

ITWS-4250/6250 Database Applications and Systems

Homework 3

For this assignment, you will construct a *single SQL query* to generate the answer to each of the below questions. Use the baseball database provided here: (<https://gitlab.com/samuelbjohnson/baseball>). A brief overview of the database can be found in the `notes.md` file in the `homework-3` directory.

You should put your answers in a series of `.sql` files, each named with the number of the question it answers (`1.sql`, `2.sql`, etc.), and then upload all of your solutions to Submitty. The assignment will be auto-graded, expecting an exact match for output. The auto-grader will consider only the first valid SQL statement in each file. A schema, specifying name and order of columns, is provided for each question.

Default Submitty rules for multiple submissions will apply (you get twenty free attempts, after that, each additional attempt will cost you 0.1 points up to a maximum of 5 points).

A few useful notes and clarifications:

- Batting average is defined as h/ab (hits divided by at-bats)
- Salary data is only available starting in 1985. Any questions that involve salary information assume a natural join (i.e., don't try to pull in data from years without salary data).
- Players are chosen for the "All Star Game" each year. A given player may appear in the all star game in multiple years.

Query efficiency isn't considered, except that the total runtime for all of your queries is capped, and the queries are run in numerical order. Therefore, a query that is too slow may cause the process to timeout, leading to no credit for that and subsequent queries.

Each question is worth four points, for a total of sixty points. **This assignment is due at 23:59 PM on Monday October 12.**

1. The number of pitchers with more than ($>$) 20 saves (`sv`) in a single year since (\geq) 1970. Note that a pitcher may play multiple stints in a given year. (*count*)
2. The number of players under 6 feet ($< 72inches$) whose batting average was above .300 in 2003. Note that integer arithmetic will apply and needs to be taken into account. Ignore players with fewer than 75 at-bats. (*count*)
3. List the years since (\geq) 1965 where the manager with the most wins also managed the team that won their league (`teams.lgwin`), in descending order with the earliest year last. (*years*)
4. The combined team salaries for each year's world series winner in descending order by yearly attendance. (*year, salary*)
5. Franchise name and average total yearly franchise attendance in the 1990's (use attendance from teams table, and don't worry about the attendance of games played elsewhere), in descending order of attendance (*franchise, attendance*)
6. First and last name and number of All Star appearances by the eight players *on the ballot* for the Hall of Fame in 2000 with the most All Star appearances, using the number of votes received as a tie-breaker if needed. Note that having a tuple in the `halloffame` table for a given year indicates that a player is on the ballot for that year, and that having a tuple in the `allstarfull` table indicates an All Star appearance (this does not require use of the `appearances` table). (*first, last, appearances*)

7. Career (total) batting average (label the column `career_avg`) of players born in 1960, who have more than 300 at-bats, ordered by average, highest to lowest, using `playerid` as a tie-breaker if needed. Round averages to three digits after the decimal point. (*playerid, career_avg*)
8. Ratio of average All Star player salaries to average player salaries for each year, in order of year, earliest years first, with the ratio rounded to five digits after the decimal point. (*year, ratio*)
9. Average combined team salary for each year, in order of year, earliest years first. (*year, salary*)
10. The school name and number of All Star players for the ten colleges with the most All Star players. List the school with the most players first. A player counts as an All Star player regardless of how many times they've played in the All Star game. In case of ties, use the school name in alphabetical order. (*school_name, count*)
11. The park names (`teams.park`) that do not include "field", "park", or "stadium" for teams that won their division series (use the `seriespost` table for this). (*park*)
12. The names of the teams that have lost more than 2 (>) league championships. (*name*)
13. The first and last name of players who have a second stint where they had more at-bats (>) but fewer hits (<) than their first stint of the same year, in order from heaviest to lightest weight. (*first, last*)
14. The awards (use `awardsplayers.awardid`) that were won by the same player three consecutive years between 1950 and 1959, inclusive. (*award*)
15. The first and last name and birth city of all players who were born in New York State and played for the New York Yankees franchise (`franchid='NYY'`). (*first, last, city*)