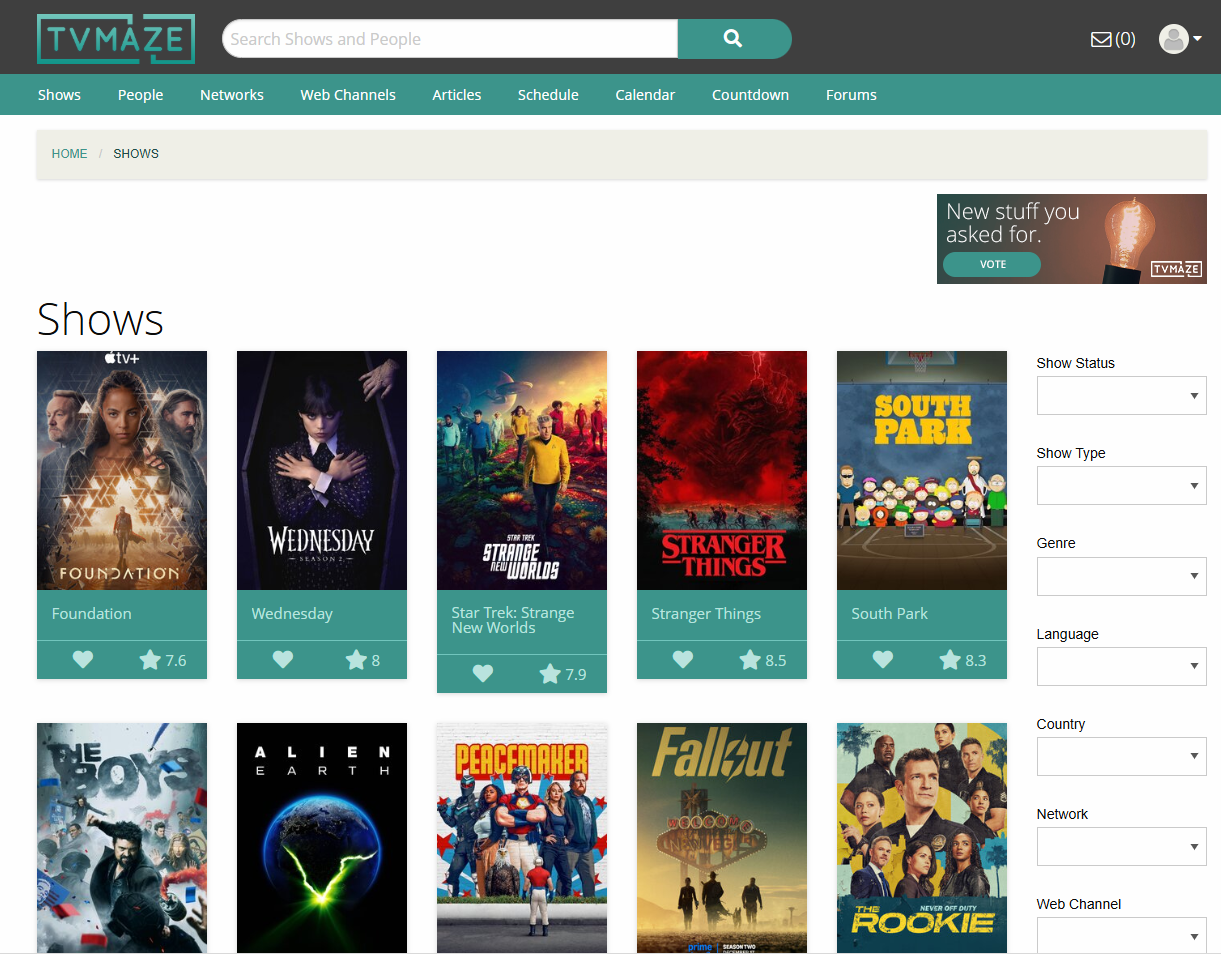
**Assignment 1: Web Scraping & Data Analysis**

Sep 9, 2025

In this assignment, you should work with data from

*https://www.tvmaze.com/shows*（Online Popular TV Shows & Movies Platform）



TV Maze is a well-known platform (with access API) for film and television enthusiasts, featuring a comprehensive library of TV shows (primary focus) and movies across genres, regions, and release eras. TV Maze provides users with detailed information including show/movie titles, premiere dates, cast & crew lists, genres, user ratings, popularity scores, content ratings, and plot descriptions. It serves as a key resource for accessing details about classic TV series, ongoing shows, upcoming releases, and standalone movies.

Everyone is interested in high-quality TV shows and movies, but with thousands of titles released annually, how can we identify the best ones? Scraping structured data from reliable entertainment platforms is essential. In this project, we will utilize the skills we've learned with requests and regular expressions to scrape key show/movie details from the TV Maze website (or its open API, for stability), allowing us to build a comprehensive dataset for further analysis.

