

# Dataset understanding 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 .....	1
Lesson Description.....	2
Lesson Contents .....	2
Learning Intentions.....	3
Success Criteria .....	3
Knowledge Prerequisites.....	3
Lesson Requirements.....	4
Task-types.....	5
Worksheet .....	6
How you can use this lesson.....	7
Alternative format.....	7

## Version Control

Version number	Purpose/Change	By	Date
1.0	Published by Effini	Emma Nylk	03 Feb 2022

## Lesson Description

<b>Lesson Overview</b>	Data dictionaries/metadata Size and shape Missing values and outliers
<b>Topic</b>	Analysis
<b>Book Chapter(s)</b>	Analysing data

<b>NPA level</b>	4, 5, 6
<b>PDA level</b>	7, 8
<b>Data skills for work level</b>	Core, Analysis

## Lesson Contents

This lesson consists of:

- A lesson plan (this document)
- A PowerPoint/PDF presentation, 'Dataset understanding in Excel'
- Question worksheet (for learners) on 'Dataset understanding in Excel' in Excel
- Answers worksheet (for teachers) on 'Dataset understanding in Excel' in Excel/PDF

## Learning Intentions

We will be learning about the data understanding part of the analysis process, specifically,

- what is metadata and the importance of a data dictionary
- how to identify the shape and size of a dataset in Excel and data types of variables
- how to identify missing values and outliers in Excel

## Success Criteria

I can *describe* what metadata is and how it can be used.

I can *describe* what is a data dictionary and how it can be used.

I can *describe* the size and format of datasets in Excel.

I can *identify* missing values and outliers of a dataset using `min()`, `max()` and `unique()` functions in Excel.

## 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

## Lesson Requirements

	<b>PDA</b>	<b>NPA</b>	<b>Data Skills for work</b>
<b>Qualification</b>	Yes	Yes	Yes
<b>Outcome ID(s)</b>	CD7.1c, CD7.1f, WD8.1e, WD8.1f	DC4.2b, DC5.2b, DC6.2b	A1.2, A1.3, C2.1
<b>Outcome description(s)</b>	CD7.1c Types of data CD7.1f Data quality WD8.1e Data quality WD8.1f Stages in the data analysis process	DS4.2b Describe how data can be analysed, DC5.2b Explain how data can be analysed, DC6.2b Explain how data can be analysed	A1.2 Data quality A1.3 Interpretation and insight C2.1 Vocabulary used in data science and analytics
<b>Level</b>	7, 8	4, 5, 6	Core, Analysis
<b>Software language</b>	N/A	N/A	N/A
<b>Required equipment /software for student</b>	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
<b>1. Recall</b>	To be able to recognise definitions or procedures.
<b>2. Define</b>	To be able to define definitions or procedures.
<b>3. Rephrase</b>	To be able to use their own words to describe definitions or procedures.
<b>4. Apply</b>	To be able to apply definitions or procedures to problem-solving activities.
<b>5. Create</b>	To be able to apply definitions or procedures and create their own solutions to a defined problem.
<b>6. Active</b>	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	Metadata and data dictionaries	Recall	2
1.2	Metadata and data dictionaries	Define	3
1.3	Metadata and data dictionaries	Apply	2
1.4	Metadata and data dictionaries	Active	1
2.1	Shape, size and format	Recall	1
2.2	Shape, size and format	Apply	5
3.1	Identifying outliers	Recall	1
3.2	Identifying outliers	Rephrase	2
3.3	Identifying outliers	Apply	3
4.1	Identifying missing values	Recall	1
4.2	Identifying missing values	Apply	5
5.1	Extension	Apply	1
<b>Total</b>			<b>27</b>

## 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](https://creativecommons.org/licenses/by-nc-sa/4.0/).



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**