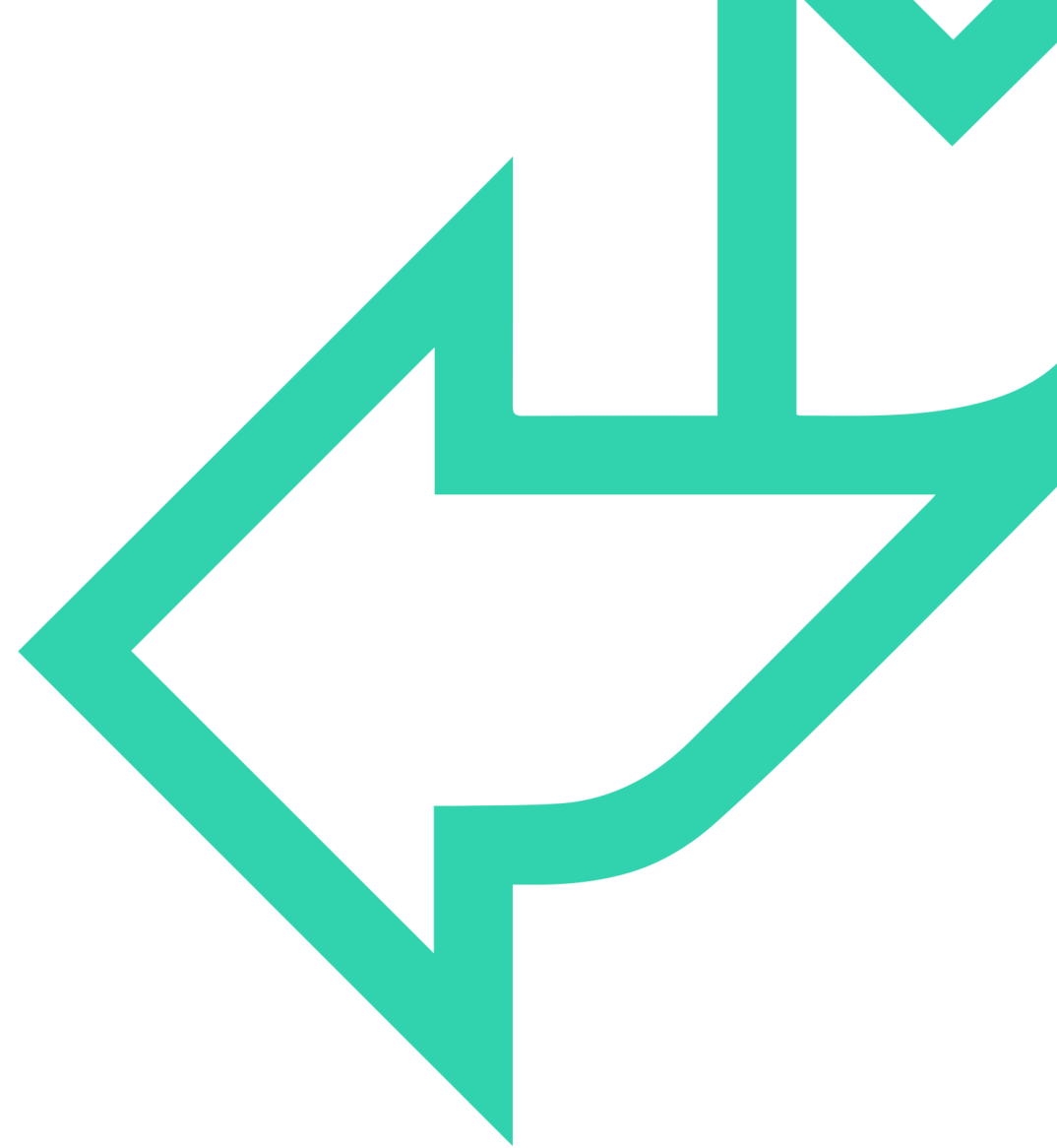




Data Essentials

L3

Module 2: From Data to Insight





KSBS

Knowledge (1 of 2)

K3₁ How to collate and format data in line with industry standards.

K4₂ Data formats and their importance for analysis. Management and presentation tools to visualise and review the characteristics of data. Communication tools and technologies for collaborative working.

K6₁₊₂ The value of data to the business. How to undertake blending of data from multiple sources.

Subscript = Assessment method

- 1: 'Scenario Demonstrations with Questioning'
- 2: 'Professional Discussion Underpinned by a Portfolio'



KSBS

Knowledge (2 of 2)

K8₁ How to filter details, focusing on information relevant to the data project.

K9₁ Basic statistical methods and simple data modelling to extract relevant data and normalise unstructured data.

K10₂ The range of common data quality issues that can arise e.g. misclassification, duplicate entries, spelling errors, obsolete data, compliance issues and interpretation/ translation of meaning.

Subscript = Assessment method

1: 'Scenario Demonstrations with Questioning'

2: 'Professional Discussion Underpinned by a Portfolio'



KSBS

Skills (1 of 2)

S2₁ Collect, format, and save datasets.

S3₁ Summarise and explain gathered data.

S4₁ Blend data sets from multiple sources and present in format appropriate to the task.

S5₁ Manipulate and link different data sets as required.

Subscript = Assessment method

- 1: 'Scenario Demonstrations with Questioning'
- 2: 'Professional Discussion Underpinned by a Portfolio'



KSBS

Skills (2 of 2)

S6₁ Use tools and techniques to identify trends and patterns in data.

S10₂ Demonstrate the different ways of communicating meaning from data in line with audience requirements.

S13₂ Explain data and results to different audiences in a way that aids understanding.

S17₂ Operate as part of a multi-functional team.

Subscript = Assessment method

1: 'Scenario Demonstrations with Questioning'

2: 'Professional Discussion Underpinned by a Portfolio'



KSBS

Distinction criteria

D2₁ Justifies why we undertake crosschecking of data.

D7₂ Evaluates why we need to store, manage, and distribute data, and justifies the importance of maintaining ethical and security standards.

Subscript = Assessment method

- 1: 'Scenario Demonstrations with Questioning'
- 2: 'Professional Discussion Underpinned by a Portfolio'



SCHEDULE

DAY 1

AM

- ⇒⇒ Types of data file
- ⇒⇒ Formatting data

PM

- ⇒⇒ Filtering data
- ⇒⇒ Formulas and functions





SCHEDULE

DAY 2

AM

⇒ Lookup functions

⇒ Joining tables

PM

⇒ Joining tables (cont'd)





SCHEDULE

DAY 3

⇒ Power Query





Day 1





Types of data file



TYPES OF DATA FILE

KSBs

K4₂ Data formats and their importance for analysis.
Management and presentation tools to visualise and review the characteristics of data. Communication tools and technologies for collaborative working.





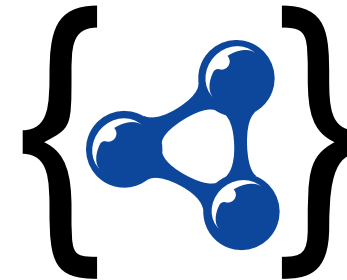
COMMON TYPES OF DATA FILE



CSV



XLSX



JSON



CSV FORMAT

CSV: Comma-separated values

A screenshot of a text editor window titled 'qa_diy_sales.csv'. The window contains the following text:

```
Product,Unit price,Qty sold
Long Weight,£40.00,10
Elbow Grease,£8.00,5
Tartan Paint,£26.00,2
Left-Handed Broom,£11.40,3
```

Below the text, there are three blue curly braces, each spanning one of the three columns of data. Under each brace is the word 'value' in a blue, handwritten-style font. Three orange arrows point from the word 'commas separate values' (written in orange cursive below the editor) to the commas in the first three rows of data. A blue arrow points from the last row of data to the third 'value' label.

Ln 5, Col 26 | 220% | Windows (CRLF) | UTF-8

commas separate values



KNOWLEDGE CHECK: CSV FILES

CSV: Comma-separated values

Question 1

What type of file is a CSV file?

- A: Image file
- B: Plain-text file
- C: Zip archive
- D: Executable file
- ✓: Excel file (same as XLSX)
- ✗: It's a trick question



KNOWLEDGE CHECK: CSV FILES

CSV: Comma-separated values

Question 2

What is the usual delimiter (separator) character in a CSV file?

- A: Tab
- B: Colon :
- C: Space
- D: Comma ,
- ✓: Double-quote '
- ✗: Pipe |



KNOWLEDGE CHECK: CSV FILES

CSV: Comma-separated values

Question 3

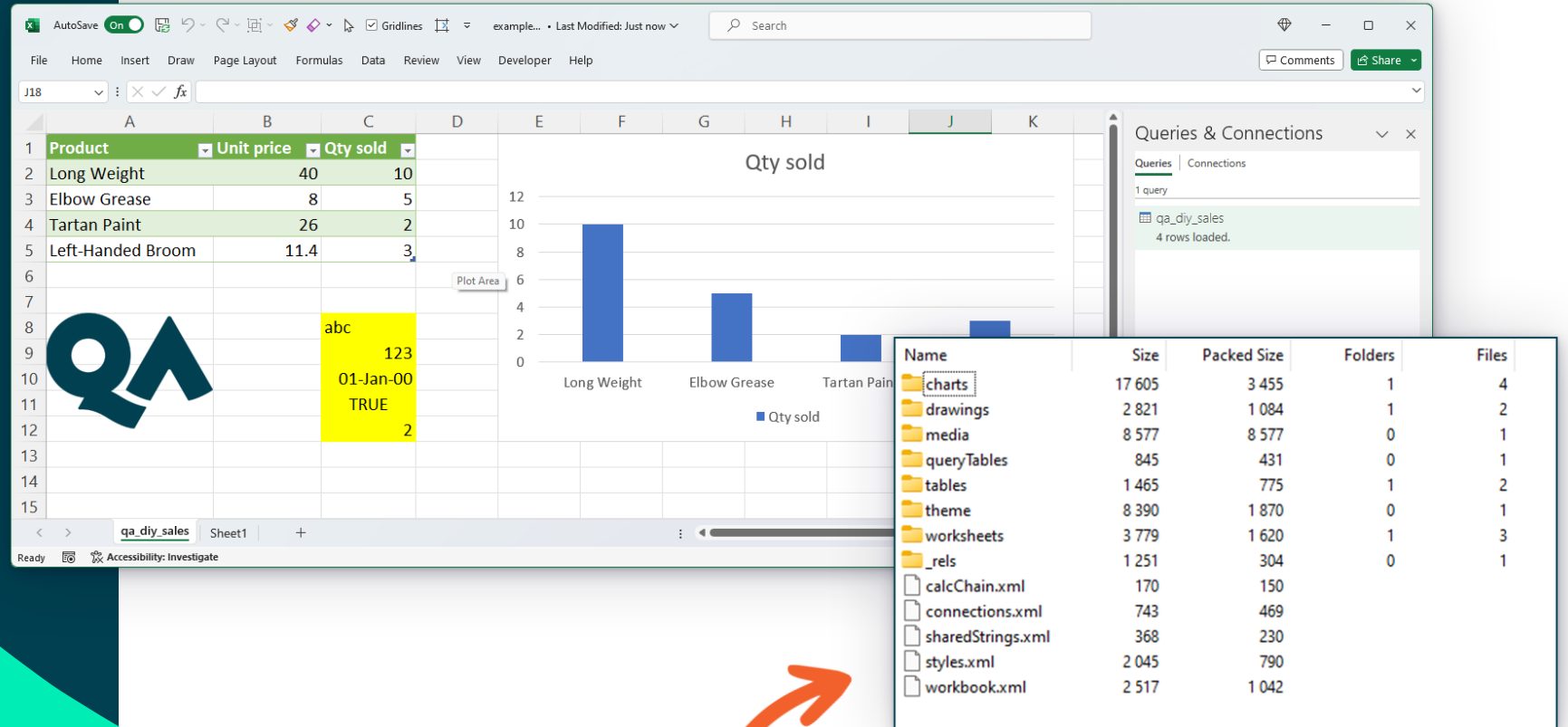
If we open a CSV file directly in Excel, what potentially unwanted transformations(s) could occur?

