Python Packages for   
Data Science



**Anish Mahapatra**

Contents

[Introduction 4](#_Toc114265932)

[Python Libraries – Overview 4](#_Toc114265933)

[Why Use Python Libraries for Data Science 4](#_Toc114265934)

[Ease of learning 4](#_Toc114265935)

[Less Code 4](#_Toc114265936)

[Prebuilt Libraries 4](#_Toc114265937)

[Platform Independent 4](#_Toc114265938)

[Massive Community Support 4](#_Toc114265939)

[List of Python Libraries for Data Science – 2022 (Note: Include Features, image, pros & cons, and applications for each library) 4](#_Toc114265940)

[Python Libraries for Math 4](#_Toc114265941)

[NumPy 4](#_Toc114265942)

[SciPy 4](#_Toc114265943)

[Theano 4](#_Toc114265944)

[Python Libraries for Data Exploration and Visualization 4](#_Toc114265945)

[Pandas 4](#_Toc114265946)

[Matplotlib 4](#_Toc114265947)

[Plotly 4](#_Toc114265948)

[Seaborn 5](#_Toc114265949)

[Ggplot 5](#_Toc114265950)

[Altair 5](#_Toc114265951)

[Autoviz 5](#_Toc114265952)

[Pydot 5](#_Toc114265953)

[Python Libraries for Machine Learning 5](#_Toc114265954)

[Keras 5](#_Toc114265955)

[SciKit-Learn 5](#_Toc114265956)

[PyTorch 5](#_Toc114265957)

[Pycaret 5](#_Toc114265958)

[TensorFlow 5](#_Toc114265959)

[Requests 5](#_Toc114265960)

[Python Libraries for Data Mining and Data Scrapping 5](#_Toc114265961)

[Scrapy 5](#_Toc114265962)

[BeautifulSoup 5](#_Toc114265963)

[SQLAlchemy 5](#_Toc114265964)

[Python Libraries For Natural Language Processing 5](#_Toc114265965)

[NLTK 5](#_Toc114265966)

[SpaCy 6](#_Toc114265967)

[Gensim 6](#_Toc114265968)

[Bonus Python Libraries! 6](#_Toc114265969)

[OpenCV 6](#_Toc114265970)

[Mahotas 6](#_Toc114265971)

[SimpleITK 6](#_Toc114265972)

[Pillow 6](#_Toc114265973)

[Selenium s 6](#_Toc114265974)

[PyTest 6](#_Toc114265975)

[PyUnit 6](#_Toc114265976)

[Conclusion 6](#_Toc114265977)

[Frequently Asked Questions (FAQs) 7](#_Toc114265978)

[Which library is most used in Python? 7](#_Toc114265979)

[Which software is best for Python for data science? 7](#_Toc114265980)

[Which Python library should I learn first? 7](#_Toc114265981)

[Which Python library is used for machine learning? 7](#_Toc114265982)

[How many libraries are in Python? 7](#_Toc114265983)

[Author 7](#_Toc114265984)

# Introduction

You might have seen different statistics on Python as being one of the best languages to learn. We are going to differ in our opinion here. Python is **the** **best** language to learn. The reason is that python is closer to a human-interpretable language than machine-level languages like C++ or Java. It is an intuitive language that can be used across a wide range of applications.

Now, what does the term packages or library even mean? Python is a plus and play language. The idea is that if you are looking to implement a simple or even a complex logic, it is likely that someone has already done it before. This logic is then put in a form that makes it reusable – this is known as a package or a library (the terms are used interchangably). So, why this blog?

With Python, ignorance is not bliss. Ignorance would mean you spend more time implementing a logic, even if it is readily available. Through this blog, we would like for you to get a brief overview on the range of packages that are available across a variety of applications. In the next section, let’s go over the fundamentals of python packages and how we can leverage them to maser Python!

# Python Libraries – Overview

The term “library” us used to collectively describe a reusable chunk of code. A python library consists of code that we can reuse while writing code for a given application. However, just to go a little bit in detail, a collection of modules is called a package and a collection of packages is called a module. Now, a fundamental question come to mind, when people are writing all this code, why would they build libraries for eveyone to use?

This is one of the reasons Python has grown to be one of the most widely used languages in the world. Besides it’s ease of use and wide applications, there is an extremely supportive community around python with millions of possible solutions for any issues you face. Python can be used for applications such as back-end, front-end, middle ware, data science, machine learning, artificial intelligence, deep learning and even something as simple as mathematics!

In the next section, let’s understand why we should leverage the libraries that are available in python.

## Why Use Python Libraries for Data Science

Let’s take a simple, fundamental example of a function and extrapolate the use of python libraries from there. Let’s say we are trying to add two numbers and we need to use this is ten places in out code.

