Data Gathering

Example of gathering image data using webcam

Note: Run this snippet using local jupyter notebook

In [2]:
```
import cv2
from google.colab.patches import cv2_imshow
key = cv2. waitKey(1)
webcam = cv2.VideoCapture(0)
while True:
    try:
        check, frame = webcam.read()
        print(check) #prints true as long as the webcam is running
        print(frame) #prints matrix values of each framecd
        cv2.imshow("Capturing", frame)
        key = cv2.waitKey(1)
        if key == ord('s'):
            cv2.imwrite(filename='saved_img.jpg', img=frame)
            webcam.release()
            img_new = cv2.imread('saved_img.jpg', cv2.IMREAD_GRAYSCALE)
            img_new = cv2.imshow("Captured Image", img_new)
            cv2.waitKey(1650)
            cv2.destroyAllWindows()
            print("Processing image...")
            img_ = cv2.imread('saved_img.jpg', cv2.IMREAD_ANYCOLOR)
            print("Converting RGB image to grayscale...")
            gray = cv2.cvtColor(img_, cv2.COLOR_BGR2GRAY)
            print("Converted RGB image to grayscale...")
            print("Resizing image to 28x28 scale...")
            img_ = cv2.resize(gray,(28,28))
            print("Resized...")
            img_resized = cv2.imwrite(filename='saved_img-final.jpg', img=img_)
            print("Image saved!")

            break
        elif key == ord('q'):
            print("Turning off camera.")
            webcam.release()
            print("Camera off.")
            print("Program ended.")
            cv2.destroyAllWindows()
            break

    except(KeyboardInterrupt):
        print("Turning off camera.")
        webcam.release()
        print("Camera off.")
        print("Program ended.")
        cv2.destroyAllWindows()
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[2], line 1
----> 1 import cv2
      2 from google.colab.patches import cv2_imshow
      3 key = cv2. waitKey(1)

ModuleNotFoundError: No module named 'cv2'
```

Example of gathering voice data using microphone

In [3]:
```
!pip3 install sounddevice
```

```
Requirement already satisfied: sounddevice in c:\users\apuyan\anaconda3\lib\site-pac
kages (0.4.6)
Requirement already satisfied: CFFI>=1.0 in c:\users\apuyan\anaconda3\lib\site-packa
ges (from sounddevice) (1.16.0)
Requirement already satisfied: pycparser in c:\users\apuyan\anaconda3\lib\site-packa
ges (from CFFI>=1.0->sounddevice) (2.21)
```

In [4]:
```
!pip3 install wavio
```

```
Requirement already satisfied: wavio in c:\users\apuyan\anaconda3\lib\site-packages
(0.0.8)
Requirement already satisfied: numpy>=1.19.0 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from wavio) (1.26.4)
```

In [5]:
```
!pip3 install scipy
```

```
Requirement already satisfied: scipy in c:\users\apuyan\anaconda3\lib\site-packages
(1.11.4)
Requirement already satisfied: numpy<1.28.0,>=1.21.6 in c:\users\apuyan\anaconda3\li
b\site-packages (from scipy) (1.26.4)
```

In [6]:
```
!apt-get install libportaudio2
```

```
'apt-get' is not recognized as an internal or external command,
operable program or batch file.
```

In [7]:
```
# import required libraries
import sounddevice as sd
from scipy.io.wavfile import write
import wavio as wv

# Sampling frequency
freq = 44100

# Recording duration
duration = 5

# Start recorder with the given values
# of duration and sample frequency
recording = sd.rec(int(duration * freq),
                   samplerate=freq, channels=2)

# Record audio for the given number of seconds
sd.wait()
```

```
# This will convert the NumPy array to an audio
# file with the given sampling frequency
write("recording0.wav", freq, recording)

# Convert the NumPy array to audio file
wv.write("recording1.wav", recording, freq, sampwidth=2)
```

Web Scraping

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. The web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis.

Image Scraping using BeautifulSoup and Request

In [8]:
```
!pip install bs4
```

```
Requirement already satisfied: bs4 in c:\users\apuyan\anaconda3\lib\site-packages
(0.0.2)
Requirement already satisfied: beautifulsoup4 in c:\users\apuyan\anaconda3\lib\site-
packages (from bs4) (4.12.2)
Requirement already satisfied: soupsieve>1.2 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from beautifulsoup4->bs4) (2.5)
```

In [9]:
```
pip install requests
```

```
Requirement already satisfied: requests in c:\users\apuyan\anaconda3\lib\site-packag
es (2.31.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\apuyan\anaconda3
\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\apuyan\anaconda3\lib\site-pa
ckages (from requests) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\apuyan\anaconda3\lib\s
ite-packages (from requests) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\apuyan\anaconda3\lib\s
ite-packages (from requests) (2024.2.2)
Note: you may need to restart the kernel to use updated packages.
```

In [10]:
```
import requests
from bs4 import BeautifulSoup

def getdata(url):
  r = requests.get(url)
  return r.text

htmldata = getdata("https://www.google.com/")
soup = BeautifulSoup(htmldata, 'html.parser')
```

```
for item in soup.find_all('img'):
  print(item['src'])
```

```
/images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
```

In [11]:
```
pip install selenium
```

```
Requirement already satisfied: selenium in c:\users\apuyan\anaconda3\lib\site-packag
es (4.18.1)
Requirement already satisfied: urllib3<3,>=1.26 in c:\users\apuyan\anaconda3\lib\sit
e-packages (from urllib3[socks]<3,>=1.26->selenium) (2.0.7)
Requirement already satisfied: trio~=0.17 in c:\users\apuyan\anaconda3\lib\site-pack
ages (from selenium) (0.25.0)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\apuyan\anaconda3\lib
\site-packages (from selenium) (0.11.1)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\apuyan\anaconda3\lib\s
ite-packages (from selenium) (2024.2.2)
Requirement already satisfied: typing_extensions>=4.9.0 in c:\users\apuyan\anaconda3
\lib\site-packages (from selenium) (4.9.0)
Requirement already satisfied: attrs>=23.2.0 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from trio~=0.17->selenium) (23.2.0)
Requirement already satisfied: sortedcontainers in c:\users\apuyan\anaconda3\lib\sit
e-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: idna in c:\users\apuyan\anaconda3\lib\site-packages
(from trio~=0.17->selenium) (3.4)
Requirement already satisfied: outcome in c:\users\apuyan\anaconda3\lib\site-package
s (from trio~=0.17->selenium) (1.3.0.post0)
Requirement already satisfied: sniffio>=1.3.0 in c:\users\apuyan\anaconda3\lib\site-
packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: cffi>=1.14 in c:\users\apuyan\anaconda3\lib\site-pack
ages (from trio~=0.17->selenium) (1.16.0)
Requirement already satisfied: wsproto>=0.14 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from trio-websocket~=0.9->selenium) (1.2.0)
Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in c:\users\apuyan\anacon
da3\lib\site-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: pycparser in c:\users\apuyan\anaconda3\lib\site-packa
ges (from cffi>=1.14->trio~=0.17->selenium) (2.21)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.14.0)
Note: you may need to restart the kernel to use updated packages.
```

Image Scraping using Selenium

In [115…
```
!pip install selenium
import sys
sys.path.insert(0,'/usr/lib/chromium-browser/chromedriver')


from selenium import webdriver
from selenium.webdriver.common.by import By
import time
import requests
import shutil
import os
import getpass
import urllib.request
```

```
import io
import time
from PIL import Image

user = getpass.getuser()
chrome_options = webdriver.ChromeOptions()
chrome_options.add_argument('--headless')
chrome_options.add_argument('--no-sandbox')
chrome_options.add_argument('--disable-dev-shm-usage')

driver = webdriver.Chrome()

def scroll_to_end(driver):
    driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
    time.sleep(5)#sleep_between_interactions


def getImageUrls(name,totalImgs,driver):
    search_url = "https://www.google.com/search?q=cat&tbm=isch&ved=2ahUKEwjNn_Gn7Yy
    driver.get(search_url)
    img_urls = set()
    img_count = 0
    results_start = 0

    while(img_count+results_start<totalImgs): #Extract actual images now
        scroll_to_end(driver)
        totalResults = driver.find_elements(By.CLASS_NAME,"Q4LuWd")
        print('total results:', len(totalResults))
        print(f"Found: {totalResults} search results. Extracting links from{results
        for img in totalResults[results_start:totalImgs]:
            img.click()
            time.sleep(5)
            image = driver.find_element(By.CLASS_NAME,'iPVvYb')
            img_urls.add(image.get_attribute('src'))
            print(img_urls)
            img_count=len(img_urls)
            print(img_count)

    return img_urls

def downloadImages(folder_path,file_name,url):
    try:
        image_content = requests.get(url).content
    except Exception as e:
        print(f"ERROR - COULD NOT DOWNLOAD {url} - {e}")
    try:
        image_file = io.BytesIO(image_content)
        image = Image.open(image_file).convert('RGB')
        file_path = os.path.join(folder_path, file_name)
        with open(file_path, 'wb') as f:
            image.save(f, "JPEG", quality=85)
        print(f"SAVED - {url} - AT: {file_path}")
    except Exception as e:
        print(f"ERROR - COULD NOT SAVE {url} - {e}")

def saveInDestFolder(searchNames,destDir,totalImgs,driver):
```

```
    for name in list(searchNames):
        path=os.path.join(destDir,name)
        if not os.path.isdir(path):
            os.mkdir(path)
        print('Current Path',path)
        totalLinks=getImageUrls(name,totalImgs,driver)
        print('totalLinks',totalLinks)

    if totalLinks is None:
        print('images not found for :',name)

    else:
        for i, link in enumerate(totalLinks):
            file_name = f"{i:150}.jpg"
            downloadImages(path,file_name,link)

searchNames=['cat']
destDir=f'C:/Users/apuyan/Desktop/HOA 7.2 Webscraping using BeautifulSoup and Reque
totalImgs=5

saveInDestFolder(searchNames,destDir,totalImgs,driver)
```

```
Requirement already satisfied: selenium in c:\users\apuyan\anaconda3\lib\site-packag
es (4.18.1)
Requirement already satisfied: urllib3<3,>=1.26 in c:\users\apuyan\anaconda3\lib\sit
e-packages (from urllib3[socks]<3,>=1.26->selenium) (2.0.7)
Requirement already satisfied: trio~=0.17 in c:\users\apuyan\anaconda3\lib\site-pack
ages (from selenium) (0.25.0)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\apuyan\anaconda3\lib
\site-packages (from selenium) (0.11.1)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\apuyan\anaconda3\lib\s
ite-packages (from selenium) (2024.2.2)
Requirement already satisfied: typing_extensions>=4.9.0 in c:\users\apuyan\anaconda3
\lib\site-packages (from selenium) (4.9.0)
Requirement already satisfied: attrs>=23.2.0 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from trio~=0.17->selenium) (23.2.0)
Requirement already satisfied: sortedcontainers in c:\users\apuyan\anaconda3\lib\sit
e-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: idna in c:\users\apuyan\anaconda3\lib\site-packages
(from trio~=0.17->selenium) (3.4)
Requirement already satisfied: outcome in c:\users\apuyan\anaconda3\lib\site-package
s (from trio~=0.17->selenium) (1.3.0.post0)
Requirement already satisfied: sniffio>=1.3.0 in c:\users\apuyan\anaconda3\lib\site-
packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: cffi>=1.14 in c:\users\apuyan\anaconda3\lib\site-pack
ages (from trio~=0.17->selenium) (1.16.0)
Requirement already satisfied: wsproto>=0.14 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from trio-websocket~=0.9->selenium) (1.2.0)
Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in c:\users\apuyan\anacon
da3\lib\site-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: pycparser in c:\users\apuyan\anaconda3\lib\site-packa
ges (from cffi>=1.14->trio~=0.17->selenium) (2.21)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\apuyan\anaconda3\lib\site-p
ackages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.14.0)
Current Path C:/Users/apuyan/Desktop/HOA 7.2 Webscraping using BeautifulSoup and Req
uests\cat
total results: 100
Found: [<selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc0
11de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2
275A54F968E1.e.10")>, <selenium.webdriver.remote.webelement.WebElement (session="49e
b72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5
533733909361D2275A54F968E1.e.12")>, <selenium.webdriver.remote.webelement.WebElement
(session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6
F9F3.d.CA0BD5533733909361D2275A54F968E1.e.14")>, <selenium.webdriver.remote.webeleme
nt.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E7
0B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.16")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.18")>, <seleni
um.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.20")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc
011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D
2275A54F968E1.e.22")>, <selenium.webdriver.remote.webelement.WebElement (session="49
eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD
5533733909361D2275A54F968E1.e.24")>, <selenium.webdriver.remote.webelement.WebElemen
t (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649B
C6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.26")>, <selenium.webdriver.remote.webele
ment.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6
```