- A: Number rounding
- B: Date-like strings converted to dates
- C: Corrupted text due to character encoding, e.g., Â£ instead of £
- D: All of the above



XLSX FORMAT

XLSX: Open Office XML



Internally, an XLSX file is mostly a collection of XML files.



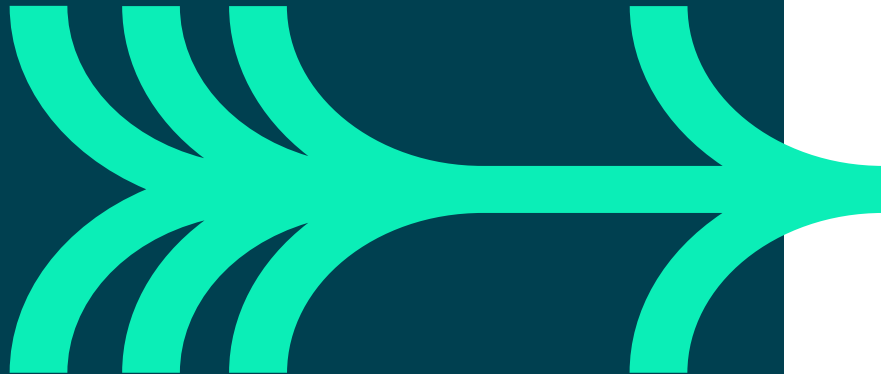
KNOWLEDGE CHECK: XLSX FILES

XLSX: Open Office XML

Question 1

Which Microsoft Office application natively supports the XLSX format?

- A: Excel
- B: PowerPoint
- C: Word
- D: All of the above





KNOWLEDGE CHECK: XLSX FILES

XLSX: Open Office XML

Question 2

Consider the XLSX, PPTX, and DOCX file formats. Of these, the XLSX format is the most suitable for storing what type of content?

- A: Documents
- B: Slideshows
- C: Spreadsheets
- D: None of the above



KNOWLEDGE CHECK: XLSX FILES

XLSX: Open Office XML

Question 3

Which of these types of content cannot be stored in an **XLSX** file but can be stored in an **XLSM** file?

A: Music

B: Mathematical formulas

C: Macros

D: Movies



JSON FORMAT

JSON: JavaScript Object Notation

```
{
  "string": "Text goes in double-quotes",
  "string_2": "Whitespace is ignored",
  "number": 42,

  "list_of_strings": ["a", "b", "c"],
  "list_of_numbers": [1, 123.45, 0.00123],
  "mixed_list": [1, 2, "a", "b"],
  "nested_lists": [1, 2, ["a", "b"], 3],

  "dictionary": {
    "another_string": "Commas separate everything",
    "another_number": 3.14,
    "nothing": null,
    "qa_is_the_best": true,
    "dict keys need underscores": false,
    "empty_dict": {},
    "empty_list": [],
    "even_this_is_fine": [[[[[{}]]]]],
    "super_nesting": [
      ["this", "is", "a", "list"],
      ["of lists"],
      ["inside a dictionary"],
      ["inside", "another", "dictionary"]
    ]
  }
}
```



