**Assignment: Problem Statement**

You have the data for the 100 top-rated movies from the past decade along with various pieces of information about the movie, its actors, and the voters who have rated these movies online. In this assignment, you will try to find some interesting insights into these movies and their voters, using Python.

This is a compulsory individual assignment wherein you will download a movie dataset, write Python code to explore the data, gain insights into the movies, actors, votes, ratings and collections, and submit the code.

About the assignment

**Where do I get the data from?**

You can download the movie dataset from the bottom of this page.

**Where do I write the code?**

Download the ‘Movie Assignment.ipynb’ file from the bottom of this page. It is a commented Jupyter IPython Notebook file in which all the instructions and tasks to be performed are mentioned.

**How do I submit the assignment?**

Upload your ‘ipynb’ solution file in the ‘submission’ section on the next page.

**Some tips before starting the assignment**

1. Go through the data dictionary 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.
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.
3. In some cases, the variable names have already been assigned, and you just need to write code against them. In other cases, the names to be given are mentioned in the instructions. We strongly advise you to use the mentioned names only.
4. Always keep inspecting your data frame after you have performed a particular set of operations.
5. There might be some subtasks which involve the use of functions which you may not have used before.  In such cases, you just need to do a simple search on Google/Stack Overflow to gain an understanding of the function. Please understand that solving this assignment is also a learning process and research is a part of this process. We have also provided some links directly in the notebook to make your search easier.
6. Always run the cells of the notebook sequentially/restart the kernel and run all the cells to avoid runtime errors. The number of commented cells provided in the notebook is an estimate based on the steps involved in each subtask. You can always add extra cells to the notebook for any additional steps you wish to perform.
7. For plot-related questions, please make sure you are using the appropriate chart size for better readability. Please refer to the links provided in the notebook (can also search for links on your own) in the notebook for better formatting of your charts. If the task involves the comparison of two charts at a time, plot them side by side using subplots to ease the comparison.
8. For the questions that ask you to write your inferences based on a plot, there are no fixed answers for these questions, however, we expect you to write some good pointers and it carries some marks.
9. There are some checkpoints given in the IPython notebook provided. They're just useful pieces of information you can use to check if the result you have obtained after performing a particular task is correct or not.
10. Optional subtasks are non-evaluative. However, we would advise solving those questions after you are done with the assignment as a good practice.