E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.27")>, <selenium.webdrive
r.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element
="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.28")>, <se
lenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08
786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E
1.e.39")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1
dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD553373390936
1D2275A54F968E1.e.40")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.41")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.42")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.43")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.44")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.45")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.46")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.47")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.48")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.50")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.52")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.54")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.56")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.58")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.60")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.62")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.64")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.66")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.68")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.70")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.72")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.74")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.76")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b

08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.78")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.80")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.38")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.83")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.85")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.87")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.89")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.91")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.93")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.95")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.97")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.99")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.101")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a0516
2c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD553373390
9361D2275A54F968E1.e.103")>, <selenium.webdriver.remote.webelement.WebElement (sessi
on="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.
d.CA0BD5533733909361D2275A54F968E1.e.105")>, <selenium.webdriver.remote.webelement.W
ebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A
6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.107")>, <selenium.webdriver.remo
te.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE
2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.109")>, <seleniu
m.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.111")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.113")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.115")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.117")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287
3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.119")>, <selenium.we
bdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", el
ement="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.121")
>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4
203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A5
4F968E1.e.123")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a
05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD55337
33909361D2275A54F968E1.e.125")>, <selenium.webdriver.remote.webelement.WebElement (s
ession="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9
F3.d.CA0BD5533733909361D2275A54F968E1.e.127")>, <selenium.webdriver.remote.webelemen

t.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70
B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.129")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.131")>, <selen
ium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.133")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.135")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.137")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.139")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287
3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.141")>, <selenium.we
bdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", el
ement="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.143")
>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4
203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A5
4F968E1.e.145")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a
05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD55337
33909361D2275A54F968E1.e.147")>, <selenium.webdriver.remote.webelement.WebElement (s
ession="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9
F3.d.CA0BD5533733909361D2275A54F968E1.e.149")>, <selenium.webdriver.remote.webelemen
t.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70
B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.151")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.153")>, <selen
ium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.155")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.157")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.159")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.161")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287
3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.163")>, <selenium.we
bdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", el
ement="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.165")
>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4
203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A5
4F968E1.e.167")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a
05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD55337
33909361D2275A54F968E1.e.169")>, <selenium.webdriver.remote.webelement.WebElement (s
ession="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9
F3.d.CA0BD5533733909361D2275A54F968E1.e.171")>, <selenium.webdriver.remote.webelemen
t.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70
B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.173")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.175")>, <selen
ium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.177")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361

D2275A54F968E1.e.179")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.181")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.183")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287
3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.185")>, <selenium.we
bdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", el
ement="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.187")
>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4
203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A5
4F968E1.e.189")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a
05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD55337
33909361D2275A54F968E1.e.191")>, <selenium.webdriver.remote.webelement.WebElement (s
ession="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9
F3.d.CA0BD5533733909361D2275A54F968E1.e.193")>, <selenium.webdriver.remote.webelemen
t.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70
B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.195")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.197")>, <selen
ium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.199")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.201")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.203")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.205")>] search results. Extracting
links from0:[<selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.10")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.12")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.14")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.16")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.18")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.20")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.22")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.24")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.26")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.27")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.28")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.39")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162

c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.40")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.41")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.42")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.43")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.44")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.45")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.46")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.47")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.48")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.50")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.52")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.54")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.56")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.58")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.60")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.62")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.64")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.66")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.68")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.70")>, <selenium.webdriver.remote.webelement.WebEl
ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5
649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.72")>, <selenium.webdriver.remote.we
belement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D
4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.74")>, <selenium.webdr
iver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", eleme
nt="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.76")>, <
selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b
08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F96
8E1.e.78")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162
c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909
361D2275A54F968E1.e.80")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.38")>, <selenium.webdriver.remote.webelement.WebEl

ement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.83")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.85")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.87")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.89")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.91")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.93")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.95")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.97")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.99")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.101")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.103")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.105")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.107")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.109")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.111")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.113")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.115")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.117")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.119")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.121")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.123")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.125")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.127")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.129")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.131")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.

