

Detailed Course Schedule



1. Introduction to Python for Data Science

2. Big Data Processing with Apache Spark

3. Introduction to Machine Learning

4. Concluding Remarks

Introduction to Python for Data Science

Prerequisites

Assumes no
prior
knowledge.

Course content

1. Why Python
2. How to run Python Programs
3. Variables and Data Structures
4. Packages and Modules
5. Popular Packages for Data Science
6. Data Analysis (home work)

Learning Outcomes

At the end of the session participants will be able to:

1. Run a Python program in Jupyter Notebook
2. Understand variables and common data types (e.g. list) in Python
3. Mention common packages for data science in Python
4. Do exploratory data analysis with Python

Introduction to Python

What we will cover

- Why, What's Python
- Python on Windows
- Python 3 Vs. Python 2
- Installing Python-the Anaconda way
- Python Development Environment
- Python Basics
- Python Data Structures (Lists, Arrays etc)
- Packages and Functions

Why Python?























Top Languages in Data Science

Top Analytics/Data Science Tools

Next table has the top 10 most popular tools in 2016 poll

| Tool | 2016 % share | % change | % alone |
|--------------|-----------------|----------|---------|
| R | 49% | +4.5% | 1.4% |
| Python | 45.8% | +51% | 0.1% |
| SQL | 35.5% | +15% | 0% |
| Excel | 33.6% | +47% | 0.2% |
| RapidMiner | 32.6% | +3.5% | 11.7% |
| Hadoop | 22.1% | +20% | 0% |
| Spark | 21.6% | +91% | 0.2% |
| Tableau | 18.5% | +49% | 0.2% |
| KNIME | 18.0% | -10% | 4.4% |
| scikit-learn | 17.2% | +107% | 0% |

Top General Purpose Programming Languages

| Language Rank | Types | Spectrum Ranking |
|---------------|---|------------------|
| 1. C |    | 100.0 |
| 2. Java |    | 98.1 |
| 3. Python |   | 98.0 |
| 4. C++ |    | 95.9 |
| 5. R |  | 87.9 |
| 6. C# |    | 86.7 |
| 7. PHP |  | 82.8 |
| 8. JavaScript |   | 82.2 |
| 9. Ruby |   | 74.5 |
| 10. Go |   | 71.9 |

Python has become kind of a defacto language for data science, big data and machine.

[About](#)[Downloads](#)[Documentation](#)[Community](#)[Success Stories](#)[News](#)[Events](#)

Tweets by @ThePSF



Python Software ✓
@ThePSF



A Rising Tide Lifts All Ships by Tom Easterbrook, PyCon UK 2018 - pyvideo.org/pycon-uk-2018/... Tom discusses the label of assistive technology & where the technology of the future has a role to play.



A Rising Tide Lifts All S...
pyvideo.org



7h



Python Software ✓
@ThePSF



What is Python? Executive Summary

What is Python? Executive Summary

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Python Vs. Stata/SPSS

1. Anyone with a computer can run Python code
2. Sharing is easy
3. General purpose programming language

1. You need to purchase the software
2. Hard to share since the other party needs to buy the software as well
3. Statistical programming only

How to Get and Use Python

Install Python for your OS

1. Directly from:
<https://www.python.org/downloads/>
2. Anaconda: Python Data Science Platform
3. Other third party platforms (e.g., home brew)

Install essential
Libraries/Packages

Python is powered by thousands of user contributed and freely available packages which can be installed easily using command:
`pip install package-name`

Use Python

For this course, we installed Anaconda on the Lab machines. You can install Anaconda on your personal machine by following instructions here:
<https://www.anaconda.com/download/>

When the purpose of installing Python is Data Science, Anaconda offers many advantages:

- 720+ data science packages for interactive data visualizations, machine learning, deep learning and Big Data
- World-class package, dependency and environment management
- Python Interactive Development Environment

Python 3 Vs. Python 2

- Python 3: Python 3.x
- Python 2: Python 2.7.x
- Why are there 2 versions: Python 2 is legacy, so Python 3 is the future
- There are syntactical differences between the two

We will use Python 3.7 in this course

Explore Anaconda

1. Open Anaconda Navigator and ensure its working

Python Development Environments

Development Environment: a place where you write and/or run a Python program

1. Terminal/Shell-Command line
2. Interactive Development Environment (IDLE-default for Python, Pycharm, Sypder(similar to R studio))
3. IPython
4. Jupyter/IPython Notebook

We will use IPython/ (Now called Jupyter Notebook) Notebook

Running Python Programs

We will test some of the Development Environments before we settle in Jupyter Notebook

1. Terminal/Shell-Command line
2. IPython
3. Jupyter/IPython Notebook

From now on, I will be working in Jupyter Notebook
and you will follow along

First, Lets Install a Plugin for Table of Contents in Jupyter Notebook

1. Open Anaconda prompt
2. Copy and paste this command: `pip install jupyter_contrib_nbextensions`
3. Copy and paste this command: `jupyter contrib nbextension install --user`
4. If everything works well, close and reopen Anaconda Navigator
5. We will check once we open Jupyter Notebook if the installation worked