

How You Can Use Open Source Materials to Learn Python & Data Science

Kamila Stępniewska, EuroPython 2018

github.com/KStepniowska/EuroPython2018

CC-BY

What can you expect?

- Sociology
- Diversity: Geek Girls Carrots, Women Who Code, She's Coding
- New Business Manager @10Clouds

Please don't ask me about:

1. How can you become a data scientist in 3 weeks?
2. Which algorithm will solve an “A” or “B” problem?

I hope that you will learn more about...

RESOURCES - Python and Data Science

github.com/KStepniowska/EuroPython2018

Open Source - basics

Data Science Workflow

PROJECTS & COOPERATION & CONTRIBUTION

Shall we?



Fernando José Ignacio
Gárate Parra

<https://bit.ly/2A5MoOW>

CC BY-NC 2.0

Open Source

“Open data and content can be freely used, modified, and shared by anyone for any purpose”

<https://opendefinition.org/>

Educational Materials

As a User

Text, Pictures, Videos, Audio Records...



Attribution
CC BY



Attribution-ShareAlike
CC BY-SA



Attribution-NoDerivs
CC BY-ND



Attribution-NonCommercial
CC BY-NC



Attribution-NonCommercial-ShareAlike
CC BY-NC-SA



Attribution-NonCommercial-NoDerivs
CC BY-NC-ND

<https://creativecommons.org/licenses/>

Popular Licenses

The following OSI-approved licenses are popular, widely used, or have strong communities:

- Apache License 2.0
- BSD 3-Clause "New" or "Revised" license
- BSD 2-Clause "Simplified" or "FreeBSD" license
- GNU General Public License (GPL)
- GNU Library or "Lesser" General Public License (LGPL)
- MIT license
- Mozilla Public License 2.0
- Common Development and Distribution License
- Eclipse Public License

<https://opensource.org/licenses>

<https://www.gnu.org/graphics/license-logos.en.html>

Pick yours

As a Creator

General: choosealicense.com

Text: creativecommons.org/licenses/

Code: opensource.org/licenses

Python

Why Python?

COMMUNITY

Welcoming & Supportive

Global & Diverse

...

If there is a problem, there is a great chance that someone has written and shared the solution already.

Learning Experience

Find Your **Project** -> learn by building

Find Your People -> **Cooperate**

Find a way to **Contribute** -> help others

<https://bugs.python.org/>

For Beginners

PEP 8 python.org/dev/peps/pep-0008/

Yes:

```
spam(ham[1], {eggs: 2})
```

No:

```
spam(ham[1], {eggs: 2})
```

“PEP 8 — the Style Guide for Python Code

This stylized presentation of the well-established PEP 8 was created by Kenneth Reitz (for humans).”

*PEP=Python Enhancement Proposal

pep8.org/#fn1

For Beginners

PEP 20 - The Zen of Python

python.org/dev/peps/pep-0020/

```
Beautiful is better than ugly.  
Explicit is better than implicit.  
Simple is better than complex.  
Complex is better than complicated.  
Flat is better than nested.
```

...

Resources: For Beginners

Python, Programming, Open Source

Knowledge: python.org -> Beginner's Guide

for Non-Programmers <https://bit.ly/1lv5glG>

for Programmers <https://bit.ly/1UIBJMJ>

How to learn: Lynn Rooth “Sink or swim”

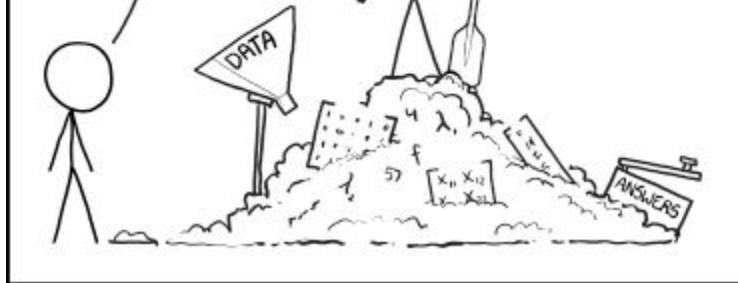
<http://www.roguelynn.com/words/The-New-Coder-A-path-to-Software-Engineering/>

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



xkcd

CC BY-NC 2.5

<https://xkcd.com/1838/>

Data Science

Use of Python in Data Science

Python Developer - Survey 2017 Results

(9,500 developers, 150 countries)

“What do you use Python for?”

(multiple answers)”

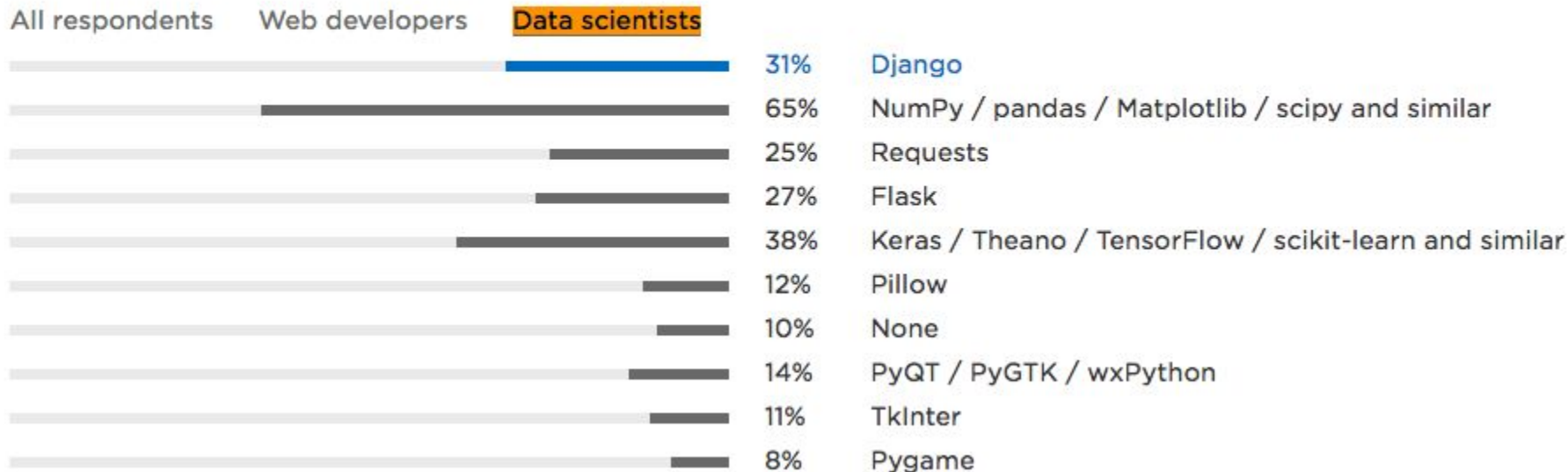
50% Data analysis, 31% Machine learning

Interesting fact

Most respondents underestimated the total number of developers involved in **data science**. It's generally perceived that web development is the major application for Python. While this was true a couple of years ago, today the number of Python **data scientists** is growing rapidly and is already on par with the number of web developers.

jetbrains.com/research/python-developers-survey-2017/

What frameworks do you use in addition to Python? (multiple answers)



jetbrains.com/research/python-developers-survey-2017/

Python & Data Science - what's more?

Jupyter Notebook jupyter.org/

PyCharm jetbrains.com/pycharm/

Spyder pythonhosted.org/spyder/

Use Python to build your tools to explore data

You need to know Python to be able to freely
build experiments.

“Data”

Gathering, cleaning and data preparation is crucial.

Typical issues:

- there is not enough data
- data is messy
- we actually don't know what is in the data set...

**Data preparation is
even 80% of a
Data Scientist Work**

Gill Press, Forbes <https://bit.ly/2OgNM4D>

“Science” - on the hunt for the right questions

- Understand what I want to achieve
- Define the problem that I want to solve
- Define what is the input and what I want to be an output

UNDERSTAND

- Looking for helpful algorithms
- Compare the chosen algorithms
- Choose the algorithm/s to be used
- Choose the evaluation metrics

SEARCH

- Choose parameters set for experiments
- Run experiments
- Analyse the results
- Define the conclusions and/or get back to previous points

EXPERIMENT

How to find the right algorithm?

The Internet...

- the resource - do you define the source as trusted? (eg. scikit-learn)
- number of stars, forks, when was the last commit? (GitHub)
- the code
 - is it aligned with the Python standards? (PEP 8)
 - check the particular functions
 - ...
- does it fit to the general architecture of a project?
- ask a friend

Anna Gut, Python Developer & Team Lead @10Clouds

Hacks - what was your steps & how did you get there

Step is a wrapper over the transformer and handles multiple aspects of the execution of the pipeline, such as saving intermediate results (if needed), checkpointing the model during training and more.

Transformer is purely computational, data scientist-defined piece that takes an input data and produces some output data. Typical Transformers are neural network, machine learning algorithms and pre- or post-processing routines.

github.com/neptune-ml/steppy

Resources

Data Science, Open Source

All... datasciencemasters.org/

Transformation from Math & Phys into Data Science:

p.migdal.pl/2016/03/15/data-science-intro-for-math-phys-background.html

Projects Cooperation Contribution

Projects

Find your project

- newcoder.io/tutorials/
- www.kaggle.com/
- devmesh.intel.com/

DRIVING
ME?

POSSIBLE?

Cooperation

Online:

- pyslackers.com (14,757 members)
- mail.python.org/mailman/listinfo/tutor
- <https://www.facebook.com/groups/python.programmers>

Offline:

- PyData, PyWaw
- PyLadies, Girl Geek, Geek Girls Carrots (Krakow)
- Django Carrots, Django Girls

Contribution

- Bag Tracker bugs.python.org
- Open Source Projects opensource.guide/how-to-contribute
- Answer questions at pyslackers.com
- Become a speaker/mentor pydata.org
- Organize Django Girls djangogirls.org/organize
- ...

Even More Resources...

Open Education

The Journal of Open Source Education

The Journal of Open Source Education (JOSE) is an **educator friendly** journal for publishing computational learning modules and educational software.

jose.theoj.org

Thank you!

github.com/KStepniowska/EuroPython2018

kamila.stepniowska@10clouds.com
[@kstepniowska](#)