

[HOME](#) > [PROGRAMMING](#)

# 5 Data Science Libraries for Python Every Data Scientist Should Use

Want to model data and create visualisations in with Python? You'll need these data science libraries.

BY GAURAV SIYAL

PUBLISHED DEC 02, 2021



Python, as a language, has become the need of the hour. It does everything from building, managing, and automating websites to analyzing and wrangling data. Its truest functionalities come to the fore when data analysts, data engineers, and data scientists trust Python to do their data's bidding.

Python's name has become synonymous with data science, as it is used extensively to manage and draw insights from burgeoning data forms.

Its series of libraries is just the tip of the iceberg; many data scientists are beginning to use the available libraries at the click of a button.

## How Can Python's Libraries Help With Data Science?

Python is a versatile, multi-faceted programming language that continues to appease

people with its simple-to-use syntax, vast arrays of purpose-specific libraries, and an extensive list of analytical-driven functionalities.

Most Python libraries are handy for performing detailed analytics, visualizations, numerical computing, and even machine learning. Since data science is all about data analysis, and scientific computing, Python has found a new home for itself within its bosom.

Some best data science libraries include:

- Pandas
- NumPy
- Scikit-Learn
- Matplotlib
- Seaborn

Let's discuss each library to see what each option offers budding data scientists.



## **RELATED:**

### **Machine Learning Project Ideas For Beginners**

## **1. Pandas**

Python Data Analysis Library or Pandas is probably one of the most common libraries used within Python. Its flexibility, agility, and series of functions have made it one of the most loved libraries within Python.

Since data science starts with data wrangling, munging, and analysis, the Pandas library lends a supportive hand to make its functionalities even more helpful. The library is all about reading, manipulating, aggregating, and visualizing data and converting everything into an easy-to-understand format.

You can connect CSV, TSV, or even SQL databases and create a data frame with Pandas. A data frame is relatively symmetrical to a statistical software table or even an Excel spreadsheet.

## Pandas in a Nutshell

Here are some things which encompass Pandas functionalities in a nutshell:

- Index, manipulate, rename, sort, and merge data sources within data frame(s)
- You can add, update, or delete columns from a data frame easily
- Assign missing files, handle missing data or NaNs
- Plot your data frame information with histograms and box plots

In short, the Pandas library forms the base on which the very essence of Python's data science concepts rests.



### RELATED:

## Pandas Operations For Beginners

## 2. NumPy

As the name aptly encapsulates, NumPy is used widely as an array-processing library. Since it can manage multi-dimensional array objects, it's used as a container for multi-dimensional data evaluations.

NumPy libraries consist of a series of elements, each of which is of the same data type. A tuple of positive integers ideally separates these data types. The dimensions are known as **axes**, while the number of axes is known as **ranks**. An array in NumPy is categorized as **ndarray**.

If you have to perform various statistical computations or work on different math operations, NumPy is going to be your first choice. When you start working with arrays in Python, you will realize how well your calculations work, and the whole process is seamless, as the evaluation time downsizes considerably.

## What Can You Do With NumPy?

NumPy is every data scientist's friend, simply because of the following reasons:

- Perform basic array operations like add, subtract, slice, flatten, index, and reshape arrays
- Use arrays for advanced procedures, including stacking, splitting, and broadcasting
- Work with Linear Algebra and DateTime operations
- Exercise Python's statistical capabilities with NumPy's functions, all with a single library



### RELATED:

#### NumPy Operations For Beginners

# Machine Learning for Beginners

## 3. Scikit-Learn

Machine Learning is an integral part of a data scientist's life, especially since almost all forms of automation seem to derive their basics from the efficiencies of machine learning.

Scikit-Learn is effectively Python's native machine learning library, which offers data scientists' the following algorithms:

- SVMs
- Random forests
- K-means clustering
- Spectral clustering
- Mean shift, and
- Cross-validation

Effectively, SciPy, NumPy, and other related scientific packages within Python draw inferences from the likes of Scikit-Learn. If you are working with Python's nuances of supervised and unsupervised learning algorithms, you should turn to Scikit-Learn.

Delve into the world of supervised learning models, including Naive Bayes, or make do with

grouping unlabeled data with KMeans; the choice is yours.

## What Can You Do With Scikit-Learn?

SciKit-Learn is a very different ball game altogether, as its features are quite different from the rest of the libraries with Python.

