CSC2001F ASSIGNMENT 4 - 2018

Two CSV files have been created that store some (fabricated) university results and information about students' matric record. You are required to create a Mysql database to store each CSV file in a separate relation, and load each with **exactly** the information contained in that CSV file. Then write SQL queries to answer the "management" questions listed below and run them against your database. Do this **individually, not in groups**.

Your queries must be correct for any instance of this database schema, and not just for the given sample data. If Mysql is unable to handle any of the queries, state this clearly in your answer, with a brief reason why. If you cannot find the precise information required in any of these "management" queries, give information as close as possible to what is being requested, explaining clearly (in words) what it is you are providing. If you are unsure about precisely what is being asked (requests are usually ambiguous) choose any one possible meaning and explain clearly the interpretation you have selected. **Please use the Vula Forum if there are problems** – emails will NOT be answered – that way everyone has the same information.

Submit 1 file with query number, SQL code and output produced, for each query in turn.

1. Output any one tuple from each of your relations (NB just one each, no more).
2. Output the number of tuples in each of your relations.
3. Show all entries where the gender disagrees with the prefix (e.g. someone is "Mr" but their gender is "F").
4. List students from school "1000929", giving first name and last name, in alphabetical order of last name.
5. For each course, give course code and number of units. A unit is 6 credits (18 credits is 3 units, etc.)
6. Give the names of all students whose citizenship status is "C" but who are missing their "examining authority", "old ex auth" and "UCT score" data.
7. Give the StuID of students who registered for both ZOO3 and ACCS1.
8. Find the course and student in the database with the highest university result, along with that mark (i.e. that mark which is the largest university mark in the database).
9. How many students have a mark (in any course) lower than some mark that student 1027 achieved?
10. . How many result categories are there in the final symbols? (i.e. how many different Symbol values?)
11. . Give the average mark over all university results in all courses (i.e. a single value in the answer, the average mark for the university).
12. . Find the mark range for each CS course (i.e. each result line/tuple has CS course name, lowest and highest mark obtained for that course).

13. . Find the highest "UCT Score" obtained each year, but only for years where there were more than 10 students who matriculated that year.

14. Which course(s) have the lowest number of fails ("F") i.e. which courses have the fewest failures?

15. . For each PSY2 student, show the number of senior Science courses they have passed.

16. . Devise a useful query of your own involving the most interesting SQL usage you can think of. Explain clearly what it is meant to find out from the database, then show the SQL and its result, and then show/explain clearly how you know it is correct.

17. . Delete all students who matriculated before the year 2000.

18. Insert a new result for 1001, who also took ZOO1 but was absent from the exam.

19. . The university has decided to use sum-of-6-best-matric-percentages as "UCT scores", instead of the old system. Change all "UCT Scores" in the database accordingly by multiplying them by 10 and adding 30 (as a rough approximation).

Note: make sure that you run 17-19 after 1-15, so that your results for 1-15 can be checked against the original spreadsheet data.

Marking rubric: 2 marks per question; except Q16 is out of 4 as it is based on how challenging the query and how clearly its correctness is shown. Total 40.