**Scraping MetaCritic**

**Introduction:** The libraries requests is used to fetch data from the website of metacritic and beautifulsoup is used to process the website html elements and access them easily. The processed data is saved to a csv file and a dataframe is created from the same. The dataframe methods are implemented to do the searching and calculating the cosine score of the parameters.

**Code Explanation:**

1.Fetching : The 5 pages for the 500 rows are fetched in a loop and added to the tr list using requests and beautiful soap. The director and genre data are in separate routes hence a new helper function is created to fetch those details.

2.Creating Dictionary from data: Iterating over obtained data and processing for the data of "Number","Movie","Genre","Date","UserScore","MetaScore","CertRating","Director" and adding them to the dictionary.

3.Csv File and dataframe: Using csv.DictWriter() and passing the header names to form a csv file with data dictionary and creating a dataframe using pandas library.

4.Main Menu: Running a loop for user menu until user exits. The options of search by movie , director and get cosine score is done in main() function by passing in the dataframe created above. The main menu gives three options 1.movie 2.director and comparison. Any other input by user is considered as invalid and looping is continued to present menu and ask input again from user.

**Comparison of three directors** –

**Genres:**

Francis Ford Coppola : genres are Drama:6,Thriller:3,Crime:2,Action:2,War:3,Mystery:2

Orson Welles : Drama:3,Mystery:1,Thriller:1,Crime:1,Film-Noir:1,History:1,Comedy:1,War:1

Alfred Hitchcock : Mystery:8,Thriller:11,Romance:4,Drama:2,Film-Noir:3,Action:1,Adventure:1,Horror:2,War:1,Crime:1

**Cosine Scores:**

Francis Ford Coppola and Orson Welles have a cosine similarity score of: 0.61397694673677.

Orson Welles and Alfred Hitchcock have a cosine similarity score of: 0.06756756756756757.

Francis Ford Coppola and Alfred Hitchcock have a cosine similarity score of: 0.26691730642466815.

**Conclusion**: As we can see the cosine score of the first two directors is > 0.5 showing a significant similarity in their genres. For the second and third directors its 0.06 can be assumed that there is no similarity of genres between them, The third and first had 0.26, showing very poor similarity. This shows us that the common genres of the first two directors are not at all common with third one as for second and third its almost zero cosine score and third and first common genres are different from the common ones of those matching first and second.