

Introduction to Python for Data Science

Dr Kieran Hughes
ATU Sligo
January 2024

Why Python for Data Science?

Python has been widely adopted as one of the primary programming languages in the field of data science and is often the language of choice in educational settings. Its popularity in this domain has continued to grow due to several factors which include:

- Rich Ecosystem of Libraries
- Relative Ease of Learning
- Open-Source Nature and Community Support
- Versatility and Job Market Demand
- Data Visualization and Integration Capabilities
- Jupyter Notebooks

Rich Ecosystem of Libraries

Python has a vast and powerful ecosystem of libraries and frameworks specifically tailored for data science and machine learning. The list of libraries include:

- NumPy
- pandas
- Matplotlib
- Seaborn
- scikit-learn
- TensorFlow
- PyTorch

These libraries provide tools for data manipulation, analysis, visualisation, and machine learning, making Python a go-to language for data scientists.

Relative Ease of Learning

- The language emphasises code readability, with a clean and straightforward syntax that reduces the cost of program maintenance and development. This makes it easier for beginners to understand and learn.
- Python allows developers to express concepts in fewer lines of code compared to languages like C++ or Java. This makes it easier to write, read, and maintain code, reducing the cognitive load for programmers.
- Python is an interpreted language, which means that there is no need for compilation before running the code. This allows for quick iteration and testing, facilitating a more immediate and interactive learning experience.
- Python is a strongly typed language, meaning that variable types are enforced, preventing unexpected behaviour. However, it is dynamically typed, allowing developers to write code more flexibly without explicitly declaring variable types.

Open-Source Nature and Community Support

Open-Source Nature

Python is an open-source language with a large and active community contributing to its development. This open-source nature fosters innovation and ensures that the language remains up-to-date with the latest advancements in data science and technology.

Community Support

Python has a large and active community of data scientists and developers. This community support is invaluable for students seeking help, sharing knowledge, and collaborating on projects. Online forums, tutorials, and open-source resources contribute to a supportive learning environment.

Versatility and Job Market Demand

Versatility

Python is a versatile language that goes beyond data science. It is widely used in web development, automation, scripting, and other domains. Learning Python provides students with a skill set that is applicable in various industries, increasing their career flexibility.

Job Market Demand

Many job postings for data science and machine learning positions specifically mention Python as a required or preferred skill. Learning Python increases a student's marketability and chances of securing relevant positions in the data science job market.

Data Visualisation and Integration Capabilities

Data Visualisation Capabilities

Python's visualisation libraries, such as Matplotlib and Seaborn, enable students to create compelling and informative data visualisations.

Visualisation is a crucial aspect of data science for conveying insights to stakeholders, making Python a valuable tool in this regard.

Integration Capabilities

Python seamlessly integrates with other technologies commonly used in data science, such as databases, big data tools (e.g., Apache Spark), and web frameworks. This integration facilitates end-to-end data science workflows and makes Python a practical choice for real-world applications.

Jupyter Notebooks

Jupyter Notebooks

Jupyter Notebooks, a popular interactive computing environment, is often used with Python in data science. It allows students to create and share documents containing live code, equations, visualisations, and narrative text, making it an effective tool for learning and presenting analyses.

Python Basics

- The `print()` statement, commenting and accessing help
- Variables and data types
- Basic arithmetic operations
- Data Structures (lists, tuples, and dictionaries)
- Control structures (if, for, while)
- Functions