Advanced data cleansing in Excel

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.

The lesson has been designed for learners using Microsoft Excel. Most of the information in the lesson will work for other spreadsheets tools. However, if another tool is being used by the learners (such as Google Sheets) the step-by-step instructions will need to be adjusted.

Contents

Version Control

Version number	Purpose/Change	By	Date
1.0	Published by Effini	Emma Nylk	11 Mar 2022

Lesson Description

Lesson Overview	*Following on from the Data cleansing in Excel lesson.* Introduction to advanced data cleansing activities as part of the analysis steps, including, Converting between different data types Fixing strings Focusing on the causes of missing/outlying values and how these impacts on how you handle them.
Topic	Data Manipulation and Data Analysis
Book Chapter(s)	Analysing data

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, 'Advanced dataset cleansing in Excel'
- Question worksheet (for learners) on 'Advanced dataset cleansing in Excel' in Excel
- Answers worksheet (for teachers) on 'Advanced dataset cleansing in Excel' in Excel/PDF

Learning Intentions

We will be learning about advanced data cleansing in Excel, specifically,

- how to convert between different data types
- how to fix strings
- understand the reasons why there may be missing or outlying values, and how these reasons affect the ways in which we handle them

Success Criteria

I can *convert* between different data types in Excel.

I can *fix* a string in Excel.

I can *describe* what might cause a missing value and why this is important for fixing them.

I can *describe* what might cause an outlying value and why this is important for fixing them.

Knowledge Prerequisites

Learners should know:

- what a dataset is
- data can be used to solve problems and find answers to questions
- that data understanding is part of the analysis steps
- that data cleansing is part of the analysis steps
- how to identify missing and outlying values
- how to filter datasets in Excel
- difference between data types and display formats
- how to perform basic data cleansing in Excel.

Lesson Requirements

	PDA	NPA	Data Skills for work
Qualification	Yes	Yes	Yes
Outcome ID(s)	WD7.2c, WD8.3e	DS5.2c, DS5.3c, DS6.2b	C2.1, A1.2, A2.1, A2.3
Outcome description(s)	WD7.2c Data cleaning WD8.3e Data cleaning	DS5.2c Describe methods of cleaning and transforming data DS5.3c Perform routine data cleaning and structuring. DS6.2b Perform data transformation to complete, correct and structure data	C2.1 Vocabulary used in data science and analytics A1.2 Data quality A2.1 Use of tools to analyse data A2.3 Data calculation and manipulation
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	Lesson: PowerPoint/PDF, Worksheet: Excel	Lesson: PowerPoint/PDF, Worksheet: Excel

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.

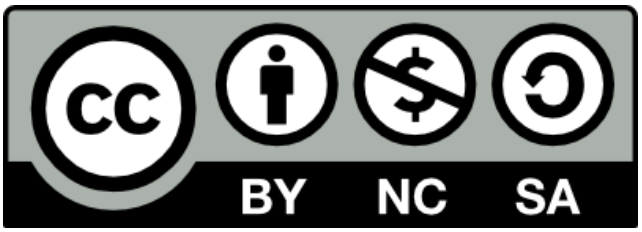
Worksheet

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	Converting data types	Recall	2
1.2	Converting data types	Define	1
1.3	Converting data types	Apply	4
1.4	Converting data types	Active	1
2.1	Fixing strings	Recall	2
2.2	Fixing strings	Apply	5
3.1	Fixing missing and outlying values	Recall	2
3.2	Fixing missing and outlying values	Apply	3
3.3	Fixing missing and outlying values	Active	1
Total			**21**

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.



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