

Stage pratique de 4 jour(s)  
Réf : BDA

## Participants

Responsables Infocentre  
(Datamining, Marketing,  
Qualité...), utilisateurs et  
gestionnaires métiers de  
bases de données.

## Pré-requis

Connaissances de base en  
statistiques ou avoir suivi le  
stages "Statistiques, maîtriser  
les fondamentaux" (Réf. STA).  
Connaissances de base en  
Python.

Prix 2021 : 2710€ HT

## Dates des sessions

### AIX

03 août. 2021, 05 oct. 2021  
02 nov. 2021

### ANGERS

10 août. 2021, 02 nov. 2021

### BORDEAUX

24 août. 2021, 19 oct. 2021  
02 nov. 2021

### BREST

10 août. 2021, 02 nov. 2021

### BRUXELLES

03 août. 2021, 19 oct. 2021  
02 nov. 2021

### CLASSE A DISTANCE

25 mai 2021, 15 juin 2021  
20 juil. 2021, 03 août. 2021  
14 sep. 2021, 05&19 oct.  
2021  
02&30 nov. 2021

### CLERMONT-FERRAND

17 août. 2021, 02 nov. 2021

### DIJON

17 août. 2021, 07 déc. 2021

### GENÈVE

15 juin 2021, 14 sep. 2021  
02 nov. 2021

### GRENOBLE

24 août. 2021, 26 oct. 2021  
02 nov. 2021

### LILLE

24 août. 2021, 19 oct. 2021  
02 nov. 2021

### LIMOGES

17 août. 2021, 07 déc. 2021

### LUXEMBOURG

07 sep. 2021, 12 oct. 2021

### LYON

17 août. 2021, 07 sep. 2021  
02 nov. 2021

### MONTPELLIER

10&31 août. 2021, 02 nov.  
2021

### NANCY

03 août. 2021, 19 oct. 2021

### NANTES

03 août. 2021, 19 oct. 2021  
02 nov. 2021

### NIORT

# Big Data Analytics avec Python

## modélisation et représentation des données

*Le Big Data Analytics repose sur la maîtrise des techniques d'exploration de données fondamentales : statistiques descriptives, prédictives ou exploratoires. Ce stage pratique vous présentera des méthodes telles que les régressions et les ACP et vous apprendra à les mettre en œuvre avec le logiciel Python.*

## OBJECTIFS PEDAGOGIQUES

Comprendre le principe de la modélisation statistique  
Choisir entre la régression et la classification en fonction du type de données  
Évaluer les performances prédictives d'un algorithme  
Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances

### 1) Introduction à la modélisation

### 2) Procédures d'évaluation de modèles

### 3) Les algorithmes supervisés

### 4) Les algorithmes non supervisés

### 5) Analyse en composantes

### 6) Analyse de données textuelles

## Travaux pratiques

*Développement/réalisation d'analyses sur le logiciel Python, avec les modules pandas, NumPy, SciPy, Matplotlib, seaborn, scikit-learn et statsmodels.*

## 1) Introduction à la modélisation

- Introduction au langage Python.
- Introduction au logiciel Jupiter Notebook.
- Les étapes de construction d'un modèle.
- Les algorithmes supervisés et non supervisés.
- Le choix entre la régression et la classification.

### Travaux pratiques

*Installation de Python 3, d'Anaconda et de Jupiter Notebook.*

## 2) Procédures d'évaluation de modèles

- Les techniques de ré-échantillonnage en jeu d'apprentissage, de validation et de test.
- Test de représentativité des données d'apprentissage.
- Mesures de performance des modèles prédictifs.
- Matrice de confusion, de coût et la courbe ROC et AUC.

### Travaux pratiques

*Mise en place d'échantillonnage de jeux de données. Effectuer des tests d'évaluations sur plusieurs modèles fournis.*

## 3) Les algorithmes supervisés

- Le principe de régression linéaire univariée.
- La régression multivariée.
- La régression polynomiale.
- La régression régularisée.
- Le Naive Bayes.
- La régression logistique.

### Travaux pratiques

*Mise en œuvre des régressions et des classifications sur plusieurs types de données.*

## 4) Les algorithmes non supervisés

- Le clustering hiérarchique.
- Le clustering non hiérarchique.
- Les approches mixtes.

### Travaux pratiques

*Traitements de clustering non supervisés sur plusieurs jeux de données.*

## 5) Analyse en composantes

- Analyse en composantes principales.
- Analyse factorielle des correspondances.
- Analyse des correspondances multiples.
- Analyse factorielle pour données mixtes.
- Classification hiérarchique sur composantes principales.

### Travaux pratiques

10 août. 2021, 02 nov. 2021

### ORLÉANS

03 août. 2021, 05 oct. 2021  
02 nov. 2021

### PARIS

25 mai 2021, 15 juin 2021  
20 juil. 2021, 03 août. 2021  
14 sep. 2021, 05 oct. 2021  
02&30 nov. 2021

### REIMS

03 août. 2021, 19 oct. 2021

### RENNES

10 août. 2021, 05 oct. 2021  
02 nov. 2021

### ROUEN

17 août. 2021, 19 oct. 2021

### SOPHIA-ANTIPOLIS

03 août. 2021, 05 oct. 2021  
02 nov. 2021

### STRASBOURG

24 août. 2021, 19 oct. 2021  
02 nov. 2021

### TOULON

17 août. 2021, 19 oct. 2021

### TOULOUSE

03 août. 2021, 19 oct. 2021  
02 nov. 2021

### TOURS

03 août. 2021, 05 oct. 2021  
02 nov. 2021

*Mise en œuvre de la diminution du nombre des variables et identification des facteurs sous-jacents des dimensions associées à une variabilité importante.*

## 6) Analyse de données textuelles

- Collecte et prétraitement des données textuelles.
- Extraction d'entités primaires, d'entités nommées et résolution référentielle.
- Étiquetage grammatical, analyse syntaxique, analyse sémantique.
- Lemmatisation.
- Représentation vectorielle des textes.
- Pondération TF-IDF.
- Word2Vec.

### Travaux pratiques

*Explorer le contenu d'une base de textes en utilisant l'analyse sémantique latente.*

## Modalités d'évaluation

L'évaluation des acquis se fait tout au long de la session au travers des multiples exercices à réaliser (50 à 70% du temps).

## Compétences du formateur

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

## Moyens pédagogiques et techniques

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour

les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.

- A l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.

- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.