

## Technical skills for MSc Data Science

Our MSc in Data Science welcomes students with a range of backgrounds. We deliberately recruit a diverse range of students with a wide variety of experiences, but expect a certain level of technical competency in both coding and mathematics. We only accept students who we think have a background that is adequately technical.

If you are less experienced in these areas you will need to do some preparation before starting the course and be prepared to work hard, particularly in the earlier stages of the programme.

Here, we explain what level you should be at to be successful.

### Will I have to write code?

Yes. You should expect to be writing code throughout this course and in more than one programming language. It is the most flexible and powerful way to manipulate data and apply techniques used in Data Science. If you don't want to write code, this is not the programme for you.

### How much coding will I be doing?

You will not be building software, but you will be using code to manipulate, analyse and visualise data and run algorithms on the data. You will be doing this mainly in **Python** and **Matlab**, but might also use R, other languages and different software tools.

The amount of code you will have to write will depend on which elective modules you choose, but it will not be at the level needed for software engineering. This is not a programming degree.

### How much coding experience do I need when I start and how do I prepare?

We expect you to already know the basics of programming languages, which includes data types, variables, conditional statements and control flow, use of functions and parameter, lists, loops, classes and file input/output. The amount of preparation will depend on your experience.

#### Python

You'll need to be familiar with the basics of Python before you start, including the Pandas, Numpy, Matplotlib and Seaborn libraries. There are many good online resources, but we recommend [LearnPython](#) or DataCamp's [Intro to Python](#) (use the free tutorial until we will give you free access to the Premium material as indicated below).

We also recommend the [Python for Data Analysis](#) book by Wes McKinney. More specific recommendations are made at the bottom of this page.

## Matlab

We will also be using Matlab for two core modules (Machine Learning and Neural Computing), so you should be familiar with the basics. Matlab is not free, so please download a [30 day Matlab trial](#) and follow these tutorials: "[Getting Started with MATLAB](#)" and "[Matlab Onramp](#)" to familiarise yourself with the Matlab environment.

We will give offer holders free access to DataCamp in the summer before the programme starts so that they can work through some of the material in advance.

## How much statistical and mathematical knowledge do I need?

Basic statistical and mathematical concepts are required. This does not need to be very advanced, but some of the topics will be easier to understand with more advanced mathematical knowledge.

Since we deal with data, we expect you to have a basic understanding of numerical distributions, basic summary statistics, correlations and probability theory and some concepts you might need to be familiar with can be found in these resources:

- [MIT: Statistics Cheat Sheet](#)
- [Maths essentials for Machine Learning](#)

You will be learning and applying algorithms during the course and an understanding of how the algorithms work will be necessary. A basic understanding of linear algebra, matrix operations, and derivatives will help here.

Some recommended resources are at the end of this page.

## Will you run preparation sessions?

Yes. We plan to run a couple of preparation sessions that go through the basics of mathematics and programming.

- **Python:** You'll need a working version of Python. We recommend [Anaconda](#).
- **Matlab:** You'll need a working copy of Matlab, probably the 30-day free trial.
- **Mathematics:** You'll need a working copy of Matlab, probably the 30-day free trial.

These sessions will run in late August or early September. More details will follow closer to the time.

## How much help will I get?

We will provide some optional preparatory sessions on coding and mathematical basics in late summer and at the start of the course.

We have three full-time teaching assistants who will provide help and will run scheduled surgery sessions throughout the programme. But we expect you to try things on your own and to form your own study groups.

We offer plenty of help, but also expect students to organise their own learning.

## Online resources

You may benefit from the following training resources provided to students on this course, free of charge.

### DataCamp

[DataCamp](#) is an online training platform providing you with tutorials and challenges in Python and other data science technologies. We will give offer holders free access to DataCamp in the summer before the programme starts so that they can work through some of the material in advance.

\*Please note this resource is subject to renewal on a six-monthly basis.

### MATLAB training

In partnership with MATLAB's MathWorks, postgraduate data science students can take the [MathWorks](#) online training.

Upon successful completion of the online training in MATLAB you will be competent to pass the MathWorksCertified [MATLAB Associate Exam](#).

## Links with industry

[Cynozure](#) - our Industry advisor - will provide talks and workshops that relate to industry-facing aspects of Data Science, involving participation from their industry contacts.

"[Data Bites](#)" is our special seminar series that regularly feature employers in the data science market presenting their companies and job opportunities.

## Suggested online courses

These resources provide some more in-depth material that you will find helpful.

### Python

- [Introduction to Python](#)
- [Introduction to Data Science in Python](#)
- [Python with anaconda](#)
- [Introduction to Python Programming](#)
- [Importing Data in Python](#)
- [Cleaning Data in Python](#)
- [Practicing Coding Interview Questions in Python](#)
- [Preprocessing for Machine Learning in Python](#)
- [Data analysis and visualisation](#)
- [Statistical Thinking in Python \(Part 1\)](#)
- [Statistical Thinking in Python \(Part 2\)](#)

- [Linear Classifiers in Python](#)
- [Big Data using PySpark](#)
- [Spark](#)
- [Analyzing marketing campaigns with Pandas](#)

## Matlab

- There are many online courses in the [Matlab Academy](#) some are free, and you could start before joining
- [Matlab Central](#) is a hub for “open exchange for the MATLAB and Simulink user community”.

## Probability

- [StatQuest](#) statistics explained
- [Khan Academy: Independent and Dependent Events](#)
- [Khan Academy: Probability and Combinatorics](#)
- [Khan Academy: Random Variables and Probability Distributions](#)

## Statistics

- [Khan Academy: Displaying and Describing data](#)
- [Khan Academy: Modeling Distributions of data](#)
- [Khan Academy: Describing relationships in quantitative data](#)
- [Khan Academy: Confidence Intervals](#)
- [Khan Academy: Significance Tests](#)

## Linear algebra

- [Khan Academy: Vectors](#)
- [Khan Academy: Matrix Multiplication](#)
- [Khan Academy: Determinants](#)

## Further enquires

Please contact [smcse.msc@city.ac.uk](mailto:smcse.msc@city.ac.uk) if you have any other questions.