

University of Toronto Mississauga  
CSC 343 Winter 2018  
Group Assignment 2  
Due: Friday March 16th, 2018 by 9:00am

Michael Liut

February 12, 2018

For this assignment, you will continue to work with *Hamilton Street Railway (HSR)*, and perform data analytics over real data, tracking bus routes over the period May 1 - May 7, 2017. On Portal and the Course Website, under Assignments, you will find the following:

- a. `createTables.ddl`, a DDL file for CREATE TABLE statements.
- b. `loadData.ddl`, a DDL file for INSERT TABLE statements.
- c. `asg2ER.pdf`, a simplified schema depicted as an ER diagram.

Please execute scripts `createTables.ddl` and `loadData.ddl` on your MySQL database on the CSC server as you will use this schema for the questions below.

Please note: much of this data is real, including routes, stops, and associated times; however, for the purposes of this assignment, assume the questions to be true (even if the routes may not always be).

## I. Structured Query Language (65 marks)

Write and provide SQL statements for each of the 12 questions. Execute each of your SQL queries against your HSR database and give the result of each query.

- (q1) [3 mark] Identify all buses (busID, age, manufacturer) with advertising revenue greater than \$9,000.
- (q2) [3 marks] Find the number of students in the database. A student is defined as a person with a student occupation or is less than 25 years old (as of 12:00AM on March 16th, 2018). Do not include duplicates. (*Hint:* you may use the `date()` function.)
- (q3) [4 marks] Find the number of students who took bus route #5 on May 3rd, 2017?

- (q4) [4 marks] For each bus route, find the total advertising revenue. Return the bus route number and the total revenue. Order the results in descending order of total revenue.
- (q5) (a) [4 marks] Find all drivers who have less than 3 infractions. Return the drivers ID, first name and last name.
- (b) [5 marks] For each driver, return the total demerit points and total fines incurred. Do not include drivers with less than 2 demerit points in the result. Sort the result such that the most offending drivers (in terms of demerit points, total fines) are listed first.
- (q6) [4 marks] Determine those buses that are the unique (only) bus made by their manufacturer. Return the busID and the manufacturer.
- (q7) (a) [4 marks] For each passenger type, find the total fares' revenue. Return the passenger type and its associated revenue (the column must be named 'revenue').
- (b) [2 marks] Extend your query in part (a); only return passenger types, and their revenue, when the total revenue is greater than \$500.
- (c) [2 marks] Extend your query in part (a); return the most profitable passenger type (in terms of total fares' revenue) on May 1, 2017.
- (q8) (a) [4 marks] Determine the most popular bus route on May 7, 2017 (according to the number of passengers). Return the route ID and the number of passenger trips (column named 'times').
- (b) [4 marks] Which day contained the largest volume of passenger trips? Return the date and the number of trips taken.
- (q9) [4 marks] Find all the people who visited a library on either May 5, 2017 or May 6, 2017. Return their occupation. Do not include duplicates.
- (q10) [5 marks] Find the drivers who have worked with HSR for more than 5 years, with a salary greater than \$80,000, and with less than 10 demerit points on their driving record. Return the drivers first name, last name, and ID.
- (q11) [6 marks] Find all students who attended the "Marauders Tennis" match and arrived via a bus on route 4. Return the student's first name, last name, and their gender.
- (q12) [7 marks] Assuming that the bus schedule has not changed since May 2017. Suppose you would like to attend the Marauders Basketball game (an event). The game starts at 5PM, and youd like to arrive at the site between 4:20PM and 4:50PM, which routes can you take? Use May 1, 2017 as a reference date for the bus schedule information. List the route ID, the bus stop name (where you should get off the bus), and the scheduled arrival time.

## II. Relational Algebra (35 marks)

For each of the SQL queries in Part I, give the corresponding relational algebra expression.

## Grading

This is a group assignment to be completed in pairs (i.e. a team of 2 people). You will work with the same group as in Assignment 1. This assignment is worth 13.3% of your final grade in this course.

## Submission

All files are to be submitted using the Blackboard platform (portal.utoronto.ca). Only one person from each group is required to submit the files. Please ensure your answers are typed and submissions are clearly legible. Include your and your partner's full name and student ID number in all files. Upload three files with the indicated file extensions (no compression based `.tar`, `.zip`, `.rar` files).

- For Part I: Submit your SQL statements in a script file called `queries.sql`.
  - Ensure your SQL statements are syntactically correct and that they are executable on the CSC MySQL server. Non-executable queries will not be marked.
- For Part I: Submit the corresponding query results in a file called `queries.results`.
  - Clearly label and comment which query corresponds to which result tuples.
- For Part II: Submit your relational algebra expressions in a file named `ra.pdf`.
  - Your relational algebra must be typed. Handwritten scans will not be accepted.

**This means you will submit a total of three files: ra.pdf, queries.sql, and queries.results.**

Please note that late assignments will be docked 20% per day of lateness and after four (4) days, the assignment will no longer be accepted.

## Plagiarism

Please refer to the course outline and introduction slides. To serve as a reminder: Turnitin will be used for all written work and MOSS for all code submissions. UTM's policy on Academic Integrity: <http://academicintegrity.utoronto.ca/>