

Python for Predictive Data Analytics

A specialist course in Sydney



Audience: This is a course for data scientists, financial analysts, researchers, statisticians, and software developers interested in learning to use Python for analysing and visualising data.

Outcome: By the end of the course, you will have all the knowledge you need to start using Python competently for processing, analysing, modelling, and visualising various kinds of data, with a focus on time series. You will have had experience with using Python for various scripting, data-manipulation and plotting tasks with data in a variety of formats, including CSV, Excel spreadsheets, SQL databases, JSON, and API endpoints, as well as unstructured text. You will have applied powerful tools for classification, regression, and clustering, in useful practical settings on small and large data sets. You will understand the elegance and power of the Python language and its powerful ecosystem of packages for data analytics, and you will be well-placed to continue learning more as you use it day-to-day.

Duration: 4 days

Dates: 11-14 September 2017

Venue: Training Choice, Level 4, 60 Clarence Street, Sydney CBD, NSW 2000

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

Instructor: Dr Robert Layton and/or Dr Edward Schofield

Prerequisites: Some experience with programming (in any language) would be beneficial. A quantitative background and familiarity with basic probability and statistics would also be beneficial.

Course Outline

Day 1: Python Basics

Day 1 covers how to use Python for basic scripting and automation tasks, including tips and tricks for making this easy. The syllabus is as follows:

- Why use Python for predictive analytics? What's possible? Python versus other languages ...
- Setting up your Python development environment (IDE, *Jupyter*)
- Modules and packages
- Python concepts: an introduction through examples
- Essential data types: strings, tuples, lists, dicts, sets
- Worked example: fetching and ranking real-time temperature data for global cities
- Raising and handling exceptions
- Handling CSV data: introduction to *Pandas*

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

The *Pandas* package is an amazingly productive tool for working with and analysing data in Python. Day 2 gives a thorough introduction to analysing data with *Pandas* and visualising it 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
- Publication-quality 2D plotting with *Seaborn* and *Matplotlib*
- Interactive visualisation with *Plotly*
- Worked example: creating automated reports with *Jupyter*, *Pandas*, and *nbconvert*

Day 3: Time-series, simulation, inference and modelling

Day 3 demonstrates more advanced features of *Pandas* for working with data, including time-series data. It then describes Monte Carlo simulation methods and walks you through using powerful Bayesian methods of inference and modelling for different kinds of data in Python:

- Time-series analysis: parsing dates, resampling, handling time-zones
- Secret weapons for *Pandas*: *searchsorted*, hierarchical indices, *unstack*, categorical, *qgrid*
- Introduction to *NumPy* for linear algebra and Monte Carlo simulation methods
- Classical statistics with *scipy.stats* and *statsmodels*
- Density estimation with *scikit-learn*
- Bayesian inference with *PyMC3*: parameter and model selection; incorporating prior information
- Bayesian regression; assessing reliabilities

Day 4: Machine learning

Day 4 introduces a more automated approach to modelling real-world data with several powerful machine learning algorithms using *scikit-learn*. The datasets are selected from a range of industries: financial, geospatial, medical, and social sciences. The syllabus is:

- Classification with *scikit-learn*: Naive Bayes, logistic regression, SVMs, random forests, with application to diagnosis, AI systems, and time-series prediction
- Nonlinear regression, with application to forecasting
- Clustering data with *DBScan*, with application to outlier detection
- Dimensionality reduction with *PCA*
- Validation and model selection
- Deploying machine learning models in production

We encourage you to bring your own data sets to the course where relevant.

Supplemental materials

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

Instructor bio

Your trainers for the course will be selected from:

Robert Layton

Robert is a data scientist who works across several industries including finance, information security, and transport. He is (2015), which has received significant praise. He is a developer for the *scikit-learn* package for machine learning and the author of the website *Learning Tensorflow*. He has presented at the last four PyCon AU conferences, at multiple international research conferences, and has given training in Python to groups of staff from companies including Cisco, Lumascape, IMC, Optus, Sportsbet, and Woolworths.

Robert has a PhD in cybercrime analytics from the Internet Commerce Security Laboratory at Federation University Australia, where he was the inaugural Young Alumni of the Year in 2014 and is now an Honorary Research Fellow. Robert is also an Official Member of the Ballarat Hackerspace, where he helps grow the future-tech sector in regional Victoria.

Edward Schofield Ed has consulted to or trained over 1500 people in Python for data analytics from dozens of organisations, including AGL, the Australian Federal Police, A*STAR, Barclays, the Bureau of Meteorology, Cisco, CSIRO, Dolby, DSTG, IMC, Macquarie Bank, Shell, Telstra, Toyota, and Verizon. 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 *future* package.

Ed holds a PhD in machine learning from Imperial College London, with application to speech and image recognition technologies. 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.

Other information

Computer: A computer will be provided for you during the course.

Exercises: There will be practical programming exercises throughout the course. These will be challenging and fun, and the solutions will be discussed after each exercise and provided as source code on the USB sticks. During the exercises, the trainer will offer individual help and suggestions.

Timing: The course will run from 9:00 to roughly 17:00 each day, with breaks of 45 minutes for lunch and 15 minutes each for morning and afternoon tea.

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. We encourage you to have your own data sets ready to use if this is relevant.

Certificate of completion: We will provide you a certificate if you complete the course and successfully answer the majority of the exercise questions.

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

Price

\$825 per day per person, including GST.

Booking

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

<https://pythoncharmners.com/training/python-for-predictive-analytics/>

Testimonials

Testimonials from past participants of similar courses are available at pythoncharmners.com/testimonials.

Questions?

You are welcome to contact us if you have any questions before the course. You can reach us at info@pythoncharmners.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 to scientists, engineers, financial engineers, data analysts, and computer scientists in the Python language. Python Charmers' delighted training clients include the ABC, Barclays, CSIRO, Dolby, Geoscience Australia, IMC, Macquarie Bank, Primary Health Care, Shell, Toyota Technical Centre, and Verizon.

Contact

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

