Creating new variables by calculation in Python (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

Lesson Overview	Creating new calculated variables, where the calculation that is used is the same for each row in the dataset
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, 'Creating new variables by calculation in Python part 1'
- Jupyter notebooks:
    - 'creating_variables_by_calculation_with_answers_part_1.ipynb' (for teachers), and
    - 'creating_variables_by_calculation_part_1.ipynb' (for learners)
- Datasets used in the Jupyter notebooks: the datasets are stored online and imported by the Jupyter notebooks.

Learning Intentions

We will be learning how to create new variables in Python, specifically to,

- understand what it means to **create a new variable by performing calculations** using existing data
- how to create simple new variables by performing a calculation in Python

Success Criteria

I can *describe* how to create a new variable by performing calculations.

I can *create* new variables in Python by performing calculations.

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 to write, edit and run Python code


Effini

THE
DATA LAB
value from data

Data Skills
FOR WORK

Data Education
in Schools

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	DS5.2c, DS5.3c, DS6.2b, DS6.3c	C2.1, A1.2, 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 including filtering, sorting CD7.3a Preparing data for visualisation *N.B. out of scope of this lesson,* *"WD8.3c … including joins"* *"WD7.2b ….combining, separating and grouping"*	DS5.2c Describe methods of cleaning and transforming data DS5.3c Perform routine data cleaning and structuring. DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling DS6.3c Perform data transformation to complete, correct and structure data *N.B. out of scope of this lesson,* *"DS5.3d …including sort, filter…, group and summarise."*	C2.1 Vocabulary used in data science and analytics A1.2 Data quality A2.3 Data calculation and manipulation *N.B. out of scope of this lesson "A1.1….quantitative and qualitative"*
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:

- 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
Create a New Variable By Performing Calculations on Existing Variables	Task 1 - Who Won?	Creating a new variable by summing the values across some columns.
	Task 2 - Best Scores	Creating a new variable that is the maximum of the values across some columns.

Datasets

The following datasets are used in this lesson.

Dataset name	Description	Link
Gym staff	The name, start date and leaving date of members of gym staff.	https://datasets.learn-data.science/gym_staff.csv
Archery	The scores in an archery competition	https://datasets.learn-data.science/archery.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.