Python for Data Science – Data speaks, Python listens!

***Free Python course with 35 real-time projects***[***Start Now!!***](https://data-flair.training/python-course/)

As you must know by now, it is a great choice to do data analysis using Python. This is why data scientists prefer Python.

Out there, there’s a battle taking place in minds of**future Data scientists** for choosing the best tools. Though there are quite a number of tools with many options, the close combat narrows down between two popular languages – Python and R.

Why Python over R for Data Science?

One of the push of choosing Python over R is from the variety of data science/data analytics libraries made available. Some of the libraries well known in the data science community – Pandas, StatsModels, NumPy, SciPy, and Scikit-Learn. It doesn’t stop there, some 72,000 of them in the Python Package Index (PyPI) and still growing constantly. After all these, I recommend you to check the [***difference between Python and R***](https://data-flair.training/blogs/r-vs-python-for-data-science/) for better understanding

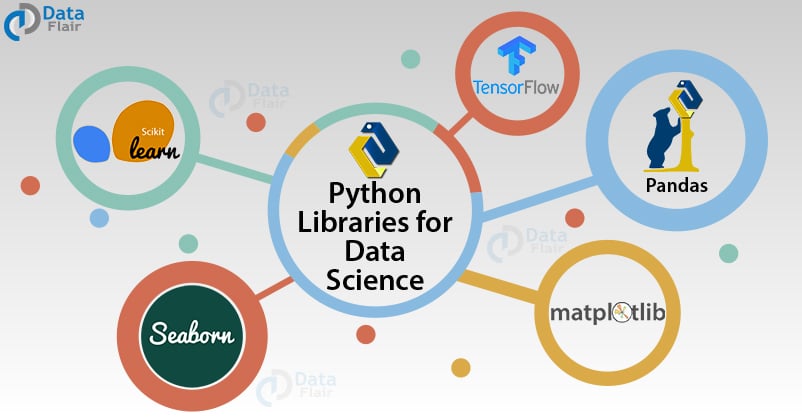
So, before we start our topic I recommend you to take a brief of what data science is?

What is Data Science?

Data science, aka data-driven science, is an interdisciplinary field of scientific methods, processes, and systems. It is used to extract knowledge or insights from data in various forms, either structured or unstructured. In this way, it is similar to data mining. With data at its heart, it employs a wide range of techniques on the data to extract essential insights from it.

Data science attracting beginners towards Python. How?

For this reason and others,[***Python is the most demanding for programmers***](https://data-flair.training/blogs/why-is-python-in-demand/). Data scientists coming from engineering or scientific backgrounds might feel a bit out of place the first time they try to use it for data analysis but when they do use it, they make most out of it. Python didn’t make its inroads to data science initially when it was conceived in the late 1980s. Tools for covering every aspect of scientific computing are now readily available in Python.

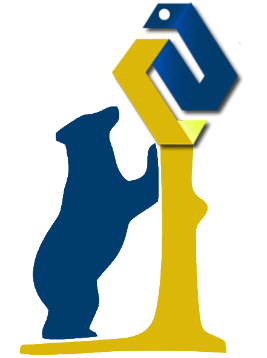
[](https://data-flair.training/blogs/wp-content/uploads/sites/2/2018/02/Python-Libraries-for-Data-Science2.jpg)

Python’s readability and simplicity make it comparatively easy to pick up. The number of dedicated and analytical libraries freely available for download today mean that data scientists present in every sector will find packages already tailored to their needs. As a jack of all trades, Python is not specialized to do statistical analysis, but in many cases, organizations already have heavily invested in extending it to that purpose as they saw advantages of standardizing on it. In short, we can say, Python has become the go-to language for data scientists. And you must **start learning Python now**! Our DataFlair team has designed a perfect[***self-paced course of Python***](https://data-flair.training/python-course/) for Data Science and the passionate learners like you. Get the course now and move a step ahead in the data science field.

Essential Python Libraries for Data Scientists

***Data science has the early benefits of these extensions and libraries!***

1. Python Pandas

[](https://data-flair.training/blogs/wp-content/uploads/sites/2/2019/07/Python-Pandas-logo.png)

Now the big daddy to all of them is [***Python Pandas***](https://data-flair.training/blogs/pandas-tutorials-home/). From importing data from spreadsheets to processing sets for time-series analysis, Pandas is used for everything. Pandas pretty much convert one data form to another on your fingertips. Hence, Pandas powerful data frames can perform both, basic cleanup and advance data manipulations.

***“One of the reasons we like to use Pandas is because we like to stay in the Python ecosystem,” Burc Arpat, a quantitative engineering manager at Facebook.***

Behind Python’s data science success story, one of the earliest libraries is Numpy (Numerical Python), on which Pandas is built. NumPy’s functions exposure is used in Pandas for advanced analysis. For more specialization, one can use Scipy which is scientifically equivalent to Numpy, offering tools and techniques for scientific data analysis.

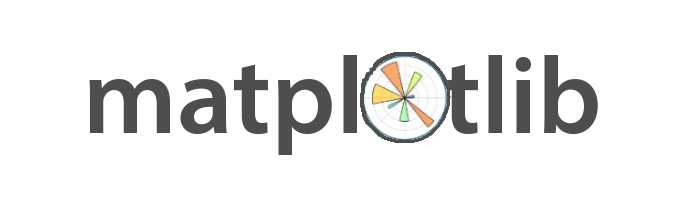
a. NumPy

NumPy facilitates easy and efficient numeric computation. It has many other libraries built on top of it. Make sure to learn NumPy arrays.

b. SciPy

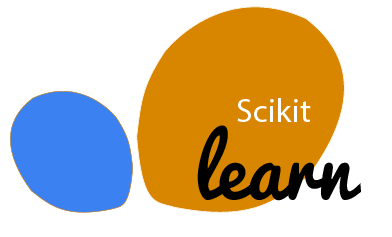
[SciPy](https://www.scipy.org/) will give you all the tools you need for scientific and technical computing. It has modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers, and other tasks.

2. Matplotlib

[](https://data-flair.training/blogs/wp-content/uploads/sites/2/2019/07/matplotlib.png)

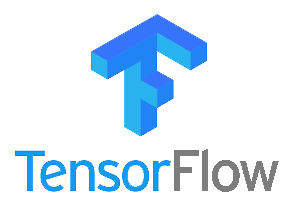
Python also provides powerful visualization libraries – [***Matplotlib***](https://data-flair.training/blogs/python-matplotlib-tutorial/). It can be used in all kinds of GUI toolkits such as python scripts, web applications as well as shell, etc. With this, you have the opportunity to use different types of plots and work with multiple plots.

3. Scikit – Learn & Pybrain

[](https://data-flair.training/blogs/wp-content/uploads/sites/2/2019/07/scikit-learn-logo.png)

Scikit – Learn & Pybrain, one of the attractions of python where you implement machine learning. With the support of simple and efficient tools in this library which can be used for data analysis and data mining. Various algorithms have their back, like — logistic regression, time series, etc.

4. TensorFlow

[](https://data-flair.training/blogs/wp-content/uploads/sites/2/2019/07/Tensorflow-Logo.png)

[***TensorFlow***](https://data-flair.training/blogs/tensorflow-tutorials-home/) is the most popular tool for Machine Learning in Python. It was developed specifically for carrying out deep learning operations. The basic data structure of TensorFlow ecosystem are the tensors. As a matter of fact, the name of TensorFlow is derived from these tensors. TensorFlow is continuously evolving owing to an open-source community who have made it a pioneering toolkit for machine learning operations. It provides support for CPUs, GPUs as well as TPUs. Due to this, it provides lightning speed execution speed for various machine learning algorithms.

[***TensorFlow has numerous applications***](https://data-flair.training/blogs/tensorflow-applications/). This is mainly because of its high processing capability. It is used for the development of speech recognition product, recommendation systems, Generative Adversarial Networks, etc. TensorFlow is basically the standardized tool for performing Deep Learning operations.

5. Seaborn

[](https://data-flair.training/blogs/wp-content/uploads/sites/2/2019/07/Seaborn-logo.png)

Welcoming the next important library of Python for Data Science – Seaborn! So whatever and whenever you will be using Python for data science, the first thing that will click to and should click will be using matplotlib (for 2D visualization) & seaborn. They have many high-level interfaces and styles in default for drawing statistical graphics.

Python is an obvious language choice for[***Data Science***](https://data-flair.training/blogs/data-science-tutorials-home/). These above-stated libraries and other specialized one’s aid everything in python, from machine learning to neural networks to data processing. Hence this flexibility has become the main benefit of choosing python at every step of the way towards data science.

Python’s large community is taking Data Science to the top!

Another plus point to above all the extensions/ libraries and properties of python that contribute to “python being the choice” is the large community of data scientist, machine learning experts, and programmers who are not only working their heart out to make it easy to learn python but also provide datasets to test one’s mastery and skillset in python. So, whether you are a social scientist needing python for advanced data analysis or a growing developer who needs inspiration, one of the parts of this python community will be ready to help you!