

## 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)
```

Note that you can also download MySQL server and WorkBench from MySQL website and install them on your laptop/PC and use them for your homework/project.

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. *select count(\*) from film where rating='PG-13' and length between 100 and 200*
- 2) Find first and last names of actors whose 2<sup>nd</sup> to the last letter of last name is 'i'. *union, last\_name like "%i", limit 1, limit 1 order by actor\_id desc*
- 3) Find the title and length of the longest films. *title, length, where length = (select max(length) from ...)*
- 4) Find out how many films there are in each category. Output category name and the number of films in the category. *rating, count(film\_id), group by rating*
- 5) Find ids of customers who have rented films at least 40 times. Return the same ids only once. *customer\_id, rental left join customer on customer\_id = ~*
- 6) Find first and last names of customers whose total payment exceeds \$200. *payment*
- 7) Find first and last names of actors who have never played in films rated R.
- 8) Find out how many films are not available in the inventory.
- 9) Find out how many actors who have the same first name but a different last name with another actor.
- 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.

Submission: a text document named **sql\_queries.txt** that contains **both the queries and their results** (copy and paste your output from mysql terminal).

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. *→ 大写小写 (str.upper())*

For example,

*→ first name*  
python3 search.py 'john'

will find customers whose first name is john.

Libraries permitted: pandas, sqlalchemy, pymysql, mysql-connector-python

\*Note:

1. You should already have the database sakila in mysql at this point.
2. In order to use the package mysql.connector, you'll need to set up the mysql.connector.connect() in your code (Please refer to lecture slides for examples). If you haven't created the user 'dsci551', please do so by following the posted slides on how to setup MySQL on EC2. After set up, run the command below:

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

Submission: `search.py`

### Checklist for Submission :

1. DO NOT return anything we didn't ask for. For example, "please enter XXX: \_\_\_\_". Please no. We have given you the EXACT output format. Please just follow them.
2. Make sure that you are able to run the code according to the execution format mentioned above in the questions.
3. Double-check your files before submitting them. Please use python3 to complete the homework and try to maintain the python version as 3.7. Do not use any libraries other than the ones specified in the handout. You can use EC2 to test your code, and python 3.7 is preinstalled on EC2.
4. You can submit it multiple times on D2L but only the latest attempt will be graded.
5. You should submit TWO files to D2L this time: A text document **sql\_queries.txt** that contains your answer to Q1, and the **search.py** for Q2.

