Homework 5

Due date: Jun 2, 2024

Submission instructions:

- **Autograder will be used for scoring, you are required to write a module in a file `hw5module.py` as in Homeworks 1 and 2.**
- You are also required to convert this notebook as a separate Python file, **`hw5.py`**.
- Also, please keep your indentifiable information (ID, name, and a list of collaborators) in a separate file **`hw5_studentinfo.txt`**.
- Submit **`hw5.ipynb`** (this notebook), `hw5.py`, `hw5module.py`, and `hw5_studentinfo.txt` on Gradescope under the window "Homework 5 - code". Do **NOT** change the file name. This will be checked by the autograder as well.
- Make sure all your code and text outputs in the problems are visible in your PDF submission.

Step 1: Setup

Locate the Starting TMDB Page

My favourite movie is My Dinner with Andre.

https://www.themoviedb.org/movie/25468-my-dinner-with-andre

Step 2: Dry-Run Navigation

we'll navigate through the steps manually to understand the structure of the web pages we'll scrape.

1. Movie Page: Start at the movie's page: https://www.themoviedb.org/movie/25468-my-dinner-with-andre
2. Full Cast & Crew: Navigate to the "Full Cast & Crew" page by appending /cast to the URL: https://www.themoviedb.org/movie/25468-my-dinner-with-andre/cast
3. Actor Page: Click on an actor's portrait to go to their profile page.
4. Acting Section: Scroll down to the "Acting" section to find the list of movies and TV shows the actor has worked on.

Step 3: Create the Module

the module code is in hw5module.py.

The first parse_actor_page function works by:

1. Forming the URL for the actor's page based on their directory.
2. Sending a request to the URL and parsing the HTML response with BeautifulSoup.

3. Extracting the actor's name from the page.
4. Finding all tables with the class credit_group that contain acting credits.
5. Iterating over each table, extracting the movie/TV show names, and adding them to the DataFrame.

The second function parse_full_credits works by:

1. Forming the URL for the movie's cast page based on its directory.
2. Sending a request to the URL and parsing the HTML response with BeautifulSoup.
3. Extracting the links to actor pages from the cast section of the page and removing duplicates.
4. Printing the number of unique actor links found.
5. Iterating over each actor link, calling the parse_actor_page function to get their acting credits, and adding the data to the DataFrame.
6. Introducing a delay to avoid getting blocked by the server.

Step 4: Test the Scraper

In [ ]:
```python
# Import necessary libraries and functions
import pandas as pd
import matplotlib.pyplot as plt
from hw5module import parse_full_credits

# Specify the movie directory for My Dinner with Andre
movie_directory = "25468-my-dinner-with-andre"

# Call the parse_full_credits function to get the DataFrame with actors and thei
df = parse_full_credits(movie_directory)

# Display the first few rows of the DataFrame
print(df)
```

```
Found 4 actors
Parsing actor: 12900-wallace-shawn
Actor name: Wallace Shawn
Found movie/TV show: The Addams Family 2
Found movie/TV show: Gossip Girl
Found movie/TV show: Waiting for Godot
Found movie/TV show: Rifkin's Festival
Found movie/TV show: A Virtual Princess Bride Reunion
Found movie/TV show: Timmy Failure: Mistakes Were Made
Found movie/TV show: Forky Asks a Question: What Is Time?
Found movie/TV show: Marriage Story
Found movie/TV show: Evil
Found movie/TV show: Toy Story 4
Found movie/TV show: Amphibia
Found movie/TV show: What We Left Behind: Looking Back at Star Trek: Deep Space N
ine
Found movie/TV show: Summer Camp Island
Found movie/TV show: Big City Greens
Found movie/TV show: Book Club
Found movie/TV show: Another Kind of Wedding
Found movie/TV show: Young Sheldon
Found movie/TV show: Cop and a Half: New Recruit
Found movie/TV show: The Only Living Boy in New York
Found movie/TV show: Animal Crackers
Found movie/TV show: The Good Fight
Found movie/TV show: Search Party
Found movie/TV show: Maggie's Plan
Found movie/TV show: Starring Austin Pendleton
Found movie/TV show: Toy Story at 20: To Infinity and Beyond
Found movie/TV show: Club Penguin Halloween Panic!
Found movie/TV show: Robo-Dog
Found movie/TV show: Life in Pieces
Found movie/TV show: Club Penguin Monster Beach Party
Found movie/TV show: Mr. Robot
Found movie/TV show: Christmas at Cartwright's
Found movie/TV show: Toy Story That Time Forgot
Found movie/TV show: The Fog of Courage
Found movie/TV show: The Mysteries of Laura
Found movie/TV show: BoJack Horseman
Found movie/TV show: The 7D
Found movie/TV show: The Night Shift
Found movie/TV show: Don Peyote
Found movie/TV show: The Double
Found movie/TV show: Mozart in the Jungle
Found movie/TV show: A Master Builder
Found movie/TV show: Toy Story of Terror!
Found movie/TV show: Andre Gregory: Before and After Dinner
Found movie/TV show: Admission
Found movie/TV show: A Late Quartet
Found movie/TV show: Partysaurus Rex
Found movie/TV show: Vamps
Found movie/TV show: Small Fry
Found movie/TV show: Hawaiian Vacation
Found movie/TV show: The Speed of Thought
Found movie/TV show: Regular Show
Found movie/TV show: Fish Hooks
Found movie/TV show: Cats & Dogs: The Revenge of Kitty Galore
Found movie/TV show: Toy Story 3
Found movie/TV show: Furry Vengeance
Found movie/TV show: Buzz Lightyear Mission Logs
```

