

Top 18 Python Libraries A Data Scientist Should Know in 2023

Maximize your data science potential with the top 18 python libraries every data scientist should know in 2023, from NumPy to PyTorch and with subsections.



As a data scientist, you should be able to do different tasks like data collection, data visualization, mathematical operations, model building in machine learning and deep learning, or using web frameworks.

To do that, a wide range of libraries are available, which have too many predefined functions. They are all defined to help Data Scientists to write neater and shorter codes and achieve tasks successfully.

This article will introduce you to the top 18 Python libraries that every data scientist should know in 2023. From data collection and visualization to web frameworks, these Python libraries will help you maximize your data science potential.

What are Python Libraries?

A Python library is a collection of custom Python codes that are pre-written. The library can be imported into a Python script to make it easier and shorter.

The libraries can include functions, classes, variables, or sometimes even datasets. These libraries have wide purposes, ranging from data analysis and scientific computing to web development.

In today's article, I'll focus on the most popular libraries that you'll find very helpful in each data science stage. Using every Python library starts with the same step: importing a library.

It's done using an import statement.

```
import pandas as pd
```

Once the library is imported, you can use its functions by adding "." after the library alias. In our case, the alias for pandas is pd.

For example, if you want to use the DataFrame function in the pandas library, you should do the following:

```
import pandas as pd  
df = pd.DataFrame("")
```

If you learn more about this, here you can read [How to Import Pandas as pd in Python](https://www.stratascratch.com/blog/how-to-import-pandas-as-pd-in-python/) (<https://www.stratascratch.com/blog/how-to-import-pandas-as-pd-in-python/>).

By using libraries, you can shorten your syntax, your code becomes neater, and you save time by using the pre-written functions.

What are the major Python Libraries used in Data Science

Python has a highly active community and a large ecosystem of libraries, specifically designed for Data Science.

Here, you can see some of the most popular Python libraries for data science.

PYTHON LIBRARIES EVERY DATA SCIENTIST SHOULD KNOW 2023



data collection

Scrapy popular web scrapping library
BeautifulSoup for parsing HTML and XML
Selenium used for automating web browsers
Requests library for making HTTP requests



data visualization

matplotlib popular data visualization library
Seaborn statistical data visualization library
plotly framework for building API with Python
pandas used for data manipulation & visualization in Python



mathematical operations, analysis, ML & DL

scikit learn library in Python for ML
Keras high-level deep-learning library
PYTORCH deep-learning framework
TensorFlow ML framework developed by Google
SciPy library for scientific computation
NumPy library for working with arrays
math built-in library in Python for math operations



web frameworks

django most popular one for building apps
Flask microweb framework for Python
FastAPI framework for building API with Python

These are only some examples of the libraries available for data science in Python. Depending on your project needs, there are many more libraries you can use (and we encourage you to do that!). Yet, these are the most popular and used ones, which will help your projects to do essentials.

Python Libraries for Data Collection

The act of fetching data from different sources is called data collection. These four Python libraries offer a range of features to assist with collecting data from various sources. Let's start by examining Scrapy.

Scrapy



Zyte created Scrapy in 2008, a Python library for web scraping. The library includes a broad variety of capabilities, including data extraction from websites or multiple pages, data export to several formats, and more.

Here is the official page of [Scrapy](https://scrapy.org/) (<https://scrapy.org/>).

BeautifulSoup



Leonard Richardson created BeautifulSoup in 2004 as a Python toolkit to extract data from HTML and XML files.

It is compatible with request and other scraping libraries. The various functions of BeautifulSoup include browsing and searching through HTML documents as well as extracting data from tags and attributes.

Here is the official page of [BeautifulSoup](https://www.crummy.com/software/BeautifulSoup/bs4/doc/) (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>).

Selenium



Thoughtworks created the Selenium Python package in 2002 to be used for browser automation, testing, and scraping. It includes a wide range of functions, including the ability to fill out forms and automate browser actions. It can also be used to scrape websites.

Here is the official page of [Selenium](https://selenium-python.readthedocs.io/) (<https://selenium-python.readthedocs.io/>).

Requests



It was created in 2011 by Kenneth Reitz.

Requests can be used to interact with APIs, send HTTP requests, and handle HTTP errors.

Here is the official page of [Requests](https://requests.readthedocs.io/) (<https://requests.readthedocs.io/>).

Python Libraries for Mathematical Operations & Analysis

There are several built-in libraries for performing mathematical operations in Python, along with other libraries developed to solve mathematical operations.

These Python libraries include functions for a wide range of mathematical operations, such as trigonometric functions, linear algebra, optimization, and statistical analysis.

Now let's start exploring them by beginning with NumPy.

Numpy



NumPy is a numerical computing library for Python. It was created by Travis Oliphant in 2005. NumPy provides functions for performing operations on arrays, including mathematical, logical, shape manipulation, basic linear algebra, basic statistical operations, and more.

Here is the official page of [NumPy](https://numpy.org/) (<https://numpy.org/>).

SciPy



SciPy is a scientific computing library for Python. It was created by Eric Jones and Travis Oliphant in 2001. SciPy builds on top of NumPy and provides a wide range of numerical and scientific computing functions such as numerical integration, optimization, signal and image processing, linear algebra, statistics, and more.

Here is the official page of [SciPy](https://scipy.org/) (<https://scipy.org/>).

math

It is a built-in Python library that offers mathematical functions. There are functions for more complex mathematical operations like trigonometric functions, logarithms, and exponentials, as well as functions for simpler mathematical operations like addition, subtraction, multiplication, and division.

Here is the official page of [math](https://docs.python.org/3/library/math.html) (<https://docs.python.org/3/library/math.html>).

Python Libraries for Machine Learning and Deep Learning

scikit-learn



scikit-learn is a machine-learning library developed by David Cournapeau in 2007.

It has many different features to build classification, regression, and clustering algorithms.

Here is the official page of [scikit-learn](https://scikit-learn.org/) (<https://scikit-learn.org/>).

Keras



François Chollet developed the machine-learning library Keras in 2015 for using in Machine Learning.

It offers various capabilities for creating and improving neural networks, as well as for processing images and texts, and more.

Here is the official page of [Keras](https://keras.io/) (<https://keras.io/>).

PyTorch



PyTorch is a machine-learning library developed by MetaAI in 2016. You can do many things with PyTorch, like building deep learning models, image classification, natural language processing, and more.

Here is the official page of [PyTorch](https://pytorch.org/) (<https://pytorch.org/>).

Tensorflow



Tensorflow is a machine learning library developed by Google in 2015. You can do many things with TensorFlow, like image classification, natural language processing, or generative modeling.

Here is the official page of [Tensorflow](https://www.tensorflow.org/) (<https://www.tensorflow.org/>).

Python Libraries for Data Visualization

Data visualization is an essential component of Data Science that helps Data Scientists to explore, analyze, and communicate data.

It is used to uncover trends, patterns, and relationships in data, which can be useful for building machine learning models or other purposes.

Let's learn how to do this in Python using several libraries, starting with Matplotlib.

Matplotlib



Matplotlib is a popular python data visualization library that enables users to create a range of visualizations in 2D.

It was developed by John D. Hunter in 2002.

Here is the official web page of [Matplotlib](https://matplotlib.org/) (<https://matplotlib.org/>).

seaborn



seaborn is a data visualization library for Python. It was created by Michael Waskom in 2014. seaborn is also built on top of Matplotlib, and often they worked together.

Here is the official web page of [seaborn](https://seaborn.pydata.org/) (<https://seaborn.pydata.org/>).

plotly



plotly is a data visualization library for Python and other programming languages. It was created by Alex Johnson, Chris Parmer, Jack Parmer, and others in 2012.

plotly is often used for its interactive visualizations, including line plots, scatter plots, bar plots, and more.

Here is the official web page of [plotly](https://plotly.com/) (<https://plotly.com/>).

pandas



pandas is a data manipulation and analysis library for Python, but it's also heavily used in data visualization.

It was created by Wes McKinney in 2008. pandas provide functions for reading and writing data, handling missing data, and performing data analysis tasks such as aggregation and reshaping.

Due to its functionalities, pandas are equally popular when it comes to manipulating data, performing mathematical operations, and visualizing data.

Here is the official web page of [Pandas](https://pandas.pydata.org/docs/user_guide/index.html) (https://pandas.pydata.org/docs/user_guide/index.html).

Python Libraries for Web Frameworks

A web framework is a set of libraries and tools that help developers to build and deploy web applications more easily. Web frameworks provide a structure for building web applications, and often include libraries for handling common tasks such as routing, authentication, and database access. Some examples of what you can do by using these frameworks are API, web Applications, and more. Let's start with django.

django



django is a web framework, developed in 2003 by Python programmers Adrian Holovaty and Simon Willison.

Here is the official web page of [django](https://www.djangoproject.com/) (<https://www.djangoproject.com/>).

Flask



Flask is a micro web framework. Like django, by using Flask you can also develop your own API or Web app. It was founded in 2004, by Armin Ronacher of Poccoo.

Here is the official web page of [Flask](https://flask.palletsprojects.com/en/2.2.x/) (<https://flask.palletsprojects.com/en/2.2.x/>).

FastAPI



FastAPI is a web framework, that allows users to create applications quickly, developed in 2018 by Sebastián Ramírez.

Here is the official web page of [FastAPI](https://fastapi.tiangolo.com/) (<https://fastapi.tiangolo.com/>).

Conclusion

As a data scientist, it is crucial to stay up-to-date with the latest tools and technologies. Of course, this list can be updated constantly, yet in 2023 these top 18 Python libraries will help you do that.

From data collection with Scrapy and BeautifulSoup to data collection to web frameworks with FastAPI and Flask, all Python libraries were introduced to you, which will help you to start with collecting and finish with deployment. By mastering these Python libraries, you will be well on your way to becoming a top-rated data scientist.