Here's what you can do with this Scikit-Learn

- Classification
- Clustering
- Regression
- Dimensional reduction
- Model selection
- Pre-processing of data

Since the discussion has moved away from importing and manipulating data, it is essential to note that Scikit-Learn **models** data and does not **manipulate** it in any form. Inferences drawn from these algorithms form an important aspect of machine learning models.

## 4. Matplotlib

Visualizations can take your data places, help you create stories, 2D figures, and embed plots into applications, all with the Matplotlib library. Data visualization can be in different forms, ranging from histograms, scatter plots, bar plots, area plots, and even pie plots.

Each plotting option has its unique relevance, thereby taking the whole idea of data visualization up a notch.

Additionally, you can use the Matplotlib library to create the following forms of charts with your data:

- Pie charts
- Stem plots
- Contour plots
- Quiver plots
- Spectrograms

## 5. Seaborn

Seaborn is another data visualization library within Python. However, the pertinent question is, how does Seaborn differ from Matplotlib? Even though both packages are marketed as data visualization packages, the actual difference lies in the type of visualizations you can perform with these two libraries.

For starters, with Matplotlib, you can only create basic plots, including bars, lines, areas, scatter, etc. However, with Seaborn, the level of visualizations is taken up a notch, as you get to create a variety of visualizations with lesser complexity and fewer syntaxes.

In other words, you can work on your visualization skills and develop them basis your task requirements with Seaborn.

## How Does Seaborn Help You?

- Determine your relationships between various variables to establish a correlation
- Compute aggregate statistics with categorical variables
- Plot linear regression models to develop dependent variables and their relationships
- Plot multi-plot grids to derive high-level abstractions

### RELATED:

## How To Learn Python For Free

## Working Smartly With Python Libraries

Python's open-source nature and package-driven efficiencies go a long way in helping data scientists perform various functions with their data. From importing and analysis to visualizations and machine learning adaptations, there is a little bit of something for every type of programmer out there.



**7 Vital Commands to Get Started With Python for Beginners**

**READ NEXT >**



### RELATED TOPICS

PROGRAMMING

### ABOUT THE AUTHOR



**Gaurav Siyal**

(9 Articles Published)

Gaurav Siyal has two years of writing experience, writing for a series of digital marketing firms and software lifecycle documents.

[More From Gaurav Siyal](#) →

## SUBSCRIBE TO OUR NEWSLETTER

*Join our newsletter for tech tips, reviews, free ebooks, and exclusive deals!*

[CLICK HERE TO SUBSCRIBE](#)

## ON THE WIRE

---



## Yes, You Can Get Microsoft Word for Free: Here's How





## 25 Funny Things to Ask Google



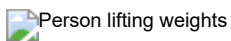
## Microsoft Has Developed an AI That Can Find and Fix Bugs in Code



## The 20 Most Popular Android Apps in the Google Play Store



## 6 Bad Remote Work Habits That Ruin Your Productivity, and How to Fix Them



## 6 Subscription Boxes to Support Your Fitness Journey

## READ NEXT

---



Persons face being removed in a photo



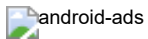
featured image for Linux server distributions



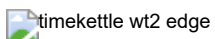
Lineup of BMWs by model year

## The 10 Best Linux Server Distributions


## 7 Websites to Help You Purchase Your Next Vehicle




## Got Pop-Up Ads on Android? How to Spot and Remove Them




## Timekettle WT2 Edge Translator Earbuds Review: It Will Get You to Where You're Going but Not Much Else

 Close-up of the Engineer Holding Laptop with CAD Component Model on Screen. In the Background Modern Factory Equipment.

## The 7 Best Laptops for 3D Modeling Engineering Software

 what is split tunneling

## What Is Split Tunneling and Should You Use It?

 check disk usage on Linux using duf

## How to Check Disk Usage on Linux With duf



## 9 DIY Projects for Your Old Hard Drive

[Write For Us](#)   [Home](#)   [Contact Us](#)   [Terms](#)   [Privacy](#)   [Copyright](#)   [About Us](#)   [Fact Checking Policy](#)  
[Corrections Policy](#)   [Ethics Policy](#)   [Ownership Policy](#)   [Partnership Disclaimer](#)

Copyright © 2021 [www.makeuseof.com](http://www.makeuseof.com)