



# Introduction to Python

This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Agenda

1. Overview of Python
2. Why Python for Data Science
3. Exploring Python tools for data science
4. Case study

This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

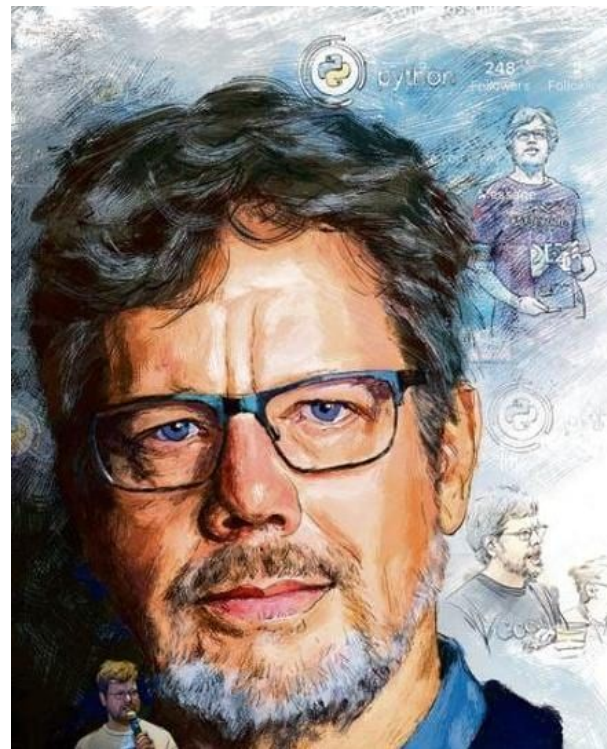
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Python: Overview

- Invented in the Netherlands, early 90s by Guido van Rossum
- Named after Monty Python
- Open-sourced from the beginning
- Scalable, object oriented and functional
- Used by Google from the beginning
- Highly popular and already the core of Data Science and AI solutions

**“Python is an experiment in how much freedom programmers need. Too much freedom and nobody can read another's code; too little and expressiveness is endangered.”**

**- Guido van Rossum**



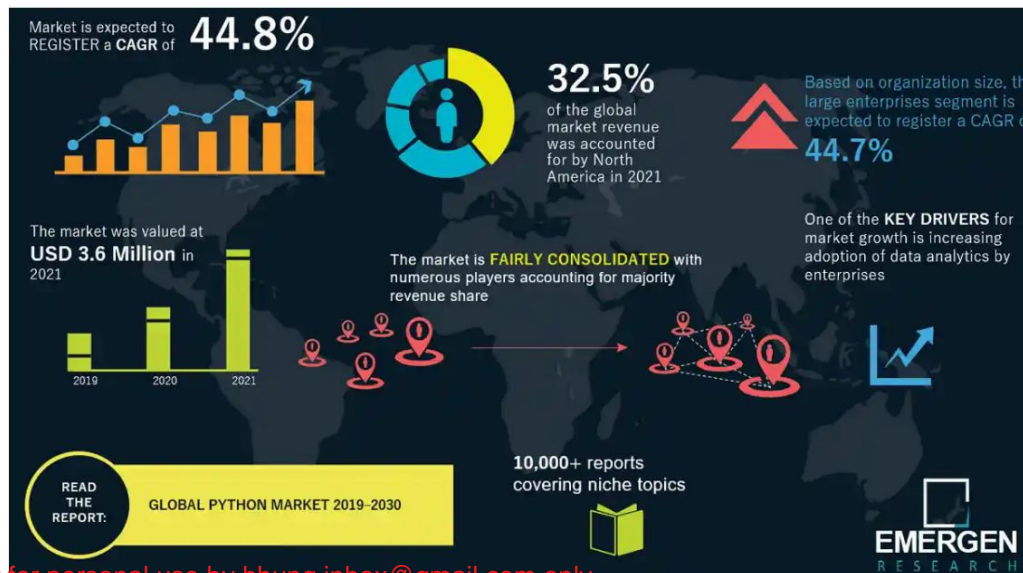
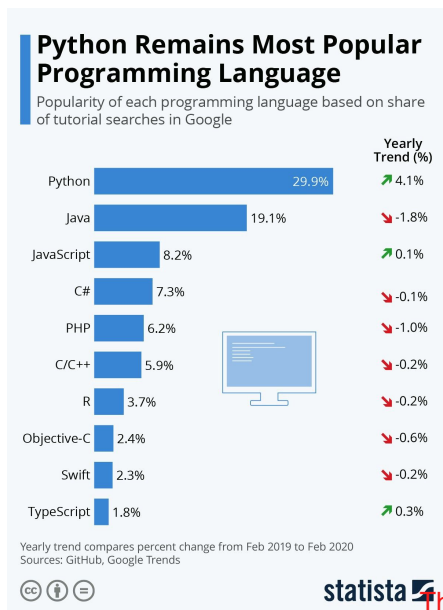
This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

