Task 3

Task: Movie Data Web Scraping and Analysis

Description:

You are tasked with scraping movie data from two different websites, MovieMeter and top250, and performing some basic analytics on the retrieved data.

This task will be performed as teams. The team pairing is found at the end of the document.

You are required to use GitHub to merge your work. Try to increase the number of commits based on the work done.

The code is to be written as functional python script (Do not use Jupyter notebooks).

Requirements:

1. Scraping Movie Data:

- Website 1 - MovieMeter:

- Scrape the top 250 best movies of all time from [MovieMeter's Top 250] (https://www.moviemeter.com/movies/top-250-best-movies-of-all-time) page.

- Retrieve movie titles, alt. titles, genres, release year, movie rank and ratings.

- Website 2 - 250Films:

- Scrape the IMDb Top 250 list from [250Films' IMDb Top 250] (http://top250.info/charts/?2023/09/25) page.

- Retrieve movie titles, years of release, movie rank and IMDb ratings.

- Implement error handling for web scraping in case of network issues or changes in the website structure.

2. Data Analysis:

- Combine the data collected from both websites into a single dataset. (Combine only movies found in both websites).

- Perform the following analytics tasks:

- Identify the movie(s) with the highest IMDb and moviemeter rating.
- Identify the movie(s) with the lowest IMDb and moviemeter rating.
- Identify the most found genre in the list.
- Count the number of movies released in each decade (e.g., 1950s, 1960s, etc.).
- Create a bar chart to visualize the distribution of ratings (bins: 0-2, 2-4, 4-6, 6-8, 8-10).
- Create a bar chat to visualize the distribution of movie genres.

3. Reporting:

Create a summary report that includes the following:

- Any required libraries or dependencies.
- The findings from the data analysis. (Results of analytics tasks)
- Include visualizations to help illustrate the results.
- Present the report in a clear and organized format.
- Save the retrieved data and analysis results to a CSV file. (Merged dataset)

Submission:

Provide the GitHub project link which should include the following:

1. Python script(s) for scraping data and performing analysis.
2. A report summarizing the findings and visualizations.

Pairs

Philo - Yousef

Ahmed - Nada

Aliaa - Habiba