**Importance of Data Quality**

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](#_Toc115353167)

[Lesson Description 2](#_Toc115353168)

[Lesson Contents 2](#_Toc115353169)

[Learning Intentions 3](#_Toc115353170)

[Success Criteria 3](#_Toc115353171)

[Knowledge Prerequisites 3](#_Toc115353172)

[Lesson Requirements 4](#_Toc115353173)

[Task-types 5](#_Toc115353174)

[Worksheet 6](#_Toc115353175)

[How you can use this lesson 7](#_Toc115353176)

[Alternative format 7](#_Toc115353177)

# Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Version number | Purpose/Change | By | Date |
| 1.0 | Published by effini | Emma Nylk | 04 Oct 2022 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Lesson Description

|  |  |
| --- | --- |
| **Lesson Overview** | What is high quality data and why its important  How to assess and improve the quality of a dataset |
| **Topic** | Quality and management |
| **Book Chapter(s)** | Quality and management |

|  |  |
| --- | --- |
| **NPA level** | 4, 5, 6 |
| **PDA level** | 7, 8 |
| **Data skills for work level** | Data management, Core, Analysis |

# Lesson Contents

This lesson consists of:

* A lesson plan (this document)
* A PowerPoint/PDF presentation, ‘Importance of data quality’
* Question worksheet (for learners) on ‘Importance of data quality’ in Excel/PDF
* Answers worksheet (for teachers) on ‘Importance of data quality’ in Excel/PDF

# Learning Intentions

We will be looking at the **quality of data**, specifically

* Why it is important to have high quality data
* How to assess the quality of a dataset
* How to improve the quality of a dataset

# Success Criteria

I can *explain* what is meant by high quality data

I can *evaluate* the quality of a dataset

I can *describe* how to improve the quality of a dataset

# 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

# Lesson Requirements

|  |  |  |  |
| --- | --- | --- | --- |
|  | **PDA** | **NPA** | **Data Skills for work** |
| **Qualification** | Yes | Yes | Yes |
| **Outcome ID(s)** | WD7.1d, CD7.1f, CD7.2f,  WD8.2f, CD8.1f | DC4.2d  DC5.2a  DC6.3b | C2.1, A1.2, DM3.1 |
| **Outcome description(s)** | WD7.1d – Data quality including bias  CD7.1f – Data quality  CD7.2f – Data bias and ways to reduce bias  WD8.2f – Data quality including data bias  CD8.1f -  Data quality and data bias | DC4.2d,  DC5.2a,  DC6.3b – Evaluate a dataset in terms of its quality including potential bias | C2.1 – Vocabulary used in data science and analytics  A1.2 – Data quality  DM3.1 – Data standards |
| **Level** | 7, 8 | 4, 5, 6 | Data management, 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 | Importance of high quality data | Recall | 2 |
| 1.2 | Importance of high quality data | Apply | 3 |
| 1.3 | Importance of high quality data | Active | 1 |
| 2.1 | Identifying high quality data | Recall | 1 |
| 2.2 | Identifying high quality data | Apply | 3 |
| 2.3 | Identifying high quality data | Rephase | 1 |
| 3.1 | Improving the quality of data | Apply | 3 |
| **Total** | | | **14** |

# How you can use this lesson

This lesson has been created by effini in partnership with Data Education in Schools and Skills Development Scotland.

© 2022. This work is licensed under a [*CC BY-NC-SA 4.0 license*](https://creativecommons.org/licenses/by-nc/4.0/legalcode)*.*

A picture containing text, clipart

Description automatically generated

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**