Problem Introduction

RSVP Movies is an Indian film production company which has produced many super-hit movies. They have usually released movies for the Indian audience but for their next project, they are planning to release a movie for the global audience in 2022.

The production company wants to plan their every move analytically based on data and have approached you for help with this new project. You have been provided with the data of the movies that have been released in the past three years. You have to analyse the data set and draw meaningful insights that can help them start their new project.

You are a data analyst and an SQL expert. You have to use SQL to analyse the given data and give recommendations to RSVP Movies based on the insights. For your convenience, the entire analytics process has been divided into four segments, where each segment leads to significant insights from different combinations of tables. The questions in each segment with business objectives are written in the script given below. You have to write the solution code below every question and submit the same SQL script file with the solution in the 'Submission' segment.

The SQL script file for the questions is given below.

RSVP Movies Question Script -Download

About the assignment

Where do I get the data from?

You can get detailed information about the dataset and database creation in the next segment.

Where do I write the code?

The SQL script file given above contains all the tasks you need to perform. You are required to write the code for each task/question below the question itself.

How do I submit the assignment?

Once you are done with writing the codes in the question script file, the same script file should be submitted in the **'submission'** segment.

Some tips before starting the assignment

1. Go through the data thoroughly before starting with the assignment. It will give you a good sense of what all the columns represent which is a good practice to follow before proceeding with the analysis. You should download the database script and run it in your local system.

2. Read through each of the instructions carefully, identify the task to be performed, and only then proceed to write the required code. Don't perform any incorrect analysis or look for information that isn't required for the assignment. The solution output should be in the expected format if given in the question.

3. Try using appropriate aliases for the variable names. Avoid using names: ABC, abc, etc.

4. Always try to write optimized codes. Use appropriate tables and joins as per the questions asked.

5. There are some checkpoints given in the question SQL script provided. Keep your eye out for them as they're useful pieces of information you can use to check if the result you have obtained after performing a particular task is correct or not.

Submissions Required

1. **SQL Script File:** The question SQL script file given with the final codes (answers) written in it. This will be your solution file.

2. **Executive Summary:** Write the executive summary consisting of the important insights derived from the data and some recommendations to be given to RSVP movies for their project. Make sure this summary doesn't exceed 250 words. Save this writeup as a PDF.

Put both these files in a zip folder and submit the final zipped file in the 'Submission' segment.

Data Set and Database Creation

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.

IMDb Dataset Import Download here

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. The solution output matches the expected output. The final output is in the desired format. The query is optimised. The query is syntactically correct.	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 partially or doesn't match the expected output. The final output is not in the desired format. The query is unoptimised. 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 final output is in the desired format. The query is optimised. The query is syntactically correct.	The solution output partially or doesn't match the expected output. The final output is not in the desired format. The query is unoptimised. 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 in the question. The solution output matches the expected output. The final output is in the desired format. The Query is optimised. The query is syntactically correct.	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 partially or doesn't matches the expected output. The final output is not in the desired format. The Query is unoptimised. 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 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.
Coding Guidelines **[~5%]**	Proper aliases are given. If required any, appropriate comments are written	Proper aliases are not given. Comments are not written rendering the code difficult to understand.
	If new variables are created, the names are descriptive and unambiguous.	Variables are poorly or ambiguously named.
		The code is more complex than what is required by the problem.
	The code is written concisely wherever possible.	
	Overall, code readability is good with appropriate indentations.	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 narrative is backed by facts.	The structure of the narrative is haphazard. The content is not concise and difficult to understand. The narrative is not backed by facts.