In [25]:

```
from bs4 import BeautifulSoup
import pandas as pd
import os
```

In [26]:

```
d={'Rank':[],'Title':[],'Years':[],'Rating':[]}

def clean_text(text):
    return text.replace('\xa0', ' ').strip()

for file in os.listdir("data"):
    # print(file)

    # reading files from current dir
    with open(f"data/{file}") as f:
        html_doc=f.read()
    soup=BeautifulSoup(html_doc,'html.parser')

    # title with rank
    t=soup.find("h3")
    title=t.get_text()

    #rank
    if '.' in title:
        rank,title=title.split('.',1)
    else:
        rank=''


    # years
    y=soup.find("span" , attrs={"class":'sc-b189961a-8'})
    years=y.get_text()


    r = soup.find("span", attrs={"class": 'ipc-rating-star'})
    rating = clean_text(r.get_text())
    #print(rating)

    #print(title,years,rating)

    d['Rank'].append(rank)
    d['Title'].append(title)
    d['Years'].append(years)
    d['Rating'].append(rating)


df=pd.DataFrame(data=d)
df.Rating=df.Rating.sort_values(ascending=False)
df.Rank=df.Rank.astype(int) # change object to int
df.to_csv("Top 250 Movies List.csv",index=False)
df
```

Out[26]:

	Rank	Title	Years	Rating
0	1	The Shawshank Redemption	1994	9.3 (2.9M)
1	2	The Godfather	1972	9.2 (2M)
2	11	Forrest Gump	1994	8.8 (2.3M)
3	101	The Apartment	1960	8.3 (198K)
4	102	Incendies	2010	8.3 (204K)
...	...	...	...	...
245	96	2001: A Space Odyssey	1968	8.3 (724K)
246	97	Reservoir Dogs	1992	8.3 (1.1M)
247	98	Ikiru	1952	8.3 (88K)
248	99	Oppenheimer	2023	8.3 (759K)
249	100	Lawrence of Arabia	1962	8.3 (316K)

250 rows × 4 columns

In [ ]: