Hollywood Actors and Actresses

Bararu Robert Daniel 10 December 2024

Contents

1	Loading in of Data			
	1.1	Data Loading Method	. 3	
	1.2	Data Structure:	. 3	
	1.3	Integration with IMDbPY	. 4	
2	Pseudocode			
	2.1	About the actor/actresses	. 5	
	2.2	Average rating of their movies	. 6	
	2.3	Movie genre of actor/actresses	. 7	
3	Solution Examples			
	3.1	About the Actor/Actress	. 8	
		Average rating of their movies		
	3.3	Movie genre of actor/actresses	. 10	

1 Loading in of Data

The data for this project, consisting of the top 50 actors and actresses, is sourced from IMDb (Link), a comprehensive and trusted movie database. To streamline access to this data, the IMDbPy module was employed, allowing for programmatic interaction with IMDb to retrieve additional details about actors and actresses.

1.1 Data Loading Method

To obtain the dataset, IMDb provides a downloadable CSV file via its website, which is easily readable and manageable for data analysis purposes. The Pandas library is used to load this file into the program, leveraging its read_csv() method to create a DataFrame, a versatile data structure that simplifies data manipulation and extraction.

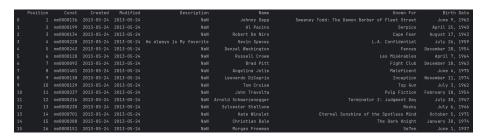


Figure 1: Example of the first 15 actors from the loaded DataFrame

1.2 Data Structure:

The dataset follows a standardized format and contains the following fields:

- **Position:** The actor/actress's ranking in the list.
- Const: A unique identifier for the actor/actress on IMDb. The first two characters are always nm.
- Created: The date the actor/actress was added to the list.
- Modified: The last time the actor/actress entry was updated in the list.
- Description: Description of the actor/actress
- Name: The full name of the actor/actress.
- **Know For:** The movie or project for which the actor/actress is most recognized.
- Birth Date: The actor/actress's date of birth.

1.3 Integration with IMDbPY

To retrieve the remaining data such as biographies, filmographies, ratings, and genres, we utilize the IMDbPy Python library. This tool allows for programmatic interaction with IMDb to fetch detailed actor-specific information directly. It streamlines the process of extracting comprehensive data, enabling more efficient and automated analyses within the project. The integration of IMDbPy is crucial for handling large datasets and extracting valuable insights regarding the actors career and filmography.

2 Pseudocode

2.1 About the actor/actresses

This algorithm outlines the process of retrieving and displaying information about actors and actresses, allowing the user to choose between viewing details for all actors or a specific one.

Algorithm 1 Retrieve Actor Information

```
1: Start
 2: Importing the necessary libraries for data handling
 4: Load the dataset from a CSV
 5: Initialize the IMDbPv object
7: Display action option
 8: if User chooses "About actor/actress" then
       Display options
9:
       if User chooses "All actors/actresses" then
10:
          for each actor in the dataset do
11:
              Extract the IMDb ID from the 'Const' column
12:
13:
              Fetch actor information from IMDb using the IMDb ID
              if information is successfully retrieved then
14:
                 Print actor's name and biography
15:
                 Print the top movie
16:
              else
17:
18:
                 Print an error message
              end if
19:
          end for
20:
       else if User chooses "A specific actor/actress" then
21:
22:
          Prompt user to enter the actor's name
          Fetch the actor's information
23:
24:
          if information is successfully retrieved then
              Print actor's name and biography
25:
              Print the top movie
26:
          else
27:
              Print an error message
28:
          end if
29:
       end if
30:
31: else
       Print "Invalid action selected"
32:
33: end if
34: End
```

2.2 Average rating of their movies

This algorithm calculates and visualizes the yearly average movie ratings for a chosen actor.

