Python For All

Python skillshare workshop details

v2.0 by Wendy Olsen, with Rad Siddiqui and co-organisers.

Dates: Friday 27 Feb., 12.00-1.30 (room is booked from 11.30 onward)

Friday 27 March, 2026, 12.00-1.30 (room is booked from 11.30 onward)

And Friday 24 April, 2026, 12.00-1.30 (room is booked from 11.30 onward)

And advanced, Monday 25 May, 2026, 9.30-11 am (provisional) *ALL HYBRID*

This event is held in Meeting Rooms 1 & 2 in the Manchester office. You can bring your own drink and/or lunch.

FYI: All staff in the Manchester ONS office are invited.

Brown-bags: We also occasionally have lunch as a Python buddy lunch / skillshare, about once a month.

Which Type of Python?: If you have to choose, you might choose 'Python version 3.12' or v3.10 – this is your 'interpreter'. You may pick a code editor such as VS Code or Jupyter Inside VS Code. Or simply VS Code. Or, use Spyder; Spyder is okay for home laptop use.

Quick Summary of Six Workshops in 2026

Workshop 1 Basic Python – installation and coding - basic resources:

PSD_Capability_Building – It's all about CODING in basic Python*.

Workshop 2 Doing Loops in Python – see McKinney 3rd ed., "For" loops in coding.

Workshop 3 link to basic resources: Session Resources - Folder containing materials for some coding upskilling sessions (Defines what "DAP" "CATS" Wiki is; shows how one runs Python in DAP by demonstration, step by step)

Workshop 4 Starting Spark in Python - link to information about using big data: See Chambers and Zaharia, either Chapter 1 or Chapter 5.

Workshop 5 Outputting Tables, Percents, Graphs After PySpark – we provide code for you. This is an advanced Python session.

Workshop 6 Filters, Grouping, Sorting in PySpark – demonstration of advanced big-data coding.

General Advice: When you go outside with your ONS Python-friends: Keep Card Hidden, Don't Trust Strangers, and Keep Voice Private! This includes other civil servants from the upper floors of Manchester office.

