excel désirant manipuler des données

Prérequis :

- Aucun prérequis

Objectifs pédagogiques

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

Programme détaillé

- Présentation d'OpenRefine et bases de la manipulation de données
    - Les bonnes pratiques de manipulation de données
    - L'outil OpenRefine pour manipuler des données
    - Licence, utilisation, limites
- Le chargement des données
    - Les différents types de fichiers pris en charge
    - Le processus de chargement et de visualisation des données dans OpenRefine
- Les principes de base de OpenRefine
    - Les facettes et filtres
    - L'historique des actions
    - Les colonnes et leurs informations
    - Le travail dans un projet
    - Les étoiles et les drapeaux
    - Cas pratique de manipulation de données
- La manipulation des données avec OpenRefine
    - Tri, filtres et facettes...
- Le nettoyage de données

  - Au niveau d'une cellule
  - Au niveau d'une colonne
  - Au niveau d'une table (script)
- Mise en pratique de transformations courantes
  - Modification de textes
  - Changement de format
  - Dédoublonnage
  - Transposition de données
  - Concaténation de colonnes
  - Renommage dans une colonne
- L'exportation des données transformées
- Aller plus loin - L'application de la personnalisation
  - Gestion de l'historique des traitements
  - Introduction au langage GREL (General Refine Expression Language) et son application dans OpenRefine

Organisation de la formation

Moyens pédagogiques et techniques

- Support PDF projeté
- Cas pratiques sous forme de données aux différents formats pris en charge par l'outil

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Cas pratiques en classe validés par le formateur
- Quizz en ligne permettant de valider la bonne compréhension

Dataiku DSS - Fondamentaux

Description

Que vous soyez data scientist, data analyst, analyste SAS ou utilisateur métier, et que vous désirez découvrir la solution Data Science Studio (DSS) de DataIku, cette formation DataIku est faite pour vous.
Deux jours intensifs pour découvrir le logiciel DSS de DataIku, la manipulation de données, la création de modèles jusqu'à la mise en production de modèles.

Après une introduction à l'outil, vous apprendrez à maîtriser les flux de traitements disponibles dans DataIku DSS. Basée sur de nombreux exercices pratiques, cette formation DataIku vous permet d'être immédiatement opérationnel pour le traitement de vos données grâce à DataIku.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-dataiku/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Tous public avec quelques connaissances en traitement de données.

Prérequis :

- Avoir quelques connaissances en traitement de données.

Objectifs pédagogiques

- Découvrir l'interface de DSS et de son utilisation
- Comprendre les installations et les systèmes de licensing
- Construire des projets et des flux de manipulation de données
- Mettre en place des premiers modèles de machine learning
- Comprendre la mise en production des modèles et des flux

Programme détaillé

- Introduction à Data Science Studio (DSS)
    - Qu'est-ce que DSS ?
    - Comment installer DSS ?
    - Quels usages et quels utilisateurs ?
    - DSS dans le paysage des plateformes data science

- Quelles solutions possibles pour DSS ? Mode de licensing et d'utilisation

- L'interface DSS
  - Découverte de l'interface
  - La notion de projet
  - La notion de flux
  - Travailler dans le Lab
  - Comment travailler en équipe avec DataIku : Le partage
    Le versionnement
- La manipulation des données avec DSS
  - Principe de la préparation des données
  - Importer des données de différents types (csv, xls, sql…)
  - Appliquer des recettes pour transformer des données : Transformation de données
    Gestion des types
    Manipulation des données textuelles
    Jointure de fichiers
  - Ateliers sur des données d'open data
- La data visualisation avec DSS
  - Présentation des principes de la data visualisation
  - Quelques graphiques simples avec DataIku DSS
  - Présentation de visualisations plus avancées
  - Ateliers sur des données réelles :Manipulation de données pour afficher des séries temporelles
    Affichage de cartographie avec DSS
  - Le développement et le partage de tableaux de bord (Dashboard) avec DSS
- Introduction au machine learning : présentation des principes de l'apprentissage automatique
  - Construction d'un premier modèle avec Dataiku : Le machine learning supervisé
    Le machine learning non supervisé
  - Les autres modèles disponibles dans DSS
  - Atelier sur des données : Préparation et construction de modèles
    Ajustement des hyper-paramètres
- La mise en production d'un projet DSS
  - Les bundles
  - Les automation node et les API node
  - Les scénarios
  - Le monitoring
- Aller plus loin :
  - L'intégration de code en Python ou en R dans DataIku DSS (les Notebooks) :
    Utilisation de Notebook et de recettes sous forme de code
  - L'utilisation d'infrastructures Big Data (Hadoop / Spark) :
    Le choix du moteur de calcul
  - Le deep learning dans DSS

