

### Durée de la formation ?

2 jours – soit 14 heures.

### À qui s'adresse cette formation ?

Aux actuaires et data scientists, informaticiens qui gèrent les bases de données, 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 contrôler automatiquement la qualité des données.

Analyser les besoins d'un géocodeur et présenter les données externes pour enrichir les données internes

### Quels objectifs pédagogiques ?

**Analyser** la qualité des données.

**Développer** des techniques avancées de correction d'anomalies.

**Maîtriser et intégrer** les données géographiques externes.

### Quelles méthodes mobilisées ?

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

### Quels sont les prérequis ?

La connaissance du langage de programmation « R ».

### Quelles modalités d'évaluation ?

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

### Qui anime cette formation ?

#### Kezhan SHI,

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 Diot Siaci en 2022, au titre de Responsable du Data Lab.



### La formation en pratique...

#### Quand et où ?

**20 et 21 novembre 2023**

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

Caritat, Paris 8<sup>e</sup>

#### Combien ça coûte ?

2 300 € HT + TVA 20%, soit 2 760 € 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
- Types de risques et d'impacts financiers

#### Bonnes pratiques et organisation des données

- Types de fichiers et de données
- Organisation des fichiers et des données.

#### Principes de traitements d'anomalies

- Explorer les données pour détecter les valeurs aberrantes
- Traiter les anomalies des variables numériques, des variables catégoriques, des variables textuelles.
- Calculer un score de vraisemblance

#### Techniques de data science mises en œuvre

- Webscraping, text-mining, Machine learning, séries temporelles.
- Clustering pour détecter des valeurs aberrantes
- Méthodes de classifications pour calculer un score de qualité

#### Cas d'application

- Vérification textuelle
- Incohérence de catégories
- Règles de calculs déterministes
- Règles multi-critères
- Segmentation des indicateurs multiples

#### Manipulation de données géographiques

- Sources des données géographiques, format des données
- Cartes thématiques, cartes avec fonds divers, cartes interactives
- Segmentation des données et zoniers
- Utilisation dans la classification et la régression

#### Construction d'un géocodeur

- Principe de géocodage
- Programme d'un géocodeur
- Déploiement de la solution

#### Étude de cas

- Étude de densité et de concentration
- Zones inondables et assurés sinistrés
- Géolocalisation des accidents automobiles
- Étude des courtiers et agents

01 44 51 04 00  
[info@caritat.fr](mailto: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.