from bs4 import BeautifulSoup

import requests

import re

import pandas as pd

# Downloading imdb top 250 movie's data

url = 'http://www.imdb.com/chart/top'

response = requests.get(url)

soup = BeautifulSoup(response.text, "html.parser")

movies = soup.select('td.titleColumn')

crew = [a.attrs.get('title') for a in soup.select('td.titleColumn a')]

ratings = [b.attrs.get('data-value')

for b in soup.select('td.posterColumn span[name=ir]')]

# create a empty list for storing

# movie information

list = []

# Iterating over movies to extract

# each movie's details

for index in range(0, len(movies)):

# Separating movie into: 'place',

# 'title', 'year'

movie\_string = movies[index].get\_text()

movie = (' '.join(movie\_string.split()).replace('.', ''))

movie\_title = movie[len(str(index))+1:-7]

year = re.search('\((.\*?)\)', movie\_string).group(1)

place = movie[:len(str(index))-(len(movie))]

data = {"place": place,

"movie\_title": movie\_title,

"rating": ratings[index],

"year": year,

"star\_cast": crew[index],

}

list.append(data)

# printing movie details with its rating.

for movie in list:

print(movie['place'], '-', movie['movie\_title'], '('+movie['year'] +

') -', 'Starring:', movie['star\_cast'], movie['rating'])

##.......##

df = pd.DataFrame(list)

df.to\_csv('imdb\_top\_250\_movies.csv',index=False)