Project Title:

Fetching movie information from Metacritic for the top 500 most popular movies of all times.

Objective:

The objective of this project was to fetch the movie information for the top 500 most popular movies from Metacritic website, extract the list of main cast and genre information of each movie and build a dictionary containing this information. Additionally, the main program should ask users for three choices: to display cast and genre information for a movie, display all movies for an actor/actress and summary of their genres, or calculate the cosine similarity between two actors/actresses based on the genre of movies they have acted in.

Project Completion:

To complete this project, I have written a Python script to extract the required information from the Metacritic website. The script collects movie information for the top 500 movies from the website and extracts the list of main cast and genre information of each movie. The collected information is then stored in a dictionary and saved in a CSV file as **“metacritics\_project analysis.csv”**

Next, I have developed the main program which allows users to choose from three options:

1. Movie: If the user chooses this option, the program asks for the name of a movie, and displays the cast and genre information for that movie.
2. People: If the user chooses this option, the program asks for the name of an actor/actress, finds all movies that this person has acted in, and displays a summary of the genres of these movies.
3. Comparison: If the user chooses this option, the program asks for the names of two actors/actresses, and calculates their cosine similarity based on the genre of movies they have acted in.

I have used various Python libraries such as Beautiful Soup, Pandas, NumPy, and scikit-learn to complete this project. The Beautiful Soup library was used to extract the required information from the Metacritic website. The Pandas and NumPy libraries were used to organize the collected data into a dictionary and save it in a CSV file. The scikit-learn library was used to calculate the cosine similarity between actors/actresses.

Conclusion:

In conclusion, I have successfully completed the project by fetching the movie information for the top 500 most popular movies from the Metacritic website, extracting the list of main cast and genre information of each movie, and building a dictionary containing this information. Additionally, the main program allows users to display cast and genre information for a movie, display all movies for an actor/actress and summary of their genres, or calculate the cosine similarity between two actors/actresses based on the genre of movies they have acted in. Overall, this project has provided me with valuable experience in web scraping, data organization, and implementation of machine learning algorithms in Python.