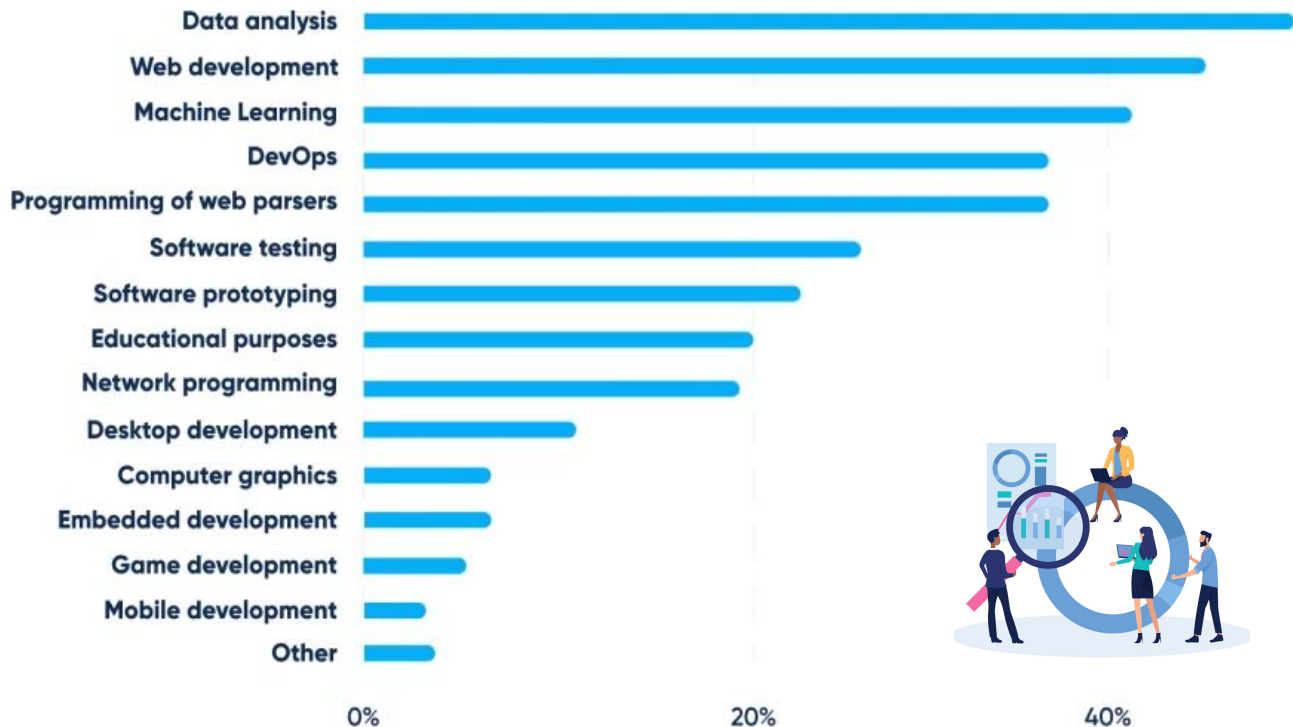
# Market Synopsis

- The global Python market size is expected to reach **USD 100.6 million by 2030** and register a revenue **CAGR of 44.8%** during the forecast period.
- The rising demand for Python in end-use applications such as web development, the rapid adoption of industry 4.0, the increasing use of data analytics and powerful computing are some of the factors projected to drive this revenue growth.





