

## CSCI-4380 Database 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 (<https://gitlab.com/samuelbjohnson/baseball>) introduced in class.

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. 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.

Each question is worth five points, for a total of fifty points. **This assignment is due at 15:59 PM on Thursday February 22.**

1. The number of pitchers with more than ( $>$ ) 40 saves (`sv`) in a single year since ( $\geq$ ) 1975. 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 2016. Note that integer arithmetic will apply and needs to be taken into account. (*count*)
3. List the years since ( $\geq$ ) 1975 where the manager with the most wins also managed the team that won the world series, in ascending order with the earliest year first. (*years*)
4. The combined team salaries for each year's world series winner in descending order, with the highest combined team salary first. (*year, salary*)
5. Franchise name and average total yearly franchise attendance from ( $\geq$ ) 1997 to the present (use attendance from teams table, only consider the number of home games) (*franchise, attendance*)
6. First and last name and number of All Star appearances by the ten players *on the ballot* for the Hall of Fame in 2000 with the most All Star appearances. Note that having a tuple in the `halloffame` table for a given year indicates that a player is on the ballot for that year. (*first, last, appearances*)
7. Career (total) batting average (label the column `career_avg`) of players born after in or after ( $\geq$ ) 1970, ordered by average, highest to lowest. *Note that your calculations should exclude any players who have no at-bats, as they have no batting average.* (*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. (*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. (*school\_name, count*)