Logiciel KNIME – les fondamentaux

Description

Le logiciel KNIME est aujourd'hui l'un des leaders parmi les plateformes de Data Science. Son système évolué de flux permet de construire des traitements avancés en utilisant différentes technologies. Que vous soyez Data Scientist, analyste, consultant, notre formation au logiciel KNIME vous permet d'acquérir une compétence supplémentaire compétitive.

Stat4decision vous propose une formation complète combinant une initiation et du perfectionnement.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-logiciel-knime/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst
- Data scientist

Prérequis :

- Avoir des bases en traitement de données

Objectifs pédagogiques

- Maîtriser l'utilisation de l'interface du logiciel KNIME
- Savoir 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
- Etre capable d'utiliser d'autres outils de Data Science dans KNIME

Programme détaillé

- Présentation de l'outil et de ses principes : KNIME Analytics Platform
  - Présentation du système et du fonctionnement
  - Les extensions
  - KNIME Server
  - Premiers nœuds et flux
- Chargement et préparation des données

- Utiliser des fichiers plats
- Utiliser des bases de données
- Principe des connecteurs « big data »

- Exploration et analyses descriptives des données avec KNIME
  - Obtention de data visualisations
  - Transformation et traitement des données - Construction de flux de transformation
- Mise en place de modèles prédictifs (machine learning) :
  - Utilisation de modèles prédictifs
  - Amélioration et évaluation des modèles
- Automatiser vos flux
  - Utilisation des variables de flux pour contrôler vos flux
  - Utilisation de boucles dans un flux pour répéter des traitements au sein d'un flux
  - Utilisation de conditions (SWITCH)
- Importation et modification de données avancé
  - Requêtes SQL
  - Données du web
- Utilisation d'autres applications dans KNIME
  - L'exemple de Python
  - L'exemple de R
  - Les traitements big data avec KNIME- Hadoop / Spark / Machine Learning
- Les extensions de KNIME
  - Présentation et manipulation d'extensions
  - Principe de développement d'extensions

Logiciel RapidMiner

Description

Que vous soyez analyste, consultant, data scientist, cette formation vous permet de devenir autonome pour l'analyse de vos données avec RapidMiner.

RapidMiner est une plateforme data science qui permet de mettre en place de manière simple des flux de traitement pouvant être passés en production. Cette formation logiciel RapidMiner vous permet d'acquérir les bases pour le traitement de vos données et l'automatisation des flux avec RapidMiner.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-logiciel-rapidminer/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Analyste
- Data scientist

Prérequis :

- Avoir des bases en traitement de données.

Objectifs pédagogiques

- Maîtriser la préparation et le traitement des données avec RapidMiner
- Comprendre les principes de déploiement de modèle data science avec RapidMiner

Programme détaillé

- Présentation de l'outil et de ses principes : RapidMiner Studio
- Chargement et préparation des données
    - Chargement des données
    - Analyses descriptives
    - Les premiers graphiques
    - Transformation et traitement des données
- Construction de flux
    - Utilisation de modèles prédictifs
    - Amélioration et évaluation des modèles

- Déploiement de modèles
- Aller plus loin : présentation des possibilités en traitement de données textuelles et en traitement de données « big data »

Langage Julia – les fondamentaux

Description

Julia est un jeune langage développé au MIT qui tente de combiner tous les avantages des différents langages de programmation utilisés pour le calcul scientifique et plus spécifiquement pour le traitement des données.

Cette formation au langage Julia se base sur des applications pratiques et des exposés sur les principes du langage et son fonctionnement.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-au-langage-julia/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data scientist
- Développeur
- Chercheur

Prérequis :

- Avoir déjà utilisé au moins un langage de programmation pour le calcul scientifique ou statistique (R, python, matlab, SAS…)

Objectifs pédagogiques

- Pouvoir lancer dans des projets en Julia
- Maîtriser les forces et les faiblesses de ce langage
- Identifier les outils nécessaires et les principes de codage en Julia

Programme détaillé

- Un nouveau langage : Julia
  - Principes du langage
  - Comparatif Julia vs les autres langages de calcul scientifique
  - Les bibliothèques de Julia
- Les outils pour coder en Julia
  - Un compilateur sophistiqué
  - Des outils encore basiques
  - Un écosystème en constante évolution

- Quelques exemples
    - Un premier programme avec Julia
    - Les fonctions mathématiques et statistiques
    - Les bibliothèques pour l'algèbre linéaire
    - Le calcul parallélisé avec Julia
- Un cas pratique complet de projet en Julia
    - Quand choisir Julia ?
    - Quels sont les réels avantages de Julia sur des cas réels ?

Microsoft Azure Machine Learning

Description

Cette formation Azure Machine Learning vous permet d'acquérir en 3 jours les compétences nécessaires à l'utilisation du nouveau service de traitement des données de Microsoft sur le cloud : Azure Machine Learning.

Intégré à la solution cloud de Microsoft, Azure Machine Learning Studio est une interface web permettant de construire la chaîne de traitement de la donnée. Relativement simple d'utilisation, cet environnement prend toute sa valeur par le choix judicieux des méthodes et l'optimisation de leur paramétrage. Une fois l'algorithme validé, un service web pourra traiter les nouvelles données et fournir, en retour, une prévision à l'utilisateur.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-azure-machine-learning/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data analyst

Prérequis :

- Connaissances statistiques de base (centrage, dispersion, corrélation, tests d'hypothèses)
- Connaissance de base du principe des API

Objectifs pédagogiques

- Prendre en main l'interface de 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.

Programme détaillé

- Prise en main de l'interface Azure Machine Learning Studio
    - Découvrir l'ensemble de l'offre Azure et son mode de facturation à l'usage
    - Découvrir l'interface Machine Learning Studio
    - Créer un jeu de données (dataset) à partir d'un fichier local ou se connecter à une source de données

- Construire une nouvelle expérience de machine learning
- Définir un Web Service prédictif
- Proposer ses algorithmes sur le Marketplace

- Les algorithmes de Machine Learning
  - Utiliser l'arbre de choix des algorithmes
  - Détecter les valeurs aberrantes
  - Choisir les variables de l'algorithme (features selection)
  - Initialiser le modèle, entrainer le modèle, évaluer le modèle
  - Paramétrer les algorithmes de régression (Linéaire, Logistique, Ordinale)
  - Paramétrer les algorithmes de classification :
    (Arbres, forêts, jungles d'arbres de décision (forest decision, jungle decision);
    Réseaux de neurones;
    Machine à vecteurs de support
  - Paramétrer les algorithmes de clustering : Choisir le nombre de clusters
  - Evaluer et comparer les différents algorithmes à l'aide de la courbe ROC
- Traiter d'autres types de données : texte ou images
  - Convertir le texte en mesures numériques
  - Appliquer un algorithme Vowpal Wabbit (online learning)
  - Procédure de configuration Import Images

FORMATIONS STATISTIQUE ET ANALYSE DE DONNÉES

Statistique et analyse de données

Description

Cette formation statistique et analyse de données de 3 jours vous permet de découvrir et maîtriser les bases de la statistique et de l'analyse de données.

