Where data lives

Three shapes, six sources and one dataset of your own Names _____

You need a browser and a text editor. No account, no password. Part A is filled in with the whole class; Part B in pairs. Write here.

Reference

Shape	What it means	Example
Structured	Fixed rows and columns, every row the same shape.	A database table, a clean CSV.
Semi-structured	Labelled, but the shape can vary.	JSON, XML.
Unstructured	No data model at all.	Photos, audio, a PDF of a letter, a chat.

Part A · Six sources (with the class)

For each source: its shape, and what you need to get in.

Source	Everyday example	Shape	What you need to reach it
Relational DB	The school's student records; a shop's orders.		
CSV	An export from a bank app; last week's quartet.		
Excel	The spreadsheet your boss emails you.		
JSON	What a weather app downloads.		
API	Spotify, a bus timetable.		
Open-data portal	GOIB portal, TidyTuesday.		

A1 Which sources ask for credentials?

Always _____ Usually _____ Never _____

A2 "CSV is Excel." Open one of each in a text editor. What do you see?

CSV _____ `.xlsx` _____

Part B · The hunt (14 min)

Choose one portal and find **one dataset that interests you**: tourism, housing, beaches, buses, music, games...

- **GOIB open-data portal**: `dadesobertes.caib.es`
- **TidyTuesday**: `github.com/rfordatascience/tidytuesday`, a new public dataset every week since 2018. Look in `data/<year>/<date>/`, open the `.csv` file and press **Download raw file**. Some picks:

Topic	Folder	File
Spotify songs	`2020/2020-01-21`	`spotify_songs.csv`: popularity, danceability, energy, tempo...
Taylor Swift	`2023/2023-10-17`	`taylor_album_songs.csv`: every album track with its audio features.
Pokémon	`2025/2025-04-01`	`pokemon_df.csv`: types, stats, height, weight.
Steam games	`2021/2021-03-16`	`games.csv`: monthly average and peak players per game.

Download it and open it in a **text editor**, not in a spreadsheet: you have to see what separates the values.

Field	Your answer
Title and web address	
Formats offered (CSV, JSON, XLSX...)	
Shape	
Separator and decimal mark	
Rows × columns, roughly	
Licence	
One question it could answer	

B1 Compare with a pair that used the other portal. Do your files separate values and write decimals the same way?

B2 What would you need to get your dataset into a tool like Superset? Does it ask for credentials?

Done when The table of Part A is full · the Part B card is complete · your question is one a chart could answer.
Fast pair? Open `pokemon_df.csv` in the text editor: how many columns does the header have, how many lines does the file have, and what is written in the `type_2` of `charmander`?

At home A gym wants a dashboard of how many people train each hour. The data is in their booking app's MySQL database and in a weekly Excel file from reception. One paragraph: for each source, its shape and what you need to get the data out.