Python for Geospatial Analysis

A specialist course in Canberra

Audience: This is a course for scientists, engineers, and analysts working with geospatial data sets.

Outcome: By the end of the course, you will have all the knowledge you need to start programming competently in Python for scientific and engineering applications, with a focus on geospatial applications. You will know what's available with Python and how to use its powerful data types and amazing libraries to write clean, efficient code. You will have had experience with using Python for various scientific data manipulation tasks and solving a range of analytical tasks, including easily creating beautiful plots, performing Monte Carlo simulations, constructing statistical models, regression, optimisation, analysing images, time-series data, geospatial data, and plotting statistical data on maps.

Duration: 5 days

Dates: 26–30 November 2018

Venue: Training Choice, Level 4, 54 Marcus Clarke Street, Civic, Canberra

Format: Each topic is a mixture of hands-on exercises and expert instruction.

Instructors: Henry Walshaw, Edward Schofield, and/or Robert Layton

Prerequisites: Familiarity with programming concepts (in any language) is assumed.

Course Outline

Day 1: Introduction to Python

Day 1 covers the basics of using the Python language and standard library, with a focus on scientific and engineering applications, including tips and tricks for making this easy:

- Why use Python? What's possible?
- The *Jupyter* notebook and shell for rapid prototyping
- Modules and packages
- Python syntax and concepts: an introduction through examples
- Essential data types, tips and tricks
- Handling exceptions
- Tour of the amazing standard library
- Worked example: fetching and ranking real-time temperature data for global cities from a web API

Day 2: Handling, Analysing, and Presenting Data in Python

The Pandas package is an amazingly productive tool for working with different kinds of data in Python. Day 2 gives a comprehensive introduction to reading and writing some of the most important data formats in science and engineering and how to analyse and visualise data easily:

- Reading and writing essential data formats: CSV, Excel, SQL databases, JSON, time-series
- Indexing and selecting data in *Pandas*
- Data fusion: joining & merging datasets
- Summarisation with "group by" operations; pivot tables
- Time-series analysis: parsing dates, resampling
- Visualisation and statistical graphics with *Seaborn*
- Worked example: creating automated reports with *Jupyter*, *Pandas* and *nbconvert*

Day 3: Essentials of Scientific Computing with Python

Day 3 teaches the use of Python for numerical and scientific computing. It covers array and matrix manipulation and an overview of available scientific routines, including an introduction to statistical modelling:

- Introduction to manipulating vectors and matrices with *NumPy*
- Tour of *SciPy* and related packages for scientific data manipulation, with fancy demos: unit conversions, clustering with *scikit-learn*, interpolation, dense & sparse linear algebra, signal processing, image processing
- Statistics in Python, with scientific applications: modelling, confidence intervals, hypothesis testing, linear regression, Monte Carlo simulation

Day 4: Scaling up

Day 4 focuses on techniques for creating larger codebases in teams and scaling up from small datasets to large ones that are too big for memory or too slow for one computer to process. It also introduces further kinds of machine learning (classification and regression) for automatically inferring models from large datasets:

Morning: best practices

- Defining your own classes: custom data types
- Integrated development environments; Python idioms and style
- Sharing code with others: creating scripts, modules and packages; revision control with Git

Afternoon: scaling up

- Efficiency: profiling, vectorization, and JIT compilation for speeding up code by 4x to 10,000x
- Parallel computing with *Dask*
- Machine learning with *scikit-learn*: classification and nonlinear regression

Day 5: Spatial Analysis in Python

Maps are a powerful tool for data visualisation. Spatial data is ubiquitous and location analytics are more important than ever. A well drawn map is not only beautiful to look at, but can change how you see the world. In the last 10 years Python has become the go-to language for spatial science. This day will provide a tutorial in working with geospatial data using Python. It will cover spatial data access, spatial analysis, and visualising the results on a map.

- Reading & writing vector data with Geopandas and GDAL
- Reading and writing rasters with Rasterio
- Working with NetCDF data with xarray
- Projections with Geopandas, pyproj and shapely
- Creating beautiful maps with Cartopy and overlaying statistical data
- Introduction to vector and raster image analysis with PySAL and SciPy
- Introduction to network analysis with NetworkX (on request)

Supplemental materials

We will supply you with printed course notes and a USB stick containing kitchen-sink Python installers for multiple platforms, electronic versions of the course notes, solutions to the programming exercises, several written tutorials, and reference documentation on Python and the third-party packages covered in the course.

Instructor bios

Your trainers for the course will be selected from:

Henry Walshaw

Henry has over 10 years of experience in GIS, spatial analysis and application development, particularly in the natural resource management field. Henry's core technical expertise relates to the development and analysis of large scale spatial datasets (primarily using Python), and communicating this understanding to people including subject matter experts and the general public.

Henry has worked in government at federal and state levels, at Geoscience Australia (GA), the Victorian Government Department of Sustainability and Environment (DSE), and the Environmental Protection Agency (EPA). He has also worked in the private sector as Senior Spatial Consultant with Sinclair Knight Merz (SKM) and we-do-IT. He holds a Bachelors in Computational Science.

Dr Edward Schofield

Ed has consulted to or trained dozens of organisations in scientific applications of Python, including ANU, A*STAR, the Bureau of Meteorology, Cisco, CSIRO, Dolby, Geoscience Australia, Optus, and Toyota Technical Centre. Ed is the co-chair of the Python for Data Science miniconf for PyCon AU, co-organises the Python user group in Melbourne, and regularly presents at conferences related to Python and data analytics in Australia and internationally. He is a former release manager of *SciPy* and the author of the widely used *future* package.

Ed holds a PhD in computer science from Imperial College London, where his thesis was in machine learning. He also holds BA and MA (Hons) degrees in maths and computer science from Cambridge University. He has 20+ years of experience in programming, teaching, and public speaking.

Dr Robert Layton

Robert is the author of the book "Data Mining in Python", published by O'Reilly. He provides analysis, consultancy, research and development work to businesses primarily using Python. Robert has worked with government, financial and security sectors, in both a consultancy and academic role. He is also a Research Fellow at the Internet Commerce Security Laboratory, Federation University Australia, investigating cybercrime analytics and data-mining algorithms for attribution and profiling.

Robert is a regular contributor to the Python-based scikit-learn open source project for machine learning and writes regularly on data mining for a number of outlets. He has presented regularly at a number of international conferences in Python, data analysis, and its applications.

Other information

Computer: A computer with an internet connection will be provided for you for the course.

Personal help: Your trainer(s) will be available after the course each day for you to ask any one-on-one questions you like — whether about the course material and exercises or about specific problems you face in your work and how to use Python to solve them.

Food and drink: We will provide lunch, morning and afternoon tea, and drinks.

Timing: The course will run from 9:00 to roughly 17:00 each day, with breaks of 1 hour for lunch and 15 minutes each for morning and afternoon tea.

Abbreviated course option: People who have already taken our Python for Scientists and Engineers course (first 4 days) or have extensive prior experience in NumPy, SciPy, Pandas, and the core Python language may choose to attend only some days. Please contact us.

Price

$825 per day per person, including GST. Group discounts are available.

Booking

To book places on the course, please contact us, or visit:

https://pythoncharmers.com/training/python-for-geospatial-analysis

Testimonials

Testimonials from past participants of this course are available at pythoncharmers.com/testimonials.

Questions?

You are welcome to contact us if you have any questions before the course. You can reach us at info@pythoncharmers.com.

About Python Charmers

Python Charmers is the leading provider of Python training in the Asia-Pacific region, based in Australia and Singapore. Python Charmers specialises in teaching programming for scientists, engineers, data analysts, and computer scientists in the Python language.

Contact

Phone: +61 1300 963 160
Email: info@pythoncharmers.com
Web: pythoncharmers.com

PYTHON CHARMERS