

# Capstone Project: Scraping IMDb's Top Movies

## Project Overview

In this project, you'll collect and clean movie data directly from **IMDb**, one of the world's most popular movie databases.

You'll use **web scraping** to extract movie titles, ratings, and years from IMDb's **Top 250 Movies** page and then load your cleaned data into **PostgreSQL**.

By the end, you'll have a clean, structured table of the **Top 20 movies** with their corresponding IMDb ratings and release years.

## Learning Objectives

By completing this project, you'll:

- Practice **static web scraping** using BeautifulSoup.
- Learn to clean scraped text and numbers in pandas.
- Understand how to transform raw HTML data into tabular format.
- Load your cleaned data into a **PostgreSQL database**.

## Data Source

**Website:** [IMDb Top 250 Movies](#)

This page contains a table listing IMDb's top-rated movies. Each row includes:

- Movie title
- Release year
- IMDb rating

## Deliverables

Each student should submit:

1. **Python script or Jupyter Notebook (.py / .ipynb)** containing:
  - Web scraping code
  - Data cleaning steps
  - Database loading code
2. **Screenshot** showing the data successfully loaded in PostgreSQL.