```
#1. 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 [15]:

```
import pandas as pd
import selenium
from selenium import webdriver
from bs4 import BeautifulSoup
import warnings
warnings.filterwarnings("ignore")
import time
from selenium.common.exceptions import NoSuchElementException
```

In [39]:

```
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
driver.get("https://en.wikipedia.org/wiki/List_of_most-viewed_Indian_YouTube_videos#Top_vid
```

In [40]:

```
urls= []
for i in driver.find_elements_by_xpath('//a[@class="mw-jump-link"]'):
    urls.append(i.get_attribute("href"))
```

In [41]:

```
len(urls)
```

Out[41]:

```
2
```

In [53]:

```
Rank= []
Name= []
Artist= []
Upload_date= []
Views= []

for i in urls:
    driver.get(i)

    try:
        Rank_all = driver.find_element_by_xpath("/html/body/div[3]/div[3]/div[5]/div[1]/tab
        Rank.append(Rank_all.text)
    except:
        Rank.append('_')
    try:
        Name_title = driver.find_element_by_xpath("/html/body/div[3]/div[3]/div[5]/div[1]/t
        Name.append(Name_title.text)
    except:
        Name.append('_')
    try:
        Artist_total = driver.find_element_by_xpath('//*[@id="mw-content-text"]/div[1]/tabl
        Artist.append(Artist.text)
    except:
        Artist.append('-')
    try:
        Date= driver.find_element_by_xpath('//*[@id="mw-content-text"]/div[1]/table[1]/thea
        Upload_date.append(Date.text)
    except:
        Upload_date.append('-')
    try:
        Views_all = driver.find_element_by_xpath('//*[@id="mw-content-text"]/div[1]/table[1
        Views.append(Views.text)
    except:
        Views.append('-')



len(Rank)
len(Name)
len(Artist)
len(Upload_date)
len(Views)
```

Out[53]:

2

In [50]:

```
#making a dataframe

product_page = pd.DataFrame({})

product_page['Rank']= Rank
product_page['Name']= Name
product_page['Artist']= Artist
product_page['Upload_date']= Upload_date
product_page['Views']= Views

product_page
```

Out[50]:

	Rank	Name	Artist	Upload_date	Views
0	No.	-	-	Upload date	-
1	No.	-	-	Upload date	-

```
#
2.) Scrape the details team India’s international fixtures from bcci.tv. Url =
https://www.bcci.tv/.
You need to find following details:
A)  Match title (I.e. 1st ODI)
B)  Series
C)  Place
D)  Date
E)  Time
Note: - From bcci.tv home page you have reach to the international fixture page through
code.
```

In [16]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [17]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get("https://www.bcci.tv/")
from bs4 import BeautifulSoup
```

In [18]:

```
names=[]
```

In [19]:

```
name_tags=browser.find_elements_by_xpath("//p[@class='fixture__additional-info']")
for i in name_tags:
    if i.text is None :
        names.append("no rating")
    else:
        names.append(i.text)
names
```

Out[19]:

```
[]
```

In [20]:

```
series=[]
```

In [21]:

```
series_tags=browser.find_elements_by_xpath("//span[@class='u-unskewed-text fixture__format'
for i in series_tags:
    if i.text is None :
        series.append("no rating")
    else:
        series.append(i.text)
series
```

Out[21]:

```
[]
```

In [22]:

```
date=[]
```

In [23]:

```
date_tags=browser.find_elements_by_xpath("//span[@class='fixture__datetime tablet-only']")
for i in date_tags:
    if i.text is None :
        date.append("no rating")
    else:
        date.append(i.text)
date
```

Out[23]:

```
[]
```

In [24]:

```
import pandas as pd
bcci=pd.DataFrame({})
bcci['Names']=names[0:32]
bcci['Series']=series[0:32]
bcci['Date']=date[0:32]
```

In [25]:

```
bcci
```

Out[25]:

	Names	Series	Date

```
#
3.  Scrape the details of selenium exception from guru99.com. Url =
https://www.guru99.com/
You need to find following details:
A)  Name
B)  Description
Note: - From guru99 home page you have to reach to selenium exception handling page
through code.
```

In [26]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [27]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get(" https://www.guru99.com/")
from bs4 import BeautifulSoup
```

In [32]:

```
names=[]
```

In [38]:

```
name_tags=browser.find_elements_by_xpath("//li[@class='fa fa-chevron-circle-right']")
for i in name_tags:
    if i.text is None :
        names.append("no rating")
    else:
        names.append(i.text)
```

In [39]:

```
names[24]
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/108518708.py in <module>
----> 1 names[24]

IndexError: list index out of range
```

In [40]:

```
summary=[]
```

In [41]:

```
urls=browser.find_elements_by_tag_name('a')
UR=[]
for i in urls[:1]:
    UR.append(i.get_attribute('href'))
for url in UR:
    browser.get(url)

try:
    summaries=browser.find_element_by_tag_name('p')
    summary.append(summaries.text)
except NoSuchElementException   as e:
    summary.append("No rating")
```

In [37]:

```
summary
```

Out[37]:

```
['We make tons of efforts to take boredom out of learning and make it fun']
```

In [42]:

```
import pandas as pd
books=pd.DataFrame({})
books['Names']=names[24:25]
books['Summary']=summary[0:1]
```

In [43]:

```
books
```

Out[43]:

	Names	Summary
0	NaN	We make tons of efforts to take boredom out of...

```
#
4.  Scrape the details of State-wise GDP of India from statisticstime.com. Url =
http://statisticstimes.com/
```

```
You have to find following details:
A)  Rank
B)  State
C) GSDP(18-19)
D) GSDP(17-18)
E)  Share(2017)
F)  GDP($ billion)
Note: - From statisticstimes home page you have to reach to economy page through code.
```

In [73]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [74]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get("https://m.statisticstimes.com/economy/india/indian-states-gdp.php")
from bs4 import BeautifulSoup
```

In [75]:

```
ranks=[]
```

In [76]:

```
rank_tags=browser.find_elements_by_xpath("//td[@class='data1']")
for i in rank_tags:
    if i.text is None :
        ranks.append("no rating")
    else:
        ranks.append(i.text)
```

In [77]:

```
ranks[0:33]
```

Out[77]:

```
['1',
 '2',
 '3',
 '4',
 '5',
 '6',
 '7',
 '8',
 '9',
 '10',
 '11',
 '12',
 '13',
 '14',
 '15',
 '16',
 '17',
 '18',
 '19',
 '20',
 '21',
 '22',
 '23',
 '24',
 '25',
 '26',
 '27',
 '28',
 '29',
 '30',
 '31',
 '32',
 '33']
```

In [78]:

```
names=[]
```

In [247]:

```
name_tags=browser.find_elements_by_xpath("//td[@class='name']")
for i in name_tags:
    if i.text is None :
        names.append("no rating")
    else:
        names.append(i.text)
```

```
---------------------------------------------------------------------------
-
ConnectionRefusedError                    Traceback (most recent call las
t)
C:\ProgramData\Anaconda3\lib\site-packages\urllib3\connection.py in _new_c
onn(self)
    173         try:
--> 174             conn = connection.create_connection(
    175                 (self._dns_host, self.port), self.timeout, **extra
_kw

C:\ProgramData\Anaconda3\lib\site-packages\urllib3\util\connection.py in c
reate_connection(address, timeout, source_address, socket_options)
     95     if err is not None:
---> 96         raise err
     97

C:\ProgramData\Anaconda3\lib\site-packages\urllib3\util\connection.py in c
reate_connection(address, timeout, source_address, socket_options)
```

In [ ]:

```
names[0:33]
```

In [ ]:

```
gsdp19=[]
```

In [ ]:

```
gsdp19_tags=browser.find_elements_by_xpath("//td[@class='data']")
for i in gsdp19_tags:
    if i.text is None :
        gsdp19.append("no rating")
    else:
        gsdp19.append(i.text)
```

In [ ]:

```
gsdp19
```

In [ ]:

```
gsdp19.pop(1)
```

In [ ]:

```
gsdp19.pop(1)
```

In [ ]:

```
gsdp19.pop(1)
```

In [ ]:

```
gsdp19.pop(1)
```

In [ ]:

```
gsdp19.pop(2)
```

In [ ]:

```
gsdp19.pop(2)
```

In [ ]:

```
gsdp19.pop(2)
```

In [ ]:

```
gsdp19.pop(2)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(3)
```

In [ ]:

```
gsdp19.pop(4)
```

In [ ]:

```
gsdp19.pop(4)
```

In [ ]:

```
gsdp19.pop(4)
```

In [ ]:

```
gsdp19.pop(4)
```

In [ ]:

```
gsdp19.pop(5)
```

In [ ]:

```
gsdp19.pop(5)
```

In [ ]:

```
gsdp19.pop(5)
```

In [ ]:

```
gsdp19.pop(5)
```

In [ ]:

```
gsdp19.pop(6)
```

In [ ]:

```
gsdp19.pop(6)
```

In [ ]:

```
gsdp19.pop(6)
```

In [ ]:

```
gsdp19.pop(6)
```

In [ ]:

```
gsdp19.pop(7)
```

In [ ]:

```
gsdp19.pop(7)
```

In [ ]:

```
gsdp19.pop(7)
```

In [ ]:

```
gsdp19.pop(7)
```

In [ ]:

```
gsdp19.pop(8)
```

In [ ]:

```
gsdp19.pop(8)
```

In [ ]:

```
gsdp19.pop(8)
```

In [ ]:

```
gsdp19.pop(8)
```

In [ ]:

```
gsdp19.pop(9)
```

In [ ]:

```
gsdp19.pop(9)
```

In [ ]:

```
gsdp19.pop(9)
```

In [ ]:

```
gsdp19.pop(9)
```

In [ ]:

```
gsdp19.pop(10)
```

In [ ]:

```
gsdp19.pop(10)
```

In [ ]:

```
gsdp19.pop(10)
```

In [ ]:

```
gsdp19.pop(10)
```

In [ ]:

```
gsdp19[0:10]
```

In [ ]:

```
share=[]
```

In [ ]:

```
share_tags=browser.find_elements_by_xpath("//td[@class='data']")
for i in share_tags:
    if i.text is None :
        share.append("no rating")
    else:
        share.append(i.text)
```

In [ ]:

```
share.pop(0)
```

In [ ]:

```
share.pop(1)
```

In [ ]:

```
share.pop(1)
```

In [ ]:

```
share.pop(1)
```

In [ ]:

```
share.pop(1)
```

In [ ]:

```
share.pop(2)
```

In [ ]:

```
share.pop(2)
```

In [ ]:

```
share.pop(2)
```

In [ ]:

```
share.pop(2)
```

In [ ]:

```
share.pop(3)
```

In [ ]:

```
share.pop(3)
```

In [ ]:

```
share.pop(3)
```

In [ ]:

```
share.pop(3)
```

In [ ]:

```
share.pop(4)
```

In [ ]:

```
share.pop(4)
```

In [ ]:

```
share.pop(4)
```

In [ ]:

```
share.pop(4)
```

In [ ]:

```
share.pop(5)
```

In [ ]:

```
share.pop(5)
```

In [ ]:

```
share.pop(5)
```

In [248]:

```
share.pop(5)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/1302928070.py in <module>
----> 1 share.pop(5)

IndexError: pop from empty list
```

In [249]:

```
share.pop(6)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/2405628314.py in <module>
----> 1 share.pop(6)

IndexError: pop from empty list
```

In [250]:

```
share.pop(6)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/2405628314.py in <module>
----> 1 share.pop(6)

IndexError: pop from empty list
```

In [251]:

```
share.pop(6)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/2405628314.py in <module>
----> 1 share.pop(6)

IndexError: pop from empty list
```

In [252]:

```
share.pop(6)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/2405628314.py in <module>
----> 1 share.pop(6)

IndexError: pop from empty list
```

In [253]:

```
share.pop(7)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3622724915.py in <module>
----> 1 share.pop(7)

IndexError: pop from empty list
```

In [254]:

```
share.pop(7)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3622724915.py in <module>
----> 1 share.pop(7)

IndexError: pop from empty list
```

In [255]:

```
share.pop(7)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3622724915.py in <module>
----> 1 share.pop(7)

IndexError: pop from empty list
```

In [256]:

```
share.pop(7)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3622724915.py in <module>
----> 1 share.pop(7)

IndexError: pop from empty list
```

In [257]:

```
share.pop(8)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3390945467.py in <module>
----> 1 share.pop(8)

IndexError: pop from empty list
```

In [ ]:

```
share.pop(8)
```

In [258]:

```
share.pop(8)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3390945467.py in <module>
----> 1 share.pop(8)

IndexError: pop from empty list
```

In [259]:

```
share.pop(8)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3390945467.py in <module>
----> 1 share.pop(8)

IndexError: pop from empty list
```

In [260]:

```
share.pop(9)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3934022383.py in <module>
----> 1 share.pop(9)

IndexError: pop from empty list
```

In [261]:

```
share.pop(9)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3934022383.py in <module>
----> 1 share.pop(9)

IndexError: pop from empty list
```

In [262]:

```
share.pop(9)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3934022383.py in <module>
----> 1 share.pop(9)

IndexError: pop from empty list
```

In [263]:

```
share.pop(9)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/3934022383.py in <module>
----> 1 share.pop(9)

IndexError: pop from empty list
```

In [264]:

```
share.pop(10)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/22330097.py in <module>
----> 1 share.pop(10)

IndexError: pop from empty list
```

In [265]:

```
share.pop(10)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/22330097.py in <module>
----> 1 share.pop(10)

IndexError: pop from empty list
```

In [266]:

```
share.pop(10)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/22330097.py in <module>
----> 1 share.pop(10)

IndexError: pop from empty list
```

In [267]:

```
share.pop(10)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/22330097.py in <module>
----> 1 share.pop(10)

IndexError: pop from empty list
```

In [268]:

```
share[0:10]
```

Out[268]:

```
[]
```

In [269]:

```
gsdp18=[]
```

In [270]:

```
gsdp18_tags=browser.find_elements_by_xpath("//td[@class='data sorting_1']")
for i in gsdp18_tags:
    if i.text is None :
        gsdp18.append("no rating")
    else:
        gsdp18.append(i.text)
```

```
---------------------------------------------------------------------------
-
ConnectionRefusedError                    Traceback (most recent call las
t)
C:\ProgramData\Anaconda3\lib\site-packages\urllib3\connection.py in _new_c
onn(self)
    173         try:
--> 174             conn = connection.create_connection(
    175                 (self._dns_host, self.port), self.timeout, **extra
_kw

C:\ProgramData\Anaconda3\lib\site-packages\urllib3\util\connection.py in c
reate_connection(address, timeout, source_address, socket_options)
     95     if err is not None:
---> 96         raise err
     97

C:\ProgramData\Anaconda3\lib\site-packages\urllib3\util\connection.py in c
reate_connection(address, timeout, source_address, socket_options)
```

In [ ]:

```
gsdpbi.pop(0)
```

In [ ]:

```
gsdpbi.pop(0)
```

In [ ]:

```
gsdpbi.pop(1)
```

In [ ]:

```
gsdpbi.pop(1)
```

In [ ]:

```
gsdpbi.pop(1)
```

In [ ]:

```
gsdpbi.pop(1)
```

In [ ]:

```
gsdpbi.pop(2)
```

In [ ]:

```
gsdpbi.pop(2)
```

In [ ]:

```
gsdpbi.pop(2)
```

In [ ]:

```
gsdpbi.pop(2)
```

In [ ]:

```
gsdpbi.pop(3)
```

In [ ]:

```
gsdpbi.pop(3)
```

In [ ]:

```
gsdpbi.pop(3)
```

In [ ]:

```
gsdpbi.pop(3)
```

In [ ]:

```
gsdpbi.pop(4)
```

In [ ]:

```
gsdpbi.pop(4)
```

In [ ]:

```
gsdpbi.pop(4)
```

In [ ]:

```
gsdpbi.pop(4)
```

In [ ]:

```
gsdpbi.pop(5)
```

In [ ]:

```
gsdpbi.pop(5)
```

In [ ]:

```
gsdpbi.pop(5)
```

In [ ]:

```
gsdpbi.pop(5)
```

In [ ]:

```
gsdpbi.pop(6)
```

In [ ]:

```
gsdpbi.pop(6)
```

In [ ]:

```
gsdpbi.pop(6)
```

In [ ]:

```
gsdpbi.pop(6)
```

In [ ]:

```
gsdpbi.pop(7)
```

In [ ]:

```
gsdpbi.pop(7)
```

In [ ]:

```
gsdpbi.pop(7)
```

In [ ]:

```
gsdpbi.pop(7)
```

In [ ]:

```
gsdpbi.pop(8)
```

In [ ]:

```
gsdpbi.pop(8)
```

In [ ]:

```
gsdpbi.pop(8)
```

In [ ]:

```
gsdpbi.pop(8)
```

In [ ]:

```
gsdpbi.pop(9)
```

In [ ]:

```
gsdpbi.pop(9)
```

In [ ]:

```
gsdpbi.pop(9)
```

In [ ]:

```
gsdpbi.pop(9)
```

In [ ]:

```
gsdpbi.pop(10)
```

In [ ]:

```
gsdpbi.pop(10)
```

In [ ]:

```
gsdpbi.pop(10)
```

In [271]:

```
gsdpbi.pop(10)
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/1010801503.py in <module>
----> 1 gsdpbi.pop(10)

IndexError: pop from empty list
```

In [272]:

```
gsdpbi[0:10]
```

Out[272]:

```
[]
```

In [273]:

```
import pandas as pd
stats=pd.DataFrame({})
stats['Rank']=ranks[0:10]
stats['Names']=names[0:10]
stats['gsdp19']=gsdp19[0:10]
stats['gsdp18']=gsdp18[0:10]
stats['gsdpbi']=gsdp18[0:10]
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/1266626521.py in <module>
      4 stats['Names']=names[0:10]
      5 stats['gsdp19']=gsdp19[0:10]
----> 6 stats['gsdp18']=gsdp18[0:10]
      7 stats['gsdpbi']=gsdp18[0:10]

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in __setitem
__(self, key, value)
   3610         else:
   3611             # set column
-> 3612             self._set_item(key, value)
   3613
   3614     def _setitem_slice(self, key: slice, value):

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _set_item
(self, key, value)
   3782         ensure homogeneity.
   3783         """
-> 3784         value = self._sanitize_column(value)
   3785
   3786         if (

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _sanitize
_column(self, value)
   4507
   4508         if is_list_like(value):
-> 4509             com.require_length_match(value, self.index)
   4510         return sanitize_array(value, self.index, copy=True, allow_2d
=True)
   4511

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\common.py in require_
length_match(data, index)
    529     """
    530     if len(data) != len(index):
--> 531         raise ValueError(
    532             "Length of values "
    533             f"({len(data)}) "

ValueError: Length of values (0) does not match length of index (10)
```

In [274]:

```
stats
```

Out[274]:

	Rank	Names	gsdp19
0	1	Maharashtra	-
1	2	Tamil Nadu	1,845,853
2	3	Uttar Pradesh	1,687,818
3	4	Gujarat	1,631,977
4	5	Karnataka	1,253,832
5	6	West Bengal	1,020,989
6	7	Rajasthan	969,604
7	8	Andhra Pradesh	117.703
8	9	Telangana	111.519
9	10	Madhya Pradesh	80.562

```
#
5.  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 [291]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [292]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get("https://github.com/trending")
from bs4 import BeautifulSoup
```

In [293]:

```
title=[]
```

In [294]:

```
title_tags=browser.find_elements_by_xpath("//h1[@class='h3 lh-condensed']")
for i in title_tags:
    if i.text is None :
        title.append("no rating")
    else:
        title.append(i.text)
```

In [295]:

```
title
```

Out[295]:

```
['pocketbase / pocketbase',
 'WongKinYiu / yolov7',
 'All-Cups / aicup22',
 'vinta / awesome-python',
 'codecrafters-io / build-your-own-x',
 'sunface / rust-by-practice',
 'qiangmzsx / Software-Engineering-at-Google',
 'public-apis / public-apis',
 'h5bp / Front-end-Developer-Interview-Questions',
 'Developer-Y / cs-video-courses',
 'Jarred-Sumner / bun',
 'ziglang / zig',
 'yangshun / tech-interview-handbook',
 'jinfagang / yolov7',
 'EbookFoundation / free-programming-books',
 'PINTO0309 / PINTO_model_zoo',
 'goabstract / Awesome-Design-Tools',
 'ossu / computer-science',
 'ShareX / ShareX',
 'bradtraversy / design-resources-for-developers',
 'facebookresearch / fairseq',
 'microsoft / Web-Dev-For-Beginners',
 'moment / moment',
 'sensity-ai / dot',
 'sindresorhus / awesome']
```

In [296]:

```
description=[]
```

In [297]:

```
description_tags=browser.find_elements_by_xpath("//p[@class='col-9 color-text-secondary my-
for i in description_tags:
    if i.text is None :
        description.append("no rating")
    else:
        description.append(i.text)
```

In [298]:

```
description
```

Out[298]:

```
[]
```

In [299]:

```
language=[]
```

In [300]:

```
language_tags=browser.find_elements_by_xpath("//span[@class='d-inline-block ml-0 mr-3']")
for i in language_tags:
    if i.text is None :
        language.append("no rating")
    else:
        language.append(i.text)
```

In [301]:

```
language
```

Out[301]:

```
['Go',
 'Python',
 'Java',
 'Python',
 'Rust',
 'HTML',
 'Python',
 'Nunjucks',
 'Zig',
 'Zig',
 'JavaScript',
 'Python',
 'Python',
 'JavaScript',
 'C#',
 'Python',
 'JavaScript',
 'JavaScript',
 'Python']
```

In [302]:

```
count=[]
```

In [303]:

```
count_tags=browser.find_elements_by_xpath("//svg[@class='octicon octicon-star']")
for i in count_tags:
    if i.text is None :
        count.append("no rating")
    else:
        count.append(i.text)
```

In [304]:

```
count[21:]
```

Out[304]:

```
[]
```

In [308]:

```
import pandas as pd
github=pd.DataFrame({})
github['title']=title[0:20]
github['description']=description[0:20]
github['language']=language[0:20]
github['count']=count[21:41]
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/4027418294.py in <module>
      2 github=pd.DataFrame({})
      3 github['title']=title[0:20]
----> 4 github['description']=description[0:20]
      5 github['language']=language[0:20]
      6 github['count']=count[21:41]

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in __setitem
__(self, key, value)
   3610         else:
   3611             # set column
-> 3612             self._set_item(key, value)
   3613
   3614     def _setitem_slice(self, key: slice, value):

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _set_item
(self, key, value)
   3782         ensure homogeneity.
   3783         """
-> 3784         value = self._sanitize_column(value)
   3785
   3786         if (

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _sanitize
_column(self, value)
   4507
   4508         if is_list_like(value):
-> 4509             com.require_length_match(value, self.index)
   4510         return sanitize_array(value, self.index, copy=True, allow_2d
=True)
   4511

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\common.py in require_
length_match(data, index)
    529     """
    530     if len(data) != len(index):
--> 531         raise ValueError(
    532             "Length of values "
    533             f"({len(data)}) "

ValueError: Length of values (0) does not match length of index (20)
```

In [307]:

```
github
```

Out[307]:

	title
0	pocketbase / pocketbase
1	WongKinYiu / yolov7
2	All-Cups / aicup22
3	vinta / awesome-python
4	codecrafters-io / build-your-own-x
5	sunface / rust-by-practice
6	qiangmzsx / Software-Engineering-at-Google
7	public-apis / public-apis
8	h5bp / Front-end-Developer-Interview-Questions
9	Developer-Y / cs-video-courses
10	Jarred-Sumner / bun
11	ziglang / zig
12	yangshun / tech-interview-handbook
13	jinfagang / yolov7
14	EbookFoundation / free-programming-books
15	PINTO0309 / PINTO_model_zoo
16	goabstract / Awesome-Design-Tools
17	ossu / computer-science
18	ShareX / ShareX
19	bradtraversy / design-resources-for-developers

```
#

7.  Scrape the details of Data science recruiters from naukri.com. Url =
https://www.naukri.com/
You have to find the following details:
A)  Name
B)  Designation
C)  Company
D)  Skills they hire for
E)  Location
Note: - From naukri.com homepage click on the recruiters option and the on the search
pane type Data science and click on search. All this should be done through code
```

In [ ]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [348]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get("https://www.naukri.com/")
from bs4 import BeautifulSoup
```

In [349]:

```
search_field_designation=browser.find_element_by_id("qsb-keyword-sugg")
search_field_location=browser.find_element_by_id("qsb-location-sugg")
search_field_designation.send_keys("Data Analyst")
search_field_location.send_keys("Bangalore")
```

```
---------------------------------------------------------------------------
NoSuchElementException                    Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_29492/2087876277.py in <module>
----> 1 search_field_designation=browser.find_element_by_id("qsb-keyword-sug
g")
      2 search_field_location=browser.find_element_by_id("qsb-location-sugg"
)
      3 search_field_designation.send_keys("Data Analyst")
      4 search_field_location.send_keys("Bangalore")

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_id(self, id_)
    476             stacklevel=2,
    477         )
--> 478         return self.find_element(by=By.ID, value=id_)
    479
    480     def find_elements_by_id(self, id_) -> List[WebElement]:

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
   1249             value = '[name="%s"]' % value
   1250
-> 1251         return self.execute(Command.FIND_ELEMENT, {
   1252             'using': by,
   1253             'value': value})['value']

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428         response = self.command_executor.execute(driver_command, par
ams)
    429         if response:
--> 430             self.error_handler.check_response(response)
    431             response['value'] = self._unwrap_value(
    432                 response.get('value', None))

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245                 alert_text = value['alert'].get('text')
    246             raise exception_class(message, screen, stacktrace, alert
_text)  # type: ignore[call-arg]  # mypy is not smart enough here
--> 247         raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:

NoSuchElementException: Message: no such element: Unable to locate element:
 {"method":"css selector","selector":"[id="qsb-keyword-sugg"]"}
  (Session info: chrome=103.0.5060.114)
Stacktrace:
Backtrace:
        Ordinal0 [0x00F5D953+2414931]
        Ordinal0 [0x00EEF5E1+1963489]
        Ordinal0 [0x00DDC6B8+837304]
        Ordinal0 [0x00E09500+1021184]
```

```
        Ordinal0 [0x00E0979B+1021851]
        Ordinal0 [0x00E36502+1205506]
        Ordinal0 [0x00E244E4+1131748]
        Ordinal0 [0x00E34812+1198098]
        Ordinal0 [0x00E242B6+1131190]
        Ordinal0 [0x00DFE860+976992]
        Ordinal0 [0x00DFF756+980822]
        GetHandleVerifier [0x011CCC62+2510274]
        GetHandleVerifier [0x011BF760+2455744]
        GetHandleVerifier [0x00FEEABA+551962]
        GetHandleVerifier [0x00FED916+547446]
        Ordinal0 [0x00EF5F3B+1990459]
        Ordinal0 [0x00EFA898+2009240]
        Ordinal0 [0x00EFA985+2009477]
        Ordinal0 [0x00F03AD1+2046673]
        BaseThreadInitThunk [0x77216739+25]
        RtlGetFullPathName_UEx [0x77D38FEF+1215]
        RtlGetFullPathName_UEx [0x77D38FBD+1165]
```

In [ ]:

```
# clicking the search button
search_button=browser.find_element_by_xpath("//div[@class='search-btn']/button")
search_button.click()
```

In [ ]:

```
titles=[]
```

In [ ]:

```
title_tags=browser.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
for i in name_tags:
    if i.text is None :
        titles.append("no rating")
    else:
        titles.append(i.text)
```

In [350]:

```
titles
```

Out[350]:

```
[]
```

In [351]:

```
company=[]
```

In [354]:

```
company_tags=browser.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
for i in company_tags:
    if i.text is None :
        company.append("no rating")
    else:
        company.append(i.text)
```

In [355]:

```
company
```

Out[355]:

```
[]
```

In [356]:

```
company=[]
```

In [386]:

```
<a href="https://www.naukri.com/glaxosmithkline-pharmaceuticals-jobs-careers-16369" class="
```

```
  File "C:\Users\hamsa\AppData\Local\Temp/ipykernel_29492/1063067744.py", li
ne 1
    <a href="https://www.naukri.com/glaxosmithkline-pharmaceuticals-jobs-car
eers-16369" class="subTitle ellipsis fleft" target="_blank" title="GlaxoSmit
hKline Pharmaceuticals Limited">GlaxoSmithKline Pharmaceuticals Limited</a>
    ^
SyntaxError: invalid syntax
```

In [382]:

```
company_tags=browser.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
for i in company_tags:
    if i.text is None :
        company.append("no rating")
    else:
        company.append(i.text)
```

```
---------------------------------------------------------------------------
-
WebDriverException                        Traceback (most recent call las
t)
~\AppData\Local\Temp/ipykernel_29492/3137245688.py in <module>
----> 1 company_tags=browser.find_elements_by_xpath("//a[@class='subTitle
 ellipsis fleft']")
      2 for i in company_tags:
      3     if i.text is None :
      4         company.append("no rating")
      5     else:

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdr
iver.py in find_elements_by_xpath(self, xpath)
    547             stacklevel=2,
    548         )
--> 549         return self.find_elements(by=By.XPATH, value=xpath)
    550
    551     def find_element_by_link_text(self, link_text) -> WebElement:
```

In [367]:

```
company
```

Out[367]:

```
[]
```

In [368]:

```
skill=[]
```

In [369]:

```
skill_tags=browser.find_elements_by_xpath("//ul[@class='fleft fs12 grey-text lh16 dot']/li[
for i in skill_tags:
    if i.text is None :
        skill.append("no rating")
    else:
        skill.append(i.text)
```

In [370]:

```
job_location=[]
```

In [371]:

```
#scraping the job-location
locations=browser.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHolderLi lo
for i in locations:
    if i.text is None :
        job_location.append("--")
    else:
        job_location.append(i.text)
job_location
```

Out[371]:

```
[]
```

In [372]:

```
import pandas as pd
naukri=pd.DataFrame({})
naukri['title']=titles[0:20]
naukri['company']=company[0:20]
naukri['location']=job_location[0:20]
naukri['skill']=skill[0:20]
naukri
```

Out[372]:

	title	company	location	skill

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

In [2]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [3]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get("https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-all-t
from bs4 import BeautifulSoup
```

In [4]:

```
books=[]
```

In [5]:

```
book_tags=browser.find_elements_by_xpath("//td[@class='left']")
for i in book_tags:
    if i.text is None :
        books.append("no rating")
    else:
        books.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/1463867140.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  book_tags=browser.find_elements_by_xpath("//td[@class='left']")
```

In [6]:

```
books.pop(0)
```

Out[6]:

```
'1'
```

In [7]:

```
books.pop(1)
```

Out[7]:

```
'Brown, Dan'
```

In [8]:

```
books.pop(1)
```

Out[8]:

```
'5,094,805'
```

In [9]:

```
books.pop(1)
```

Out[9]:

```
'Transworld'
```

In [11]:

```
books.pop(1)
```

Out[11]:

```
'2'
```

In [12]:

```
books.pop(2)
```

Out[12]:

```
'4,475,152'
```

In [13]:

```
books.pop(2)
```

Out[13]:

```
'Bloomsbury'
```

In [14]:

```
books.pop(2)
```

Out[14]:

```
'3'
```

In [15]:

```
books.pop(2)
```

Out[15]:

```
"Harry Potter and the Philosopher's Stone"
```

In [16]:

```
books.pop(3)
```

Out[16]:

```
'4,200,654'
```

In [17]:

```
books.pop(3)
```

Out[17]:

```
'Bloomsbury'
```

In [18]:

```
books.pop(3)
```

Out[18]:

```
'4'
```

In [19]:

```
books.pop(3)
```

Out[19]:

```
'Harry Potter and the Order of the Phoenix'
```

In [20]:

```
books.pop(4)
```

Out[20]:

```
'4,179,479'
```

In [21]:

```
books.pop(4)
```

Out[21]:

```
'Bloomsbury'
```

In [22]:

```
books.pop(4)
```

Out[22]:

```
'5'
```

In [23]:

```
books.pop(4)
```

Out[23]:

```
'Fifty Shades of Grey'
```

In [24]:

```
books.pop(5)
```

Out[24]:

```
'3,758,936'
```

In [25]:

```
books.pop(5)
```

Out[25]:

```
'Random House'
```

In [26]:

```
books.pop(5)
```

Out[26]:

```
'6'
```

In [27]:

```
books.pop(5)
```

Out[27]:

```
'Harry Potter and the Goblet of Fire'
```

In [28]:

```
books.pop(6)
```

Out[28]:

```
'3,583,215'
```

In [29]:

```
books.pop(6)
```

Out[29]:

```
'Bloomsbury'
```

In [30]:

```
books.pop(6)
```

Out[30]:

```
'7'
```

In [31]:

```
books.pop(6)
```

Out[31]:

```
'Harry Potter and the Chamber of Secrets'
```

In [32]:

```
books.pop(7)
```

Out[32]:

```
'3,484,047'
```

In [33]:

```
books.pop(7)
```

Out[33]:

```
'Bloomsbury'
```

In [34]:

```
books.pop(7)
```

Out[34]:

```
'8'
```

In [35]:

```
books.pop(7)
```

Out[35]:

```
'Harry Potter and the Prisoner of Azkaban'
```

In [36]:

```
books.pop(8)
```

Out[36]:

```
'3,377,906'
```

In [37]:

```
books.pop(8)
```

Out[37]:

```
'Bloomsbury'
```

In [38]:

```
books.pop(8)
```

Out[38]:

```
'9'
```

In [39]:

```
books.pop(8)
```

Out[39]:

```
'Angels and Demons'
```

In [40]:

```
books.pop(9)
```

Out[40]:

```
'3,193,946'
```

In [41]:

```
books.pop(9)
```

Out[41]:

```
'Transworld'
```

In [42]:

```
books.pop(9)
```

Out[42]:

```
'10'
```

In [43]:

```
books.pop(9)
```

Out[43]:

```
"Harry Potter and the Half-blood Prince:Children's Edition"
```

In [44]:

```
books.pop(10)
```

Out[44]:

```
'2,950,264'
```

In [45]:

```

books.pop(10)
```

Out[45]:

```
'Bloomsbury'
```

In [46]:

```
books.pop(10)
```

Out[46]:

```
'11'
```

In [47]:

```
books.pop(10)
```

Out[47]:

```
'Fifty Shades Darker'
```

In [48]:

```
books.pop(11)
```

Out[48]:

```
'2,479,784'
```

In [49]:

```
books.pop(11)
```

Out[49]:

```
'Random House'
```

In [50]:

```
books.pop(11)
```

Out[50]:

```
'12'
```

In [51]:

```
books.pop(11)
```

Out[51]:

```
'Twilight'
```

In [52]:

```
books.pop(12)
```

Out[52]:

```
'2,315,405'
```

In [53]:

```
books.pop(12)
```

Out[53]:

```
'Little, Brown Book'
```

In [54]:

```
books.pop(12)
```

Out[54]:

```
'13'
```

In [55]:

```
books.pop(12)
```

Out[55]:

```
'Girl with the Dragon Tattoo,The:Millennium Trilogy'
```

In [56]:

```
books.pop(13)
```

Out[56]:

```
'2,233,570'
```

In [57]:

```
books.pop(13)
```

Out[57]:

```
'Quercus'
```

In [58]:

```
books.pop(13)
```

Out[58]:

```
'14'
```

In [59]:

```
books.pop(13)
```

Out[59]:

```
'Fifty Shades Freed'
```

In [60]:

```
books.pop(14)
```

Out[60]:

```
'2,193,928'
```

In [61]:

```
books.pop(14)
```

Out[61]:

```
'Random House'
```

In [62]:

```
books.pop(14)
```

Out[62]:

```
'15'
```

In [63]:

```
books.pop(14)
```

Out[63]:

```
'Lost Symbol,The'
```

In [64]:

```
books.pop(15)
```

Out[64]:

```
'2,183,031'
```

In [65]:

```
books.pop(15)
```

Out[65]:

```
'Transworld'
```

In [66]:

```
books.pop(15)
```

Out[66]:

```
'16'
```

In [67]:

```
books.pop(15)
```

Out[67]:

```
'New Moon'
```

In [68]:

```
books.pop(16)
```

Out[68]:

```
'2,152,737'
```

In [69]:

```
books.pop(16)
```

Out[69]:

```
'Little, Brown Book'
```

In [70]:

```
books.pop(16)
```

Out[70]:

```
'17'
```

In [71]:

```
books.pop(16)
```

Out[71]:

```
'Deception Point'
```

In [72]:

```
books.pop(17)
```

Out[72]:

```
'2,062,145'
```

In [73]:

```
books.pop(17)
```

Out[73]:

```
'Transworld'
```

In [74]:

```
books.pop(17)
```

Out[74]:

```
'18'
```

In [75]:

```
books.pop(17)
```

Out[75]:

```
'Eclipse'
```

In [76]:

```
books.pop(18)
```

Out[76]:

```
'2,052,876'
```

In [77]:

```
books.pop(18)
```

Out[77]:

```
'Little, Brown Book'
```

In [78]:

```
books.pop(18)
```

Out[78]:

```
'19'
```

In [79]:

```
books.pop(18)
```

Out[79]:

```
'Lovely Bones,The'
```

In [80]:

```
books.pop(19)
```

Out[80]:

```
'2,005,598'
```

In [81]:

```
books.pop(19)
```

Out[81]:

```
'Pan Macmillan'
```

In [82]:

```
books.pop(19)
```

Out[82]:

```
'20'
```

In [83]:

```
books.pop(19)
```

Out[83]:

```
'Curious Incident of the Dog in the Night-time,The'
```

In [84]:

```
books.pop(20)
```

Out[84]:

```
'1,979,552'
```

In [85]:

```
books.pop(20)
```

Out[85]:

```
'Random House'
```

In [86]:

```
books.pop(20)
```

Out[86]:

```
'21'
```

In [87]:

```

books.pop(20)
```

Out[87]:

```
'Digital Fortress'
```

In [88]:

```
books[0:20]
```

Out[88]:

```
['Da Vinci Code,The',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'James, E. L.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Brown, Dan',
 'Rowling, J.K.',
 'James, E. L.',
 'Meyer, Stephenie',
 'Larsson, Stieg',
 'James, E. L.',
 'Brown, Dan',
 'Meyer, Stephenie',
 'Brown, Dan',
 'Meyer, Stephenie',
 'Sebold, Alice',
 'Haddon, Mark']
```

In [89]:

```
author=[]
```

In [90]:

```
author_tags=browser.find_elements_by_xpath("//td[@class='left']")
for i in author_tags:
    if i.text is None :
        author.append("no rating")
    else:
        author.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/3162954424.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  author_tags=browser.find_elements_by_xpath("//td[@class='left']")
```

In [91]:

```
author
```

Out[91]:

```
['1',
 'Da Vinci Code,The',
 'Brown, Dan',
 '5,094,805',
 'Transworld',
 '2',
 'Harry Potter and the Deathly Hallows',
 'Rowling, J.K.',
 '4,475,152',
 'Bloomsbury',
 '3',
 "Harry Potter and the Philosopher's Stone",
 'Rowling, J.K.',
 '4,200,654',
 'Bloomsbury',
 '4',
 'Harry Potter and the Order of the Phoenix',
 'Rowling, J.K.',
```

In [92]:

```
author.pop(0)
```

Out[92]:

```
'1'
```

In [93]:

```
author.pop(0)
```

Out[93]:

```
'Da Vinci Code,The'
```

In [94]:

```
author.pop(1)
```

Out[94]:

```
'5,094,805'
```

In [95]:

```
author.pop(1)
```

Out[95]:

```
'Transworld'
```

In [96]:

```
author.pop(1)
```

Out[96]:

```
'2'
```

In [97]:

```
author.pop(1)
```

Out[97]:

```
'Harry Potter and the Deathly Hallows'
```

In [98]:

```
author.pop(2)
```

Out[98]:

```
'4,475,152'
```

In [99]:

```
author.pop(2)
```

Out[99]:

```
'Bloomsbury'
```

In [100]:

```
author.pop(2)
```

Out[100]:

```
'3'
```

In [101]:

```
author.pop(2)
```

Out[101]:

```
"Harry Potter and the Philosopher's Stone"
```

In [102]:

```
author.pop(3)
```

Out[102]:

```
'4,200,654'
```

In [103]:

```
author.pop(3)
```

Out[103]:

```
'Bloomsbury'
```

In [104]:

```
author.pop(3)
```

Out[104]:

```
'4'
```

In [105]:

```
author.pop(3)
```

Out[105]:

```
'Harry Potter and the Order of the Phoenix'
```

In [106]:

```
author.pop(4)
```

Out[106]:

```
'4,179,479'
```

In [107]:

```
author.pop(4)
```

Out[107]:

```
'Bloomsbury'
```

In [108]:

```
author.pop(4)
```

Out[108]:

```
'5'
```

In [109]:

```
author.pop(4)
```

Out[109]:

```
'Fifty Shades of Grey'
```

In [110]:

```
author.pop(5)
```

Out[110]:

```
'3,758,936'
```

In [111]:

```
author.pop(5)
```

Out[111]:

```
'Random House'
```

In [112]:

```
author.pop(5)
```

Out[112]:

```
'6'
```

In [113]:

```
author.pop(5)
```

Out[113]:

```
'Harry Potter and the Goblet of Fire'
```

In [114]:

```
author.pop(6)
```

Out[114]:

```
'3,583,215'
```

In [115]:

```

author.pop(6)
```

Out[115]:

```
'Bloomsbury'
```

In [116]:

```
author.pop(6)
```

Out[116]:

```
'7'
```

In [117]:

```
author.pop(6)
```

Out[117]:

```
'Harry Potter and the Chamber of Secrets'
```

In [118]:

```
author.pop(7)
```

Out[118]:

```
'3,484,047'
```

In [119]:

```
author.pop(7)
```

Out[119]:

```
'Bloomsbury'
```

In [120]:

```
author.pop(7)
```

Out[120]:

```
'8'
```

In [121]:

```
author.pop(7)
```

Out[121]:

```
'Harry Potter and the Prisoner of Azkaban'
```

In [122]:

```
author.pop(8)
```

Out[122]:

```
'3,377,906'
```

In [123]:

```
author.pop(8)
```

Out[123]:

```
'Bloomsbury'
```

In [124]:

```
author.pop(8)
```

Out[124]:

```
'9'
```

In [125]:

```
author.pop(8)
```

Out[125]:

```
'Angels and Demons'
```

In [126]:

```
author.pop(9)
```

Out[126]:

```
'3,193,946'
```

In [127]:

```
author.pop(9)
```

Out[127]:

```
'Transworld'
```

In [128]:

```
author.pop(9)
```

Out[128]:

```
'10'
```

In [129]:

```

author.pop(9)
```

Out[129]:

```
"Harry Potter and the Half-blood Prince:Children's Edition"
```

In [130]:

```
author.pop(10)
```

Out[130]:

```
'2,950,264'
```

In [131]:

```
author.pop(10)
```

Out[131]:

```
'Bloomsbury'
```

In [132]:

```
author.pop(10)
```

Out[132]:

```
'11'
```

In [133]:

```
author.pop(10)
```

Out[133]:

```
'Fifty Shades Darker'
```

In [134]:

```
author.pop(11)
```

Out[134]:

```
'2,479,784'
```

In [135]:

```
author.pop(11)
```

Out[135]:

```
'Random House'
```

In [136]:

```
author.pop(11)
```

Out[136]:

```
'12'
```

In [137]:

```
author.pop(11)
```

Out[137]:

```
'Twilight'
```

In [138]:

```
author.pop(12)
```

Out[138]:

```
'2,315,405'
```

In [139]:

```
author.pop(12)
```

Out[139]:

```
'Little, Brown Book'
```

In [140]:

```
author.pop(12)
```

Out[140]:

```
'13'
```

In [141]:

```
author.pop(12)
```

Out[141]:

```
'Girl with the Dragon Tattoo,The:Millennium Trilogy'
```

In [142]:

```
author.pop(13)
```

Out[142]:

```
'2,233,570'
```

In [143]:

```
author.pop(13)
```

Out[143]:

```
'Quercus'
```

In [144]:

```
author.pop(13)
```

Out[144]:

```
'14'
```

In [145]:

```
author.pop(13)
```

Out[145]:

```
'Fifty Shades Freed'
```

In [146]:

```
author.pop(14)
```

Out[146]:

```
'2,193,928'
```

In [147]:

```
author.pop(14)
```

Out[147]:

```
'Random House'
```

In [148]:

```
author.pop(14)
```

Out[148]:

```
'15'
```

In [149]:

```
author.pop(14)
```

Out[149]:

```
'Lost Symbol,The'
```

In [150]:

```
author.pop(15)
```

Out[150]:

```
'2,183,031'
```

In [151]:

```
author.pop(15)
```

Out[151]:

```
'Transworld'
```

In [152]:

```
author.pop(15)
```

Out[152]:

```
'16'
```

In [153]:

```
author.pop(15)
```

Out[153]:

```
'New Moon'
```

In [154]:

```
author.pop(16)
```

Out[154]:

```
'2,152,737'
```

In [155]:

```
author.pop(16)
```

Out[155]:

```
'Little, Brown Book'
```

In [156]:

```
author.pop(16)
```

Out[156]:

```
'17'
```

In [157]:

```
author.pop(16)
```

Out[157]:

```
'Deception Point'
```

In [158]:

```
author.pop(17)
```

Out[158]:

```
'2,062,145'
```

In [159]:

```
author.pop(17)
```

Out[159]:

```
'Transworld'
```

In [160]:

```
author.pop(17)
```

Out[160]:

```
'18'
```

In [161]:

```
author.pop(17)
```

Out[161]:

```
'Eclipse'
```

In [162]:

```
author.pop(18)
```

Out[162]:

```
'2,052,876'
```

In [163]:

```
author.pop(18)
```

Out[163]:

```
'Little, Brown Book'
```

In [164]:

```
author.pop(18)
```

Out[164]:

```
'19'
```

In [165]:

```
author.pop(18)
```

Out[165]:

```
'Lovely Bones,The'
```

In [166]:

```
author.pop(19)
```

Out[166]:

```
'2,005,598'
```

In [167]:

```
author.pop(19)
```

Out[167]:

```
'Pan Macmillan'
```

In [168]:

```
author.pop(19)
```

Out[168]:

```
'20'
```

In [169]:

```
author.pop(19)
```

Out[169]:

```
'Curious Incident of the Dog in the Night-time,The'
```

In [170]:

```
author.pop(20)
```

Out[170]:

```
'1,979,552'
```

In [171]:

```
author.pop(20)
```

Out[171]:

```
'Random House'
```

In [172]:

```
author.pop(20)
```

Out[172]:

```
'21'
```

In [173]:

```
author.pop(20)
```

Out[173]:

```
'Digital Fortress'
```

In [174]:

```
author[0:20]
```

Out[174]:

```
['Brown, Dan',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'James, E. L.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Brown, Dan',
 'Rowling, J.K.',
 'James, E. L.',
 'Meyer, Stephenie',
 'Larsson, Stieg',
 'James, E. L.',
 'Brown, Dan',
 'Meyer, Stephenie',
 'Brown, Dan',
 'Meyer, Stephenie',
 'Sebold, Alice',
 'Haddon, Mark']
```

In [175]:

```
author[0:20]
```

Out[175]:

```
['Brown, Dan',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'James, E. L.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Rowling, J.K.',
 'Brown, Dan',
 'Rowling, J.K.',
 'James, E. L.',
 'Meyer, Stephenie',
 'Larsson, Stieg',
 'James, E. L.',
 'Brown, Dan',
 'Meyer, Stephenie',
 'Brown, Dan',
 'Meyer, Stephenie',
 'Sebold, Alice',
 'Haddon, Mark']
```

In [176]:

```
volume=[]
```

In [177]:

```
volume_tags=browser.find_elements_by_xpath("//td[@class='left']")
for i in volume_tags:
    if i.text is None :
        volume.append("no rating")
    else:
        volume.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/3346901027.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  volume_tags=browser.find_elements_by_xpath("//td[@class='left']")
```

In [178]:

```
volume.pop(0)
```

Out[178]:

```
'1'
```

In [179]:

```
volume.pop(0)
```

Out[179]:

```
'Da Vinci Code,The'
```

In [180]:

```
volume.pop(0)
```

Out[180]:

```
'Brown, Dan'
```

In [181]:

```
volume.pop(1)
```

Out[181]:

```
'Transworld'
```

In [182]:

```
volume.pop(1)
```

Out[182]:

```
'2'
```

In [183]:

```
volume.pop(1)
```

Out[183]:

```
'Harry Potter and the Deathly Hallows'
```

In [184]:

```
volume.pop(1)
```

Out[184]:

```
'Rowling, J.K.'
```

In [185]:

```
volume.pop(2)
```

Out[185]:

```
'Bloomsbury'
```

In [186]:

```
volume.pop(2)
```

Out[186]:

```
'3'
```

In [187]:

```
volume.pop(2)
```

Out[187]:

```
"Harry Potter and the Philosopher's Stone"
```

In [188]:

```
volume.pop(2)
```

Out[188]:

```
'Rowling, J.K.'
```

In [189]:

```
volume.pop(3)
```

Out[189]:

```
'Bloomsbury'
```

In [190]:

```
volume.pop(3)
```

Out[190]:

```
'4'
```

In [191]:

```
volume.pop(3)
```

Out[191]:

```
'Harry Potter and the Order of the Phoenix'
```

In [192]:

```
volume.pop(3)
```

Out[192]:

```
'Rowling, J.K.'
```

In [193]:

```
volume.pop(4)
```

Out[193]:

```
'Bloomsbury'
```

In [194]:

```
volume.pop(4)
```

Out[194]:

```
'5'
```

In [195]:

```
volume.pop(4)
```

Out[195]:

```
'Fifty Shades of Grey'
```

In [196]:

```
volume.pop(4)
```

Out[196]:

```
'James, E. L.'
```

In [197]:

```
volume.pop(5)
```

Out[197]:

```
'Random House'
```

In [198]:

```
volume.pop(5)
```

Out[198]:

```
'6'
```

In [199]:

```
volume.pop(5)
```

Out[199]:

```
'Harry Potter and the Goblet of Fire'
```

In [200]:

```
volume.pop(5)
```

Out[200]:

```
'Rowling, J.K.'
```

In [201]:

```
volume.pop(6)
```

Out[201]:

```
'Bloomsbury'
```

In [202]:

```

volume.pop(6)
```

Out[202]:

```
'7'
```

In [203]:

```
volume.pop(6)
```

Out[203]:

```
'Harry Potter and the Chamber of Secrets'
```

In [204]:

```
volume.pop(6)
```

Out[204]:

```
'Rowling, J.K.'
```

In [205]:

```
volume.pop(7)
```

Out[205]:

```
'Bloomsbury'
```

In [206]:

```
volume.pop(7)
```

Out[206]:

```
'8'
```

In [207]:

```
volume.pop(7)
```

Out[207]:

```
'Harry Potter and the Prisoner of Azkaban'
```

In [208]:

```
volume.pop(7)
```

Out[208]:

```
'Rowling, J.K.'
```

In [209]:

```
volume.pop(8)
```

Out[209]:

```
'Bloomsbury'
```

In [210]:

```
volume.pop(8)
```

Out[210]:

```
'9'
```

In [211]:

```
volume.pop(8)
```

Out[211]:

```
'Angels and Demons'
```

In [212]:

```
volume.pop(8)
```

Out[212]:

```
'Brown, Dan'
```

In [213]:

```
volume.pop(9)
```

Out[213]:

```
'Transworld'
```

In [214]:

```
volume.pop(9)
```

Out[214]:

```
'10'
```

In [215]:

```
volume.pop(9)
```

Out[215]:

```
"Harry Potter and the Half-blood Prince:Children's Edition"
```

In [216]:

```
volume.pop(9)
```

Out[216]:

```
'Rowling, J.K.'
```

In [217]:

```
volume.pop(10)
```

Out[217]:

```
'Bloomsbury'
```

In [218]:

```
volume.pop(10)
```

Out[218]:

```
'11'
```

In [219]:

```
volume.pop(10)
```

Out[219]:

```
'Fifty Shades Darker'
```

In [220]:

```
volume.pop(10)
```

Out[220]:

```
'James, E. L.'
```

In [221]:

```
volume.pop(11)
```

Out[221]:

```
'Random House'
```

In [222]:

```
volume.pop(11)
```

Out[222]:

```
'12'
```

In [223]:

```
volume.pop(11)
```

Out[223]:

```
'Twilight'
```

In [224]:

```
volume.pop(11)
```

Out[224]:

```
'Meyer, Stephenie'
```

In [225]:

```
volume.pop(12)
```

Out[225]:

```
'Little, Brown Book'
```

In [226]:

```
volume.pop(12)
```

Out[226]:

```
'13'
```

In [227]:

```
volume.pop(12)
```

Out[227]:

```
'Girl with the Dragon Tattoo,The:Millennium Trilogy'
```

In [228]:

```
volume.pop(12)
```

Out[228]:

```
'Larsson, Stieg'
```

In [229]:

```
volume.pop(13)
```

Out[229]:

```
'Quercus'
```

In [230]:

```
volume.pop(13)
```

Out[230]:

```
'14'
```

In [231]:

```
volume.pop(13)
```

Out[231]:

```
'Fifty Shades Freed'
```

In [232]:

```
volume.pop(13)
```

Out[232]:

```
'James, E. L.'
```

In [233]:

```
volume.pop(14)
```

Out[233]:

```
'Random House'
```

In [234]:

```
volume.pop(14)
```

Out[234]:

```
'15'
```

In [235]:

```
volume.pop(14)
```

Out[235]:

```
'Lost Symbol,The'
```

In [236]:

```
volume.pop(14)
```

Out[236]:

```
'Brown, Dan'
```

In [237]:

```
volume.pop(15)
```

Out[237]:

```
'Transworld'
```

In [238]:

```
volume.pop(15)
```

Out[238]:

```
'16'
```

In [239]:

```
volume.pop(15)
```

Out[239]:

```
'New Moon'
```

In [240]:

```
volume.pop(15)
```

Out[240]:

```
'Meyer, Stephenie'
```

In [241]:

```
volume.pop(16)
```

Out[241]:

```
'Little, Brown Book'
```

In [242]:

```
volume.pop(16)
```

Out[242]:

```
'17'
```

In [243]:

```
volume.pop(16)
```

Out[243]:

```
'Deception Point'
```

In [244]:

```
volume.pop(16)
```

Out[244]:

```
'Brown, Dan'
```

In [245]:

```
volume.pop(17)
```

Out[245]:

```
'Transworld'
```

In [246]:

```
volume.pop(17)
```

Out[246]:

```
'18'
```

In [247]:

```
volume.pop(17)
```

Out[247]:

```
'Eclipse'
```

In [248]:

```
volume.pop(17)
```

Out[248]:

```
'Meyer, Stephenie'
```

In [249]:

```
volume.pop(18)
```

Out[249]:

```
'Little, Brown Book'
```

In [250]:

```
volume.pop(18)
```

Out[250]:

```
'19'
```

In [251]:

```
volume.pop(18)
```

Out[251]:

```
'Lovely Bones,The'
```

In [252]:

```
volume.pop(18)
```

Out[252]:

```
'Sebold, Alice'
```

In [253]:

```
volume.pop(19)
```

Out[253]:

```
'Pan Macmillan'
```

In [254]:

```
volume.pop(19)
```

Out[254]:

```
'20'
```

In [255]:

```
volume.pop(19)
```

Out[255]:

```
'Curious Incident of the Dog in the Night-time,The'
```

In [256]:

```
volume.pop(19)
```

Out[256]:

```
'Haddon, Mark'
```

In [257]:

```
volume.pop(20)
```

Out[257]:

```
'Random House'
```

In [258]:

```
volume.pop(20)
```

Out[258]:

```
'21'
```

In [259]:

```
volume.pop(20)
```

Out[259]:

```
'Digital Fortress'
```

In [260]:

```
volume[0:20]
```

Out[260]:

```
['5,094,805',
 '4,475,152',
 '4,200,654',
 '4,179,479',
 '3,758,936',
 '3,583,215',
 '3,484,047',
 '3,377,906',
 '3,193,946',
 '2,950,264',
 '2,479,784',
 '2,315,405',
 '2,233,570',
 '2,193,928',
 '2,183,031',
 '2,152,737',
 '2,062,145',
 '2,052,876',
 '2,005,598',
 '1,979,552']
```

In [261]:

```
publisher=[]
```

In [262]:

```
publisher_tags=browser.find_elements_by_xpath("//td[@class='left']")
for i in publisher_tags:
    if i.text is None :
        publisher.append("no rating")
    else:
        publisher.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/488239590.py:1: Deprecatio
nWarning: find_elements_by_xpath is deprecated. Please use find_elements(by=
By.XPATH, value=xpath) instead
  publisher_tags=browser.find_elements_by_xpath("//td[@class='left']")
```

In [263]:

```
publisher.pop(0)
```

Out[263]:

```
'1'
```

In [264]:

```
publisher.pop(0)
```

Out[264]:

```
'Da Vinci Code,The'
```

In [265]:

```
publisher.pop(0)
```

Out[265]:

```
'Brown, Dan'
```

In [266]:

```
publisher.pop(0)
```

Out[266]:

```
'5,094,805'
```

In [267]:

```
publisher.pop(1)
```

Out[267]:

```
'2'
```

In [268]:

```
publisher.pop(1)
```

Out[268]:

```
'Harry Potter and the Deathly Hallows'
```

In [269]:

```
publisher.pop(1)
```

Out[269]:

```
'Rowling, J.K.'
```

In [270]:

```
publisher.pop(1)
```

Out[270]:

```
'4,475,152'
```

In [271]:

```
publisher.pop(2)
```

Out[271]:

```
'3'
```

In [272]:

```
publisher.pop(2)
```

Out[272]:

```
"Harry Potter and the Philosopher's Stone"
```

In [273]:

```
publisher.pop(2)
```

Out[273]:

```
'Rowling, J.K.'
```

In [274]:

```
publisher.pop(2)
```

Out[274]:

```
'4,200,654'
```

In [275]:

```
publisher.pop(3)
```

Out[275]:

```
'4'
```

In [276]:

```
publisher.pop(3)
```

Out[276]:

```
'Harry Potter and the Order of the Phoenix'
```

In [277]:

```
publisher.pop(3)
```

Out[277]:

```
'Rowling, J.K.'
```

In [278]:

```
publisher.pop(3)
```

Out[278]:

```
'4,179,479'
```

In [279]:

```
publisher.pop(4)
```

Out[279]:

```
'5'
```

In [280]:

```
publisher.pop(4)
```

Out[280]:

```
'Fifty Shades of Grey'
```

In [281]:

```
publisher.pop(4)
```

Out[281]:

```
'James, E. L.'
```

In [282]:

```
publisher.pop(4)
```

Out[282]:

```
'3,758,936'
```

In [283]:

```
publisher.pop(5)
```

Out[283]:

```
'6'
```

In [284]:

```
publisher.pop(5)
```

Out[284]:

```
'Harry Potter and the Goblet of Fire'
```

In [285]:

```
publisher.pop(5)
```

Out[285]:

```
'Rowling, J.K.'
```

In [286]:

```
publisher.pop(5)
```

Out[286]:

```
'3,583,215'
```

In [287]:

```
publisher.pop(6)
```

Out[287]:

```
'7'
```

In [288]:

```
publisher.pop(6)
```

Out[288]:

```
'Harry Potter and the Chamber of Secrets'
```

In [289]:

```
publisher.pop(6)
```

Out[289]:

```
'Rowling, J.K.'
```

In [290]:

```
publisher.pop(6)
```

Out[290]:

```
'3,484,047'
```

In [291]:

```
publisher.pop(7)
```

Out[291]:

```
'8'
```

In [292]:

```
publisher.pop(7)
```

Out[292]:

```
'Harry Potter and the Prisoner of Azkaban'
```

In [293]:

```
publisher.pop(7)
```

Out[293]:

```
'Rowling, J.K.'
```

In [294]:

```
publisher.pop(7)
```

Out[294]:

```
'3,377,906'
```

In [295]:

```
publisher.pop(8)
```

Out[295]:

```
'9'
```

In [296]:

```
publisher.pop(8)
```

Out[296]:

```
'Angels and Demons'
```

In [297]:

```
publisher.pop(8)
```

Out[297]:

```
'Brown, Dan'
```

In [298]:

```
publisher.pop(8)
```

Out[298]:

```
'3,193,946'
```

In [299]:

```
publisher.pop(9)
```

Out[299]:

```
'10'
```

In [300]:

```
publisher.pop(9)
```

Out[300]:

```
"Harry Potter and the Half-blood Prince:Children's Edition"
```

In [301]:

```
publisher.pop(9)
```

Out[301]:

```
'Rowling, J.K.'
```

In [302]:

```
publisher.pop(9)
```

Out[302]:

```
'2,950,264'
```

In [303]:

```
publisher.pop(10)
```

Out[303]:

```
'11'
```

In [304]:

```
publisher.pop(10)
```

Out[304]:

```
'Fifty Shades Darker'
```

In [305]:

```
publisher.pop(10)
```

Out[305]:

```
'James, E. L.'
```

In [306]:

```
publisher.pop(10)
```

Out[306]:

```
'2,479,784'
```

In [307]:

```
publisher.pop(11)
```

Out[307]:

```
'12'
```

In [308]:

```
publisher.pop(11)
```

Out[308]:

```
'Twilight'
```

In [309]:

```
publisher.pop(11)
```

Out[309]:

```
'Meyer, Stephenie'
```

In [310]:

```
publisher.pop(11)
```

Out[310]:

```
'2,315,405'
```

In [311]:

```
publisher.pop(12)
```

Out[311]:

```
'13'
```

In [312]:

```
publisher.pop(12)
```

Out[312]:

```
'Girl with the Dragon Tattoo,The:Millennium Trilogy'
```

In [313]:

```
publisher.pop(12)
```

Out[313]:

```
'Larsson, Stieg'
```

In [314]:

```
publisher.pop(12)
```

Out[314]:

```
'2,233,570'
```

In [315]:

```
publisher.pop(13)
```

Out[315]:

```
'14'
```

In [316]:

```
publisher.pop(13)
```

Out[316]:

```
'Fifty Shades Freed'
```

In [317]:

```
publisher.pop(13)
```

Out[317]:

```
'James, E. L.'
```

In [318]:

```
publisher.pop(13)
```

Out[318]:

```
'2,193,928'
```

In [319]:

```
publisher.pop(14)
```

Out[319]:

```
'15'
```

In [320]:

```
publisher.pop(14)
```

Out[320]:

```
'Lost Symbol,The'
```

In [321]:

```
publisher.pop(14)
```

Out[321]:

```
'Brown, Dan'
```

In [322]:

```
publisher.pop(14)
```

Out[322]:

```
'2,183,031'
```

In [323]:

```
publisher.pop(15)
```

Out[323]:

```
'16'
```

In [324]:

```
publisher.pop(15)
```

Out[324]:

```
'New Moon'
```

In [325]:

```
publisher.pop(15)
```

Out[325]:

```
'Meyer, Stephenie'
```

In [326]:

```
publisher.pop(15)
```

Out[326]:

```
'2,152,737'
```

In [327]:

```
publisher.pop(16)
```

Out[327]:

```
'17'
```

In [328]:

```
publisher.pop(16)
```

Out[328]:

```
'Deception Point'
```

In [329]:

```
publisher.pop(16)
```

Out[329]:

```
'Brown, Dan'
```

In [330]:

```
publisher.pop(16)
```

Out[330]:

```
'2,062,145'
```

In [331]:

```
publisher.pop(17)
```

Out[331]:

```
'18'
```

In [332]:

```
publisher.pop(17)
```

Out[332]:

```
'Eclipse'
```

In [333]:

```
publisher.pop(17)
```

Out[333]:

```
'Meyer, Stephenie'
```

In [334]:

```
publisher.pop(17)
```

Out[334]:

```
'2,052,876'
```

In [335]:

```
publisher.pop(18)
```

Out[335]:

```
'19'
```

In [336]:

```
publisher.pop(18)
```

Out[336]:

```
'Lovely Bones,The'
```

In [337]:

```
publisher.pop(18)
```

Out[337]:

```
'Sebold, Alice'
```

In [338]:

```
publisher.pop(18)
```

Out[338]:

```
'2,005,598'
```

In [339]:

```
publisher.pop(19)
```

Out[339]:

```
'20'
```

In [340]:

```
publisher.pop(19)
```

Out[340]:

```
'Curious Incident of the Dog in the Night-time,The'
```

In [341]:

```
publisher.pop(19)
```

Out[341]:

```
'Haddon, Mark'
```

In [342]:

```
publisher.pop(19)
```

Out[342]:

```
'1,979,552'
```

In [343]:

```
publisher.pop(20)
```

Out[343]:

```
'21'
```

In [344]:

```
publisher.pop(20)
```

Out[344]:

```
'Digital Fortress'
```

In [345]:

```
publisher.pop(20)
```

Out[345]:

```
'Brown, Dan'
```

In [346]:

```
publisher.pop(20)
```

Out[346]:

```
'1,928,900'
```

In [347]:

```
publisher.pop(21)
```

Out[347]:

```
'22'
```

In [348]:

```
publisher.pop(21)
```

Out[348]:

```
'Short History of Nearly Everything,A'
```

In [349]:

```
publisher.pop(21)
```

Out[349]:

```
'Bryson, Bill'
```

In [350]:

```
publisher.pop(21)
```

Out[350]:

```
'1,852,919'
```

In [351]:

```
publisher[0:20]
```

Out[351]:

```
['Transworld',
 'Bloomsbury',
 'Bloomsbury',
 'Bloomsbury',
 'Random House',
 'Bloomsbury',
 'Bloomsbury',
 'Bloomsbury',
 'Transworld',
 'Bloomsbury',
 'Random House',
 'Little, Brown Book',
 'Quercus',
 'Random House',
 'Transworld',
 'Little, Brown Book',
 'Transworld',
 'Little, Brown Book',
 'Pan Macmillan',
 'Random House']
```

In [361]:

```
genre=[]
```

In [362]:

```
<td id="table-cell-10943-0-0" class="left" style="color: #000000"> 1 </td>
```

```
  File "C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/2293968143.py", li
ne 1
    <td id="table-cell-10943-0-0" class="left" style="color: #000000"> 1 </t
d>
    ^
SyntaxError: invalid syntax
```

In [363]:

```
genre_tags=browser.find_elements_by_xpath("//td[@id='table-cell-10943-20-5']")
for i in genre_tags:
    if i.text is None :
        genre.append("no rating")
    else:
        genre.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/3187598671.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  genre_tags=browser.find_elements_by_xpath("//td[@id='table-cell-10943-20-
5']")
```

In [364]:

```
genre
```

Out[364]:

```
['Crime, Thriller & Adventure']
```

In [365]:

```
genre[0:20]
```

Out[365]:

```
['Crime, Thriller & Adventure']
```

In [367]:

```
import pandas as pd
guardian=pd.DataFrame({})
guardian['author']=author[0:20]
guardian['book']=books[0:20]
guardian['volume']=volume[0:20]
guardian['publisher']=publisher[0:20]
guardian['genre']=genre[0:20]
guardian
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_19328/1166435369.py in <module>
      5 guardian['volume']=volume[0:20]
      6 guardian['publisher']=publisher[0:20]
----> 7 guardian['genre']=genre[0:20]
      8 guardian

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in __setitem
__(self, key, value)
   3610         else:
   3611             # set column
-> 3612             self._set_item(key, value)
   3613
   3614     def _setitem_slice(self, key: slice, value):

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _set_item
(self, key, value)
   3782         ensure homogeneity.
   3783         """
-> 3784         value = self._sanitize_column(value)
   3785
   3786         if (

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _sanitize
_column(self, value)
   4507
   4508         if is_list_like(value):
-> 4509             com.require_length_match(value, self.index)
   4510         return sanitize_array(value, self.index, copy=True, allow_2d
=True)
   4511

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\common.py in require_
length_match(data, index)
    529     """
    530     if len(data) != len(index):
--> 531         raise ValueError(
    532             "Length of values "
    533             f"({len(data)}) "

ValueError: Length of values (1) does not match length of index (20)
```

```
#

9.  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 [368]:

```
import pandas as pd
import selenium
from selenium import webdriver
from selenium.common.exceptions import StaleElementReferenceException, NoSuchElementExcepti
```

In [369]:

```
import requests
from selenium import webdriver

browser = webdriver.Chrome()

browser.get("https://www.imdb.com/list/ls095964455/")
from bs4 import BeautifulSoup
```

In [370]:

```
names=[]
```

In [371]:

```
name_tags=browser.find_elements_by_xpath("//a[@href='/title/tt2193021/?ref_=ttls_li_tt']")
for i in name_tags:
    if i.text is None :
        names.append("no rating")
    else:
        names.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/597610653.py:1: Deprecatio
nWarning: find_elements_by_xpath is deprecated. Please use find_elements(by=
By.XPATH, value=xpath) instead
  name_tags=browser.find_elements_by_xpath("//a[@href='/title/tt2193021/?ref
_=ttls_li_tt']")
```

In [372]:

```
names
```

Out[372]:

```
['Arrow']
```

In [373]:

```
year=[]
```

In [374]:

```
year_tags=browser.find_elements_by_xpath("//span[@class='lister-item-year text-muted unbold
for i in year_tags:
    if i.text is None :
        year.append("no rating")
    else:
        year.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/4228166199.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  year_tags=browser.find_elements_by_xpath("//span[@class='lister-item-year
text-muted unbold']")
```

In [375]:

```
year[0:10]
```

Out[375]:

```
['(2011–2019)',
 '(2016– )',
 '(2010–2022)',
 '(2017–2020)',
 '(2014–2020)',
 '(2013–2019)',
 '(2017–2022)',
 '(2005– )',
 '(2014– )',
 '(2012–2020)']
```

In [376]:

```
genre=[]
```

In [377]:

```
genre_tags=browser.find_elements_by_xpath("//span[@class='genre']")
for i in genre_tags:
    if i.text is None :
        genre.append("no rating")
    else:
        genre.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/4197743509.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  genre_tags=browser.find_elements_by_xpath("//span[@class='genre']")
```

In [378]:

```
genre[0:10]
```

Out[378]:

```
['Action, Adventure, Drama',
 'Drama, Fantasy, Horror',
 'Drama, Horror, Thriller',
 'Drama, Mystery, Thriller',
 'Drama, Mystery, Sci-Fi',
 'Comedy, Crime, Drama',
 'Crime, Drama, Mystery',
 'Drama, Romance',
 'Action, Adventure, Drama',
 'Action, Adventure, Crime']
```

In [379]:

```
runtime=[]
```

In [380]:

```
runtime_tags=browser.find_elements_by_xpath("//span[@class='runtime']")
for i in runtime_tags:
    if i.text is None :
        runtime.append("no rating")
    else:
        runtime.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/726389388.py:1: Deprecatio
nWarning: find_elements_by_xpath is deprecated. Please use find_elements(by=
By.XPATH, value=xpath) instead
  runtime_tags=browser.find_elements_by_xpath("//span[@class='runtime']")
```

In [381]:

```
runtime[0:10]
```

Out[381]:

```
['57 min',
 '51 min',
 '44 min',
 '60 min',
 '43 min',
 '59 min',
 '45 min',
 '41 min',
 '43 min',
 '42 min']
```

In [382]:

```
rating=[]
```

In [383]:

```
rating_tags=browser.find_elements_by_xpath("//span[@class='ipl-rating-star__rating']")
for i in rating_tags:
    if i.text is None :
        rating.append("no rating")
    else:
        rating.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/2217403016.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  rating_tags=browser.find_elements_by_xpath("//span[@class='ipl-rating-star
__rating']")
```

In [384]:

```
rating
```

Out[384]:

```
['9.2',
 '',
 'Rate',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
 '',
```

In [564]:

```
rating.pop(1)
```

Out[564]:

```
'8.2'
```

In [565]:

```
rating.pop(1)
```

Out[565]:

```
''
```

In [566]:

```
rating.pop(1)
```

Out[566]:

''

In [567]:

```
rating.pop(1)
```

Out[567]:

''

In [568]:

```
rating.pop(1)
```

Out[568]:

''

In [569]:

```
rating.pop(1)
```

Out[569]:

''

In [570]:

```
rating.pop(1)
```

Out[570]:

''

In [571]:

```
rating.pop(1)
```

Out[571]:

''

In [572]:

```
rating.pop(1)
```

Out[572]:

''

In [573]:

```
rating.pop(1)
```

Out[573]:

```
''
```

In [574]:

```
rating.pop(1)
```

Out[574]:

```
''
```

In [575]:

```
rating.pop(1)
```

Out[575]:

```
''
```

In [576]:

```
rating.pop(1)
```

Out[576]:

```
''
```

In [577]:

```
rating.pop(1)
```

Out[577]:

```
''
```

In [578]:

```
rating.pop(1)
```

Out[578]:

```
'7.6'
```

In [579]:

```
rating.pop(1)
```

Out[579]:

```
''
```

In [580]:

```
rating.pop(1)
```

Out[580]:

```
'Rate'
```

In [581]:

```
rating.pop(1)
```

Out[581]:

```
''
```

In [582]:

```
rating.pop(1)
```

Out[582]:

```
''
```

In [583]:

```
rating.pop(1)
```

Out[583]:

```
''
```

In [584]:

```
rating.pop(1)
```

Out[584]:

```
''
```

In [585]:

```
rating.pop(1)
```

Out[585]:

```
''
```

In [586]:

```
rating.pop(2)
```

Out[586]:

```
''
```

In [587]:

```
rating.pop(2)
```

Out[587]:

''

In [588]:

```
rating.pop(2)
```

Out[588]:

''

In [589]:

```
rating.pop(2)
```

Out[589]:

''

In [590]:

```
rating.pop(2)
```

Out[590]:

''

In [591]:

```
rating.pop(2)
```

Out[591]:

''

In [592]:

```
rating.pop(2)
```

Out[592]:

''

In [593]:

```
rating.pop(2)
```

Out[593]:

''

In [594]:

```
rating.pop(2)
```

Out[594]:

''

In [595]:

```
rating.pop(2)
```

Out[595]:

''

In [596]:

```
rating.pop(2)
```

Out[596]:

''

In [597]:

```
rating.pop(2)
```

Out[597]:

''

In [598]:

```
rating.pop(2)
```

Out[598]:

''

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(2)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3)
```

In [ ]:

```
rating.pop(3).
```

In [ ]:

```
rating.pop(3)
```

In [406]:

```
rating.pop(4)
```

Out[406]:

```
''
```

In [407]:

```
rating.pop(4)
```

Out[407]:

```
'Rate'
```

In [408]:

```
rating.pop(4)
```

Out[408]:

```
''
```

In [409]:

```
rating.pop(4)
```

Out[409]:

```
''
```

In [410]:

```
rating.pop(4)
```

Out[410]:

```
''
```

In [411]:

```
rating.pop(4)
```

Out[411]:

```
''
```

In [412]:

```
rating.pop(4)
```

Out[412]:

```
''
```

In [413]:

```
rating.pop(4)
```

Out[413]:

''

In [414]:

```
rating.pop(4)
```

Out[414]:

''

In [415]:

```
rating.pop(4)
```

Out[415]:

''

In [416]:

```
rating.pop(4)
```

Out[416]:

''

In [417]:

```
rating.pop(4)
```

Out[417]:

''

In [418]:

```
rating.pop(4)
```

Out[418]:

''

In [419]:

```
rating.pop(4)
```

Out[419]:

''

In [420]:

```
rating.pop(4)
```

Out[420]:

''

In [421]:

```
rating.pop(4)
```

Out[421]:

''

In [422]:

```
rating.pop(4)
```

Out[422]:

''

In [423]:

```
rating.pop(4)
```

Out[423]:

''

In [424]:

```
rating.pop(4)
```

Out[424]:

''

In [425]:

```
rating.pop(4)
```

Out[425]:

''

In [426]:

```
rating.pop(4)
```

Out[426]:

''

In [427]:

```
rating.pop(4)
```

Out[427]:

```
''
```

In [428]:

```
rating.pop(5)
```

Out[428]:

```
''
```

In [429]:

```
rating.pop(5)
```

Out[429]:

```
'Rate'
```

In [430]:

```
rating.pop(5)
```

Out[430]:

```
''
```

In [431]:

```
rating.pop(5)
```

Out[431]:

```
''
```

In [432]:

```
rating.pop(5)
```

Out[432]:

```
''
```

In [433]:

```
rating.pop(5)
```

Out[433]:

```
''
```

In [434]:

```
rating.pop(5)
```

Out[434]:

''

In [435]:

```
rating.pop(5)
```

Out[435]:

''

In [436]:

```
rating.pop(5)
```

Out[436]:

''

In [437]:

```
rating.pop(5)
```

Out[437]:

''

In [438]:

```
rating.pop(5)
```

Out[438]:

''

In [439]:

```
rating.pop(5)
```

Out[439]:

''

In [440]:

```
rating.pop(5)
```

Out[440]:

''

In [441]:

```
rating.pop(5)
```

Out[441]:

''

In [442]:

```
rating.pop(5)
```

Out[442]:

''

In [443]:

```
rating.pop(5)
```

Out[443]:

''

In [444]:

```
rating.pop(5)
```

Out[444]:

''

In [445]:

```
rating.pop(5)
```

Out[445]:

''

In [446]:

```
rating.pop(5)
```

Out[446]:

''

In [447]:

```
rating.pop(5)
```

Out[447]:

''

In [448]:

```
rating.pop(5)
```

Out[448]:

```
''
```

In [449]:

```
rating.pop(5)
```

Out[449]:

```
''
```

In [450]:

```
rating.pop(6)
```

Out[450]:

```
''
```

In [451]:

```
rating.pop(6)
```

Out[451]:

```
'Rate'
```

In [452]:

```
rating.pop(6)
```

Out[452]:

```
''
```

In [453]:

```
rating.pop(6)
```

Out[453]:

```
''
```

In [454]:

```
rating.pop(6)
```

Out[454]:

```
''
```

In [455]:

```
rating.pop(6)
```

Out[455]:

''

In [456]:

```
rating.pop(6)
```

Out[456]:

''

In [457]:

```
rating.pop(6)
```

Out[457]:

''

In [458]:

```
rating.pop(6)
```

Out[458]:

''

In [459]:

```
rating.pop(6)
```

Out[459]:

''

In [460]:

```
rating.pop(6)
```

Out[460]:

''

In [461]:

```
rating.pop(6)
```

Out[461]:

''

In [462]:

```
rating.pop(6)
```

Out[462]:

''

In [463]:

```
rating.pop(6)
```

Out[463]:

''

In [464]:

```
rating.pop(6)
```

Out[464]:

''

In [465]:

```
rating.pop(6)
```

Out[465]:

''

In [466]:

```
rating.pop(6)
```

Out[466]:

''

In [467]:

```
rating.pop(6)
```

Out[467]:

''

In [468]:

```
rating.pop(6)
```

Out[468]:

''

In [469]:

```
rating.pop(6)
```

Out[469]:

```
''
```

In [470]:

```
rating.pop(6)
```

Out[470]:

```
''
```

In [471]:

```
rating.pop(6)
```

Out[471]:

```
''
```

In [472]:

```
rating.pop(7)
```

Out[472]:

```
''
```

In [473]:

```
rating.pop(7)
```

Out[473]:

```
'Rate'
```

In [474]:

```
rating.pop(7)
```

Out[474]:

```
''
```

In [475]:

```
rating.pop(7)
```

Out[475]:

```
''
```

In [476]:

```
rating.pop(7)
```

Out[476]:

''

In [477]:

```
rating.pop(7)
```

Out[477]:

''

In [478]:

```
rating.pop(7)
```

Out[478]:

''

In [479]:

```
rating.pop(7)
```

Out[479]:

''

In [480]:

```
rating.pop(7)
```

Out[480]:

''

In [481]:

```
rating.pop(7)
```

Out[481]:

''

In [482]:

```
rating.pop(7)
```

Out[482]:

''

In [483]:

```
rating.pop(7)
```

Out[483]:

''

In [484]:

```
rating.pop(7)
```

Out[484]:

''

In [485]:

```
rating.pop(7)
```

Out[485]:

''

In [486]:

```
rating.pop(7)
```

Out[486]:

''

In [487]:

```
rating.pop(7)
```

Out[487]:

''

In [488]:

```
rating.pop(7)
```

Out[488]:

''

In [489]:

```
rating.pop(7)
```

Out[489]:

''

In [490]:

```
rating.pop(7)
```

Out[490]:

```
''
```

In [491]:

```
rating.pop(7)
```

Out[491]:

```
''
```

In [492]:

```
rating.pop(7)
```

Out[492]:

```
''
```

In [493]:

```
rating.pop(7)
```

Out[493]:

```
''
```

In [494]:

```
rating.pop(8)
```

Out[494]:

```
''
```

In [495]:

```
rating.pop(8)
```

Out[495]:

```
'Rate'
```

In [496]:

```
rating.pop(8)
```

Out[496]:

```
''
```

In [497]:

```
rating.pop(8)
```

Out[497]:

''

In [498]:

```
rating.pop(8)
```

Out[498]:

''

In [499]:

```
rating.pop(8)
```

Out[499]:

''

In [500]:

```
rating.pop(8)
```

Out[500]:

''

In [501]:

```
rating.pop(8)
```

Out[501]:

''

In [502]:

```
rating.pop(8)
```

Out[502]:

''

In [503]:

```
rating.pop(8)
```

Out[503]:

''

In [504]:

```
rating.pop(8)
```

Out[504]:

''

In [505]:

```
rating.pop(8)
```

Out[505]:

''

In [506]:

```
rating.pop(8)
```

Out[506]:

''

In [507]:

```
rating.pop(8)
```

Out[507]:

''

In [508]:

```
rating.pop(8)
```

Out[508]:

''

In [509]:

```
rating.pop(8)
```

Out[509]:

''

In [510]:

```
rating.pop(8)
```

Out[510]:

''

In [511]:

```
rating.pop(8)
```

Out[511]:

```
''
```

In [512]:

```
rating.pop(8)
```

Out[512]:

```
''
```

In [513]:

```
rating.pop(8)
```

Out[513]:

```
''
```

In [514]:

```
rating.pop(8)
```

Out[514]:

```
''
```

In [515]:

```
rating.pop(8)
```

Out[515]:

```
''
```

In [516]:

```
rating.pop(9)
```

Out[516]:

```
''
```

In [517]:

```
rating.pop(9)
```

Out[517]:

```
'Rate'
```

In [518]:

```
rating.pop(9)
```

Out[518]:

''

In [519]:

```
rating.pop(9)
```

Out[519]:

''

In [520]:

```
rating.pop(9)
```

Out[520]:

''

In [521]:

```
rating.pop(9)
```

Out[521]:

''

In [522]:

```
rating.pop(9)
```

Out[522]:

''

In [523]:

```
rating.pop(9)
```

Out[523]:

''

In [524]:

```
rating.pop(9)
```

Out[524]:

''

In [525]:

```
rating.pop(9)
```

Out[525]:

''

In [526]:

```
rating.pop(9)
```

Out[526]:

''

In [527]:

```
rating.pop(9)
```

Out[527]:

''

In [528]:

```
rating.pop(9)
```

Out[528]:

''

In [529]:

```
rating.pop(9)
```

Out[529]:

''

In [530]:

```
rating.pop(9)
```

Out[530]:

''

In [531]:

```
rating.pop(10)
```

Out[531]:

''

In [532]:

```
rating.pop(10)
```

Out[532]:

```
''
```

In [533]:

```
rating.pop(10)
```

Out[533]:

```
''
```

In [534]:

```
rating.pop(10)
```

Out[534]:

```
''
```

In [535]:

```
rating.pop(10)
```

Out[535]:

```
''
```

In [536]:

```
rating.pop(10)
```

Out[536]:

```
''
```

In [537]:

```
rating.pop(11)
```

Out[537]:

```
''
```

In [538]:

```
rating.pop(11)
```

Out[538]:

```
'Rate'
```

In [539]:

```
rating.pop(9)
```

Out[539]:

```
''
```

In [540]:

```
rating[0:10]
```

Out[540]:

```
['9.2', '', '', '8.7', '8.2', '7.5', '7.6', '8.1', '6.6', '7.6']
```

In [604]:

```
votes=[]
```

In [601]:

```
voting_tags=browser.find_elements_by_xpath("//span[@name='nv']")
for i in voting_tags:
    if i.text is None :
        votes.append("no rating")
    else:
        votes.append(i.text)
```

```
C:\Users\hamsa\AppData\Local\Temp/ipykernel_19328/3893642565.py:1: Deprecati
onWarning: find_elements_by_xpath is deprecated. Please use find_elements(by
=By.XPATH, value=xpath) instead
  voting_tags=browser.find_elements_by_xpath("//span[@name='nv']")
```

In [602]:

```
votes[0:10]
```

Out[602]:

```
['2,006,280',
 '1,079,874',
 '953,951',
 '285,133',
 '244,857',
 '297,437',
 '140,679',
 '297,953',
 '340,277',
 '427,367']
```

In [603]:

```
import pandas as pd
idmb=pd.DataFrame({})
idmb['names']=names[0:5]
idmb['genre']=genre[0:5]
idmb['rating']=rating[0:5]
idmb['votes']=votes[0:5]
idmb['year']=year[0:5]
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_19328/1582931528.py in <module>
      2 idmb=pd.DataFrame({})
      3 idmb['names']=names[0:5]
----> 4 idmb['genre']=genre[0:5]
      5 idmb['rating']=rating[0:5]
      6 idmb['votes']=votes[0:5]

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in __setitem
__(self, key, value)
   3610         else:
   3611             # set column
-> 3612             self._set_item(key, value)
   3613
   3614     def _setitem_slice(self, key: slice, value):

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _set_item
(self, key, value)
   3782         ensure homogeneity.
   3783         """
-> 3784         value = self._sanitize_column(value)
   3785
   3786         if (

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _sanitize
_column(self, value)
   4507
   4508         if is_list_like(value):
-> 4509             com.require_length_match(value, self.index)
   4510         return sanitize_array(value, self.index, copy=True, allow_2d
=True)
   4511

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\common.py in require_
length_match(data, index)
    529     """
    530     if len(data) != len(index):
--> 531         raise ValueError(
    532             "Length of values "
    533             f"({len(data)}) "

ValueError: Length of values (5) does not match length of index (1)
```

In [606]:

```
idmb
```

Out[606]:

	names
0	Arrow

```
#

6.  Scrape the details of top 100 songs on billiboard.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 [608]:

```
from email import header
import spotipy
from spotipy.oauth2 import SpotifyOAuth
from bs4 import BeautifulSoup
import requests

url="https://www.billboard.com/charts/hot-100/"

date=input("what year you would like to travel to in YYY-MM-DD")
#*******************************************Code Dealing with Getting authenticated with
CLIENT_ID="29b4aa34873d43e7a5fb0133a098c832"
CLIENT_SECRET="63c8b9a9dbe747f8814d4143b83c56be"

scope = "playlist-modify-private"

sp = spotipy.Spotify(auth_manager=SpotifyOAuth(client_id=CLIENT_ID,client_secret=CLIENT_SEC

USER_ID=sp.current_user()['display_name']

#*******************************************Code Dealing with Getting Billiboard Data In d
response=requests.get(url=url+date)
webpage=response.text

soup=BeautifulSoup(webpage,"html.parser")
raw_songs=soup.find_all(name="h3",id="title-of-a-story",class_="c-title")

songs=[song.getText() for song in raw_songs]
del songs[0:6]
start=1
end=start+3
while (end<len(songs)):
    del songs[start:end]
    start=start+1
    end=start+3

del songs[len(songs)-4:len(songs)]

final_songs=[song.split("\n")[1] for song in songs]

#*****************************************Code Dealing with Getting url of every Song from
songs_uri=[]
for song in final_songs:
    string_to_search="track:"+song+" year:"+date[0:4]
    result = sp.search(q=string_to_search, type="track")

    try:
        uri=result["tracks"]["items"][0]["uri"]
        songs_uri.append(uri)
    except IndexError:
        print(f"{song} does not exist in Spotify")

#creating playlist
PLAYLIST_ID=sp.user_playlist_create(user=USER_ID,name=f"{date} Billiboard 100", public=Fals

#adding tracks to the playlist
sp.playlist_add_items(playlist_id=PLAYLIST_ID, items=songs_uri, position=None)
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_19328/8884534.py in <module>
      1 from email import header
----> 2 import spotipy
      3 from spotipy.oauth2 import SpotifyOAuth
      4 from bs4 import BeautifulSoup
      5 import requests

ModuleNotFoundError: No module named 'spotipy'
```

```
#

10. 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
Note: - from the home page you have to go to the ShowAllDataset page through code.
```

Notebook for crawling the UCI Machine Learning repository website + Convinient functions for downloading, summarizing or viewing information about various datasets

In [625]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re
import requests
import os
```

In [626]:

```
import urllib.request, urllib.parse, urllib.error
from bs4 import BeautifulSoup
import ssl

# Ignore SSL certificate errors
ctx = ssl.create_default_context()
ctx.check_hostname = False
ctx.verify_mode = ssl.CERT_NONE

# Read the HTML from the URL and pass on to BeautifulSoup
url = 'https://archive.ics.uci.edu/ml/datasets.html?format=&task=&att=&area=&numAtt=&numIns
print("Opening the file connection...")
uh= urllib.request.urlopen(url, context=ctx)
print("HTTP status",uh.getcode())
html =uh.read()
print(f"Reading done. Total {len(html)} characters read.")

# Soupify!
soup = BeautifulSoup(html, 'html5lib')
```

```
Opening the file connection...
---------------------------------------------------------------------------
-
HTTPError                                 Traceback (most recent call las
t)
~\AppData\Local\Temp/ipykernel_19328/30607957.py in <module>
     11 url = 'https://archive.ics.uci.edu/ml/datasets.html?format=&task=&
att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=list'
     12 print("Opening the file connection...")
---> 13 uh= urllib.request.urlopen(url, context=ctx)
     14 print("HTTP status",uh.getcode())
     15 html =uh.read()

C:\ProgramData\Anaconda3\lib\urllib\request.py in urlopen(url, data, timeo
ut, cafile, capath, cadefault, context)
    212     else:
    213         opener = _opener
--> 214     return opener.open(url, data, timeout)
    215
    216 def install_opener(opener):

C:\ProgramData\Anaconda3\lib\urllib\request.py in open(self, fullurl, dat
a, timeout)
    521         for processor in self.process_response.get(protocol, []):
    522             meth = getattr(processor, meth_name)
--> 523             response = meth(req, response)
    524
    525         return response

C:\ProgramData\Anaconda3\lib\urllib\request.py in http_response(self, requ
est, response)
    630         # request was successfully received, understood, and accep
ted.
    631         if not (200 <= code < 300):
--> 632             response = self.parent.error(
    633                 'http', request, response, code, msg, hdrs)
    634
```

```
C:\ProgramData\Anaconda3\lib\urllib\request.py in error(self, proto, *arg
s)
    559         if http_err:
    560             args = (dict, 'default', 'http_error_default') + orig_
args
--> 561             return self._call_chain(*args)
    562
    563 # XXX probably also want an abstract factory that knows when it ma
kes

C:\ProgramData\Anaconda3\lib\urllib\request.py in _call_chain(self, chain,
kind, meth_name, *args)
    492         for handler in handlers:
    493             func = getattr(handler, meth_name)
--> 494             result = func(*args)
    495             if result is not None:
    496                 return result

C:\ProgramData\Anaconda3\lib\urllib\request.py in http_error_default(self,
req, fp, code, msg, hdrs)
    639 class HTTPDefaultErrorHandler(BaseHandler):
    640     def http_error_default(self, req, fp, code, msg, hdrs):
--> 641         raise HTTPError(req.full_url, code, msg, hdrs, fp)
    642
    643 class HTTPRedirectHandler(BaseHandler):

HTTPError: HTTP Error 404: Not Found
```

In [ ]:

```
lst=[]
for tag in soup.find_all('p'):
    lst.append(tag.contents)

i=0
description_dict={}
dataset_list=[]
for l in lst:
    if len(l)>2:
        if str(l[1]).find('datasets/')!=-1:
            string=str(l[1])
            s=re.search('">.*</a>',string)
            x,y=s.span()
            description_dict[string[x+2:y-4]]=(l[2])[2:]
            s=re.search("\".*\"",string)
            x,y=s.span()
            dataset_list.append(string[x+10:y-1])
            i+=1
print(f"{i} datasets read")
```

In [ ]:

```
def dataset_page_crawl(dataset):
    dataset_dict={}
    baseurl='https://archive.ics.uci.edu/ml/datasets/'
    url = baseurl+dataset
    dataset_dict['Dataset Page']=url
    #print("Opening the page:", url)
    try:
        uh= urllib.request.urlopen(url, context=ctx)
        html =uh.read().decode()
        soup=BeautifulSoup(html,'html5lib')
        #print(soup.text[:200])
        if soup.text.find("does not appear to exist")!=-1:
            print(f"{dataset} not found")
            return None
        else:
            dataurls = []
            for pclass in soup.find_all('p'):
                if (pclass.get_text().find('Abstract: '))!=-1:
                    dataset_dict['Abstract']=str(pclass.get_text())[len('Abstract: '):]
                    break # Breaking here is crucial. Otherwise the loop will progress thro
                    # and put 'Not found' in the Abstract field
                else:
                    dataset_dict['Abstract']='Not found!'
            for link in soup.find_all('a'):
                if link.attrs['href'].find('machine-learning-databases')!=-1:
                    a=link.attrs['href']
                    a=a[2:]
                    dataurl="https://archive.ics.uci.edu/ml/"+a
                    dataurls.append(dataurl)

            # After finishing the for-loop with a-tags, the first dataurl is added to the d
            dataset_dict['dataurl']=dataurls[0]

        return dataset_dict
    except:
        print("Could not retrieve")
        return None
```

In [ ]:

```
i=0
baseurl='https://archive.ics.uci.edu/ml/datasets/'
dataset_dicts=[]
for dataset in dataset_list:
    a=dataset_page_crawl(dataset)
    if a!=None:
        dataset_dicts.append(a)
        i+=1
        print(f"Dataset {i} processed",end=', ')

print("\nTotal datasets analyzed: ",i)
```

In [627]:

```
df_dataset = pd.DataFrame(data=dataset_dicts)
```

In [628]:

```
df_description=pd.DataFrame(data=list(description_dict.items()),columns=['Dataset','Abstrac
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_19328/78833370.py in <module>
----> 1 df_description=pd.DataFrame(data=list(description_dict.items()),colu
mns=['Dataset','Abstract'])

NameError: name 'description_dict' is not defined
```

In [629]:

```
df_joined=df_description.merge(df_dataset,on='Abstract')
df_joined
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_19328/1967391846.py in <module>
----> 1 df_joined=df_description.merge(df_dataset,on='Abstract')
      2 df_joined

NameError: name 'df_description' is not defined
```

In [ ]:

In [ ]: