In [ ]:
```
# first, connect to the webdriver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url = 'https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos'
driver.get(url)
```

In [ ]:
```
# creating empty list for scraping the data

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

In [1]:
```
# scraping Rank of the videos
try:
    for i in driver.find_elements_by_xpath("//table[@class='wikitable sortable jquery-
        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 jquery-
        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 jquery-
        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 jquery-
        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 jquery-
        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('"'))
Wiki
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[1], line 3
      2 try:
----> 3     for i in driver.find_elements_by_xpath("//table[@class='wikitable sortabl
e jquery-tablesorter'][1]/tbody/tr/td[1]"):
      4         Rank.append(i.text)

NameError: name 'driver' is not defined

During handling of the above exception, another exception occurred:

NameError                                 Traceback (most recent call last)
Cell In[1], line 5
      3     for i in driver.find_elements_by_xpath("//table[@class='wikitable sortabl
e jquery-tablesorter'][1]/tbody/tr/td[1]"):
      4         Rank.append(i.text)
----> 5 except NoSuchElementException:
      6     Rank.append("-")
      8 # Scraping Name of the videos

NameError: name 'NoSuchElementException' is not defined
```

In [2]:
```
print(len(Rank),
len(Name),
len(Artist),
len(Date),
len(Views))
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[2], line 1
----> 1 print(len(Rank),
      2 len(Name),
      3 len(Artist),
      4 len(Date),
      5 len(Views))

NameError: name 'Rank' is not defined
```

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

In [ ]:
```

```

In [ ]:
```
# connecting to the webdriver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url=('https://www.bcci.tv/')
driver.get(url)
```

In [ ]:
```
btn=driver.find_element_by_xpath("//div[@class='navigation__drop-down drop-down drop-d
driver.get(btn.get_attribute("href"))
time.sleep(3)
```

```
# creating empty lists for scraping the data
Match_Title = []
Series = []
Place = []
Date = []
Time = []
```

In [3]:
```
for i in driver.find_elements_by_xpath("//div[@class='fixture__format-strip']/span[@cl
    Match_Title.append(i.text)

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

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

for i in driver.find_elements_by_xpath("//span[@class='fixture__datetime tablet-only']
    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
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[3], line 1
----> 1 for i in driver.find_elements_by_xpath("//div[@class='fixture__format-stri
p']/span[@class='u-unskewed-text fixture__format']"):
      2     Match_Title.append(i.text)
      4 for i in driver.find_elements_by_xpath("//div[@class='fixture__format-stri
p']/span[@class='u-unskewed-text fixture__tournament-label u-truncated']"):

NameError: name 'driver' is not defined
```

In [4]:
```
len(url)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[4], line 1
----> 1 len(url)

NameError: name 'url' is not defined
```

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

In [ ]:
```

```

In [ ]:
```
# connecting to the web driver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive
```

```
# getting the webpage of mentioned url
url = ("https://statisticstimes.com/")
driver.get(url)
```

In [ ]:
```
# clicking on Economy button
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").click
time.sleep(3)
```

In [5]:
```
# creating empty list
Rank = []
State = []
GSDP1 = []
GSDP2 = []
Share = []
GDP_billion = []

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

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

# scraping GDP $ billion
```

```
try:
    for i in driver.find_elements_by_xpath("//table[@class='display dataTable']/tbody/
        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
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[5], line 11
     10 try:
---> 11     for i in driver.find_elements_by_xpath("//table[@class='display dataTabl
e']/tbody/tr/td[1]"):
     12         Rank.append(i.text)

NameError: name 'driver' is not defined

During handling of the above exception, another exception occurred:

NameError                                 Traceback (most recent call last)
Cell In[5], line 13
     11     for i in driver.find_elements_by_xpath("//table[@class='display dataTabl
e']/tbody/tr/td[1]"):
     12         Rank.append(i.text)
---> 13 except NoSuchElementException:
     14     Rank.append("_")
     16 # scraping State

NameError: name 'NoSuchElementException' is not defined
```

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

In [ ]:
```

```

In [ ]:
```
# connecting to the web driver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url = ("https://github.com/")
driver.get(url)
```

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

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

In [6]:

```
# creating empty list
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(),'    Contribut
        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//s
            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
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[6], line 10
      7 lang = []
      9 # fetching urls for each repository
---> 10 repository = driver.find_elements_by_xpath("//h1[@class='h3 lh-condensed']//
a")
     11 for i in repository:
     12     URLs.append(i.get_attribute("href"))

NameError: name 'driver' is not defined
```

In [7]:
```
driver.close()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[7], line 1
----> 1 driver.close()

NameError: name 'driver' is not defined
```

In [ ]:
```
```

In [8]:
```
# connecting to the web driver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url = ("https://www.billboard.com/")
driver.get(url)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[8], line 2
      1 # connecting to the web driver
----> 2 driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/ch
romedriver.exe")
      4 # getting the webpage of mentioned url
      5 url = ("https://www.billboard.com/")

NameError: name 'webdriver' is not defined
```

In [9]:
```
# clicking on option button
charts=driver.find_element_by_xpath("//a[@class='header__main-link header__main-link--
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[9], line 2
      1 # clicking on option button
----> 2 charts=driver.find_element_by_xpath("//a[@class='header__main-link header__ma
in-link--charts']").click()

NameError: name 'driver' is not defined
```

In [10]:
```
# creating empty lists
Song_Name = []
Artist_Name =[]
Last_week_rank = []
Peak_rank = []
Weeks_on_board = []
```

