The stat card

Centre, spread and shape of `games_clean` Names _____

Superset → **SQL → SQL Lab**, database `Dades classe`, schema `public`, table `games_clean` (26 685 games). Pairs: one types, one writes; swap at Task 3. Write the numbers here.

Reference · the queue

Put the data in a queue, smallest first. Five friends earn 1 200, 1 300, 1 400, 1 500 and 30 000: the **mean** (add everything, split it equally) is 7 080; the **median** (the one in the middle of the queue) is 1 400.

Question	In the queue	SQL
Centre	Where is the middle?	`AVG(x)` · `PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY x)`
Spread	How far apart are the first and the last? And the middle half?	`MIN(x)` · `MAX(x)` · `STDDEV(x)` · percentiles 0.25 and 0.75, the **quartiles** Q1 and Q3
Shape	Where does the queue bunch up?	`width_bucket(x, from, to, n)` + `COUNT(*)`

Read `WITHIN GROUP (ORDER BY price)` as “make the queue by price” and `PERCENTILE_CONT(0.5)` as “walk to 50% of the queue and read the value”. The **standard deviation** is the typical distance from the mean. `width_bucket` says which of `n` equal boxes between `from` and `to` a value falls in; anything above `to` goes to box `n + 1`.

Task 1 · Centre and extremes of `price` (8 min)

1.1 Before you run anything: will the mean price be above or below the median? _____ Then run:

```
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` _____ Were you right? _____

1.2 Who is at the very end of the queue? Find the three most expensive rows (`ORDER BY price DESC LIMIT 3`).

The last one in the queue: _____ Is it a game? _____

1.3 A classmate says: “The median is halfway between the lowest and the highest.” Compute that number.

Halfway: _____ Is it the median? Why not? _____

Task 2 · Spread (7 min)

```
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` _____ Half of all games cost between _____ and _____

2.2 Count the games that cost more than the mean (`WHERE price >` your mean).

Games above the mean: _____ of 26 685. Is that half? _____

Task 3 · Shape of `price` (9 min)

Six boxes of 10 from 0 to 60; anything above goes to box 7. The last column draws the bar: one `#` per 500 games.

```sql
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;
```

Copy the `games` column into the left table below and sketch the seven bars. Then mark on the sketch where the mean (**M**) and the median (**Md**) fall.

Task 4 · A second column: `metascore` (9 min)

Repeat Tasks 1 to 3 for `metascore`, in two queries. Changes: `ROUND(AVG(metascore)::numeric, 1)`; boxes `width_bucket(metascore, 20, 100, 8)`; one `#` per 25 games.

4.1 Run the boxes query as it is. One row too many comes out.

Its bucket: _____ Its games: _____ Who are they? _____

4.2 Add `WHERE has_metascore` to both queries and fill in the card and the right table.

`n` _____ lowest _____ highest _____ mean _____ median _____ std. dev. _____ Q1 _____ Q3 _____

Box	Games	`price`: your bar
1		
2		
3		
4		
5		
6		
7		

Box	Games	`metascore`: your bar
1		
2		
3		
4		
5		
6		
7		
8		

4.3 Look at the two sketches side by side.

Which column has mean ≈ median? _____ What does its sketch look like? _____

And the other one? _____

Done when Every blank is filled · both sketches have **M** and **Md** marked. **Fast pair?** Add `GROUP BY owners_tier` to the Task 1 query: do Hit games cost more? Then the stat card of one numeric column of your own open dataset.

Stuck?

- **`function round(double precision, integer) does not exist`**: cast first, `ROUND(metascore::numeric, 1)`.
- **The `lowest` of a box is not a round number**: correct. It is the cheapest game **in** the box.
- **`relation "games_clean" does not exist`**: ask the teacher for `games_clean.sql` (18/11) and rerun it.

At home Finish Task 4 if needed. Keep this sheet: on Monday it becomes the “key figures” section of your report.