INF 551 – Spring 2017

Homework #3

Due: March 12
100 points

1. [50 points] Consider again the Los Angeles International Airport (LAX) traffic data *lax.json*.
   a. [25 points] Write a Python script *load.py* with embedded SQL insert statement, to load the data in *lax.json* into a MySQL table defined as follows:

```
create table LAX (
       report_period datetime,
       terminal enum ('Imperial Terminal', 'Terminal 2',
'Terminal 1', 'Terminal 6', 'Terminal 7', 'Terminal 4',
'Terminal 3', 'Terminal 8', 'Terminal 5', 'Misc. Terminal',
'Tom Bradley International Terminal'),
       arrv_dept enum ('Arrival', 'Departure'),
       dom_inter enum ('Domestic', 'International'),
       passenger_count int,
       primary key (report_period, terminal, arrv_dept,
dom_inter)
);
```

You may refer to the below link for more details on the *ENUM* data type in MySQL:
https://dev.mysql.com/doc/refman/5.7/en/enum.html

Note: You should first run the above *"create table"* statement **manually** to create the table in the *"inf551"* database (the same one as you created in class), then create a user *"inf551"* with password *"inf551"* and grant access to *"inf551"* database to the user (see *mysql connector* slide). **Do not** include *create table* and *grant privilege* statements, which should be done **separately**, in your Python code other than insertion.

Your script should obtain the database connection using the following call (same as that in the slide).

```
cnx = mysql.connector.connect(user='inf551',
password='inf551',
host='127.0.0.1', database='inf551')
```

Your script takes only one argument (*lax.json*):
Execution: *python load.py lax.json*

Submission: <FirstName>_<LastName>_load.py

b. [25 points] Write a Python script *lax.py* that takes two keywords (that specify the search condition) as the input, and outputs the total number of passengers in the records that satisfy the condition.

   The first keyword is either "*Departure*" or "*Arrival*"; the second is a four-digit year, e.g., *2006*. There're **only** and **always** these two kinds of keywords.
   For example, "*python lax.py Arrival 2015*" will return the number of passengers that arrived in 2015.

   Your script should turn user search condition into an SQL query and execute it to obtain the result. The query will be executed on the LAX table you populated in the previous task.

   Submission: <FirstName>_<LastName>_lax.py

2. [50 points] In this question, we will use the Employees Sample Database for MySQL: https://dev.mysql.com/doc/employee/en/

   You need to first follow the instructions in the "*Set up Employees Sample Databases*" slides to create and populate the database.

   Use the database, write an SQL query for each of the following questions.

   a. Find name(s) of the employee(s) with the highest salary. Show names **in one column** in this format: *"last_name, first_name"*, i.e. "Smith, John" (**separated by comma and space**) rather than in two independent columns, which would be *"last_name", "first_name".* **Show employee names in this format in all following questions.**
   b. Find name(s) of the department(s) with the largest number of employees.
   c. Find out how many employees work only for a single department.
   d. Find out how many employees work for more than one department.
   e. Find out how many employees work for **both** Development and Production departments.
   f. Find out how many employees work for Development department but **not** for Production department.
   g. Find out how many male employees in each department (order result by department name, ascending).
   h. Find out the number of "Senior Engineer" in each department (order result by department name, ascending).
   i. Find out average salary of staff (including senior staff) in each department (order result by department name, ascending).
   j. List names of departments with more than 100 employees (order result by department name, ascending).

Submission: Submit a text file, *<FirstName>_<LastName>_sql.txt*, containing a list of SQL queries numbered by questions. Right after each query, copy and paste the query result obtained from MySQL. For example,

Question 2.a: SELECT COUNT(*) FROM Beers;

```
+----------+
| COUNT(*) |
+----------+
|        5 |
+----------+
```