

Durée de la formation ?

2 jours – soit 14 heures.

À qui s'adresse cette formation ?

Aux actuaire et data scientists, informaticiens qui gèrent les bases de données, managers et directeurs responsables de la qualité des données.

Pour obtenir quoi ?

Comprendre différents types d'anomalies et de risques liés à la qualité des données ;
Utiliser des techniques data science pour contrôler la qualité des données ;
Comment le machine learning peut aider à prédire la qualité des données ;
Comment contrôler automatiquement la qualité des données.

Quels objectifs pédagogiques ?

Analyser les problématiques de qualité de données.
Présenter les enjeux data science pour la qualité des données.

Développer des techniques avancées de correction d'anomalies.

Maîtriser les techniques prédictives et processus de correction.

Quelles méthodes mobilisées ?

Par des études de cas et cas pratiques avec « R ».

Quels sont les prérequis ?

Aucun. Cependant la connaissance du langage de programmation « R » peut aider dans les cas pratiques.

Quelles modalités d'évaluation ?

Une évaluation des acquis des objectifs sera réalisée durant la formation.

Qui anime cette formation ?

Kezhan SHI

Membre certifié de l'IA. Il est diplômé de l'École Centrale Paris et titulaire d'un master en actuariat de l'Université Paris Dauphine. Il a travaillé chez Axa Global Direct et Allianz, avant de rejoindre Capgemini en 2019, au titre de Manager Data Scientist.



La formation en pratique...

Quand et où ?

17 et 18 mai 2021

9 h 00 - 12 h 30 et 14 h 00 - 17 h 30

Caritat, Paris 8°

Combien ça coûte ?

2 100 € HT + TVA 20%, soit 2 520 € TTC.

Les frais de participation couvrent les journées de formation, la documentation complète, les déjeuners et les pauses café.

Qu'allez-vous apprendre ?

Objectifs et enjeux de la qualité des données

- Types d'anomalies (erreurs de saisies, fautes d'orthographe, erreurs intentionnelles, défaillance du système, etc.)
- Différents types de risques et d'impacts financiers

Bonnes pratiques et organisation des données

- Comment mieux organiser les fichiers
- Comment organiser les données
- Types de fichiers et de données

Principes de traitements d'anomalies

- Comment explorer les données pour détecter les valeurs aberrantes
- Comment traiter les anomalies des variables numériques
- Comment traiter les anomalies des variables catégoriques
- Comment traiter les anomalies des variables textuelles
 - Identifier des clients doublons
 - Identifier les anomalies des adresses
 - Incohérences par rapport aux clauses des contrats
- Calculer un score de vraisemblance

Techniques de data science mises en œuvre

- Découvrir comment les techniques de data science permettent de contrôler la qualité des données
 - Webscraping
 - Text-mining
 - Machine learning
 - Analyse des séries temporelles
- Comment les méthodes clustering permettent de détecter des valeurs aberrantes
- Comment les méthodes classifications permettent de calculer un score de qualité

Cas d'application

- Vérification textuelle (noms clients, adresses, clauses contrats, et autres saisies manuelles de textes)
- Incohérence de catégories
- Règles de calculs déterministes
- Règles multi-critères
- Segmentation des indicateurs multiples

01 44 51 04 00
info@caritat.fr

! Chaque participant se munira d'un ordinateur portable pour les travaux pratiques.

Qu'en disent les stagiaires ?

Cette formation est une nouveauté du catalogue Caritat.