

Data Scientist, Bootcamp (3 months)

by DataScientest

Practical course - 57d - 399h00 - Ref. 4IK

Price : 7190 € E.T.

Devenez expert en développement de modèles d'analyse de données avec ce parcours proposé par notre partenaire DataScientest. Un Data Scientist développe des modèles d'analyse complexes visant à extraire des informations à partir de bases de données. Cette formation certifiante se déroule à distance dans un format hybride mêlant temps d'échanges synchrones avec un formateur expert, exercices pratiques et modules E-learning. Basée sur la pédagogie Learning By Doing, vous réaliserez un projet fil rouge en équipe afin de mettre en pratique vos connaissances. Lors de votre inscription, vous serez rattaché à l'une des promotions DataScientest. A l'issue de cette formation, vous obtiendrez un certificat « Data Scientist » des Mines Paris- PSL Executive ainsi que le bloc de compétence 2 de la certification RNCP « Expert en ingénierie de l'intelligence artificielle ». Contactez-nous dès maintenant pour connaître les prochaines dates !



Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Clean data.
- ✓ Analyze data using visualization and statistical tools and methods.
- ✓ Design predictive artificial intelligence models.

Intended audience

Anyone with an interest in programming and data manipulation.

Prerequisites

Un diplôme ou un titre de niveau bac+3 ainsi qu'un bon niveau en mathématiques.

Pour les candidats ne présentant pas le niveau de qualification requis, une dérogation est possible sur dossier.

PARTICIPANTS

Anyone with an interest in programming and data manipulation.

PREREQUISITES

Un diplôme ou un titre de niveau bac+3 ainsi qu'un bon niveau en mathématiques.

Pour les candidats ne présentant pas le niveau de qualification requis, une dérogation est possible sur dossier.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.

Participants also complete a placement test before and after the course to measure the skills they've developed.

Certification

Pour clôturer la formation, l'équipe pédagogique évaluera le projet fil rouge de l'apprenant à l'aide d'un rapport écrit et d'une soutenance à distance. La validation des compétences développées au cours de la formation Data Scientist vous permettra d'obtenir : • Un certificat de l'École des Mines Paris - PSL Executive Education • Le bloc de compétence 2 de la certification RNCP de niveau 7 "Expert en ingénierie de l'intelligence artificielle" enregistrée au RNCP sous le n°RNCP38587.

Practical details

Digital activities

Online courses and exercises, group masterclasses, question/answer sessions, support classes, e-mail coaching, red thread projects, individualized career coaching, social learning.

Mentoring

Un formateur expert accompagne l'apprenant tout au long de sa formation. Il échange régulièrement avec lui sur son projet fil rouge et l'accompagne lors de points de mentorat (individuel). Plusieurs formateurs animent également les différentes masterclass (classes collectives) et répondent aux questions des apprenants à tout moment depuis un forum dédié. En complément, de nombreuses séances de questions-réponses peuvent être organisées pour aider les apprenants.

Pedagogy and practice

Lors de l'inscription, l'apprenant est affecté à une promotion (dates à définir lors de l'inscription) et reçoit son calendrier de formation. Le parcours de formation est découpé en « Sprint » de plusieurs semaines sur une thématique dédiée. Chaque semaine l'apprenant est convié à un temps d'échange avec le formateur qui se présente sous la forme de masterclass (classe collective) ou de points de mentorat (individuel). Pendant 80% du temps, l'apprenant travaille en autonomie sur la plateforme d'enseignement. Tous les modules intègrent des exercices pratiques permettant de mettre en œuvre les concepts développés en cours. L'apprenant doit également travailler en binôme ou trinôme sur un projet fil rouge tout au long de la formation. Cela lui permettra de développer et faire reconnaître ses compétences. En complément, des événements et ateliers thématiques sont régulièrement proposés pour permettre aux apprenants de découvrir les dernières innovations en matière de Data Science. Afin de suivre efficacement la formation, nous estimons le temps travail nécessaire entre 35 et 40 heures par semaine.

TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

Course schedule

1 Upcoming session dates

- November 2025: Start date 04/11/25
- January 2026: Start date 01/13/26
- March 2026: Start date 03/03/26

2 Python fundamentals

- Programming: variable types, lists, tuples, loops, functions, classes and modules.
- NumPy: NumPy Array, matrix operations, statistical indicator and operations.
- Pandas: introduction to the library, loading a dataset, Data Cleaning, Data Processing.

3 Data Visualization

- Matplotlib: graph types, graph customization.
- Plotly: interactive web graphics and widgets.
- Seaborn: distribution analysis, statistical analysis, multivariate analysis.

4 Programming tools

- Git: Git version management, Git repository, branches, tag and merge.
- GitHub: Github platform, Fork, Pull Request, Issues, sharing modifications, Github Actions.
- Unit testing: Pytest, integration and function tests, benefits of testing.
- Linux systems and Bash scripting: introduction to Linux systems, terminal handling, Bash scripting.

5 Machine Learning

- Classification models and algorithms: Scikit-learn, logistic regression, KNN, SVM, bagging and boosting.
- Clustering methods: unsupervised classification models, evaluation metrics for clustering.
- Regression methods: simple and multiple linear regression, regularized linear regression.

6 Advanced Machine Learning

- Time series with Statsmodels: ARIMA and SARIMA models.
- Anomaly detection: anomaly detection issues, supervised models, Robust Covariance, DBScan.
- Dimension reduction methods: feature selection process, component analysis, Manifold Learning approach.

7 Applied Machine Learning

- Text mining: regular expressions, text data, WordCloud, sentiment analysis.
- Webscraping: web language, BeautifulSoup, Google scraping.
- Ethics & Interpretability: sensitive data, ethical issues, bias analysis, model interpretability.

8 Complex models

- Recommender systems: introduction to the system, content-based methods, recommendation mathematics.
- Reinforcement Learning: Monte-Carlo method, temporal Difference Learning, SARSA and Q-Learning.
- Graph theory: Kruskal and Dijkstra algorithms, community detection, PageRank algorithm, NetworkX.

9 Deep Learning fundamentals

- Convolution networks: convolution image analysis, LeNet, transfer learning.
- TensorFlow: optimization with TensorFlow, callbacks and custom templates, TensorBoard.
- Dense networks: neural networks, Keras framework, optimization of a dense network.

10 Data Engineering

- SQL: database queries with SQL, join types, nested queries.
- PySpark: distributed computing with PySpark, Spark RDD and DataFrame APIs, processing pipeline, MLLib.
- APIs: Introduction to APIs, HTTP methods and functions.

11 MLOps

- MLFlow: Machine Learning project life cycle, Model training, MLflow.
- Docker: introduction to containerization, how Docker works, images and containers, Dockerfile and Docker Hub.