CS 327E Project 2, due Thursday, 09/14.

1. Open a terminal window in JupyterLab and download the pagila dataset from Google Cloud Storage. Run the following commands to download and extract the dataset:
   ```
   gsutil cp gs://cs327e-open-access/pagila.zip .
   unzip pagila.zip
   ```

   Open the pagila folder and browse the two files `pagila-schema.sql`, `pagila-insert-data.sql`.

2. Create a new Python Jupyter notebook and name it `project2.ipynb`. Implement the following logic in your Jupyter notebook:

   - Create the pagila schema and objects by running `pagila-schema.sql`.
   - Populate the pagila tables by running `pagila-insert-data.sql`.
   - Describe each table by running `\d {table}` in psql.
   - Sign up for a Lucidchart account by following the steps in our guide.
   - Open Lucidchart and create an ERD with the information you gathered in the previous step. Make sure you include the keys, column names and data types for each relation and draw the proper relationships between the relations. Name the diagram `pagila-erd.pdf` and download it to your computer.
   - Write 3 queries that use an INNER JOIN between two or more tables. The queries should also include a WHERE clause and ORDER BY clause. Feel free to include a LIMIT clause if the output returns > 10 rows. Add a short Markdown comment above each SQL statement.
   - Write 3 queries that use an OUTER JOIN between two or more tables. The queries should also include a WHERE clause and ORDER BY clause. Feel free to include a LIMIT clause if the output returns > 10 rows. Add a short Markdown comment above each SQL statement.
   - Write 1 query that uses a self join on any table. Feel free to include a LIMIT clause if the output returns > 10 rows. Add a short Markdown comment above each SQL statement.

CS 327E Project 2 Rubric
Due Date: 09/14/23

Download and extract the pagila dataset to your jupyter notebook instance. **-5** no dataset or incorrect dataset found in Jupyter instance	5
Create a new Python Jupyter notebook named `project2.ipynb`. **-5** incorrect file name	5
Create the schema and database objects by running `pagila-schema.sql`. **-10** missing pagila schema **-2** for each missing database object up to -10	10
Populate the pagila tables from the `pagila-insert-data.sql` file. **-5** for each empty table up to -10 **-3** for each error in the data load	10
Describe each table in the pagila schema in psql using `\d {table}`. **-1** for each missing describe command	5
Create an ERD of the tables and relationships in the pagila schema. Include column names, data types, and keys. Draw proper relationships between the tables. **-5** for each missing entity type **-5** for each incomplete entity type (e.g. missing columns, data types, keys) **-3** for each missing or incorrect relationship	20
Write 3 SQL queries that use an inner join. Each query must also contain a `WHERE` clause and `ORDER BY` clause. Each SQL query should have a comment preceding it, describing its function.	20
Write 3 SQL queries that use an outer join. Each query must also contain a `WHERE` clause and `ORDER BY` clause. Each SQL query should have a comment preceding it, describing its function.	20
Write 1 SQL query that uses a self join. Add a comment preceding it, describing its function.	5
`project2.ipynb` pushed to your group's private repo on GitHub. Your project **will not** be graded without this submission.	**Required**
`pagila-erd.pdf` pushed to your group's private repo on GitHub. Your project **will not** be graded without this submission.	**Required**

`submission.json` submitted into Canvas. Your project **will not** be graded without this submission. The file should have the following schema:
 	
`{`
`"commit-id": "your most recent commit ID from GitHub",`
	Required

`"project-id": "your project ID from GCP"`
`}`
 	
Example:
 	
`{`
`"commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9",`
`"project-id": "some-project-id"`
`}`	

Total Credit:	**100**