EXPLORER

LEARN_WITH_GEEKSTER

- GeeksterWorkVansh
- datascrap-CW1.py
- datascrap-CW2.py
- datascrap-HW.py
- file2.txt
- first.txt
- req.py
- top250movies.csv
- wt2.py

OUTLINE

TIMELINE

datascrap-CW1.py | datascrap-CW2.py | datascrap-HW.py

datascrap-CW2.py > ...

```python
import requests
import pandas as pd
from bs4 import BeautifulSoup
url = requests.get('https://www.imdb.com/chart/top/',headers = {"U
soup = BeautifulSoup(url.text,"html.parser")
data = soup.find('script',id='__NEXT_DATA__')
data = str(data.text)
data = data.replace('false','False')
data = data.replace('true','True')
data = data.replace('null','None')
data = eval(data)
data_fetched = data['props']['pageProps']['pageData']['chartTitles
movie_name = data_fetched[0]['node']['titleText']['text']
year = data_fetched[0]['node']['releaseYear']['year']
ratings = data_fetched[0]['node']['ratingsSummary']['aggregateRati
l1=[]
l2=[]
l3=[]
for i in range(len(data_fetched)):
    movie_name = data_fetched[i]['node']['titleText']['text']
    year = data_fetched[i]['node']['releaseYear']['year']
    rating = data_fetched[i]['node']['ratingsSummary']['aggregateR
    l1.append(movie_name)
    l2.append(year)
    l3.append(rating)

d = {"movie name":l1,"year":l2,"ratings":l3}
df = pd.DataFrame(d)
df.to_csv('top250movies.csv')
print(df)
```

PROBLEMS OUTPUT TERMINAL

Python

```
vanshtanwar@Vanshs-MacBook-Air Learn_with_geekster % /usr/local/bin/python3 /Users/vanshtanwar/Downloads/Learn_with_geekster/datascrap-CW2.py
                movie name  year  ratings
0    The Shawshank Redemption  1994      9.3
1               The Godfather  1972      9.2
2             The Dark Knight  2008      9.0
3      The Godfather: Part II  1974      9.0
4                12 Angry Men  1957      9.0
..                        ...   ...      ...
245            Cool Hand Luke  1967      8.1
246     It Happened One Night  1934      8.1
247        Gangs of Wasseypur  2012      8.2
248                  Drishyam  2015      8.2
249                 Andhadhun  2018      8.2

[250 rows x 3 columns]
vanshtanwar@Vanshs-MacBook-Air Learn_with_geekster %
```

	A	B	C
1	movie name	year	ratings
2	The Shawshank Redemption	1994	9.3
3	The Godfather	1972	9.2
4	The Dark Knight	2008	9
5	The Godfather: Part II	1974	9
6	12 Angry Men	1957	9
7	The Lord of the Rings: The Return of the King	2003	9
8	Schindler's List	1993	9
9	Pulp Fiction	1994	8.9
10	The Lord of the Rings: The Fellowship of the Ring	2001	8.9
11	Il Buono, Il Brutto, Il Cattivo	1966	8.8
12	Forrest Gump	1994	8.8
13	The Lord of the Rings: The Two Towers	2002	8.8
14	Fight Club	1999	8.8
15	Inception	2010	8.8
16	Star Wars: Episode V - The Empire Strikes Back	1980	8.7
17	The Matrix	1999	8.7
18	GoodFellas	1990	8.7
19	One Flew Over the Cuckoo's Nest	1975	8.7
20	Interstellar	2014	8.7
21	Se7en	1995	8.6
22	It's a Wonderful Life	1946	8.6
23	Shichinin No Samurai	1954	8.6
24	The Silence of the Lambs	1991	8.6
25	Saving Private Ryan	1998	8.6
26	City of God	2002	8.6
27	Life Is Beautiful	1997	8.6
28	The Green Mile	1999	8.6
29	Terminator 2: Judgment Day	1991	8.6
30	Star Wars: Episode IV - A New Hope	1977	8.6
31	Back to the Future	1985	8.5
32	Spirited Away	2001	8.6
33	The Pianist	2002	8.5
34	Parasite	2019	8.5
35	Psycho	1960	8.5
36	Gladiator	2000	8.5
37	The Lion King	1994	8.5
38	Spider-man: Across the Spider-verse	2023	8.6
39	The Departed	2006	8.5

EXPLORER

LEARN_WITH_GEEKS...

- GeeksterWorkVansh
- datascrap-CW1.py
- datascrap-CW2.py
- datascrap-HW.py
- file2.txt
- first.txt
- req.py
- states-population.csv
- top250movies.csv
- wt2.py

OUTLINE

TIMELINE

datascrap-CW1.py | datascrap-CW2.py | datascrap-HW.py

datascrap-CW1.py > ...

```python
import requests
import pandas as pd
from bs4 import BeautifulSoup

url = "https://en.wikipedia.org/wiki/List_of_states_and_territorie
response = requests.get(url)
soup = BeautifulSoup(response.content, "html.parser")

table = soup.find("table", {"class": "wikitable"})
state_names = []
populations = []
if table:
    rows = table.find_all("tr")
    for row in rows[1:]:
        cells = row.find_all("td")
        if len(cells) >= 6:
            state_name = cells[0].text.strip()
            population = cells[5].text.strip()
            state_names.append(state_name)
            populations.append(population)
else:
    print("Table not found on the page.")

d = {"state_name": state_names, "population": populations}
df = pd.DataFrame(d)
df.to_csv('states-population.csv', index=False)
print(df)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

Python

```
vanshtanwar@Vanshs-MacBook-Air Learn_with_geekster % /usr/local/bin/python3 /Users/vanshtanwar/Downloads/Learn_with_geekster/datascrap-CW1.py
   state_name population
0          AL     52,420
1          AK    665,384
2          AZ    295,234
3          AR    137,732
4          CA    163,695
5          CO    269,601
6          CT      5,543
7          DE      2,489
8          FL     65,758
9          GA    153,910
10         HI     28,313
11         ID    216,443
12         IL     57,914
13         IN     94,326
14         IA    145,746
15         KS     82,278
16         KY     40,408
17         LA     52,378
18         ME     35,380
19         MD     12,406
20         MA     27,336
21         MI     96,714
22         MN     86,936
23         MS    125,438
24         MO     69,707
25         MT    147,040
26         NE     77,348
27         NV    110,572
28         NH      9,349
29         NJ      8,723
30         NM    121,590
31         NY     54,555
32         NC     53,819
33         ND     70,698
34         OH    116,098
35         OK    181,037
36         OR     98,379
37         PA     46,054
38         RI      4,001
39         SC     32,020
40         SD     77,116
41         TN    109,153
42         TX    268,596
43         UT    219,882
44         VT      9,616
45         VA     42,775
46         WA     71,298
47         WV     62,756
48         WI     65,496
49         WY    253,335
vanshtanwar@Vanshs-MacBook-Air Learn_with_geekster %
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	state_name	population																				
2	AL	52,420																				
3	AK	6,65,384																				
4	AZ	2,95,234																				
5	AR	1,37,732																				
6	CA	1,63,695																				
7	CO	2,69,601																				
8	CT	5,543																				
9	DE	2,489																				
10	FL	65,758																				
11	GA	1,53,910																				
12	HI	28,313																				
13	ID	2,16,443																				
14	IL	57,914																				
15	IN	94,326																				
16	IA	1,45,746																				
17	KS	82,278																				
18	KY	40,408																				
19	LA	52,378																				
20	ME	35,380																				
21	MD	12,406																				
22	MA	27,336																				
23	MI	96,714																				
24	MN	86,936																				
25	MS	1,25,438																				
26	MO	69,707																				
27	MT	1,47,040																				
28	NE	77,348																				
29	NV	1,10,572																				
30	NH	9,349																				
31	NJ	8,723																				
32	NM	1,21,590																				
33	NY	54,555																				
34	NC	53,819																				
35	ND	70,698																				
36	OH	1,16,098																				
37	OK	1,81,037																				
38	OR	98,379																				
39	PA	46,054																				