Variables sintètiques

Columnes que mai no van ser al fitxer Noms _____

Superset, base de dades `Dades classe`, esquema `public`. Partiu de la taula `games` i l'script `games_clean.sql` (demaneu-lo al professor). Parelles: un clica, l'altre llegeix; canvi a la Tasca 3. Resultats aquí.

Referència · el tiquet

Al tiquet	Què és	A SQL / Superset
Total de línia (unitats × preu)	Un valor nou **per a cada fila**, fet només a partir d'aquella fila.	**Columna calculada**: `owners_min * price`
TOTAL de baix de tot	Un valor **per a tot el grup**; les files desapareixen.	**Mètrica**: `SUM(owners_min * price)`
"Aquesta línia és el 40% del tiquet"	Un valor de grup escrit **al costat de cada fila**; les files es queden.	**Funció de finestra**: `SUM(...) OVER ()`

Una **variable sintètica** és una columna que ningú no va recollir: es calcula a partir de les que sí. Llegiu `OVER ()` com "sobre totes les files, sense col·lapsar-les"; `PARTITION BY` és el `GROUP BY` d'una finestra.

Part A · Munteu `games_clean` (8 min)

Pas	Feu això	Hauríeu de veure / escriviu
A1	**SQL → SQL Lab**, esquema `public`: enganxeu tot el `games_clean.sql` i executeu-lo. Després executeu `SELECT COUNT(*) FROM games_clean;`	Files: _____ `games` en té 26 688. Quina línia de l'script ha eliminat la resta? _____
A2	Llegiu l'última sentència de l'script (`ALTER TABLE ...`). Afegeix una columna que mai no va ser al fitxer.	El seu nom: _____ Calculada a partir de: _____
A3	**Datasets → + Dataset** → Database `Dades classe`, Schema `public`, Table `games_clean` → **Create dataset** (la fletxa al costat de **Create and explore dataset**).	`games_clean` és a la vostra llista de Datasets. No comenceu la Part B sense això.

Part B · Tasca 1 · Una columna calculada (6 min)

Pas	Feu això	Hauríeu de veure / escriviu
1.1	**Datasets** → fila `games_clean`, llapis d'**Actions** (**Edit**) → pestanya **Calculated columns** → **Add item**. Nom: `est_revenue`. **SQL expression**: en lloc del text d'exemple, `owners_min * price`. **Data type**: `NUMERIC`. **Save** → **OK**.	Quina línia del tiquet és? _____

1.2 Comproveu la mateixa expressió a SQL Lab:

```
SELECT game, owners_min, price, owners_min * price AS est_revenue
FROM games_clean ORDER BY est_revenue DESC LIMIT 5;
```

Primer joc: _____ `est_revenue`: _____ Milions o milers de milions? _____

Encallats?

- **`This database does not allow for DDL/DML...`**: aviseu el professor. És una casella als ajustos de la connexió.
- **L'script s'ha executat a mitges**: torneu-lo a executar sencer. Sempre reconstrueix la taula des de zero.
- **`This table already has a dataset`**: ja està registrada. Mireu a la llista de Datasets.
- Encara encallats després de dos intents? Pregunteu als veïns i després al professor.

Tasca 2 · Dues mètriques (7 min)

Pas	Feu això	Hauríeu de veure / escriviu

2.1	Mateix editor, pestanya **Metrics** → **Add item**, dues vegades.
	`total_est_revenue` = `SUM(owners_min * price)`
	`pct_free` = `100.0 * SUM(CASE WHEN is_free THEN 1 ELSE 0 END) / COUNT(*)`
	Save → **OK**.
		Feu una predicció abans del pas 2.2: quin `owners_tier` (Hit, Mid, Niche) té la proporció més alta de jocs gratuïts? _____

2.2	A SQL Lab, calculeu les dues expressions per nivell: escriviu la consulta vosaltres mateixos, amb `GROUP BY owners_tier`.	Hit: _____ · _____ % Mid: _____ · _____ % Niche: _____ · _____ %
2.3	Executeu `SELECT SUM(est_revenue) FROM games_clean;` Falla **expressament**.	L'error diu: _____ Heu creat `est_revenue` a la Tasca 1. On viu? _____

Tasca 3 · Funcions de finestra (11 min)

3.1 Executeu-la tal com està escrita: la proporció de jocs de cada franja de preu.

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

Franja més gran: _____ amb _____ % Al tiquet, què és `SUM(COUNT(*)) OVER ()`? _____

3.2 Cada joc comparat amb el preu mitjà del seu any.

```sql
SELECT game, release_year, price,
       ROUND(AVG(price) OVER (PARTITION BY release_year), 2)         AS year_avg,
       ROUND(price - AVG(price) OVER (PARTITION BY release_year), 2) AS vs_year_avg
FROM games_clean ORDER BY owners_min DESC, number LIMIT 5;
```

Quin dels cinc costa més que la mitjana del seu any? _____ Quant més? _____

3.3 Adapteu la consulta de 3.1: en lloc de la proporció de **jocs** per `price_band`, la proporció d'**ingressos estimats** per `owners_tier`.

El nivell Hit és el _____ % dels jocs i el _____ % dels ingressos estimats.

3.4 Mireu la fila Niche de 3.3. De debò van guanyar això, aquests jocs? Doneu un motiu pel qual `est_revenue` **no** és el que un joc va guanyar realment.

_____

_____

Heu acabat quan Tots els espais en blanc estan omplerts · `games_clean` és a la vostra llista de Datasets. **Parella ràpida?** Cliqueu el nom del dataset `games_clean` per obrir-lo a Explore. Tipus de gràfic **Table**; **Dimensions**: `owners_tier`; **Metrics**: trieu `total_est_revenue` a la pestanya **Saved**; **Create chart**. No el deseu. Després proveu les mateixes tres variables al vostre conjunt de dades obert.

A casa Acabeu aquest full. Una finestra més: amb `RANK() OVER (PARTITION BY release_year ORDER BY owners_min DESC)`, trobeu els tres jocs amb més propietaris del 2017.