Vous avez besoin de comprendre les analyses statistiques de bases et interpréter les résultats, cette formation répond à ce besoin. Il s'agit d'une formation pratique qui s'appuie sur des exemples concrets afin d'illustrer les logiques de la statistique et de l'analyse de données.

Cette formation se base sur différents outils dont la maîtrise n'est pas nécessaire (Excel, R, XLSTAT…).

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-statistique-analyse-de-donnees/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Business analyste
- Spécialiste BI
- Utilisateur de la statistique

Prérequis :

- Avoir des bases en statistique (moyenne, variance...)

Objectifs pédagogiques

- Comprendre les principes de bases de la statistique.
- Savoir décrire des données en utilisant des statistiques descriptives.
- Comprendre les bases de la modélisation et de l'analyse de données

Programme détaillé

- Introduction à la statistique et à l'analyse de données
- Les principes de bases : échantillonnage, population vs. échantillon…
- La préparation des données (transformation, traitement des données manquantes, filtrage…)
- Cas du traitement d'une enquête
- Les statistiques descriptives pour décrire vos données (moyenne, variance, médiane, mode…)
- Les intervalles de confiance et les tests statistiques
- Les tests de comparaison de moyennes

- Les tests paramétriques et non paramétriques
- Cas pratiques : applications sur des données réelles
- Principes de l'analyse de données multivariée
- L'analyse en composantes principales
- Les autres méthodes d'analyse factorielles (AFC, ACM, analyse discriminantes…)
- Présentation des méthodes de clustering (apprentissage non supervisé)
- La classification ascendante hiérarchique
- Les k-means
- Les principes de la modélisation
- Le modèle linéaire : régression, analyse de la variance
- La régression logistique
- Vers le machine learning et le data mining
- Les principes et les différences avec la statistique classique
- Quelques exemples : Les arbres, les forets…

Statistique et analyse de données avec XLSTAT

Description

Cette formation XLSTAT complète de 3 jours vous permet de découvrir et maîtriser les bases de la statistique et de l'analyse de données avec le logiciel XLSTAT.

Vos données sont au format Excel et vous voulez effectuer des analyses avancées ? Pourquoi importer les données dans un autre logiciel ? Excel couplé avec la puissance de XLSTAT vous permet de faire toutes vos analyses directement dans Excel.
Venez participer à notre formation XLSTAT !

Stat4decision est partenaire officiel de Luminero-XLSTAT

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-xlstat/

Durée

3.00 jours (21.00 heures)

Profils des stagiaires et prérequis

Profils :

- Analyse
- Utilisateur d'Excel
- Expert BI

Prérequis :

- Avoir déjà utilisé Microsoft 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

Programme détaillé

- Introduction à XLSTAT-Base
- La préparation des données avec Excel et XLSTAT-Base (transformation, traitement des données manquantes, filtrage…)
- Cas du traitement d'une enquête
- La description des données avec XLSTAT
- Les intervalles de confiance et les tests statistiques

- Les tests de comparaison de moyennes
- Les tests paramétriques et non paramétriques
- Cas pratiques : applications sur des données réelles
- Présentation de l'analyse de données multivariée
- L'analyse en composantes principales
- Les autres méthodes d'analyse factorielles (AFC, ACM, analyse discriminantes…)
- Présentation des méthodes de clustering (apprentissage non supervisé)
- La classification ascendante hiérarchique
- Les k-means
- Les principes de la modélisation
- Le modèle linéaire : régression, analyse de la variance
- La régression logistique
- Les méthodes de machine learning (arbres, SVM…)
- Aller plus loin avec XLSTAT
    - Utilisation de la programmation pour automatiser vos analyses
    - XLSTAT-R : utilisation de fonctions R dans XLSTAT

Approche PLS (PLS Path Modeling)

Description

En deux jours, maîtrisez l'approche PLS. L'approche PLS ou PLS Path Modeling (ou parfois PLS-SEM) est une méthode d'analyse de données proche des modèles d'équations structurelles à variables latentes. Il s'agit d'estimer les paramètres d'un modèle comprenant des variables observées et des variables latentes.

L'approche PLS (PLS Path Modeling) est une méthode novatrice qui trouve des applications dans de nombreux domaines notamment l'analyse de la satisfaction des consommateurs.

Cette formation vous permettra d'appréhender l'approche PLS de manière détaillée et d'apprendre à l'utiliser dans un contexte opérationnel.
Les applications de cette formation se font avec le logiciel XLSTAT mais d'autres outils peuvent être utilisés en formation intra (R, SmartPLS…)

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-approche-pls/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Statisticien
- Analyste

Prérequis :

- Avoir de bonnes connaissances en analyse de données

Objectifs pédagogiques

- Comprendre les principes de l'approche PLS pour analyser des modèles à variables latentes
- Savoir utiliser les outils permettant d'appliquer l'approche PLS
- Être capable d'interpréter les résultats d'une analyse par l'approche PLS

Programme détaillé

- Principes et concepts des modèles à variables latentes
- L'approche PLS (PLS Path Modeling) – Algorithme, validation et interprétation
- Introduction aux outils pour l'approche PLS
- Cas pratiques :

- Application à l'analyse de la satisfaction
- Application aux sciences sociales
- Autres applications

- Utilisation avancée de l'approche PLS :
  - Effets médiateurs
  - Effets modérateurs
  - Comparaisons multi-groupes
  - Segmentation
- Discussions d'applications sur les données des participants

Les équations structurelles à variables latentes

Description

Maîtrisez le traitement des modèles à équations structurelles à variables latentes.

Cette formation vous donne tous les outils pour bien appréhender ce type de modèles en présentant les concepts importants que sont les variables latentes, les notions de causalité…
Vous apprendrez aussi à utiliser ce type de modèles avec des applications pratiques avec le logiciel SPSS-AMOS ou avec R.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-equations-structurelles/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Statisticiens
- Analystes
- Chercheurs en sciences sociales
- Spécialistes marketing

Prérequis :

- Avoir de bonnes bases en statistique

Objectifs pédagogiques

- Comprendre les principes théoriques des modèles à équations structurelles
- Savoir se servir des outils permettant d'analyser un modèle à équations structurelles
- Pouvoir interpréter les résultats d'un modèle à variables latentes

Programme détaillé

- Les concepts pour l'analyse des modèles d'équations structurelles à variables latentes
    - Variables latentes / variables manifestes
    - Modèle de mesure / modèle structurel
    - Variables exogènes / variables endogènes
- Hypothèses et méthodologie
    - Les méthodes d'estimation et leurs hypothèses
- Validation et tests
- Application avec des logiciels adaptés

- Interprétation des résultats par l'exemple
- Aller plus loin
  - Effets modérateurs, effets médiateurs, comparaison de modèles…
- En fonction des apprenants, les thèmes abordés pourront évoluer (autres domaines d'application, présentation de concepts plus avancés…).

Régression PLS

Description

En une journée, maîtrisez l'utilisation de la régression PLS ainsi que celle de l'analyse discriminante PLS (PLS-DA) pour modéliser vos données et apprenez à interpréter les résultats de vos analyses avec le logiciel XLSTAT.
La régression PLS ou régression par les moindres carrés partiels est une méthode d'analyse de données extrêmement puissantes permettant de remplacer la régression linéaire ordinaire dans de nombreux cas. Elle s'applique lorsqu'il existe une forte multicolinéarité entre les variables explicatives, lorsque le nombre de colonnes est plus grand que le nombre de lignes ou encore lorsque des données manquantes sont présentes.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-regression-pls/

Durée

1.00 jours (7.00 heures)

Profils des stagiaires et prérequis

Profils :

- Statisticiens
- Analystes
- Data scientists

Prérequis :

- Avoir des connaissances de base en statistique et en analyse de données

Objectifs pédagogiques

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

Programme détaillé

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

Analyse conjointe (méthodes de trade-off)

Description

Vous voulez comprendre les préférences de vos clients afin de leur proposer le produit le plus adapté à leurs attentes ?
L'analyse conjointe vous permet de comprendre et de trouver le produit le plus adapté à vos clients.