```
# getting urls for top 100 songs
urls = driver.find_element_by_xpath("//li[@class='header__submenu__list__element']//a"
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("//span[@class='chart-element__information__sor
    Song_Name.append(i.text)

# scraping data of artist names
for i in driver.find_elements_by_xpath("//span[@class='chart-element__information__art
    Artist_Name.append(i.text)

# scraping data of last week ranks
for i in driver.find_elements_by_xpath("//div[@class='chart-element__meta text--center
    Last_week_rank.append(i.text)


# scraping data of peak ranks
for i in driver.find_elements_by_xpath("//div[@class='chart-element__meta text--center
    Peak_rank.append(i.text)


# scraping data of weeks on board
for i in driver.find_elements_by_xpath("//div[@class='chart-element__meta text--center
    Weeks_on_board.append(i.text)

    # creating dataframe for scraped data
billiboard = pd.DataFrame({})
billiboard['Name'] = Song_Name
billiboard['Artist'] = Artist_Name
billiboard['Last Week Rank'] = Last_week_rank
billiboard['Peak Rank'] = Peak_rank
billiboard['Weeks on board'] = Weeks_on_board
billiboard
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[10], line 9
      6 Weeks_on_board = []
      8 # getting urls for top 100 songs
----> 9 urls = driver.find_element_by_xpath("//li[@class='header__submenu__list__elem
ent']//a")
     10 page_url = urls.get_attribute("href")
     11 driver.get(page_url)

NameError: name 'driver' is not defined
```

In [11]:
```
driver.close()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[11], line 1
----> 1 driver.close()

NameError: name 'driver' is not defined
```

In [ ]:

In [ ]:
```
# connecting to the web driver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url = ("https://www.theguardian.com/news/datablog/2012/aug/09/best-selling-books-all-t
driver.get(url)
time.sleep(3)
```

In [12]:
```
# creating empty lists
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(3)


# 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
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[12], line 10
      6 Genre = []
      9 # scraping book names data
---> 10 for i in driver.find_elements_by_xpath("//tbody//tr//td[2]"):
     11     Book_name.append(i.text)
     14 # scraping author names data

NameError: name 'driver' is not defined
```

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

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[13], line 1
----> 1 driver.close()

NameError: name 'driver' is not defined
```

In [ ]:
```
```

In [ ]:
```
# connecting to the web driver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url = ("https://www.imdb.com/list/ls095964455/")
driver.get(url)
```

In [14]:
```
# creating empty lists
Name = []
Year_span = []
Genre = []
Run_time = []
Ratings = []
Votes = []

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


# scraped data of Year span
for i in driver.find_elements_by_xpath("//span[@class='lister-item-year text-muted unb
    Year_span.append(i.text)


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


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


# scraped data of Ratings
for i in driver.find_elements_by_xpath("//div[@class='ipl-rating-star small']//span[2]
```

```
    Ratings.append(i.text)


# scraped data of Votes
for i in driver.find_elements_by_xpath("//div[@class='lister-item-content']//p[4]/spar
    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
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[14], line 10
      7 Votes = []
      9 # scraped data of Names
---> 10 for i in driver.find_elements_by_xpath("//h3[@class='lister-item-header']/
a"):
     11     Name.append(i.text)
     14 # scraped data of Year span

NameError: name 'driver' is not defined
```

In [15]:
```
driver.close()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[15], line 1
----> 1 driver.close()

NameError: name 'driver' is not defined
```

In [ ]:
```

```

In [ ]:
```
# connecting to the web driver
driver=webdriver.Chrome(r"C:/Users/Lenovo/Downloads/chromedriver_win32 (1)/chromedrive

# getting the webpage of mentioned url
url = (" https://archive.ics.uci.edu/")
driver.get(url)
```

In [ ]:
```
# fetching 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 [ ]:
```
# fetching page urls of all datasets
view_list = driver.find_element_by_xpath("/html/body/table[2]/tbody/tr/td[2]/table[1]/
list_url = view_list.get_attribute("href")
driver.get(list_url)
time.sleep(3)
```

In [ ]:
```
# 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 [16]:
```
# 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(3)



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



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



    # scraping No of Instances
    try:
        instances = driver.find_element_by_xpath("//table[@border='1']//tbody/tr/td[4]
        if instances.text == "N/A": raise NoSuchElementException
        No_of_instances.append(instances.text)
```

```
    except NoSuchElementException:
        No_of_instances.append('-')
    time.sleep(3)



    # scraping No of Arrtibutes
    try:
        attribute = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[2]/td
        if attribute.text == "N/A": raise NoSuchElementException
        No_of_attributes.append(attribute.text)
    except NoSuchElementException:
        No_of_attributes.append('-')
    time.sleep(3)


    # scraping Year
    try:
        year = driver.find_element_by_xpath("//table[@border='1']//tbody/tr[2]/td[6]")
        if year.text == "N/A": raise NoSuchElementException
        Year.append(year.text[:4])
    except NoSuchElementException:
        Year.append('-')
    time.sleep(3)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[16], line 10
      7 No_of_attributes = []
      8 Year = []
---> 10 for i in urls:
     11     driver.get(i)
     12     time.sleep(3)

NameError: name 'urls' is not defined
```

In [17]:
```
# creating dataframe for scraped data
ML = pd.DataFrame({})
ML['Data Name'] = Data_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
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[17], line 2
      1 # creating dataframe for scraped data
----> 2 ML = pd.DataFrame({})
      3 ML['Data Name'] = Data_name
      4 ML['Data Type '] = Data_type

NameError: name 'pd' is not defined
```

In [18]:
```
driver.close()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[18], line 1
----> 1 driver.close()

NameError: name 'driver' is not defined
```

In [ ]: