L'inventari de brutícia

Mireu abans de netejar: què li passa a `games`? Noms _____

Superset → **SQL → SQL Lab**, base de dades `Dades classe`, esquema `public`, taula `games`: el cofre del botí que vau pujar dilluns, un joc per fila. Per parelles, un tecleja i l'altre escriu; canvieu a la Tasca 3. Executeu **una sentència cada vegada**: seleccioneu-la i després **Run**. Avui no arregleu res: obriu el cofre, compteu què hi ha malament i ho apunteu.

Referència · quatre tipus de brutícia

Tipus	Al cofre	En una taula	Com trobar-lo
Valor absent	Una ranura buida.	Un `NULL`: cap valor.	`COUNT(*) - COUNT(col)`
Duplicat	El mateix objecte dues vegades.	La mateixa cosa real en dues files.	`GROUP BY` + `HAVING COUNT(*) > 1`
Tipus o format incorrecte	Una espasa guardada com a imatge d'una espasa.	Una data o un nombre guardats com a text.	Mireu el tipus; ordeneu-la; `LIKE`
Valor atípic	Un objecte amb 9,999 de dany.	Un valor lluny de tots els altres. Un error, o real?	`MIN`, `MAX`, `ORDER BY ... DESC`

`COUNT(*)` compta files · `COUNT(col)` compta les files on `col` **no** és buida · en un patró `LIKE`, `%` és qualsevol text i `_` és exactament un caràcter.

Tasca 1 · Mida i tipus

```sql
SELECT COUNT(*) FROM games;
SELECT column_name, data_type FROM information_schema.columns
WHERE table_name = 'games' ORDER BY ordinal_position;
```

Files _____ Columnes _____ Dues columnes són `text` i no ho haurien de ser: _____ i _____

Tasca 2 · Valors absents

Un `COUNT(*) - COUNT(col)` per columna, totes deu en un sol `SELECT`. Comença així:

```sql
SELECT COUNT(*) - COUNT(price) AS no_price, COUNT(*) - COUNT(game) AS no_game   -- ...and eight more
FROM games;
```

`number` _____ `game` _____ `release_date` _____ `price` _____ `owners` _____

`developer` _____ `publisher` _____ `average_playtime` _____ `median_playtime` _____ `metascore` _____

La pitjor columna és _____ : buida en el _____ % de les files.

Ara els valors absents disfressats:

```sql
SELECT COUNT(*) FROM games WHERE average_playtime = 0;
SELECT COUNT(*) FROM games WHERE price = 0;
SELECT game, price FROM games WHERE game = 'Dota 2';
```

`average_playtime = 0` en _____ files. No hi va jugar ningú, o no ho va mesurar ningú? _____

`price = 0` en _____ files. Dota 2 és gratuït. Aleshores, què vol dir un preu buit? _____

Tasca 3 · Formats incorrectes

Aquesta consulta hauria de llistar els tres jocs més antics. Executeu-la.

```
SELECT game, release_date FROM games ORDER BY release_date LIMIT 3;
```

Primera data _____ Segona data _____ Tercera data _____

Surt primer el més antic? Per què la base de dades les ha ordenat així? _____

La majoria de dates són com `Nov 16, 2004`. Trobeu les que no ho són i després mireu `owners`:

```
SELECT release_date, COUNT(*) AS n FROM games
WHERE release_date NOT LIKE '___ %, ____'       -- 3 letters, space, day, comma, 4 digits
GROUP BY release_date ORDER BY n DESC;
SELECT owners, COUNT(*) AS n FROM games GROUP BY owners ORDER BY n DESC;
SELECT game, owners FROM games ORDER BY owners DESC LIMIT 3;
```

Què els falta a la majoria de dates estranyes? _____ Una ho és d'una altra manera: _____

Files amb una data estranya (`SELECT COUNT(*)` amb el mateix `WHERE`) _____ Valors diferents d'`owners` _____

La tercera consulta hauria de donar els jocs amb més propietaris. Diu _____

Per què és incorrecte? _____

Tasca 4 · Duplicats: hi ha algun joc dues vegades a la taula?

```
SELECT game, COUNT(*) AS n FROM games WHERE game IS NOT NULL
GROUP BY game HAVING COUNT(*) > 1 ORDER BY n DESC, game;
SELECT number, game, release_date, developer FROM games WHERE game = 'Dark Matter';
SELECT COUNT(*) AS all_rows, COUNT(DISTINCT number) AS numbers FROM games;
```

Noms que apareixen més d'una vegada _____ Files que es diuen `Dark Matter` _____

Són el mateix joc? Com ho sabeu? _____

`all_rows` _____ `numbers` _____ És `number` un identificador? _____

Tasca 5 · Valors atípics (els preus són en dòlars dels EUA)

```
SELECT MIN(price), MAX(price), ROUND(AVG(price)::numeric, 2) AS mean FROM games;
SELECT game, price, developer FROM games ORDER BY price DESC NULLS LAST LIMIT 8;
```

Mín. _____ Màx. _____ Mitjana _____ La fila més cara: _____

És un error de tecleig? El vostre motiu: _____

Heu acabat quan Tots els espais estan omplerts i l'inventari, en un full a part, té almenys vuit files: columna, tipus de brutícia, evidència (la consulta i el seu nombre). Deixeu **Decisió** buida: l'omplirem junts. **Parella ràpida?** Feu la Tasca 2 amb el vostre conjunt de dades obert i afegiu-hi les seves tres pitjors columnes.

Encallats?

- **`relation "games" does not exist`**: comproveu l'esquema, ha de ser `public`.
- **Només un resultat per a tres consultes**: SQL Lab mostra l'últim. Executeu-les d'una en una. **Només tornen 1,000 files**: és el límit de visualització, no la taula. Compteu amb `COUNT(*)`.
- "Hi ha NULLs" no és evidència. Sense nombre, no hi ha evidència. Encara encallats després de dos intents? Pregunteu als veïns i després al professor.

A casa Acabeu l'inventari amb la columna **Decisió**. Després, tres línies d'inventari per al vostre conjunt de dades obert: algun dels seus `NULL` vol dir alguna cosa?