Python Programming - ACCTG 579 B

Instructor:	Ties de Kok	University of Washington	**Quarter:**	Sept. & Oct.
Email:	tdekok@uw.edu	**Location:**	Virtual

Workshop Page:

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

This repository is located here: *https://github.com/TiesdeKok/acctg-579B*

Main Resources:

This course uses the following two resources as core foundation:

- Ties de Kok, *Learn Python for Research*, GitHub

- Ties de Kok, *Python Natural Language Processing (NLP) Tutorial*, GitHub

Course summary:

This programming course is designed to introduce PhD students to the basic principles needed to use Python for their research projects. It consists of two parts, the first part will introduce the participants to the basic fundamentals and the second part will discuss more advanced topics. Each session will consist of recorded lecture material (asynchronous) and a 2.5 hour synchronous session. The synchronous session will start with a Q&A portion after which the participants will work on a set of problems related to the session.

Deliverables:

The problem set for each session is a deliverable that needs to be handed in. Deadlines will be communicated via the Github repository. A good part of this can be completed during the class hours, however, the participant might be required to complete the remaining problems of the problem set in their own time after class. The final grade will be based on the quality of the deliverables.

Learning objectives:

At the end of the part 1, an active participant should be comfortable to:

- set up a workflow to efficiently incorporate Python into their projects (including Jupyter).

- comprehend and implement basic Python programming operations.

- use `Pandas` and `Numpy` for basic data handling tasks.

- create visualizations to present data and results using Python.

- execute basic web scraping tasks using `Requests` and `Requests-HTML`.

- process and analyze basic text documents using `Spacy`.

At the end of the part 2, an active participant should be comfortable to:

- perform advanced web scraping using browser automation and Selenium.

- perform advanced NLP techniques to process and analyze text documents.

- use `Scikit-learn` to perform supervised machine learning (e.g. classification).

- use `Scikit-learn` to perform unsupervised machine learning (e.g. topic clustering).

- integrate Python with Stata (`IPyStata`) or R (`rpy2`) for a full research workflow.

- incorporate version control into their Python workflow using Git and Github.

Prerequisites:

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

☞   Make sure to prepare your computer accordingly, see `setup instructions`.

Content overview

Part 1: Basic topics

Python introduction (Class 1)

- Structure of the course

- Python Programming Language

- Python eco-system

- Using Python

- Jupyter Notebook

- Python syntax

Basic data handling with Pandas (Class 2)

- Introduction to Pandas

- Opening / Closing various file types

- Pandas operations

- Basic visualizations

Gathering data from the web (Class 3)

- Terminology / Ethics / Tools

- Interacting with an API

- Web scraping a page

Basics Natural Language Processing (Class 4)

- What is NLP / Textual Analysis

- Terminology / Tools

- Processing and Cleaning text

- Direct feature extraction (Regular expressions / dictionary counting)

Part 2: Advanced topics

Advanced web scraping (Class 5)

- Reverse-engineer HTTP requests
- Browser automation with Selenium

Advanced Natural Language Processing (Class 6)

- Representing text numerically
- Setting up a machine learning pipeline with Spacy
- Word embeddings with Gensim

Machine learning I (Class 7)

- Introduction to machine learning
- Terminology / Tools
- Basic classification with scikit-learn
- How/Where to obtain training data
- NLP and Machine learning
- Hyper-parameter optimization

Machine learning II (Class 8)

- Introduction to unsupervised machine learning
- Topic modelling with sci-kit learn and pyLDAvis
- NLP and Machine learning

Misc. Topics (Class 9 - Recording only)

- Combine Python with Stata (IPyStata)
- Combine Python with R (rpy2)
- Speed up code with multi-processing
- Version control with GitHub

Class Schedule

Part 1: Basic topics

Date	Time	Topic	Class Type
W - Sept. 8th	3pm - 5:30pm	Python Introduction	Remote
M - Sept. 13th	3pm - 5:30pm	Data handling with Pandas	Remote
W - Sept. 15th	3pm - 5:30pm	Gathering data from the web	Remote
T - Sept. 28th	3pm - 5:30pm	Basics Natural Language Processing	Remote

Part 2: Advanced topics

Date	Time	Topic	Class Type
F - Oct. 1st	3pm - 5:30pm	Advanced web scraping	Remote
M - Oct. 4th	3pm - 5:30pm	Advanced NLP	Remote
W - Oct. 6th	3pm - 5:30pm	Machine learning I	Remote
T - Oct. 12th	8:45am - 11am	Machine learning II	Remote
TBD	TBD	Happy Hour	In-person