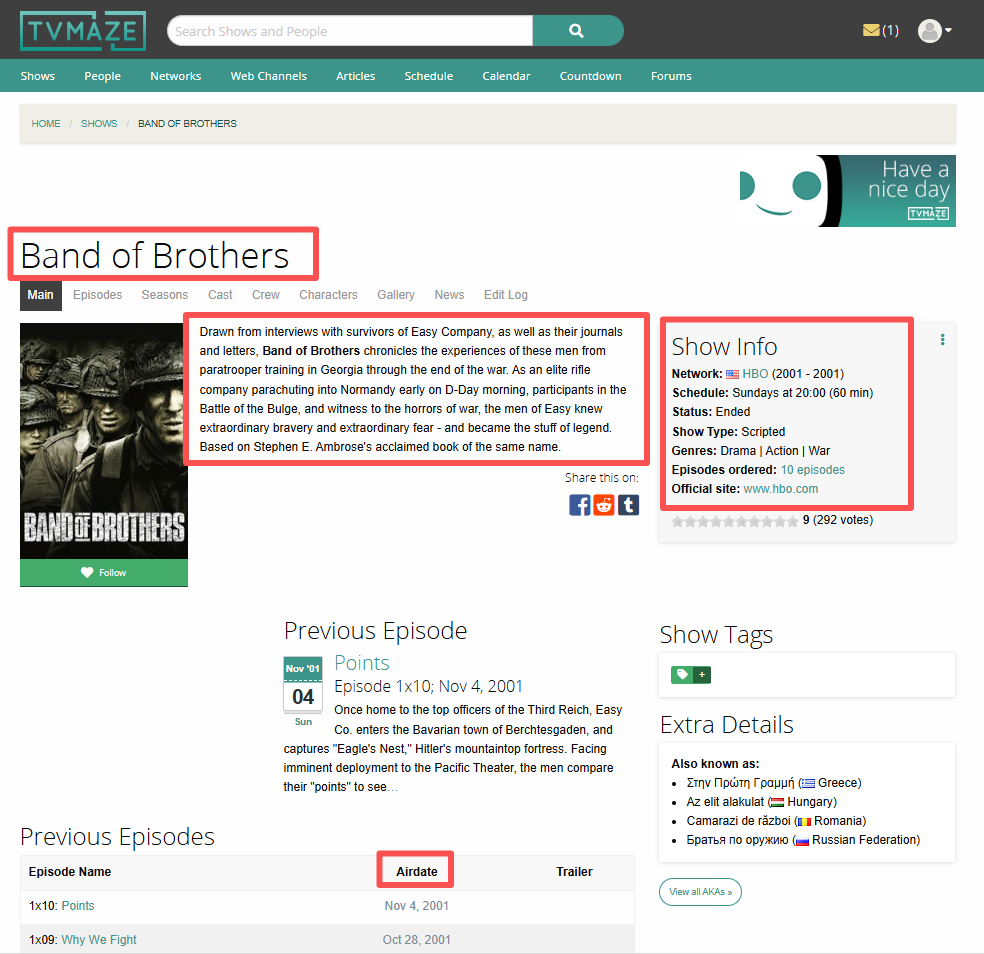
***Task 1: Web Scraping (40 marks)***

You are required to scrape 200 titles (TV shows or movies) from the website and save the result into your\_name+id.csv. This file must contain data with the following columns (each column accounts for 5 marks),you are also free to explore data **with more properties** if needed:

|  |  |
| --- | --- |
| Title | 5 marks |
| First air date  (or other property you think is important for subsequent mining) | 5 marks |
| End date | 5 marks |
| Rating | 5 marks |
| Genres | 5 marks |
| Status | 5 marks |
| Network | 5 marks |
| Summary | 5 marks |

Below is an example (with screenshot):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Title | First air date | End date | Rating | Genres | Status | Network | Summary |
| Band of Brothers | 2001/9/9 | 2001/11/4 | 9 | ['Drama', 'Action', 'War'] | Ended | HBO | Drawn from interviews with survivors of Easy Company, as well as their journals and letters... |



***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 Tv show you might watch? Does the type of show affect show rating? Which category of show rates the best?

*Note: This is an open topic project. You are required to provide a novel topic and demonstrate your hypotheses (viewpoints) 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)** | 1. Presents a novel idea 2. Clearly demonstrate your viewpoints. 3. Demonstrates good understanding of the topic. |
| **Discussion (30 marks)** | 1. Provide convincing arguments to your viewpoints. 2. Backs up arguments with appropriate data analysis results. 3. Visualize data analysis results by using more than 5 figures. |
| **Organization (20 marks)** | 1. Use of figures to support ideas discussed in the report. 2. The quality of the figures. 3. These figures should be informative. 4. Use of sub-titles and/or clear topic sentences. 5. Use multiple visualization methods (line, bar, pie chart, etc, ). |
| **Writing Style (5 marks)** | 1. Concise writing style 2. Strong scientific writing without grammatical errors. |