Data cleaning - Lesson overview

Learning objectives:

- Understand the concept of data cleaning and how it contributes to accurate and reliable data analysis.

- Know about common data issues that can arise and the data cleaning techniques that can be used to address them.

- Understand how to practically implement these data cleaning techniques in Google Sheets.

- Understand the role that data integrity plays in maintaining the quality of data.

- Know how we can implement data integrity practices in Google Sheets.

Slide deck **Knowledge questions**

Walk-through **Reference card**

5 min — Find and remove outliers

25 min / 5 min — Practical data integrity

6 min — Find and remove duplicate observations

4 min — Irrelevant observations

7 min — NaNs and Nulls

6 min — Imputation

15 min / 5 min — Cleaning techniques

Data quality checklist

Description

Data preparation efforts and approaches are crucial and **significantly impact** the outcomes of our **analysis**, sometimes even more than the analysis methods themselves**.** Keen attention needs to be applied to the data cleaning process to ensure that only **high-quality data** are utilised **for effective analysis**. Moreover, it is equally important to maintain data quality consistently throughout its entire lifecycle. In this lesson, we are introduced to the concepts of **data cleaning** and **data integrity** and the crucial role they play in **enhancing** and **maintaining data quality**. We will also see how we can implement these concepts practically in Google Sheets.