

# Data Science Foundations Syllabus

## Specialisation description

This specialization is aimed at a general audience who may be interested in training for a career in Data Science either through an academic or a professional route. It could lead straight to employment in data-rich environments or into a degree course in Computer Science or Data Science. The specialization will introduce learners to data science, statistics, programming, computational thinking, machine learning, and the structure and uses of computers. Students will learn current Data Science and Computer Science topics from a professional and from an academic perspective.

This specialisation consists of eight courses aiming to help you to become confident working in Python within the Data Science field. During the eight courses, you will work with multiple instructors on various aspects of Data Science: Whilst taking the courses, you will learn about a range of programming techniques from variables, basic control flow and input/output to data structure, algorithms, and Data Science techniques to solve real-life problems.

## Specialisation learning outcomes

Upon successful completion of this Specialisation, you will be able to:

1. Demonstrate skills to begin data science careers
2. Choose appropriate basic datatypes to represent different data
3. Understand and use clustering algorithms such as K-means clustering
4. Explain in broad strokes what problems and algorithms are in Computer Science
5. Identify the users and benefits of Python
6. Describe how to use a learning model to solve a problem
7. Understand and use techniques such as flowcharts and pseudocode
8. Develop Python code to manipulate data
9. Utilize data science methodologies to define and formulate a real-world business problem
10. Understand a career in Data Science either through an academic or a professional route

## Specialisation outline

The Specialisation consists of 8 courses, each of which spans two weeks.

<p><b>Course 1 - This course</b></p> <p>This course introduces you to Data Science from the viewpoint of a student studying at university. This course, designed by an academic team from Goldsmiths, University of London, will quickly introduce you to the core concepts of Data Science to prepare you for intermediate and advanced Data Science courses. You will also hear from students who are currently or just completed their studies in Data Science.</p>	<p><b>Key concepts:</b></p> <ul style="list-style-type: none"> <li>• Data</li> <li>• Machine Learning</li> <li>• Clustering</li> </ul> <p><b>Learning outcomes:</b></p> <ul style="list-style-type: none"> <li>• What we mean by data</li> <li>• What we mean by machine learning</li> <li>• How the K-Means clustering algorithm works</li> <li>• Tips and experiences from other students</li> </ul>
<p><b>Course 2</b></p> <p>This course introduces you to Data Science from the viewpoint of a professional practitioner. The course comes from the IBM Professional Certificate course <i>What is Data Science</i>. That course gives a general overview of professional data science practice. Part of that is provided by interviews with practicing professionals. This course will also give a taste of the more hands-on learning to come in the specialization.</p>	<p><b>Key concepts:</b></p> <ul style="list-style-type: none"> <li>• Definition of Data Science</li> <li>• Big data</li> <li>• Watson Studio</li> </ul> <p><b>Learning outcomes:</b></p> <ul style="list-style-type: none"> <li>• Describe the range of professional data science practice</li> <li>• Describe the uses and limitations of some important Data Science tools and systems, including Watson Studios</li> <li>• Explain the difference between deep learning and machine learning</li> </ul>
<p><b>Course 3</b></p> <p>This is a course on the IBM Professional Certificate. In this course, you will learn some of the most important tools in data science, taught from an IBM perspective. You will learn about Jupyter Notebooks, JupyterLab, RStudio, and continue your study of Watson Studio. You will also learn about what each tool is used for, what programming languages they can execute, their features and limitations.</p>	<p><b>Key concepts:</b></p> <ul style="list-style-type: none"> <li>• IBM Toolkit</li> <li>• Languages of Data Science</li> <li>• Jupyter Notebooks and Watson Studio</li> </ul> <p><b>Learning outcomes:</b></p> <ul style="list-style-type: none"> <li>• Recognise and do some rudimentary design in several programming languages, such as Python, R, and SQL</li> </ul>

<b>Course 4</b>  This course will introduce you to your first algorithms and data structures, as well as the tools of abstraction required to help you decide which of these concepts to use. In this way you will not only enhance your box of problem solving tools, you will be able to critically compare, and assess, the advantages and disadvantages of these tools. The course starts at a basic level by describing problems and algorithms in computer science. It then goes on to introduce the tools for describing algorithms: flowcharts and pseudocode.	<b>Key concepts:</b> <ul style="list-style-type: none"> <li>• Algorithmic thinking</li> <li>• Problem Solving</li> <li>• Flowcharts and pseudocode</li> </ul> <b>Learning outcomes:</b> <ul style="list-style-type: none"> <li>• Explain in broad strokes what problems and algorithms are in Computer Science</li> <li>• Recall the basic elements and construction of flowcharts</li> <li>• Express elements of simple algorithms as flowcharts and pseudocode</li> </ul>
--	---

<b>Course 5</b>  This course is taken from the IBM Professional Certificate in Data Science. It provides an introduction to the python programming needed in data science. You will learn Python fundamentals, including data structures and data analysis through a series of lectures and practical hands-on exercises.	<b>Key concepts:</b> <ul style="list-style-type: none"> <li>• Python basics</li> <li>• Data Structures</li> <li>• Programming fundamentals</li> </ul> <b>Learning outcomes:</b> <ul style="list-style-type: none"> <li>• Build a program in JupyterLab to demonstrate your knowledge of types, expressions, and variables</li> <li>• Create sets to demonstrate understanding of the differences between sets, tuples, and lists.</li> <li>• Use exception handling in Python</li> <li>• Read text files using Python libraries</li> </ul>
---	--

<b>Course 6</b>  This course gives another account of core concepts of Data Science, this one more focussed on algorithm design. It develops the core mathematics, statistics, and programming skills necessary for typical data analysis tasks. You will do a series of mathematical and programming exercises throughout the course and complete a small data clustering project for a given dataset using the K-mean algorithm.	<b>Key concepts:</b> <ul style="list-style-type: none"> <li>• Mean and Deviations</li> <li>• One and two-dimensional data</li> <li>• Pandas and K-means</li> </ul> <b>Learning outcomes:</b> <ul style="list-style-type: none"> <li>• Understand key mathematical and statistical concepts pertinent to data clustering</li> <li>• Integrate the key steps of data clustering with K-means.</li> <li>• Analyse and explain the results of the data clustering.</li> </ul>
--	---

<b>Course 7</b>	<b>Key concepts:</b>
-----------------	----------------------

<p>In this course, you will create a complete system in Python which aims to solve a particular data-science problem using regression. Through the course, you will learn the life cycle of a project from planning to implementation.</p>	<ul style="list-style-type: none"> <li>• Project Life Cycle</li> <li>• Correlation</li> <li>• Regression</li> </ul> <p><b>Learning outcomes:</b></p> <ul style="list-style-type: none"> <li>• Plan a substantial project</li> <li>• Clean and prepare data for analysis</li> <li>• Implement a data science system</li> <li>• Communicate the plan and implementation of a project in writing</li> </ul>
--	--

<p><b>Course 8</b></p> <p>This is the final course in the IBM Data Science Professional Certificate as well as the Applied Data Science with Python Specialization. This capstone project course will give you the chance to practice the work that data scientists do in real life when working with datasets. In this course you will assume the role of a Data Scientist working for a startup intending to compete with SpaceX, and in the process follow the Data Science methodology involving data collection, data wrangling, exploratory data analysis, data visualization, model development, model evaluation, and reporting your results to stakeholders.</p>	<p><b>Key concepts:</b></p> <ul style="list-style-type: none"> <li>• Real-world data</li> <li>• Data Science methodology</li> <li>• Exploratory data analysis</li> </ul> <p><b>Learning outcomes:</b></p> <ul style="list-style-type: none"> <li>• Demonstrate proficiency in data science and machine learning techniques using a real-world data set and prepare a report for stakeholders</li> <li>• Apply your skills to perform data collection, data wrangling, exploratory data analysis, data visualization model development, and model evaluation</li> </ul>
---	--

### Activities in the specialisation

The course is comprised of the following elements:

- **Lecture videos.** In each course, you will find a sequence of theoretical and practical videos in which the instructors explain the main takeaways of each individual course.
- **Readings.** Each course may include several suggested readings. These are a core part of your learning, and, together with the videos, will cover all of the concepts you need for this module.
- **Practice Quizzes.** Each course will include practice quizzes, intended for you to assess your understanding of the topics. You will be allowed unlimited attempts at each practice quiz. There is no time limit on how long you take to complete each attempt at the quiz. These quizzes do not contribute toward your final score in the class.
- **Programming Activities.** Each course includes programming activity worksheets. These take you through the steps you have seen in the videos and provide code

excerpts. They also contain challenges activities which challenge you develop the program beyond the functionality seen in the lecture videos.

- Discussion Prompts. Each course includes discussion prompts. You will see the discussion prompt alongside other items in the lesson. Each prompt provides a space for you to respond. After responding, you can see and comment on your peers' responses. All prompts and responses are also accessible from the general discussion forum and the module discussion forum.
- Assessed coursework. There are several assessed activities for each course. They consist of summative quizzes.

### How to pass this specialization

Each course has two or more quiz assessments each worth a percentage of your final grade.

Activity	Required?	Estimated time per course	% of final grade
Summative Quiz	Yes	Between 10 and 40 minutes	Depends on each course format

### Specialisation Readings

There are no specific textbooks for this Specialisation that you need to read to successfully obtain your certification. Instead, there are reading activities written by the course author, some of which involve coding exercises. The specific URL links for the reading activities will be given in the platform, and there is no need to read beyond to recommended pages. There will also be discussion prompts asking you to do some independent research using online sources.