KNOWLEDGE CHECK: JSON FILES

JSON: JavaScript Object Notation

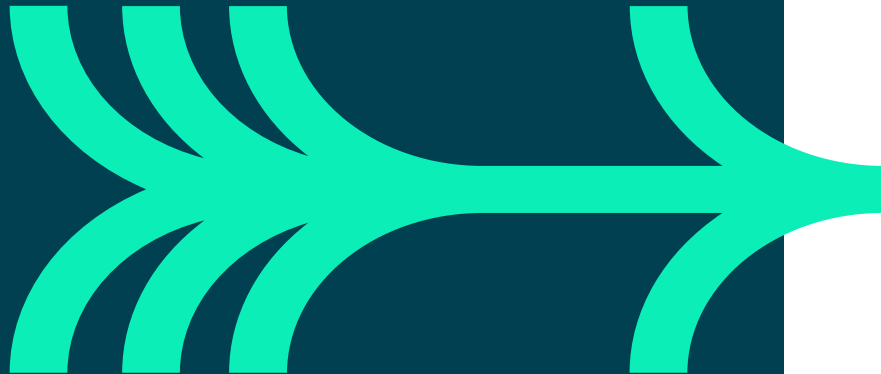
Question 1

What type of file is a JSON file?

- A: Image file
- B: Plain-text file
- C: Zip archive
- D: Executable JavaScript file



KNOWLEDGE CHECK: JSON FILES



JSON: JavaScript Object Notation

Question 2

What types of **collection** are supported by the JSON standard?

- A: Dictionaries and lists
- B: Thesauruses and sets
- C: Tables and cells
- D: Files and folders



KNOWLEDGE CHECK: JSON FILES

JSON: JavaScript Object Notation

Question 2

How can collections be nested in a JSON file?

A: Lists can go inside lists

B: Dictionaries can go inside dictionaries

C: Lists can go inside dictionaries

D: Dictionaries can go inside lists

✓: A and B only

✗: A, B, C, and D



Formatting data



FORMATTING DATA

KSBs

K3₁ How to collate and format data in line with industry standards.

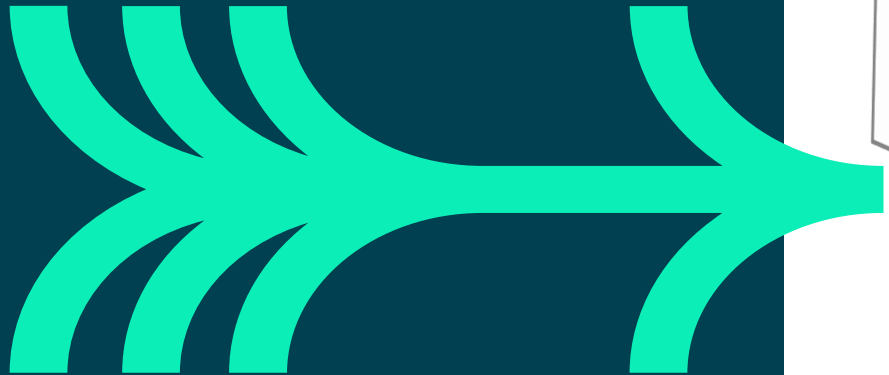
K4₂ Data formats and their importance for analysis. Management and presentation tools to visualise and review the characteristics of data. Communication tools and technologies for collaborative working.

K10₂ The range of common data quality issues that can arise e.g. misclassification, duplicate entries, spelling errors, obsolete data, compliance issues and interpretation / translation of meaning.

S6₁ Use tools and techniques to identify trends and patterns in data.