Durant ces deux jours, vous apprenez à mener une analyse conjointe (ou un trade-off) avec XLSTAT afin d'identifier les préférences des consommateurs, de simuler un marché et de proposer de nouveaux produits.
L'analyse conjointe se base sur un ensemble d'outils statistiques du plan d'expérience aux méthodes d'analyse de choix en passant par des méthodes de simulation, nous aborderons chacun de ces points afin de maîtriser le processus de choix de vos clients.

PLUS D'INFORMATIONS :
https://www.stat4decision.com/fr/formations/formation-analyse-conjointe/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Chargé d'étude
- Responsable marketing
- Public intéressé par les applications marketing

Prérequis :

- Aucun pré-requis

Objectifs pédagogiques

- Comprendre les principes de l'analyse conjointe
- Savoir 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 savoir lire et interpréter les résultats obtenus

Programme détaillé

- Les préférences des consommateurs : principes et notions
- Modéliser les préférences de vos client. Comment y arriver ?
- Mise en place d'une analyse conjointe
    - Choix du type d'analyse

- Choix des attributs et des modalités
- Description de toutes les étapes de l'analyse conjointe

- Présentation de la l'analyse conjointe avec XLSTAT
- Cas pratique : La génération des produits
  - Les plans d'expériences
  - Le cas d'un processus de classement
  - Le cas d'un processus de choix
- Cas pratique : Le recueil des données
- Cas pratique : L'interprétation des préférences
  - La notion d'utilité et d'importance
  - Les modèles
  - L'interprétation
- Cas pratique : Construction et simulation de marché :
  - Principes
  - Méthode de simulation
  - Interprétation des simulations
- Aller plus loin
  - La méthode MaxDiff
  - L'analyse TURF
- Discussions et évaluation de la formation

XLSTAT-R et programmation avec XLSTAT

Description

Le logiciel XLSTAT est un outil pour faire de l'analyse de données directement dans Microsoft Excel. Si vous voulez aller plus loin avec XLSTAT, XLSTAT propose deux approches extrêmement puissantes pour accéder à de nouvelles méthodes et pour automatiser vos analyses : XLSTAT-R et l'utilisation du VBA.
Stat4decision est partenaire officiel de Luminero-XLSTAT.

Plus d'informations :
https://www.stat4decision.com/fr/formations/formation-programmation-avec-xlstat/

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Statisticien
- Analyste

Prérequis :

- Ayant suivi la formation introductive à XLSTAT ou ayant un bon niveau d'utilisation d'XLSTAT.
- Bonnes connaissances d'Excel
- Pas de pré-requis sur R ou le langage VBA

Objectifs pédagogiques

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

Programme détaillé

- Jour 1 : XLSTAT-R
  - Présentation des principes de XLSTAT-R
  - Utilisation de XLSTAT-R
  - Automatiser des scripts R depuis XLSTAT
  - Créer de nouvelles applications dans Excel en utilisant XLSTAT et R
- Jour 2 : Automatisation des traitements
  - Quelques bases de VBA
  - La syntaxe, les objets, le débogage

- Les macros, les fonctions
- XLSTAT et le VBA : Les appels loadrun pour l'appel de fonctionnalités XLSTAT
- La récupération des sorties
- L'automatisation des appels
- Vers la création d'une interface utilisant les fonctionnalités d'XLSTAT

Stat4decision

Data science company

python™

data iku

scikit learn

R

APACHE Spark™

jupyter

www.stat4decision.com