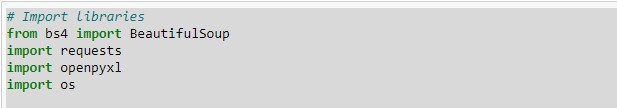
**Scraping IMDb Top Rated Movies and Saving to Excel**

I was tasked by Hamoye to build a project by scraping data from the web using Python libraries. And my project must contain a storytelling report and code implementation that shows how the data is obtained, analyzed, and processed.

I decided to explore the IMDB website to view their top-rated movies so far…. I decided this would be a nice project, scraping it and building visuals to gain insights. Below are steps I took to perform the web scraping and a preview of my dashboard.

Step 1: Import Libraries

First, import the necessary libraries for web scraping and working with Excel files.



Step 2: Create Excel Workbook and Add a Sheet

Create an Excel workbook and add a sheet named 'Top Rated Movies'. We'll use this sheet to store the movie data.

A picture containing text, screenshot, font

Description automatically generated

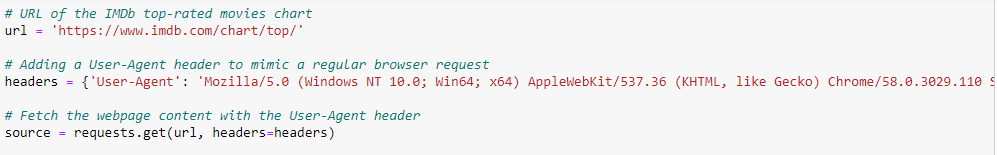
Step 3: Set Current Working Directory (Optional)

You can set the current working directory to a specific path where you want to save the Excel file. This step is optional. If you skip this, the file will be saved in the default directory.



Step 4: Fetch IMDb Top Rated Movies Chart

Next, we'll fetch the IMDb top-rated movies chart webpage using the **requests** library and parse it using **BeautifulSoup**.



Step 5: Parse HTML Content

Check if the request was successful, and if so, parse the HTML content using **BeautifulSoup**. We'll extract the movie information from the webpage.



Step 6: Extract Movie Information and Save to Excel

Loop through each movie row, extract the movie rank, name, year, and IMDb rating. Then, print the movie details and append them to the Excel sheet.

A screenshot of a computer

Description automatically generated with medium confidence

Step 7: Run the Script

Finally, run the Python script. It will fetch the IMDb top-rated movies chart, extract the required movie information, print it to the console, and save the data to an Excel file named 'IMDB Top Rated Movies.xlsx'.

A screenshot of a computer

Description automatically generated with medium confidence

That's it! You now have a script that can scrape the IMDb top-rated movies chart and store the movie details in an Excel file.

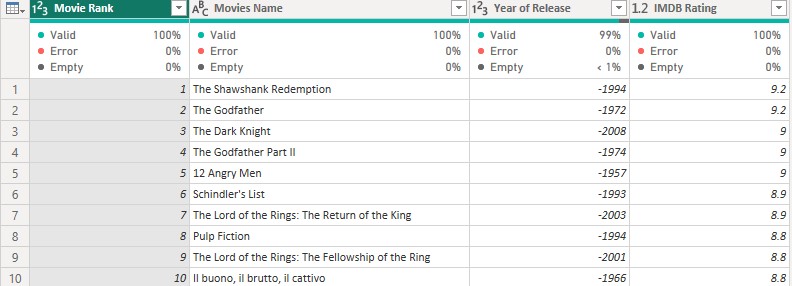
**Excel Preview of our dataset**

A screenshot of a computer

Description automatically generated with medium confidence

I will be taking my dataset to Powerbi to clean, analyze and visualize.

