

SNYK REPORT

Python security insights

September 2021

Table of contents

tl;dr	3
Introduction	4
The security footprint of a typical Python project	5
The most common security issues in Python projects	7
Dealing with vulnerabilities in Python containers	9
Vulnerability spotlight: ipaddress	10
The most commonly downloaded Python packages	11
The most common vulnerable packages in Python projects	13
Package spotlight: urllib3	14
Fixing vulnerabilities	15
Keep your code secure	16
Tackling known vulnerabilities	17
Picking the right container image	18
Conclusion	19



tl;dr*

PyPI stats

- 316,360 packages
- 2,739,363 version releases
- 4,595,611 package files
- 8.3TB total size of packages on PyPI

Python known vulnerabilities

- On average, 60 new Python vulnerabilities are added to Snyk's vulnerability database on a monthly basis
- 5% of Python vulnerabilities are critical severity issues, 27% are high severity, 56% are medium severity and 12% are low severity
- 87% of vulnerabilities in Python packages have available remediation

Typical types of Python issues:

- (1) XSS
 (2) TLS Certificate verification disabled
 (3) Path traversal
 (4) Hard coded secret (non crypto)
- (5) SQL Injection

9% of XSS, Path Traversal, and SQL Injections were interfile. Especially, Django and Flask XSS and Path Traversal fall into this category.

Python projects

- An average project has 35 dependencies with an almost 50/50 split between direct and indirect dependencies
- An average vulnerable project consists of 33 known vulnerabilities

Python containers

- Over 1500 different Python image tags were pulled down from Docker in the last month alone
- python:latest (AKA python:3.9), has 431 preinstalled packages with 353 vulnerabilities.
- Most vulnerabilities can quickly be "fixed" by using slimmer images:
 - python:slim (Debian-based) has only 94 packages and 78 vulnerabilities
 - python:alpine has just 37 packages and 0 vulnerabilities

Security takeaways

Bad news

- 47% of Python projects contain known vulnerabilities
- 33% of all known Python vulnerabilities are high and critical severity issues
- In over 60% of Python projects, code related elements of OWASP Top 10 2021 list of issues can be found (e.g. Command Injection, XSS).

Good news

- You can eliminate 87% of known vulnerabilities by upgrading the vulnerable package
- Most container vulnerabilities can be fixed using slimmer images
- The most popular Python packages are healthy, receiving an average Snyk Advisor health score of 81%

Data taken from the Snyk Intel Vulnerability Database, hundreds of thousands of Python projects monitored by Snyk, the same number of projects used as a training set for Snyk Code, and Snyk Advisor.

snyk 🔁

Introduction

With a huge and ever-growing ecosystem of libraries, frameworks, and tools used by a community of over 10 million developers worldwide, the stage is set for Python to overtake contenders to become the most popular programming language.

Being a high-level, versatile, object-oriented language, Python is easy to learn while also being useful and powerful. This makes it an ideal choice by programmers of all backgrounds for a variety of projects: data analysis, web development, internet of things development (IoT), machine learning, DevOps, scripting, and plenty of other computing uses.

Python application development has changed over the years. Today, a typical Python repository will consist of much more than just the Python code written by the developer, but will also pull in more open source packages. It will likely include container images as well as configuration files used for provisioning the infrastructure required to run them. A lot of what used to be the responsibility of IT is now code built and managed by developers. From a security perspective, this means one thing - the attack surface is much wider than before.

Scanning a Python project will shed light on all the different nuts and bolts of your application - the vulnerabilities they introduce and (in the best case) the remediation path. But it's likely you will still have questions. How do your Python projects compare to others? How are other Python developers using open source packages? Are there any security trends in the Python ecosystem as a whole? This report aims to cover these questions, and includes the following:

- The security footprint of a typical Python project
- The most common vulnerabilities seen in Python applications
- · The most commonly used Python packages as well as the common vulnerable packages



The security footprint of a typical Python project

Snyk performs millions of monthly scans of hundreds of thousands of Python projects. We now have the ability to describe what an average Python project looks like and can give you an idea of what you might find when you scan one of your Python projects.

Known vulnerabilities

Python projects monitored by Snyk contain 1.3M distinct direct dependencies and 1.4M distinct transitive dependencies. An average Python project has around 35 dependencies. Out of these, 17 are direct dependencies and 18 are indirect dependencies.

In 47% of these projects, dependencies are introducing vulnerabilities. An average vulnerable project consists of 33 known vulnerabilities, out of which 10% are critical severity vulnerabilities, 26% are high severity, 26% medium severity, and 38% low severity

Container issues

Running Python applications in containers is a popular option as it eases the management of Python versions and dependencies. Docker provides official Python images going back as far as Python 2.7. There are also images available in popular Python data science packages like pandas, NumPy, and Jupyter.





Snyk's customers scan 10s of thousands of new containers each month and Python is the 3rd most popular languagebased container in use, which aligns with the overall popularity of Python. Images built directly from Docker's official Python image family are the most popular: 93% of the readily identifiable images we scan are built on Docker's Official images. Unfortunately, while the use of slimmer images is a well-known best practice, and Docker provides many slim options, relatively heavy images like python:3.7 are still the most common Python base images detected by Snyk. Python:3.7 has over 400 packages in it, resulting in over 300 vulnerabilities detected. Multistage builds are a great way to allow for the bigger images in development, while ensuring the final production image is as light and secure as possible.

Security weaknesses

Snyk Code support for Python is relatively new but Snyk is already performing static code analysis for thousands of Python projects. Roughly speaking, 20% of the security weaknesses identified by Snyk Code are related to Python projects. Out of these, 25% are high severity issues, 60% are medium severity issues and 15% are low severity issues. Overall, issues show up in over 60% of the Python projects scanned by Snyk Code.

Relative popularity of language-specific container images

Python versions in

use based on Snyk

Container scans



Based on the use of Docker's official images, as detected by Snyk Container. Python is the fourth most popular language-based container and Python 3.7 is the most popular version amongst Snyk's users.





The most common security issues in Python projects

Let's start our deeper dive into the state of security for Python projects with a look at the most common security issues found in Python projects.

The findings you see here are based on the training set used for Snyk Code -Snyk's Static Application Security Testing (SAST) solution, which numbered over 120k public Python projects on GitHub. A comparison to a recent study comparing vulnerabilities types (CWE) found in PyPI, confirms the pattern.

Unsurprisingly, you can see OWASP's Top 10 reflected in the results. There are two issues specific to Python, though:

- Unicode issues: Like most programming languages older than 10 years, Python has a history of issues related to handling encoding in strings.
- Closing API calls: When interacting with external resources such as file or network streams, calling the close function signals the system to flush the content as well as to free any handles. Developers in other languages such as C++ or Java - are a bit more disciplined in managing these resources.

Pro Tip: Snyk's Python cheatsheet provides a list of typical issue types with concrete advice.



Top 5 Python Issues Reported by Snyk Code



Туре:	Rel. within all issues	Explanation	CWE
XSS (Cross Site Scripting)	20%	Attacker injects client-side scripts into website.	CWE-79
TLS Certificate Verification Disabled	16%	Certificate verification is disabled by disabling verification calling certain functions. Possible man-in-the-middle attacks.	CWE-295
Path Traversal	16%	User supplied file names or paths are used and attackers could traverse the filesystem.	CWE-23
Hard Coded Secret (non-crypto)	10%	Hard coded values like account names, passwords, paths or file names should be kept a secret.	CWE-798 CWE-259
SQL Injection	9%	User supplied string is used to construct SQL queries and could be used for SQL injection attacks	CWE-89
Handled Unicode	7%	Calling functions that need encoded parameters or provides encoded results without proper formatting it	CWE-176
Missing closing call to API	7%	Access to external or buffered resource without proper closing may lead to data loss (e.g. when storing a resource as global)	CWE-400
Binary Write	7%	When opening a file to write, not using binary mode may result in issues if you are using Python 3 (or on Windows).	CWE-116
Wait can Deadlock	4%	When working with child processes, wait can lead to deadlocks on long running processes.	CWE-1322
Command Injection	3%	Unsanitized external data is used to build a command line call. Attackers might inject commands to be executed.	CWE-78



Dealing with vulnerabilities in Python containers

Running Python apps in containers is quite common, because it's an elegant way of dealing with Python's dependencies and virtual environments. In fact, the open source project that would go on to become Docker was first announced and demonstrated on stage by Solomon Hykes at PyCon 2013.

While it's pretty easy to get a Python app running in a container, as with so much in life the "easy way" isn't always the best way. Containers come with pre-installed Linux packages, which may or may not be important to making your app run but will impact your vulnerability reports.

Let's take a quick look at some Python base image options to see how they vary in size and vulnerabilities:



All of these images are very popular, ranking in the top tags pulled from Docker's Official Images for Python. But "fat" images like :3.9, :3.8, and :3.7 are the most common, despite all the best practices stating you should use a slimmer base image to reduce attack surface. You can get both worlds in containers through the use of multi-stage builds, starting with the bigger images to simplify building and testing your code and then moving the required production packages to a slim image in the final stage.

The key takeaway here is that smaller usually is better, when it comes to security. But diligence is still required. In addition, as fixes for vulnerabilities are published by the Linux maintainers, Docker updates their images. For this reason, even if your code doesn't change, it is worthwhile to have a process that rebuilds, tests, and redeploys your container images.

For comparison purposes, we included Google's distroless image as well. Distroless images remove the Linux shell, package manager, and other components resulting in a very small image. But Docker's python:3.9-alpine image, while slightly larger, actually has fewer vulnerabilities.

snyk 🔁

Vulnerability spotlight: ipaddress

In March this year, security researchers Sick Codes, Victor Viale, Kelly Kaoudis, John Jackson, and Nick Sahler uncovered a critical IP validation vulnerability in the netmask package. The vulnerability existed in both npm and Perl versions of netmask but was also identified later as impacting the ipaddress stdlib package in Python.

ipaddress enables developers to easily create IP addresses, networks, and interfaces, and to parse/normalize IP addresses inputted in different formats.

The uncovered vulnerability in ipaddress is related to the way the package parses IP addresses, improperly validating octal strings, and rendering IPv4 addresses that contain certain octal strings as integers. This opens up those Python applications using ipaddress to Server-Side Request Forgery (SSRF), Remote File Inclusion (RFI), and Local File Inclusion (LFI) attacks.

The vulnerability affects Python version 3.8.0 through 3.10. As reported in Snyk Advisor, ipaddress continues to be downloaded millions of times a week.

Package data:

Vulnerability data:

- Name: ipaddress
- Last release: Oct 18, 2019
- Popularity: 1.4M weekly downloads
- Security: severity vulnerabilities affecting all versions - 1 high severity, 2 medium severity
- Community: 6 contributors
- Direct inclusion rate*: 19%

- CVE: CVE-2021-29921
- Severity level: High
- CVSS score: 7.4
- Disclosure date: April 30, 2021
- Fix version: no fixed version availab
- * Calculated based on the number of projects which include this package as a direct dependency.

Timeline

ΥS

DA

767

- 2019-03-30 vulnerability introduced
 2021-03-29 vulnerability discovered
 2021-03-29 CVE requested
 2021-04-30 CVE Assigned CVE-2021-29921
 2021-05-01 PoC exploit published
 2021-05-02 commit 60ce8f0 (branch 3.10)
 2021-05-02 commit 5374fbc (branch 3.9)
 2021-05-03 Python 3.9.5 released
 2021-05-06 CVE added to NVD
- It's worth noting that the vulnerability was actually first introduced in 2019, two full years before it was discovered. During this time, attackers could have exploited this vulnerability as a 0-day without this becoming public knowledge!

snyk

The most commonly downloaded Python packages

What open source packages are Python developers using? How healthy are these projects? Are these packages secure? How are they included in projects?

To answer these questions we looked at two key datasets - the 1000 most downloaded packages from PyPI and the 1000 most common packages used in the Python projects monitored by Snyk.

Python packages in general are healthy!

To examine PyPI package downloads, we used Snyk Advisor - a free, online, research tool that helps you decide which open source packages or container base images to use to build your Python project. Containing all projects from PyPI, Snyk Advisor calculates a health score based on packages' popularity, security, maintenance, and community strength. For the 1000 most downloaded Python packages from PyPI, Snyk Advisor suggests an average high health score of 81%!

Usage patterns: Direct vs. Indirect

Taking a closer look at the 1000 most common packages used in projects monitored by Snyk, we found some interesting findings related to usage patterns.

The average direct inclusion rate for these packages is 63% implying that Python packages are mostly included in projects as a direct dependency.

Drilling a bit deeper into inclusion rates, popular utility packages, such as boto3 (91%) or markupsafe (96%), are more likely to be pulled into a project as a direct dependency. Other popular packages, such as urllib3 (31%), six (32%) and certifi (29%) have a much lower direct inclusion rate and are mostly introduced as an indirect dependency.



Package	# of downloads*	Health score	Last release date**	Age	License	Contributors	Dependencies
urllib3	158,893,228	98.50%	Jun 25, 2021	12 years	MIT	250	0
boto3	134,136,171	95.50%	Jul 28, 2021	7 years	Apache-2.0	110	6
six	128,572,888	86.93%	May 5, 2021	11 years	MIT	60	0
botocore	128,304,032	97.00%	Jul 27, 2021	9 years	Apache-2.0	130	4
requests	116,360,262	97.00%	Jul 13, 2021	10 years	Apache-2.0	410	7
certifi	113,424,297	85.70%	May 30, 2021	10 years	MPL-2.0	30	0
setuptools	113,270,733	95.64%	Jul 19, 2021	15 years	MIT	380	0
idna	112,275,170	83.60%	May 29, 2021	8 years	BSD-3-Clause	20	0
chardet	107,130,519	88.43%	Dec 10, 2020	15 years	LGPL-2.1	40	0
python-dateutil	100,233,788	94.14%	Jul 14, 2021	13 years	Apache-2.0 OR BSD-2-Clause	110	1

* Average monthly downloads for the period of March 21-June 21.

** At the time of writing.



The most common vulnerable packages in Python projects

We've taken a look at the packages that are downloaded more frequently. Of course, not all packages are vulnerable. Let's now look at the top ten vulnerable Python packages that are currently impacting Snyk's users the most, because they most frequently appear in Snyk project scans.

Included in the table are the minimum version upgrades you need to make to move to a vulnerability-free version.

As seen in the data, some particularly popular packages are vulnerable. urllib3 stands out here (we take a closer look at this package in the next section) but so do others.

pillow - the popular (8.3M weekly downloads) imaging library - is riddled with vulnerabilities. While the latest version is safe to use, previous and widely-used versions include 1 critical severity vulnerability and 16 high severity vulnerabilities.

The same goes for Dj ango - a popular web framework for Python. Versions 3.1.13 and above are safe to use but older versions are not as secure.

ipaddress and cryptography - downloaded together over 12.7M times a week, currently have no remediation path.

Package	Use	Vulnerabilities	Minimum known vuln free version	Weekly downloads
urllib3	HTTP client	Critical - 0 High - 3 Medium - 6 Low - 1	1.26.6 (latest)	38M
pillow	Imaging library	Critical - 1 High - 18 Medium - 7 Low -	8.3.1 (latest)	8.3M
PyYAML	YAML parser and emitter	Critical - 4	5.4	21M
ipaddress	IPv4/IPv6 manipulation library	Critical - 0 High - 1 Medium - 2 Low - 0	-	1.5M
cryptography	Cryptographic recipes and primitives	Critical - 0 High - 4 Medium - 3 Low - 0		11.2M
django	High-level Python Web framework	Critical - 0 High - 8 Medium - 15 Low - 5	3.1.13	1.6M
jinja2	A fast and expressive template engine	Critical - 0 High - 1 Medium - 4 Low - 5	2.11.3	17M
pygments	A syntax highlighting package	Critical -1 High - 2 Medium - 0 Low - 0	2.7.4	6M
requests	HTTP client	Critical -1 High - 0 Medium - 5 Low - 0	2.20.0	37M
rsa	RSA implementation	Critical -0 High - 2 Medium - 3 Low - 0	4.7	16M

Data as of Aug/2021



Package spotlight: urllib3

#1

Most downloaded package on PyPl

100%

Snyk Advisor Health Score



Most used package in projects monitored by Snyk

snyk

urllib3 is an extremely popular HTTP client for Python, supporting a lot of greatly-needed functionality missing in Python libraries such as thread safety, connection pooling, client-side SSL/TLS verification, file uploads with multipart encoding, and plenty more.

With millions of downloads a week, urllib3 is the most downloaded Python package on PyPI. It is also the 3rd most used package in the projects monitored by Snyk. From a health perspective, urllib receives top scores with an impressive 100/100 Snyk Advisor health score.

- Popularity urllib3 is downloaded over 42 million times a week, marking it as a key
 project in the Python ecosystem
- Maintenance with 146 open issues, 13 open PRs, a recent release under a month ago and a commit from just a few days ago, urllib3 shows strong vitality
- Security the latest version of urllib3, version 1.26.6, is free of known security vulnerabilities and safe to use
- Community with 250 developers collaborating on the project, urllib is clearly receiving strong external contributions

The devil, as always, is in the details. When looking at usage patterns for urllib3, two worrying data points immediately stand out:

Older vulnerable versions - while the latest version of the package is safe to use, previous versions include security vulnerabilities, including high and medium severity issues. Version 1.24.3, downloaded over 2 million times a week, includes a high severity CRLF injection vulnerability. Version 1.26.3, downloaded over 1.5 million times a week contains an Improper Certificate Validation vulnerability.

Indirect inclusion - in the projects monitored by Snyk, urllib3 is pulled in as a direct dependency in only 31% of the cases. This means that vulnerabilities in older versions of the package are more difficult to identify and fix.

Fixing vulnerabilities

Choose your dependencies wisely

Adding dependencies using PyPI (or other Python package managers for that matter) is extremely easy. But PyPI does not scan packages before listing them nor does it guarantee the authenticity of a package. Before you add a dependency into your project, some due diligence is a good best practice. Using Snyk Advisor for example, you will be able to check for known vulnerabilities or license issues as well as gauge how strong the community behind the package is.

snyk.io/advisor/

Pro tip: Snyk Advisor provides you with the command line to install the package. By copy and pasting it, you prevent typosquatting.





Keep your code secure

A modern application consists of only 10-20% proprietary code - the rest of the code base is actually made up of open source code brought in by open source dependencies. But it is this 10-20% that makes the application unique. Containing the intellectual property of the organization, this portion of code is not maintained by a community of open source developers -- it is all yours to take care of!

Python provides some build-in mechanisms that can help you develop stable apps - support for virtual environments is one example.

Reviewing the top 10 security issues found in application code listed above, leads to six general points of advice:

For more information on best practices for keeping your Python code secure, check out Snyk's Python cheatsheet

- 1. Use modern static code analysis: Linters like Pylint and scanners like Bandit are a good start. But nasty problems are interfile (aka the issue happens as the application execution flows between various source files). Finding these kinds of issues manually is near impossible.
- 2. Sanitization of data: Try to sanitize inflowing data from any external sources (including databases) at the entry point in the application.
- 3. ORM: Use modern Object Relational Mapping (ORM) tools to abstract the database interactions and prevent SQL injection opportunities. If you are using packages like Django or Flask, use libraries like Django ORM or SQLAlchemy which are well-vetted.
- 4. Unicode: If possible, standardize all strings to a certain unicode encoding we recommend UTF-8. Be careful when converting unicode strings into ASCII.
- 5. Close APIs: Make sure to close your network connections (e.g. external read and writes). This ensures that data written in their buffers are actually stored, the state is stored correctly and it frees up handles in your system.
- 6. Guard your secrets: This is not Python specific but we see secrets like usernames, passwords, API access tokens, but also file paths or file names leak into the source code. It is a good practice to keep them in separate files, or better yet, secret stores like HashiCorp Vault, AWS Key Management Service, etc.

snyk

Tackling known vulnerabilities

Once a vulnerability is found, project maintainers will typically include a fix (if possible) in a future version, so keeping your dependencies up to date is generally a good way to stay on top of known security vulnerabilities. In some cases, though, upgrading a dependency is challenging because of the way dependencies interact with each other and your code.

Fixing vulnerabilities in direct dependencies is usually straightforward. Upgrade the dependency to the minimum version that includes the fix. Fixing vulnerabilities in indirect dependencies requires two things: a fixed version of the indirect dependency and a version of the direct dependency that utilizes that fixed version.

If these two conditions are met, upgrading the associated direct dependency to a version that utilizes the fixed version of the indirect dependency will fix the issue. If no fix is available at the level of the direct dependency, you can upgrade the indirect dependency to resolve the issue. Note, however, that this has the potential to break your code due to compatibility issues between the dependencies.

Pro tip: The Snyk CLI can be used to scan your Python dependencies locally or as part of a CI/CD pipeline. A new fix command (currently in beta) will also fix the vulnerability by updating your requirement.txt file.

goof@1.0.1 adm-zip@0.4.7 H body-parser@1.9.0 ♥ H gs@2.2.4 cfenv@1.2.2 underscore@1.9.1 dustjs-linkedin@2.5.0 9 H 𝘵 M ejs@1.0.0 1 H



Picking the right container image

At first glance, one might see a container vulnerability report and think that dealing with 375 vulnerabilities is an overwhelming task or that containers are far too risky. Or one might wonder why Docker doesn't fix these vulnerabilities themselves. Fundamentally, this thinking usually stems from the belief that every vulnerability needs a patch and diligent sysadmins work to stay on top of these 375 issues.

Instead, the theory behind containers is that they should only contain what your app needs to run, and nothing else. That means "fixing" is not strictly limited to installing a patch; rather, removing unneeded packages is also a fix. Which brings us back to the best practice of using slimmer images. But there are literally thousands of image tags within the Docker Official Image Python repository alone -if you're using python:latest, how do you know which alternatives might be better, short of testing them one-by-one?

To pick a good container base image:

- 1. Use Snyk Advisor to look up good starting points. Tags with the word "slim" and Alpine images will be the smallest, reducing the likelihood of vulnerabilities.
- 2. Use Snyk Container for scanning your Dockerfile from your git repo to get immediate recommendations and fix PRs even before you build and scan any images.
- 3. Rebuild, scan and redeploy often, even if there are no code changes, to get the latest updates from Docker and the latest versions of your tools. Snyk Container will monitor for new vulnerabilities, out-of-date images, and incorporate the running configuration to prioritize issues to fix.
- Multistage builds not only help produce small, secure images for production, but are a great way to provide a single source of image instructions for every stage of development and release.



Conclusion

As applications get more complex, so does the task of securing them. Malicious actors have a wide variety of attack vectors to use when attacking a Python app - whether via known vulnerabilities introduced via direct and indirect dependencies, security issues in the app's proprietary code, or container vulnerabilities. Just under half of all Python projects monitored by Snyk are indeed vulnerable.

But not all is bad. In fact, the data here suggests the Python ecosystem is well equipped to successfully tackle security. An impressive 87% of known vulnerabilities have a fix. Not only are the most commonly used packages generally healthy, but 63% of these packages are also included via a direct dependency and so applying a fix is relatively simple. Using the python:slim image will quickly remove hundreds of container vulnerabilities.

At Snyk, our goal is to help development and security teams develop fast while staying secure. A big part of that is providing developer-first tools that help these teams find, prioritize and fix issues quickly and efficiently. Snyk Advisor helps developers research packages on PyPi. And we facilitate the reporting of new vulnerabilities. Currently the Python ecosystem does not have a centralized place to report vulnerabilities in open source libraries. Snyk is a CVE Numbering Authority (CNA), which means we are able to assign a new vulnerability a CVE number and add the vulnerability to relevant databases. As a CNA, Snyk can help you responsibly report vulnerabilities.

Secure your Python projects with Snyk

Scan your Python code for quality and security issues, and get fix advice right in your IDE. Get started with Snyk for free.

Get started for free

