Practise Reshaping Datasets 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

Version Control

Version number	Purpose/Change	By	Date
1.0	Published by Effini	John Bell, Effini	27 Jan 2022

Lesson Description

Lesson Overview	Reshape datasets from long to wide and wide to long formats.
Topic	Data manipulation
Book Chapter(s)	“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, 'Practise Reshaping Data in Python'
- 2 Jupyter notebooks:
    - 'practise_reshaping_data.ipynb' (for learners)
    - 'practise_reshaping_data_with_answers.ipynb' (for teachers)

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

Learning Intentions

We are learning to reshape data in Python, specifically to,

- understand how pandas **indexes** work
- reshape a dataset from a **long to wide** format
- reshape a dataset from a **wide to long** format

Success Criteria

I can *describe* how pandas indexes work

I can *reshape* a dataset from a long to wide format in Python

I can *reshape* a dataset from a wide to long format in Python

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)	WD8.3b, WD8.3c, CD8.1g, WD7.2a, WD7.2b, CD7.3a	DS5.3c, DS6.2b, DS6.3c	C2.1, A1.2, A1.3, A2.1, A2.3
Outcome description(s)	WD8.3b Types of data transformation, WD8.3c Transformations […] CD8.1g Preparing data for visualisation WD7.2a Types of data transformation. WD7.2b Common transformations […] CD7.3a Preparing data for visualisation N.B. out of scope of this lesson, *"WD7.2b …filtering, sorting, combining, separating and grouping"* *"WD8.3c … including joins"*	DS5.3c Perform routine data cleaning and structuring. DS6.2b Explain techniques for data capture, cleaning and transformation DS6.3c Perform data transformation to complete, correct and structure data.	C2.1 Vocabulary used in data science and analytics A1.2 Data quality A1.3 Interpretation and insight A2.1 Use of tools to analyse data A2.3 Data calculation and manipulation
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

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:

- 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
Indexes	Task 1- GBBO index	Find some properties of a pandas index.
Reshape from Long to Wide	Task 2 - Which columns to use?	Choose the correct columns to use when reshaping a data frame using pivot().
	Task 3 - Pivoting makes you happy	Reshape a data frame from long to wide.
	Extension Task 1 - Change in Happiness	Add a new variable to the wide data frame just created, using a calculation involving other variables in the data frame.
Reshape from Wide to Long	Task 4 - It's melting in here	Choose the correct columns to use when reshaping a data frame using melt().
	Task 5 - Melting makes you happy	Reshape a data frame from wide to long.

Datasets

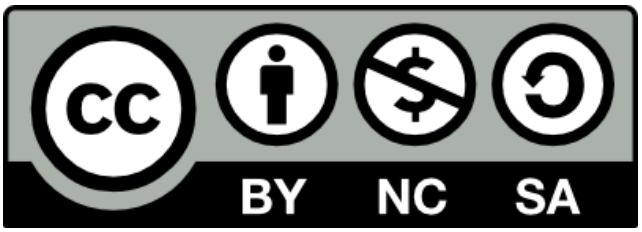
The following datasets are used in this lesson.

Dataset name	Description	Link
forbes_celebrities_simple	Forbes Celebrity list - the 100 highest-earning celebrities from 1999-2020	https://datasets.learn-data.science/forbes_celebrities_simple.csv
climate_data_per_city_uk	The average yearly temperature for cities in the UK from 1743 to 2013	https://datasets.learn-data.science/climate_data_per_city_uk.csv
climate_data_simple	The average yearly temperature for Aberdeen and Edinburgh from 2011 to 2013	https://datasets.learn-data.science/climate_data_simple.csv

gbbo_ingredients	Ingredients used by contestants in The Great British Bakeoff, Series 12	https://datasets.learn-data.science/gbbo_ingredients.csv
world_happiness_scores	The World Happiness Report averaged 'Happiness Scores' of sampled individuals from around the world from 2015 to 2019	https://datasets.learn-data.science/world_happiness_scores.csv
climate_data_wide_simple	A wide version of the climate_data_simple dataset	https://datasets.learn-data.science/climate_data_wide_simple.csv
forbes_celebrities_wide_simple	A wide version of the forbes_celebrities_simple dataset	https://datasets.learn-data.science/forbes_celebrities_wide_simple.csv

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.



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.

Alternative format

If you require this document in an alternative format, such as large print or a coloured background, please contact

hello@effini.com
or
4th Floor, The Bayes Centre
47 Potterrow
Edinburgh
EH8 9BT