MLDS-413 Introduction to Databases and Information Retrieval
Homework 4: INNER JOINs

Name 1: _____

NetID 1: _____

Name 2: _____

NetID 2: _____

Instructions

You should submit this homework assignment via Canvas. Acceptable formats are word files, text files, and pdf files. Paper submissions are not allowed and they will receive an automatic zero.

As explained during lecture and in the syllabus, assignments are done in groups. The groups have been created and assigned. Each group needs to submit only one assignment (i.e., there is no need for both partners to submit individually the same homework assignment).

Each group can submit solutions multiple times (for example, you may discover an error in your earlier submission and choose to submit a new solution set). We will grade only the last submission and ignore earlier ones.

Make sure you submit your solutions before the deadline. The policies governing academic integrity, tardiness and penalties are detailed in the syllabus.

Homework Instructions

For this assignment, you will use the program "DB Browser for SQLite" (available at http://sqlitebrowser.org/). This is the same software we have worked with in class. I posted several sample database files on Canvas in the Lecture Materials page. These database files can be opened with the DB Browser for SQLite. The database files we will use for this homework are:

- EntertainmentAgency.sqlite
- BowlingLeague.sqlite
- SchoolScheduling.sqlite

For every question, we expect to see both your SQL code and the resulting data. Copy and paste both the SQL code and the results into a document and submit it following the submission instructions.

Here is an example question that applies to the SalesOrders.sqlite database:

Question: What bikes cost more than $1000?

Answer:
```
SELECT ProductName, RetailPrice
FROM Products JOIN Categories
     ON Products.CategoryID = Categories.CategoryID
WHERE CategoryDescription = "Bikes"
      AND RetailPrice > 1000;

Output:
"Trek 9000 Mountain Bike"  "1200"
"Eagle FS-3 Mountain Bike" "1800"
"GT RTS-2 Mountain Bike"   "1650"
```

You must answer each question with a single query.

You may find it helpful to use the "Basic SQL Cheat Sheet" posted on Canvas.

Each one of the questions below is worth **10 points**.

SchoolScheduling.sqlite

1) How many students are majoring in English or Mathematics? To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

2) What is the full name of the instructor of the class that has the highest average students' grade? Your output should list the full name of the instructor, the Class ID, and the average students' grade of that class. The full name of the instructor can be formed by concatenating the last name, a comma, a space, and the first name. For example, the full name of the professor in this MLDS-413 class is the string "Hardavellas, Nikos". To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

3) What is the percentage of students with majors in English or Mathematics? To receive full credit for this question you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all). To receive partial credit you **must** use the JOIN operator and you **must not** use subqueries in the WHERE clause (subqueries elsewhere are fine).

EntertainmentAgency.sqlite

4) What percentage of all entertainer members are male entertainer members whose musical style is Jazz, and what percentage of all entertainer members are female entertainer members whose musical style is Jazz? You should provide a single query that outputs the percentages of each gender separately and indicates which is which. To receive full credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all). To receive partial credit you **must** use the JOIN operator and you **must not** use subqueries in the FROM and WHERE clauses (subqueries elsewhere are fine).

5) What is the full name (in the form "LastName, FirstName") of the top 3 agents who have the highest average commission per engagement? The commission can be calculated by multiplying the contract price with the agent's commission rate. To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

6) What is the total income of the Jazz entertainers (i.e., the sum of all Jazz entertainers' income across all of their engagements) and the total income of the Salsa entertainers? The income of each entertainer for each engagement is the ContractPrice of the engagement minus the agent's commission. To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

7) What are the top 5 musical styles that have the highest number of unique customers, and how many customers each of these styles has? To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

BowlingLeague.sqlite

8) Which teams have captains with the same last name? Each such team must be listed exactly once, along with the team captain's full name (in the form "LastName, FirstName"). To receive full credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

9) In question 8 you identified the bowling teams that have captains with the same last name. List all the matches in which any of these teams participates. The output should provide the TourneyDate, TourneyLocation, odd and even Team Names, and Lanes. You can use the team names identified in question #8 here to make the query easier. To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).

10) How many teams have different players in the same team with the same last name? To receive credit you **must not** use subqueries anywhere (i.e., no nested SELECT clauses at all).