Found movie/TV show: Jack and the Beanstalk
Found movie/TV show: The Good Wife
Found movie/TV show: Capitalism: A Love Story
Found movie/TV show: The Windmill Movie
Found movie/TV show: André Gregory and Wallace Shawn
Found movie/TV show: Life on Mars
Found movie/TV show: Scooby-Doo! and the Goblin King
Found movie/TV show: Kit Kittredge: An American Girl
Found movie/TV show: USIDent TV: Surveilling the Southland
Found movie/TV show: Southland Tales
Found movie/TV show: Gossip Girl
Found movie/TV show: The Pixar Story
Found movie/TV show: Phineas and Ferb
Found movie/TV show: I Could Never Be Your Woman
Found movie/TV show: Strange Culture
Found movie/TV show: Happily N'Ever After
Found movie/TV show: Air Buddies
Found movie/TV show: Tom and Jerry: Shiver Me Whiskers
Found movie/TV show: Eureka
Found movie/TV show: Howard Zinn: Voices of a People's History of the United Stat
es
Found movie/TV show: Chicken Little
Found movie/TV show: Love Thy Neighbor
Found movie/TV show: Fat Actress
Found movie/TV show: Karroll's Christmas
Found movie/TV show: Melinda and Melinda
Found movie/TV show: The Incredibles
Found movie/TV show: Desperate Housewives
Found movie/TV show: The L Word
Found movie/TV show: Teacher's Pet
Found movie/TV show: The Haunted Mansion
Found movie/TV show: Duplex
Found movie/TV show: Monte Walsh
Found movie/TV show: Personal Velocity
Found movie/TV show: Monsters, Inc.
Found movie/TV show: Law & Order: Criminal Intent
Found movie/TV show: Crossing Jordan
Found movie/TV show: As You Wish: The Story of 'The Princess Bride'
Found movie/TV show: The Curse of the Jade Scorpion
Found movie/TV show: Blonde
Found movie/TV show: The Prime Gig
Found movie/TV show: Teacher's Pet
Found movie/TV show: Buzz Lightyear of Star Command: The Adventure Begins
Found movie/TV show: Toy Story 2
Found movie/TV show: Law & Order: Special Victims Unit
Found movie/TV show: My Favorite Martian
Found movie/TV show: Family Guy
Found movie/TV show: The Diary of the Hurdy-Gurdy Man
Found movie/TV show: Noah
Found movie/TV show: The Jungle Book: Mowgli's Story
Found movie/TV show: The Lionhearts
Found movie/TV show: Sex and the City
Found movie/TV show: Critical Care
Found movie/TV show: The Wonderful World of Disney
Found movie/TV show: Ally McBeal
Found movie/TV show: Stargate SG-1
Found movie/TV show: Just Write
Found movie/TV show: Vegas Vacation
Found movie/TV show: King of the Hill
Found movie/TV show: The Story Behind 'Toy Story'

```
Found movie/TV show: House Arrest
Found movie/TV show: All Dogs Go to Heaven 2
Found movie/TV show: Toy Story Treats
Found movie/TV show: Toy Story
Found movie/TV show: Just Like Dad
Found movie/TV show: Clueless
Found movie/TV show: Napoleon
Found movie/TV show: Canadian Bacon
Found movie/TV show: The Wife
Found movie/TV show: Driving Mr. Pink
Found movie/TV show: A Goofy Movie
Found movie/TV show: Vanya on 42nd Street
Found movie/TV show: Mrs. Parker and the Vicious Circle
Found movie/TV show: The Nanny
Found movie/TV show: The Pink Panther
Found movie/TV show: The Meteor Man
