

# Intermediate Python Skills

## Online course learning objectives

This course teaches the required skills to manipulate and visualize a variety of data types using Python.

### This course will help learners to:

- Review essential elements of Python programming.
- Understand data extraction and visualization techniques applied to social science research using Python.
- Make appropriate decisions when dealing with and manipulating data taken from the web and APIs.

**Language:** English

**Time to complete:** 16 hours

**Instructors:** Dr. Rob Mastrodomenico and Dr. Phillip Brooker

## Online course full syllabus

### MODULE ONE: OVERVIEW

Learn some core functionality in Python to allow you to perform the rest of the course.

#### Introduction

- Why the Intermediate Python course?
- Course goal and structure.
- Prerequisites.

#### Installing Python

- Anaconda for Windows & MAC.
- PyCharm for Windows & MAC.
- PRAW.
- PIP.

#### Functions and scripts

- What is a function?
- Creating a function.
- Writing a script.
- Using a function.

#### Reading from the web and files

- Getting data into Python from the web.
- Opening and writing files.

# Intermediate Python Skills

## MODULE TWO: MANIPULATING DATA

Learn how to manipulate data and produce visualizations.

### Pandas series

- Defining a Series using a Dictionary.
- Defining a Series using Lists.
- Operating on elements of a series using Numpy.
- Other methods of Series.

### Matplotlib series

- Importing Matplotlib.
- Types of plots in Series (line, histogram, box, horizontal box, density).
- How to edit the plots in Matplotlib.

### Pandas DataFrames

- Transforming data structures (lists, series, tuples and dictionary) in DataFrames.
- How to access elements of the DataFrame.
- Editing DataFrames.
- Introduction the ILOC method.
- Additional ways to access DataFrames.

### Matplotlib DataFrames

- Types of plots in DataFrames (box, density, line, scatter).
- Preparing the data for plotting.
- Editing the data for plotting.

## MODULE THREE: EXTRACTING DATA

Learn how to process different data types and authenticate in Python.

### Authentication

- The definition of authentication.
- The methods and process of authentication.
- Using PRAW to access Reddit.

### Introduction to JSON

- What is JSON?
- JSON in Python.
- Writing a JSON file.
- Reading a JSON file.
- Application: Websites that use JSON.
- Extracting data from JSON API.

# Intermediate Python Skills

## Introduction to XML

- What is XML and lxml?
- Writing a XML file.
- Reading a XML file.
- Application: Websites that use XML.
- Extracting data from XML API.

## Web scraping using HTML

- What is HTML?
- HTML Tags.
- HTML Page Structure.

## Web scraping using BeautifulSoup4

- What is Web Scraping?
- Scraping Rules.
- Inspecting the Page.
- BeautifulSoup4.
- Application: Identity site for scraping.
- Web Scraping.

## MODULE FOUR: PRACTICAL APPLICATION

Gives you the skills to effectively plan and execute a task in Python.

### Planning a research task

- Examine the research methodology and explore how Python can be used for a research task.

### 3-part structured activity: Data analysis challenge

- **Part 1:** Data acquisition using reddit.
- **Part 2:** Data examination using Pandas.
- **Part 3:** Data visualization using Matplotlib.

### Conclusion

- Wrapping up. What's next?