

```
In [1]: from bs4 import BeautifulSoup
import pandas as pd
import time
import os
```

```
In [24]: d={'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
        t=soup.find("h3")
        title=t.get_text()

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

df=pd.DataFrame(data=d)
df.to_csv("Top 250 Movies List.csv",index=False)
```