Algorithm 2 Retrieve and Analyze Movie Ratings

```
1: Start
 2: Importing the necessary libraries for data handling
 4: Load the dataset from a CSV
6: Initialize the IMDbPy object
7:
8: Display action options
 9: if User chooses "Average rating of movies" then
       Prompt user to enter the actor's name
10:
       Fetch the actor's information using IMDbPy
11:
       Extract the 'filmography' from the actor's information
12:
13:
       Initialize an empty list to store movie ratings and years
       for each movie in the filmography do
14:
          Retrieve the movie details
15:
          if 'rating' and 'year' are in movie details then
16:
              Append the movie's rating and year to the list
17:
18:
          else
19:
              Print: "No rating or year available for this movie"
          end if
20:
       end for
21:
       Convert the list into a DataFrame
22:
       Calculate the yearly mean rating using 'groupby' method
23:
24:
       Print yearly mean rating
       Display the results in a plot
25:
26: else
       Print "Invalid action selected"
27:
28: end if
29: End
```

Bararu Robert Daniel

2.3 Movie genre of actor/actresses

This algorithm retrieves and organizes the titles and genres of a specific actor's films, presenting the information in a tabular format.

Algorithm 3 Retrieve Genre of Actor/Actress's Movies

```
1: Start
 2: Import necessary libraries for data handling
 4: Load the dataset from a CSV file
 5: Initialize the IMDbPy
6:
 7: Display action options
8: if User chooses "Movie genres" then
9:
       Prompt user to enter the actor's name
       Fetch the actor's information using IMDbPy
10:
       Extract the 'filmography' from the actor's information
11:
       Initialize an empty list to store movie titles and genres
12:
       for each movie in the filmography do
13:
14:
          Retrieve the movie details
          if 'genre' is available in movie details then
15:
              Append the movie's title and its genre(s) to the list
16:
          else
17:
              Print: "No genre information available for this movie"
18:
19:
          end if
20:
       end for
       if information is successfully retrieved then
21:
          Print title and genre movies
22:
       else
23:
24:
          Print an error message
       end if
25:
26: else
       Print "Invalid action selected"
27:
28: end if
29: End
```

Bararu Robert Daniel

3 Solution Examples

3.1 About the Actor/Actress

This section presents the solution to retrieve and display detailed information about actors and actresses.

Simulated User Interaction:

```
Welcome to the IMDb Information System
Please choose an action:
1. About actor/actress
2. Average rating of movies
3. Movie genres
> 1
Do you want information about:
1. All actors/actresses
2. A specific actor/actress
> 2
Please enter the actors name:
> Johnny Depp
Fetching information for Johnny Depp...
Name: Johnny Depp
Biography: John Christopher Depp II was born on
   June 9, 1963...
Top Movie: Pirates of the Caribbean: The Curse of
   the Black Pearl
```

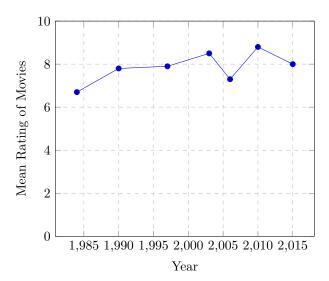
This output shows the required details about the selected actor, including the name, a short biography and the most famous film.

3.2 Average rating of their movies

This section presents the solution to retrieve and display detailed information about ratings of movies.

Simulated User Interaction:

```
Welcome to the IMDb Information System
Please choose an action:
1. About actor/actress
2. Average rating of movies
3. Movie genres
Do you want information about which actor?
Please enter the actors name:
> Johnny Depp
Fetching information for Johnny Depp...
   Year
    1984
                6.7
    1990
               7.8
    2015
               8.0
```



The table lists the years and average ratings for a specific actor's films. Provides a clear numerical summary of annual ratings. On the other hand, the line graph displays the table data in a visual format. It is useful for analyzing the evolution of the quality of films over time.

3.3 Movie genre of actor/actresses

This section presents the solution to retrieve and display detailed information about genre of movies.

Simulated User Interaction:

```
Welcome to the IMDb Information System
Please choose an action:
1. About actor/actress
2. Average rating of movies
3. Movie genres
> 3
Do you want information about which actor?
Please enter the actors name:
> Johnny Depp
Fetching information for Johnny Depp...
                                       Genre
  A Nightmare on Elm Street |
                                 Horror, Thriller
        Benny an Joon
                                  Comedy, Romance
      Finding Neverland
                                  Drama, Family
```

This output presents a table that maps movie titles to their genres. It is especially useful for understanding the variety and type of roles played by the selected actor.