**Combining Datasets**

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 1](#_Toc132350193)

[Lesson Description 2](#_Toc132350194)

[Lesson Contents 2](#_Toc132350195)

[Learning Intentions 3](#_Toc132350196)

[Success Criteria 3](#_Toc132350197)

[Knowledge Prerequisites 3](#_Toc132350198)

[Lesson Requirements 4](#_Toc132350199)

[Task-types 5](#_Toc132350200)

[Worksheet 6](#_Toc132350201)

[How you can use this lesson 7](#_Toc132350202)

[Alternative format 7](#_Toc132350203)

# Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Version number | Purpose/Change | By | Date |
| 1.0 | Published by effini | Emma Nylk | 14 Apr 2023 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Lesson Description

|  |  |
| --- | --- |
| **Lesson Overview** | Overview of the theory of combining datasets by appending rows and joining columns. |
| **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/PDF presentation, ‘Combining Datasets’
* Question worksheet (for learners) on ‘Combining Datasets’ in Excel/PDF
* Answers worksheet (for teachers) on ‘Combining Datasets’ in Excel/PDF

# Learning Intentions

We will be learning **how to combine datasets**, specifically

* what we mean by **combining datasets**
* to add rows to a dataset by **appending**
* to add columns to a dataset by **joining**
* understand common **types of joins**

# Success Criteria

I can *describe* how to append rows to a dataset

I can *describe* how to join columns to a dataset

I can *explain* the difference between the types of joins (left, right, inner, outer)

I can *explain* what are the common causes of issues when combining datasets

# Knowledge Prerequisites

Learners should know:

* what data is
* data can be transformed into valuable information
* data can be used to solve problems and find answers to questions
* data can be stored in different types
* how to rows and columns in a dataset can manipulated

Learners should complete the **Data types and storage** lesson before undertaking this lesson.

Also, the **Manipulating dataset rows/columns** lessons in the either Excel or Python.

# Lesson Requirements

|  |  |  |  |
| --- | --- | --- | --- |
|  | **PDA** | **NPA** | **Data Skills for work** |
| **Qualification** | Yes | Yes | Yes |
| **Outcome ID(s)** | WD8.3c, CD8.1g, WD7.2a, WD7.2b, CD7.3a | DS5.3c, DS5.3d, DS6.2b, DS6.3c | A2.1, A2.3, C2.1, |
| **Outcome description(s)** | WD8.3c - Transformations including joins,  CD8.1g - Preparing data for visualisation,  WD7.2a - Types of data transformation,  WD7.2b - Common transformations including filtering, sorting, combining, separating and grouping,  CD7.3a - Preparing data for visualisation | DS5.3c - Perform routine data cleaning and structuring,  DS5.3d - Perform analyses including query, sort, filter, consolidate, group and summarise,  DS6.2b - Explain techniques for data capture, cleaning and transformation including data modelling,  DS6.3c - Perform data transformation to complete, correct and structure data. | A2.1 - Use of tools to analyse data,  A2.3 - Data calculation and manipulation,  C2.1 - Vocabulary used in data science and analytics, |
| **Level** | 7/8 | 5/6 | Core, Analysis |
| **Software language** | N/A | N/A | N/A |
| **Required equipment /software for student** | Lesson: PowerPoint/PDF,  Worksheet: Excel/PDF | Lesson: PowerPoint/PDF,  Worksheet: Excel/PDF | Lesson: PowerPoint/PDF,  Worksheet: Excel/PDF |

# Task-types

In the worksheet for this lesson, there are up to 6 task-types to that become increasingly challenging to support the students learning. Based on the student’s previous knowledge it is possible to select the task-types that are relevant to their stage.

|  |  |
| --- | --- |
| **Task-type** | **Description** |
| **1. Recall** | To be able to recognise definitions or procedures. |
| **2. Define** | To be able to define definitions or procedures. |
| **3. Rephrase** | To be able to use their own words to describe definitions or procedures. |
| **4. Apply** | To be able to apply definitions or procedures to problem-solving activities. |
| **5. Create** | To be able to apply definitions or procedures and create their own solutions to a defined problem. |
| **6. Active** | Using knowledge from the lesson which they apply to scenarios they have researched/designed themselves. |

# Worksheet

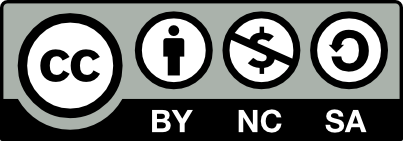
The worksheet associated with this lesson is available either in Excel or as a PDF that can be printed. The answer worksheet is available in both formats too.

|  |  |  |  |
| --- | --- | --- | --- |
| **Worksheet section ID** | **Description** | **Task-type** | **Number of questions** |
| 1.1 | Appending | Recall | 2 |
| 1.2 | Appending | Define | 1 |
| 1.3 | Appending | Apply | 6 |
| 1.4 | Appending | Active | 1 |
| 2.1 | Joining and keys | Recall | 1 |
| 2.2 | Joining and keys | Define | 1 |
| 2.3 | Joining and keys | Apply | 4 |
| 3.1 | Types of joins | Recall | 1 |
| 3.2 | Types of joins | Apply | 8 |
| 4.1 | More joining | Rephase | 1 |
| 4.2 | More joining | Apply | 3 |
| **Total** | | | **29** |

# How you can use this lesson

This lesson has been created by effini in partnership with The Data Lab.

© 2023. This work is licensed under a [*CC BY-NC-SA 4.0 license*](https://creativecommons.org/licenses/by-nc/4.0/legalcode)*.*



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](https://creativecommons.org/licenses/by-nc/4.0/), provide a link to the license, and [indicate if changes were made](https://creativecommons.org/licenses/by-nc/4.0/). 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](https://creativecommons.org/licenses/by-nc/4.0/).
* **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](https://creativecommons.org/licenses/by-nc-sa/4.0/) 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**