5 Steps to Learn Python for Data Science

data-flair.training/blogs/python-for-data-science

February 22, 2018

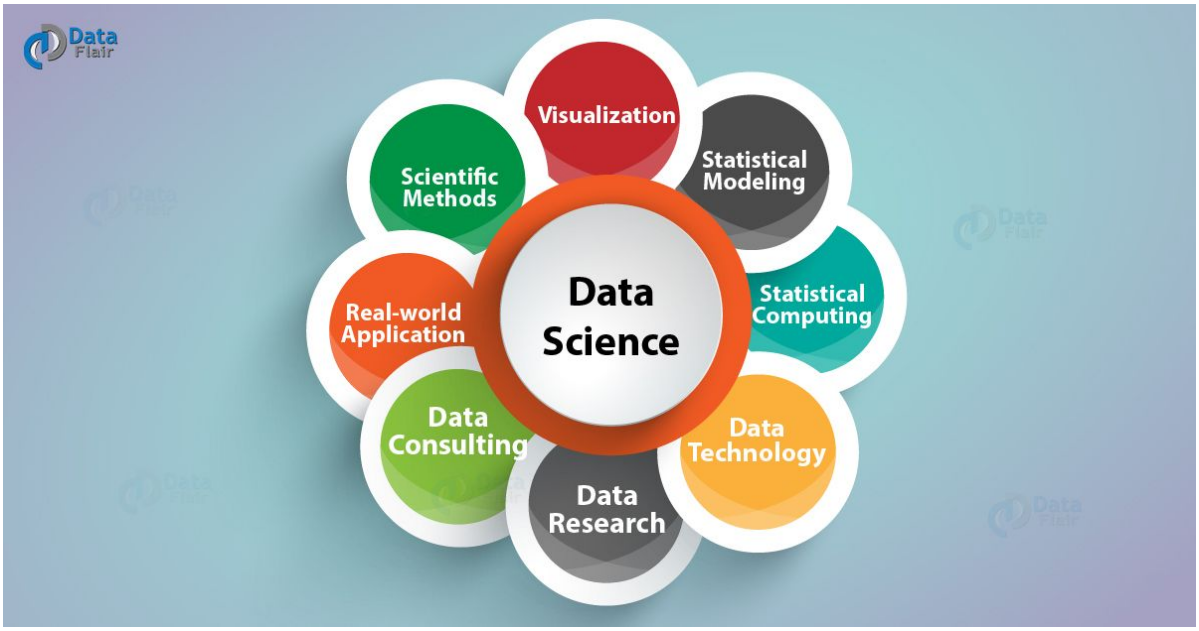
Contents

1. Python for Data Science

As you must know by now, it is a great choice to do data analysis using Python. This is why data scientists prefer Python. We talked about this when we discussed **Career Opportunities in Python**; let's see why Python for Data Science is preferred.

2. 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.



Python for Data Science

This was a brief Introduction to Data Science. If you choose to set out on**Python for Data Science**, we've compiled a to-do list for you:

3. Learn Python for Data Science – The Basics

To step into the world of Python for Data Science, you don't need to know Python like your own kid. Just the basics will be enough.

- **Python Lists**
- **List Comprehensions**
- **Python Tuples**
- **Python Dictionaries and Dictionary Comprehensions**
- **Decision Making in Python**
- **Loops in Python**

4. Set up Your Machine

To gear up with Python for Data Science, we suggest **Anaconda**. It is a freemium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing. You can download it from Continuum.io. Anaconda has all you need for your data science journey with Python.

5. Learn Regular Expressions

If you work on text data, regular expressions will come in handy with data cleansing. It is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database. It identifies incomplete, incorrect, inaccurate or irrelevant parts of the data, and then replaces, modifies, or deletes the dirty or coarse data. We will discuss regular expressions in detail in a later tutorial.

6. Essential Libraries of Python used for Data Science

Like we mentioned, there are some libraries with Python that are used for data science journey. A library is a bundle of pre-existing functions and objects that you can import into your script to save time and effort. Here, we list the important libraries that you mustn't forgo if you want to go anywhere for Python with data science.

Python for Data Science – Python Libraries

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. Pandas

One such library built on top of NumPy is Pandas. It comes in handy with data structures and exploratory analysis. Another important feature it offers is DataFrame, a 2-dimensional data structure with columns of potentially different types. Pandas will be one of the most important libraries you will need all the time.

c. SciPy

SciPy 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.

d. Matplotlib

A flexible plotting and visualization library, Matplotlib is powerful. However, it is cumbersome, so, you may go for Seaborn instead.

e. scikit-learn

scikit-learn is the primary library for machine learning. It has algorithms and modules for

pre-processing, cross-validation, and other such purposes. Some of the algorithms deal with regression, decision trees, ensemble modeling, and non-supervised learning algorithms like clustering.

f. Seaborn

With Seaborn, it is easier than ever to plot common data visualizations. It is built on top of Matplotlib, and offers a more pleasant high-level wrapper. You should learn effective data visualization.

7. Projects and Further Learning

To really get to know a technology and to learn Python for Data Science, you must build something in it. Chances are, you will get stuck on your way, and every time you get stuck, you will find your way out on your own. Start with problems available on the Internet, and build your skills. Then, come up with your own problems, and define and solve them. We also suggest that you take a good look at deep learning. It is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

This is all on Python for Data Science.