SYLLABUS DATA SCIENTIST

Learn how to master the workings behind Artificial Intelligence

DataScientest

SUMMARY

DATASCIENTEST IN NUMBERS

We are DataScientest. Not only are we the market leader in Data Science training in France. We also are the learning institute who offer business-oriented training to professionals and individuals who want to upgrade their skills in the data field.

+70	**Fortune Global 500** Companies trained
+15k	**Alumni**
+3k	Hours of **content**

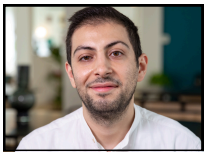
Key figures for the Data Scientist in 2023

723	Number of **subscribers**
93%	**Satisfaction** rate

They trust us

OUR TEAM

DataScientest's teaching team is made exclusively of internal professors. They are fully dedicated to teaching and are consistently updating our various training and expert courses.



Charles S.

CTO & Academic Manager
(9 *years of experience)*

Having graduated of the "École Polytechnique", Charles is specialized in programming, Machine Learning and Deep Learning. As CTO of DataScientest, he leads both the faculty and developers working on the platform.

Raphaël K.

Pedagogical Director
(9 years of experience)

Raphaël has a Master's degree in "Statistical Learning and Data Science" from the University of Paris-Dauphine. He designed the Data Analyst course thanks to his knowledge of programming, data visualization and machine learning.

Lecturers

The Data Scientist course is run by lecturers with degrees from leading universities in France. They are selected for their expertise and teaching skills.

Manon G.

Data Scientist Training Coordinator - 3 years experience

Manon holds a Master's degree in Applied Mathematics and Data Science from the Ecole Normale Supérieure and the Université Paris-Dauphine, and specialises in machine learning. She is currently in charge of Data Scientist training.

Gaspard G.

Data Scientist training referent - 3 years of experience

With a Master's degree in Mathematics and a Master's degree in Data Science from the Ecole Normale Supérieure and the Université Paris-Dauphine he specializes in Computer Vision and NLP. He participated in the development of our expert curricula and supervised many Deep Learning projects and today, he is in charge of the Data Scientist training.

Thomas B.

Deep Learning training referent - 6 years of experience

After his studies at the "École des Mines de Paris", Thomas joined DataScientest. He specialized in Deep Learning and is in charge of our Natural Language Processing and Computer Vision expert courses.

Book an appointment

OUR PEDAGOGICAL APPROACH

DataScientest offers you a course which is a 100% in English and distance learning with a pedagogy based on Learning By Doing.

Training Goals

- You are going to **clean up** data relevant to the company's production process, sales or a set of customer data.
- You will be able to analyse **data** using visualisation and statistical tools and methods
- You are going to **design** predictive modules and incorporating artificial intelligence to anticipate change or determine a target value for the company.

Hybrid Format

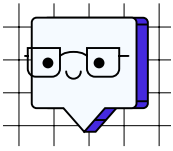
- **Practical work:** For 80% of your time, you are going to be on our personalized learning platform **Learn**, which was developed by DataScientest. The platform includes online exercises enabling you to gradually implement the concepts developed in the course.

- **Masterclass:** 20% of the training take place in Masterclasses. These face-to-face live-web training sessions with our professors allow you to address the all the current problems relating to technologies, methods and tools in the field.

- **Project-based teaching:** In the **backbone project,** you are going to directly apply your newly acquired skills.

Support and assistance

Every day of the week from 9am to 5pm, all the data expert take turns on a dedicated forum to offer personalized technical assistance to you. Educational support is also offered via the "**Slack**", our communication network, and throughout our Q&A sessions.



"We quickly realised that DataScientest had the same vision of teaching and learning as Orange, and that they would be a partner that would listen to our specific needs."

Anne Beaugendre-Frénot
Director, Orange Campus Data IA @Orange

Backbone project

Right from the beginning of your training, you are going to carry out a concrete project. The goal of the project is to gradually combine all the skills you obtain during your course. It requires an investment of about 120 hours of work throughout the training.

You will be able to select a project from our list in groups of two or four people. Our topics are updated monthly and are inspired by the work we do in companies. You can also propose a personal project, as long as the data is accessible, and our teaching team confirms it!

This part of our course is crucial to make you fully operational as a Data Scientist. You are going to work on uncleared data sets and produce a professional project. Some coaching sessions are regularly organized by your project mentor in order to guide and coach you.

This allows you to move efficiently from theory to practice and ensures 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 in Data Science with a successful project during your job interviews.

If you want to know more about our learners' projects, we have created Data Days, a live broadcast of their projects.

Check the Data Days replay

"DataScientest has genuine expertise in Data Science, delivered with tailor-made support and a constant focus on customer satisfaction."

