

# SQL HW 2

University of San Francisco HS 611

Think carefully about each problem before beginning. There might be caveats to various methods or edge cases to consider. Your answers should be written in a PDF document or plain text file, and should be numbered according to the problem numbers in this assignment. For those of you familiar with R, it is suggested that you use RMarkdown to knit a clean-looking PDF. For each problem, you are expected to write a single SQL query as your solution *and* on the next line copy and paste the output of that query. Only include the first FIVE lines of output for any queries that output more than five lines. For all problems, *only return the information and columns requested* and only use SQL commands shown in class. Every query should produce at minimum one column.

1. Using the table `cmsclaims`, write a query to answer each of the following:
  - a. For every possible value in `hmo_mo`, what is the average carrier reimbursement? Order by number of HMO months in ascending order.
  - b. For every value of carrier reimbursement, how many *total* months of HMO coverage were provided? Order by carrier reimbursements in descending order.
  - c. Using your query from part b, create a query that finds the largest carrier reimbursement when the total months of HMO coverage is zero.
2. Using the table `cmspop`:
  - a. Return the number of instances of depression by sex, ordered by sex in ascending order.
  - b. Return the most common month to be born in.
  - c. Return the average birth year for all records. Hint: use the `round()` function.
  - d. Return the average age (should contain a decimal) of those who died under the age of 50.
  - e. Now return the average age of those who died under the age of 50 to just two decimal places. Hint: if you get stuck, Google the error message.
3. Using the table `cmspop`:
  - a. Return the state and number of heart failures for the state with the greatest number of heart failures.
  - b. Find which county in California has the most hispanics. Return columns for `county`, `race`, and the number of hispanics in that county. Hint: you can group by more than one column at a time.
4. The data in `cmspop` was collected on Jan 1st, 2010. With that in mind:
  - a. Return the youngest age in *integer years* at the time of data collection by sex for those who had cancer and were alive, ordered by sex in ascending order. To be clear, age in integer years means that age should be calculated like when you tell your age to someone: for example, "I'm 62 years old." You don't say, "I'm 62.7456 years old". Hint: use the `floor()` function.
  - b. Return the `id`, `sex`, `race`, and age in integer years at the time of data collection for the oldest living person in Massachusetts that has all complications (Alzheimers, depression, cancer, and heart failure).
5. Assume that, for the deceased, the flags for Alzheimers, depression, cancer, and heart failure indicate cause of death. Using the table `cmspop`, consider the following scenario:
  - a. In the 0th, 80th, 150th, 160th, 180th, 190th, and 220th counties within Wyoming, each of the deceased had a single cause of death. Count the number of deaths caused by each complication for those individuals. Hints: you can use parenthesis in `WHERE` statements to group conditions and, also in `WHERE` statements, you can pass a list of values to `IN`.