

DATA SCIENCE FRANCE

English & French Syllabi

 **GENERAL ASSEMBLY**



OVERVIEW

Create robust predictive models with statistics and Python programming. Build confidence and credibility to tackle complex machine-learning problems on the job.

Ever wonder how the Netflix recommendation engine works, or how Amazon.com determines what items “you may also like?” These functionalities are driven by training a computer how to learn using large data sets. Machine learning is powering innovation in everything from growth marketing to lending models to fraud detection.

General Assembly’s Data Science course is a practical introduction to the interdisciplinary field of data science and machine learning, which lies at the intersection of computer science, statistics, and business. You will learn to use the Python programming language to acquire, parse, and model data, distilling predictions to influence business strategy. A significant portion of the course is a hands-on approach to fundamental modelling techniques and machine learning algorithms. These enable you to build robust predictive models of real-world data and test their validity. You’ll also practise communicating your results and insights by compiling technical documentation and a stakeholder presentation.

By the end of the course, you will be able to:

- Perform exploratory data analysis with Python.
- Build and refine machine learning models to predict patterns from data sets.
- Communicate data-driven insights to technical and non-technical audiences alike.

To assess your skills, you will need to:

- Attend all class sessions throughout the course.
- Complete and present your final project to your instructor. A minimum score of 66% must be reached to pass the project.

Upon successfully meeting all of the course requirements, you will receive a certificate of completion.



WHAT TO EXPECT

Engage in hands-on, project-based learning that's designed to cover the most important data science concepts and tools. You'll be given access to 12 hours of pre-work to prepare for the course.

As a student, you'll:

- Explore new concepts and tools through expert-led lectures and discussions.
- Complete hands-on exercises with real-world data sets to reinforce newly learned skills.
- Receive individualised feedback and support from your instructional team.
- Gain access to pre-work lessons via the myGA platform.
- Apply what you've learned to create a portfolio project: a predictive model that addresses a real-world data problem.
- Collaborate with your peers, instructors, and alumni via Slack, a collaborative messaging platform.





PREREQUISITES

This is a fast-paced course with some prerequisites:

- Students should be comfortable with programming fundamentals, core Python syntax, and basic statistics. You'll complete a short onboarding task and, based on your results, may be advised to take an introductory Python workshop.
- Have access to a laptop (PC or Mac). It must be no more than four years old and able to run the most recent operating system, updated prior to starting the course class.
- If remote, a webcam, headphones, and good access to the internet are required.

Note: The course also recommends a good level of English proficiency (for non-native speakers, a CEFR level of B2 or a TOEFL score of 90 is recommended). Check this [self-assessment grid](#). Once enrolled, you'll access up to 12 hours of online preparatory lessons that will ensure you have the foundations to dive into rigorous coursework. Our [Admissions team](#) can discuss your background and learning goals to advise if this course is a good fit for you. A level placement test must be taken online if your native language is not English.

THE FINAL PROJECT

For the Data Science final project, you will address a data-related problem in your professional field or a topic of interest to create a predictive model. You will acquire a real-world data set, form a hypothesis about it, and then clean, parse, and apply modelling techniques and data analysis principles.

Students present their results and each write a report that includes:

- A clearly articulated problem statement.
- A summary of the data acquisition, cleaning, and parsing stages.
- A clear presentation of your predictive model and the processes you took to create it.
- A presentation style that's appropriate to your audience.

Your instructional team will help you scope out a project that is commensurate with the skills taught in the course.



COURSE STRUCTURE

Duration: 60 Hours

Pre-Work

Get up to speed on the fundamentals of data science by completing a series of self-paced preparatory lessons via our online learning platform, myGA.

Instructor-Led Sessions

Learn by doing with help from an expert. Explore key concepts and tools, then put them to work through hands-on activities grounded in real-world scenarios. Data Science courses may feature additional flexible sessions or final project studio time based on format, class interests, and trends in local employer requirements.

Homework

Apply what you've learned in class to cumulative assignments that build toward the final project.

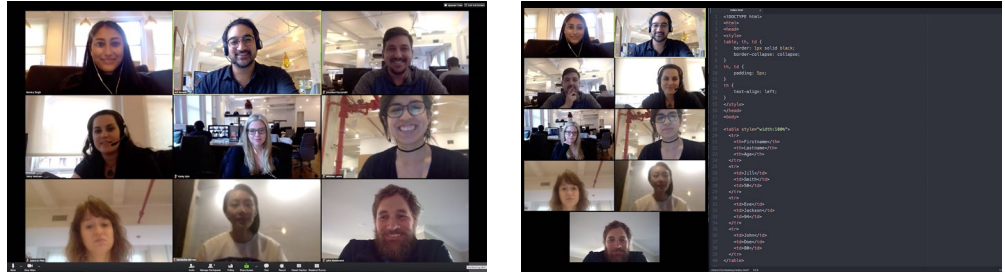
Presentations

Share your final presentation and receive feedback from your instructional team and classmates.



