

## À qui s'adresse cette formation ?

Aux actuaires 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.

## Comment ?

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.

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

## 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 (**un petit exemple**)
  - incohérences par rapport aux clauses des contrats
- Calculer un score de vraisemblance

### Techniques de data science mises en oeuvre

- 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

## Qui anime cette formation ?

Kezhan SHI est membre certifié de l'IA. Il est titulaire d'un master en actuariat de l'Université Paris Dauphine. Il a également fait l'École Centrale Paris. Il a travaillé chez Prim'Act, Direct Assurance et Aviva, avant de rejoindre Allianz en 2017, au titre de Data Scientist.



## La formation en pratique...

### Quand et où ?

**17 et 18 septembre 2018**

9 h 00 - 12 h 30 et 14 h 00 - 17 h 30  
Caritat, Paris 8<sup>e</sup>

### Combien ça coûte ?

2 000 € HT + TVA 20%, soit 2 400 € TTC.  
Les frais de participation couvrent les deux journées de formation, la documentation complète, les deux déjeuners et les pauses café.

### Où vous loger ?

Si vous venez de province ou de l'étranger, pour toute réservation d'hôtel, contactez notre partenaire :

Elysées West Hôtel - 01 85 34 72 00

Précisez que vous venez de la part de Caritat.

<http://elysees.hotusa.com/caritat/>

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

## Qu'en disent les stagiaires ?

Cette formation est une nouveauté du catalogue Caritat.