

```
In [1]: import re
import selenium
import pandas as pd
from selenium import webdriver
from bs4 import BeautifulSoup
import requests

import time
from selenium.common.exceptions import NoSuchElementException, StaleElementReferenceError
from selenium.webdriver.support.ui import WebDriverWait

import warnings
warnings.filterwarnings("ignore")
```

1. Scrape the details of most viewed videos on YouTube from Wikipedia. Url =
https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos
(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 [2]: driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")
```

```
In [3]: url = 'https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos'
driver.get(url)
```

```
In [4]: Rank = []
Name = []
Artist = []
Date = []
Views = []
```

```
In [5]: # scraping Rank of the videos
try:
    for i in driver.find_elements_by_xpath("//table[@class='wikitable sortable js-wikitable']//tr[1]/td"):
        Rank.append(i.text)
except NoSuchElementException:
    Rank.append("-")

# Scraping Name of the videos
try:
    for i in driver.find_elements_by_xpath("//table[@class='wikitable sortable js-wikitable']//tr[1]/td"):
        Name.append(i.text)
except NoSuchElementException:
    Name.append("-")

# Scraping Artist of the videos
try:
    for i in driver.find_elements_by_xpath("//table[@class='wikitable sortable js-wikitable']//tr[1]/td"):
        Artist.append(i.text)
except NoSuchElementException:
    Artist.append("-")

# Scraping Upload Date of the videos
try:
    for i in driver.find_elements_by_xpath("//table[@class='wikitable sortable js-wikitable']//tr[1]/td"):
        Date.append(i.text)
except NoSuchElementException:
    Date.append("-")

# Scraping Views of the videos
try:
    for i in driver.find_elements_by_xpath("//table[@class='wikitable sortable js-wikitable']//tr[1]/td"):
        Views.append(i.text)
except NoSuchElementException:
    Views.append("-")

# creating DataFrame for scraped data
Wiki = pd.DataFrame({})
Wiki['Rank'] = Rank
Wiki['Name'] = Name
Wiki['Artist'] = Artist
Wiki['Upload Date'] = Date
Wiki['Views (in Billions)'] = Views

# removing stray numbers from Name column
Wiki.Name = Wiki.Name.apply(lambda x:x[:-4].strip(''))
```

Out[5]:

	Rank	Name	Artist	Upload Date	Views (in Billions)
0	1.	Baby Shark Dance	Pinkfong Baby Shark - Kids' Songs & Stories	June 17, 2016	10.80
1	2.	Despacito	Luis Fonsi	January 12, 2017	7.87

Rank		Name	Artist	Upload Date	Views (in Billions)
2	3.	Johny Johny Yes Papa	LooLoo Kids	October 8, 2016	6.37
3	4.	Shape of You	Ed Sheeran	January 30, 2017	5.74
4	5.	See You Again	Wiz Khalifa	April 6, 2015	5.54
5	6.	Bath Song	Cocomelon – Nursery Rhymes	May 2, 2018	5.44
6	7.	Phonics Song with Two Words	ChuChu TV	March 6, 2014	4.67
7	8.	Uptown Funk	Mark Ronson	November 19, 2014	4.60
8	9.	Learning Colors – Colorful Eggs on a Farm	Miroshka TV	February 27, 2018	4.58
9	10.	Masha and the Bear – Recipe for Disaster	Get Movies	January 31, 2012	4.50
10	11.	Gangnam Style	Psy	July 15, 2012	4.45
11	12.	Wheels on the Bus	Cocomelon – Nursery Rhymes	May 24, 2018	4.12
12	13.	Dame Tu Cosita	EI Chombo	April 5, 2018	3.95
13	14.	Sugar	Maroon 5	January 14, 2015	3.71
14	15.	Roar	Katy Perry	September 5, 2013	3.60
15	16.	Counting Stars	OneRepublic	May 31, 2013	3.59
16	17.	Sorry	Justin Bieber	October 22, 2015	3.55
17	18.	Thinking Out Loud	Ed Sheeran	October 7, 2014	3.46
18	19.	Axel F	Crazy Frog	June 16, 2009	3.38
19	20.	Girls Like You	Maroon 5	May 31, 2018	3.30
20	21.	Faded	Alan Walker	December 3, 2015	3.29
21	22.	Dark Horse	Katy Perry	February 20, 2014	3.29
22	23.	Baa Baa Black Sheep	Cocomelon – Nursery Rhymes	June 25, 2018	3.26
23	24.	Let Her Go	Passenger	July 25, 2012	3.24
24	25.	Bailando	Enrique Iglesias	April 11, 2014	3.22
25	26.	Lean On	Major Lazer	March 22, 2015	3.22
26	27.	Perfect	Ed Sheeran	November 9, 2017	3.18
27	28.	Shake It Off	Taylor Swift	August 18, 2014	3.18

	Rank	Name	Artist	Upload Date	Views (in Billions)
28	29.	Waka Waka (This Time for Africa)	Shakira	June 4, 2010	3.16
29	30.	Mi Gente	J Balvin	June 29, 2017	3.09

In [6]:

```
print(len(Rank),
len(Name),
len(Artist),
len(Date),
len(Views))
```

30 30 30 30 30

In [7]:

```
driver.close()
```

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 [18]:

```
driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")
```

In [19]:

```
from bs4 import BeautifulSoup
import urllib.request
import pandas as pd
import requests

from selenium import webdriver
from selenium.webdriver.chrome.options import Options
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.action_chains import ActionChains

import time

def india_fixtures(url):

    opts = Options()
    driver = webdriver.Chrome(ChromeDriverManager().install(),chrome_options=opts)

    opts.add_argument("user-agent=Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4453.102 Safari/537.36")

    driver.get(url)

    tab_hover = driver.find_element_by_xpath("//div[@class='navigation__dropdown-item']")

    ActionChains(driver).move_to_element(tab_hover).perform()

    driver.find_element_by_xpath("//a[@class='navigation__link navigation__link--active']")

    time.sleep(2)

    title_l,series_l,place_l,date_l,time_l = [],[],[],[],[]

    for date in driver.find_elements_by_xpath("//div[@class='fixture__datetime']"):
        date_l.append(' '.join(date.text.replace('\n',' ').split(' ',3)[-1]))
    for title in driver.find_elements_by_xpath("//strong[@class='fixture__name fixture__name--bold']"):
        title_l.append(title.text)
    for series in driver.find_elements_by_xpath("//div[@class='fixture__teams']"):
        series_l.append(series.text.replace('\n',' '))
    for place in driver.find_elements_by_xpath("//p[@class='fixture__additional-information']"):
        place_l.append(place.text)

    df = pd.DataFrame(list(zip(title_l,series_l,place_l,date_l,time_l)),columns=['Title','Series','Place','Date','Time'])

    return df

india_fixtures('https://www.bcci.tv/')
```

```
ModuleNotFoundError
```

```
Traceback (most recent call last)
```

```
~\AppData\Local\Temp\ipykernel_8852\3466755244.py in <module>
```

```
6 from selenium import webdriver
```

```
7 from selenium.webdriver.chrome.options import Options
```

```
----> 8 from webdriver_manager.chrome import ChromeDriverManager  
      9 from selenium.webdriver.common.action_chains import ActionChains  
     10
```

ModuleNotFoundError: No module named 'webdriver_manager'

```
In [39]: for i in driver.find_elements_by_xpath("//div[@class='fixture__format-strip']/span"):
    Match_Title.append(i.text)

for i in driver.find_elements_by_xpath("//div[@class='fixture__format-strip']/span"):
    Series.append(i.text)

for i in driver.find_elements_by_xpath("//div[@class='fixture__description u-unstyleable']"):
    Place.append(i.text)

for i in driver.find_elements_by_xpath("//span[@class='fixture__datetime tablet-dotlist-item']"):
    Date.append(i.text.replace('\n', ' '))

date=[i.split(' ',3)[:3] for i in Date]
date=[' '.join(i) for i in date]
Time=[i.split(' ',3)[-1] for i in Date]

# creating data frame
fixture=pd.DataFrame({'Match Title': Match_Title,
                      "Series": Series,
                      "Place": Place,
                      "Date": date,
                      "Time": Time})
fixture
```

```
-----  

ValueError                                     Traceback (most recent call last)  

~\AppData\Local\Temp\ipykernel_11544\411839399.py in <module>  

16  

17 # creating data frame  

--> 18 fixture=pd.DataFrame({'Match Title': Match_Title,  

19                      "Series": Series,  

20                      "Place": Place,  

  

~\anaconda3\lib\site-packages\pandas\core\frame.py in __init__(self, data, index, columns, dtype, copy)  

612             elif isinstance(data, dict):  

613                 # GH#38939 de facto copy defaults to False only in non-di  

ct cases  

--> 614                 mgr = dict_to_mgr(data, index, columns, dtype=dtype, copy  

=copy, typ=manager)  

615                 elif isinstance(data, ma.MaskedArray):  

616                     import numpy.ma.mrecords as mrecords  

  

~\anaconda3\lib\site-packages\pandas\core\internals\construction.py in dict_t  

o_mngr(data, index, columns, dtype, typ, copy)  

462             # TODO: can we get rid of the dt64tz special case above?  

463  

--> 464         return arrays_to_mngr(  

465                 arrays, data_names, index, columns, dtype=dtype, typ=typ, con  

solidate=copy  

466             )  

  

~\anaconda3\lib\site-packages\pandas\core\internals\construction.py in arrays  

_to_mngr(arrays, arr_names, index, columns, dtype, verify_integrity, typ, cons  

olidate)  

117             # figure out the index, if necessary
```

```
118     if index is None:
--> 119         index = _extract_index(arrays)
120     else:
121         index = ensure_index(index)

~\anaconda3\lib\site-packages\pandas\core\internals\construction.py in _extra
ct_index(data)
633     lengths = list(set(raw_lengths))
634     if len(lengths) > 1:
--> 635         raise ValueError("All arrays must be of the same leng
th")
636
637     if have_dicts:

ValueError: All arrays must be of the same length
```

In [40]: `driver.close()`

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 [41]: `driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")`

In [10]: `url = ("https://www.guru99.com/")`
`driver.get(url)`

In [11]: `Name = []`
`Description = []`

```
In [12]: driver.find_element_by_xpath("//li//a[@title='Selenium']").click()

# clicking on Exception Handling button
driver.find_element_by_xpath('//a[@title="Selenium Exception Handling (Common Exceptions)"]')

# scraping Name
for i in driver.find_elements_by_xpath("//table[@class='table table-striped']/tbody/tr/td[1]"):
    Name.append(i.text)

# scraping Description
for i in driver.find_elements_by_xpath("//table[@class='table table-striped']/tbody/tr/td[2]"):
    Description.append(i.text)

# creating the dataframe from the scraped data
Selenium = pd.DataFrame({})
Selenium['Exception_Name'] = Name
Selenium['Description'] = Description
Selenium
```

Out[12]:

	Exception_Name	Description
0	ElementNotVisibleException	This type of Selenium exception occurs when an element is not visible.
1	ElementNotSelectableException	This Selenium exception occurs when an element is not selectable.
2	NoSuchElementException	This Exception occurs if an element could not be located by the provided locator.
3	NoSuchFrameException	This Exception occurs if the frame target to be switched to does not exist.
4	NoAlertPresentException	This Exception occurs when you switch to no present alert.
5	NoSuchWindowException	This Exception occurs if the window target to be switched to does not exist.
6	StaleElementReferenceException	This Selenium exception occurs happens when the element reference becomes stale.
7	SessionNotFoundException	The WebDriver is acting after you quit the browser.
8	TimeoutException	Thrown when there is not enough time for a command to complete.
9	WebDriverException	This Exception takes place when the WebDriver fails to execute some command.
10	ConnectionClosedException	This type of Exception takes place when there is a connection closed.
11	ElementClickInterceptedException	The command may not be completed as the element is being interacted with.
12	ElementNotInteractableException	This Selenium exception is thrown when any element is not interactable.
13	ErrorInResponseException	This happens while interacting with the Firefox browser.
14	ErrorHandler.UnknownServerException	Exception is used as a placeholder in case if an unknown server exception occurs.
15	ImeActivationFailedException	This expectation will occur when IME engine activation fails.
16	ImeNotAvailableException	It takes place when IME support is unavailable.
17	InsecureCertificateException	Navigation made the user agent to hit a certificate error.
18	InvalidArgumentException	It occurs when an argument does not belong to the expected type.
19	InvalidCookieDomainException	This happens when you try to add a cookie under an invalid domain.
20	InvalidCoordinatesException	This type of Exception matches an interacting with coordinates.
21	InvalidElementStateException	It occurs when command can't be finished when the element is not found.

	Exception_Name	Description
22	InvalidSessionIdException	This Exception took place when the given sessi...
23	InvalidSwitchToTargetException	This occurs when the frame or window target to...
24	JavascriptException	This issue occurs while executing JavaScript g...
25	JsonException	It occurs when you afford to get the session w...
26	NoSuchAttributeException	This kind of Exception occurs when the attribu...
27	MoveTargetOutOfBoundsException	It takes place if the target provided to the A...
28	NoSuchContextException	ContextAware does mobile device testing.
29	NoSuchCookieException	This Exception occurs when no cookie matching ...
30	NotFoundException	This Exception is a subclass of WebDriverExcep...
31	RemoteDriverServerException	This Selenium exception is thrown when the ser...
32	ScreenshotException	It is not possible to capture a screen.
33	SessionNotCreatedException	It happens when a new session could not be suc...
34	UnableToSetCookieException	This occurs if a driver is unable to set a coo...
35	UnexpectedTagNameException	Happens if a support class did not get a web e...
36	UnhandledAlertException	This expectation occurs when there is an alert...
37	UnexpectedAlertPresentException	It occurs when there is the appearance of an u...
38	UnknownMethodException	This Exception happens when the requested comm...
39	UnreachableBrowserException	This Exception occurs only when the browser is...
40	UnsupportedCommandException	This occurs when remote WebDriver does n't sen...

In [13]: `driver.close()`

4. 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)
- 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 [20]: `driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")`

In [21]: `url = ("https://statisticstimes.com/")`
`driver.get(url)`

```
In [22]: driver.find_element_by_xpath("//div[@class='navbar']/div[2]/button").click()

# clicking on India
driver.find_element_by_xpath("//div[@class='dropdown-content']/a[3]").click()
time.sleep(3)

# clicking on GDP of Indian Economy
GDP = driver.find_element_by_xpath("//html/body/div[2]/div[2]/div[2]/ul/li[1]/a")
time.sleep(3)
```

```
NoSuchElementException                                     Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_8852\1002729594.py in <module>
      6
      7 # clicking on GDP of Indian Economy
----> 8 GDP = driver.find_element_by_xpath("//html/body/div[2]/div[2]/div[2]/ul/
      li[1]/a").click()
      9 time.sleep(3)

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_el
ement_by_xpath(self, xpath)
    519         stacklevel=2,
    520     )
--> 521         return self.find_element(by=By.XPATH, value=xpath)
    522
    523     def find_elements_by_xpath(self, xpath) -> List[WebElement]:
```

```
~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_el
ement(self, by, value)
    1246         value = '[name="%s"]' % value
    1247
-> 1248         return self.execute(Command.FIND_ELEMENT, {
    1249             'using': by,
    1250             'value': value})['value']

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute
(self, driver_command, params)
    423         response = self.command_executor.execute(driver_command, params
)
    424         if response:
--> 425             self.error_handler.check_response(response)
    426             response['value'] = self._unwrap_value(
    427                 response.get('value', None))

~\anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in chec
k_response(self, response)
    245             alert_text = value['alert'].get('text')
    246             raise exception_class(message, screen, stacktrace, alert_te
xt) # 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, defau
lt: _VT) -> _VT:
```

NoSuchElementException: Message: no such element: Unable to locate element: {"m
ethod":"xpath","selector":"/html/body/div[2]/div[2]/div[2]/ul/li[1]/a"}
(Session info: MicrosoftEdge=102.0.1245.44)

Stacktrace:

Backtrace:

```
Microsoft::Applications::Events::EventProperties::unpack [0x00007FF73B7
ECBB2+24658]
    Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7
3B72D2A2+539938]
        Ordinal0 [0x00007FF73B230F05+659205]
        Ordinal0 [0x00007FF73B26C1F1+901617]
        Ordinal0 [0x00007FF73B26C403+902147]
        Ordinal0 [0x00007FF73B29E397+1106839]
        Ordinal0 [0x00007FF73B2876DF+1013471]
        Ordinal0 [0x00007FF73B29B9E7+1096167]
        Ordinal0 [0x00007FF73B2874B3+1012915]
        Ordinal0 [0x00007FF73B25E440+844864]
        Ordinal0 [0x00007FF73B25F848+849992]
    Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F73B4F17D8+120584]
        Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F73B4DCE1E+36174]
        Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F73B4E001C+48972]
        Microsoft::Applications::Events::EventProperty::to_string [0x00007FF73B
330716+30806]
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7
3B734571+569329]
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7
3B738FA4+588324]
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7
3B7390FD+588669]
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7
3B74288E+627470]
    BaseThreadInitThunk [0x00007FF925757034+20]
    RtlUserThreadStart [0x00007FF925942651+33]
```

In [23]:

```

Rank = []
State = []
GSDP1 = []
GSDP2 = []
Share = []
GDP_billion = []

# scraping Rank
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tr[1]/td[1]"):
        Rank.append(i.text)
except NoSuchElementException:
    Rank.append("_")

# scraping State
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tr[1]/td[2]"):
        State.append(i.text)
except NoSuchElementException:
    State.append("_")

# scraping GSDP at current price (19-20)
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tr[2]/td[1]"):
        GSDP1.append(i.text)
except NoSuchElementException:
    GSDP1.append("_")

# scraping GSDP at current price (18-19)
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tr[2]/td[2]"):
        GSDP2.append(i.text)
except NoSuchElementException:
    GSDP2.append("_")

# scraping Share (18-19)
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tr[3]/td[1]"):
        Share.append(i.text)
except NoSuchElementException:
    Share.append("_")

# scraping GDP $ billion
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tr[3]/td[2]"):
        GDP_billion.append(i.text)
except NoSuchElementException:
    GDP_billion.append("_")

# creating DataFrame from the scraped data
GDP = pd.DataFrame({})
GDP['Rank'] = Rank
GDP['State'] = State
GDP['GSDP at current price (19-20)'] = GSDP1
GDP['GSDP at current price (18-19)'] = GSDP2

```

```
GDP['Share (18-19)'] = Share
GDP['GDP($ billion)'] = GDP_billion
GDP
```

Out[23]:

Rank	State	GSDP at current price (19-20)	GSDP at current price (18-19)	Share (18-19)	GDP(\$ billion)
------	-------	-------------------------------	-------------------------------	---------------	-----------------

In [24]: driver.close()

In []:

In []:

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 []: driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")

In []: url = ("https://github.com/")
 driver.get(url)

```
# getting explore button and clicking on it
explore = driver.find_element_by_xpath("/html/body/div[1]/header/div/div[2]/nav/u

# selecting trending option
trend_url = driver.find_element_by_xpath("/html/body/div[1]/header/div/div[2]/nav/u
urls = trend_url.get_attribute("href")
driver.get(urls)
```

```
In [ ]: URLs = []
repository_title = []
Description = []
Contributors = []
Language = []
lang = []

# fetching urls for each repository
repository = driver.find_elements_by_xpath("//h1[@class='h3 lh-condensed']//a")
for i in repository:
    URLs.append(i.get_attribute("href"))

# scraping Repository title data
title = driver.find_elements_by_xpath("//h1[@class = 'h3 lh-condensed']")
for i in title:
    repository_title.append(i.text)

# scraping data from all repository page
for i in URLs:
    driver.get(i)
    time.sleep(5)

# scraping Repository Description data
try:
    desc = driver.find_element_by_xpath("//p[@class='f4 mt-3']")
    Description.append(desc.text)
except NoSuchElementException:
    Description.append('-')

# scraping Contributors Count data
try:
    contributor = driver.find_element_by_xpath("//*[contains(text(), 'Contributors')]")
    Contributors.append(contributor.text.replace('Contributors', ''))
except NoSuchElementException:
    Contributors.append('-')

# scraping Languages used data
try:
    for i in driver.find_elements_by_xpath("//ul[@class= 'list-style-none']//li"):
        lang.append(i.text)
    Language.append(lang)
except NoSuchElementException:
    Language.append('-')

# Data Framing
Github = pd.DataFrame({})
Github['Repository Title'] = repository_title
Github['Repository Description'] = Description
Github['Contributors Count'] = Contributors
Github['Language Used'] = Language
Github
```

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

```
In [ ]:
```

```
In [ ]:
```

6. 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 [ ]: driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")
```

```
In [ ]: url = ("https://www.billboard.com/")
driver.get(url)
```

```
In [28]: charts=driver.find_element_by_xpath("/html/body/div[3]/div[6]/div/div/ul/li[1]
```

```
In [32]: Song_Name = []
Artist_Name = []
Last_week_rank = []
Peak_rank = []
Weeks_on_board = []

# getting urls for top 100 songs
urls = driver.find_element_by_xpath("/html/body/div[3]/div[6]/div/div/div/ul/li[1]")
page_url = urls.get_attribute("href")
driver.get(page_url)
time.sleep(4)

# scraping data of song names
for i in driver.find_elements_by_xpath("/html/body/div[3]/main/div[2]/div[3]/div"):
    Song_Name.append(i.text)

# scraping data of artist names
for i in driver.find_elements_by_xpath("/html/body/div[3]/main/div[2]/div[3]/div"):
    Artist_Name.append(i.text)

# scraping data of last week ranks
for i in driver.find_elements_by_xpath("/html/body/div[3]/main/div[2]/div[3]/div"):
    Last_week_rank.append(i.text)

# scraping data of peak ranks
for i in driver.find_elements_by_xpath("/html/body/div[3]/main/div[2]/div[3]/div"):
    Peak_rank.append(i.text)

# scraping data of weeks on board
for i in driver.find_elements_by_xpath("/html/body/div[3]/main/div[2]/div[3]/div"):
    Weeks_on_board.append(i.text)

# creating dataframe for scraped data
billboard = pd.DataFrame({})
billboard['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[32]:

Name	Artist	Last Week Rank	Peak Rank	Weeks on board
------	--------	----------------	-----------	----------------

In []:

In []:

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 [38]: driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")
```

```
In [39]: # getting the webpage of mentioned url
url = ("https://www.naukri.com/")
driver.get(url)
time.sleep(3)
```

There is no "RECRUITERS OPTION"

In [40]: `driver.close()`

```
-----
WebDriverException                                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_8852\3193570824.py in <module>
----> 1 driver.close()

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in close(self)
    936         driver.close()
    937     """
--> 938     self.execute(Command.CLOSE)
    939
    940     def quit(self) -> None:

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)
    423         response = self.command_executor.execute(driver_command, params)
)
    424         if response:
--> 425             self.error_handler.check_response(response)
    426             response['value'] = self._unwrap_value(
    427                 response.get('value', None))

~\anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)
    245             alert_text = value['alert'].get('text')
    246             raise exception_class(message, screen, stacktrace, alert_te
xt) # 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, defau
lt: _VT) -> _VT:

WebDriverException: Message: chrome not reachable
(Session info: MicrosoftEdge=102.0.1245.44)
Stacktrace:
Backtrace:
    Microsoft::Applications::Events::EventProperties::unpack [0x00007FF73B7
ECBB2+24658]
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7
3B72D2A2+539938]
            Ordinal0 [0x00007FF73B230DBB+658875]
            Ordinal0 [0x00007FF73B220052+589906]
            Ordinal0 [0x00007FF73B210AD8+527064]
            Ordinal0 [0x00007FF73B28F301+1045249]
            Ordinal0 [0x00007FF73B287660+1013344]
            Ordinal0 [0x00007FF73B25E440+844864]
            Ordinal0 [0x00007FF73B25F848+849992]
        Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F73B4F17D8+120584]
            Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F73B4DCE1E+36174]
            Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F73B4E001C+48972]
            Microsoft::Applications::Events::EventProperty::EventProperty [0x00007FF73B
330716+30806]
```

```
Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7  
3B734571+569329]  
    Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7  
3B738FA4+588324]  
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7  
3B7390FD+588669]  
            Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF7  
3B74288E+627470]  
                BaseThreadInitThunk [0x00007FF925757034+20]  
                    RtlUserThreadStart [0x00007FF925942651+33]
```

In []:

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 [41]: `driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")`In [42]: `# Getting the webpage of mentioned url
url='https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-a'
driver.get(url)`

```
In [43]: #Creating empty list
Book_name = []
Author_name = []
Volumes_sold = []
Publisher = []
Genre = []

#scraping book names data
for i in driver.find_elements_by_xpath("//tbody//tr/td[2]"):
    Book_name.append(i.text)

#Scraping author names data
for i in driver.find_elements_by_xpath("//tbody/tr/td[3]"):
    try:
        if i.text == '0' : raise NoSuchElementException
        Author_name.append(i.text)
    except NoSuchElementException:
        Author_name.append('-')
time.sleep(1)

#Scraping data of volumes sold
for i in driver.find_elements_by_xpath("//tbody/tr/td[4]"):
    Volumes_sold.append(i.text)

#Scraping data of publisher names
for i in driver.find_elements_by_xpath("//tbody/tr/td[5]"):
    Publisher.append(i.text)

#Scraping data of genre
for i in driver.find_elements_by_xpath("//tbody/tr/td[6]"):
    Genre.append(i.text)

#Creating dataframe for scraped data
Novels=pd.DataFrame({})
Novels['Book Name'] = Book_name
Novels['Author'] = Author_name
Novels['Volume sold'] = Volumes_sold
Novels['Publisher'] = Publisher
Novels['Genre'] = Genre
Novels
```

Out[43]:

	Book Name	Author	Volume 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

	Book Name	Author	Volume sold	Publisher	Genre
4	Fifty Shades of Grey	James, E. L.	3,758,936	Random House	Romance & Sagas
...
95	Ghost, The	Harris, Robert	807,311	Random House	General & Literary Fiction
96	Happy Days with the Naked Chef	Oliver, Jamie	794,201	Penguin	Food & Drink: General
97	Hunger Games, The: Hunger Games Trilogy	Collins, Suzanne	792,187	Scholastic Ltd.	Young Adult Fiction
98	Lost Boy, The: A Foster Child's Search for the L...	Pelzer, Dave	791,507	Orion	Biography: General
99	Jamie's Ministry of Food: Anyone Can Learn to C...	Oliver, Jamie	791,095	Penguin	Food & Drink: General

100 rows × 5 columns

In [44]: `driver.close()`

In []:

9. Scrape the details most watched tv series of all time from [imdb.com](https://www.imdb.com/list/ls095964455/).

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 [45]: `driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")`

In [46]: `# Getting the webpage of mentioned url
url="https://www.imdb.com/list/ls095964455/"
driver.get(url)`

In [47]: #Creating empty lists.

```
Name = []
Year_span = []
Genre = []
Run_time = []
Ratings = []
Votes = []

#Scraping data of Names
for i in driver.find_elements_by_xpath("//h3[@class='lister-item-header']/a"):
    Name.append(i.text)

#Scraping data of Year span
for i in driver.find_elements_by_xpath("//span[@class='lister-item-year text-mute']"):
    Year_span.append(i.text)

#Scraping data of genre
for i in driver.find_elements_by_xpath("//span[@class='genre']"):
    Genre.append(i.text)

#Scraping data of Run time
for i in driver.find_elements_by_xpath("//span[@class='runtime']"):
    Run_time.append(i.text)

#Scraping data of Ratings
for i in driver.find_elements_by_xpath("//div[@class='ipl-rating-star small']//span"):
    Ratings.append(i.text)

#Scraping data of votes
for i in driver.find_elements_by_xpath("//div[@class='lister-item-content']//p[4]"):
    Votes.append(i.text)

# Creating dataframe for scraped data
TV_series=pd.DataFrame({})
TV_series['Name'] = Name
TV_series['Year Span'] = Year_span
TV_series['Genre'] = Genre
TV_series['Run Time'] = Run_time
TV_series['Ratings'] = Ratings
TV_series['Votes'] = Votes
TV_series
```

Out[47]:

	Name	Year Span	Genre	Run Time	Ratings	Votes
0	Game of Thrones	(2011–2019)	Action, Adventure, Drama	57 min	9.2	1,998,535
1	Stranger Things	(2016–)	Drama, Fantasy, Horror	51 min	8.7	1,046,879
2	The Walking Dead	(2010–2022)	Drama, Horror, Thriller	44 min	8.2	951,014
3	13 Reasons Why	(2017–2020)	Drama, Mystery, Thriller	60 min	7.5	284,236
4	The 100	(2014–2020)	Drama, Mystery, Sci-Fi	43 min	7.6	244,022

	Name	Year Span	Genre	Run Time	Ratings	Votes
...
95	Reign	(2013–2017)	Drama, Fantasy	42 min	7.4	48,942
96	A Series of Unfortunate Events	(2017–2019)	Adventure, Comedy, Drama	50 min	7.8	59,650
97	Criminal Minds	(2005–2020)	Crime, Drama, Mystery	42 min	8.1	190,916
98	Scream: The TV Series	(2015–2019)	Comedy, Crime, Drama	45 min	7.1	40,378
99	The Haunting of Hill House	(2018)	Drama, Horror, Mystery	572 min	8.6	229,433

100 rows × 6 columns

In [48]: `driver.close()`

In []:

In []:

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 cod

In [2]: `driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")`

```
-----
SessionNotCreatedException                                     Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_14472\3992560675.py in <module>
----> 1 driver=webdriver.Edge(r"C:/Python/msedgedriver.exe")

~\anaconda3\lib\site-packages\selenium\webdriver\edge\webdriver.py in __init__
(self, executable_path, port, options, service_args, capabilities, service_log_
path, service, keep_alive, verbose)
    60             service = Service(executable_path, port, service_args, serv
ice_log_path)
    61
---> 62         super(WebDriver, self).__init__(DesiredCapabilities.EDGE['brows
erName'], "ms",
    63                                         port, options,
    64                                         service_args, capabilities,

~\anaconda3\lib\site-packages\selenium\webdriver\chromium\webdriver.py in __ini
t__(self, browser_name, vendor_prefix, port, options, service_args, desired_cap
abilities, service_log_path, service, keep_alive)
    91
    92     try:
---> 93         RemoteWebDriver.__init__(
    94             self,
    95             command_executor=ChromiumRemoteConnection(

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in __init__
(self, command_executor, desired_capabilities, browser_profile, proxy, keep_al
ive, file_detector, options)
    267         self.file_detector = file_detector or LocalFileDetector()
    268         self.start_client()
---> 269         self.start_session(capabilities, browser_profile)
    270
    271     def __repr__(self):

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in start_s
ession(self, capabilities, browser_profile)
    358         parameters = {"capabilities": w3c_caps,
    359                         "desiredCapabilities": capabilities}
---> 360         response = self.execute(Command.NEW_SESSION, parameters)
    361         if 'sessionId' not in response:
    362             response = response['value']

~\anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute
(self, driver_command, params)
    423             response = self.command_executor.execute(driver_command, params
)
    424         if response:
---> 425             self.error_handler.check_response(response)
    426             response['value'] = self._unwrap_value(
    427                 response.get('value', None))

~\anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in chec
k_response(self, response)
    245                 alert_text = value['alert'].get('text')
    246             raise exception_class(message, screen, stacktrace, alert_te
```

```

xt) # 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, defau
lt: _VT) -> _VT:

SessionNotCreatedException: Message: session not created: This version of MSEdg
eDriver only supports MSEdge version 101
Current browser version is 103.0.1264.37 with binary path C:\Program Files (x8
6)\Microsoft\Edge\Application\msedge.exe
Stacktrace:
Backtrace:
    Microsoft::Applications::Events::EventProperties::unpack [0x00007FF60AE
8CBB2+24658]
        Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF6
0ADCD2A2+539938]
            Ordinal0 [0x00007FF60A8D0F05+659205]
            Ordinal0 [0x00007FF60A8FB6F+834927]
            Ordinal0 [0x00007FF60A8F7214+815636]
            Ordinal0 [0x00007FF60A8F2F3B+798523]
            Ordinal0 [0x00007FF60A92CD9D+1035677]
            Ordinal0 [0x00007FF60A9274B3+1012915]
            Ordinal0 [0x00007FF60A8FE440+844864]
            Ordinal0 [0x00007FF60A8FF848+849992]
            Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F60AB917D8+120584]
            Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F60AB7CE1E+36174]
            Microsoft::Applications::Events::EventProperty::EventProperty [0x00007F
F60AB8001C+48972]
            Microsoft::Applications::Events::EventProperty::to_string [0x00007FF60A
9D0716+30806]
            Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF6
0ADD4571+569329]
            Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF6
0ADD8FA4+588324]
            Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF6
0ADD90FD+588669]
            Microsoft::Applications::Events::time_ticks_t::time_ticks_t [0x00007FF6
0ADE288E+627470]
            BaseThreadInitThunk [0x00007FFAFA867034+20]
            RtlUserThreadStart [0x00007FFAFAAE2651+33]

```

In [3]: # Getting the webpage of mentioned url
url="https://archive.ics.uci.edu/"
driver.get(url)

In [4]: # Finding view all dataset button from the webpage
viewall_dataset = driver.find_element_by_xpath("//tbody[1]//tr/td[2]/span[2]/a")
page_url = viewall_dataset.get_attribute("href")
driver.get(page_url)
time.sleep(3)

```
In [5]: # Fetching urls for each dataset
dataset_url = driver.find_elements_by_xpath("//p[@class='normal']//b/a")

urls = []
for i in dataset_url:
    urls.append(i.get_attribute("href"))
```

```
In [ ]: #Creating empty lists
Dataset_name = []
Data_type = []
Task = []
Attribute_type = []
No_of_instances = []
No_of_attributes = []
Year = []

for i in urls:
    driver.get(i)
    time.sleep(3)

#Scraping Dataset name
try:
    dataset_name = driver.find_element_by_xpath("//span[@class='heading']")
    Dataset_name.append(dataset_name.text)
except NoSuchElementException:
    Dataset_name.append('-')
time.sleep(2)

#Scraping data type
try:
    data_type = driver.find_element_by_xpath("//table[@border='1']//tbody/tr")
    if data_type.text == "N/A": raise NoSuchElementException
    Data_type.append(data_type.text)
except NoSuchElementException:
    Data_type.append('-')
time.sleep(2)

#scraping Task
try:
    task = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[3]/td")
    if task.text == "N/A": raise NoSuchElementException
    Task.append(task.text)
except NoSuchElementException:
    Task.append('-')
time.sleep(2)

# Scraping Attribute type
try:
    attribute_type = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[4]/td")
    if attribute_type.text == "N/A": raise NoSuchElementException
    Attribute_type.append(attribute_type.text)
except NoSuchElementException:
    Attribute_type.append('-')
time.sleep(2)

# Scraping No of instances
try:
    instances = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[5]/td")
    if instances.text == "N/A": raise NoSuchElementException
    No_of_instances.append(instances.text)
except NoSuchElementException:
    No_of_instances.append('-')
time.sleep(2)
```

```
# Scraping No of attributes
try:
    attribute = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[1]")
    if attribute.text == "N/A": raise NoSuchElementException
    No_of_attributes.append(attribute.text)
except NoSuchElementException:
    No_of_attributes.append('-')
time.sleep(2)

# Scraping year
try:
    year = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[2]/td[1]")
    if year.text == "N/A": raise NoSuchElementException
    Year.append(year.text[:4])
except NoSuchElementException:
    Year.append('-')
time.sleep(2)
```

In []:

```
# Creating dataframe for scraped data
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_attributes
ML['Year'] = Year
ML
```

In []: