Introduction to Jupyter notebooks

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

Lesson Overview	An introduction to Jupyter notebooks as a tool for writing code 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, 'Jupyter Notebooks'
- A Jupyter notebook, 'intro_to_jupyter.ipynb'
- A dataset used in the Jupyter notebook: the dataset is stored online and imported by the Jupyter notebook.

Learning Intentions

We are learning about Jupyter notebooks, specifically to understand:

- what a Jupyter notebook is
- why they are useful for data science
- how to use a Jupyter notebook, including opening/closing and saving a notebook, importing data, running code and adding text and code to a notebook

Success Criteria

I can *describe* what a Jupyter notebook is.

I can *describe* why Jupyter notebooks are useful for data science.

I can *use* a Jupyter notebook to open, edit and save a document, import data and run code.

Knowledge Prerequisites

There are no knowledge prerequisites for this lesson.

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 Environments

There are several software environments which support Jupyter notebooks.

Examples include:

- Jupyter Notebook
- JupyterLab
- Google Colab
- Noteable

The Jupyter notebook which accompanies this lesson, and the Jupyter notebooks which accompanies all the other lessons, can be run in any software environment which supports Jupyter notebooks.

As each environment provides a slightly different user interface, it's not possible within the resources provided to include detailed instructions on how to use each environment. The approach that has been adopted here is to make the instructions as generally-applicable as possible, whilst providing learners with practical guidance that they can follow to learn the basics of using the environment.

Consequently, the Jupyter notebook for this lesson contains some instructions that relate to the Google Colab user interface. The contents of the Powerpoint presentation are relevant for all Jupyter notebook environments, except for the instructions on opening and closing Jupyter notebooks on slides 15 and 18, which relate to the Google Colab user interface.

Whilst other Jupyter notebook environments are not dissimilar to Google Colab, teachers who have chosen to use an environment other than Google Colab may chose to adapt the lesson materials for their chosen environment.

Choosing a Jupyter Notebook Environment

This section provides a short overview of some environments which support Jupyter notebooks. All are free and run within a web browser.

Environment	Requires Software Installation?	Notes	Link
Jupyter Notebook	No (if used with binder)	The original Jupyter notebook environment.	https://jupyter.org/
JupyterLab	No (if used with binder)	An updated version of the Jupyter Notebook with a more modern user interface.	https://jupyter.org/
Google Colab	No	Users need to have a Google account in order to access the service.	https://colab.research.google.com/
Noteable	No	A service provided by The University of Edinburgh. Teachers and learners with Glow accounts can access the service using a Single Sign On.	https://dataed.in/aboutnoteable

		The user interface is the same as Jupyter Notebook.	

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:

- pandas - for data manipulation
- s3fs - an API to AWS S3 (Simple Storage Service), used to import datasets
- matplotlib – a data visualization library
- seaborn – a data visualization library

The tasks are described in the table below.

Notebook section	Task	Description
Editing Text Cells	Change Me!	Change the wording in a plain-text markdown cell
	Change me as well!	Add some simple markdown to a markdown cell
Editing and Running Code Cells	Change “hello world” to "hello Ducky / beautiful / lovely or your name!".	Change a string in a very simple line of Python (a print() function) and run the code cell
Adding New Cells	Write about what you enjoy	Create a new text cell and add some text to it.
	Simple Python	Create a new code cell, add some simple Python code to it, and run it.
Adding New Sections	My New Section	Create a new markdown cell and using markdown headings (e.g. ##) to create a new section in the notebook.
Importing Data	Import fictional characters dataset	Run a code cell to import a dataset in csv format from the internet into the notebook.
Outputting Graphs	Cities Graph	Run a code cell to output a graph.

Saving the Notebook	Save the Notebook	Save the notebook by downloading it.

Datasets

The following dataset is used in this lesson.

Dataset name	Description	Link
Fictional characters	The names and addresses of fictional characters from books and films.	https://datasets.learn-data.science/fictional_characters.csv