

## Homework #3: SQL

Due: November 3, Sunday (end of day)

100 points

Consider again the LA Restaurants & Market Health data set available at Kaggle:

<https://www.kaggle.com/cityofLA/la-restaurant-market-health-data>. In particular, we consider the two CSV files: one for [inspections](#); the other for [violations](#).

1. [20 points] Write an SQL script “`firstname_lastname_load.sql`” that does the following:
  - a. Creates a table, “inspections” for the inspection data set; and a table “violations” for the violation data set. Your tables should be stored in a database called “inf551” with both user & password being “inf551”.
  - b. Loads the data in the csv files into the respective tables. You may refer to: <https://dev.mysql.com/doc/refman/5.7/en/load-data.html> for details on “load data” statement in MySQL.

Note that load.sql will assume the two data sets are located at the same directory in the name of “violations.csv” and “inspections.csv”, no need to give arguments from the command line.

查询语句

2. [50 points] Write an SQL query for each of the following questions. Submit the same in sql files, “`firstname_lastname_a.sql`”, “`firstname_lastname_b.sql`”, and so on.
  - a. Find out names of facilities whose name contains “cafe” (case insensitive) and had a violation with code “F030”.
  - b. Find out names of facilities that have the highest inspection scores.
  - c. Find out which facility (by id) has the largest number of violations. Output the names of such facilities (ascending order).
  - d. Find out which facilities that had inspections done but do not have any violations (as recorded in the violations data set). Output names of such facilities (ascending order).
  - e. For each different letter grade in inspections, output the average score of facilities receiving the letter grade.
3. [30 points] Write a Python script “`firstname_lastname_good.py`” that answers the question 2.d above. Note that your script should use Python MySQL connector to connect to the “inf551” database mentioned above. Output the results to a file whose name is specified in command line.

## INF 551 – Fall 2019

Execution format: `firstname_lastname_good.py` `output_file_name.txt`

Output format: One restaurant name per line.

Eg.

ALL INDIA CAFÉ

ANDY'S DONUTS

BIOBAR

...

**NOTE: Submit all files in a zip file named as `firstname_lastname_hw3.zip`**

### Grading Criteria

1. If your programs cannot be executed with the command specified above, there will be 40% penalty.
2. If your programs cannot be executed with the required Python version (Python 3), there will be 30% penalty.
3. If you use non-standard python packages (except for mysql-connector package), there will be 30% penalty.
4. If your .py takes more than 5 minutes for each to complete, there will be 20% penalty.
5. Please do not keep any "print" statements, they will lead to 10% penalty.
6. Please do not hard-code file name for Q3, else 10% penalty.
7. Please submit all files under 1 zip file in the format mentioned in then note.
8. Late homework will be deducted by 10% for every 24 hours that it is late. (no credit after 72 hours)