**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**

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.**

**IMDb Dataset Import**

**Download**

Let's move to the next segment.

**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 2018, 11:59:00 PM IST, the submissions on 28 October 2018, 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.

1. If you are 100% sure that you will not need to make any more changes in the assignment, **click “Submit for Grading”** to submit your assignment for evaluation.

**Important note: You must click “Submit for Grading” as the assignment will not be automatically submitted at the deadline.**

Here are the steps that you must follow during submitting any assignment-

1. Collect all the files (if there are multiple files) and compress them together.
2. Try to upload this compressed file latest by 11:30 PM
3. Download your submission and check that you have included all the required files.
4. Check that none of the files or the zip is corrupt. If it is found to be corrupt during grading, you will NOT be allowed to re-submit.

Create a zip file consisting of the SQL file and the PDF file and submit it below.