**Method A**

a = 1, b = 6

We will mention c = a + b in ten places in our code.

**Method B**

We can define a function,

*Def calculation\_function(a, b):   
 c = a + b   
 return c*

Now, we will mention this function is ten places in our code, instead of the code directly like in Method A.

**Use-Case**

Now, let’ say that the logic is changed by the business and we need to make it multiplication (\*), instead of addition (+).

In Method A, you will have to go to ten places and change the code manually. This is error-prone and inefficient.

In Method B, you will have to change one character in the function. This will apply in 10 places in a consistent and efficient manner.

Now, let’s scale this up to thousands, maybe even millions of lines of code. Everytime you are trying to implement a new logic, would you rather re-write so many lines of code that is error prone or would you rather use near-perfect, well-documented, versioned code that is compliant to world-class global coding standards? Unless you are doing something extraordinarly unique, the best route for 99% of people is to use packages in python.

### Ease of learning

Python code is easier than most languages to learn. A major factor is how well python is documented. It is completely free to download, install and work with. In under 30 minutes, it is possible to download Python and even write your first Python program. There are multiple

### Less Code

The more we use packages, the less time we need to write and maintain code. Great coders know how to use Python packages efficiently, so that more of their focus can be on the logic and thinking, rather than writing code. This is something important to know about geat programmers, thep spend more time thinking and optimizing, rather than just writing simple code.

### Prebuilt Libraries

If you can think it, it’s most likely built. The logic exists in the form of Python packages, we need to apply it to our context of what we are trying to achieve. These pre-builtlibraries can be used to make our lives much easier. Not just that, even for complex applications such as image recognition, deepl learning etc., years and years of research and effort by many peeople has been condensed into simple functions that we can use in a jiffy.

### Platform Independent

Whether you are a fan of one IDE (Integratd Development Environment) or the other, packages have no barriers and can be installed on any platform. This is one of the biggest advatages of using Python – you can use it almost any technology stack you have

### Massive Community Support

# List of Python Libraries for Data Science – 2022 (Note: Include Features, image, pros & cons, and applications for each library)

## Python Libraries for Math

### NumPy

* Features
* Pros
* Cons
* Applications

### SciPy

* Features
* Pros
* Cons
* Applications

### Theano

* Features
* Pros
* Cons
* Applications

## Python Libraries for Data Exploration and Visualization

### Pandas

* Features
* Pros
* Cons
* Applications

### Matplotlib

* Features
* Pros
* Cons
* Applications

### Plotly

* Features
* Pros
* Cons
* Applications

### Seaborn

* Features
* Pros
* Cons
* Applications

### Ggplot

* Features
* Pros
* Cons
* Applications

### Altair

* Features
* Pros
* Cons
* Applications

### Autoviz

* Features
* Pros
* Cons
* Applications

### Pydot

* Features
* Pros
* Cons
* Applications

## Python Libraries for Machine Learning

### Keras

* Features
* Pros
* Cons
* Applications

### SciKit-Learn

* Features
* Pros
* Cons
* Applications

### PyTorch

* Features
* Pros
* Cons
* Applications

### Pycaret

* Features
* Pros
* Cons
* Applications

### TensorFlow

* Features
* Pros
* Cons
* Applications

### Requests

* Features
* Pros
* Cons
* Applications

## Python Libraries for Data Mining and Data Scrapping

### Scrapy

* Features
* Pros
* Cons
* Applications

### BeautifulSoup

* Features
* Pros
* Cons
* Applications

### SQLAlchemy

* Features
* Pros
* Cons
* Applications

## Python Libraries For Natural Language Processing

### NLTK

* Features
* Pros
* Cons
* Applications

### SpaCy

* Features
* Pros
* Cons
* Applications

### Gensim

* Features
* Pros
* Cons
* Applications

## Bonus Python Libraries!

### OpenCV

* Features
* Pros
* Cons
* Applications

### Mahotas

* Features
* Pros
* Cons
* Applications

### SimpleITK

* Features
* Pros
* Cons
* Applications

### Pillow

* Features
* Pros
* Cons
* Applications

### Selenium

* Features
* Pros
* Cons
* Applications

### PyTest

* Features
* Pros
* Cons
* Applications

### PyUnit

* Features
* Pros
* Cons
* Applications

# Conclusion

# Frequently Asked Questions (FAQs)

## Which library is most used in Python?

## Which software is best for Python for data science?

## Which Python library should I learn first?

## Which Python library is used for machine learning?

## How many libraries are in Python?

# Author

**Anish Mahapatra**   
*Senior Data Scientist*   
LinkedIn: [https://www.linkedin.com/in/anishmahapatra](about:blank) ([LinkedIn](about:blank))   
Email: anishmahapatra01@gmail.com