Principles of Urban Informatics
Assignment 6

Posted on: 10/13/2014
Due Date: 10/20/2014

Assignment Description

In this assignment we are going to continue to use SQL to explore MTA subway data as in the previous assignment.

You are encouraged to use the Web as a resource to find more information about MySQL. You can exchange ideas with your classmates, but the work you submit should be your own. Copying is not allowed.

In the following we present instructions to set up the database and the data for the assignment (Section Setting Up MySQL), describe the data being used (Section Data) and describe the problems that you need to solve and submit (section Problems).

Setting Up MySQL

Installing MySQL

Note: only install MySQL if you have not done it before for the previous assignment.

First, download mySQL from `http://dev.mysql.com/downloads`. There are many different versions, select MySQL Community Server.

MySQL has an extensive documentation, including installation instructions for different OSes, for example, MacOS: `https://dev.mysql.com/doc/refman/5.7/en/macosx-installation.html` and for Windows: `http://www.mysql.com/why-mysql/windows/`.

In the past, students had trouble installing the MySQL Startup Item, do not install this option! During the installation you can create a password for your database administrator, if you do this you need to remember the password for future use.

User and database creation

Now, you should execute the following steps (the following is an example on Linux):

- Login into mysql shell as root:

```
$ mysql -u root -p
<enter password, if no password leave blank>
mysql>
```

This indicates that MySQL is running.

- Create database: Notice, MySQL is a DB management system and can manage multiple databases. To create a database you should execute

```
mysql> create database assignment6;
Query OK, 1 row affected (0.00 sec)
```

- Create an user account: Notice, MySQL is a DB management system and can manage multiple users. To create an user account you execute

```
mysql> create user 'user_assignment6'@'localhost';
Query OK, 1 row affected (0.00 sec)
```

This is an user account without password that is going to be used for testing.

- Grant permission to *user_assignment6* to use the database *assignment6*: To do this execute

```
mysql> grant all privileges on assignment6.* to
user_assignment6@localhost;
Query OK, 1 row affected (0.00 sec)
```

- Finish: you can now quit from mysql shell:

```
mysql> quit;
Bye
```

Loading data

To create the tables used in the assignment and populate (insert data into) them, first download the dump file from: `http://vgc.poly.edu/projects/gx5003-fall2014/week6/lab/data/mta.sql` Then, from a terminal window invoke MySQL and issue the following commands:

```
$ mysql -u user_assignment6
mysql> use assignment6;
Database changed
mysql> source mta.sql
mysql> show tables;
+-----------------------------+
| Tables_in_assignment6       |
+-----------------------------+
| fares_feb1                  |
| fares_jan18                 |
| stations                    |
+-----------------------------+
3 rows in set (0.00 sec)
```

This shows that 3 tables were created (from the code that is inside the file *mta.sql*). If you are not running MySQL from the same directory where the file mta.sql is located, you need to specify the In case of doubt, send us an email.

Data

Note: this should not be new to you if you did the previous assignment, as no modification has been made to the data.

In this section we briefly describe the data used in the assignment. The data is stored in 3 different tables:

```
table fares_jan18(
 remote varchar(10),
 station varchar(100),
 ff int,
 sendis int(11),
 7d int(11),
 30d int(11),
 students int(11)
);

table fares_feb1(
 remote varchar(10),
 station varchar(100),
 ff int(11),
 sendis int(11),
 7d int(11),
 30d int(11),
 students int(11)
);
```

where:

- remote: id of station entrance
- station: name of station
- ff: number of full fares
- sendis: number of senior citizen / disabled fares
- 7d: number of 7 day pass fares
- 30d: number of 30 day pass fares
- students: number of student fares

```
table stations(
  name  varchar(100),
  lat float,
  lng float,
  line varchar(50),
  lines varchar(50)
);
```

where:

- name: name of station
- lat,lng: latitude and longitude of station
- line: name of the line (e.g. Broadway)
- lines: lines that go through the station (e.g. 1,2,3)

Problems

Your task is to write SQL queries to answer the following questions. Make sure the return of your query matches the names we supplied between brackets. For instance, in item 1, the result of your query must be a table with column nyc_ff_jan18. For more information, see `http://www.tutorialarena.com/mysql/mysql-as-keyword.php`.

1. How many full-fare tickets were sold in NYC in the week of Jan 18? [as nyc_ff_jan18]
2. List the names of the stations whose decrease in the number of full-fare tickets sold between Jan 18 and Feb 1 is greater than 1000. [as name]. You might need to trim trailing spaces for the station name.
3. What was the average difference in the number of full-fare tickets sold between Jan 18 and Feb 1 across the stations of the Broadway line? [as broadway_ff_avg_diff]
4. Which station had the largest increase in the number of full-fare tickets sold between Jan 18 and Feb 1? [as station_with_largest_increase]. You might need to trim trailing spaces for the station name.
5. Which station had the largest decrease in the number of full-fare tickets sold between Jan 18 and Feb 1? [as station_with_largest_decrease]. You might need to trim trailing spaces for the station name.

Grading

The grading is going to be done by running the queries on MySQL. Your query has to run (i.e., no syntax error) and return the right solution (otherwise you will get a 0 for the query).

Similar to the previous assignment, we provide you with a test code that automatically grades your submission and gives you your grade, in this way you will know your grade beforehand. Download and uncompress from `http://vgc.poly.edu/projects/gx5003-fall2014/week6/lab/data/assignment6.zip`.

Only if you have not done this before for the previous assignment, to execute the grading code you need to install MySQL-python `https://pypi.python.org/`

pypi/MySQL-python/1.2.5 (for more details see http://mysql-python.blogspot.no/2012/11/is-mysqldb-hard-to-install.html).

In the following we include examples of the execution of the grading code.

```
> python grade_assignment 1
Your grade in problem 1 is 10
Your output:
(12905094)
> python grade_assignment
  Your grade in problem 1 is 10
  Your grade in problem 2 is 10
  Your grade in problem 3 is 10
  Your grade in problem 4 is 10
  Your grade in problem 5 is 10
```

This code tests the query files (*query1.sql* to *query5.sql*) that are also provided. You should not change the test code, as we will use a fresh version to test your queries.

Questions

Any questions should be sent to the teaching staff (Instructor Role and Teaching Assistant Role) through the NYU Classes system.

How to submit your assignment?

Your assignment should be submitted using the NYU Classes system. Create a zip file with your SQL queries in the files *query1.sql, query2.sql, query3.sql, query4.sql* and *query5.sql* (**only these files**). Name the zip file as NetID_assignment_6.zip, changing NetID to your NYU Net ID.