

**Le diplôme universitaire (DU) Data Analyst est une formation professionnelle visant à former les Data Analysts de demain. Entièrement en distanciel, son rythme flexible est conçu pour s'adapter aux impératifs d'un public d'actifs.**

## Présentation

```
/**/ .actions-fiche__item:first-of-type { margin-top: 0; display: none; } /**/
```

### Enjeux

Le Big data, et plus généralement l'analyse de données, occupent une place de plus en plus importante au sein des stratégies de nombreuses organisations. Suivi de performance, analyse des comportements, découvertes de nouvelles opportunités de marché : les applications sont multiples, et intéressent des secteurs variés. Du e-commerce à la finance, en passant par les transports et la santé, les entreprises ont besoin de talents formés à la collecte, au stockage, mais aussi au traitement et à la modélisation des données.

### Spécificités

Le diplôme obtenu est un diplôme universitaire (DU) CY Cergy Paris Université. Gage d'une formation d'excellence reconnue sur le marché du travail, un diplôme universitaire ne peut être délivré que par une université française.

## Durée de la formation

10 mois

## Lieu(x) de la formation

100% distanciel

## Stage(s)

Non

## Rythme

- Compatible avec une activité professionnelle

## Modalités

- Totalement à distance

## Renseignements

[du-data-analyst@cyu.fr](mailto:du-data-analyst@cyu.fr)

Sélectif

Oui

## Admission

### Pré-requis

#### Formation(s) requise(s)

- Maîtrise des bases de Python ou R : les objets, les listes, les fonctions et packages, les boucles for et while, les filtres et opérateurs logiques. Ce qui correspond :
  - Pour Python, au cours Introduction to Python et aux chapitres 3 et 4 de Intermediate Python sur Datacamp

- Pour R, au cours Introduction to R et aux chapitres 1 à 3 de Intermediate R du Datacamp
- Maîtrise suffisante de l'anglais (niveau B1) pour pouvoir comprendre et effectuer des exercices de code en anglais. Aucun niveau d'expression orale n'est requis.
- L'accès à un ordinateur muni d'un micro et d'une webcam, ainsi que d'une connexion internet suffisante.

*Un test de positionnement sera soumis au candidat durant sa procédure d'admission. En cas d'une nécessité d'une remise à niveau, des ressources externes seront proposées, à valider avant inscription au parcours.*

## Candidature

### Modalités de candidature

**Prochaine session de formation : 27 février 2024.**

1. Pour candidater, merci de compléter le [dossier de candidature disponible ici](#).  
**Clôture de la phase de candidatures au 21/12/2023.**
2. Vous serez informés des résultats de la première pré-sélection sur dossier à partir du 15/01/2024.
3. Une fois votre dossier pré-sélectionné, effectuez le test de positionnement (R ou Python au choix) qui vous sera transmis via la plateforme Datacamp\*.
4. Rencontrez individuellement le responsable de formation lors d'un entretien en ligne pour mieux vous connaître et vérifier l'adéquation du programme avec vos attentes et vos compétences.
5. Validez votre dossier de financement et inscrivez-vous administrativement auprès de la gestionnaire pédagogique du DU.

\*Un accès temporaire gratuit vous sera fourni par l'université.

## Et après ?

### Niveau de sortie

#### Niveau de sortie

- Diplôme université niv. form. bac +4

# Programme

Au cours de cette formation, vous travaillerez à partir des bases d'un langage de programmation (R ou Python) sur la visualisation de données, la maîtrise des tests statistiques et les techniques de base du machine learning et du deep learning. En plus des compétences techniques, vous serez formé à l'interprétation et à l'analyse critique des données.

**À la fin de cette formation, vous aurez acquis le niveau d'autonomie suffisant pour être opérationnel sur un poste de Data Analyst.**

## Blocs de compétences

BDC 1 : Collecter, organiser et nettoyer des données

### Module 1.1 : Manipulation et pré-traitement de données :

- Savoir charger et décrire un jeu de données à travers R ou Python
- Savoir manipuler une base de données relationnelles (SQL)
- Savoir traiter des valeurs manquantes ou aberrantes
- Manipuler et représenter la distribution de variables de différents types (continues, binaires, dates, etc.)

BDC 2 : Réaliser l'analyse exploratoire d'un jeu de données

### Module 2.1 : Réduction de dimensionnalité et algorithmes de clustering :

- Réaliser une analyse multivariée de variables continues (PCA) ou catégorielles (MCA)
- Appliquer une méthode d'apprentissage non supervisé pour partitionner des données (k-means, CAH)

### Module 2.2 : Techniques de visualisation de données

- Pouvoir réaliser avec R ou Python les principaux types de graphiques
- Savoir identifier quel type de graphique est le plus adapté à une situation donnée

BDC 3 : Identifier et mettre en œuvre des modèles adaptés à une problématique donnée

### Module 3.1 : Introduction aux statistiques

- Savoir mettre en œuvre des analyses statistiques simples (Student, ANOVA, Régression logistique)
- Identifier les situations pour lesquels des tests non-paramétriques sont requis
- Savoir transformer des données pour retomber dans les conditions de validité d'un test

### Module 3.2 : Introduction au Machine Learning

- Appliquer une méthode d'apprentissage supervisé (forêts aléatoires) pour réaliser une tâche de classification
- Utiliser des métriques (AUC, OOB) pour estimer la qualité d'un modèle

BDC 4 : Communiquer sur une analyse de données

### Module 4.1 : Initiation à la Business Intelligence

- Concevoir le cahier des charges d'un tableau de bord
- Maîtriser les différentes fonctionnalités de Power BI pour pouvoir mettre en place un tableau de bord

### Module 4.2 : Rédaction d'un rapport d'analyse

- Explorer un jeu de données en autonomie pour mettre en œuvre les différentes techniques maîtrisées
- Produire un document en Latex
- Structurer de manière logique la restitution des résultats des analyses

