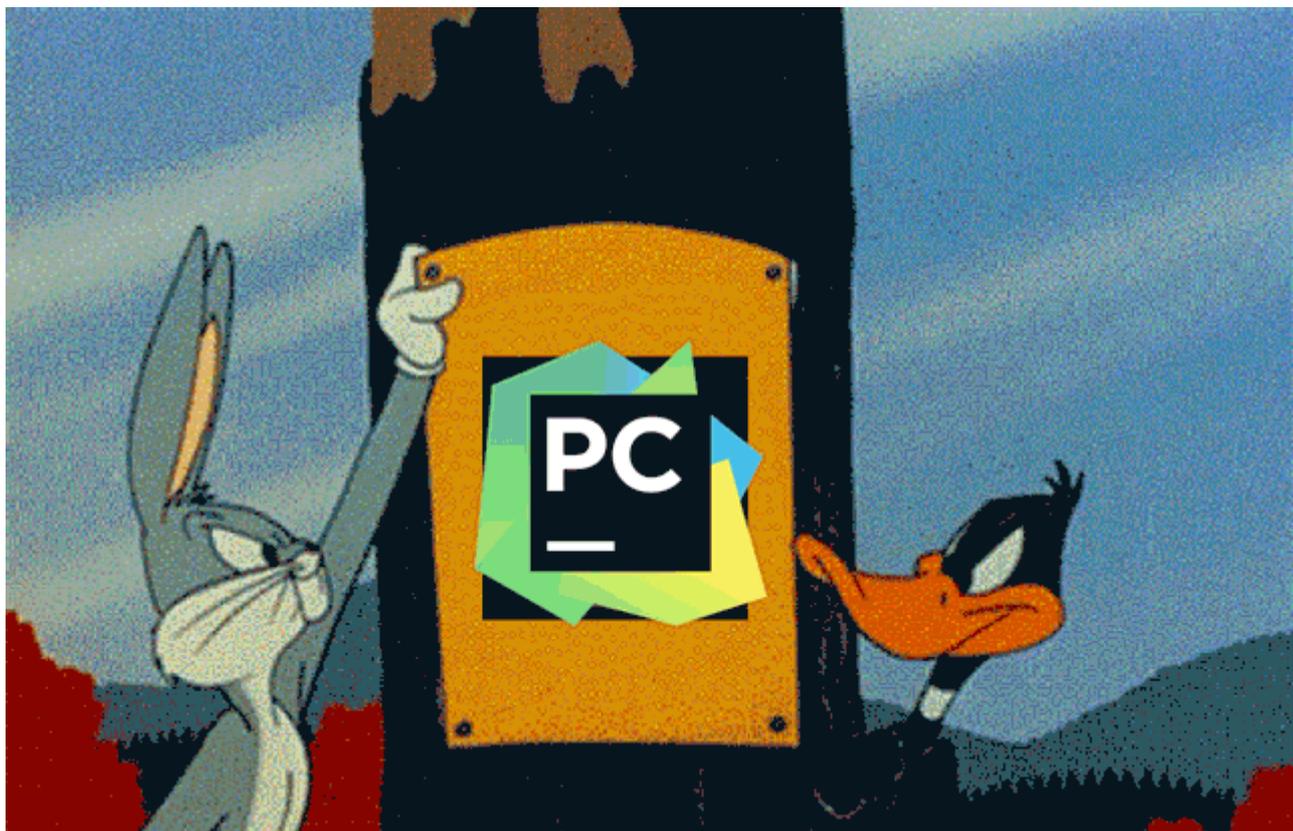


5 Must-have skills in Python for every Data Scientist

Python is a ready-to-use programming language with different packages for loading and playing around with data, visualizing the data, transforming inputs into a numerical matrix, or actual machine learning and assessment.

Data Science is intimidating and exceedingly challenging without mentorship. We hope you will learn all the important skills and also get the personalized guidance you need to land the job you want.

Most Data Scientists and Machine Learning Engineers prefer using Python for Data Science and developing artificial intelligence and machine learning apps.



If your understanding about Data Science is a big question mark, I've got a few

practical reads for you. One about [Python for Data Science Courses](#) from the World-Class Educators and one about Probability and [Statistics for Data Science](#).

Now, without further ado, let's get started !

5 Must-have skills in Python for every Data Scientist

Here are five critical skills you need to develop as a beginner and to help you develop these skills, I've linked some of the best available resources to help you become a creative data practitioner.

— Data Scraping

Gathering data from websites is one of the most logical and easily accessible sources of data.

You'll need to learn how to use Python packages like [urllib2](#), [requests](#), [simplejson](#), [regular expression operations](#), [selenium](#) and [beautiful soup](#) to make handling web requests and data

formats easier.

— SQL

You need to learn how to turn raw data into actionable insights and once you have a large amount of structured data, you will want to store and process it.

To be an effective data scientist or an engineer, you should be able to wrangle and extract data from relational databases using SQL.

I've just published a piece on Mastering [SQL for Data Science](#), give it a read if you want to learn SQL.

— Data Frames

SQL is important in data science and great for handling large amounts of data however it lacks Machine Learning and Data Visualization.

So you will have to go through the painful process of enabling Machine Learning

services in SQL Server or use MapReduce to get data to a manageable size and then process it using [Pandas](#).

— Machine Learning

A lot of data science can be done with select, join, and group by (or equivalently, map and reduce) but sometimes you need to do some non-trivial machine-learning.

Before you jump into fancier algorithms, try out simpler algorithms like [Naive Bayes](#) and [regularized, linear regression](#). In Python, these are implemented in [scikit-learn](#).

If you want to learn Machine Learning, i've got you a practical reads for you. One about A Beginner's Guide to [Machine Learning with Python](#) and one about Best [Machine Learning Courses](#) on the internet.

— Data Visualization

Data science is about communicating your findings, and data visualization is incredibly valuable part of that.

Python offers Matlab-like plotting via [matplotlib](#), which is functional, even if it is ascetically lacking and if you are really serious about dynamic visualizations, try [d3](#)

If you want to get learn Data Visualization Course, I've got you covered in this piece about [Data Visualization Courses](#).

“All these skills are taught excellently in [Data Scientist with Python Career Track](#).”



You will learn faster through DataCamp's immediate and personalized feedback on every exercise.

Thanks for making it to the end ☐

If you liked this article, I've got a few very practical reads for you. One about Learning how to [Learn Data Science](#) and one about Learning [Maths for Data Science](#).

You may also be interested in reading [How to Hire a Great Big Data Architect](#) by Toptal.

I've also got this [data science newsletter](#) that you might be into. I send a tiny email once every fortnight (if that) with some useful and cool stuff I've found/made.

Don't worry, I hate spam as much as you. 🗨️