Bristol Python Workshop

Spring 2018

Instructor:	Ties de Kok | Tilburg University	**Date:**	23 to 25 May 2018
Email:	t.c.j.dekok@uvt.nl	**Place:**	University of Bristol

Workshop Page:

All workshop-specific materials are made available through a companion repository hosted on GitHub.

This repository is located here: *Bristol Python Workshop 2018*

Main Resources:

This workshop uses the following two resources as core foundation:

- Ties de Kok, *Learn Python for Research*, GitHub, 2017.
- Ties de Kok, *Python Natural Language Processing (NLP) Tutorial*, GitHub, 2017.

Additional Resources:

This is a restricted list of various interesting and useful resources that are worth highlighting:

- Al Sweigart, *Automate the boring stuff with Python* (Link to free HTML version), No Starch Press, 2015.
- Brandon Rhodes, *PyCon Pandas Tutorial* (GitHub page, Video), 2015.

Objectives:

This workshop is primarily designed to introduce the participants to the basic principles needed to use Python for Accounting and Finance research. We will discuss the following core elements: an efficient Python workflow, Python for data-handling, Python for gathering data from the web, and using Python for natural language processing (NLP). Each element will be introduced by a brief lecture, followed by a hands-on session where the participants will work on a mini-task relating to that element.

At the end of the workshop, an active participant should be comfortable to:

- set up a workflow to efficiently incorporate Python into their projects,
- comprehend and implement basic Python programming operations,
- use `Pandas` and `Numpy` for basic data handling tasks,
- execute basic web scraping tasks using `Requests` and `Requests-HTML`,
- process and analyze text documents using common Python NLP packages.

Prerequisites:

Prior knowledge of the Python programming language is not required to participate in this workshop.

Tentative Workshop Planning

	Wednesday	Thursday	Friday
09:00 - 10:00	09:15 – 10:45 Python introduction + Python workflow TBD	09:15 – 10:45 Gathering data from the web TBD	09:15 – 10:45 Intro to NLP with Python TBD
10:00 - 11:00			
11:00 - 12:00	11:00 – 12:15 Setup + personal help TBD	11:00 – 12:15 Setup + get started TBD	11:00 – 12:15 Setup + get started TBD
12:00 - 13:00			
13:00 - 14:00	13:00 – 13:45 Handling data TBD	13:00 – 15:00 Mini task: web scraping TBD	13:00 – 15:00 Mini task: Textual Analysis TBD
14:00 - 15:00	13:45 – 15:15 Mini task: handling data TBD		
15:00 - 16:00			15:15 – 16:30 Q&A for own projects TBD
16:00 - 17:00	16:00 – 17:00 Seminar TBD	16:00 – 17:00 Seminar TBD	

Preparation | hardware:

Large parts of the workshop involve so-called "mini tasks", these hands-on parts require a personal computer. For the instructions I will assume that you are using the Windows operating system, however, it should be no problem to participate with a computer running Mac OS or any of the Linux distributions.

Preparation | software:

We will be using the Python 3.6 version of the Anaconda Distribution as a starting point. The `Anaconda Distribution` is the most convenient way to get started with Python for data science purposes as it makes it easy to install, run, and upgrade a comprehensive Python environment.

☞ We will be using Python 3 exclusively, however, I will include a note whenever an important difference between Python 3 and Python 2 comes up.

Step 1: Install Anaconda on Windows/macOS/Linux:

Please make sure that you have the Python 3.6 Anaconda Distribution installed on your computer. Downloads are available here: Anaconda Distribution

☞ Not all Python packages/libraries that we will be using come pre-installed with Anaconda. Please follow step 2 to install all the necessary packages.

Step 2: Install additional requirements:

Installing each package manually is tedious and prone to errors, a better approach is to create a new `Conda` environment with the provided `environment.yml` file.

Please follow these steps:

1. Download the `environment.yml` file to your system: `download environment.yml`
2. Open a command prompt / shell and `cd` (change dir) to the folder containing the `environment.yml`
3. Run the following command: `conda env create -f environment.yml`

   ☞ Installing everything will take a while.
4. Activate the `bristol-py` environment by typing:
   - `activate bristol-py` on Windows
   - `source activate bristol-py` on Mac OS or Linux.

Note, if you want to use Spacy, NLTK, and/or Textblob then it is important to also download the corresponding language models. Without the language model these packages will not be very useful.

Install them as follows:

☞ I can help you during the workshop to get everything setup if you run into problems.

- NLTK (Link to docs)

  In a Jupyter Notebook run:

```
1 import nltk
2 nltk.download()
```

- TextBlob (Link to docs)

  In the command line / terminal run:

```
python -m textblob.download_corpora
```

- Spacy (Link to docs)

  ☞ If you installed using `requirements.yml` you can skip this step as the Spacy models are included.

  In the command line / terminal run:

```
python -m spacy download en
```

Text editor: We will primarily be using the `Jupyter Notebook` as our Python interface, this only requires a browser. However, it would be convenient to also have a basic text editor installed. For Windows I recommend installing `Notepad++` as a good first basic editor.

Complete overview of all additional packages:

☞ You don't need to run the commands below if you followed the steps above!

```
$ conda install spacy
$ conda install textacy
$ conda install textblob
$ conda install nltk
$ conda install tqdm
$ conda install deepdish
$ conda install xlrd
$ conda install openpyxl
$ conda install pytables
$ conda install qgrid
$ pip install pyldavis
$ pip install fuzzywuzzy
$ pip install git+https://github.com/kennethreitz/requests-html
$ pip install https://github.com/explosion/spacy-models/releases/download/
    en_core_web_sm-2.0.0/en_core_web_sm-2.0.0.tar.gz#en_core_web_sm
```