El informe descriptivo

La página de la tienda de `games_clean` Nombres _____

Superset → **SQL → SQL Lab**, base de datos `Dades classe`, esquema `public`, tabla `games_clean`. Escribid el informe en un documento (plantilla del profesor): seis partes con su encabezado, **250 palabras como máximo**. Por parejas: uno ejecuta consultas, el otro escribe; cambiad en la parte 4. Esta hoja es para vuestras notas y para la revisión.

Referencia · las seis partes

Un informe descriptivo es la página de la tienda de un conjunto de datos: lo leéis antes de «instalar» los datos en un gráfico.

Parte	Pregunta a la que responde	En la página de la tienda
1 · Fuente	¿Quién los recogió, dónde, cuándo? ¿Quién los limpió? De una API: endpoint, parámetros, fecha de extracción.	Desarrollador y fecha de lanzamiento.
2 · Tamaño	¿Cuántas filas y columnas? ¿Qué es una fila? ¿Qué periodo?	Tamaño de la descarga.
3 · Variables	¿Cuáles son las columnas importantes, su tipo y su unidad?	Lista de características.
4 · Calidad	¿Qué falta, qué está estimado o es dudoso? ¿Cuánto?	Problemas conocidos.
5 · Cifras clave	Centro, dispersión y distribución de lo que importa.	Puntuación de las reseñas y etiquetas.
6 · Límites	¿Qué **no** pueden deciros estos datos?	Requisitos del sistema: dónde no funcionará.

Tarea 0 · Dos informes, los mismos datos (4 min)

Informe A	Informe B
«El conjunto de datos tiene muchos juegos. La mayoría son baratos. El precio medio es 7.91. Faltan algunas puntuaciones. Los juegos son cada vez más populares.»	«26 685 juegos de Steam lanzados 2004–2018. La mitad cuesta 4.99 USD o menos (mediana); la media es 7.91 USD por unos pocos productos de hasta 595.99 USD. Solo el 10.7% tiene Metascore.»

0.1 ¿De cuál os fiaríais? Subrayad las palabras exactas que marcan la diferencia y después decid tres cosas que tiene y que al otro le faltan.

0.2 Una frase del otro informe dice algo que los datos no pueden demostrar. ¿Cuál, y por qué?

Tarea 1 · Partes 1 a 3 (7 min)

La fuente y las variables salen de vuestros apuntes de la UT2. El tamaño necesita una consulta:

```sql
SELECT COUNT(*) AS n_games, MIN(release_date) AS first_release,
       MAX(release_date) AS last_release
FROM games_clean;
```

Columnas: _____ Primera fecha: _____ Última fecha: _____ Una fila es: _____

Tarea 2 · Partes 4 y 5 (11 min)

Reutilizad las consultas del miércoles pasado (la ficha estadística). Patrón de una distribución con porcentajes:

```sql
SELECT price_band, COUNT(*) AS n,
       ROUND(100.0 * COUNT(*) / SUM(COUNT(*)) OVER (), 1) AS pct
FROM games_clean GROUP BY price_band ORDER BY n DESC;
```

El mínimo, cada cosa con su **unidad** (USD, juegos, %):

2.1 Una cifra de datos ausentes con su porcentaje: _____

2.2 Centro y dispersión de `price`: _____

¿Qué centro ponéis en el informe, y por qué? _____

2.3 Una distribución (`price_band` u `owners_tier`): _____

Tarea 3 · Parte 6, sin ninguna consulta (5 min)

¿Qué concluiría mal un periodista de esta tabla? Pensad en `owners_min`, `metascore` y `price`. Dos límites:

3.1 Estos datos no pueden decir... _____

3.2 Cuidado: _____

Revisión entre parejas (11 min)

Dad esta hoja y vuestro informe a otra pareja. **Revisores**: marcad sí o no. Cada «No» necesita la línea a la que se refiere. Sin notas y sin adjetivos: solo sí, no y dónde. Revisado por _____

#	Comprobación	Sí / No	Si es no: ¿qué línea?
1	Están las seis partes, en orden, cada una con su encabezado.		
2	La fuente dice quién recogió los datos, de dónde y qué años cubren. Para una API: endpoint, parámetros y fecha de extracción.		
3	El tamaño da filas **y** columnas, y dice qué es una fila.		
4	Cada cifra tiene una unidad y su consulta está adjunta.		
5	Para `price`, el centro es la mediana (o ambos), y el texto dice por qué.		
6	Calidad: al menos dos problemas, cada uno con recuento o porcentaje.		
7	Los límites tienen al menos dos cosas que los datos no pueden decir.		
8	Ninguna afirmación va más allá de los datos. 250 palabras o menos.		

R **Autores**, cuando vuelva la hoja: corregid **un** «No». ¿Qué habéis cambiado?

_____

Habéis acabado cuando Seis partes con encabezado, 250 palabras o menos · cada cifra tiene unidad y su SQL bajo el informe · la otra pareja ha rellenado la lista · vuestro cambio está escrito aquí · informe entregado.

¿Atascados?

- **`relation "games_clean" does not exist`**: pedid `games_clean.sql` al profesor y ejecutadlo, o juntaos.
- **`function round(double precision, integer) does not exist`**: convertid: `ROUND(metascore::numeric, 1)`.
- **`SUM(COUNT(*)) OVER ()` da un error**: falta el `GROUP BY`.
- **¿Seguís con las consultas a mitad del tiempo?** Dejad de consultar, empezad a escribir.

En casa Acabad el informe si hace falta y entregadlo. Para el miércoles: releed vuestros apuntes sobre `COALESCE`, `CASE WHEN`, `CREATE TABLE AS` y las seis partes. Abrid la tabla `pokemon` y mirad `type_2` y `generation_id`.