

# DS 5003 Project: Movie Analysis

## 1. Basic Problem Statement

1. You are provided with a CSV file containing detailed information about movies, including metadata such as titles, genres, cast, directors, budget, revenue, release dates, and more.
2. Perform the necessary data pre-processing and analysis on the given dataset to clean, normalize, and extract meaningful insights.
3. Design a relational database schema to store this movie data, organizing it in a way that optimizes queries and supports complex relationships between entities (such as actors, directors, production companies, etc.).
4. Develop a dashboard that can connect to the designed database and visualize various aspects of the data.
5. The dashboard must support two user roles:
  - Administrators: Can view, modify, update, and delete data.
  - Users: Can view and interact with visualizations but are not allowed to modify any data.

## 2. Analysis

You are required to analyze the data and answer the following questions as a baseline. Additional insights and custom analysis are encouraged to provide a deeper understanding of the dataset:

1. Who are the top 5 directors by the number of movies they have directed?
2. Which are the top 5 movies with the highest profit margin (revenue-to-budget ratio)?
3. Which directors have the most movies in the top 100 grossing films?
4. Which actors have the highest average vote in Sci-Fi movies?
5. For each director, which actor have they collaborated with the most, and what is the highest-grossing movie from their collaboration?
6. Is there a correlation between higher popularity scores and higher box office revenue across different genres?
7. Which actors have the highest difference between their average ratings in Drama vs. Comedy movies?
8. Which are the top 5 production companies by total revenue across different genres?
9. Which actors have the most appearances in high-rated movies across different genres?
10. Which genres have the most profitable sequels and prequels?

### **3. Deliverables**

#### **1. Database**

- a) Database should contain at least 3 tables
- b) Different roles and privileges for the database
- c) Accessing the database functions through a frontend will be commendable.
- d) Prepare at least five more queries that use different functions.

#### **2. Dashboard / Website (Website is optional but contain bonus)**

- a) User login (for website)
- b) Visualizing the output
- c) Statistical analysis
- d) User interaction

#### **3. Final report**

The report should have the following points :

- a) Any pre-processing you have done, with appropriate justification
- b) Relational design of your database including constraints, functions, triggers, roles and privileges

#### **4. Evaluation**

Marks will be provided for developing a unique and creative design.

- a) Demo [10] (group, slides and live demonstration expected)
- b) Report [5] (group)
- c) Viva [5] (individual, each team member is expected to be familiar with all
- d) aspects of the project including the code and database queries)