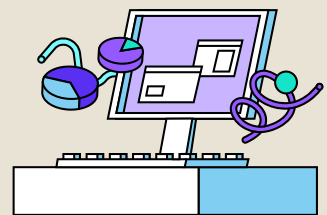




SYLLABUS DATA SCIENTIST



DISCOVER THE DIFFERENT JOBS IN DATA

OUR STORY

DataScientest is the leading training organization in Europe for Data Science. We offer business-oriented training courses for upskilling and reskilling to professionals and individuals alike.

TRAINING DESIGNED WITH AND FOR COMPANIES

Data has become the main resource to exploit for companies. It allows them to **guide their decision-making** and to predict the behaviour of their environment. Therefore, staying **competitive** means exploiting data to its fullest potential. Thus, given that data is everywhere, it is essential to **increase workforce capabilities** on these new skills, tools and technologies to ensure stable development.

This is why DataScientest has been training the Data teams of more than **70 Fortune 500 groups** as well as many other SMEs and Startups for over **6 years**. Since 2021, more than **5500 professionals** have been trained with us.

74%

of the heads of Data in the largest groups wish to strengthen their teams by recruiting Data Scientists*

*Survey conducted among our forty partner companies.

A CURRICULUM ADAPTED TO INDIVIDUALS

Having been designed for our **partner companies**, our training courses are professionally oriented, as close as possible to the **needs of the market**. Thus, in order to face the shortage of Data profiles and bridge this gap, we have decided to open our training courses to **individuals**. Under this impulse, DataScientest is constantly developing thanks to **new training courses** (Data Manager) or **new specialized courses** (Deep Learning).



"DataScientest is a rich, personalized learning experience that responds to the real problems of the company."

Samir Ait Idir
Chief Data Officer & Head of Data Intelligence @Orange Bank

DATASCIENTEST KEY FIGURES

+70

**FORTUNE 500 COMPANIES
AS CLIENTS**

98%

**COMPLETION
RATE**

94%

**SATISFACTION
RATE**

+5500

ALUMNI

2000h

**OF EXCLUSIVE
CONTENT**

THEY TRUST US





OUR PARTNERS

DataScientest has developed partnerships with academic institutions on the one hand and with software publishers on the other. The content of our training courses on Business Intelligence or Cloud tools is henceforth certified by the companies that own them. Here are our different partnerships:



AMAZON WEB SERVICES Technological Partner

Today, DataScientest enjoys the exclusive status of **Amazon Training Partner**. We are therefore authorized by Amazon to train teams on the products and services of the American company. Thanks to this partnership, we have established **several training courses** that prepare for **official AWS certifications**. The registration fee for the official exam is included in the price of the course.

In some of our courses, you will benefit from a day of preparation for an official Amazon certification. For instance, with the **Data Scientist** training course, you will have the opportunity to prepare for the fundamental **AWS certification: Cloud Practitioner**.



MICROSOFT - POWER BI Technological Partner

DataScientest is a **Microsoft Learning Partner**, which means that we can prepare you for **official Microsoft certifications**. These certifications attest to a level of expertise in Azure, the set of cloud computing products and services, and in **Power BI**, Microsoft's business intelligence tool. The registration fee for the official exam is once more included in the price of the course.

In some of its training courses, DataScientest prepares you to pass the fundamental official certifications from **Microsoft: PL-900 or AZ-900**.



L'ÉCOLE DES MINES - PARISTECH EXED Academic Partner

Ecole des **Mines - PSL Executive Education** is a **top-ranked engineering school** with one of the most dynamic Data Science academies in Europe. DataScientest is proud to be an **official partner**, which attests to the **quality of our Data Scientist and Data Engineer programs**, which are eligible for **official certification by the school**.



QUALIOP Quality Criteria

This label recognises that the processes implemented by the training organization meet the quality requirements demanded by the State. **All DataScientest training courses are Qualiopi certified.**

OUR TEAM

OUR TEACHERS

DataScientest's teaching team is made up of internal professors. They are dedicated to teaching and research for our various courses and expert curricula.

All our professors are always there for you. They have created and continue to update all our courses. They will accompany you by video-conference and on our platform throughout your training. With their excellent academic background and varied professional experience, they are the Data Science experts who will enable you to join the Data team in the sector that interests you most (Banking & Insurance, Industry, Finance, Medical and many others).

Charles S.
Head of Academic
8 years of experience



A graduate of the École Polytechnique, Charles has been involved in the development of the courses from day one. He is specialized in programming, Machine Learning and Deep Learning.

Raphael K.
Pedagogical Director
8 years of experience



Holder of the ISF master's degree specializing in Statistical Learning and Data Science from the University of Paris-Dauphine, Raphaël designed the Data Analyst and Data Management courses thanks to his knowledge in programming, dataviz and Machine Learning.

OUR TEAM

Gaspard G.
Data Scientist course director
2 years of experience



Graduate of a Master in Research in Mathematics and a Master in Data Science from the Ecole Normale Supérieure and the University Paris-Dauphine University, Gaspard quickly joined DataScientest. Specialized in Computer Vision and NLP, he actively participated in the development of our expert curricula and has supervised numerous Deep Learning projects. Today, he is in charge of the Data Scientist training.

Thomas B.
Deep Learning course director
6 years of experience



After attending a top tier engineering school, Thomas quickly entered DataScientest's ranks. His array of knowledge stretches from programming and Dataviz to NLP, Deep Learning and Computer vision.



DANIEL

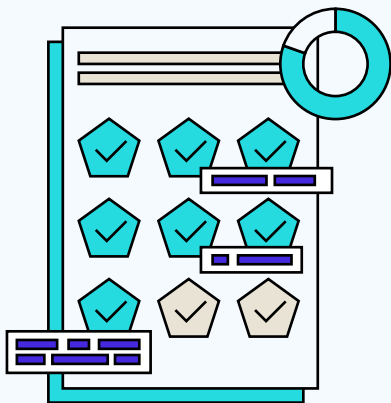
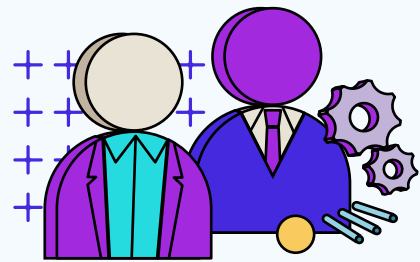
During your training, you will progress on our platform. However, **you will never be alone**, Daniel is there for you. All our teachers and Program Managers take turns to **answer all your questions live via Slack** and on a dedicated support platform. With support every day of the week, from 9am to 5pm, you are guided throughout your course.

OUR PEDAGOGY

HYBRID FORMAT

DataScientest offers 100% distance learning in a hybrid format :

During approximately 20% of the duration, you are accompanied by your teacher and your cohort during **coaching sessions via videoconference**. The rest of the time, you work on our teaching platform and are **guided** via Slack by our **Data Scientists**. This format gives you the flexibility to organize yourself. We also monitor your progress and accompany you to ensure the **successful completion** of your training.



MASTERCLASS

Each training period on the platform (called a Sprint) is accompanied by one or more **masterclasses by videoconference**. The objectives and teaching formats of each masterclass change according to the sprint: correction of concrete use cases, lectures on specific topics, competitions between cohorts, etc.

EXAMENS

At DataScientest, there are no automated MCQs. **Each paper is reviewed and corrected** by hand by our teachers. They go over all your difficulties with you.

TRAINING LEADING TO CERTIFICATION - STATE RECOGNITION

The validation of the skills developed during our Data Scientist training will allow you to obtain a certificate from **Mines ParisTech PSL Executive Education** certificate and to validate the 3 of the RNCP 36129 certification "**Project Manager in Artificial Intelligence**" recognized by the recognized by the State and constituting a strong signal on the the job market.



BACKBONE PROJECT

From the beginning of your training, you will carry out a concrete project with the aim to deploy it. It requires an investment of about 120 hours of work throughout the training.



You will work in pairs or in triads on a project that you will select from our regularly updated **project catalogue**. Our subjects are inspired by the work we do with companies.

This is a crucial step in your career path that will make you fully **operational**. You will be working on un-cleaned data sets, and will be expected to produce work of professional quality.

Reviews are done at regular intervals by your teacher to guide you, and **coach you**. This allows you to move efficiently from theory to practice and to ensure that you master the skills required on the different modules.

It is also a project that is **highly valued by companies**. It confirms your skills and knowledge acquired at the end of the **Data Scientist** course. You will then be able to justify your skills with a **successful Data Science project** during your interviews



DataScientest is a rich, personalized course, perfectly adapted to today's corporate data science topics.

Xavier Bocher
Head of Credit Risk Internal Models & Operational Research @Groupe Crédit Agricole

OUR PEDAGOGY

PREREQUISITES

In order to join the Data Scientist training course, it is necessary to have obtained a bachelor's degree in mathematics, statistics or science (European level 6). You must also have good communication skills.

To understand and know the necessary pre-requisites, [discover this article !](#)

CONTINUOUS FORMAT

Only one format is available for the Data Scientist training :



CONTINUOUS FORMAT

Flexible, it is the format adapted for those that work a full-time profession.



DURATION 9 months

RYTHM 10-12h / week

THE CURRICULUM

Data Scientist

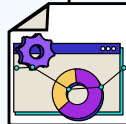
1 PROGRAMMING

 **Fundamentals**
Numpy
 Pandas






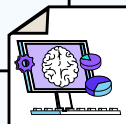
2 DATAVIZ

Matplotlib 
Seaborn 
Bokeh 





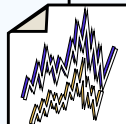
3 MACHINE LEARNING

 **Classification methods and algorithms**
 **Advanced models of classification**
 **Clustering methods**
Regression models





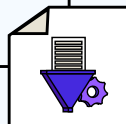
4 ADVANCED MACHINE LEARNING

Time series with Statsmodels 
Text Mining
Machine Learning and graph theory with NetworkX 
Size/dimension reduction methods





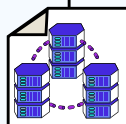
5 DATA ENGINEERING

 **SQL**
Data Processing
 **Machine Learning with PySpark**



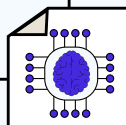
6 DEEP LEARNING

Tensorflow - Keras 
CNN-RNN- Gans 



7 COMPLEX SYSTEMS AND IA

 **Reinforcement Learning**
 **Deep Reinforcement Learning**



PYTHON FUNDAMENTALS - duration 48h

Python Fundamentals

- Discovery of the different variables, lists and Tuples
- Initiation to the concept of a programming loop and its different types
- Introduction to functions and their documentation
- Creation of classes and use of modules

Pandas

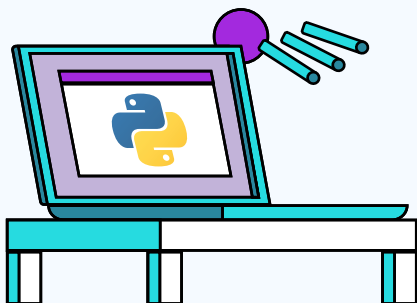
- Introduction to Pandas library
- Loading and first exploration of a dataset
- Introduction to Data Cleaning
- Introduction to Data Processing

NumPy

- Creation and manipulation of a NumPy Array
- Presentation of matrix operations and management of a NumPy Array
- Creation of a statistical indicator and operations on a NumPy Array

Skills acquired at the end

- Reading and understanding a Python code, the most used language in Data Science
- Handling and managing data tables
- Querying, managing, ordering and modifying a dataset with Python
- Mastering the software libraries NumPy and Pandas



DATA VISUALISATION - duration 33h

Matplotlib

- Presentation of different types of graphs :
 - Curves
 - Charts
 - Point clouds
 - Histograms
- Introduction to graph customization

Bokeh (optional)

- Training in all types of interactive graphics that can be integrated into a Web page
- Visualization of geographical data
- Discovery and creation of Widgets

Seaborn

- Control of distribution analysis
- Implementation of statistical analysis
- Introduction to multivariate analysis

Skills acquired at the end

- Control and customization of a large set of graph types, which is fundamental for Data Visualization
- Being able to use Data Visualization for Data Analysis
- Creating simple statistical graphs that mix Data Visualization and Data Analysis
- Mastering of data visualization best practices and data storytelling techniques
- Creating an interactive graph and integrating it to a website



MACHINE LEARNING - duration 45h

Classification models and algorithms

- Introduction to Scikit-learn
- Presentation of classic algorithms: Logistic regression, KNN, SVM...
- Bagging and Boosting techniques

Advanced classification of models

- Selection of models
- Semi-supervised Classification
- Anomaly detection

Clustering methods

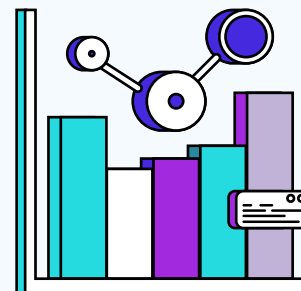
- Unsupervised classification models (K-Means, CAH, Mean Shift...)
- Evaluation metrics for clustering

Regression methods

- Simple and multiple linear regression
- Regularized linear regression

Skills acquired at the end

- Mastery of the main Machine Learning algorithms
- Preparing data for modeling and prediction
- Identifying the right algorithm for a given problem
- Knowing and mastering the evaluation metrics of Machine Learning algorithms



ADVANCED MACHINE LEARNING - duration 52h

Time Series with Statsmodels

- Discovery of ARIMA and SARIMA models
- Analysis and decomposition of a temporal signal

Text mining

- Introduction to regular expressions
- Management of text data
- Creation of a WordCloud

Machine Learning and Graph Theory with Network X (optional)

- Introduction to graph theory
- Application of fundamental algorithms: Kruskal and Dijkstra
- Detection of communities
- Application of the PageRank algorithm (classify webpages)

Size reduction methods

- Feature selection process
- Introduction to principal component analysis
- Application of the Manifold Learning approach

Skills acquired at the end

- Mastering data management and text data pre-treatment
- Reading and using regular expressions
- Using a Machine Learning model on text data, an area which is continuously growing in Data Analysis
- Understanding the structure of an Internet website
- Automatizing the extraction of data coming from diverse webpages

DATA ENGINEERING - duration 25h

Langage SQL

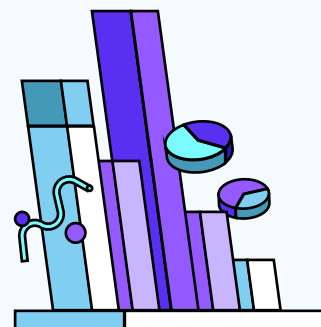
- Discover Relational Databases
- Implementation of SQL queries

PySpark

- Discovery of PySpark's different functionalities :
 - Large databases management
 - Machine Learning
 - Pipelines
 - Optimization

Skills acquired at the end

- Reading and interrogating relational databases
- Mastery of SQL queries' syntax
- Training to treat massive databases thanks to distributed computing
- Applying efficiently Machine Learning models to big databases



DEEP LEARNING - duration 60h

Deep Learning with Keras framework

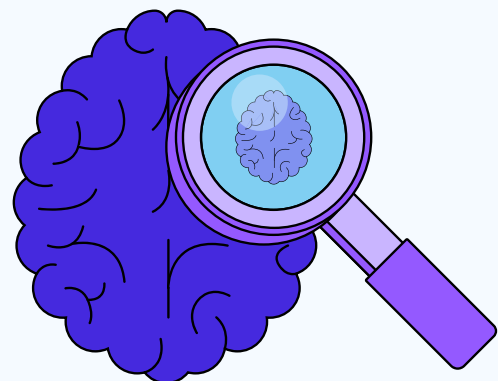
- Discovery of fundamental concepts:
 - Dense Neural Networks
 - Convolutional Neural Networks
 - LeNet Architecture
 - Transfer Learning

TensorFlow

- Linking TensorFlow with Keras
- Application of Word Embedding with Word2vec
- Presentation of the Recurrent Neural Networks (GRU, LSTM...)
- Presentation of the Generative adversarial Network

Skills acquired at the end

- Understanding and implementing neural networks
- Training and evaluating the results obtained by a neural network
- Using the methods of Transfer Learning, very useful in Deep Learning
- Mastery of the syntax and peculiarities of the TensorFlow framework
- Text and image treatment with deep learning (translation, objects detection)



COMPLEX SYSTEM AND AI - durée 22h

Introduction to Reinforcement Learning

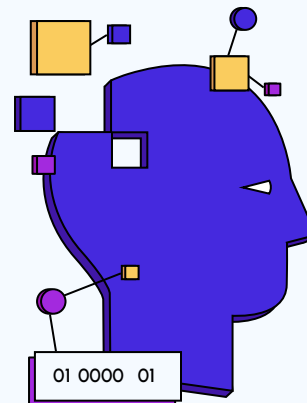
- Development of mathematics for Reinforcement Learning
- Application of the Monte-Carlo method
- Discovery of the Temporal Difference
- Comparison of learnings: SARSA and Q-Learning

Skills acquired at the end

- Understanding the mathematical fundamentals of Reinforcement Learning
- Understanding the main algorithms used in Reinforcement Learning
- Knowing how to choose a Reinforcement Learning algorithm according to the task at hand

Deep Reinforcement Learning

- Presentation of Deep Q Learning
- Introduction to the Policy Gradient method



ALUMNI TESTIMONIES



Karina CASTILLO
Data Expert Consultant at Expertime



I followed the Data Scientist bootcamp, and I must say that even if it is demanding, the teachers are very friendly and always listening, in addition to being very professional. I would also like to underline the fact that they are always up to date with the technologies and topics related to the data business in general. The training combines perfectly both theory and practice throughout each step of the way. One of their strong points is that they understand very well the needs of companies and it shows in the content of their training. Finally, once I finished the training, the career department accompanied me throughout the year by giving me advice for interviews, my CV, salary ranges and even proposed me as a candidate for a job offer which means I was able to land a job as a full-time Data Expert Consultant. I fully recommend DataScientest's training courses.



Cyrille Chemama



I went through the Data Scientist training in bootcamp mode in March 2021 and despite my amateurism in the field of computer science and programming languages, the curriculum is very educational and covers all the concepts necessary for the position of data scientist: from learning Python to data acquisition, especially via web scraping, to the most advanced skills in machine learning, deep learning and reinforcement learning. Datascientest's learning platform allows you to work at your own pace and even in case of difficulty, the support and hotline are always available to help you out. Glad I signed up!

ALUMNI TESTIMONIES



Thomas GOSSART



I followed the Data Scientist course in bootcamp, and I must say that even if it is demanding, it is complete and very well explained, especially for a novice in Data Science. Thanks to this course, I noticed an extremely rapid rise in competence. Many means of support are available to answer our questions, which is why I warmly recommend this training course, even to total beginners because the personal investment is worth it..



Emeline SILVESTRE



Datascientest allows me today to realize my dream, to become a Data Scientist!

The continuous training model allows you to progress at your own pace and whenever I wanted. The courses available on the online platform are very educational. You gain in skill little by little on complex concepts. We have access to many resources to help us, a forum, an instant messaging system and the possibility of making regular calls with the supervisors. We also have a class with whom we share, in particular, the project. In addition to the training itself, many events are organized, from meditation sessions to CV workshops, as well as data workshops, which are very enriching.

For me, with their learning method, this is the best way to combine fun and effectiveness. You learn thanks to concrete examples, through experiences and mistakes.

TO GO FURTHER

If you wish to strengthen your skills, DataScientest has a variety of training courses on publisher certifications such as Microsoft or AWS that enable you to deepen your knowledge and improve your skills in Data!

MLOPS

If you want to acquire advanced programming skills, know how to put models into production and be independent in deployment, this training is for you !

DEEP LEARNING

Are you interested in how Siri works? The creation of autonomous cars? This course is for you! Through two specialization tracks, you will learn all about language processing with Natural Language Processing (NLP), or about image processing with Computer Vision.

Following the Cloud Practitioner certification, you can specialize by preparing for a more technical AWS certification or train on Azure :

ARCHITECTING ON AWS

If you want to know how to make architectural decisions in accordance with AWS, this is the certification for you! Earn the status of «**AWS Certified Solutions Architect Associate**».

AZURE DATA ENGINEER

If you want to improve on the AZ-900 training and know how to implement solutions on Azure in an enterprise, this training is for you! At the end of the training, pass the official Microsoft certification and become an «**Azure Data Engineer Associate**».

AZURE ADMINISTRATOR

If you want to have skills and knowledge about managing storage and virtual networks on Azure, this training is what you are looking for! Train and get the official certification «**Microsoft Certified Azure Administrator Associate**».

DEVELOPPING ON AWS

If you want the ability to write and deploy cloud-based applications, look no further! This course is for you. Upon completion you will be certified as an **AWS Certified Developer Associate**.

POWER BI

Do you want to provide a complete analysis of a dataset and improve your dashboarding skills? This training is for you! Learn to master Power BI and get your official Microsoft certification by becoming a «**Power BI Data Analyst Associate**».



**DO YOU WANT TO BECOME
A DATA SCIENTIST ?**

