assignmentQuery

August 24, 2017

1 Assignment 1 -- SQL 1

1.1 Due September 7, 2017, by 11:45 PM.

The goal of this assignment is to write several SQL queries that will answer questions about a set of surgical cases. The data is based on the Anaesthetic Shoulder Arthroscopy Cases (ASAC) Dataset from David Cumin. I have made some minor modificiations to the data, so be sure to use the version of the data available on the course Canvas site. If you are interested, you can learn more about the dataset at https://researchspace.auckland.ac.nz/handle/2292/5378.

There are 6 tables:

1. AN_CASE(ID, LASTNAME, FIRSTNAME, WEIGHT, HEIGHT, SEX, DOB, ASA, OPDATE, PROC)
2. AN_COMORBID(ID, ICD, DESCR, DETAILS)
3. AN_DRUGCATEGORY(DRNAME, DRCAT)
4. AN_DRUG(ID, DRNAME, DRTIME, DRVALUE, DRUNIT, DRROUTE, DRCOMMENT)
5. AN_EVENT(ID, EVENTTIME, EVENTDESC)
6. AN_VITALS(ID, SIGNALTIME, SIGNALNAME, VALUE)
7. AN_CASE contains an entry for each surgical case. It includes the patient's name, gender, height and weight. The ASA is the American Society of Anesthesiologist physical status classification, and reflects the patient's ability to tolerate the surgery. DOB is the patient's date of birth and OPDATE is the date the operation happened. PROC describes the procedures performed.
8. AN_COMORBID contains a list of patient conditions, such as smoking or asthma. Patients can have different numbers of comborbidities. Comorbidities are described using an ICD code (ICD-10-CM Diagnosis Code) and a text description. Some conditions include additional information in the DETAILS field.
9. AN_DRUGCATEGORY contains a list of different medications used during the surgical cases and their categories.
10. AN_DRUG contains a list of medications administered during each surgical case along with the details of the administration (dose, time, method, etc.).

11. AN_EVENT contains key events during the surgical case. Each row includes the case id, time the event occurred and description of the event.

12. AN_VITALS contains the vital sign data (heart rate, systolic and diastolic blood pressure) collected during each case.

1.1.1 What to turn in

You must turn in your Jupyter Notebook on Canvas.

1.1.2 Grading

Each query is worth 10 points. Points will be assigned for each query based on the following guidelines: * 0 points: Query not attempted, query does not give any results, or it does not compile * 5 points: Query compiles, runs and is most of the way towards a correct answer * 8 points: The query and answer it produces are almost correct, but there is a slight or subtle bug in the query * 10 points: The query is correct and gives the right answer

1.1.3 What's In and Out of Scope

This is intended to be a SQL query assignment. Therefore, you must write queries in SQL (not stored procedures or functions or python code). You may use VIEWs as needed and you may use standard built-in MySQL functions (e.g. ROUND, IF or CASE statements). If you're not sure if something is allowed, ask!

1.2 Acadmic Honesty

The following level of collaboration is allowed on this assignment: You may discuss the assignment with your classmates at a high level. Any issues getting Jupyter Notebooks or MySQL running is totally fine. What is not allowed is direct examination of anyone else's SQL code (on a computer, email, whiteboard, etc.) or allowing anyone else to see your SQL code. You MAY post and discuss query results with your classmates.

You may use the search engine of your choice to lookup the syntax for SQL commands, but may not use it to find answers to queries.

It does not matter whether or not you show the output of each code block. Submit your resulting .ipynb file on Canvas.

First, the standard preliminary steps. For security reasons, don't include your ricedb password.

