Introduction to Python for Data Science (Part 1)

This planning document is intended to support teachers who are delivering the NPA/PDA Data Science or for students who are learning independently. It also aligns with the Data Skills for Work framework.

Contents

Lesson Description....1
Lesson Contents....2
Learning Intentions....2
Success Criteria....2
Knowledge Prerequisites....2
Lesson Requirements....3
Jupyter Notebook....3
Datasets....4
Python Learning Resources....5
How you can use this lesson....5

Lesson Description

Lesson Overview	An introduction to the use of Python for Data Science projects.
Topic	Tools and Languages
Book Chapter(s)	"Tools and Languages"

NPA level	5, 6
PDA level	7, 8
Data skills for work level	Analysis

Lesson Contents

This lesson consists of:

- A lesson plan (this document)
- A Powerpoint presentation, 'Introduction to Python for Data Science (Part 1)'
- 2 Jupyter notebooks:
    - 'intro_to_python_for_data_science_part_1.ipynb' (for learners)
    - 'intro_to_python_for_data_science_with_answers_part_1.ipynb' (for teachers)

The Jupyter notebook for teachers contains answers to the tasks set for learners.

This Learning Intentions

We are learning about Python, specifically to:

- understand why Python is widely used in data science
- install, import and use Python packages
- understand how to get help when using Python
- name variables clearly and consistently

Success Criteria

I can *state* why Python is widely used in data science

I can *install*, *import* and *use* Python packages

I can *use* in-built help and documentation to help me when using Python

I can *name* variables clearly and consistently

Knowledge Prerequisites

- Python programming to at least the level defined in SQA Computer Programming Level 5 (HY2C 45)
- How to use a Jupyter notebook to write, edit and run Python code

For learners who do not have the prerequisite Python skills, a list of external learning resources that could be used to develop these skills is provided in the Python Learning Resources section below.

Lesson Requirements

	PDA	NPA	Data Skills for work
Qualification	Yes	Yes	Yes
Outcome ID(s)	WD7.3a, WD8.1j	DS5.1d, DC6.2b	A2.1, A5.1, A5.2
Outcome description(s)	WD7.3a Types of software for data analysis WD8.1j Tools for data analysis	DS5.1d Describe the tools that can be used at each stage in the life cycle DC6.2b Explain how data can be analysed and the tools that can be used to perform analysis	A2.1 Use of tools to analyse data A5.1 Use of programming languages A5.2 Programming for analysis
Level	7, 8	5, 6	Analysis
Software language	Python	Python	Python
Required equipment /software for student	Lesson: PowerPoint Python notebook: Jupyter notebook environment	Lesson: PowerPoint Python notebook: Jupyter notebook environment	Lesson: PowerPoint Python notebook: Jupyter notebook environment

Jupyter Notebook

There is a Jupyter notebook for this lesson that provides examples and programming tasks for learners, drawn from the examples in the lesson Powerpoint.

The notebook uses Python 3.x and the following packages:

- emoji – for displaying emojis
- sys – for accessing system-specific parameters and functions
- numpy – for scientific computing
- pandas - for data manipulation
- s3fs - an API to AWS S3 (Simple Storage Service), used to import datasets

The tasks are described in the table below.

Notebook section	Task	Description
Install Packages	Task 1 - Install numpy	Install a package using the !pip command.
	Task 2 - Install pandas and s3fs	Install 2 packages using the !pip command.
Import Modules	Task 3 - Import emoji	Import a module using **import *module***
	Task 4 - Import numpy	Import a module using an alias
Use Modules	Task 5 - Keep on Smiling	Write simple code to access a function in a module that has been imported.
	Task 6 - What's My Python	Write simple code to access a variable in a module that has been imported.
	Task 7 – Are You Being Series?	Write simple code to access a class in a module that has been imported, and create an object using that class's constructor function.
Name Variables	Task 8 - And This Week's Star Namer is...	Choose a good name for a variable.
	Task 9 - Think of a Name	Given a description of a variable, decide on a suitable name for it.
	Task 10 - I'm Bad	Given some code with poorly-named variables: • State why the choice of names is poor • Suggest better alternatives for 3 of them
	Extension Task 1	Rewrite the code provided in 'I'm Bad', using better names for the variables.

Datasets

The following datasets are used in this lesson.

Dataset name	Description	Link
archery	The scores in an archery competition	https://datasets.learn-data.science/archery.csv

Python Learning Resources

- Python For Everybody (https://www.py4e.com/html3/
- W3 Schools Python Tutorial (https://www.w3schools.com/python/)
- Data Carpentry Introduction to Python (https://swcarpentry.github.io/python-novice-gapminder/)
- Python.Land (https://python.land/)

How you can use this lesson

This lesson has been created by Effini in partnership with Data Education in Schools, The Data Lab and Data Skills for Work, with funding from the Scottish Government.

© 2021. This work is licensed under a *CC BY-NC-SA 4.0 license*.

You are free to:

- **Share** – copy and redistribute the material in any medium or format
- **Adapt** – remix, transform and build upon the material

Under the following terms:

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NonCommercial** — You may not use the material for commercial purposes.
- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.