El registre d'imputació

Què escrivim a les cel·les buides de `games`? Noms _____

Superset → **SQL → SQL Lab**, base de dades `Dades classe`, esquema `public`, taula `games`. Una cel·la buida no es pot netejar: només podeu decidir què en feu. Això és la **imputació**. En parelles, un tecleja i l'altre escriu; canvieu a la Tasca 3. Avui no es modifica res: tot són `SELECT`. Executeu **una sentència cada vegada**: seleccioneu-la i **Run**.

Referència · el quadern del professor

Opció	Al quadern	És justa quan…
Una constant	Escriure un 0.	**Sabeu** què vol dir el buit: el treball no es va lliurar.
Un valor central	Escriure la mitjana de la classe (o la mediana).	Hi ha pocs buits, i els qui els tenen no tenen res d'especial.
Deixar-ho buit + indicador	No escriure res, afegir una nota "no avaluat".	No ho sabeu, o la major part de la columna és buida.

`COUNT(*)` compta files · `COUNT(col)` compta les files on `col` **no** és buida · `COALESCE(a, b)` retorna el primer argument que no és `NULL`: "`a`, i si és buit, `b`".

Tasca 1 · Mesureu el forat

```sql
SELECT COUNT(*)                 AS total_rows,
       COUNT(price)             AS with_value,
       COUNT(*) - COUNT(price)  AS missing,
       ROUND(100.0 * (COUNT(*) - COUNT(price)) / COUNT(*), 1) AS pct_missing
FROM games;
```

Repetiu-la amb les altres tres. És el vostre **registre**: ara dues columnes; les altres dues, en acabar cada tasca.

Columna	Buides	% de files	Opció (fila del quadern)	Per què, en una frase
`price`				
`metascore`				
`developer`				
`publisher`				

Tasca 2 · `price`: buit vol dir gratuït? Primer, les proves

```sql
SELECT COUNT(*) FROM games WHERE price = 0;        -- how many games cost exactly 0?
SELECT MIN(price) FROM games;                      -- the cheapest price on record
SELECT game, owners FROM games
WHERE price IS NULL ORDER BY number LIMIT 10;      -- who are they?
```

Jocs que costen exactament 0 _____ Preu mínim _____ Dos que coneixeu de la llista: _____

Aleshores, què vol dir un preu buit en aquesta taula? _____

Ara ompliu-lo i deixeu una nota al marge: l'**indicador** recorda quins valors eren buits.

```sql
SELECT game, COALESCE(price, 0) AS price, (price IS NULL) AS is_free
FROM games WHERE game IN ('Half-Life 2', 'Dota 2', 'Portal') ORDER BY game;
```

`Dota 2`: preu _____ `is_free` _____ `Half-Life 2`: preu _____ `is_free` _____

Sense `is_free`, com distingiríeu un 0 real d'un d'omplert? _____

Dues mitjanes de la mateixa columna:

```sql
SELECT ROUND(AVG(price)::numeric, 2)              AS avg_ignoring_null,
       ROUND(AVG(COALESCE(price, 0))::numeric, 2) AS avg_free_as_zero
FROM games;
```

`avg_ignoring_null` _____ `avg_free_as_zero` _____ Entre quantes files divideix cadascuna? _____

Quina publicaríeu, i per a quina pregunta? _____

Tasca 3 · `metascore`: intenteu omplir-la i mireu els danys

La puntuació mitjana és 71.9. Ompliu els buits amb 72 i compteu.

```sql
SELECT COUNT(*) FROM games WHERE metascore = 72;                -- real 72s
SELECT COUNT(*) FROM games WHERE COALESCE(metascore, 72) = 72;  -- after filling
```

72 reals _____ 72 després d'omplir _____ Com seria un gràfic de la columna omplerta? _____

Ara mireu **a qui** li falta. Primer, els jocs petits:

```sql
SELECT (owners LIKE '0%')  AS under_20k_owners,
       COUNT(*)            AS games,
       COUNT(metascore)    AS with_score,
       ROUND(100.0 * COUNT(metascore) / COUNT(*), 1) AS pct_with_score
FROM games GROUP BY 1;
```

Jocs petits amb puntuació _____ % La resta _____ % Els buits es reparteixen a l'atzar? _____

Per què 72 és una mentida per a un joc que ningú no ha ressenyat? _____

La columna que us quedeu en lloc seu: `(metascore IS NOT NULL) AS has_metascore`.

Tasca 4 · `developer` i `publisher`: una etiqueta per al buit

D'una columna de text no se'n pot fer la mitjana. Doneu al buit una categoria pròpia, perquè `GROUP BY` el mostri en lloc d'amagar-lo.

```sql
SELECT COALESCE(developer, 'Unknown') AS developer, COUNT(*) AS games
FROM games GROUP BY 1 ORDER BY games DESC LIMIT 3;
```

Primera fila _____ amb _____ jocs. És el desenvolupador més gran de la taula?

Per què és una trampa? _____

Heu acabat quan Tots els espais en blanc estan omplerts · el registre de la Tasca 1 té les quatre línies. **Parella ràpida?** El problema contrari: `average_playtime` és 0 a la majoria de jocs, perquè ningú no ho va mesurar. `NULLIF(x, 0)` torna a convertir un 0 en `NULL`: compareu `AVG(average_playtime)` amb `AVG(NULLIF(average_playtime, 0))`.

Encallats?

- **`pct_missing` mostra 0**: divisió entera. Escriviu `100.0 *`, amb el decimal, abans de dividir.
- **`COALESCE types text and integer cannot be matched`**: el substitut ha de tenir el tipus de la columna.
- **`function round(double precision, integer) does not exist`**: afegiu `::numeric` abans d'arrodonir, com a la Tasca 2.

A casa Al vostre conjunt de dades obert, trieu la columna amb més cel·les buides. Mesureu-la (files i %), trieu una de les tres opcions i escriviu el `COALESCE`, o expliqueu per què la deixeu buida.