DSCI 551 – Spring 2022

HW3, 100 points

Due: March 11, Friday (end of day, 11:59pm)

You will need to install MySQL Sakila database for this homework. You can either install the database as described in https://dev.mysql.com/doc/sakila/en/; or you may follow these steps to install it on EC2.

- Download package:
  - wget https://downloads.mysql.com/docs/sakila-db.tar.gz
- Unzip it:
  - tar xvf sakila-db.tar.gz
- Install:
  - cd sakila-db
  - mysql -u root -p
    - source sakila-schema.sql
    - source sakila-data.sql
    - use sakila

Note that two source commands above need to be executed after you log into your MySQL as root.

```
[ec2-user@ip-172-31-15-144 ~]$ wget https://downloads.mysql.com/docs/sakila-db.tar.gz
--2022-02-06 19:27:28--  https://downloads.mysql.com/docs/sakila-db.tar.gz
Resolving downloads.mysql.com (downloads.mysql.com)... 137.254.60.14
Connecting to downloads.mysql.com (downloads.mysql.com)|137.254.60.14|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 732134 (715K) [application/x-gzip]
Saving to: 'sakila-db.tar.gz'

100%[======================================================>] 732,134     1.43MB/s   in 0.5s

2022-02-06 19:27:29 (1.43 MB/s) - 'sakila-db.tar.gz' saved [732134/732134]

[ec2-user@ip-172-31-15-144 ~]$ tar xvf sakila-db.tar.gz
sakila-db/
sakila-db/sakila-data.sql
sakila-db/sakila-schema.sql
sakila-db/sakila.mwb
[ec2-user@ip-172-31-15-144 ~]$ cd sakila-db/
[ec2-user@ip-172-31-15-144 sakila-db]$ ls
sakila-data.sql  sakila.mwb  sakila-schema.sql
[ec2-user@ip-172-31-15-144 sakila-db]$ sudo service mysqld start
Redirecting to /bin/systemctl start mysqld.service
[ec2-user@ip-172-31-15-144 sakila-db]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 8
Server version: 8.0.28 MySQL Community Server - GPL

Copyright (c) 2000, 2022, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
```

```
mysql> source sakila-schema.sql
Query OK, 0 rows affected (0.00 sec)
```

- Run the following command in mysql, if you haven't created a user named “dsci551” with password “Dsci-551” in mysql, refer to the posted lecture note on how to setup MySQL on EC2:

  GRANT ALL PRIVILEGES ON sakila.* TO 'dsci551'@'localhost';

- Download “hw3_grade.sh” from blackboard and put it in the directory (e.g. LASTNAME_FIRSTNAME_HW3) you are working on:
  a. cd LASTNAME_FIRSTNAME_HW3
  b. chmod 707 hw3_grade.sh

1. [70 points] Write an SQL query for each of the following questions.
   1) Find out how many films are rated ‘PG-13’ and last between 100 and 200 minutes. Your column names should look EXACTLY like:

```
+--------+
| Total |
+--------+
```

   2) Find first and last names of actors whose 2nd last letter of last name is ‘i’. (ex- Jolie, Davis). Your column names should look EXACTLY like:

```
+---------------+----------------+
| first_name | last_name |
+---------------+----------------+
```

   3) Find the title and length of the longest films. Your column names should look EXACTLY like:

```
+-------+-----------+
| Title | Length |
+-------+-----------+
```

   4) Find out how many films there are in each category. Output category name and the number of films in the category. Your column names should look EXACTLY like:

```
+--------------+-------+
| Category | Total |
+-------------+--------+
```

   5) Find ids of customers who have rented films at least 40 times. Return the same ids only once. Your column names should look EXACTLY like:

```
+------------------+
| Customer_id |
+------------------+
```

   6) Find first and last names of customers whose total payment exceeds $200. Your column names should look EXACTLY like:

```
+---------------+----------------+
| first_name | last_name |
+---------------+----------------+
```

7) Find first and last names of actors who have never played in films rated R. Your column names should look EXACTLY like:

```
+--------------+---------------+
| first_name | last_name |
+--------------+---------------+
```

8) Find out how many films are not available in the inventory. Your column names should look EXACTLY like:

```
+--------+
| Total |
+--------+
```

9) Find out how many actors who have the same first name but a different last name with another actor. Your column names should look EXACTLY like:

```
+--------+
| Total |
+--------+
```

10) Show the first name, last name, and city of the customers whose first name is either Jamie, Jessie, or Leslie. Order the result by first name. Your column names should look EXACTLY like:

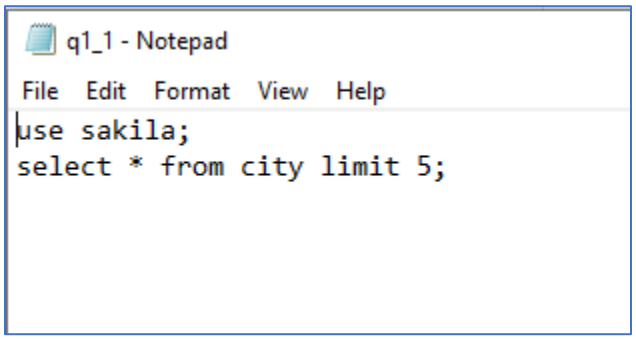
```
+--------------+--------------+------------+
| first_name | last_name | city       |
+--------------+--------------+------------+
```

Files to submit: A .sql file for **each** question. File name format = q1_1.sql, q1_2.sql, …… , q1_10.sql

Each .sql file should have 2 queries –

- use sakila;
- <<mysql query for that question>>

Inside of your .sql file should look like this:



2. [30 points] Write a Python script search.py that searches for customers using their first name (case insensitive). It should return first name, last name, and city of found customers.

For example, we will run the following query in the command line:

python3 search.py 'john'

(Note the quotes around the first name input.)

Example output format (note capitalization of each word):

JOHN FARNSWORTH PARBHANI

If no customer name was found, print:

Customer does not exist

(If there are multiple customers with the same first name, then print each customer in a new line.)

Use "dsci551" as username and "Dsci-551" as password.
Libraries permitted: pandas, sqlalchemy, pymysql, mysql-connector-python

File name: search.py

Submission

1. Your submission folder should contain **12 files** and look EXACTLY like this (PLEASE INCLUDE hw3_grade.sh, otherwise 10 pts will be deducted), any extra files like "README" will be ignored

```
[ec2-user@ip-172-31-44-49 Tandon_Utkarsh_HW3]$ ls
hw3_grade.sh  q1_10.sql  q1_1.sql  q1_2.sql  q1_3.sql  q1_4.sql  q1_5.sql  q1_6.sql  q1_7.sql  q1_8.sql  q1_9.sql  search.py
```

   Please understand how TA will run your sql files for q1 and q2.

   The TAs will simply run.

   ./hw3_grade.sh

   And then the command will generate a bunch of ".res" files. Then TAs will grade based on those ".res" files. If your filename is incorrect or your username and password is incorrect for the database, points will be deducted. Test your files with the given grading script before you submit. If you change a single byte in hw3_grade.sh, **50 pts** will be deducted.

   After running the grading script your directory should look **EXACTLY** like

```
[ec2-user@ip-172-31-44-49 Tandon_Utkarsh_HW3]$ ls
hw3_grade.sh  q1_10.sql  q1_1.sql  q1_2.sql  q1_3.sql  q1_4.sql  q1_5.sql  q1_6.sql  q1_7.sql  q1_8.sql  q1_9.sql  search.py
[ec2-user@ip-172-31-44-49 Tandon_Utkarsh_HW3]$ ./hw3_grade.sh
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
mysql: [Warning] Using a password on the command line interface can be insecure.
[ec2-user@ip-172-31-44-49 Tandon_Utkarsh_HW3]$ ls
hw3_grade.sh  q1_10.sql.res  q1_1.sql.res  q1_2.sql.res  q1_3.sql.res  q1_4.sql.res  q1_5.sql.res  q1_6.sql.res  q1_7.sql.res  q1_8.sql.res  q1_9.sql.res
q1_10.sql     q1_1.sql       q1_2.sql      q1_3.sql      q1_4.sql      q1_5.sql      q1_6.sql      q1_7.sql      q1_8.sql      q1_9.sql      search.py
```

   Do NOT submit the .res files

2. Put all files in the same directory and compress it into a zip file.

Zip file name format: **LASTNAME_FIRSTNAME_HW3.zip**

3. If you modify a column or delete a record or drop a table from TA's database, your homework will be graded 0.