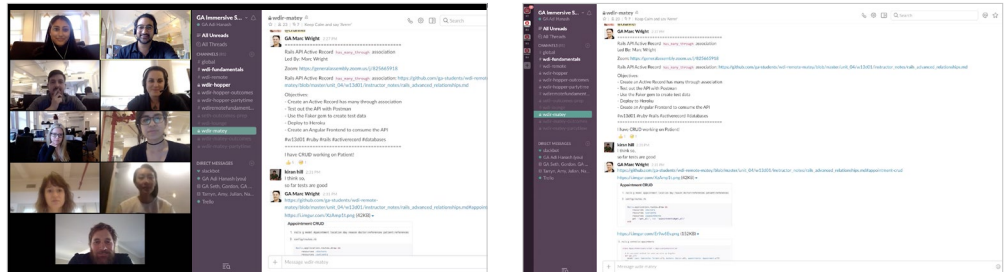
Inside the Remote Classroom

Engage in live, online, project-based learning that's designed to transform your career — from anywhere. Powered by Zoom, the major video conferencing tool, and the collaborative messaging platform Slack, our interactive classroom mirrors what you'll encounter as part of the modern workforce.



Instructor-led courses with Zoom.

You'll be able to not only share your screen and present your work to classmates but also collaborate in small groups via breakout rooms. Additionally, you can revisit recordings of class sessions in the future.



Chat and share with Slack.

Beyond class hours, get guidance, feedback, technical assistance, and more during frequent one-on-one check-ins and office hours.



WHAT YOU'LL LEARN

Pre-Work

Data Science Fundamentals

Explore the essentials of Python for data science and applied math through a series of self-paced online preparatory lessons.

- Define basic Python programming concepts and data types, including variables, lists, dictionaries, loops, and functions.
- Create functions that accept multiple arguments and return multiple values.
- Understand the purpose of iterators in real-world data science workflows.
- Describe the use and purpose of DataFrames and how they can be used to manipulate data with pandas.
- Plot visualisations with Matplotlib and Seaborn.
- Get acquainted with descriptive and inferential statistics and how to calculate them.
- Calculate combinations and permutations.
- Familiarise yourself with developer tools for data science, including GitHub basics and working with the command line.
- Calculate linear algebra and regression equations.

Unit 1

Data Foundations

Discover the fundamentals of evidential science by executing basic functions in Python.

What Is Data Science?

- Define the workflow, tools, and approaches data scientists use to analyse data.
- Apply the Data Science Workflow to solve a task.

Your Development Environment

- Navigate through directories using the command line.
- Use Git and GitHub to share repositories.

Python Foundations

- Conduct arithmetic and string operations in Python.
- Assign variables.
- Implement loops and conditional statements.
- Use Python to clean and edit data sets.

Project: Complete coding challenges that often appear in data science job interviews, further developing your Python programming skills.



Unit 2 Working With Data

Practise exploratory data analysis for cleaning and aggregating data, and understand the basic statistical testing values of your data.

Exploratory Data Analysis in pandas

- Use DataFrames and Series to read data.
- Rename, remove, combine, select, and JOIN data.
- Identify and handle null and missing values.

Data Visualisation in Python

- Define key principles of data visualisation.
- Create line plots, bar plots, histograms, and box plots using Seaborn and Matplotlib.

Statistics in Python

- Use NumPy and pandas libraries to analyse data sets using basic summary statistics.
- Create data visualisations to discern characteristics and trends in a data set.
- Identify a normal distribution within a data set using summary statistics and visualisation.

Experiments and Hypothesis Testing

- Determine causality and sampling bias.
- Test a hypothesis using a sample case study.
- Validate your findings using statistical analysis (e.g., p values, confidence intervals).

Project: Apply your growing Python and analytical skills to conduct a basic exploratory data analysis and answer questions about a real-world data set.

Unit 3 Data Science Modelling

Branch from traditional statistics into machine learning, and explore supervised learning techniques including classification and regression.

Linear Regression

- Define data modelling and linear regression.
- Differentiate between categorical and continuous variables.
- Build a linear regression model for prediction using the scikit-learn library.

Train/Test Split

- Describe errors of bias and variance.
- Define overfitting and underfitting.
- Explore k-folds, LOOCV, and three-split methods.

KNN and Classification

- Build a k-nearest neighbours model using the scikit-learn library.
- Evaluate and tune the model using metrics such as classification accuracy/error.



Unit 3 Data Science Modelling (Cont.)

Logistic Regression

- Build a logistic regression classification model using the scikit-learn library.
- Describe the sigmoid function, odds, and odds ratios and how they relate to logistic regression.
- Evaluate a model using metrics such as classification accuracy/error, confusion matrix, ROC/AOC curves, and loss functions.

Project: Build and validate linear regression and KNN models based on a provided data set.

Unit 4 Data Science Applications

Learn and implement core machine learning models to evaluate complex problems.

Working With API Data

- Access public APIs and get information back.
- Read and write data in JSON.
- Use the requests library.

Natural Language Processing

- Demonstrate how to tokenise natural language text.
- Categorise and tag unstructured text data.
- Perform text classification model using scikit-learn, CountVectorizer, TfidfVectorizer, and TextBlob.

Time Series Data

- Create rolling means and plot time series data.
- Examine autocorrelation on time series data.

Flex Sessions

- Explore an additional data science topic based on class interest. Options include clustering, decision trees, robust regression, and deploying models with Flask.



FREQUENTLY ASKED QUESTIONS

Why are data science skills relevant today?

Companies of all stripes use data science to take on today's biggest challenges, tackling everything from public policy and robotics to dating and eCommerce. As a result, organisations are moving quickly to build robust in-house teams of data scientists and advanced analysts, and there's not enough talent to go around.

According to [Burning Glass](#), "Data science and analytics skills are now widely in demand in decision-making roles, including managers across a range of industries. In fact, our data shows that more than 1.7 million job postings asked for data science skills in 2018." Learning this future-proof skill set can help you enter the next stage of your career, whether that's advancing in your current profession or exploring an exciting and lucrative field.

What are the professional backgrounds of data science students?

This course is designed for data professionals who want to perform complex analysis to power predictions and add marketable skills to their resume. You'll find a diverse range of students in the classroom, including:

- Data analysts, marketing analysts, BI analysts, or consultants who work with big data and need to upgrade their skills.
- Software engineers who want to apply their programming skills toward a new career.
- Other professionals with a quantitative background eyeing a transition to tech.

Ultimately, this programme attracts a community of eager learners who have an interest in manipulating large data sets and forecasting to impact strategy and the bottom lines.

What skills are assessed by the end of the programme?

In order to earn a certificate, you will be evaluated based on your ability to:

- Perform exploratory data analysis with Python.
- Build and refine machine learning models to predict patterns from data sets.
- Communicate data-driven insights to technical and non-technical audiences alike.

What does my tuition cover?

Here are just some of the benefits you can expect as a GA student:

- 60 hours of expert instruction designed to build a well-rounded foundational data science skill set.
- 12 hours of self-paced pre-work to brush up on programming fundamentals and statistics.
- Robust data science coursework, including expert-vetted lesson decks, lab materials, and more. Refresh and refine your knowledge throughout your professional journey as needed.
- A portfolio-ready capstone project built with support from your instructor.
- Individual feedback and guidance from instructors and TAs during office hours. Stay motivated and make the most of your experience with the help of GA's dedicated team.



- Access to the collaborative chat tool Slack to communicate with your instructors, peers, and alumni throughout the course.
- Exclusive access to alumni discounts, networking events, and career workshops.
- A certificate of completion to showcase your new skill set on LinkedIn.
- Connections with a professional network of instructors and peers that lasts well beyond the course. The global GA community can help you navigate and succeed in the data science field.

Are there any prerequisites?

This is a fast-paced course with some prerequisites.

- Students should be comfortable with programming fundamentals, core Python syntax, and basic statistics. You'll complete a short onboarding task and, based on your results, may be advised to take an introductory Python workshop.
- Have access to a laptop (PC or Mac). It must be no more than four years old and able to run the most recent operating system, updated prior to starting the course class.
- If remote, a webcam, headphones, and good access to the internet are required.

Note: The course also recommends a good level of English proficiency (for non-native speakers, a CEFR level of B2 or a TOEFL score of 90 is recommended). Check this [self-assessment grid](#). Once enrolled, you'll access up to 12 hours of online preparatory lessons that will ensure you have the foundations to dive into rigorous coursework. Our [Admissions team](#) can discuss your background and learning goals to advise if this course is a good fit for you. A level placement test must be taken online if your native language is not English.

Will I earn a certificate?

Upon successfully meeting all of the course requirements, you will receive a Data Science certificate. To pass, you will need to:

- Attend all class sessions.
- Complete and present your final project to your instructor. A minimum score of 66% must be reached to pass the project.

Thousands of GA alumni use their course certificate to demonstrate skills to employers and their LinkedIn networks. GA's Data Science course is well-regarded by many top employers who contribute to our curriculum and use our data programmes to train their own teams.

What does student success look like for this course?

In 2021, no student followed this course, on the French Campus.

In 2022, on the French Campus, our evening and 1-week courses have an average of 100% completion rate and consistently earn an average of +50 Net Promoter Score.*

*NPS: This metric from students themselves evaluates their satisfaction to recommend this course to a friend.

To get more details, click on this [link](#).



What projects will I work on during this course?

For your capstone project, you'll apply machine learning techniques to solve a real-world problem. You'll develop a model, technical documentation, and stakeholder presentation, and graduate with a polished, portfolio-ready data science project to showcase your skills. We encourage you to tackle a problem that's related to your work or a passion project you've been meaning to carve out time for.

Throughout the course, you'll also complete a number of smaller projects designed to reinforce what you've learned in each unit.

Can I work full-time while enrolled in this course?

Yes! All of our evening and 1-week courses are designed for busy professionals with full-time work commitments. Depending on your location, classes meet twice per week in the evenings or all day on Saturday. If you're able to dedicate a full week to learning, we also have a full-time, accelerated 1-week option. Please keep in mind that our remote course is live online, which means that if you take the accelerated programme, you'll be connected live with your instructor and classmates the whole day.

You'll be expected to spend time working on homework and projects outside of class each week, but the workload is designed to be manageable with a full-time job.

If you need to miss a session or two, we offer resources to help you catch up. We recommend you discuss any planned absences with your instructor.

In what language is this programme taught?

Our courses will be delivered in two ways:

- Fully in English
- In French, with English materials

Please check with the Admissions team to confirm which is true of your course of interest and what is relevant for your situation. In both cases, if you are not a native speaker, a CEFR level of B2 or a TOEFL score of 90 is recommended. (Review this [self-assessment grid in English](#).) A level placement test must be taken online if your native language is not English.

How does this course relate to GA's other data programmes?

This Data Science course assumes some prerequisite knowledge and is designed for professionals who already work with data and want to perform more complex analyses involving computation.

If you're searching for a more entry-level course, [Data Analytics](#) teaches beginners how to perform rigorous data analysis with Excel, SQL, and Tableau.

For those committed to a career change, the full-time [Data Science Bootcamp](#) programme provides the most direct pathway to data science and other advanced analytics roles.

What are my financing options?

We offer a variety of different [financing options](#), including instalment plans, so you can focus on what counts — your education. Plus, more than 45% of our evening and 1-week students receive full or partial tuition reimbursement from their employers. Speak with our [Admissions team](#) to learn more and find the best fit for you and your finances.



TAKE THE NEXT STEP

Have questions about our Data Science course? Our [Admissions team](#) is here to help and can advise on if this programme is right for you and your learning goals. You can also:

- Attend an info session [online](#) or at your [local campus](#).
- Explore your [financing options](#).
- [Apply](#) to enrol in the course.*
- Email us directly at paris-inscription@generalassemb.ly or call at +33 1.88.24.42.51.

Course Enrolment Availability

Enrolments are available up to 14 days before the course starts.

Course Dates & Prices

GA delivers about four sessions per year. To learn about our course dates and prices, please explore [this page](#) or contact our Admissions team.

** Course modality options vary by location, pending market availability. Please contact our Admissions team to discuss what version is available in your location.*

ACCESSIBILITY

GA provides access to its courses for people with disabilities. Please contact our Admissions team to discuss your needs and learn how we could support you throughout your learning journey.



PRÉSENTATION DE LA FORMATION

Utilisez les statistiques et la programmation Python pour créer des modèles prédictifs fiables. Développez votre confiance en vous et votre crédibilité, en vous entraînant à résoudre des problèmes complexes.

Vous êtes-vous déjà demandé comment fonctionne le système de recommandations de Netflix, ou comment Amazon.com sait quels articles « vous pourriez aussi aimer » ? Ces fonctionnalités sont obtenues grâce à un ordinateur qui a appris à utiliser de gros volumes de données. Le machine learning (ou apprentissage automatique) révolutionne l'innovation dans tous les secteurs, du growth marketing aux modèles de prêts en passant par la détection des fraudes.

La formation Data Science de General Assembly est une initiation pratique au domaine interdisciplinaire de la data science et du machine learning, entre IT, statistiques, et business. Vous apprendrez à utiliser le langage de programmation Python pour acquérir, analyser, et modéliser des données, puis distiller des prédictions qui influencent toute la stratégie de l'entreprise.

L'une des parties essentielles de la formation est la mise en pratique des principales techniques de modélisation et des algorithmes du machine learning. Ils vous permettent de développer des modèles prédictifs fiables basés sur des données réelles et de tester leur validité. Vous communiquerez vos résultats et observations via la rédaction d'un document technique et une présentation aux différentes parties prenantes.

À la fin de cette formation, vous serez capable de :

- Réaliser une analyse exploratoire de données avec Python
- Construire et affiner des modèles de machine learning pour faire des prédictions à partir d'ensembles de données
- Communiquer des observations basées sur des données à des publics avertis comme novices

Pour valider la formation, il faut :

- Être assidu à toutes les journées de cours
- Réaliser et soutenir votre projet final face au formateur. Pour attester de la maîtrise de l'ensemble des compétences visées, vous devez obtenir un score de 66%.

Le respect de ces deux critères donne accès à la validation de la formation et à son certificat.



VOTRE EXPÉRIENCE

En ligne, comme sur nos campus, vous participez à une formation très pratique, basée sur un projet professionnalisant et conçue pour couvrir tous les concepts et outils de base. Notre expert vous apporte toutes les compétences nécessaires pour réaliser pas à pas, votre projet, mobiliser vos compétences et ainsi, faire avancer votre carrière.

Pour préparer la formation, vous avez accès au “pre-work,” des cours en ligne sur notre plateforme, entièrement dédiés aux bases de la data science. Ce travail préparatoire, d’une durée d’environ 12h, est à compléter avant le début de la formation.

Se former avec General Assembly, c’est :

- Découvrir de nouveaux concepts et outils lors de conférences et de débats menés par des experts
- Accéder à notre plateforme de travail préparatoire myGA. Avec ce pre-work, bénéficiez de nombreux cours en ligne afin de vous préparer à la formation ou de renforcer vos acquis
- Plonger au cœur de sujets en mettant en pratique très concrètement vos compétences lors d’exercices interactifs, de programmation, et d’études de cas, à réaliser seul et en groupe
- Être accompagné par nos experts et avoir leur soutien personnalisé, ainsi que celui de notre équipe pédagogique
- Appliquer, pas à pas, ce que vous avez appris en réalisant un projet type : un modèle prédictif qui résout un problème lié à des données réelles
- Échanger avec notre communauté d’étudiants, d’experts, et d’alumni via l’outil de communication collaboratif, Slack. Retrouvez de nombreux conseils, événements, ou opportunités même après votre formation



PRÉ-REQUIS

Cette formation accélérée nécessite quelques prérequis.

- Les étudiants doivent maîtriser les bases de la programmation, de la syntaxe Python, et des statistiques. Vous devrez réaliser un test auprès de notre équipe Admission. Selon vos résultats, il pourra vous être conseillé de participer à un workshop d'initiation à la programmation en Python.
- Avoir un ordinateur portable PC ou Mac (le système d'exploitation doit être le plus récent - moins de 4 ans - ou mis à jour avant la formation)
- Si la formation se déroule à distance, une connexion à Internet, webcam et des écouteurs sont nécessaires.

Lorsque la formation est délivrée en anglais, il est recommandé de :

- Disposer d'un niveau B2 en anglais (CECRL) ou un score au TOEFL d'au moins 90. En effet, en fonction des possibilités et de votre choix, la formation peut être entièrement en anglais ou en français avec des supports pédagogiques en anglais. Vous pouvez vous auto-évaluer à partir de ces grilles ([description en français](#)/ [description en anglais](#)). Un test de placement de niveau devra être passé en ligne si votre langue maternelle n'est pas l'anglais.

N'hésitez pas à contacter notre [équipe d'admission](#) (ou écrivez directement à paris-inscription@generalassembly.ly) pour faire le point sur votre parcours et vos objectifs afin de vérifier si ce bootcamp est adapté à vos besoins.



VOTRE PROJET FINAL

Grâce aux compétences acquises pendant la formation Data Science, vous devrez résoudre un problème relatif à des données, en lien avec votre secteur professionnel ou un sujet qui vous intéresse, pour créer un modèle prédictif. Vous acqueriez un ensemble de données réelles, vous en formulez une hypothèse, puis vous les nettoyez, les analysez, et leur appliquez les techniques de modélisation et les principes d'analyse des données.

Les étudiants présentent leurs résultats et chacun d'eux rédige un rapport avec :

- Un énoncé du problème clairement articulé
- Une synthèse des phases d'acquisition, de nettoyage, et d'analyse des données
- Une présentation claire de votre modèle prédictif et des processus suivis pour le créer
- Un style de présentation adapté à votre public

Votre équipe pédagogique vous aidera à définir un projet mettant en valeur les compétences acquises lors de votre formation.



DÉROULÉ DE LA FORMATION

Durée : 60 heures

Pre-work : Travail préparatoire

Pour préparer la formation, vous avez accès au “pre-work,” des cours en ligne sur notre plateforme, entièrement dédiés aux bases de la programmation en Python.

Ce travail préparatoire, conçu pour vous initier aux compétences et au contexte des données fondamentales, est d’une durée d’environ 12h entièrement en anglais. Il est à compléter avant le début de la formation. Il vous aidera à partir du bon pied dès le premier jour de cours.

Animé par un expert métier

Passez de la théorie à la pratique avec un expert métier. Mettez en pratique les principaux concepts et outils abordés via des activités basées sur des projets concrets. La formation Data Science peut varier en termes de format, de temps dédié au projet final et de contenus qui sont adaptés aux intérêts manifestés par l’ensemble des participants ou aux besoins identifiés des employeurs. Nos formations en ligne sont dispensées sur Zoom, notre outil de visioconférence.

Travaux pratiques

Appliquez vos acquis et mobilisez les compétences vues lors des cours pour réaliser les TP pendant et après les sessions de cours. Ils vous permettront de construire pas à pas, le projet professionnalisant final.

Projet final

En fin de formation, vous présenterez votre projet face aux autres participants et le formateur expert.

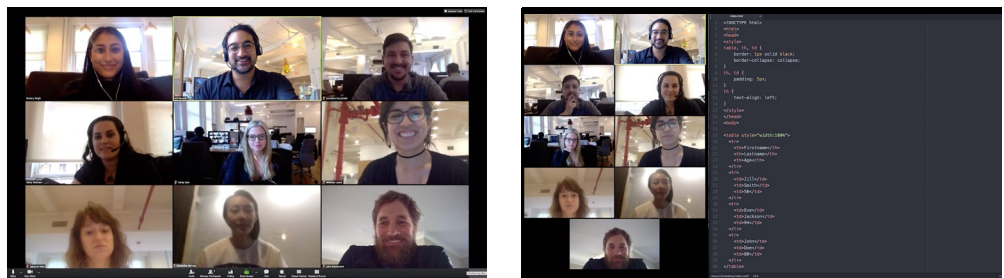


Si vous suivez la formation entièrement en ligne

Vous vivrez l'expérience campus, depuis chez vous!

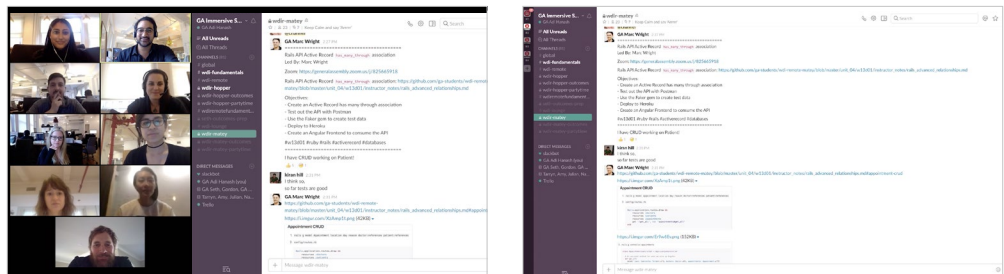
Comment? Grâce aux sessions en classe virtuelle, avec tous les participants. Notre expert animera la formation via [Zoom](#), notre outil de visioconférence.

Vous assisterez en temps réel à la formation et pourrez collaborer avec tous les autres participants en classe entière ou en groupe. Ces sessions sont interactives. Vous pourrez produire, interagir, partager votre écran, et présenter vos travaux, travailler en petit groupe via nos classes virtuelles et le chat intégré, pour suivre la formation, comme si vous y étiez. Vous pouvez d'ailleurs visionner les vidéos en replay pour voir, ou revoir, le cours.



Les classes virtuelles en groupe

Grâce à [Slack](#), une application de messagerie téléchargeable sur votre ordinateur et votre mobile, vous pourrez chatter pendant ou en dehors du cours avec votre formateur ou les participants. Cet espace d'échange et de collaboration permettra de partager des documents, discuter, débattre, ou même faire un point sur votre avancement.



Le chat collaboratif



CE QUE VOUS APPRENDREZ

Travail de préparation (pre-work)

Fondamentaux de la data science

Explorez les fondamentaux de Python pour la data science et les mathématiques appliquées via une série de cours en ligne que vous suivez en toute autonomie.

- Définir les concepts de base de la programmation Python et des types de données, notamment les variables, listes, dictionnaires, boucles, et fonctions
- Créer des fonctions qui acceptent des arguments multiples et renvoient des valeurs multiples
- Comprendre le rôle des itérateurs dans des processus concrets de data science
- Décrire l'utilisation et le rôle des DataFrames et apprendre à les utiliser pour exploiter des données avec Pandas
- Tracer des représentations graphiques avec Matplotlib et Seaborn
- Se familiariser avec les statistiques descriptives et inférentielles et apprendre à les calculer
- Calculer des combinaisons et des permutations
- S'initier aux outils de développement pour la data science, notamment à GitHub et au travail en ligne de commande
- Calculer un algèbre linéaire et des équations de régression

Module 1

Principes fondamentaux de la data

Découvrez les fondamentaux de la science probante en exécutant des fonctions de base en Python.

Qu'est-ce que la data science ?

- Définir les processus, outils, et approches utilisés par les data scientists (ou scientifiques des données) pour analyser des données
- Appliquer des processus de data science pour résoudre une tâche

Votre environnement de développement

- Naviguer d'un répertoire à l'autre grâce à la ligne de commande
- Utiliser Git et GitHub pour partager les répertoires

Bases de Python

- Effectuer des opérations arithmétiques et sur des chaînes de caractères en Python
- Affecter des variables
- Implémenter des boucles et des instructions conditionnelles
- Utiliser Python pour nettoyer et modifier des ensembles de données



Module 1 Principes fondamentaux de la data (Cont.)

Projet : Perfectionner vos compétences en programmation Python pour relever les défis auxquels les candidats font face lors d'un entretien d'embauche pour un emploi dans le domaine de la data science.

Module 2 Exploitation des données

Entraînez-vous à l'analyse exploratoire de données pour nettoyer et agréger des données et découvrez les principaux tests statistiques de vos données.

Analyse exploratoire de données en Pandas

- Utiliser DataFrames et Series pour lire des données
- Renommer, supprimer, combiner, sélectionner, et consolider des données
- Identifier et traiter les valeurs nulles et manquantes

Visualisation des données en Python

- Définir les principes clés de la visualisation des données
- Créer des graphiques linéaires, des graphiques à barres, des histogrammes, et des graphiques en boîtes avec Seaborn et Matplotlib

Statistiques en Python

- Utiliser les bibliothèques NumPy et Pandas pour analyser des ensembles de données en utilisant des statistiques synthétiques de base
- Créer des visualisations de données pour distinguer les caractéristiques et les tendances d'un ensemble de données
- Identifier une distribution normale dans un ensemble de données en utilisant des statistiques sommaires et des visualisations

Expérimentations et vérification des hypothèses

- Déterminer la causalité et les biais d'échantillonnage
- Tester une hypothèse via une étude de cas
- Valider vos conclusions via une analyse statistique (ex. : valeurs-p, intervalles de confiance)

Projet : Appliquer vos compétences en Python et en analyse pour effectuer une analyse exploratoire de données et répondre à des questions relatives à un ensemble concret de données.



Module 3 **Modélisation de la data science**

Passez des statistiques classiques au *machine learning* et explorez des techniques d'apprentissage supervisées telles que la *classification* et la *régression*.

Régression linéaire

- Définir la modélisation des données et la régression linéaire
- Faire la distinction entre variables catégorielles et continues
- Construire un modèle de régression linéaire pour la prédiction avec la bibliothèque *scikit-learn*

Fractionnement train/test

- Décrire les erreurs de biais et de variance
- Définir *overfitting* and *underfitting*
- Explorer les méthodes de validation croisée à k blocs, *LOOCV* (*Leave One-Out Cross Validation*), et en trois sections

KNN et classification

- Construire un modèle de *k-nearest neighbors* avec la bibliothèque *scikit-learn*
- Évaluer et ajuster le modèle en utilisant des indicateurs de mesure comme l'exactitude/erreur de classification

Régression logistique

- Construire un modèle de classification par régression logistique avec la bibliothèque *scikit-learn*
- Décrire la fonction sigmoïde, les cotes et les rapports de cotes, et leur relation avec la régression logistique
- Évaluer un modèle à l'aide d'indicateurs de mesure comme l'exactitude/erreur de classification, la matrice de confusion, les courbes *ROC/AUC*, et les fonctions de perte

Projet : Construire et valider une régression linéaire et des modèles *k-nearest neighbors* à partir d'un ensemble de données fourni.



Module 4 **Application de la data science**

Découvrez et implémentez des modèles basiques de *machine learning* pour évaluer des problèmes complexes.

Exploitation de données API

- Accéder aux API publiques et récupérer des informations
- Lire et écrire des données en JSON
- Utiliser la bibliothèque de requêtes

Traitement du langage naturel

- Montrer comment tokeniser un texte en langage naturel
- Catégoriser et baliser des données textuelles non structurées
- Réaliser un modèle de classification de texte en utilisant scikit-learn, CountVectorizer, TfidfVectorizer, et TextBlob



QUESTIONS FRÉQUEMMENT POSÉES

Pourquoi les compétences en data science sont-elles pertinentes aujourd'hui ?

Toutes sortes d'entreprises utilisent la data science pour faire face aux défis majeurs auxquels elles sont confrontées, que ce soit dans le domaine public ou celui de la robotique, pour des sites de rencontres comme de eCommerce. Résultat : les organisations s'empressent de constituer de solides équipes internes de data scientists et d'analystes expérimentés, et la pénurie de talents se fait déjà sentir.

Selon [Burning Glass](#), « les compétences en data science et analytics sont aujourd'hui très recherchées pour les rôles décisionnaires, notamment de managers, dans divers secteurs d'activité. En effet, nos données montrent que des compétences en data science étaient demandées dans plus d'1,7 million d'offres d'emploi publiées en 2018 ». L'apprentissage de cet ensemble de compétences d'avenir peut vous permettre d'évoluer professionnellement, que ce soit dans votre emploi actuel ou en découvrant un autre domaine passionnant et lucratif.

Quel est le profil professionnel des étudiants qui suivent la formation Data Science ?

Cette formation est destinée aux professionnels des données qui souhaitent réaliser des analyses complexes pour renforcer leurs prédictions et ajouter des compétences monnayables sur leur CV. Vous découvrirez que les profils des élèves sont très variés :

- Analystes de données, analystes marketing, analystes BI (business intelligence), ou consultants qui manipulent des big data et doivent améliorer leurs compétences
- Ingénieurs logiciels qui souhaitent utiliser leurs compétences en programmation dans une nouvelle aventure professionnelle
- Autres professionnels ayant une expérience quantitative qui envisagent une transition vers la technologie. Enfin, ce programme attire aussi des passionnés qui aiment manipuler de gros volumes de données et faire des prédictions pour influencer sur les stratégies et les résultats

Quelles sont les compétences évaluées ?

Pour obtenir votre attestation de formation, vous êtes évalué-e sur votre capacité à :

- Réaliser une analyse exploratoire de données avec Python
- Construire et affiner des modèles de machine learning pour faire des prédictions à partir d'ensembles de données
- Communiquer des observations basées sur des données à des publics avertis comme novices



Comment se déroule la formation et que couvrent les frais d'inscription ?

Suivre une formation chez General Assembly, c'est accéder à de nombreux services.

- Une formation dirigée par des experts, sur les compétences clés en data science en campus ou entièrement à distance (en classe virtuelle)
- Un accès à notre plateforme de cours en ligne myGA pour effectuer le travail préparatoire pre-work afin d'actualiser et d'affiner vos connaissances en data science et statistiques tout au long de la formation
- Un projet concret consistant à appliquer les techniques du machine learning en créant un modèle, une documentation technique, et une présentation destinée aux différentes parties prenantes
- Un accompagnement personnalisé par notre équipe pédagogique (expert et équipe de student experience)
- Un accès à notre outil de communication collaboratif, Slack, avant, pendant, et après la formation pour échanger avec notre communauté d'étudiants, d'experts, et d'alumni et retrouver ainsi de nombreux conseils, événements, ou opportunités même après votre formation
- Une attestation et un certificat GA pour mettre en valeur vos nouvelles compétences sur LinkedIn
- Un accès exclusif aux avantages en tant qu'alumni : promotions, invitations aux événements de réseautage, et aux ateliers de développement professionnel pour booster votre carrière

Quelles sont les conditions d'admission et pré-requis?

Cette formation accélérée nécessite quelques prérequis.

- Les étudiants doivent maîtriser les bases de la programmation, de la syntaxe Python, et des statistiques. Vous devrez réaliser et valider un test d'entrée auprès de notre équipe d'admission. Selon vos résultats, il pourra vous être conseillé de participer à un workshop d'initiation à Python.
- Avoir un ordinateur portable PC ou Mac (le système d'exploitation doit être le le plus récent - moins de 4 ans - ou mis à jour avant la formation)
- Si la formation se déroule à distance, une connexion à internet, webcam, et des écouteurs sont nécessaires

Pour suivre la formation, il est recommandé de :

- Disposer d'un niveau B2 en anglais (CECRL) ou un score au TOEFL d'au moins 90. En effet, en fonction des possibilités et de votre choix, la formation peut être entièrement en anglais ou en français avec des supports pédagogiques en anglais. Vous pouvez vous auto-évaluer à partir de ces grilles ([description en français](#)/ [description en anglais](#)). Un test de placement de niveau devra être passé en ligne si votre langue maternelle n'est pas l'anglais.

N'hésitez pas à contacter notre [équipe d'admission](#) (ou écrivez directement à paris-inscription@generalassembly.com) pour faire le point sur votre parcours et vos objectifs afin de vérifier si cette formation est adaptée à vos besoins.



Comment valider la formation et obtenir le certificat ?

À la fin de la formation, si vous remplissez les conditions de validation de la formation, vous obtenez une attestation de réussite et son certificat. Pour valider la formation, il vous faut :

- Être assidu aux journées de cours
- Réaliser et soutenir votre projet final face au formateur. Pour attester de la maîtrise de l'ensemble des compétences visées, vous devez obtenir un score de min. 66%.

Le respect de ces deux critères donne accès à la validation de la formation et à son certificat.

Des milliers d'anciens étudiants de GA utilisent leur certification pour démontrer leurs compétences aux employeurs et à leurs réseaux sur LinkedIn. La formation Data Science de GA est appréciée et reconnue par de nombreuses grandes entreprises qui participent à la conception de notre programme et/ou l'utilisent afin de former leurs propres équipes.

Quelle est la satisfaction de cette formation par les alumni ?

En 2021, aucun apprenant n'a suivi ce cours sur le Campus français.

En 2022, sur le Campus français, nos cours du soir et nos cours intensif sur une semaine ont un taux d'achèvement moyen de 100 % et obtiennent régulièrement un score de recommandation net moyen de +50 (NPS*).

*NPS : Cette mesure, réalisée par les apprenants eux-mêmes, évalue leur satisfaction à recommander ce cours à un tiers.

Pour obtenir plus de détails ou accéder à plus d'indicateurs de satisfaction et de résultats, cliquez sur [ce lien](#).

Sur quels projets vais-je travailler pendant la formation ?

Pour votre projet final, vous appliquez les techniques du machine learning pour résoudre un problème concret. Vous créez un modèle, une documentation technique, et une présentation destinée aux différentes parties prenantes. Vous obtenez votre certificat en développant un projet concret de data science qui met en valeur vos compétences acquises.

Tout au long de votre formation, vous participez également à divers petits ateliers destinés à renforcer les connaissances acquises dans chaque module.

Quel est le lien entre cette formation et les autres programmes relatifs à la data ?

Cette formation en Data Science suppose quelques connaissances préalables et est destinée aux professionnels qui maîtrisent déjà les données et souhaitent réaliser des analyses plus complexes impliquant des calculs.

Si vous recherchez un cours de niveau débutant, la formation [Data Analytics](#) est faite pour vous. Vous apprendrez à réaliser des analyses complètes de données avec Excel, SQL, et Tableau.

Pour ceux qui sont en reconversion professionnelle, le programme [Bootcamp Data Science](#) à temps plein offre un accès direct à la data science et à divers métiers dans le domaine de l'analyse.



Puis-je travailler à temps plein pendant que je suis cette formation ?

Oui, c'est possible. Nous proposons plusieurs modalités de suivi de formation : entièrement en ligne, à temps partiel, cours du soir ou week-end ou à temps complet sur une semaine. Selon votre région, les sessions de formation peuvent avoir lieu deux fois par semaine le soir, ou toute la journée le samedi. Si vous avez la possibilité de consacrer une semaine complète à votre formation, nous proposons également l'option accélérée d'une semaine à temps plein en ligne. Gardez bien à l'esprit que vous devez être en ligne avec votre expert et les autres participants toute la journée.

Lors de la formation, vous devez réaliser une série de travaux pratiques intermédiaires conduisant au projet final en dehors des heures de formation mais la charge de travail est conçue pour être réalisable avec un emploi à temps plein.

Si vous devez manquer une session ou deux, nous proposons des ressources pour vous permettre de les rattraper. Nous vous recommandons de prévenir votre formateur de toute absence prévue.

Dans quelle langue a lieu la formation ?

Nous proposons la formation sous les deux formes suivantes :

- Entièrement en anglais
- En français, sur la base de supports pédagogiques en anglais

Dans les deux cas, nous vous recommandons d'avoir le niveau suivant en anglais :

- Un niveau B2 en anglais (CECRL) ou un score au TOEFL d'au moins 90. En effet, en fonction de votre choix, la formation peut être entièrement en anglais ou en français avec des supports pédagogiques en anglais. Vous pouvez vous auto-évaluer à partir de ces grilles ([description en français](#) / [description en anglais](#)) Un test de placement de niveau devra être passé en ligne si votre langue maternelle n'est pas l'anglais.

Notez que le domaine de la data science est très largement exposé à l'anglais, c'est pourquoi il est important que vous soyez à l'aise avec cette langue.

Quelles sont mes options de financement ?

Nous proposons différentes [options de financement](#), y compris des programmes de paiements échelonnés, afin que vous puissiez vous concentrer sur ce qui compte : votre formation.

Discutez avec notre [équipe d'admission](#) pour en savoir plus et trouver la solution la mieux adaptée à vous et vos finances.



PASSEZ À LA PROCHAINE ÉTAPE

Vous avez des questions sur notre formation “Data Science” ? Notre équipe d'admission est là pour vous aider à déterminer si cette formation est faite pour vous et adaptée à vos objectifs.

Vous pouvez aussi :

- Assister à une séance d'information [en ligne](#) ou sur [votre campus local](#)
- Découvrir vos [options de financement](#)
- [Envoyer votre candidature](#) pour vous inscrire au cursus*
- Nous contacter par email à paris-inscription@generalassembly.ly ou par téléphone au +33 1.88.24.42.51.

Délai d'accès de la formation

Les inscriptions peuvent avoir lieu jusqu'à 14 jours avant le début de la formation. Contactez-nous via ce [formulaire](#).

Dates et prix de la formation

Nous proposons régulièrement des sessions de formation : en moyenne, 4 sessions de formation par an. Pour obtenir les dates et le tarif, consultez directement notre [site web](#) ou contactez notre équipe Admission.

** Les options de modalités de formation varient selon le lieu géographique, dans l'attente de la disponibilité du marché. N'hésitez pas à contacter notre équipe d'admission pour savoir quelles modalités sont disponibles dans votre région.*

ACCESSIBILITÉ AUX PERSONNES EN SITUATION DE HANDICAP

General Assembly soutient la diversité, l'équité et l'inclusion des personnes en situation de handicap. Pour cela, nos formations sont disponibles sous différentes modalités (à distance ou en présentiel). Notre équipe peut vous proposer une assistance spécifique sur demande et étude de vos besoins.

Contactez notre équipe Admission afin de faire le point sur vos besoins et discuter avec notre représentant au handicap sur la meilleure façon de vous accompagner lors de votre expérience d'apprentissage.