# Where is the industry using Python?



This file is meant for personal use by hhung.inbox@gmail.com only.  
Sharing or publishing the contents in part or full is liable for legal action.



# Why Python for Data Science?

This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Python is easy to learn

Python syntax is very similar to English, which makes it relatively easy to read and understand. For instance, let's try to print out "Hello, World" in 4 different programming languages.

## C

```
#include
<stdio.h>
int main()
{
    printf("\n Hello,
World");
    return 0;
}
```

## C++

```
int main()
{
    // prints hello
world
    cout <<
    "Hello, World";

    return 0;
}
```

## Java

```
class HelloWorld
{
    public static void
main(String args[])
    {

        System.out.println("Hello,
World");
    }
}
```

## Python

```
print("Hello, World")
```



**Easy and  
powerful**

# How does that help Data Scientists?

Python's primary focus is on **keeping things simple and readable**. That is why it's an **ideal “entry-point” programming language** for Data Scientists, who often come from a mixture of technical and non-technical backgrounds. Rather than paying too much attention to learning the syntax, it allows the focus to move to quickly **writing code to solve the problem statement**.

The learning curve of Python is hence relatively easy. Python is more concise and has fewer lines of code in comparison to other languages such as C, C++ and Java, giving you the chance to **prototype and iterate fast**.



This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

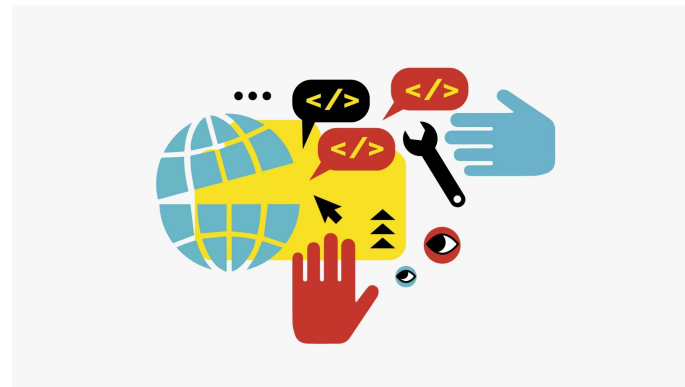
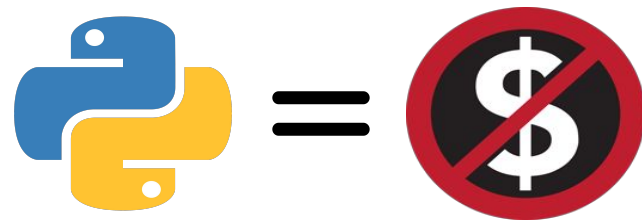


# Python is Free and Open-Source

Python and its source code are freely available. You can modify, improve / extend what is essentially open-source software, with no limit to your creativity.

Python has been developed under an OSI-approved open source license, making it freely usable and distributable even for commercial use

**As such, Python is ideal for anyone who's looking to take their first steps in Data Science without burning a hole in your pocket or worrying about licenses.**



This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Cross-Platform Language

Python is a cross-platform programming language. **You can execute almost any Python program on Windows, Mac, Linux hardware, and even on Android and iOS.** A Python program written on a Macintosh computer will run on a Linux system and vice versa.

So what would a Data Scientist use a cross-platform language?

- Code Reusability
- Faster Customization
- Uniformity in Design
- Cost-Effectiveness
- Easy Maintenance



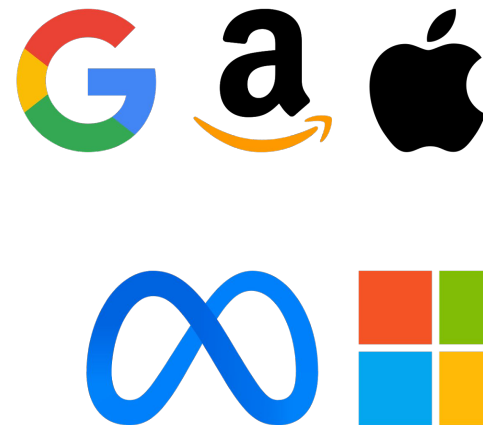
This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Python is Powerful

