

Data Science Graduate Programme Handbook



Table of contents

Welcome	2
Introduction and expectations	3
Learning outcomes	5
Course structure	6
Curriculum overview	8
What to do if you have a problem	9
Communicating with the faculty on Slack	9
IT requirements	10
Frequently asked questions	11

Welcome

Congratulations on securing your place on the Data Science Graduate Programme.

I am delighted to welcome you to the programme and look forward to seeing the contributions you will make to support society and the economy during this programme, and as you develop your careers.

The need for data science skills across the public sector has never been more visible, informing the most important policy questions of the day and playing a critical role in improving inclusion and efficiency. This Data Science Graduate Programme will help you gain the skills and experience you need to play your part in this.

Please do make the most of this opportunity to learn and make use of the network of data scientists across the public sector created through this programme. As the largest cohort to date, across more than 30 organisations, you will have an insight into how data science can benefit society and contribute to the public good.

I look forward to working with you and celebrating your success during the programme.

M. Gregory

Mary Gregory,
Deputy Director





Introduction and expectations

The aim of this programme is to prepare you with the knowledge and skills needed to derive crucial insight from a range of data sources.

Our goal is to equip the analysts of the future with the capacity to develop and use data products in the service of public interest. This handbook is your first port of call when you have a question about the programme. If the answer is not in this book, it will point you in the direction of whom to ask.

Expectations

The Data Science Graduate Programme is managed by the Office for National Statistics (ONS) Data Science Campus. It offers an innovative curriculum, supported by the Data Science Campus faculty, the Government Analysis Function, Ordnance Survey and ONS Geography. On this page, we outline what you can expect from the faculty, and what we, the faculty, expect of you.

We, the faculty, agree to:

- Provide a diverse and challenging training schedule. Occasionally, circumstances beyond our control may require us to reschedule planned training.
- Communicate with learners as soon as is feasible if there are changes to the planned training schedule.
- Provide opportunities for learners to practise the application of training content to real questions in the form of mini projects.
- Where possible, differentiate training material according to the specific needs of our learners.
- Offer learners opportunities to ask questions and get clarification on concepts they don't understand through our Slack channel and during Question and Answer sessions.

You, the learner, agree to:

- Make every effort to attend all training sessions, communicating to the faculty in an appropriate way where this is not possible and complete independent learning by agreed target dates.
- Fully engage with and participate in training, discussions, workshops and equivalent active-learning opportunities.
- Complete pre-reading material and ensure prerequisites are met before training days, ensuring all relevant software and dependencies operate as required.
- Share your knowledge with your peers and teams.
- Engage in Slack channel comms and extracurricular activities when possible.
- Raise any issues you encounter in the programme in a timely manner through an appropriate channel.
- Model professional conduct and uphold the Civil Service values at work.



Learning outcomes

- To develop the necessary coding skills in relevant programming languages.
- To utilise the required packages for a data scientist in government and the public sector.
- To use tools for building efficient and reproducible workflows to facilitate collaborative development.
- To demonstrate proficiency in exploring, analysing and drawing insight from complex data that can be both structured and unstructured.
- To apply data visualisation techniques to explore and summarise findings when analysing data.
- To apply algorithms to build data-based mathematical models.
- To understand, assess and compare predictive models.
- To select appropriate analytical techniques to solve problems relevant to a data scientist in government and the public sector.
- To apply quantitative modelling techniques to the solution of real-world problems.
- To employ technologies to analyse Big Data.
- To communicate findings from data to allow an organisation to make more informed business or policy decisions.
- To gain knowledge in the appropriate and responsible use of data in government and the wider public sector.
- To be able to continue learning and using new data science technologies.

Course structure

Training usually takes place during the second week of each month, and includes:

Guest lectures

We provide case studies from industry and related fields professionals to supplement trainer-led sessions and independent learning. This helps to introduce topics, reinforce the learner's understanding and embed new concepts within domain-specific contexts.

Ethics readings

The digital and data sectors are currently undergoing a period of rapid innovation. With this increased capability, opportunities to use data for public good are becoming more frequent. However, we must also mitigate risk where potential to misuse data and techniques present themselves. Algorithmic decision-making is a burgeoning area of innovation with consequences for us all.

The Graduate Programme presents opportunities to reflect upon and discuss the potential use and misuse of the skills, tools and techniques in data science. We provide readings to give participants awareness of ethical issues beyond the technical challenges of data science. These readings are intended to complement and build on ethical issues and topics presented in many of the courses, including Statistics in R or Python and Ethics and Governance. Through these readings, we aim to promote a questioning and thoughtful culture of reflection that students can apply to their technical assignments.

Trainer-led sessions

Participants will join a group call at a designated time with their cohort where a trainer will facilitate an interactive session. The trainer supports the participants working through the material together.

Independent learning

Participants will access the learning material on the Learning Hub and work through it at their preferred pace during their designated learning time.



Review question and answer sessions

Following an independent learning session, participants will be able to ask questions relating to the material with their cohort and trainers. Questions can be submitted in advance on the course Slack channel or in person on the day.

Assignment or practice activities

Participants will be given case studies, example projects and questions to tackle. These practical assignments will help cement knowledge and embed the skills learned. The trainers will either provide direct feedback or solutions or a rubric for self-assessment. There will also be some assignments where presentations, peer feedback and discussion will take place. Any assignments are for the benefit of the learner and to provide experience tackling real problems and working as a team. Assignments do not contribute to an overall grade but are used to help the learner build transferable skills and understanding.

Mentoring

If interested, participants will meet with a mentor on a regular basis, in small allocated groups across their cohort.

Curriculum overview:

Intro to Python / R

Command line basics

Introduction to Git

Clean code

Data Viz in Python / R

Statistics in Python / R

Introduction to RAP

Reproducible reporting using RMarkdown

Modular code

Unit testing

Machine learning in Python / R

NLP in Python / R

Quality assurance of predictive modelling

Introduction to Geospatial for data science

Foundations of SQL

Introduction to Pyspark / SparklyR

Group and individual projects

What to do if you have a problem

Your first port of call if you have a question is the frequently asked questions section of this booklet, on [page 10](#). If you don't find the answer to your question in the frequently asked questions, you can get help from your cohort, your home organisation, or the faculty.

Your cohort can help if...

- You want to double check a date or prerequisite for upcoming training.
- You want advice about a project you are working on at work.

Your home organisation can help if...

- You are having trouble downloading a package.
- You can't access a website, resource or dataset that you need for upcoming training.
- Software that you need for upcoming training is not available or not working.
- You have a different IT issue.
- You have a question about an HR matter.

The faculty can help if...

- You don't understand some of the course content.
- You don't understand the brief for an assignment.

Communicating with the faculty on Slack



A dedicated Slack workspace will be set up for communication with your cohort and the faculty. Slack is the main platform to be used for contacting the faculty. The Slack workspace contains a variety of channels.

General: for general enquiries and course announcements.

Resources: to share resources and interesting articles.

Module channels: a channel per module where you can post questions related to the course content and briefs for an assignment. The faculty will endeavour to answer these questions during the Q&A session. This channel will also provide a space where peers can answer each other's questions.

There will also be a course channel per action learning set where you can connect with your peers and plan your action learning set sessions.

IT requirements

If you do not have access to any of the minimum requirements outlined below, contact your line manager or the IT department of your home organisation.

Description	Minimum requirements
R distribution, RStudio preferred	R v4.0+ An IDE for R (RStudio, Visual Studio Code etc.)
Python distribution, Anaconda preferred	Jupyter Notebook access An IDE for Python (Spyder, Visual Studio Code, PyCharm etc.) Python 3.6+
Python & R packages / libraries	Ability to install relevant Python and R packages for data science. This may require permissions from IT departments.
Command line interface	Access to a command line interface, with appropriate permissions for programming tasks. Preferred: Windows Command Prompt, Anaconda Prompt
Version control	Access to Git, through a command line interface. Preferred: GitBash
IT support	Installation, package management, troubleshooting and environment configuration. A data scientist within the organisation to support with problems raised with the above.
ONS Learning Hub	Access to the Learning Hub, appropriate network permissions to download files from the website.

Frequently asked questions

How do training sessions work?

Training is delivered as a mixture of self-directed study, trainer-led sessions or workshops and group activities. Independent study courses are followed by a Q&A session where you have opportunities to review the course material and problem-solve as a group.

What do I do if I can't attend an upcoming training session?

You should make every effort to attend all training sessions, but if you are unable to attend you should contact Data.Science.Campus.Faculty@ons.gov.uk with the subject "Graduate Programme" as soon as possible. There may be self-study materials that you can use to catch up on the key points of the session.

What do I do if I don't have access to software needed for a training session?

You should promptly run any scripts or commands provided by the trainer to test the functionality of your installed software. Technical support on the day of training cannot be provided by trainers as this will delay the course for others and detract from their training experience. If you are missing a required software or package, contact your line manager or your home organisation's IT department.

Should I focus on learning Python or R?

Although all of the courses are offered in either Python or R, it would be beneficial for you to become comfortable using both languages. For courses where you already have some familiarity with the subject matter, you would benefit most from taking the course that is not in your preferred language.

What does day to day working look like?

Your work-based experience will vary considerably depending on the context of the organisation where you are based. However, you should expect to be treated in line with the organisation's policies for fixed-term employees. You should also be provided with the necessary support from your line manager in complying with your responsibilities and managing your workload. You should expect to be allocated work that complements the graduate training programme, providing you with an opportunity to promptly apply your learning from training, helping to further your understanding and skills.

What kind of support do graduates get?

We hope that you will support and discuss your learning with fellow graduates from your cohort. The faculty are here to support you during training sessions, so you can ask for help if you are stuck on an assignment or do not understand the materials. Your home organisation will have further support in place, including your line manager, the HR team, and other pastoral support (for example, PAM assist for ONS staff). Speak to your line manager about your specific needs to find out who can help.

Where can I find the course materials?

All your course details and materials can be found on the Learning Hub Graduate Programme page. Please be aware that the materials may be updated during the year so what you view before training sessions is subject to change.

As part of this Graduate Programme, you will be given a license to use the learning platform. If you have difficulties logging on to the platform or accessing training materials, please email Data.Science.Campus.Faculty@ons.gov.uk with "Graduate" Programme on the subject line.

How does year 2 of the programme differ from year 1?

Year 2 of the programme focuses on building your professional data science skills experience while you develop a portfolio of evidence in support of your career as a data scientist. There will be planned learning days, as there was in your first year, and these will cover topics related to working within a team, management of a programme and other skills needed to work in data science in government or public organisations.

How do I get the required packages for each course?

The packages required for each course are listed on the Learning Hub. Your organisation will also have a list of all packages you will need, as well as installation instructions. Please be aware that each organisation may curate and allow access to software differently; if you run into issues, please consult someone in your organisation first. To prevent problems with different package versions, we recommend using a package management system, such as ``conda env`` or ``venv`` in Python or ``renv`` in R.

Questions about the Data Science Graduate Programme

Please email Data.Science.Campus.Faculty@ons.gov.uk
with the subject "Graduate Programme"

Contact us

Data Science Campus, Office for National Statistics, Government Buildings
Cardiff Road, Newport, NP10 8XG

To find out more information about our projects and services, visit
datasciencecampus.ons.gov.uk, email atacampus@ons.gov.uk
or follow us on Twitter [@DataSciCampus](https://twitter.com/DataSciCampus)