

Manipulating dataset columns in Python

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 Intention	2
Success Criteria	2
Knowledge Prerequisites	2
Lesson Requirements.....	3
Python Notebook	4
How you can use this lesson.....	5

Lesson Description

Lesson Overview	Selecting, reordering and reformatting columns
Topic	Data manipulation
Book Chapter(s)	“8. Data Transformation and Manipulation”

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

Lesson Contents

This lesson consists of:

- A lesson plan (this document)
- A Powerpoint presentation, 'Manipulating dataset columns in Python'
- Jupyter notebooks:
 - 'data_manipulation_of_columns_with_answers.ipynb' (for teachers), and
 - 'data_manipulation_of_columns.ipynb' (for learners)
- Datasets used in the Jupyter notebook: the datasets are stored in 'the cloud' and imported by the Jupyter notebook.

Learning Intention

We will be learning how to manipulate data in Python, specifically to be able to:

- select and reorder columns
- reformat columns

Success Criteria

I can *describe* what it means to select, reformat and reorder columns in a data frame.

I can *manipulate* data by selecting, reformatting and reordering columns in Python.

Knowledge Prerequisites

Learners should know:

- Data is held in structured data frames
- Python is a programming language that can be used for data analysis
- How to use a Jupyter notebook to write, edit and run Python code
- How to open a Jupyter notebook

Lesson Requirements

	PDA	NPA	Data Skills for work
Qualification	Yes	Yes	Yes
Outcome ID(s)	WD8.3b, WD8.3c, CD8.1g, WD7.2a, WD7.2b, CD7.3a	DS4.2c, DS4.3a, DS5.2c, DS5.3c, DS6.2b, DS6.3c	C2.1, A1.2, A2.3
Outcome description(s)	<p>WD8.3b Types of data transformation</p> <p>WD8.3c Transformations</p> <p>CD8.1g Preparing data for visualisation</p> <p>WD7.2a Types of data transformation</p> <p>WD7.2b Common transformations including filtering, sorting</p> <p>CD7.3a Preparing data for visualisation</p> <p><i>N.B. out of scope of this lesson:</i></p> <p><i>“WD8.3c ... including joins”</i></p> <p><i>“WD7.2bcombining, separating and grouping”</i></p>	<p>DS4.2c Describe simple methods of cleaning and transforming data</p> <p>DS4.3a Perform simple data cleaning and structuring</p> <p>DS5.2c Describe methods of cleaning and transforming data</p> <p>DS5.3c Perform routine data cleaning and structuring.</p> <p>DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling</p> <p>DS6.3c Perform data transformation to complete, correct and structure data</p>	<p>C2.1 Vocabulary used in data science and analytics</p> <p>A1.2 Data quality</p> <p>A2.3 Data calculation and manipulation</p>

Level	7, 8	5, 6	Core, 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

Python Notebook

There is a Python 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

The notebooks can be used with any Jupyter notebook environment. The tasks are described in the table below.

Notebook section	Task	Description
Select columns	Task 1 - Volume and Depth of Lochs	Select named columns from a data frame
	Extension Task 1 - Cities	Select columns of your choice from a data frame
Reorder columns	Task 2 - Reorder Cities Differently	Reorder columns in a data frame in a specific order
	Extension Task 2 - Reorder Lochs	Reorder columns in a data frame in an order of your choice
Reformat columns	Task 3 - How Much for the Car?	Reformat a column in a data frame using an existing Python dictionary
	Task 4 - Car Ratings	Reformat a column in a data frame using a Python dictionary you create yourself (with step-by-step guidance)
	Extension Task 3 - Simpler Numbers Please	Reformat a column in a data frame using a Python dictionary you create yourself (with less guidance)

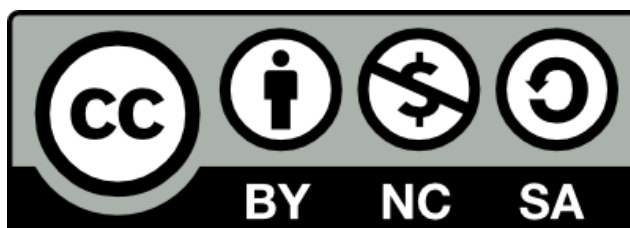
How you can use this lesson

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** – the medium
- **Adapt** – remix, transform and build upon the material



copy and redistribute material in any or format

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.