An overview of Python
A language for most of your applications

Alex de Siqueira, Data Science Outreach Lead
Ciera Martinez, Biodiversity and Environmental Sciences Lead
Maryam Vareth, I4H Co-Director, Health and Life Sciences Lead
Liliana Cardile, Industry Relations and Events Specialist
Marsha Fenner, Communications/Program Manager
Ann Cleaveland, Interim Executive Director
Alexandre (Alex) de Siqueira

- Graduated in Mathematics (2007)
- Master’s, PhD in Materials Science (2009–2015)
- Native language: Portuguese
Part 1. Why Python?
The history of Python

Guido van Rossum, the creator of Python. From Wikipedia
A function to collect all words in a document

**ABC:**

**HOW TO RETURN words document:**

- PUT {} IN collection
- FOR line IN document:
  - FOR word IN split line:
    - IF word not in collection:
      - INSERT word IN collection
- RETURN collection

**Python:**

```python
def words(document):
    collection = []
    for line in document:
        for word in line.split():
            if word not in collection:
                collection.append(word)
    return collection
```

M87 – the first image of a black hole


NumPy was one of the software used!

From NumPy

https://github.com/achael/eht-imaging
Open Source on Mars: Community powers NASA’s Ingenuity Helicopter

DEVELOPERS AROUND THE WORLD CONTRIBUTE TO HISTORIC FLIGHT.

KLINT FINLEY // APRIL 14, 2021

Today, nearly 12,000 developers will see a new badge on their GitHub profile celebrating their contributions to the specific versions of projects and libraries used by NASA to fly the Ingenuity Helicopter on Mars. Learn
Part 2. Python for Data Science
Part 3. Some tools to get you started
Love people, use tools; not the other way around!
Individual Edition

Your data science toolkit

With over 25 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for
Code editing. Redefined.
Free. Built on open source. Runs everywhere.

By using VS Code, you agree to its license and privacy statement.
Part 4. How to start creating your portfolio
Competition

Titanic - Machine Learning from Disaster
Start here! Predict survival on the Titanic and get familiar with ML basics
Getting Started · Ongoing

Predict Future Sales
Final project for "How to win a data science competition" Coursera course
Playground · a year to go

House Prices - Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting
Getting Started · Ongoing

At Kaggle
Getting started tutorials

- What kind of data does pandas handle?
- How do I read and write tabular data?
- How do I select a subset of a DataFrame?
- How do I create plots in pandas?
- How to create new columns derived from existing columns?
- How to calculate summary statistics?
- How to reshape the layout of tables?
- How to combine data from multiple tables?
- How to handle time series data with ease?
- How to manipulate textual data?

At Pandas

At scikit-learn

NumPy tutorials

This set of tutorials and educational materials is being developed in the numpy-tutorials repository, and is not a part of the NumPy source tree. The goal of this repository is to provide high-quality resources by the NumPy project, both for self-learning and for teaching classes with. If you're interested in adding your own content, check the Contributing section.

To open a live version of the content, click the launch Binder button above. To open each of the .ipynb files, right click and select "Open with -> Notebook". You can also launch individual tutorials on Binder by clicking on the rocket icon that appears in the upper-right corner of each tutorial. To download a local copy of the .ipynb files, you can either clone this repository or use the download icon in the upper-right corner of each tutorial.

At Numpy

Scipy Lecture Notes

One document to learn numerics, science, and data with Python

Tutorials on the scientific Python ecosystem: a quick introduction to central tools and techniques. The different chapters each correspond to a 1 to 2 hours course with increasing level of expertise, from beginner to expert.

Release: 2020.2

At Scipy Lecture Notes
Upcoming Events
Browse our upcoming events below, or see past events.

Want to host a conference?

At PyData

At Meetup
OPEN DATA NETWORK

Publish data and share. Find data and build. Answer questions.

Datasets
Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.

+ New Dataset  Your Work
# Portfolio

A portfolio is a compilation of materials that exemplifies your beliefs, skills, qualifications, education, training, and experiences. It provides insight into your personality and work ethic. A portfolio website is generally a static website featuring work samples and other details as explained above. It is closely related to Online Resume or C.V.

Here are 12,672 public repositories matching this topic...

Language: All  ▼  Sort: Best match  ▼

Hi, my name is John Doe

At GitHub
Going further

Python

- [Python website](#)
- [Visualizing the commit history of the Python scripting language project](#)
- [Guido van Rossum - Interviews](#)
  - [Interview with Guido van Rossum (July 1998)](#)
- [Python Success Stories](#)
- [Python Standard Library](#)
- [Python Package Index](#)
- [Astronomers Capture First Image of a Black Hole](#)
- [List of qualifying repositories for Mars 2020 Helicopter Contributor badge](#)
- [2020] Scipy Lecture Notes: basic NumPy, Matplotlib, Scipy, and more
Going further

Anaconda

- Anaconda Installation
  - Anaconda: How Do I Become a Data Scientist?

VSCode

- Setting up Visual Studio Code
- Python extension for Visual Studio Code
  - Python in Visual Studio Code
  - Getting Started with Python in VS Code
- Jupyter Extension for Visual Studio Code

Project Jupyter

- Project Jupyter Documentation
Going further

Python communities

- PyLadies
- PyData Meetups
  - PyData on YT

Python packages

- NumPy: https://numpy.org/
- SciPy: https://www.scipy.org/
- matplotlib: https://matplotlib.org/
- pandas: https://pandas.pydata.org/
- scikit-learn: https://scikit-learn.org/stable/
Going further

Open Data

- Open Data Network
- DATA.GOV
- Kaggle Datasets

Misc

- Nice flowers on slide 14 from PublicDomainPictures.net.
Thank you!