

# **Project Title: Exploratory Data Analysis on Film Data Using PostgreSQL**

## **Project Description:**

In this project, students will apply the SQL skills they have learned throughout the course to perform exploratory data analysis on the "film" database, which contains information about films, people, reviews, and roles. The project will focus on retrieving, filtering, and analyzing data to answer real-life business questions related to the film industry.

## **Project Objectives:**

- Apply SQL queries to retrieve and analyze data from the "film" database.
- Gain practical experience in using SQL for data analysis.
- Answer real-life business questions using SQL.
- Present findings in a clear and organized manner.

## **Project Tasks:**

### **Task 1: Data Retrieval**

1. Connect to the "film" database using PostgreSQL and PgAdmin.

### **Task 2: Film Analysis**

2. What are the top 10 highest-grossing films in the database, and when were they released?
3. How many films in the database were released in each country, and what are the top five countries?
4. How many films are available in each language, and what are the top three languages represented?
5. What is the average IMDb score for films in the database?
6. Which country has made the highest profit from movies? [Tip: To avoid a biased result, use average instead of sum.]
7. Which movie made the highest profit in the 21st century?
8. How many people in the database are still alive (based on birthdate and death date)?
9. Which year has the highest number of movie releases?

### **Task 3: People and Roles Analysis**

10. Determine the top 10 people with the most roles in the database.
11. Who are the top 10 actors or directors with the most roles in the database?
12. Identify how many people in the database are still alive.

### **Task 4: User Reviews and Votes Analysis**

13. Calculate the average number of user and critic reviews for films.
14. Identify films with the highest number of user and critic reviews.
15. Which films have the most Facebook likes, and is there a correlation with their IMDb scores?

### **Task 5: Presentation and Report**

- Create a report or presentation summarizing their findings. If you want to create slides, you can use either PowerPoint or Google Slides. If you want to create a report, it should be in PDF format.
- The report should contain a screenshot of the SQL code and another screenshot of the result. Then you explain your findings briefly.
- Create a visualization using Tableau to show your key findings.