To run the code, click on it, and press SHIFT+ENTER.

```
In [ ]: %load_ext sql
```

```
In [ ]: %sql mysql+pymysql://ricedb:yourpassword@localhost/ricedb
```

1.3 Preparation

Start fresh, if you need to recreate the tables.

```
In [ ]: %%sql
        DROP TABLE IF EXISTS an_case;
        DROP TABLE IF EXISTS an_vitals;
        DROP TABLE IF EXISTS an_comorbid;
        DROP TABLE IF EXISTS an_event;
        DROP TABLE IF EXISTS an_drug;
        DROP TABLE IF EXISTS an_drugCategory;
```

Create the tables you will need for this assignment by excuting the following code:

```
In [ ]: %%sql
        CREATE TABLE an_case (
            id INTEGER,
            lastName varchar(50),
            firstName varchar(50),
            weight FLOAT null,
            height FLOAT null,
            sex CHAR(1),
            dob DATE,
            asa INTEGER,
            ebl INTEGER,
            opDate DATE,
            proc VARCHAR(110)
        );

        CREATE TABLE an_vitals (
            id INTEGER,
            signaltime INTEGER,
            signalname CHAR(3),
            value FLOAT
        );

        CREATE TABLE an_comorbid (
            id INTEGER,
            icd VARCHAR(10),
            descr VARCHAR(100),
            details VARCHAR(100)
        );


        CREATE TABLE an_event (
            id INTEGER,
            eventtime INTEGER,
            eventdescr VARCHAR(100)
        );

        CREATE TABLE an_drug (
            id INTEGER,
```

```
    drname VARCHAR(25),
    drtime INTEGER,
    drvalue FLOAT,
    drunit VARCHAR(10),
    drroute VARCHAR(50),
    drcomment VARCHAR(100)
);

CREATE TABLE an_drugCategory (
    drname VARCHAR(25),
    drcat VARCHAR(25)
);
```

Now load the data needed for the assignment. You have to do this in MySQL Workbench. The files are located on the class Canvas site in https://canvas.rice.edu/courses/3600/files/folder/HW1.

1. Download the files to your computer
2. Update the statements below to point to your file location
3. Execute the statements in MySQL Workbench

LOAD DATA LOCAL INFILE '[your file location]/an_case.txt' INTO TABLE an_case fields terminated by '|' escaped by '\' OPTIONALLY ENCLOSED BY '"' lines terminated by '\r\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '[your file location]/an_vitals.txt' INTO TABLE an_vitals fields terminated by '|' escaped by '\' OPTIONALLY ENCLOSED BY '"' lines terminated by '\r\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '[your file location]/an_comorbid.txt' INTO TABLE an_comorbid fields terminated by '|' escaped by '\' OPTIONALLY ENCLOSED BY '"' lines terminated by '\r\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '[your file location]/an_event.txt' INTO TABLE an_event fields terminated by '|' escaped by '\' OPTIONALLY ENCLOSED BY '"' lines terminated by '\r\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '[your file location]/an_drug.txt' INTO TABLE an_drug fields terminated by '|' escaped by '\' OPTIONALLY ENCLOSED BY '"' lines terminated by '\r\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '[your file location]/an_drugCategory.txt' INTO TABLE an_drugCategory fields terminated by '|' escaped by '\' OPTIONALLY ENCLOSED BY '"' lines terminated by '\r\n' IGNORE 1 LINES;

If you accidentally load the data more than once, or run into some other problem, you can rerun these steps after dropping the tables and recreating and loading them as described above.

1.4 Queries

Answer all of the questions below by writing and executing SQL queries. The queries must contain ONLY the answer to the question (no extra rows or columns). You many only use SQL to answer the questions. You may need to explore the database a bit prior to generating your final solutions. You can limit the number of records returned from a query by using the ending term "LIMIT N", where N is the number of records to turn.

For example, "SELECT * FROM AN_VITALS LIMIT 100;" returns the 100 records from the AN_VITALS table.

1.4.1 Query 1

How many patients are male?

```
In [ ]: %%sql
```

1.4.2 Query 2

What are the names of the different signals recorded in the vitals table (in alphabetical order)?

```
In [ ]: %%sql
```

1.4.3 Query 3

How old was each patient at the time of the operation? (show id and age in years). Sort in order by age from youngest to oldest.

```
In [ ]: %%sql
```

1.4.4 Query 4

Which patients are either smokers or have allergies? Use a UNION operator in this query. List the relevant patient ids, first and last names in order by last then first name.

```
In [ ]: %%sql
```

1.4.5 Query 5

Which woman's highest systolic blood pressure was over 170? Show her first and last name.

```
In [ ]: %%sql
```

1.4.6 Query 6

The function TIMESTAMPDIFF(unit,startingDatetime,endingDatetime) can be used to calculate elapsed time in different units. You may use this function to help answer this question.

Which patients who are at least 40 years old at the time of surgery, had a max SBP < 120? Show lastname, firstname, max SBP

```
In [ ]: %%sql
```

1.4.7 Query 7

What is the average number of comorbidities? (to two decimal places)

```
In [ ]: %%sql
```

1.4.8 Query 8

What are the 3 most frequent comorbidity ICD codes? Show the ICD code, description and the number of occurrences. List in descending order by frequency.

```
In [ ]: %%sql
```

1.4.9 Query 9

What is the eventdescr value for the last event for Patient 3?

```
In [ ]: %%sql
```

1.4.10 Query 10

Which patient(s) do not have a "knife to skin" event? List id(s) in numerical order. Do not repeat Ids.

```
In [ ]: %%sql
```

1.4.11 Query 11

How long was each patients' surgery (in minutes)? Round to the nearest minute using the SQL ROUND command. List the case id and the number of minutes, in order from shortest to longest, then by id.

```
In [ ]: %%sql
```

1.4.12 Query 12

Which case had the longest surgical time? (Knife to skin to Surgery / operation over events). Give the case id.

```
In [ ]: %%sql
```

1.4.13 Query 13

How long was the case? (Knife to skin to Surgery / operation over events). Give the answer in whole minutes.

```
In [ ]: %%sql
```

1.4.14 Query 14

The hospital wants to reduce it's inventory. So, it wants to review drugs that are used infrequently.

Find all the named drugs (from the drugs table or from the drug category table) used in less than 2 cases. List the drug name and the number of cases it was used in. Sort by drug name.

```
In [ ]: %%sql
```

1.4.15 Query 15

The hospital wants to identify patients who might have complications. One way to do this is to use to do this is to use the Surgical Apgar score (http://www.atulgawande.com/documents/AnApgarScoreforSurgery.pdf).

The score uses estimated blood loss (the EBL column in our case table), the patient's minimum mean blood pressure (they want arterial, but our noninvasive measure will suffice), and lowest heart rate.

	0 points	1 point	2 points	3 points	4 points
EBL		> 1,000	601-1,000	101-600	<= 100	-
Lowest Mean BP		< 40	40-54	55-69	>=70	-
Lowest Heart Rate		> 85	76-85	66-75	56-65	<= 55

Mean Blood Pressure = $\frac{SBP+2*DBP}{3}$. It weights the diastolic blood pressure twice as much as the systolic. Note that valid (non-artifact) values for DBP are > 20 and SBP > 40.

Show the case id, EBL points, BP points, HR points and the final Surgical Apgar Score for each case. Sort in order by id.

Your SQL code MUST handle all possible ranges of values.

```
In [ ]: %%sql
```

1.5 Short answer questions

These questions will be graded based on the thought and effort put into the answers (yes, I know that's a bit vague). Trivial or very minor answers that don't significantly differ from the existing solution will be given minimal points. Answers should be reasonable approaches, without resorting to overloading fields, etc.

1.5.1 Short answer 1

Describe another way of structuring the vital sign data within the database. (5 points)

What the are advantages and disadvantages? (5 points)

Specifically address storage space and how you can access values for more than one signal at any given time point.

1.5.2 Short answer 2

Describe a more compact way to store the BP values. (5 points)

Calculate the savings, based on the following storage requirements: https://dev.mysql.com/doc/refman/5.7/en/storage-requirements.html#data-types-storage-reqs-numeric

You must explain how you got to your space saving answer. (5 points)

1.5.3 Short answer 3

Some of the event times are negative numbers.

Why might this be? (3 points)

What are the implications? (4 points)
What might you choose to do about this? (3 points)

1.6 Survey (5 points)

It took me approximately N hours to complete this assignment, where N is: