

```
In [1]: import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings('ignore')
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementException, ElementNotInteract
from selenium.webdriver.common.by import By
import time
```

In []:

Scrape the details of most viewed videos on YouTube from Wikipedia. Url = https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos You need to find following details: A) Rank B) Name C) Artist D) Upload date E) Views

In []:

```
In [31]: driver=webdriver.Chrome(r'C:/chromedriver.exe')
```

```
In [32]: driver.maximize_window()
time.sleep(2)
```

```
In [33]: url=driver.get(' https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos')
```

```
In [34]: for _ in range(5):
        driver.execute_script("window.scrollTo(0,100)")
```

```
In [35]: Rank=[]
Name=[]
Artist=[]
Upload_date=[]
Views=[]
```

```
In [36]: rank=driver.find_elements(By.XPATH,'//table[@class="wikitable sortable jquery-tablesorter"]/tbody/tr/td[1]')
for i in rank[:30]:
    Rank.append(i.text)
name=driver.find_elements(By.XPATH,'//table[@class="wikitable sortable jquery-tablesorter"]/tbody/tr/td[2]')
for i in name[:30]:
    Name.append(i.text)
artist=driver.find_elements(By.XPATH,'//table[@class="wikitable sortable jquery-tablesorter"]/tbody/tr/td[3]')
for i in artist[:30]:
    Artist.append(i.text)
views=driver.find_elements(By.XPATH,'//table[@class="wikitable sortable jquery-tablesorter"]/tbody/tr/td[4]')
for i in views[:30]:
    Views.append(i.text)
date=driver.find_elements(By.XPATH,'//table[@class="wikitable sortable jquery-tablesorter"]/tbody/tr/td[5]')
for i in date[:30]:
    Upload_date.append(i.text)
```

```
In [38]: print(len(Rank),len(Name),len(Artist),len(Views),len(Upload_date))

30 30 30 30 30
```

```
In [39]: youtube=pd.DataFrame({'Rank':Rank,'Name':Name,'Artist':Artist,'Upload_Date':Upload_date,'Views':Views})
youtube
```

Out[39]:	Rank	Name	Artist	Upload_Date	Views
0	1.	"Baby Shark Dance"[4]	Pinkfong Baby Shark - Kids' Songs & Stories	June 17, 2016	12.73
1	2.	"Despacito"[7]	Luis Fonsi	January 12, 2017	8.14
2	3.	"Johnny Johnny Yes Papa"[14]	LooLoo Kids	October 8, 2016	6.69
3	4.	"Bath Song"[15]	Cocomelon – Nursery Rhymes	May 2, 2018	6.15
4	5.	"Shape of You"[16]	Ed Sheeran	January 30, 2017	5.97
5	6.	"See You Again"[18]	Wiz Khalifa	April 6, 2015	5.86
6	7.	"Phonics Song with Two Words"[23]	ChuChu TV	March 6, 2014	5.26
7	8.	"Wheels on the Bus"[24]	Cocomelon – Nursery Rhymes	May 24, 2018	5.14
8	9.	"Uptown Funk"[25]	Mark Ronson	November 19, 2014	4.89
9	10.	"Learning Colors – Colorful Eggs on a Farm"[26]	Miroshka TV	February 27, 2018	4.87
10	11.	"Gangnam Style"[27]	Psy	July 15, 2012	4.77
11	12.	"Masha and the Bear – Recipe for Disaster"[32]	Get Movies	January 31, 2012	4.55
12	13.	"Dame Tu Cosita"[33]	El Chombo	April 5, 2018	4.32
13	14.	"Axel F"[34]	Crazy Frog	June 16, 2009	3.87
14	15.	"Sugar"[35]	Maroon 5	January 14, 2015	3.86
15	16.	"Roar"[36]	Katy Perry	September 5, 2013	3.78
16	17.	"Counting Stars"[37]	OneRepublic	May 31, 2013	3.77
17	18.	"Sorry"[38]	Justin Bieber	October 22, 2015	3.65
18	19.	"Baa Baa Black Sheep"[39]	Cocomelon – Nursery Rhymes	June 25, 2018	3.61
19	20.	"Thinking Out Loud"[40]	Ed Sheeran	October 7, 2014	3.58
20	21.	"Waka Waka (This Time for Africa)"[41]	Shakira	June 4, 2010	3.56
21	22.	"Dark Horse"[42]	Katy Perry	February 20, 2014	3.50
22	23.	"Faded"[43]	Alan Walker	December 3, 2015	3.44
23	24.	"Perfect"[44]	Ed Sheeran	November 9, 2017	3.42
24	25.	"Lakdi Ki Kathi"[45]	Jingle Toons	June 14, 2018	3.42
25	26.	"Let Her Go"[46]	Passenger	July 25, 2012	3.42
26	27.	"Girls Like You"[47]	Maroon 5	May 31, 2018	3.40
27	28.	"Humpty the train on a fruits ride"[48]	Kiddiestv Hindi – Nursery Rhymes & Kids Songs	January 26, 2018	3.38
28	29.	"Lean On"[49]	Major Lazer	March 22, 2015	3.37
29	30.	"Bailando"[50]	Enrique Iglesias	April 11, 2014	3.37

```
In [40]: youtube.to_csv('Youtube.csv')
```

```
In [41]: driver.close()
```

```
In [ ]:
```

