ES 113: DCC

Assignment 2: Introduction to SQL

Total Points: **25**	**Submission deadline: 23:59:59 Hrs, ~~March 15, 2024~~ March 22, 2024**

Assignment Instructions

Please refer to the following assignment instructions:

1. Regarding the late submission, we will be following the penalty as per the table:

Late Submission	Penalty (Out of 100) [Scaled accordingly] to maximum points
Till 1-hour past deadline	5 points
1 to 12 hours past deadline	10 points
12 to 24 hours past deadline	20 points
24 to 36 hours past deadline	40 points
36+ hours past deadline	100 points

2. We will follow the zero plagiarism policy, and any act of plagiarism will result in a zero for the assignment.
3. No Generative AI tools such as ChatGPT are allowed; if found guilty of using these tools, serious actions will be taken.

Problem Statement

We will be using the complete The Movie Database in this assignment. Please download the database files from here. You need to import the database file into your local **MySQL** installation and then write queries for the following questions (Please look at references for the installation commands):

1. How many number of movies are available in the *moviedetails* table?
2. Which is the least popular movie?
3. Find the *pid* and their names involved in the production of the most popular movie (based on popularity).
4. List the count of jobs (unique) in each movie department from the *moviecredits* table. E.g.: For the ‘Directing’ department, there are nine unique jobs (Director, Assistant Director, etc.)
5. Count the number of rows in the full outer join of *moviedetails* and *moviecountry.*

6. How many movies are there whose popularity is more than '*Toy Story 2*'?
7. How many movies are there whose popularity is lesser than '*Toy Story 2*' and has popularity greater than 1?
8. Save the full outer join of *moviedetails* and *moviecountry* in table *moviedetailscountry* and find the movies whose *countryid* is 'US', along with their count.
9. Alter table to add a new column, *movie_release_date* (Date format data type), to the *moviedetails* table. The values have to be taken from the *release_data* (stored as varchar) column. To submit the output file for this alter command, print 10 tuples and the schema of the table (use *describe*).

Points Split:

1. 10 Points [Importing the SQL dump]
2. 10 Points [Using Join functions]
3. 5 Points [Exploratory Results and documentation]

Submission

1. You have to prepare a Word document where you answer the above questions and paste the screenshots of your output for every question.
2. You have to upload the PDF file from your Word document to this form. **No partial marks will be awarded for any of the questions.**
3. Please ensure that the TAs can view the PDF document before submitting the form. For this change, the general access of the notebook and spreadsheet to IIT Gandhinagar as shown below:

References

- To install MySQL, please refer to the installation guide for both Windows and Mac. We have added an installation video at the end of the document.
- Practice Queries are added here.
- **Importing the SQL Dump: here.**

FAQs

- If the SQL workbench is not running, it is fine to use the MySQL CLI.

- PLEASE REFER **here** for the MySQL dump import.
- In the Exploratory Results and Documentation part you need to put the screenshot and write a justification of how you got those results.