Introduction to Python
Data Science Summer School 2022

Tom Paskhalis

tom@paskhal.is

- **Duration:** 1 day (4 hours overall)

- **Date:** 27 July 2022

- **Location:** Zoom/YouTube

- **Instructor:** Tom Paskhalis (tom@paskhal.is)

- **Course Website:** bit.ly/DS3-Python

- **Course Registration:** https://ds3.ai/2022/python.html#register-form

Overview

Python is a versatile, multi-purpose, open-source programming language which has become one of the top choices for data engineering and advanced analytics. Due to its interactive nature and easy syntax Python is increasingly used in academia, business, and even journalism. The purpose of this course is to provide a gentle introduction to Python with a focus on its data analysis capabilities.

This workshop is aimed at participants who would like an overview of/introduction to data manipulation and statistical analysis in Python. It can also serve as an overview/refresher of the basics of Python programming. In the class we will focus on understanding the core Python objects and getting hands-on experience with the main aspects of data analysis.

Prerequisites

This is an introductory class and no prior experience with programming is required. Familiarity with some foundational statistical analysis (e.g. types of variables and measures of central tendency) and R/Stata would be an advantage, but is not essential.

Software

For this workshop I recommend using one of the 2 online platforms for working with Jupyter Noteboks:

- Google Colab, a cloud platform for hosting Jupyter Notebooks. You need to have a Google account, but it does not require any local installations.

- Kaggle Code, a platform for sharing and exploring data-science-focussed Jupyter Notebooks. Although technically owned by Google, you can register just for Kaggle website.

If you would prefer to install Jupyter Notebook on your local machine, there are two main ways to do this: pip and conda. Unless you have prior experience with Python, I recommend installing Anaconda distribution, which contains all the packages required for this course.

Class Schedule

Class will take place on 27 July.

Date	Time (CEST)	Topic
27 July	15:00-16:45	Introduction to Python objects and data types
	16:45-17:00	Exercise I
	17:00-17:15	Break
	17:15-18:00	Introduction to Pandas
	18:00-19:00	Exploratory data analysis and data visualization
	19:00-19:15	Exercise II

Materials

The workshop is designed to be self-contained. All the materials are available in this GitHub repository. However, some students might find additional materials helpful. The following texts provide a good introduction to Python programming with a focus on data analysis applications:

- John Guttag. 2016. *Introduction to Computation and Programming Using Python: With Application to Understanding Data.* 2nd ed. Cambridge, Massachusetts: The MIT Press

- Wes McKinney. 2017. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.* 2nd ed. Sebastopol, CA: O'Reilly Media

- Al Sweigart. 2019. *Automate the Boring Stuff with Python.* 2nd ed. San Francisco, CA: No Starch Press

Additional online resources:

- [Python For You and Me](#)

- [Python Wikibook](#)

- [Python 3 documentation](#) (intermediate and advanced)

Short bio

Tom Paskhalis is an Assistant Professor in Political Science and Data Science at [Trinity College Dublin](#), [Department of Political Science](#). Previously, he was a Postdoctoral Fellow at [New York University](#), [Center for Social Media and Politics](#). He received his PhD from [London School of Economics and Political Science](#), [Department of Methodology](#). His research focuses on political communication, comparative politics, and computational analysis.

More information:

🌐[tom.paskhal.is](#)

🐦[tpaskhalis](#)

🐙[tpaskhalis](#)