La ficha estadística

Centro, dispersión y forma de `games_clean` Nombres _____

Superset → **SQL → SQL Lab**, base de datos `Dades classe`, esquema `public`, tabla `games_clean` (26 685 juegos). Por parejas: uno teclea, el otro escribe; cambiad en la Tarea 3. Escribid los números aquí.

Referencia · la cola

Poned los datos en cola, el menor primero. Cinco amigos ganan 1 200, 1 300, 1 400, 1 500 y 30 000: la **media** (sumar todo y repartir a partes iguales) es 7 080; la **mediana** (el que está en el medio de la cola) es 1 400.

Pregunta	En la cola	SQL
Centro	¿Dónde está el medio?	`AVG(x)` · `PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY x)`
Dispersión	¿A qué distancia están el primero y el último? ¿Y la mitad central?	`MIN(x)` · `MAX(x)` · `STDDEV(x)` · percentiles 0.25 y 0.75, los **cuartiles** Q1 y Q3
Forma	¿Dónde se amontona la cola?	`width_bucket(x, from, to, n)` + `COUNT(*)`

Leed `WITHIN GROUP (ORDER BY price)` como «haced la cola por precio» y `PERCENTILE_CONT(0.5)` como «caminad hasta el 50% de la cola y leed el valor». La **desviación estándar** es la distancia típica a la media. `width_bucket` dice en cuál de las `n` cajas iguales entre `from` y `to` cae un valor; todo lo que pasa de `to` va a la caja `n + 1`.

Tarea 1 · Centro y extremos de `price` (8 min)

1.1 Antes de ejecutar nada: ¿la media del precio será mayor o menor que la mediana? _____ Después ejecutad:

```sql
SELECT COUNT(*) AS n, MIN(price) AS lowest, MAX(price) AS highest,
       ROUND(AVG(price), 2) AS mean,
       PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY price) AS median
FROM games_clean;
```

`n` _____ `lowest` _____ `highest` _____ `mean` _____ `median` _____ ¿Habéis acertado? _____

1.2 ¿Quién está al final del todo de la cola? Buscad las tres filas más caras (`ORDER BY price DESC LIMIT 3`).

El último de la cola: _____ ¿Es un juego? _____

1.3 Un compañero dice: «La mediana está a medio camino entre el más bajo y el más alto.» Calculad ese número.

A medio camino: _____ ¿Es la mediana? ¿Por qué no? _____

Tarea 2 · Dispersión (7 min)

```sql
SELECT ROUND(STDDEV(price), 2) AS std_dev,
       PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY price) AS q1,
       PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY price) AS q3
FROM games_clean;
```

`std_dev` _____ `q1` _____ `q3` _____ La mitad de los juegos cuesta entre _____ y _____

2.2 Contad los juegos que cuestan más que la media (`WHERE price >` vuestra media).

Juegos por encima de la media: _____ de 26 685. ¿Es la mitad? _____

Tarea 3 · Forma de `price` (9 min)

Seis cajas de 10 de 0 a 60; todo lo que pasa de 60 va a la caja 7. La última columna dibuja la barra: un `#` por cada 500 juegos.

```
SELECT width_bucket(price, 0, 60, 6) AS bucket,
       MIN(price) AS lowest, MAX(price) AS highest, COUNT(*) AS games,
       REPEAT('#', (COUNT(*) / 500)::int) AS bar
FROM games_clean GROUP BY bucket ORDER BY bucket;
```

Copiad la columna `games` en la tabla de la izquierda de más abajo y dibujad las siete barras. Después marcad en el dibujo dónde caen la media (**M**) y la mediana (**Md**).

Tarea 4 · Una segunda columna: `metascore` (9 min)

Repetid las Tareas 1 a 3 para `metascore`, en dos consultas. Cambios: `ROUND(AVG(metascore)::numeric, 1)`; cajas `width_bucket(metascore, 20, 100, 8)`; un `#` por cada 25 juegos.

4.1 Ejecutad la consulta de las cajas tal como está. Sale una fila de más.

Su caja: _____ Sus juegos: _____ ¿Quiénes son? _____

4.2 Añadid `WHERE has_metascore` a las dos consultas y rellenad la ficha y la tabla de la derecha.

n _____ mínimo _____ máximo _____ media _____ mediana _____ desv. est. _____ Q1 _____ Q3 _____

Caja	Juegos	`price`: vuestra barra
1		
2		
3		
4		
5		
6		
7		

Caja	Juegos	`metascore`: vuestra barra
1		
2		
3		
4		
5		
6		
7		
8		

4.3 Mirad los dos dibujos uno al lado del otro.

¿Qué columna tiene media ≈ mediana? _____ ¿Qué forma tiene su dibujo? _____

¿Y la otra? _____

Habéis acabado cuando Todos los huecos están rellenos · los dos dibujos tienen **M** y **Md** marcadas. **¿Pareja rápida?** Añadid `GROUP BY owners_tier` a la consulta de la Tarea 1: ¿los juegos Hit cuestan más? Después, la ficha estadística de una columna numérica de vuestro conjunto de datos abierto.

¿Atascados?

- **`function round(double precision, integer) does not exist`**: convertid primero el tipo, `ROUND(metascore::numeric, 1)`.
- **El `lowest` de una caja no es un número redondo**: es correcto. Es el juego más barato **de dentro** de la caja.
- **`relation "games_clean" does not exist`**: pedid `games_clean.sql` (18/11) al profesor y volved a ejecutarlo.

En casa Acabad la Tarea 4 si hace falta. Guardad esta hoja: el lunes se convierte en la sección de «cifras clave» de vuestro informe.