LLM TOOLING · CHEAT SHEET

The querychat data description

Text-to-SQL is only as good as the description it reads. Fill in the template, then run the three test questions yourself and compare with what the model writes before anyone else uses it.

Template The data description, filled in for the season game log.

```
Table game_log: one row per game of one team's 162-
game season.
- game: integer, 1 to 162, the order games were
played.
- home: 1 for a home game, 0 away.
- runs_scored, runs_allowed: runs in that game.
- weekend: 1 for a Friday, Saturday or Sunday game.
- attendance: paid attendance; 0 on away games (not
missing).
Rules: a win is runs_scored > runs_allowed.
Attendance questions use home games only.
```

SQL The three test questions, run against the data with sqlite3.

```
-- 1. How many home games did the team win?
SELECT COUNT(*) AS home_wins FROM game_log WHERE
home = 1 AND runs_scored > runs_allowed;

-- 2. Average attendance, weekend against weekday
(home games only)
SELECT weekend, ROUND(AVG(attendance)) AS
avg_attendance FROM game_log WHERE home = 1 GROUP
BY weekend;

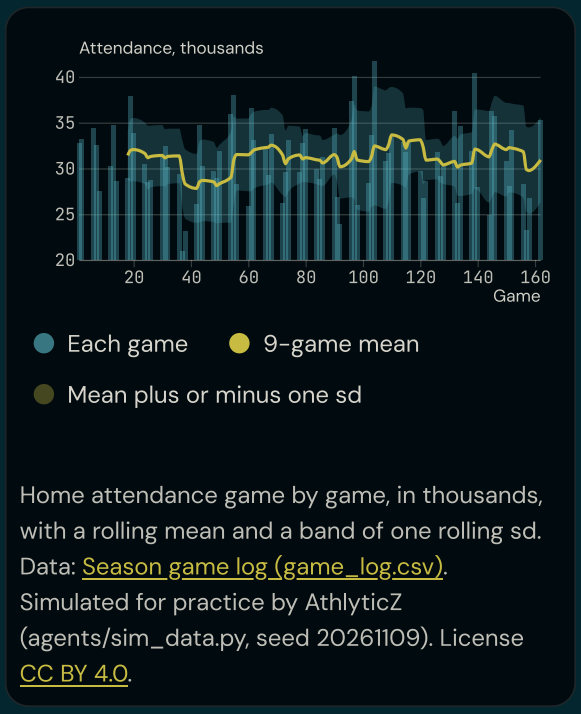
-- 3. The longest winning streak
WITH g AS (SELECT game, runs_scored > runs_allowed
AS win,
                  game - ROW_NUMBER() OVER
(PARTITION BY runs_scored > runs_allowed ORDER BY
game) AS grp
           FROM game_log)
SELECT MAX(n) AS longest_win_streak FROM (SELECT
COUNT(*) AS n FROM g WHERE win = 1 GROUP BY grp);
```

Ran when this sheet was built. Output:

```
home_wins
       45

weekend  avg_attendance
      0         29450.0
      1         35061.0

longest_win_streak
                   7
```



Home attendance game by game, in thousands, with a rolling mean and a band of one rolling sd. Data: Season game log (game_log.csv). Simulated for practice by AthlyticZ (agents/sim_data.py, seed 20261109). License CC BY 4.0.

One line per column
Name, type, unit, and what a value means, in plain words.

Business rules
The things a newcomer gets wrong: attendance is zero on away games, not missing.

Grain
One row is one game. Say it, or the model will sum the wrong thing.

Three test questions
A count, a grouped average, and something that needs a window. Run them yourself first.

From Ask the Box Score Anything. Every code block on this sheet ran when it was built; the data is simulated for practice.