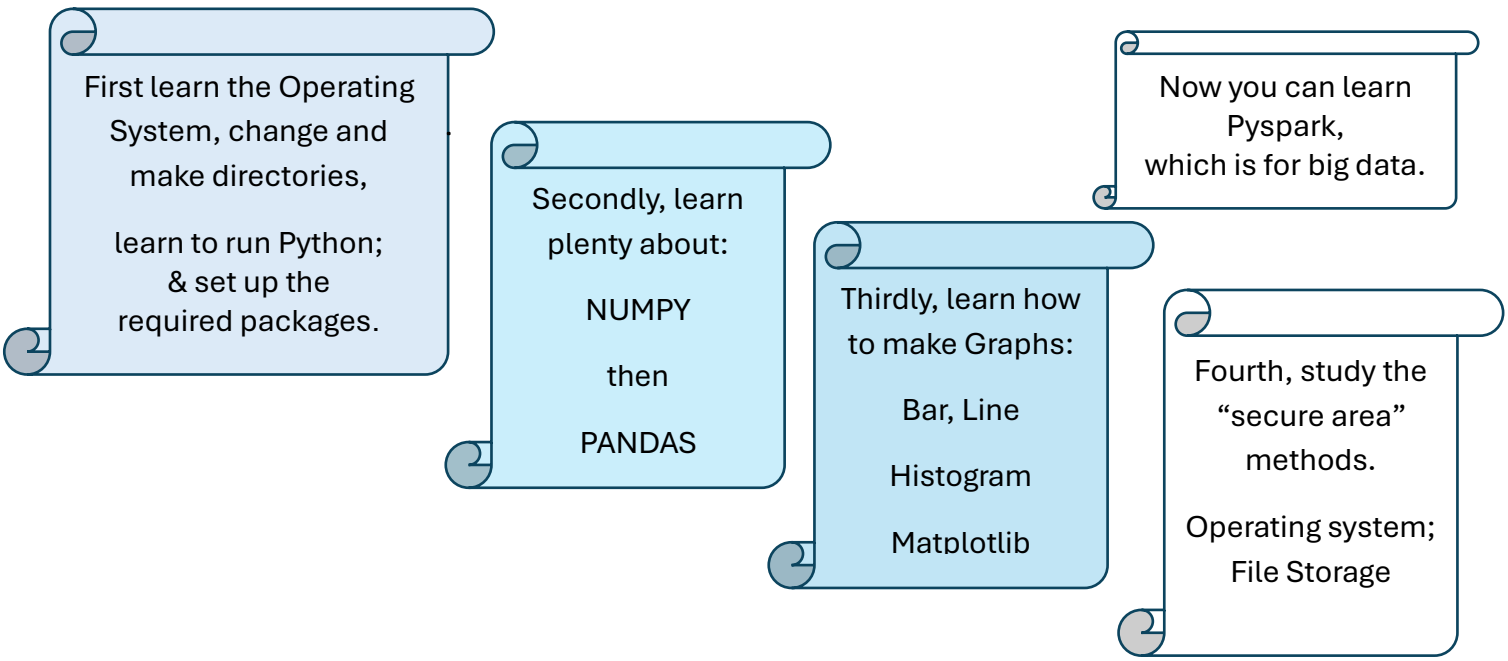
Python For All

Workshop 1 Link Python to your Operating system and use it!

We will help you learn how to use, or bypass, the ONS method of installing Python. You can just run Python at the command line – batch files – or as an interactive terminal in VS code – install – pip- packages and requirements file – how to edit the list of packages - finally getting output for the first time.

Outline of the workshop 12.00 Introduction, 12.10 Demonstration of Python for population (or investment & savings) growth forecasting 12.20 Installing & Setup Q&A; speedqueries. 12.45 Demonstration of setup and 7 basic python skills. These are how to output a CSV file; make a line graph; add a package; do algebra; make a list and transform it; what is the 'pandas dataframe' punctuation vs. the Pandas Series? Subset a data table. 1.15 activity on your own laptop.

(Participants:* Learning task 1: You are a Winner if you can calculate an answer to this problem using Python: 😊 Suppose the world population is 8 billion. Suppose the net pop growth rate is +0.85% per year. Calculate what year the world population would reach 14 billion. Later on, using a loop, substitute more accurate starter populations such as 8.2 billion, 8.45b. You could gather the sum from this kaggle file of country populations, https://www.kaggle.com/datasets/asadullahcreative/world-population-by-country-2025 , accessed Dec. 2025. Print a list of the starting population and year of reaching the target.) *answers will be released on request wendy.olsen@ons.gov.uk*



Python For All

The Python program needs to read/write from your operating system, eg from Terminal or from Linux. We examine the linkages together. A supportive book is *Learn Python the Hard Way*, various editions, Zed Shaw, PDF not free, available at https://www.informit.com/store/learn-python-the-hard-way-9780138270575 , see **also https://learnpythonthehardway.org/)**

Prepare by getting Python to run using the ONS guidance on Python. (see Learning Hub) - or ask a colleague on our organiser group to get you started. General learning Hub https://learninghub.ons.gov.uk/login/index.php and specific one is

"Installing a Python Package" if you are working inside the secure area 'DAP' ie Dapcats https://learninghub.ons.gov.uk/course/view.php?id=529 ; or for beginners, use a Laptop, and see an Introduction to Python course; one of these is at URL https://learninghub.ons.gov.uk/course/view.php?id=1168 and also has webpage file:///C:/Users/OlseNW/Downloads/ITP_course_intro%20(1).html. Without studying one of these links veeeery carefully, you might end up wasting time trying to run python just by guessing.

Learning task 2: can you write a bit of code, which sets the path for output files out of Python in your Python interpreter? You will have a chance to ask questions about key terms like interpreter, path, and output files so it is a workshop open to all, including beginners. You may want to try xlsxwriter package (you must use pip to install this, then import it) and write a csv or xlsx file. I myself use VS Code instead of Spyder but it's up to you. Spyder, Jupiter and VS Code are known as 'interfaces' or 'VDI's for Python. Choose one.

Learning task to jump into this as a Newbie: RunMS Copilot by using https://m365.cloud.microsoft/chat/?home=1&auth=2 [You will need to log in to MS 365 beforehand. Do not worry, it is fine to login to MS 365 but do it in your laptop. Don't do this in your secure area (DAP).] Ask copilot to write you some code to set the path for output files using your Python interpreter version 3.11 or which version you use. A beginner book: McKinney, Wes (2012), *Python for Data Analysis.* *3rd edition available https://wesmckinney.com/book , 3rd ed. Pub. 2022.(Open Access)*

Stretch aim: Some clever person might pool four tables into an Excel workbook, automating this process with xlsxwriter. This currently has a barrier inside our Secure Area, as it requires raz and boto3 packages there. But on your own laptop it should work all right. See https://xlsxwriter.readthedocs.io/tutorial01.html, note tutorial 1 :

```
import xlsxwriter

workbook = xlsxwriter.Workbook('Expenses01.xlsx')
```

```
worksheet = workbook.add_worksheet() # Write a total using a
formula

worksheet.write(row, 0, 'Total')

worksheet.write(row, 1, '=SUM(B1:B4)')

workbook.close()
```

Workshop 2 Python For Loops:

This time, let's share skills on doing a 'For-loop', which is a short programme code chunk that starts with For.

Outline of the workshop 12.00 Introductions & speedqueries, 12.15 Demonstration of Python 'for' loops, for savings growth forecasting 12.20 Practice with loops, Q&A. 12.45 Demonstration of 5 basic python skills. These are how to make a bar chart and output that to image format; how to input data; why we use synthetic data; query the Copilot; transpose a dataframe for publishing in Powerpoint. 1.15 activities on your laptop.

Learning Task 1:

Please prepare by using Range or similar before/during the workshop. For example use for i in visalist: lengthname[i] = str.len(visalist[i]).

Learning Task 2:

You can then loop through a list of many names to shorten them all to 6-letter abbreviations. Check out 'regular expressions' using https://www.datacamp.com/courses/regular-expressions-in-python, and Cheat Sheet URL https://www.datacamp.com/cheat-sheet/regular-expresso, which omits String commands, see also URL https://www.w3schools.com/python/python_ref_string.asp (Imagine if these were Value Label strings on a key variable; you are making a simple typology.) Or you can loop through a list of six numbers to pick out those >3000 and place them into a new vector. Which package would you use (we will help you decide; and DataCamp could illustrate; or look at this book that uses Python), or do we use "base" python.

Learning task 3:

can you set up a for-loop which breaks if Age is >90 but otherwise, adapts age to a new vector, Agein2020, based on an age variable which refers to age in 2025? We will explain how for-loops are not the same as new vector definitions, such as age20 = age-5, and yet their effect can be the same.

Resources for growing skills:

- We will provide a synthetic dataset to get you started. It is a csv file. You can use the Python outside of our secure area.
- See: https://www.taylorfrancis.com/books/mono/10.1201/9781003159834/foundations-statistics-data-scientists-alan-agresti-maria-kateri and this legitimate

personal author website with the Python appendix, https://stat4ds.rwth-aachen.de/

Workshop 3 When to Move from Python to PySpark and Back: [Spark is a big-data coding language]

This workshop relies on those who have 'secure access' in the Data Access Platform (DAP) of Cloudera at ONS. We will discuss the PySpark software which currently is not available unless one is inside the Cloudera platform. It is known as DAP. The PySpark is for big data manipulations, eg 251,000,000 cases. A set of packets is arranged so that each file-open action only refers to a small amount of memory, and parallel activities with 20 or even 200 packets can occur simultaneously. This can be called distributed or parallel processing. The PySpark is Apache Spark software, which works with Python, but can also work with R.

Learning task 1: Look up the admin data catalog of the ONS. Check out the size of some of the files there. Imagine using larger datasets to supplement your current data source.

Outline of the workshop 12.00 Introductions & speedqueries about the Secure Area 12.15 Demonstration of Pyspark on a Laptop, which is 'local' Pyspark. 12.30 Demonstration of 5 advanced pyspark skills. These are how to make a kernel density histogram of dates, and put that image into the files area; how to convert data from Pyspark 'spark dataframe' back to a Pandas dataframe or series; using Laptop code based on synthetic data & Copilot to develop Secure Area code by copy/paste and editing. 1:00 to 1:30 challenge exercises on your own laptop.

Stretch activity for beginners: Create a csv holding a series of strings. Open this csv as a dataframe in Python or Pyspark. Truncate all these to 6 letters, now do a frequency count of the distinct strings. This can now be drawn as a bar chart: for each distinct string, sorted by frequency, a vertical bar: we see how skewed or uniform this distribution is (see also McKinnery, 2012, pages 204-5 and 210). The suggested possible dataset is called 'retail' available via Kaggle at: `df = pd.read_csv('/kaggle/input/online-retails-sale-dataset/Online Retail.csv')` (one blog at online-retail-sales-analysis) [Not possible to download this inside the organisation.]

A more advanced learning task: Try creating an output such as a table inside DAP's python, using pyspark. Now drop this file onto your S: drive. Next find a way to run Python so that you graph that file inside DAP. This is a method which can work to quickly create graphs using matplotlib package. So to prepare, if unsure, study the matplotlib package, see McKinney (2012: Chapter 8). Try:

https://matplotlib.org/stable/users/index.html Hope you like this package but beware. Keep it simple at first.

Fun exercise for those new to Pyspark:

In your Laptop, ask Copilot to develop code to create a large dataframe (400K rows) containing five variables, of which 3 are correlated, 1 is binary, and 1 is multinomial. Now ask Copilot to develop the matplotlib graphs for 3 graphic images about these data. (1 barchart, 1 scattergram, and 1 date histogram.) If you have Secure Access, go into DAP and copy your new code into there. Edit the code (we are happy to help you), till it runs in DAP and creates 3 images inside DAP. Well done!

Resource for growing skills:

- The ONS has the DAP-CAT, a series of activities that help you learn about DAP activities.
- Chambers and Zaharia, *Spark: The Definitive Guide*. https://www.oreilly.com/library/view/spark-the-definitive/9781491912201/ (Use their ***Read Now*** button)

ANY FURTHER REQUESTS, send to wendy.olsen@ons.gov.uk

GENERAL INFORMATION:

Mode of workshops: These workshops are face-to-face pretty much, and not too often or too long. Using our site as a physical base, we also keep a virtual set of files for self-study.

Can I Work From Home? Yes: We aim to hold a hybrid version and make a videorecording. Plenty of time for Question-Answers. We help also with file locations.

Audience: anyone at all interested, from total beginners, up to expert.

Facilitator email wendy.olsen@ons.gov.uk

References:

Chambers and Zaharia, *Spark: The Definitive Guide,* https://www.oreilly.com/library/view/spark-the-definitive/9781491912201/ helps you study chunks of the book. (Use the ***Read Now button***)

McKinney, Wes (2012), *Python for Data Analysis: Data Wrangling with Pandas, Numpy, and Ipython*. Singapore: O'Reilly. *(3rd edition available https://wesmckinney.com/book , 3rd ed. Pub. 2022.(Open Access))*

Shaw, Zed (2018) *Learn Python the Hard Way*, various editions, PDF is not free, available at https://www.informit.com/store/learn-python-the-hard-way-9780138270575 , see also https://learnpythonthehardway.org/) 5th ed., ISBN 0-13-827057-0, London: Addison-Wesley. Don't forget to scroll down….

Practice skills: https://www.datacamp.com/courses/intro-to-python-for-data-science.

Acknowledgements- we thank Eleanor Woods and Tom Hodgson of ONS for their help with gauging the task/difficulty levels. We thank Rad Siddiqui and Roisin Aldag for the inputs about the arrangements, too.

Well done to you all, dear reader, for studying Python and doing all that self-learning and training. Maybe, try for a Certificate from DataCamp or Coursera! They don't cost much. https://www.datacamp.com/courses/intro-to-python-for-data-science

Or https://www.coursera.org/learn/python-programming-intro#modules is one example in Coursera (MOOC type of learning, mass open online course).

#End, this document version date is 20/02/2026.