Found movie/TV show: The Cemetery Club
Found movie/TV show: Homicide: Life on the Street
Found movie/TV show: Star Trek: Deep Space Nine
Found movie/TV show: The Double 0 Kid
Found movie/TV show: Mom and Dad Save the World
Found movie/TV show: The Magic Bubble
Found movie/TV show: Nickel & Dime
Found movie/TV show: Shadows and Fog
Found movie/TV show: The Simpsons
Found movie/TV show: Scenes from the Class Struggle in Beverly Hills
Found movie/TV show: We're No Angels
Found movie/TV show: She's Out of Control
Found movie/TV show: The Moderns
Found movie/TV show: The Princess Bride
Found movie/TV show: Prick Up Your Ears
Found movie/TV show: Nice Girls Don't Explode
Found movie/TV show: Radio Days
Found movie/TV show: The Bedroom Window
Found movie/TV show: Head Office
Found movie/TV show: Heaven Help Us
Found movie/TV show: Micki + Maude
Found movie/TV show: How to Be a Perfect Person in Just Three Days
Found movie/TV show: The Cosby Show
Found movie/TV show: My Dinner with Louis
Found movie/TV show: The Bostonians
Found movie/TV show: The Hotel New Hampshire
Found movie/TV show: Crackers
Found movie/TV show: Saigon: Year Of The Cat
Found movie/TV show: Deal of the Century
Found movie/TV show: Strange Invaders
Found movie/TV show: The First Time
Found movie/TV show: Lovesick
Found movie/TV show: A Little Sex
Found movie/TV show: My Dinner with Andre
Found movie/TV show: Strong Medicine
Found movie/TV show: Cheaper to Keep Her
Found movie/TV show: Atlantic City
Found movie/TV show: Simon
Found movie/TV show: All That Jazz
Found movie/TV show: Starting Over
Found movie/TV show: Manhattan
Found movie/TV show: Taxi
Parsing actor: 109470-andre-gregory
Actor name: Andre Gregory
```

```
Found movie/TV show: A Master Builder
Found movie/TV show: Andre Gregory: Before and After Dinner
Found movie/TV show: André Gregory and Wallace Shawn
Found movie/TV show: Goodbye Lover
Found movie/TV show: Celebrity
Found movie/TV show: Hudson River Blues
Found movie/TV show: Who Is Henry Jaglom?
Found movie/TV show: Last Summer in the Hamptons
Found movie/TV show: Vanya on 42nd Street
Found movie/TV show: The Shadow
Found movie/TV show: Demolition Man
Found movie/TV show: The Linguini Incident
Found movie/TV show: The Bonfire of the Vanities
Found movie/TV show: Location Production Footage: The Last Temptation of the Chri
st
Found movie/TV show: Some Girls
Found movie/TV show: The Last Temptation of Christ
Found movie/TV show: Street Smart
Found movie/TV show: The Mosquito Coast
Found movie/TV show: Follies: In Concert
Found movie/TV show: Always … But Not Forever
Found movie/TV show: Protocol
Found movie/TV show: The Soldier's Tale
Found movie/TV show: Alice in Wonderland
Found movie/TV show: Author! Author!
Found movie/TV show: My Dinner with Andre
Found movie/TV show: Great Performances
Parsing actor: 1757529-roy-butler
Actor name: Roy Butler
Found movie/TV show: My Dinner with Andre
Parsing actor: 143806-jean-lenauer
Actor name: Jean Lenauer
Found movie/TV show: My Dinner with Andre
```