- Python is being used in real revenue generation by very successful companies:
  - Google makes use of Python in its web search system and has even employed Python's creator Guido van Rossum in the past
  - Intel & Cisco use Python extensively for hardware testing
  - YouTube's video sharing service is built on Python
  - Reddit, Pinterest and Instagram are largely written in Python
  - Python is also used heavily in academic research, particularly in Bioinformatics, Biology and Mathematics



# Convenience & Flexibility for Data Scientists

Python not only lets you create software but also enables you to deal with the analysis, computing of numeric and logical data, and web development.

Python's flexibility also allows Python code to make ML/AI scalability possible without requiring distributed systems expertise and lots of invasive code changes.

Python enables Data Scientist programmers to focus on the **tasks to be completed or action to be performed without worrying about how to accomplish the specific objective.**



This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Python's Community Support

Great software is supported by great people, and Python is no exception. **Python's user base is very enthusiastic and dedicated to spreading the use of the language far and wide.** The Python community can help support either the beginner or the expert, and is adding to the ever-increasing open-source knowledge base.

1. <https://www.python.org/community>
2. <https://discord.gg/python>
3. <https://pyslackers.com/web>



As a Data Scientist in the fast-evolving space of Machine Learning and AI, it is important to learn about the new features that are getting added to the language, and if possible, also make contributions to the community with your own findings.

This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Popular Python Packages for Data Science

Even with Python's fairly brief history, the community has already created a range of excellent libraries, packages, frameworks and development tools for Data Scientist programmers to choose from.

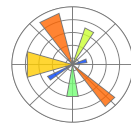
Data Scientists can avail of libraries such as NumPy, SciPy, Pandas, Matplotlib, Seaborn, Scikit-Learn and TensorFlow to simplify the development of statistical and Data Science applications. These tools have contributed hugely towards making Python more widely used than any other programming languages for Applied Data Science, and you will see much more of them over the course of this program.



NUMPY



PANDAS



MATPLOTLIB



SEABORN



SKLEARN



SCIPY



TENSORFLOW



KERAS

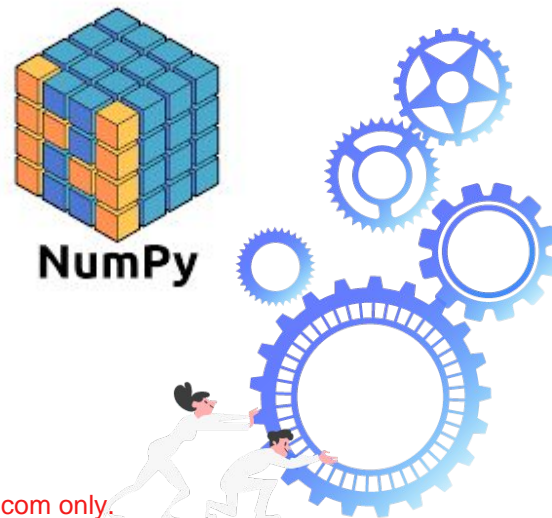


# Applications of a few of these tools

**1. TensorFlow:** TensorFlow is an all-purpose Machine Learning package developed by Google. It has proved highly useful especially in applications centered around Deep Learning and cutting-edge AI, such as Object Detection and Speech Recognition.



**2. NumPy:** NumPy is a general-purpose array-processing package that provides high-performance multi-dimensional objects called arrays, as well as tools to work with them. These NumPy arrays support several useful mathematical operations, and are the core data structure that modern Data Science in Python is based on.



**3. Scikit-Learn:** Scikit-Learn is a Machine Learning library in Python that provides almost all the algorithms you will need for basic Applied Data Science. Scikit-Learn is designed to be integrated to work with NumPy arrays, and it includes algorithms for clustering, regression, classification and other ML domains.

This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Now let's get our hands dirty with some code

This file is meant for personal use by [hhung.inbox@gmail.com](mailto:hhung.inbox@gmail.com) only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.