

Python for Finance and Data Analysis (2nd edition)

(for the 1st edition click [here](#))

Venue: [ISEG](#), Lisbon

Date: 9-11 Dezembro 2015

Instructor: José Pedro C. M. da Silva e-mail: silva@math.uni-wuppertal.de

Abstract:

This workshop is aimed at users with no or low proficiency in Python, who wish to apply numerical methods to financial and data-rich problems. The workshop will introduce the basic syntax as well as the PyData stack, i.e., the main python packages (NumPy, SciPy, Matplotlib, Pandas, Jupyter) necessary to start extracting the most out of your data.

Detailed Syllabus

Day 1 - "Unconscious Incompetence":

- Introduction to vcs and why it is important and relevant. Reproducibility. Diff. Log. Code History. Examples of situations adapted to the participants.
- Introduction git and GitHub. Create repository. Initial commit. Change. Commit. Have someone else change and commit. Probably base it on GitHub tutorial.
- IDE and virtual environments. Different solutions to manage packages. Anaconda, pip, docker. Where and how to run python. Conflicting versions. Test different versions. Spyder, PyCharm, Ipython, Geany, Emacs, vim, Python IDLE, Text Editor. Hello World example and IDE installation.
- Basic Python tutorial. Variables. Loops. Functions. Data Structures. Style. Python history.
- Basic IPython/Jupyter Tutorial. Notebook. Nbviewer. Slideshow mode. Sharing and securing notebooks.
- NumPy. Basic Array manipulations. Linear Algebra. Data crunching (mean, max, min, std, etc..). Comparison with Matlab.
- Matplotlib. Basic Plotting. Time Series plotting. Lots of examples (Pie Chart, Time Series, Bar Plots, 3D Plots, Violin plots, etc.) .

Day 2 - "Conscious Incompetence":

- SciPy. Probability distributions. Random number generation. Depending on the background of the participants, either go into ODE and Special Functions or go into statistical analysis and regression.
- Pandas. The main focus of the day. Explore data. Timeseries data. Upsampling, downsampling, missing data. Example with stock data, fx data, census data, insurance data, etc...
- Seaborn. Statistical data looking good. Publication ready, higher-level than matplotlib plots, where the emphasis is not on all the available methods but on the look.
- Bokeh. Interactive data. Create interactive plot both in Ipython and outside of with, possibly, real-time data.
- Sphinx and reportlab. Sphinx for automatic documentation and reportlab for automatically create templated reports. An end-of-day report example will be given.

Day 3 - "Conscious Competence":

- Introduction to sklearn, the de facto machine learning package for python. Simple demonstration of the most common models, i.e., decision Trees, SVM, K-Means and other clustering examples. Easiness of testing, cross-validating and optimizing parameters.
- Introduction to scraping tools, both for webpages (scrapy) and for text documents. Examples of data extraction from a relevant webpage and from a pdf.
- Numba+Cython+PySpark+Dask. High Performance Python. Depth will depend on the audience.

Short Syllabus

Day 1 - "Unconscious Incompetence"

- vcs with git and [GitHub](#)
- [Integrated Development Environments](#) and [Virtual Environments](#)
- [Python](#)
- [Jupyter](#) (former IPython)
- [Numpy](#)
- [Matplotlib](#)

2º dia - "Conscious Incompetence"

- [SciPy](#)
- [pandas](#)
- [seaborn](#)
- [bokeh](#)
- [Sphinx](#) and [reportlab](#)

3º dia - "Conscious Competence"

- [sklearn](#)
- [Scraping](#)
- [Webscraping](#)
- [Numba](#) - [Cython](#) - [Ibis](#) - [Dask](#) - [PySpark](#)