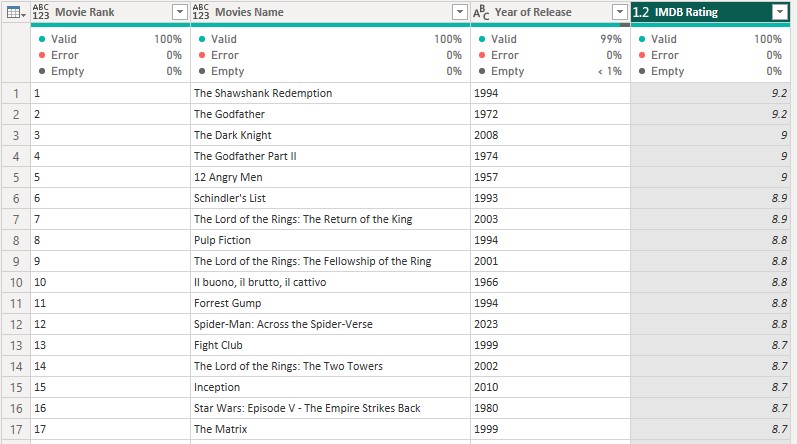
**Importing to Power query**



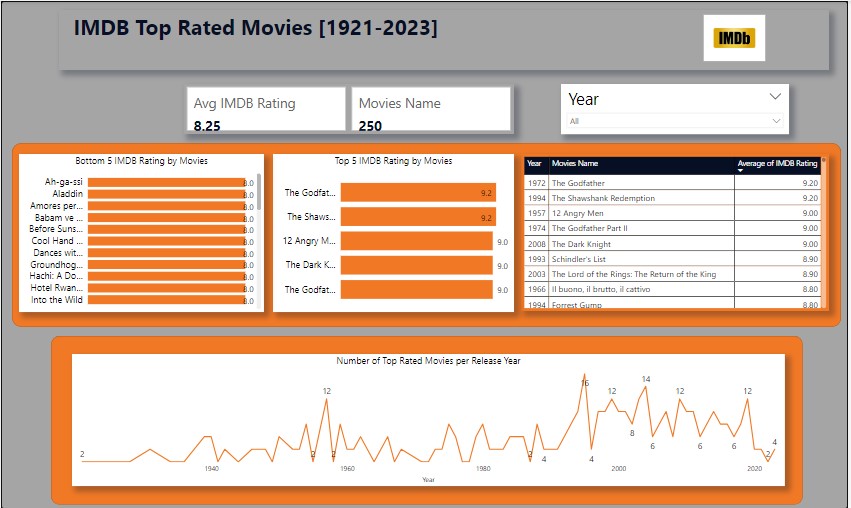
After importing I discovered that the power query changed my ‘Year of Release’ by removing the brackets and adding a negative sign to the year. This could be a result of converting the Years to numeric data.

I will fix this issue by going to the applied steps and deleting the data type change and my Year is back with brackets. Then I will use the Find and Replace to clean off the brackets. Then convert the Rating to decimal type, in order to allow me perform aggregation.

**Cleaned Dataset**



**Dashboard Preview**



**Insight:**

Movie Production Over the Years: The data shows that there has been a decline in the number of top-rated movies over the years. In 1995, there were 16 movies in the top-rated list, while in 2023, there are only 4 movies. This could indicate a change in movie production trends, viewer preferences, or a shift in the film industry landscape.

Classic Movie with High Rating: "The Godfather" released in 1972 stands out with the highest rating of 9.20. This classic movie continues to be highly regarded by audiences and critics alike, showcasing the enduring impact of certain timeless films on popular culture.

Recent Movie with High Rating: "Spider-Man: Across the Spider-Verse" released in 2023 also has a noteworthy rating of 8.80. This indicates that recent movies can still receive critical acclaim and perform well among top-rated movies, despite the lower number of recent releases in the list.

**Recommendations:**

**Promote Classic Movies**: Based on the high rating of "The Godfather" from 1972, film enthusiasts and streaming platforms could consider promoting classic movies to introduce newer audiences to timeless cinematic masterpieces.

**Foster Diversity in Movie Production**: With the decline in the number of top-rated movies over the years, the film industry could focus on fostering diversity in movie production to bring in fresh and unique storytelling. Encouraging diverse perspectives may lead to more top-rated movies that resonate with a broader audience.

**Focus on Quality:** The example of "Spider-Man: Across the Spider-Verse" demonstrates that recent movies can still receive high ratings. Filmmakers and studios should continue to prioritize quality storytelling, compelling characters, and technical excellence to produce top-rated movies.

**Analyze Audience Preferences**: Understanding the reasons behind the changes in top-rated movies over the years requires deeper analysis of audience preferences and changing trends. Analyzing viewer feedback, ratings, and engagement data can provide valuable insights for the film industry to adapt and create content that resonates with audiences.