EXCEL NUMBER FORMATS



What we see

What Excel uses for calculations

General	12345.6
Currency	£12,345.60
Scientific	£1.23E+04
Date	18/10/1933



CONDITIONAL FORMATTING

Data
<div><div></div></div> 8
<div><div></div></div> 7
<div><div></div></div> 5
<div><div></div></div> 6
<div><div></div></div> 7
<div><div></div></div> 18
<div><div></div></div> 7
<div><div></div></div> 3
<div><div></div></div> 3
<div><div></div></div> 5

Data
Unique 1
Duplicate 1
Unique 2
Duplicate 1
Duplicate 2
Unique 3
Duplicate 2
Duplicate 1
Unique 4
Unique 5

Data
Complete
Pending
Pending
Error
Complete
Complete
Pending
Complete
Pending
Error

Data
★ 5.0
★ 2.5
★ 5.0
★ 1.0
★ 3.0
★ 0.5
★ 4.0
★ 3.0
★ 4.5
★ 3.0



Filtering data



FILTERING DATA

KSBs

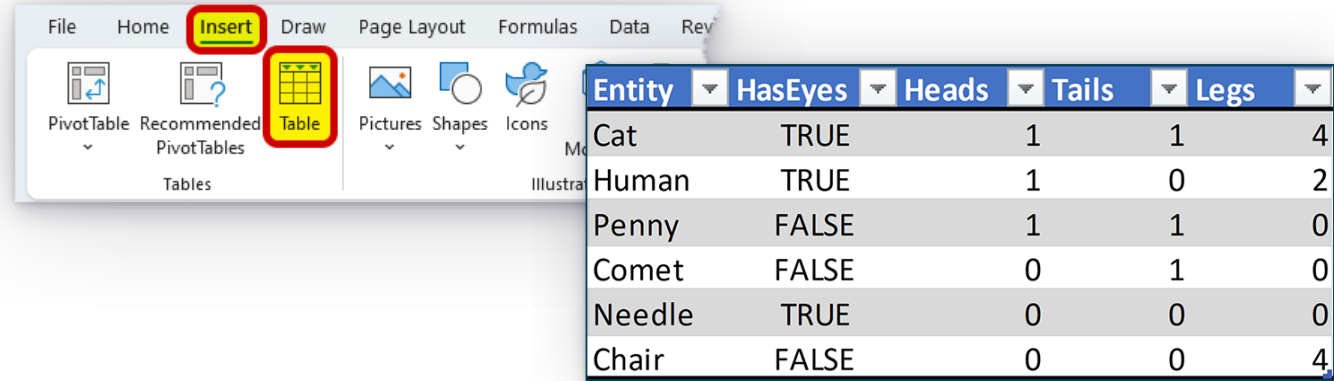
K3₁ How to collate and format data in line with industry standards.

K8₁ How to filter details, focusing on information relevant to the data project.

S2₁ Collect, format, and save datasets.



EXCEL TABLES



Entity	HasEyes	Heads	Tails	Legs
Cat	TRUE	1	1	4
Human	TRUE	1	0	2
Penny	FALSE	1	1	0
Comet	FALSE	0	1	0
Needle	TRUE	0	0	0
Chair	FALSE	0	0	4

In Excel, the **Table** is a convenient container for structured data.

Tables offer many useful functions, including:

- Sorting and **filtering**.
- Excel formula integration.
Tables and Table columns can be referenced by name.
- Dynamic referencing.
Formula references to Table columns always point to the entire column, even when the Table changes size.



FILTERING IN EXCEL

Open file

'Learner Guide - Basic Filtering.pdf'

In this activity, we will learn how to apply various types of filter to an Excel Table, including:

- text and numeric filters.
- multi-column filters.
- stacking filters via AND.





FILTERING IN EXCEL

Open file

'Learner Guide - Advanced Filtering.pdf' **OPTIONAL**

In this activity, we will learn how to work with Excel's advanced filter commands, including:

- defining the criteria range.
- specifying criteria.
- defining an advanced filter.
- applying Unique, Sort, and Filter functions.





FORMULAS AND FUNCTIONS IN EXCEL



Open file

'Learner Guide - Formulas and Functions.pdf'

In this activity, we will learn how to work with Excel's formulas and functions, including:

- Data input.
- Sorting data.
- IF, COUNTIF, SUM IF.
- AND, OR.
- Conditional formatting.
- VLOOKUP.
- SUM, MAX, MIN.



POWER QUERY ACCESS

Check if you have the Power Query feature in your Excel.