	actor	movie_or_TV_name
0	Andre Gregory	A Master Builder
1	Andre Gregory	Alice in Wonderland
2	Andre Gregory	Always … But Not Forever
3	Andre Gregory	Andre Gregory: Before and After Dinner
4	Andre Gregory	André Gregory and Wallace Shawn
..	...	...
194	Wallace Shawn	Vegas Vacation
195	Wallace Shawn	Waiting for Godot
196	Wallace Shawn	We're No Angels
197	Wallace Shawn	What We Left Behind: Looking Back at Star Trek...
198	Wallace Shawn	Young Sheldon

```
[199 rows x 2 columns]
```

Step 5: Make Recommendations

In [ ]:

```python
# Make recommendations based on the shared actors
def make_recommendations(df):
    """
    Generates recommendations based on the shared actors between movies or TV sh

    Assumptions:
    - DataFrame contains columns 'actor' and 'movie_or_TV_name'.

    Effect:
```

```
    - Groups the DataFrame by 'movie_or_TV_name' and counts unique actors.
    - Sorts the movies or TV shows by the number of shared actors.

    Args:
    df (pd.DataFrame): DataFrame with actors and their credits.

    Returns:
    pd.DataFrame: Sorted DataFrame with recommendations.
    """
    shared_actors = df.groupby('movie_or_TV_name')['actor'].nunique().reset_inde
    shared_actors.columns = ['movie_or_TV_name', 'shared_actors_count']
    recommendations = shared_actors.sort_values(by='shared_actors_count', ascend
    return recommendations

# Generate and display recommendations
recommendations = make_recommendations(df)
print("Top movies and TV shows with shared actors:")
print(recommendations)
```

```
Top movies and TV shows with shared actors:
                          movie_or_TV_name  shared_actors_count
92                     My Dinner with Andre                    4
14          André Gregory and Wallace Shawn                    2
3                          A Master Builder                    2
185                    Vanya on 42nd Street                    2
13   Andre Gregory: Before and After Dinner                    2
..                                      ...                  ...
67                   Jack and the Beanstalk                    1
68                            Just Like Dad                    1
69                               Just Write                    1
70                     Karroll's Christmas                    1
191                          Young Sheldon                    1

[192 rows x 2 columns]
```

Based on the analysis of shared actors between movies and TV shows, it is clear that "André Gregory and Wallace Shawn," "A Master Builder," "Vanya on 42nd Street," and "Andre Gregory: Before and After Dinner," each of which have two shared actors. These results indicate that there is a significant overlap in the cast of these productions. For someone who enjoyed "My Dinner with Andre," these titles are highly recommended as they feature a similar ensemble of actors, likely contributing to a comparable style and quality of performance.

Additionally, other productions such as "Jack and the Beanstalk," "Just Like Dad," "Just Write," "Karroll's Christmas," and "Young Sheldon" each feature one shared actor from "My Dinner with Andre." These recommendations, while not as closely linked in terms of shared actors, may still offer an enjoyable experience due to the presence of familiar talent. Thus, viewers can explore these recommendations to find similar thematic or stylistic elements in the performances of the shared actors.

Step 6: Visualization

In [ ]:
```
# Visualization of the shared actors
def plot_shared_actors(recommendations):
    """
```

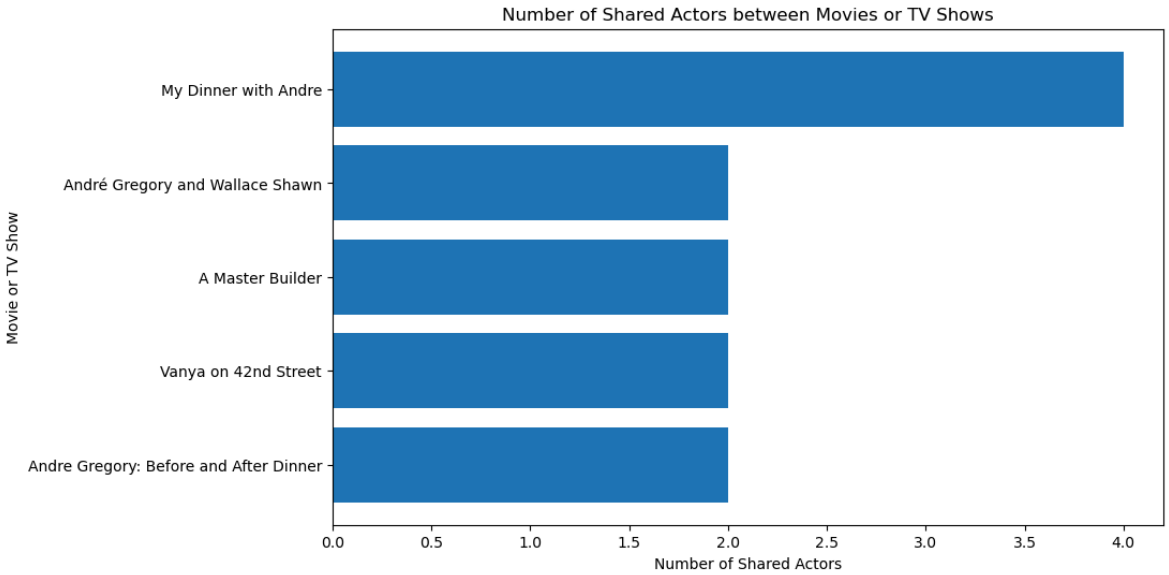
```
    Plots a horizontal bar chart of the top movies or TV shows with shared actor

    Assumptions:
    - The recommendations DataFrame contains 'movie_or_TV_name' and 'shared_acto

    Effect:
    - Displays a horizontal bar chart with the number of shared actors on the x-

    Args:
    recommendations (pd.DataFrame): DataFrame with recommendations.
    """
    plt.figure(figsize=(10, 6))
    plt.barh(recommendations['movie_or_TV_name'], recommendations['shared_actors
    plt.xlabel('Number of Shared Actors')
    plt.ylabel('Movie or TV Show')
    plt.title('Number of Shared Actors between Movies or TV Shows')
    plt.gca().invert_yaxis()
    plt.show()

# Plot the top 10 shared actors
plot_shared_actors(recommendations.head(5))
```



Description:

1. Title: Number of Shared Actors between Movies or TV Shows
2. X-Axis: Number of Shared Actors (representing how many actors are shared between "My Dinner with Andre" and other movies or TV shows)
3. Y-Axis: Movie or TV Show names
4. Bars: Each bar represents a movie or TV show that shares actors with "My Dinner with Andre."

Observations:

1. My Dinner with Andre: This movie has the highest number of shared actors, with 4 shared actors, indicating that it is the primary focus of the dataset.
2. André Gregory and Wallace Shawn: This entry has 2 shared actors, indicating a significant overlap with the primary movie.
3. A Master Builder: Also shares 2 actors with "My Dinner with Andre."

4. Vanya on 42nd Street: Shares 2 actors.
5. Andre Gregory: Before and After Dinner: Another entry with 2 shared actors.

The horizontal bar chart effectively visualizes the number of shared actors between the specified movie "My Dinner with Andre" and other movies or TV shows. The use of integer ticks on the x-axis ensures clarity in the number of shared actors, making the data easy to interpret. The bars are sorted in descending order, showcasing the movies or TV shows with the highest number of shared actors at the top.