Scrape the details of State-wise GDP of India from statisticstimes.com. Url = <http://statisticstimes.com/> You have to find following details: A) Rank B) State C) GSDP(18-19)- at current prices D) GSDP(19-20)- at current prices E) Share(18-19) F) GDP(\$ billion) Note: - From statisticstimes home page you have to reach to economy page through code.

```
In [ ]:
```

```
In [28]: driver=webdriver.Chrome(r'C:/chromedriver.exe')
```

```
In [29]: driver.maximize_window()
time.sleep(2)
```

```
In [30]: Url=driver.get('http://statisticstimes.com/')
```

```
In [31]: economy=driver.find_element(By.XPATH, '/html/body/div[2]/div[1]/div[2]/div[2]/button')
economy.click()
```

```
In [32]: india=driver.find_element(By.XPATH, '/html/body/div[2]/div[1]/div[2]/div[2]/div/a[3]')
driver.get(india.get_attribute('href'))
```

```
In [33]: GDP=driver.find_element(By.XPATH, '/html/body/div[2]/div[2]/div[2]/ul/li[1]/a')
driver.get(GDP.get_attribute('href'))
```

```
In [34]: for _ in range(17):
driver.execute_script("window.scrollTo(0,100)")
```

```
In [35]: Rank=[]
State=[]
GSDP1=[]
GSDP2=[]
Share=[]
GDP=[]
```

```
In [36]: Rank=[]
rank=driver.find_elements(By.XPATH,'//table[@id="table_id"]/tbody/tr/td[1]')
for i in rank:
    Rank.append(i.text)

state=driver.find_elements(By.XPATH,'//table[@id="table_id"]/tbody/tr/td[2]')
for i in state:
    State.append(i.text)
gsdp1=driver.find_elements(By.XPATH,'//table[@id="table_id"]/tbody/tr/td[4]')
for i in gsdp1:
    GSDP1.append(i.text)
gsdp2=driver.find_elements(By.XPATH,'//table[@id="table_id"]/tbody/tr/td[3]')
for i in gsdp2:
    GSDP2.append(i.text)
share=driver.find_elements(By.XPATH,'//table[@id="table_id"]/tbody/tr/td[5]')
for i in share:
    Share.append(i.text)
gdp1=driver.find_elements(By.XPATH,'//table[@id="table_id"]/tbody/tr/td[6]')
for i in gdp1:
    GDP.append(i.text)
```

```
In [37]: print(len(Rank),len(State),len(GSDP1),len(GSDP2),len(Share),len(GDP))

33 33 33 33 33 33
```

```
In [38]: GDP=pd.DataFrame({'Rank':Rank, 'State':State, 'GSDP(18-19)':GSDP1, 'GSDP(19-20)':GSDP2, 'Share(19-20)':Share, 'GDP($
GDP
```

Out[38]:

	Rank	State	GSDP(18-19)	GSDP(19-20)	Share(19-20)	GDP(\$ billion)
0	1	Maharashtra	2,632,792	-	13.94%	399.921
1	2	Tamil Nadu	1,630,208	1,845,853	8.63%	247.629
2	3	Uttar Pradesh	1,584,764	1,687,818	8.39%	240.726
3	4	Gujarat	1,502,899	-	7.96%	228.290
4	5	Karnataka	1,493,127	1,631,977	7.91%	226.806
5	6	West Bengal	1,089,898	1,253,832	5.77%	165.556
6	7	Rajasthan	942,586	1,020,989	4.99%	143.179
7	8	Andhra Pradesh	862,957	972,782	4.57%	131.083
8	9	Telangana	861,031	969,604	4.56%	130.791
9	10	Madhya Pradesh	809,592	906,672	4.29%	122.977
10	11	Kerala	781,653	-	4.14%	118.733
11	12	Delhi	774,870	856,112	4.10%	117.703
12	13	Haryana	734,163	831,610	3.89%	111.519
13	14	Bihar	530,363	611,804	2.81%	80.562
14	15	Punjab	526,376	574,760	2.79%	79.957
15	16	Odisha	487,805	521,275	2.58%	74.098
16	17	Assam	315,881	-	1.67%	47.982
17	18	Chhattisgarh	304,063	329,180	1.61%	46.187
18	19	Jharkhand	297,204	328,598	1.57%	45.145
19	20	Uttarakhand	245,895	-	1.30%	37.351
20	21	Jammu & Kashmir	155,956	-	0.83%	23.690
21	22	Himachal Pradesh	153,845	165,472	0.81%	23.369
22	23	Goa	73,170	80,449	0.39%	11.115
23	24	Tripura	49,845	55,984	0.26%	7.571
24	25	Chandigarh	42,114	-	0.22%	6.397
25	26	Puducherry	34,433	38,253	0.18%	5.230
26	27	Meghalaya	33,481	36,572	0.18%	5.086
27	28	Sikkim	28,723	32,496	0.15%	4.363
28	29	Manipur	27,870	31,790	0.15%	4.233
29	30	Nagaland	27,283	-	0.14%	4.144
30	31	Arunachal Pradesh	24,603	-	0.13%	3.737
31	32	Mizoram	22,287	26,503	0.12%	3.385
32	33	Andaman & Nicobar Islands	-	-	-	-

```

In [39]: GDP.to_csv('GDP.csv')

In [40]: driver.close()

In [ ]:

Scrape the details of trending repositories on Github.com. Url = https://github.com/ You have to find the following details: A) Repository title B) Repository description C) Contributors count D) Language used

In [ ]:

In [90]: driver=webdriver.Chrome(r'C:/chrome.exe')
driver.maximize_window()
time.sleep(2)

In [91]: Url =driver.get(' https://github.com/')

In [92]: source=driver.find_element(By.XPATH, '/html/body/div[1]/div[1]/header/div/div[2]/div/nav/ul/li[3]/button').click

In [93]: trending=driver.find_element(By.XPATH, '/html/body/div[1]/div[1]/header/div/div[2]/div/nav/ul/li[3]/div/div[3]/u
urls=trending.get_attribute('href')
driver.get(urls)

In [94]: urls=[]
Contributors_count=[]
Language_used=[]
repo_title=[]

```

```

repo_desc=[]

urls=[]
rep=driver.find_elements(By.XPATH,'//h2[@class="h3 lh-condensed"]/a')
for i in rep:
    urls.append(i.get_attribute('href'))

```

In [95]:

```

urls
Out[95]: ['https://github.com/artidoro/qlora',
'https://github.com/gptlink/gptlink',
'https://github.com/ThePrimeagen/ts-rust-zig-deez',
'https://github.com/geohot/tinygrad',
'https://github.com/JushBJJ/Mr.-Ranedeer-AI-Tutor',
'https://github.com/microsoft/devicescript',
'https://github.com/xorvoid/sectorc',
'https://github.com/OpenGVLab/InternGPT',
'https://github.com/microsoft/PowerToys',
'https://github.com/Zeqiang-Lai/DragGAN',
'https://github.com/vanjs-org/van',
'https://github.com/kamiyaa/joshuto',
'https://github.com/microsoft/semantic-kernel',
'https://github.com/AggressiveUser/AllForOne',
'https://github.com/pengzhile/pandora',
'https://github.com/openai/openai-cookbook',
'https://github.com/immersive-translate/immersive-translate',
'https://github.com/microsoft/devhome',
'https://github.com/salesforce/LAVIS',
'https://github.com/hwchase17/langchain',
'https://github.com/Azure/cognitive-search-vector-pr',
'https://github.com/iperov/DeepFaceLive',
'https://github.com/chinese-poetry/chinese-poetry',
'https://github.com/GoogleCloudPlatform/generative-ai',
'https://github.com/ApriI NEA/ChatGPT-Admin-Web']

```

In [96]:

```

repo_title=[]
try:
    repos=driver.find_elements(By.XPATH,'//article[@class="Box-row"]/h2/a')
    for i in repos:
        repo_title.append(i.text)
except NoSuchElementException:
    repo_title.append('No details available')
time.sleep(3)

for i in urls:
    driver.get(i)
    l=[]
    time.sleep(3)

    try:
        repo = driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/main/turbo-frame/div/div/div/div[2]/div")
        repo_desc.append(repo.text)
    except NoSuchElementException:
        repo_desc.append('-')

    try:
        count=driver.find_element(By.XPATH,"//h2[@class='h4 mb-3']/a[contains(text(),'Contributors')]/span")
        Contributors_count.append(count.text)
    except NoSuchElementException:
        Contributors_count.append('No details available')
    time.sleep(2)

    languages=driver.find_elements(By.XPATH,"//li[@class='d-inline']/a//span[1]")
    if languages:
        for i in languages:
            l.append(i.text)
    else:
        l.append('No languages used')
    Language_used.append(l)

time.sleep(2)

```

In [97]:

```

print(len(repo_title),len( repo_desc),len(Contributors_count),len(Language_used))
25 25 25 25

```

In [100]:

```

Github = pd.DataFrame({})
Github['Repository Title'] = repo_title
Github['Repository Description'] = repo_desc
Github['Contributors Count'] = Contributors_count
Github['Language Used'] = Language_used
Github

```

Out[100]:

	Repository Title	Repository Description	Contributors Count	Language Used
0	artidoro / qlora	QLoRA: Efficient Finetuning of Quantized LLMs	3	[Jupyter Notebook, Python, Shell]
1	gptlink / gptlink	10分钟搭建自己可免费商用的ChatGPT环境，搭建简单，包含用户，订单，任务，付费等功能	7	[PHP, JavaScript, HTML]
2	ThePrimeagen / ts-rust-zig-deez	-	20	[OCaml, Elixir, Fortran, TypeScript, Zig, C++]
3	geohot / tinygrad	You like pytorch? You like micrograd? You love...	114	[Python, C, C++, Objective-C++, Shell, Assembly]
4	JushBJJ / Mr.-Ranedeer-AI-Tutor	A GPT-4 AI Tutor Prompt for customizable perso...	10	[No languages used]
5	microsoft / devicescript	TypeScript for Tiny IoT Devices	10	[TypeScript, C, Python, Starlark, JavaScript, ...]
6	xorvoid / sectorc	A C Compiler that fits in the 512 byte boot se...	2	[C, Assembly, Shell]
7	OpenGVLab / InternGPT	InternGPT (iGPT) is an open source demo platfo...	6	[Python, Cuda, C++]
8	microsoft / PowerToys	Windows system utilities to maximize productivity	366	[C#, C++, PowerShell, C, HTML, Batchfile]
9	Zeqiang-Lai / DragGAN	Online Demo and Implementation of DragGAN - "D...	3	[Python, Cuda, C++, Jupyter Notebook]
10	vanjs-org / van	VanJS: World's smallest reactive UI framework	2	[JavaScript, TypeScript]
11	kamiyaa / joshuto	ranger-like terminal file manager written in Rust	38	[Rust, Shell, Dockerfile]
12	microsoft / semantic-kernel	Integrate cutting-edge LLM technology quickly ...	84	[C#, Python]
13	AggressiveUser / AllForOne	-	2	[Python]
14	pengzhile / pandora	潘多拉，一个让你呼吸顺畅的ChatGPT。Pandora, a ChatGPT that h...	5	[Python, HTML]
15	openai / openai-cookbook	Examples and guides for using the OpenAI API	82	[Jupyter Notebook, Python, TypeScript, CSS]
16	immersive-translate / immersive-translate	Immersive Dual Web Page Translation Extension ...	17	[TypeScript, JavaScript, Makefile]
17	microsoft / devhome	Windows Dev Home Application	29	[C#, PowerShell]
18	salesforce / LAVIS	LAVIS - A One-stop Library for Language-Vision...	19	[Python, Jupyter Notebook, Shell]
19	hwchase17 / langchain	🦜 Building applications with LLMs through comp...	817	[Python, Jupyter Notebook]
20	Azure / cognitive-search-vector-pr	The official documentation and code samples fo...	6	[Jupyter Notebook, C#, JavaScript, Python]
21	iperov / DeepFaceLive	Real-time face swap for PC streaming or video ...	7	[Python]
22	chinese-poetry / chinese-poetry	The most comprehensive database of Chinese poe...	56	[JavaScript, Python]
23	GoogleCloudPlatform / generative-ai	Sample code and notebooks for Generative AI on...	10	[Jupyter Notebook]
24	AprilNEA / ChatGPT-Admin-Web	带有用户管理和后台管理系统的 ChatGPT WebUI	4	[TypeScript, SCSS, JavaScript, CSS]

In [101]: Github.to_csv('Github.csv')

In [102]: driver.close()

In []:

In []:

In []:

Scrape the details of top 100 songs on billboard.com. Url = <https://www.billboard.com/> You have to find the following details: A) Song name B) Artist name C) Last week rank D) Peak rank E) Weeks on board Note: - From the home page you have to click on the charts option then hot 100-page link through code.

In []:

In [120]: driver=webdriver.Chrome(r'/chrome.exe')
driver.maximize_window()
time.sleep(2)

In [122]: Url =driver.get(' https://www.billboard.com/')

In [125]: charts=driver.find_element(By.XPATH,"/html/body/div[3]/div[9]/div/div/div/ul/li[1]/h3/a").click()

In [69]: Song_Name = []

```
Artist_Name = []
Last_week_rank = []
Peak_rank = []
Weeks_on_board = []
```

```
In [70]: urls = driver.find_element(By.XPATH, "/html/body/div[3]/main/div[2]/div[1]/div[1]/div/div/div[1]/div[1]/div[2]/s
page_url = urls.get_attribute("href")
driver.get(page_url)
time.sleep(4)

try:
    name= driver.find_elements(By.XPATH, '//li[@class="lrv-u-width-100p"]/ul/li/h3')
    for i in name:
        Song_Name.append(i.text)
except NoSuchElementException:
    pass

try:
    art= driver.find_elements(By.XPATH, '//li[@class="lrv-u-width-100p"]/ul/li[1]/span')
    for i in art:
        Artist_Name.append(i.text)
except NoSuchElementException:
    pass

try:
    rank= driver.find_elements(By.XPATH, '//li[@class="lrv-u-width-100p"]/ul/li[4]/span')
    for i in rank:
        Last_week_rank.append(i.text)
except NoSuchElementException:
    pass

try:
    peak=driver.find_elements(By.XPATH, '//li[@class="lrv-u-width-100p"]/ul/li[5]/span')
    for i in peak:
        Peak_rank.append(i.text)
except NoSuchElementException:
    pass

try:
    board=driver.find_elements(By.XPATH, '//li[@class="lrv-u-width-100p"]/ul/li[6]/span')
    for i in board:
        Weeks_on_board.append(i.text)
except NoSuchElementException:
    pass
```

```
In [71]: print(len(Song_Name),len(Artist_Name),len(Last_week_rank),len(Peak_rank),len(Weeks_on_board) )

100 100 100 100 100
```

```
In [73]: billboard = pd.DataFrame({})
billboard['Song Name'] = Song_Name
billboard['Artist'] = Artist_Name
billboard['Last Week Rank'] = Last_week_rank
billboard['Peak Rank'] = Peak_rank
billboard['Weeks on board'] = Weeks_on_board
billboard
```

```
Out[73]:
```

	Song Name	Artist	Last Week Rank	Peak Rank	Weeks on board
0	Last Night	Morgan Wallen	1	1	16
1	All My Life	Lil Durk Featuring J. Cole	-	2	1
2	Flowers	Miley Cyrus	3	1	18
3	Kill Bill	SZA	2	1	23
4	Ella Baila Sola	Eslabon Armado X Peso Pluma	4	4	9
...
95	Forever	Lil Baby Featuring Fridayy	87	8	19
96	Private Landing	Don Toliver Featuring Justin Bieber & Future	100	72	6
97	I Heard	YoungBoy Never Broke Again	-	98	1
98	Sunrise	Morgan Wallen	89	30	11
99	Happy	NF	95	54	6

100 rows × 5 columns

```
In [74]: billboard.to_csv('Billiboard.csv')
driver.close()
```

```
In [ ]:
```

Scrape the details of highest selling novels. Url = <https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-all-time-fifty-shades-greycompare> You have to find the following details: A) Book name B) Author name C) Volumes sold D) Publisher E) Genre

In []:

```
In [17]: driver=webdriver.Chrome(r'C:/chrome.exe')
```

```
In [18]: driver.maximize_window()
time.sleep(2)
```

```
In [19]: URL=driver.get('https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-all-time-fifty-shades-
```

```
In [20]: for _ in range(20):
         driver.execute_script("window.scrollTo(0,100)")
```

```
In [21]: Book_name=[]
         Author_name=[]
         Volumes_sold=[]
         Publisher=[]
         Genre=[]
```

```
In [22]: try:
         name=driver.find_elements(By.XPATH,'//table[@class="in-article sortable"]/tbody/tr/td[2]')
         for i in name:
             Book_name.append(i.text)
     except NoSuchElementException:
         pass

     try:
         name2=driver.find_elements(By.XPATH,'//table[@class="in-article sortable"]/tbody/tr/td[3]')
         for i in name2:
             Author_name.append(i.text)
     except NoSuchElementException:
         pass

     try:
         sold=driver.find_elements(By.XPATH,'//table[@class="in-article sortable"]/tbody/tr/td[4]')
         for i in sold:
             Volumes_sold.append(i.text)
     except NoSuchElementException:
         pass

     try:
         publish=driver.find_elements(By.XPATH,'//table[@class="in-article sortable"]/tbody/tr/td[5]')
         for i in publish:
             Publisher.append(i.text)
     except NoSuchElementException:
         pass

     try:
         genre=driver.find_elements(By.XPATH,'//table[@class="in-article sortable"]/tbody/tr/td[6]')
         for i in genre:
             Genre.append(i.text)
     except NoSuchElementException:
         pass
```

```
In [23]: print(len(Book_name),len(Author_name),len(Volumes_sold),len(Publisher),len(Genre))
```

100 100 100 100 100

```
In [24]: novels=pd.DataFrame({'Book_name':Book_name,'Author_name':Author_name,'Volumes_sold':Volumes_sold,'Publisher':Pu
novels.head()
```

```
Out[24]:
```

	Book_name	Author_name	Volumes_sold	Publisher	Genre
0	Da Vinci Code,The	Brown, Dan	5,094,805	Transworld	Crime, Thriller & Adventure
1	Harry Potter and the Deathly Hallows	Rowling, J.K.	4,475,152	Bloomsbury	Children's Fiction
2	Harry Potter and the Philosopher's Stone	Rowling, J.K.	4,200,654	Bloomsbury	Children's Fiction
3	Harry Potter and the Order of the Phoenix	Rowling, J.K.	4,179,479	Bloomsbury	Children's Fiction
4	Fifty Shades of Grey	James, E. L.	3,758,936	Random House	Romance & Sagas

```
In [25]: novels.to_csv('Novels.csv')
```

```
In [26]: driver.close()
```

In []:

Scrape the details most watched tv series of all time from imdb.com. Url = <https://www.imdb.com/list/ls095964455/> You have to find the following details: A) Name B) Year span C) Genre D) Run time E) Ratings F) Votes

In []:


```
In [39]: driver=webdriver.Chrome(r'C:/chrome.exe')
```

```
In [42]: driver.maximize_window()
```

```
In [43]: Url = driver.get('https://www.imdb.com/list/ls095964455/')
```

```
In [44]: Name=[]
Year_span=[]
Genre=[]
Run_time=[]
Ratings=[]
Votes=[]
```

```
In [45]: try:
name=driver.find_elements(By.XPATH,'//div[@class="lister-list"]/div/div[2]/h3/a')
for i in name:
    Name.append(i.text)
except NoSuchElementException:
    pass

try:
year=driver.find_elements(By.XPATH,'//div[@class="lister-list"]/div/div[2]/h3/span[2]')
for i in year:
    Year_span.append(i.text)
except NoSuchElementException:
    pass

try:
genre=driver.find_elements(By.XPATH,'//div[@class="lister-list"]/div/div[2]/p/span[5]')
for i in genre:
    Genre.append(i.text)
except NoSuchElementException:
    pass

try:
time=driver.find_elements(By.XPATH,'//div[@class="lister-list"]/div/div[2]/p/span[3]')
for i in time:
    Run_time.append(i.text)
except NoSuchElementException:
    pass

try:
rating=driver.find_elements(By.XPATH,'//div[@class="lister-list"]/div/div[2]/div/div/span[2]')
for i in rating:
    Ratings.append(i.text)
except NoSuchElementException:
    pass

try:
vote=driver.find_elements(By.XPATH,'//div[@class="lister-list"]/div/div[2]/p[4]/span[2]')
for i in vote:
    Votes.append(i.text)
except NoSuchElementException:
    pass
```

```
In [46]: print(len(Name),len(Year_span),len(Genre),len(Run_time),len(Ratings),len(Votes))
```

```
100 100 100 100 100 100
```

```
In [47]: IMBD=pd.DataFrame({'Name':Name,'Year_span':Year_span,'Genre':Genre,'Run_time':Run_time,'Ratings':Ratings,'Votes':Votes})
IMBD.head()
```

```
Out[47]:
```

	Name	Year_span	Genre	Run_time	Ratings	Votes
0	Game of Thrones	(2011–2019)	Action, Adventure, Drama	57 min	9.2	2,161,809
1	Stranger Things	(2016–2022)	Drama, Fantasy, Horror	51 min	8.7	1,242,081
2	The Walking Dead	(2010–2022)	Drama, Horror, Thriller	44 min	8.1	1,027,034
3	13 Reasons Why	(2017–2020)	Drama, Mystery, Thriller	60 min	7.5	302,123
4	The 100	(2014–2020)	Drama, Mystery, Sci-Fi	43 min	7.6	261,217

```
In [48]: IMBD.to_csv('IMBD.csv')
```

```
In [49]: driver.close()
```

```
In [ ]:
```

Details of Datasets from UCI machine learning repositories. Url = <https://archive.ics.uci.edu/> You have to find the following details: A) Dataset name B) Data type C) Task D) Attribute type E) No of instances F) No of attribute G) Year

```
In [ ]:
```

```
In [30]: driver=webdriver.Chrome(r'C/chrome.exe')
```

```
In [31]: driver.maximize_window()  
time.sleep(2)
```

```
In [32]: Url =driver.get(' https://archive.ics.uci.edu/')
```

```
In [33]: datasets=driver.find_element(By.XPATH, '/html/body/table[1]/tbody/tr/td[2]/span[2]/a/font/b').click()
```

```
In [34]: Dataset_name=[]  
Data_type=[]  
Task=[]  
Attribute_type=[]  
No_of_instances=[]  
No_of_attribute=[]  
Year=[]
```

```
In [35]: try:  
    name=driver.find_elements(By.XPATH, '//p[@class="normal"]/b/a')  
    for i in name:  
        Dataset_name.append(i.text)  
except NoSuchElementException:  
    pass  
  
try:  
    data_type=driver.find_elements(By.XPATH, '//table[@border="1"]/tbody/tr/td[2]/p')  
    for i in data_type[1:]:  
        Data_type.append(i.text)  
except NoSuchElementException:  
    pass  
  
try:  
    task=driver.find_elements(By.XPATH, '//table[@border="1"]/tbody/tr/td[3]/p')  
    for i in task[1:]:  
        Task.append(i.text)  
except NoSuchElementException:  
    pass  
  
try:  
    type=driver.find_elements(By.XPATH, '//table[@border="1"]/tbody/tr/td[4]/p')  
    for i in type[1:]:  
        Attribute_type.append(i.text)  
except NoSuchElementException:  
    pass  
  
try:  
    inst=driver.find_elements(By.XPATH, '//table[@border="1"]/tbody/tr/td[5]/p')  
    for i in inst[1:]:  
        No_of_instances.append(i.text)  
except NoSuchElementException:  
    pass  
  
try:  
    att=driver.find_elements(By.XPATH, '//table[@border="1"]/tbody/tr/td[6]/p')  
    for i in att[1:]:  
        No_of_attribute.append(i.text)  
except NoSuchElementException:  
    pass  
  
try:  
    year=driver.find_elements(By.XPATH, '//table[@border="1"]/tbody/tr/td[7]/p')  
    for i in year[1:]:  
        Year.append(i.text)  
except NoSuchElementException:  
    pass
```

```
In [37]: print(len(Dataset_name),len(Data_type),len(Task),len(Attribute_type),len(No_of_instances),len(No_of_attribute),  
622 622 622 622 622 622 622
```

```
In [40]: ML = pd.DataFrame({})  
ML['Data Name'] = Dataset_name  
ML['Data Type '] = Data_type  
ML['Task '] = Task  
ML['Attribute Type '] = Attribute_type  
ML['No of Instance '] = No_of_instances  
ML['No of Attributes '] = No_of_attribute
```

```
ML['Year'] = Year
ML.head()
```

Out[40]:

	Data Name	Data Type	Task	Attribute Type	No of Instance	No of Attributes	Year
0	Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
1	Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
2	Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
3	Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
4	Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998

In [41]: ML.to_csv('ML.csv')

In [42]: driver.close()

In []:

Scrape the details of Data science recruiters Url = <https://www.naukri.com/hr-recruiters-consultants> You have to find the following details:
A) Name B) Designation C)Company D)Skills they hire for E) Location

In []:

In [46]: driver=webdriver.Chrome(r'C:/chrome.exe')

In [47]: driver.maximize_window()
time.sleep(2)

In [48]: Url =driver.get(' https://www.naukri.com/hr-recruiters-consultants')

In [49]: designation=driver.find_element(By.XPATH, '/html/body/div[2]/div[2]/div[1]/div[1]/form/div[1]/div/div[1]/div[1]/
designation.send_keys('data science')

In [50]: data_science=driver.find_element(By.XPATH, '/html/body/div[2]/div[2]/div[1]/div[1]/form/div[1]/div/div[1]/div[2]

In [52]: search=driver.find_element(By.XPATH, '/html/body/div[2]/div[2]/div[1]/div[1]/form/div[1]/button').click()

In [54]: Name=[]
Designation=[]
Company=[]
Skills_they_hire=[]
Location=[]

In [55]: try:
 name=driver.find_elements(By.XPATH, '//div[@class="outerRecSec"]/div/div/div/p/a/span')
 for i in name:
 Name.append(i.text)
except NoSuchElementException:
 pass

try:
 des=driver.find_elements(By.XPATH, '//div[@class="outerRecSec"]/div/div/div/p/span[1]')
 for i in des:
 Designation.append(i.text)
except NoSuchElementException:
 pass

try:
 company=driver.find_elements(By.XPATH, '//div[@class="outerRecSec"]/div/div/div/p/a[2]')
 for i in company:
 Company.append(i.text)
except NoSuchElementException:
 pass

try:
 skill=driver.find_elements(By.XPATH, '//div[@class="outerRecSec"]/div/div/div[2]')
 for i in skill:
 Skills_they_hire.append(i.text)
except NoSuchElementException:
 pass

try:
 loc=driver.find_elements(By.XPATH, '//div[@class="outerRecSec"]/div/div/div/p/span[2]/small')
 for i in loc:
 Location.append(i.text)
except NoSuchElementException:
 pass

In [56]: Recruiters=pd.DataFrame({})
Recruiters['Name']= Name[0:49]

```

Recruiters['Designation']= Designation[0:49]
Recruiters['Company']= Company[0:49]
Recruiters['Skills_they_hire']= Skills_they_hire[0:49]
Recruiters['Location']= Location[0:49]
Recruiters

```

Out[56]:

	Name	Designation	Company	Skills_they_hire	Location
0	Aakash Harit	HR Manager	Data Science Network	Classic ASP Developer, Internet Marketing Prof...	Delhi
1	shravan Kumar Gaddam	Company Recruiter	Shore Infotech India Pvt. Ltd	.Net, Java, Data Science, Linux Administration...	Hyderabad / Secunderabad
2	MARSIAN Technologies LLP	Company HR	MARSIAN Technologies LLP	Data Science, Artificial Intelligence, Machine...	Pune
3	Anik Agrawal	Company Recruiter	Enerlytics Software Solutions Pvt Ltd	Mean Stack, javascript, angularjs, mongodb, We...	Ahmedabad
4	subhas patel	Founder CEO	LibraryXProject	Hadoop, Spark, Digital Strategy, Data Architec...	UK - (london)
5	Abhishek - Only Analytics Hiring - India and	Recruitment Lead Consultant	Apidel Technologies Division of Transpower	Analytics, Business Intelligence, Business Ana...	Vadodara / Baroda
6	Institute for Financial Management and Resear	Programme Manager	IFMR	Data Science	Chennai
7	Balu Ramesh	HR Administrator	Techvantage Systems Pvt Ltd	Machine Learning, algorithms, Go Getter, Compu...	Trivandrum
8	Asif Lucknowi	Director	Weupskill- Live Wire India	Technical Training, Software Development, Pres...	Indore
9	InstaFinancials	Human Resource	CBL Data Science Private Limited	Software Development, It Sales, Account Manage...	Bengaluru / Bangalore
10	Kalpana Dumpala	Executive Hiring	Innominds Software	Qa, Ui/ux, Java Developer, Java Architect, C++...	Hyderabad / Secunderabad
11	Mubarak	Company HR	MoneyTap	Business Intelligence, Data Warehousing, Data ...	Bengaluru / Bangalore
12	Kushal Rastogi	Company HR	QuantMagnum Technologies Pvt. Ltd.	Office Administration, Hr Administration, tele...	Mumbai
13	Priyanka Akiri	HR Manager	Infinite Software Solutions	Oracle Db, Data Science, Data Warehousing, ET...	Hyderabad
14	Kapil Devang	HR Manager	BISP Solutions	Big Data, Hadoop, Data Analytics, Data Science	Bhopal
15	Mahesh Babu Channa	HR Team Lead	SocialPrachar.com	Social Media, digital media maketing, seo, smm...	Hyderabad / Secunderabad
16	Vaishnavi Kudalkar	HR Executive	Codeachive learning	Data Science, Python, Data Analytics	Mumbai
17	Sakshi Chhikara	Assistant Manager HR	BIZ INFOTECNO PRIVATE LIMITED	React.js, Data Science, Java, Front End, Busin...	Chandigarh
18	Ruchi Dhote	Senior Executive Talent Acquisition	Bristlecone India Ltd	Qlikview, Qlik Sense, Microsoft Azure, Power B...	Pune
19	Manisha Yadav	HR Executive	Easi Tax	Telecalling, Client Interaction, Marketing, Re...	Navi Mumbai
20	Riya Rajesh	Manager Talent Acquisition	Novelworx Digital Solutions	Data Science	Cochin
21	Rashmi Bhattacharjee	HR Head	AXESTRACK SOFTWARE SOLUTIONS PRIVATE...	Corporate Sales, Software Development, Softwar...	Delhi
22	Faizan Kareem	HR MANAGER	FirstTech Consaltants Pvt.Ltd	Data Analytics, Data Science, Machine Learning...	Hyderabad / Secunderabad
23	Rithika dadwal	HR Recruiter	Affine Analytics	Data Science, Machine Learning, Python, R, Dee...	Pune
24	Sandhya Khandagale	HR Recruiter	Compumatrice Multimedia Pvt Ltd	Big Data, Data Science, Artificial Intelligenc...	Pune
25	Shaun Rao	Manager Human Resources	Exela Technologies	Java, Net, Angularjs, Hr, Infrastructure, Mana...	Pune
26	Deeparchi Sharma	Company Recruiter	ZIGRAM	Research, Digital Marketing, Analytics, Softwa...	Gurgaon
27	Azahar Shaikh	Company Recruiter	NEAL ANALYTICS SERVICES PVT LTD	Data Science, Artificial Intelligence, Machine...	Pune
28	Manas	Lead Talent acquisition	Autumn Leaf Consulting Services Private...	Software Architecture, Vp Engineering, Product...	Bengaluru / Bangalore
29	kumar	Proprietor	trainin	Data Science, Hadoop, Rpas, Devops, Python, Aw...	Bengaluru / Bangalore
30	Sunil Vedula	CEO	Nanoprecise Sci Corp	Signal Processing, Machine Learning, Neural Ne...	Delhi
31	Rajat Kumar	Founder CEO	R.S Consultancy & Services	Web Technologies, Project Management, Software...	Bengaluru / Bangalore
32	Dhruv Dev Dubey	Company Recruitment Head	NETAPS FOUNDATION	Server Administartion, Verilog, Vhdl, Digital ...	Noida

33	Avnish Mishra	Senior Technical & Corporate Recruiter	RMS Risk Management Solutions	Data Science, Machine Learning, Deep Learning,...	Mysoru / Mysore
34	Jayanth N	Project Manager	Dollarbird Information Services Pvt, Ltd	Data Analytics, Managed Services, Team Leading...	Hyderabad / Secunderabad
35	Avodha	Business Development Associate	Nikitha Palaparathi	Ethical Hacking, Security Operations Center, S...	Bengaluru / Bangalore
36	Priya Khare	Senior Manager	Independent Consultant	Data Science, Artificial Intelligence, analyti...	New Delhi
37	Amit Sharma	Consultant	ASCO consulting	Machine Learning, Artificial Intelligence, Dat...	Chennai
38	Kanan	senior technology instructor	NY INST	C, C++, Artificial Intelligence, Python, Php, ...	Aligarh
39	Shashikant Chaudhary	HR Recruiter/HR Excutive	3D India Staffing Research & Consulting...	Relationship Management, Retail Sales, Private...	Salt Lake City
40	Brad	Manager, Technical Recruiting	O.C. Tanner	Data Science, Software Engineering	Pune
41	Rutuja Pawar	Technical Recruiter	Demand Matrix	Data Science, Big Data Analytics, Digital Mark...	Bengaluru / Bangalore
42	Madhusudhan Sridhar	Erp Implementer	MADHUSUDHAN SRIDHAR	Data Science, Recruitment, Salary	Mumbai
43	Ankit Sinha	Head Analytics	Suntech Global	B.Tech, Tableau, Statistics, R, Analytics, Tim...	Indore
44	Gaurav Chouhan	Chief Technical Officer	Strategic Consulting Lab	Software Development, Business Intelligence, B...	Bengaluru / Bangalore
45	Rashi Kacker	Sr Product Manager	Impel Labs Pvt. Ltd.	Data Science, Node.js, Angularjs	MYSORE
46	Ashwini	Director Global Delivery	MRP Advisers	Data Science, Media Marketing, Resource Planni...	Hyderabad / Secunderabad
47	Balaji Kolli	Co Founder	Saras Solutions India Pvt Ltd	Data Analysis, Learning, Data Science, Compute...	Bengaluru / Bangalore
48	Rajani Nagaraj	HR Manager	WildJasmine	Java, Hadoop, R, Machine Learning, Spark, Flum...	Mumbai

```
In [57]: Recruiters.to_csv('Recruters.csv')
driver.close()
```

```
In [ ]:
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js