Franjas y niveles

Común, Raro, Épico: `price_band` y `owners_tier` Nombres _____

Superset → **SQL → SQL Lab**, base de datos `Dades classe`, esquema `public`, tabla `games`, y vuestra consulta guardada `games_step2`. Una categoría es un número que se ha vuelto legible, como la rareza de un objeto. En parejas, uno teclea y el otro escribe; cambiad en la Tarea 3. Ejecutad **una sentencia cada vez**: seleccionadla y pulsad **Run**.

Referencia · las tres reglas de `CASE`

```
CASE WHEN condition_1 THEN 'label 1'
     WHEN condition_2 THEN 'label 2'
     ELSE 'label for the rest'
END AS new_column
```

Regla	Qué significa
De arriba abajo	Cada fila se detiene en la primera rama que es verdadera.
`ELSE` se queda con el resto	Todo lo que no ha encajado en nada. Sin `ELSE`, el resto se convierte en `NULL`.
`NULL` nunca es verdadero	`NULL < 5` es desconocido: un `NULL` no pasa ninguna comparación.

Tarea 1 · `price_band`

Cinco franjas, precios en dólares de EE. UU. Escribid las dos ramas que faltan (`5 to 15` y `15 to 30`). Después contad.

```
SELECT CASE WHEN price IS NULL THEN 'Free'
            WHEN price < 5 THEN 'Under 5'
            -- your two branches here
            ELSE '30 or more' END AS price_band,
       COUNT(*) AS n
FROM games WHERE game IS NOT NULL
GROUP BY price_band ORDER BY MIN(price) NULLS FIRST;
```

Vuestras dos ramas: _____

Free _____ Under 5 _____ 5 to 15 _____ 15 to 30 _____ 30 or more _____ Total _____

¿Por qué `ORDER BY MIN(price)` y no `ORDER BY price_band`? Probadlo. _____

Ejecutad `SELECT price, COUNT(*) FROM games GROUP BY price ORDER BY 2 DESC LIMIT 3;` Los dos precios más comunes: _____ y _____. ¿En qué franja cae cada uno? ¿Por qué el límite es `< 5` y no `<= 5`?

_____

Tarea 2 · Rompedla dos veces

Copiad vuestra consulta y estropeadla, un cambio cada vez. Apuntad qué pasa **y por qué**.

1 Moved la rama `WHEN price < 30` **arriba del todo**, por encima de todas las demás.

Franjas ahora _____ La mayor: _____ con _____ juegos.

¿Adónde han ido a parar los juegos baratos? _____

2 Devolvedla a su sitio. Ahora borrad la línea `WHEN price IS NULL THEN 'Free'`.

`30 or more` tiene ahora _____ juegos.

¿Adónde han ido a parar los juegos gratuitos, y qué regla los ha enviado allí? _____

Tarea 3 · `owners_tier` en la consulta limpia

Abrid `games_step2` (**SQL → Saved Queries**, clic en el nombre) y cambiad el nombre del bloque a `step3`. Añadid dos columnas: vuestra `price_band` y esta. Una columna creada en un `SELECT` no se puede reutilizar en ese mismo `SELECT`, y por eso la expresión de `owners_min` se escribe otra vez entera:

```
CASE WHEN REGEXP_REPLACE(SPLIT_PART(owners, '..', 1), '[^0-9]', '', 'g')::bigint < 20000
       THEN 'Niche'
     WHEN REGEXP_REPLACE(SPLIT_PART(owners, '..', 1), '[^0-9]', '', 'g')::bigint < 1000000
       THEN 'Mid'
     ELSE 'Hit' END AS owners_tier,
```

Comprobadlo con `SELECT owners_tier, COUNT(*) AS n FROM step3 GROUP BY owners_tier ORDER BY MIN(owners_min);`

Niche _____ Mid _____ Hit _____ Después guardadla como `games_step3`: **Save**, cambiad el nombre, **Save as new**.

Tarea 4 · Haced una pregunta con las dos

Las categorías existen para cruzarse. Conservad el bloque `WITH step3` y cambiad el `SELECT` final:

```
SELECT price_band, COUNT(*) AS games,
       SUM(CASE WHEN owners_tier = 'Hit' THEN 1 ELSE 0 END) AS hits,
       ROUND(100.0 * SUM(CASE WHEN owners_tier = 'Hit' THEN 1 ELSE 0 END) / COUNT(*), 1) AS hit_pct
FROM step3 GROUP BY price_band ORDER BY MIN(price) NULLS FIRST;
```

Porcentaje de éxitos más alto: _____ con _____ % Más bajo: _____ con _____ %

¿Demuestra esto que un precio más alto hace un éxito? _____

Tarea 5 · Categorías sucias

Las categorías de texto también se ensucian. ¿Cuántos editores hay?

```
SELECT COUNT(DISTINCT publisher) AS as_written,
       COUNT(DISTINCT LOWER(publisher)) AS ignoring_case FROM games;
SELECT publisher, COUNT(*) FROM games
WHERE LOWER(publisher) IN ('big fish games', 'capcom') GROUP BY publisher ORDER BY 1;
```

`as_written` _____ `ignoring_case` _____ Grafías de más _____

Las cuatro grafías y sus juegos: _____

`CASE WHEN LOWER(publisher) = 'big fish games' THEN 'Big Fish Games' ELSE publisher END` fusiona dos de ellas.

¿Qué hace su `ELSE` con todos los demás editores? _____

Habéis acabado cuando Todos los huecos rellenos · `games_step3` está en las consultas guardadas. **¿Pareja rápida?** Cortad una columna numérica de vuestro conjunto de datos abierto en tres o cuatro franjas.

¿Atascados?

- **`We can't seem to resolve the column "price_band"`**: una columna nueva solo se puede usar un nivel más arriba, en el `SELECT` que lee `step3`.
- **`Error parsing near 'CASE'`**: falta la coma de delante.
- **Seis franjas en lugar de cinco**: una errata en una etiqueta. `'5 to 15 '` con un espacio es otra categoría.
- **`games_step2` no está en Saved Queries**: pedid su SQL al profesor.

En casa Diseñad tres o cuatro franjas para una columna numérica de vuestro conjunto de datos abierto. Escribid el `CASE`, los recuentos y una frase que defienda dónde habéis puesto los límites.