Dónde viven los datos

Tres formas, seis fuentes y un conjunto de datos vuestro

Nombres _____

Necesitáis un navegador y un editor de texto. Ninguna cuenta, ninguna contraseña. La Parte A se rellena con toda la clase; la Parte B, por parejas. Escribid aquí.

Referencia

Forma	Qué significa	Ejemplo
Estructurados	Filas y columnas fijas, todas las filas con la misma forma.	Una tabla de una base de datos, un CSV limpio.
Semiestructurados	Etiquetados, pero la forma puede variar.	JSON, XML.
No estructurados	Sin ningún modelo de datos.	Fotos, audio, el PDF de una carta, un chat.

Parte A · Seis fuentes (con la clase)

Para cada fuente: su forma y qué necesitáis para entrar.

Fuente	Ejemplo cotidiano	Forma	Qué necesitáis para llegar a ella
BD relacional	Los expedientes del instituto; los pedidos de una tienda.		
CSV	Una exportación de la app del banco; el cuarteto de la semana pasada.		
Excel	La hoja de cálculo que te envía el jefe por correo.		
JSON	Lo que se descarga una app del tiempo.		
API	Spotify, el horario de un bus.		
Portal de datos abiertos	Portal del GOIB, TidyTuesday.		

A1 ¿Qué fuentes piden credenciales?

Siempre _____ Normalmente _____ Nunca _____

A2 «Un CSV es un Excel.» Abrid uno de cada en un editor de texto. ¿Qué veis?

CSV _____ `.xlsx` _____

Parte B · La caza (14 min)

Elegid un portal y encontrad **un conjunto de datos que os interese**: turismo, vivienda, playas, buses, música, juegos…

- **Portal de datos abiertos del GOIB**: `dadesobertes.caib.es`
- **TidyTuesday**: `github.com/rfordatascience/tidytuesday`, un conjunto de datos público nuevo cada semana desde 2018. Mirad en `data/<year>/<date>/`, abrid el archivo `.csv` y pulsad **Download raw file**. Algunas propuestas:

Tema	Carpeta	Archivo
Canciones de Spotify	`2020/2020-01-21`	`spotify_songs.csv`: popularidad, bailabilidad, energía, tempo…
Taylor Swift	`2023/2023-10-17`	`taylor_album_songs.csv`: todas las canciones de los álbumes con sus características de audio.
Pokémon	`2025/2025-04-01`	`pokemon_df.csv`: tipos, estadísticas, altura, peso.
Juegos de Steam	`2021/2021-03-16`	`games.csv`: media y pico mensual de jugadores por juego.

Descargadlo y abridlo en un **editor de texto**, no en una hoja de cálculo: tenéis que ver qué separa los valores.

Campo	Vuestra respuesta
Título y dirección web	
Formatos ofrecidos (CSV, JSON, XLSX…)	
Forma	
Separador y marca decimal	
Filas × columnas, aproximadamente	
Licencia	
Una pregunta que podría responder	

B1 Comparad con una pareja que haya usado el otro portal. ¿Vuestros archivos separan los valores y escriben los decimales de la misma manera?

B2 ¿Qué necesitaríais para meter vuestro conjunto de datos en una herramienta como Superset? ¿Pide credenciales?

Habéis acabado cuando La tabla de la Parte A está llena · la ficha de la Parte B está completa · vuestra pregunta es una que un gráfico podría responder. **¿Pareja rápida?** Abrid `pokemon_df.csv` en el editor de texto: ¿cuántas columnas tiene la cabecera, cuántas líneas tiene el archivo y qué hay escrito en el `type_2` de `charmander`?

En casa Un gimnasio quiere un cuadro de mando de cuánta gente entrena cada hora. Los datos están en la base de datos MySQL de su app de reservas y en un archivo Excel semanal de recepción. Un párrafo: para cada fuente, su forma y qué necesitáis para sacar los datos.