DB Systems H - Week 6: Worksheet & In-class Activities

Introduction

Before you attempt solving the tasks listed in this worksheet, please make sure that you have installed the both "peewee" and "csvkit" Python packages. You may use any Python IDE or you may work using a Jupyter Notebook.

Task 1

Download the dataset that comprises the New York City school level College Board SAT results for the graduating seniors of 2010 from this GitHub repository (https://github.com/ip-iulia/DBH_Week_6_Tutorial).

This dataset is presented in CSV format and the columns have the following meaning:

- dbn is an alphanumeric code identifier of the schools;
- school_name represents the name of the school;
- number_of_test_takers comprises the number of students who took the SAT test;
- critical_reading_mean represents the mean value of all the scores obtained by the students of a school in the Critical Reading assessment;
- mathematics_mean represents the mean value of all the scores obtained by the students of a school in the Mathematics assessment;
- writing_mean represents the mean value of all the scores obtained by the students of a school in the Writing assessment.

Task 2

Part 1: Convert the dataset into a SQLite database containing a table named *schools.* Hint: you can use the *csvsql* command in your terminal / command line interface.
Part 2: In your terminal, run *sqlite3 name_of_your_database.db* to open the SQLite console.
Part 3: Using the SQLite console, look at the table and its schema. Hint: *.table* and *.schema* commands might be useful)

Task 3

Create a new Python script and import the peewee package. Then connect to the SQLite database that you created in Task 2.
After you connected to the database, we need to define our **models**. A model is a representation of a database **table**.
In other words, you need to "explain" to peewee what is in the table and what each column is. For example, school_name = CharField() means that school_name is a string, while writing_mean is an IntegerField(), because it represents a number.

Task 4

Now that you have everything set up, we can start querying our database. For example, let's find out:

- What is the name of the school with the dbn equal to *01M292* ?
- How many schools had more than 500 students who took the SAT test ?
- And what are their names ?
- Also, tell me the first 5 schools, which had more than 50 test takers, sorted in descending order by their mean Mathematics score.

Task 5

Part 1: Create a new SQLite database named "student.db" and connect to it using peewee.

Part 2: Define a Student model that has the following attributes:

- A unique username with a maximum length of 255 characters;
- An integer score;

Task 6

Part 1: Build a method, which adds students to the database.

Part 2: Implement a method, which returns the username of the student with the highest score.

Task 7

In the "student.db" database created in Task 6, create a method that deletes the student with the lowest score.