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

Evaluation methods

An evaluation system has been set up throughout your training process. From your first meeting with a DataScientest advisor, we **collect your expectations and needs** at registration. Then, before validating your entry into the course, a **placement test** is sent to you and your knowledge is evaluated using the exams at the end of each module. Finally, through **satisfaction questionnaires**, we collect your appreciation and reviews.

Certification training: Academic recognition

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Our certification is issued by The Sorbonne University in Paris. By completing our Data Scientist training, you will receive an official certificate of the French university, Paris-Sorbonne. This will greatly enhance your resume for future job applications.

Book an appointment

Prerequisites

In order to join the Data Scientist course, it is of advantage to have a bachelor's degree. Good knowledge in mathematics is necessary, as well as a little bit of prior knowledge in programming.
To take the courses, the student must have a computer with an internet connection and a webcam

In Bootcamp or Continuous

Two formats are possible for the Data Scientist training:

Bootcamp

Train quickly, following an intensive program.

- **Duration :** 13 weeks
- **Pace :** 35-40h/week
- **Total duration :** 500h

Continuous format

Flexible, following the training program while continuing to work.

- **Duration :** 9 months
- **Pace :** 10-12h/week
- **Total duration :** 400h

Book an appointment

THE CURRICULUM

1 - Python Fundamentals

Python for Data Scientist, Exploration Statistics, Data Quality, Object-oriented programming

2 - Data Visualisation

Matplotlib, Seaborn, Plotly

3 - Programming tools

Linux & Bash, Git & GitHub, Unit testing, AWS Cloud Practitioner

4 - Machine Learning

Classification, Regression, Clustering

5 - Advanced Machine Learning

Advanced Classification, Recommender Systems, Pipeline

6 - Applied Machine Learning

Dimension Reduction, Time Series, Anomaly Detection, Reinforcement Learning

7 - Complex models

Ethics, Bias & Interpretability, MLflow, Text Mining, Web Scraping with BeautifulSoup, Graph Theory with NetworkX

8 - Deep Learning Fundamentals

Dense networks, Convolution networks, Keras - TensorFlow

9 - Data Engineering

SQL, API, PySpark

10 - MLOps

Streamlit, Docker, AWS Solution Architect

Python Fundamentals - *duration 40h*

Programming

- 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

Object-Oriented Programming (optional)

- Complements the Python module
- Implementing more complex algorithms

Statistics

- Exploring numerical variables
- Exploration of categorical variables
- Study of relationships between variables

Data Quality (optional)

- Presentation of the essential principles of Data Quality
- Treatment of numerical and textual variables
- Data cleaning
- Managing missing values

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

Data Visualisation - *duration 20h*

Matplotlib

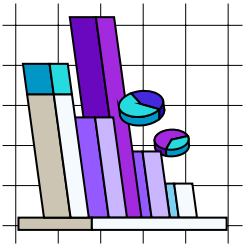
- Presentation of different types of graphs:
  - Curves
  - Charts
  - Scatter plot
  - Histograms
- Introduction to graph customization

Seaborn

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

Plotly (optional)

- Training in all types of interactive graphics that can be integrated into Web pages
- Discover and create Widgets



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

Programming tools - *duration 20h*

Git

- Introduction to the Git version control system
- Initialization of a Git repository
- Presentation and deepening of Git concepts
  - Branches
  - Tag
  - Merge

Linux System and Bash Script

- Presentation of Linux Systems
- Handling and use of a terminal
- Setting up Bash scripts

AWS Cloud Practitioner (optional)

- Overview of the AWS Cloud and the basic global infrastructure
- The platform's key services
- AWS and common use cases

GitHub

- Discovery of the GitHub platform for collaborative work on Git
- Presentation of the major features of GitHub
  - Fork
  - Pull Request
  - Issues
- Share your modifications with pull and push
- Presentation of GitHub Actions and application examples

Unit tests (optional)

- Setting up unit tests with Pytest
- Introduction to Integration Tests and their functions
- Presentation of the advantages of testing: time saving, readability, quality and quality and improvement of code

The skills acquired at the end

- Learn how to use a Terminal
- Master the Linux operating system
- Create and manage Bash executables
- Master versioning tools
- Work collaboratively and version projects with Git and GitHub
- Be able to set up unit tests

Machine Learning - *duration 30h*

Classification models and algorithms

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

Regression methods

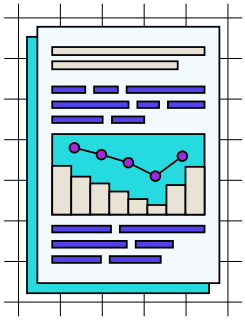
- Simple and multiple linear regression
- Regularized linear regression

Clustering methods (optional)

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

Skills acquired at the end

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



Advanced Machine Learning - *duration 30h*

Advanced Classification

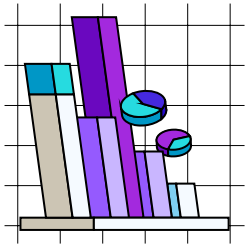
- Presentation of more complex Classification models
- Explanation of how to choose the right Classification models
- Presentation of particular problematics of Classification (imbalanced datasets)

Recommender systems

- Introduction to the functioning and challenges of recommendation systems
- contentbased and collaborative methods
- mathematics of recommender systems

Pipeline (optional)

- Implementing a webapp in Python
- Create interactive graphs and widgets
- Using a Data Storytelling API
- Processors and Estimators
- Building a Pipeline



Skills acquired at the end

- Use of advanced models for specific data
- Time series analysis and prediction
- Management of databases with many variables
- Detection of fraud, defects, etc.

Applied Machine Learning - *duration 20h*

Methods for reducing the dimensions

- Introduction to Principal Component Analysis
- Presentation of the T-SNE algorithm
- Introduction to discriminant linear analysis
- Discovery of clustering with the K-means algorithm

Anomaly Detection (optional)

- Learn how to predict rare situations
- Introducing Anomaly Detection algorithms (KNN, LOF)

Time Series

- Learning how to recognize, clean and analyse a Time Series data frame
- Doing forecasting predictions (ARIMA, SWRIMA)

Reinforcement Learning (optional)

- 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

- Master text data manipulation and pre-processing techniques
- Use a Machine Learning model on textual data
- Systematically retrieve data from various web pages
- Understand the ethical issues associated with data
- Interpret the results of a Machine Learning model

Complex models - *duration 20h*

Text mining

- Introduction to regular expressions
- Text data management
- Creating wordclouds
- Sentiment analysis

Ethics, Bias & Interpretability

- Raising awareness of the use of sensitive data
- Ethical issues of AI
- Bias analysis techniques
- Interpretability of models with SHAP

Web Scraping (optional)

- Introduction to Web Scraping with Beautiful Soup
- Learn how to navigate an HTML document and identify the data on the page

MLFlow

- Introduction to MLFlow architecture
- MLFlow Tracking, MLFlow Projects, MLFlow Models, MLFlow Registry
- Managing the lifecycle of a Machine Learning project

Graph theory (optional)

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

Skills acquired at the end

- Understand the mathematical foundations of complex models
- Master the main algorithms used in Reinforcement Learning
- Manipulate and browse data in graph form
- Build recommendation systems

Deep Learning Fundamentals - *duration 40h*

Dense networks

- Neural network theory
- Discovering the Keras framework
- Optimizing a dense network

Convolution networks

- Image analysis by convolution
- LeNet architecture
- Transfer Learning

TensorFlow (optional)

- Optimization with TensorFlow
- Callbacks and custom templates
- TensorBoard

Skills acquired at the end

- Understanding and implementing neural networks
- Training and evaluating neural network results
- Image processing in Deep Learning
- Use Transfer Learning methods
- Master the syntax and features of the TensorFlow framework

Data Engineering - *duration 30h*

SQL

- How to query a database using SQL
- Understand the different types of joins and when to use them
- Master nested queries

API

- Introduction to APIs
- Overview of the different HTTP methods and their functions

PySpark (optional)

- Introduction to distributed computing with PySpark
- Introducing Spark's RDD and DataFrame APIs
- Distributed data processing pipeline
- Distributed Machine Learning with MLLib

Skills acquired at the end

- Read and query relational databases
- Master the syntax of SQL queries
- Understand the theory of distributed systems architectures
- Master distributed computing with Spark

MLOps - *duration 30h*

AWS Solutions Architect (optional)

- Introduction to APIs and discovery of microservices architectures
- Presentation of the different HTTP methods and their functions
- Using the FastAPI and Flask libraries to develop RESTful APIs
- Documenting an API using the OpenAPI specification
- Managing API errors and performance

Docker

- Introduction to containerization and its usefulness in relation to virtualization
- Introduction to how Docker works
- Handling images and containers
- Using Dockerfile and Docker Hub

Streamlit

- Introducing how to work with Stramlit
- Handling the method codes with a Use case
- Using best practice on Streamlit
- Deploying the application on the Stramlit Cloud

Skills acquired at the end

- Master the entire lifecycle of a Machine Learning project
- Master the AWS Cloud
- Containerization with Docker

Book an appointment

AWS Cloud

With the Data Scientist course, DataScientest offers you the opportunity to prepare for the **AWS Cloud Practitioner** certification, allowing you to learn how to use the **AWS cloud** and to **discover the different services** of this platform. At the end of this training, you will have the opportunity to take the official Amazon certification exam and become an **Amazon Certified Cloud Practitioner**.

The AWS platform is now used by many large groups and companies. Passing the certification will give you a **real advantage** in your job search and will **differentiate** you from the profiles you will be put against during your applications.

The training takes place during one day and is 100% remote, with courses given by our teachers who hold the **Amazon Authorized Instructor status**.

1 - Familiarization with the Cloud

Introduction to AWS Cloud,
Getting started with the AWS Cloud

2 - Introduction to the fundamental concepts of AWS

Discovering and creating services in the AWS Cloud, Maintaining AWS resources and deploying database services, Securing your infrastructure, Assessing the compliance and security of your data, Protection of your infrastructure, Support and pricing

OUR PARTNERS

DataScientest has developed partnerships with world-renowned institutions. On one hand, with academic institutions such as Université Paris - Panthéon Sorbonne, and on the other hand, with software publishers such as AWS or Microsoft. These partnerships are designed to help learners distinguish themselves from other candidates by obtaining certifications recognized by companies."

Amazon Web Services

Software partner

Today, DataScientest enjoys the exclusive status of an Amazon Digital Partner. Therefore, we have been authorized by Amazon to train teams on their products and services. As part of this partnership, we have built several courses that prepare you for the official **AWS certifications,** such as **Cloud Practitioner**. The registration fee for the official exam is included in the course price.

Microsoft

Software partner

DataScientest is a **Microsoft Learning Partner** and therefore authorized to train you in official Microsoft certifications (**PL-900 or AZ-900**). These certifications attest to a certain level of expertise in Azure, the collection of cloud computing products and services, and in Power BI, Microsoft's business intelligence tool.

University Paris 1 - Panthéon Sorbonne

Academic partner

Our **certification** is issued by the **Sorbonne University** in Paris. By completing our Machine Learning Engineer, Data Analyst, Data Scientist, Data Engineer or DevOps training and all of our other courses, you will receive an official certificate of the French university Paris-Sorbonne. This will greatly enhance your resume for future job applications.

ALUMNI TESTIMONIES

Karina CASTILLO

Data Expert Consultant @Expertime

"One of the strong points of DataScientest 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

Data Engineer GCP @GoWizYou

"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!"

Yvonne Rippers

Data Scientist @Dr. Türck Engineering office

"DataScientest was an excellent opportunity for me to deepen my knowledge of the most important fields in data science in a flexible, systematic and thorough way, alongside my work and family life. I found the complexity of neural networks and the final team project entitled "Tree Recognition with Convolutional Neural Networks" particularly exciting.
In the meantime, I have also been offered a job in the field of data science at Dr Türck Ingenieurbüro "

Henry Tornow

Data Analyst @Ed-Tech Start-up

"Through this training, I feel well-equipped to apply my knowledge to real projects and further my career in the field of data science. The organization is exceptional which makes the learning experience very efficient and structured. I can only recommend applying to Data Scientist Paris to anyone who wants to expand their skills in this area."

Customer Care

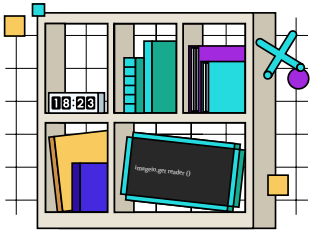
The Customer Care team is made up of **technical advisors** and program managers who work together to provide supervised **support** to each class. The technical side of the training is managed by the technical advisors, while the human side is taken care of by the programme managers.

The classes and learners are **monitored individually** to help them achieve their diploma. To this end, we organize training **follow-up interviews** which emphasize the human dimension and allow us to adapt the rules in the event of personal complications. We also offer **advices** on how to manage time and improve the way you learn throughout the course. **Events** are also organized to strengthen group cohesion and avoid the isolation of distance learners.

We are attentive to the comments of our users to continue to improve our training programmes through satisfaction questionnaires. In addition, there are profiles dedicated to supporting and dealing with each learner's questions.

"Caring for our learners is our main goal. We make sure that their training goes well through individual follow-up, advice and an attentive ear to the needs of each individual."

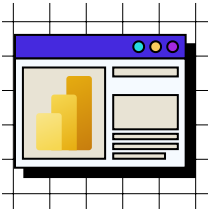
Pauline Messager
Head of Customer Care Service @DataScientest



TO GO FURTHER

If you'd like to improve your skills, DataScientest has set up a range of expert courses and publisher certifications such as Microsoft and AWS to help you deepen your knowledge and perfect your data skills!

You might also like these training programmes:



Power BI

You want to provide a complete analysis of a dataset and improve your dashboard creation? This course is for you! Learn to master Power BI and earn your official Microsoft certification by becoming a **"Power BI Data Analyst Associate"**.

Discover the curriculum »

DataScientest

DO YOU WANT TO BECOME A DATA SCIENTIST?

datascientest.com

contact@datascientest.com

+49 32 222003762