Data Set and Database Creation

You were already introduced to the dataset in the optional practice assignment of the Data Modelling module. You might have created the tables and inserted all the values for the entire schema. If you have completed this, you can verify your code and database from the SQL script file given at the bottom of the page.

For those who didn't attempt the optional Data Modelling Assignment, download the dataset given below and follow the given steps to insert all the values for the entire schema.

Note: It is highly recommended to create the SQL database on your local machine using the below dataset. However, note that this part is completely optional and will not carry any grades and if you wish, you can directly run the SQL script given at the end of this segment to load the dataset.

IMDb Dataset

Download

Steps to follow:

1. Download the IMDb dataset from above.
2. The first tab contains the ERD and the table details. Study that carefully and understand the relationships between the table.
3. Inspect each table given in the subsequent tabs and understand the features associated with each of them.

4. Open your MySQL Workbench and start writing the DDL and DML commands to create the database.

If you don't wish to perform the data loading part, you can directly download the SQL script file given below containing all the commands and data required for the database creation and start directly with the querying.

Grading Criteria

Criteria	Meets expectations	Does not meet expectations
Segment 1 (Q1-Q9) [~20%]	Problem statement task is correctly identified and executed. For example, suppose two tables are joined by inner join as needed in the question.	Problem statement task is not correctly identified and/or executed. For example, suppose two tables are joined by outer join instead of inner join.
	The solution output matches the expected output.	The solution output partially or doesn't match the expected output.
	The final output is in the desired format.	The final output is not in the desired format.
	The query is optimised.	The query is unoptimised.

	The query is syntactically correct.	**The query is syntactically incorrect.**
Segment 2 (Q10-Q17) [~25%]	**Problem statement task is correctly identified and executed. For example, suppose two tables are joined by inner join as needed in the question.**	**Problem statement task is not correctly identified and/or executed. For example, suppose two tables are joined by outer join instead of inner join.**
	The solution output matches the expected output.	**The solution output partially or doesn't match the expected output.**
	The final output is in the desired format.	**The final output is not in the desired format.**
	The query is optimised.	**The query is unoptimised.**
	The query is syntactically correct.	**The query is syntactically incorrect.**
Segment 3 (Q18-Q23) [~20%]	**Problem statement task is correctly identified and executed. For example, suppose two tables are joined by inner join as needed**	**Problem statement task is not correctly identified and/or executed. For example, suppose two tables are joined by outer**

	in the question.	**join instead of inner join.**
	The solution output matches the expected output.	**The solution output partially or doesn't matches the expected output.**
	The final output is in the desired format.	**The final output is not in the desired format.**
	The Query is optimised.	**The Query is unoptimised.**
	The query is syntactically correct.	**The query is syntactically incorrect.**
Segment 4 (Q24-Q29) [~25%]	**Problem statement task is correctly identified and executed. For example, suppose two tables are joined by inner join as needed in the question.**	**Problem statement task is not correctly identified and/or executed. For example, suppose two tables are joined by outer join instead of inner join.**
	The solution output matches the expected output.	**The solution output partially or doesn't matches the expected output.**

	The final output is in the desired format. **The Query is optimised.** **The query is syntactically correct.**	**The final output is not in the desired format.** **The Query is unoptimised.** **The query is syntactically incorrect.**
Coding Guidelines [~5%]	**Proper aliases are given.** **If required any, appropriate comments are written.** **If new variables are created, the names are descriptive and unambiguous.**	**Proper aliases are not given.** **Comments are not written rendering the code difficult to understand.** **Variables are poorly or ambiguously named.**

	The code is written concisely wherever possible. **Overall, code readability is good with appropriate indentations.**	**The code is more complex than what is required by the problem.** **Code readability is poor because of poor indentation / other reasons.**
Executive Summary (~250 words) [~5%]	**The structure of the narrative is clear.** **The content is concise and readable.**	**The structure of the narrative is haphazard.** **The content is not concise and difficult to understand.**

	The narrative is backed by facts.	**The narrative is not backed by facts.**

Submission

For submissions obtained within one week after the deadline, there will be a 30% penalty. Submissions beyond one week after the deadline will not be accepted.

You must go through these guidelines-

1. Make sure you have not made any changes to the original dataset provided to you. Your SQL code should work on the dataset given to you as part of the problem statement. You are not allowed to make modifications in the dataset using excel and then use it in your SQL code. Entire data processing must be done in SQL only. During grading we will be running your code on the dataset provided by us, in case your code gives errors with that, then marks will be deducted accordingly.
2. All penalties are automatically applied by the system based on time of submission. Hence, submissions that are late, even by a second, will attract penalties.

For e.g.- If the deadline is 28 October 2022, 11:59:00 PM IST, the submissions on 28 October 2022, 11:59:01 PM IST will attract a penalty of 30%. Hence we

recommend that assignments are submitted at least 30 minutes before the deadline to avoid any last-minute issues.

Also, note that all the deadlines are in IST (UTC +5.5), hence, if you are in a different time zone, then your deadline may vary according to local time. For eg - If you are in London and following BST (British Summer Time) which is UTC +1 then deadline for you in local time would be 7:29:00 PM BST when the deadline in India is 11:59:00 PM IST.