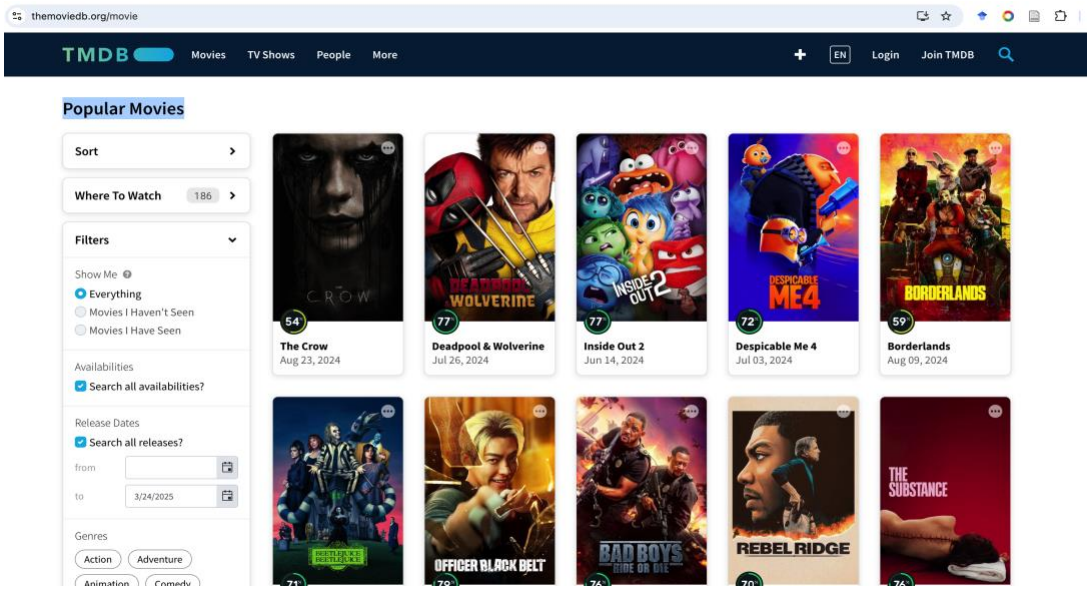
Assignment 1: Web Scraping & Data Analysis

Sep 31, 2024

In this assignment, you should work with data from
https://www.themoviedb.org/movie (Online Popular Movies Platform)

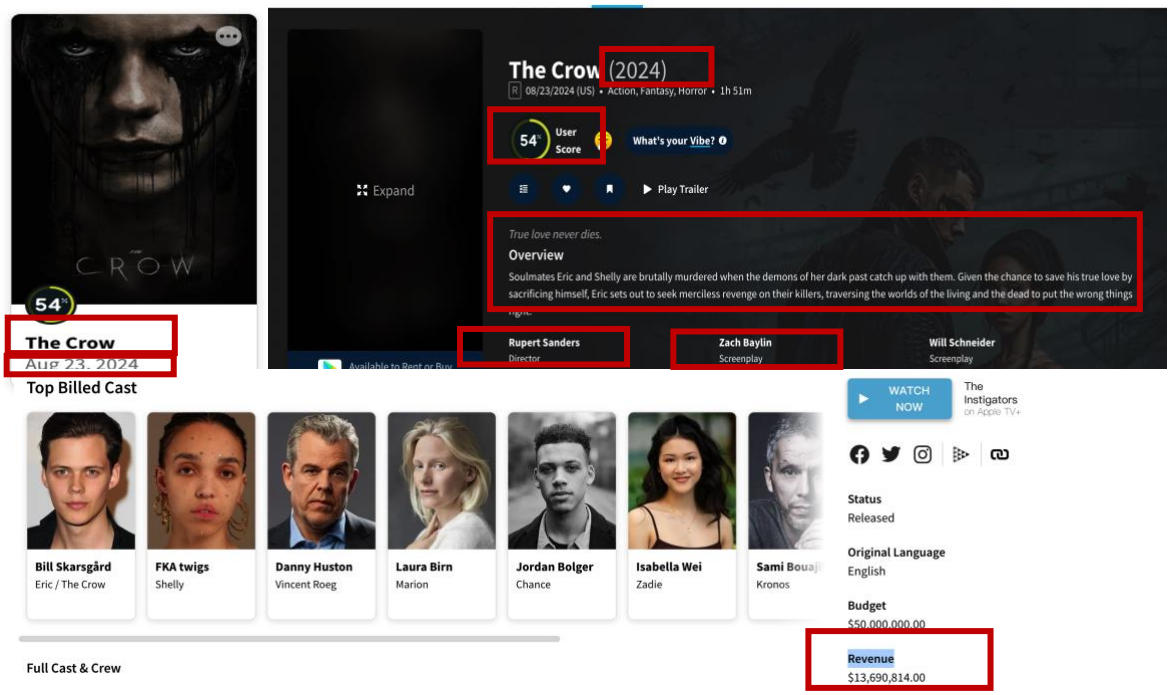


The Movie Database (TMDb) is a popular platform for movie enthusiasts, offering a vast collection of movies from all genres and regions. TMDb provides users with detailed information such as movie titles, release dates, cast, crew, genres, ratings, and more. It's a go-to source for finding information about both classic and upcoming films, as well as the latest in TV shows.

Everyone is interested in great movies, but with so many films released each year, how can we find the best ones? Scraping high-quality data from movie websites is crucial. In this project, we will utilize the skills we've learned with requests and regular **expressions to scrape essential movie details from The Movie Database (TMDb) website, allowing us to build a comprehensive dataset for further analysis.**

Task1. You are required to **scrape 200 Movies** from the website and save result into 'your_name+id.csv'. This file should contain data with the following columns: **(40 marks)**

Title of Movie	5 marks
Year	5 marks
User Score	5 marks
Description	5 marks
Director	5 marks
Screenplay	5 marks
Type	5 marks
Revenue	5 marks



You are free to explore data **with more properties** if needed.

Task2. You are required to do a data analysis on the data. What do you think is interesting about this data? Tell a story about some interesting thing you have discovered by looking at the data. **(60 marks)**

For example, which one is the best movie you might watch? Does the type of movie affect movie sales? Which category of movies sells the best?

Note: This is an open topic project. You are required to provide a novel topic and demonstrate your hypotheses (view points) with **data analysis and figures illustrations**.

The reports and running code (web scraping + data analysis) should be submitted using Jupter Notebook file.

Submission Checklist:

Yes/No	Items
	Jupyter Notenook code
	your_name+id.csv

Marking Guidelines

Marking Criteria	
Idea (5 marks)	• Presents a novel idea • Clearly demonstrate your viewpoints. • Demonstrates good understanding of the topic.
Discussion (30 marks)	• Provide convincing arguments to your viewpoints. • Backs up arguments with appropriate data analysis results. • Visualize data analysis results by • using more than 5 figures.
Organization (20 marks)	• Use of figures to support ideas discussed in the report. • The quality of the figures. • These figures should be informative. • Use of sub-titles and/or clear topic sentences. • Use multiple visualization methods (line, bar, pie chart, etc,).
Writing Style (5 marks)	• Concise writing style Strong scientific writing without grammatical errors.