e.133")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.135")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.137")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.139")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287
3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.141")>, <selenium.we
bdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", el
ement="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.143")
>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4
203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A5
4F968E1.e.145")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a
05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD55337
33909361D2275A54F968E1.e.147")>, <selenium.webdriver.remote.webelement.WebElement (s
ession="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9
F3.d.CA0BD5533733909361D2275A54F968E1.e.149")>, <selenium.webdriver.remote.webelemen
t.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70
B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.151")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.153")>, <selen
ium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.155")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.157")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.159")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.161")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287
3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.163")>, <selenium.we
bdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", el
ement="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.165")
>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4
203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A5
4F968E1.e.167")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a
05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD55337
33909361D2275A54F968E1.e.169")>, <selenium.webdriver.remote.webelement.WebElement (s
ession="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9
F3.d.CA0BD5533733909361D2275A54F968E1.e.171")>, <selenium.webdriver.remote.webelemen
t.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70
B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.173")>, <selenium.webdriver.r
emote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.
8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.175")>, <selen
ium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786
f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.
e.177")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1d
c011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361
D2275A54F968E1.e.179")>, <selenium.webdriver.remote.webelement.WebElement (session
="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.C
A0BD5533733909361D2275A54F968E1.e.181")>, <selenium.webdriver.remote.webelement.WebE
lement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF
5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.183")>, <selenium.webdriver.remote.
webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE287

3D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.185")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.187")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.189")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.191")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.193")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.195")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.197")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.199")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.201")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.203")>, <selenium.webdriver.remote.webelement.WebElement (session="49eb72a05162c1dc011de4203b08786f", element="f.8DE2873D4CF6E70B1A6CF5649BC6F9F3.d.CA0BD5533733909361D2275A54F968E1.e.205")>]
{'https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeographic_2572187_square.jpg'}
1
{'https://cdn.britannica.com/70/234870-050-D4D024BB/Orange-colored-cat-yawns-displaying-teeth.jpg', 'https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeographic_2572187_square.jpg'}
2
{'https://cdn.britannica.com/70/234870-050-D4D024BB/Orange-colored-cat-yawns-displaying-teeth.jpg', 'https://upload.wikimedia.org/wikipedia/commons/thumb/1/15/Cat_August_2010-4.jpg/1200px-Cat_August_2010-4.jpg', 'https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeographic_2572187_square.jpg'}
3
{'https://cdn.britannica.com/70/234870-050-D4D024BB/Orange-colored-cat-yawns-displaying-teeth.jpg', 'https://upload.wikimedia.org/wikipedia/commons/thumb/1/15/Cat_August_2010-4.jpg/1200px-Cat_August_2010-4.jpg', 'https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeographic_2572187_square.jpg', 'https://cdn.britannica.com/34/235834-050-C5843610/two-different-breeds-of-cats-side-by-side-outdoors-in-the-garden.jpg'}
4
{'https://upload.wikimedia.org/wikipedia/commons/thumb/1/15/Cat_August_2010-4.jpg/1200px-Cat_August_2010-4.jpg', 'https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeographic_2572187_square.jpg', 'https://cdn.britannica.com/34/235834-050-C5843610/two-different-breeds-of-cats-side-by-side-outdoors-in-the-garden.jpg', 'https://cdn.britannica.com/70/234870-050-D4D024BB/Orange-colored-cat-yawns-displaying-teeth.jpg', 'https://media.4-paws.org/5/b/4/b/5b4b5a91dd9443fa1785ee7fca66850e06dcc7f9/VIER%20PFOTEN_2019-12-13_209-2890x2000-1920x1329.jpg'}
5
totalLinks {'https://upload.wikimedia.org/wikipedia/commons/thumb/1/15/Cat_August_2010-4.jpg/1200px-Cat_August_2010-4.jpg', 'https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeographic_2572187_square.jpg', 'https://cdn.britannica.com/34/235834-050-C5843610/two-different-breeds-of-cats-side-by-side-outdoors-in-the-garden.jpg', 'https://cdn.britannica.com/70/234870-050-D4D024BB/Orange-colored-cat-yawns-displaying-teeth.jpg', 'https://media.4-paws.org/5/b/4/b/5b4b5a91dd9443fa1785ee

```
7fca66850e06dcc7f9/VIER%20PFOTEN_2019-12-13_209-2890x2000-1920x1329.jpg'}
SAVED - https://upload.wikimedia.org/wikipedia/commons/thumb/1/15/Cat_August_2010-4.
jpg/1200px-Cat_August_2010-4.jpg - AT: C:/Users/apuyan/Desktop/HOA 7.2 Webscraping u
sing BeautifulSoup and Requests\cat\
0.jpg
SAVED - https://i.natgeofe.com/n/548467d8-c5f1-4551-9f58-6817a8d2c45e/NationalGeogra
phic_2572187_square.jpg - AT: C:/Users/apuyan/Desktop/HOA 7.2 Webscraping using Beau
tifulSoup and Requests\cat\
1.jpg
SAVED - https://cdn.britannica.com/34/235834-050-C5843610/two-different-breeds-of-ca
ts-side-by-side-outdoors-in-the-garden.jpg - AT: C:/Users/apuyan/Desktop/HOA 7.2 Web
scraping using BeautifulSoup and Requests\cat\
2.jpg
SAVED - https://cdn.britannica.com/70/234870-050-D4D024BB/Orange-colored-cat-yawns-d
isplaying-teeth.jpg - AT: C:/Users/apuyan/Desktop/HOA 7.2 Webscraping using Beautifu
lSoup and Requests\cat\
3.jpg
SAVED - https://media.4-paws.org/5/b/4/b/5b4b5a91dd9443fa1785ee7fca66850e06dcc7f9/VI
ER%20PFOTEN_2019-12-13_209-2890x2000-1920x1329.jpg - AT: C:/Users/apuyan/Desktop/HOA
7.2 Webscraping using BeautifulSoup and Requests\cat\
4.jpg
```

Web Scraping of Movies Information using BeautifulSoup

We want to analyze the distributions of IMDB and Metacritic movie ratings to see if we find anything interesting. To do this, we'll first scrape data for over 2000 movies.

In [13]:
```
from requests import get
url = 'https://www.imdb.com/search/title/?release_date=2017-01-01,2017-12-31&sort=n
agent = {'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
response = get(url,headers=agent)
print(response.text[:500])
```

```
<!DOCTYPE html><html lang="en-US" xmlns:og="http://opengraphprotocol.org/schema/" xm
lns:fb="http://www.facebook.com/2008/fbml"><head><meta charSet="utf-8"/><meta name
="viewport" content="width=device-width"/><script>if(typeof uet === 'function'){ uet
('bb', 'LoadTitle', {wb: 1}); }</script><script>window.addEventListener('load', (eve
nt) => {
        if (typeof window.csa !== 'undefined' && typeof window.csa === 'function') {
            var csaLatencyPlugin = window.csa('Content', {
```

In [14]:
```
from bs4 import BeautifulSoup
html_soup = BeautifulSoup(response.text, 'html.parser')
headers = {'Accept-Language': 'en-US,en;q=0.8'}
type(html_soup)
```

Out[14]:
```
bs4.BeautifulSoup
```

In [15]:
```
movie_containers = html_soup.find_all('div', class_ = 'sc-ab6fa25a-3 bVYfLY dli-par
print(type(movie_containers))
print(len(movie_containers))
```

```
<class 'bs4.element.ResultSet'>
50
```

First Movie

Extracting the data for a single movie

We can access the first container, which contains information about a single movie, by using list notation on movie_containers.

In [16]:
```
first_movie = movie_containers[0]
first_movie
```

Out[16]:
```
<div class="sc-ab6fa25a-3 bVYfLY dli-parent"><div class="sc-ab6fa25a-2 gOsifL"><di
v class="sc-e5a25b0f-0 jQjDIb dli-poster-container"><div class="ipc-poster ipc-pos
ter--base ipc-poster--dynamic-width ipc-sub-grid-item ipc-sub-grid-item--span-2" r
ole="group"><div aria-label="add to watchlist" class="ipc-watchlist-ribbon ipc-foc
usable ipc-watchlist-ribbon--s ipc-watchlist-ribbon--base ipc-watchlist-ribbon--lo
ading ipc-watchlist-ribbon--onImage ipc-poster__watchlist-ribbon" role="button" ta
bindex="0"><svg class="ipc-watchlist-ribbon__bg" height="34px" role="presentation"
viewbox="0 0 24 34" width="24px" xmlns="http://www.w3.org/2000/svg"><polygon class
="ipc-watchlist-ribbon__bg-ribbon" fill="#000000" points="24 0 0 0 0 32 12.2436611
26.2926049 24 31.7728343"></polygon><polygon class="ipc-watchlist-ribbon__bg-hove
r" points="24 0 0 0 0 32 12.2436611 26.2926049 24 31.7728343"></polygon><polygon c
lass="ipc-watchlist-ribbon__bg-shadow" points="24 31.7728343 24 33.7728343 12.2436
611 28.2926049 0 34 0 32 12.2436611 26.2926049"></polygon></svg><div class="ipc-wa
tchlist-ribbon__icon" role="presentation"><svg class="ipc-loader ipc-loader--circl
e ipc-watchlist-ribbon__loader" data-testid="watchlist-ribbon-loader" height="48p
x" role="presentation" version="1.1" viewbox="0 0 48 48" width="48px" xmlns="htt
p://www.w3.org/2000/svg"><g class="ipc-loader__container" fill="currentColor"><cir
cle class="ipc-loader__circle ipc-loader__circle--one" cx="24" cy="9" r="4"></circ
le><circle class="ipc-loader__circle ipc-loader__circle--two" cx="35" cy="14" r
="4"></circle><circle class="ipc-loader__circle ipc-loader__circle--three" cx="39"
cy="24" r="4"></circle><circle class="ipc-loader__circle ipc-loader__circle--four"
cx="35" cy="34" r="4"></circle><circle class="ipc-loader__circle ipc-loader__circl
e--five" cx="24" cy="39" r="4"></circle><circle class="ipc-loader__circle ipc-load
er__circle--six" cx="13" cy="34" r="4"></circle><circle class="ipc-loader__circle
ipc-loader__circle--seven" cx="9" cy="24" r="4"></circle><circle class="ipc-loader
__circle ipc-loader__circle--eight" cx="13" cy="14" r="4"></circle></g></svg></div
></div><div class="ipc-media ipc-media--poster-27x40 ipc-image-media-ratio--poster
-27x40 ipc-media--base ipc-media--poster-m ipc-poster__poster-image ipc-media__im
g" style="width:100%"><img alt="Hugh Jackman in Logan (2017)" class="ipc-image" lo
ading="lazy" sizes="50vw, (min-width: 480px) 34vw, (min-width: 600px) 26vw, (min-w
idth: 1024px) 16vw, (min-width: 1280px) 16vw" src="https://m.media-amazon.com/imag
es/M/MV5BYzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJhZTAzXkEyXkFqcGdeQXVyNjc1NTYyM
jg@._V1_QL75_UX140_CR0,1,140,207_.jpg" srcset="https://m.media-amazon.com/images/
M/MV5BYzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJhZTAzXkEyXkFqcGdeQXVyNjc1NTYyMjg
@._V1_QL75_UX140_CR0,1,140,207_.jpg 140w, https://m.media-amazon.com/images/M/MV5B
Yzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJhZTAzXkEyXkFqcGdeQXVyNjc1NTYyMjg@._V1_Q
L75_UX210_CR0,2,210,311_.jpg 210w, https://m.media-amazon.com/images/M/MV5BYzc5MTU
4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJhZTAzXkEyXkFqcGdeQXVyNjc1NTYyMjg@._V1_QL75_UX2
80_CR0,3,280,414_.jpg 280w" width="140"/></div><a aria-label="View title page for
Logan" class="ipc-lockup-overlay ipc-focusable" href="/title/tt3315342/?ref_=sr_i_
1"><div class="ipc-lockup-overlay__screen"></div></a></div></div><div class="sc-b0
691f29-0 jbYPfh"><div class="ipc-title ipc-title--base ipc-title--title ipc-title-
link-no-icon ipc-title--on-textPrimary sc-b0691f29-9 klOwFB dli-title"><a class="i
pc-title-link-wrapper" href="/title/tt3315342/?ref_=sr_t_1" tabindex="0"><h3 class
="ipc-title__text">1. Logan</h3></a></div><div class="sc-b0691f29-7 hrgukm dli-tit
le-metadata"><span class="sc-b0691f29-8 ilsLEX dli-title-metadata-item">2017</span
><span class="sc-b0691f29-8 ilsLEX dli-title-metadata-item">2h 17m</span><span cla
ss="sc-b0691f29-8 ilsLEX dli-title-metadata-item">R-16</span></div><span class="sc
-b0691f29-1 grHDBY"><div class="sc-e2dbc1a3-0 ajrIH sc-b0691f29-2 bhhtyj dli-ratin
gs-container" data-testid="ratingGroup--container"><span aria-label="IMDb rating:
8.1" class="ipc-rating-star ipc-rating-star--base ipc-rating-star--imdb ratingGrou
p--imdb-rating" data-testid="ratingGroup--imdb-rating"><svg class="ipc-icon ipc-ic
on--star-inline" fill="currentColor" height="24" role="presentation" viewbox="0 0
24 24" width="24" xmlns="http://www.w3.org/2000/svg"><path d="M12 20.1l5.82 3.682c
1.066.675 2.37-.322 2.09-1.584l-1.543-6.926 5.146-4.667c.94-.85.435-2.465-.799-2.5
67l-6.773-.602L13.29.89a1.38 1.38 0 0 0-2.581 0l-2.65 6.53-6.774.602C.052 8.126-.4
```

53 9.74.486 10.59l5.147 4.666-1.542 6.926c-.28 1.262 1.023 2.26 2.09 1.585L12 20.0
99z"></path></svg>8.1 (<!-- -->827K<!-- -
->)<button aria-label="Rate Logan" class="ipc-rate-button sc-e2dbc1a
3-1 jboOQc ratingGroup--user-rating ipc-rate-button--unrated ipc-rate-button--bas
e" data-testid="rate-button"><span class="ipc-rating-star ipc-rating-star--base ip
c-rating-star--rate"><svg class="ipc-icon ipc-icon--star-border-inline" fill="curr
entColor" height="24" role="presentation" viewbox="0 0 24 24" width="24" xmlns="ht
tp://www.w3.org/2000/svg"><path d="M22.724 8.217l-6.786-.587-2.65-6.22c-.477-1.133
-2.103-1.133-2.58 0l-2.65 6.234-6.772.573c-1.234.098-1.739 1.636-.8 2.446l5.146 4.
446-1.542 6.598c-.28 1.202 1.023 2.153 2.09 1.51l5.818-3.495 5.819 3.509c1.065.643
2.37-.308 2.089-1.51l-1.542-6.612 5.145-4.446c.94-.81.45-2.348-.785-2.446zm-10.726
8.89l-5.272 3.174 1.402-5.983-4.655-4.026 6.141-.531 2.384-5.634 2.398 5.648 6.14.
531-4.654 4.026 1.402 5.983-5.286-3.187z"></path></svg><span class="ipc-rating-sta
r--rate">Rate</button> <span
class="sc-b0901df4-0 bcQdDJ metacritic-score-box" style="background-color:#54A72
A">77Metascore</d
iv> <button aria-disabled="fals
e" aria-label="See more information about Logan" class="ipc-icon-button dli-info-i
con ipc-icon-button--base ipc-icon-button--onAccent2" role="button" tabindex="0" t
itle="See more information about Logan"><svg class="ipc-icon ipc-icon--info" fill
="currentColor" height="24" role="presentation" viewbox="0 0 24 24" width="24" xml
ns="http://www.w3.org/2000/svg"><path d="M0 0h24v24H0V0z" fill="none"></path><path
d="M11 7h2v2h-2zm0 4h2v6h-2zm1-9C6.48 2 2 6.48 2 12s4.48 10 10 10 10-4.48 10-10S1
7.52 2 12 2zm0 18c-4.41 0-8-3.59-8-8s3.59-8 8-8 8 3.59 8 8-3.59 8-8 8z"></path></s
vg></button> <div class="ipc-html-con
tent ipc-html-content--base sc-ab6fa25a-0 bhexuD dli-plot-container" role="present
ation"> In a future where mutants are nearl
y extinct, an elderly and weary Logan leads a quiet life. But when Laura, a mutant
child pursued by scientists, comes to him for help, he must get her to safety.</di
v>

The name of the movie

In [17]:
```
first_movie.div
```

Out[17]:
```
<div class="sc-ab6fa25a-2 gOsifL"><div class="sc-e5a25b0f-0 jQjDIb dli-poster-cont
ainer"><div class="ipc-poster ipc-poster--base ipc-poster--dynamic-width ipc-sub-g
rid-item ipc-sub-grid-item--span-2" role="group"><div aria-label="add to watchlis
t" class="ipc-watchlist-ribbon ipc-focusable ipc-watchlist-ribbon--s ipc-watchlist
-ribbon--base ipc-watchlist-ribbon--loading ipc-watchlist-ribbon--onImage ipc-post
er__watchlist-ribbon" role="button" tabindex="0"><svg class="ipc-watchlist-ribbon_
_bg" height="34px" role="presentation" viewbox="0 0 24 34" width="24px" xmlns="htt
p://www.w3.org/2000/svg"><polygon class="ipc-watchlist-ribbon__bg-ribbon" fill="#0
00000" points="24 0 0 0 0 32 12.2436611 26.2926049 24 31.7728343"></polygon><polyg
on class="ipc-watchlist-ribbon__bg-hover" points="24 0 0 0 0 32 12.2436611 26.2926
049 24 31.7728343"></polygon><polygon class="ipc-watchlist-ribbon__bg-shadow" poin
ts="24 31.7728343 24 33.7728343 12.2436611 28.2926049 0 34 0 32 12.2436611 26.2926
049"></polygon></svg><div class="ipc-watchlist-ribbon__icon" role="presentation"><
svg class="ipc-loader ipc-loader--circle ipc-watchlist-ribbon__loader" data-testid
="watchlist-ribbon-loader" height="48px" role="presentation" version="1.1" viewbox
="0 0 48 48" width="48px" xmlns="http://www.w3.org/2000/svg"><g class="ipc-loader_
_container" fill="currentColor"><circle class="ipc-loader__circle ipc-loader__circ
le--one" cx="24" cy="9" r="4"></circle><circle class="ipc-loader__circle ipc-loade
r__circle--two" cx="35" cy="14" r="4"></circle><circle class="ipc-loader__circle i
pc-loader__circle--three" cx="39" cy="24" r="4"></circle><circle class="ipc-loader
__circle ipc-loader__circle--four" cx="35" cy="34" r="4"></circle><circle class="i
pc-loader__circle ipc-loader__circle--five" cx="24" cy="39" r="4"></circle><circle
class="ipc-loader__circle ipc-loader__circle--six" cx="13" cy="34" r="4"></circle>
<circle class="ipc-loader__circle ipc-loader__circle--seven" cx="9" cy="24" r="4">
</circle><circle class="ipc-loader__circle ipc-loader__circle--eight" cx="13" cy
="14" r="4"></circle></g></svg></div></div><div class="ipc-media ipc-media--poster
-27x40 ipc-image-media-ratio--poster-27x40 ipc-media--base ipc-media--poster-m ipc
-poster__poster-image ipc-media__img" style="width:100%"><img alt="Hugh Jackman in
Logan (2017)" class="ipc-image" loading="lazy" sizes="50vw, (min-width: 480px) 34v
w, (min-width: 600px) 26vw, (min-width: 1024px) 16vw, (min-width: 1280px) 16vw" sr
c="https://m.media-amazon.com/images/M/MV5BYzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTE
yMzJhZTAzXkEyXkFqcGdeQXVyNjc1NTYyMjg@._V1_QL75_UX140_CR0,1,140,207_.jpg" srcset="h
ttps://m.media-amazon.com/images/M/MV5BYzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJ
hZTAzXkEyXkFqcGdeQXVyNjc1NTYyMjg@._V1_QL75_UX140_CR0,1,140,207_.jpg 140w, https://
m.media-amazon.com/images/M/MV5BYzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJhZTAzXk
EyXkFqcGdeQXVyNjc1NTYyMjg@._V1_QL75_UX210_CR0,2,210,311_.jpg 210w, https://m.media
-amazon.com/images/M/MV5BYzc5MTU4N2EtYTkyMi00NjdhLTg3NWEtMTY4OTEyMzJhZTAzXkEyXkFqc
GdeQXVyNjc1NTYyMjg@._V1_QL75_UX280_CR0,3,280,414_.jpg 280w" width="140"/></div><a
aria-label="View title page for Logan" class="ipc-lockup-overlay ipc-focusable" hr
ef="/title/tt3315342/?ref_=sr_i_1"><div class="ipc-lockup-overlay__screen"></div>
</a></div></div><div class="sc-b0691f29-0 jbYPfh"><div class="ipc-title ipc-title-
-base ipc-title--title ipc-title-link-no-icon ipc-title--on-textPrimary sc-b0691f2
9-9 klOwFB dli-title"><a class="ipc-title-link-wrapper" href="/title/tt3315342/?re
f_=sr_t_1" tabindex="0"><h3 class="ipc-title__text">1. Logan</h3></a></div><div cl
ass="sc-b0691f29-7 hrgukm dli-title-metadata"><span class="sc-b0691f29-8 ilsLEX dl
i-title-metadata-item">2017</span><span class="sc-b0691f29-8 ilsLEX dli-title-meta
data-item">2h 17m</span><span class="sc-b0691f29-8 ilsLEX dli-title-metadata-ite
m">R-16</span></div><span class="sc-b0691f29-1 grHDBY"><div class="sc-e2dbc1a3-0 a
jrIH sc-b0691f29-2 bhhtyj dli-ratings-container" data-testid="ratingGroup--contain
er"><span aria-label="IMDb rating: 8.1" class="ipc-rating-star ipc-rating-star--ba
se ipc-rating-star--imdb ratingGroup--imdb-rating" data-testid="ratingGroup--imdb-
rating"><svg class="ipc-icon ipc-icon--star-inline" fill="currentColor" height="2
4" role="presentation" viewbox="0 0 24 24" width="24" xmlns="http://www.w3.org/200
0/svg"><path d="M12 20.1l5.82 3.682c1.066.675 2.37-.322 2.09-1.584l-1.543-6.926 5.
146-4.667c.94-.85.435-2.465-.799-2.567l-6.773-.602L13.29.89a1.38 1.38 0 0 0-2.581
0l-2.65 6.53-6.774.602C.052 8.126-.453 9.74.486 10.59l5.147 4.666-1.542 6.926c-.28
```

1.262 1.023 2.26 2.09 1.585L12 20.099z"></path></svg>8.1 (<!-- -->827K<!-- -->)<button aria-label="Rate Logan" class="ipc-rate-button sc-e2dbc1a3-1 jboOQc ratingGroup--user-rating ipc-rate-button--unrated ipc-rate-button--base" data-testid="rate-button"><svg class="ipc-icon ipc-icon--star-border-inline" fill="currentColor" height="24" role="presentation" viewbox="0 0 24 24" width="24" xmlns="http://www.w3.org/2000/svg"><path d="M22.724 8.217l-6.786-.587-2.65-6.22c-.477-1.133-2.103-1.133-2.58 0l-2.65 6.234-6.772.573c-1.234.098-1.739 1.636-.8 2.446l5.146 4.446-1.542 6.598c-.28 1.202 1.023 2.153 2.09 1.51l5.818-3.495 5.819 3.509c1.065.643 2.37-.308 2.089-1.51l-1.542-6.612 5.145-4.446c.94-.81.45-2.348-.785-2.446zm-10.726 8.89l-5.272 3.174 1.402-5.983-4.655-4.026 6.141-.531 2.384-5.634 2.398 5.648 6.14.531-4.654 4.026 1.402 5.983-5.286-3.187z"></path></svg>Rate</button> 77Metascore <button aria-disabled="false" aria-label="See more information about Logan" class="ipc-icon-button dli-info-icon ipc-icon-button--base ipc-icon-button--onAccent2" role="button" tabindex="0" title="See more information about Logan"><svg class="ipc-icon ipc-icon--info" fill="currentColor" height="24" role="presentation" viewbox="0 0 24 24" width="24" xmlns="http://www.w3.org/2000/svg"><path d="M0 0h24v24H0V0z" fill="none"></path><path d="M11 7h2v2h-2zm0 4h2v6h-2zm1-9C6.48 2 2 6.48 2 12s4.48 10 10 10 10-4.48 10-10S17.52 2 12 2zm0 18c-4.41 0-8-3.59-8-8s3.59-8 8-8 8 3.59 8 8-3.59 8-8 8z"></path></svg></button>

In [18]:
```
first_movie.a
```

Out[18]:
```
<a aria-label="View title page for Logan" class="ipc-lockup-overlay ipc-focusable" href="/title/tt3315342/?ref_=sr_i_1"><div class="ipc-lockup-overlay__screen"></div></a>
```

In [19]:
```
first_movie.h3
```

Out[19]:
```
<h3 class="ipc-title__text">1. Logan</h3>
```

In [20]:
```
first_movie.h3.a
```

In [21]:
```
first_name = first_movie.find('h3',class_='ipc-title__text').text[3:]
first_name
```

Out[21]:
```
'Logan'
```

The year of the movie's release

In [22]:
```
first_year = first_movie.find('span',class_='sc-b0691f29-8 ilsLEX dli-title-metadat
first_year
```

Out[22]:
```
'2017'
```

The IMDB rating

In [23]:
```
first_movie.strong
```

In [28]:
```
first_imdb = first_movie.find('span',class_='ipc-rating-star ipc-rating-star--base
first_imdb
```

Out[28]: '8.1'

The Metascore

In [30]:
```
first_mscore = first_movie.find('span',class_='sc-b0901df4-0 bcQdDJ metacritic-scor
first_mscore
```

Out[30]: '77'

The number of votes

In [36]:
```
first_votes = first_movie.find('span', class_='ipc-rating-star--voteCount').text[1:
first_votes
```

Out[36]: '(827K)'

The script

In [64]:
```
#Lists to store the scraped data in
names = []
years = []
imdb_ratings = []
metascores = []
votes = []
# Extract data from individual movie container

for container in movie_containers:
# If the movie has a Metascore, then extract:
    if container.find('span',class_='sc-b0901df4-0 bcQdDJ metacritic-score-box') is

        name = container.find('h3',class_='ipc-title__text').text[3:]
        names.append(name)

        year = container.find('span', class_='sc-b0691f29-8 ilsLEX dli-title-metada
        years.append(year)

        imdb_rating = container.find('span',class_='ipc-rating-star ipc-rating-star
        imdb_ratings.append(imdb_rating)

        metascore = int(container.find('span', class_='sc-b0901df4-0 bcQdDJ metacri
        metascores.append(metascore)

        vote = container.find('span', class_='ipc-rating-star--voteCount').text[1:]
        votes.append(vote)
```

In [65]:
```
import pandas as pd
test_df = pd.DataFrame({'movie': names,
'year': years,
'imdb': imdb_ratings,
'metascore': metascores,
```

```
'votes': votes
})
print(test_df.info())
test_df
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41 entries, 0 to 40
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   movie      41 non-null     object
 1   year       41 non-null     object
 2   imdb       41 non-null     object
 3   metascore  41 non-null     int64 
 4   votes      41 non-null     object
dtypes: int64(1), object(4)
memory usage: 1.7+ KB
None
```

Out[65]:

	movie	year	imdb	metascore	votes
0	Logan	2017	8.1	77	(827K)
1	Thor: Ragnarok	2017	7.9	74	(813K)
2	Guardians of the Galaxy Vol. 2	2017	7.6	67	(756K)
3	Dunkirk	2017	7.8	94	(736K)
4	Spider-Man: Homecoming	2017	7.4	73	(716K)
5	Wonder Woman	2017	7.3	76	(698K)
6	Get Out	2017	7.8	85	(691K)
7	Star Wars: Episode VIII - The Last Jedi	2017	6.9	84	(670K)
8	Blade Runner 2049	2017	8.0	81	(658K)
9	Baby Driver	2017	7.5	86	(605K)
10	It	2017	7.3	69	(603K)
11	Coco	2017	8.4	81	(586K)
12	Three Billboards Outside Ebbing, Missouri	2017	8.1	88	(553K)
13	John Wick: Chapter 2	2017	7.4	75	(509K)
14	Justice League	2017	6.1	45	(477K)
15	The Shape of Water	2017	7.3	87	(446K)
16	Jumanji: Welcome to the Jungle	2017	6.9	58	(436K)
17	Kingsman: The Golden Circle	2017	6.7	44	(361K)
18	Kong: Skull Island	2017	6.7	62	(345K)
19	Pirates of the Caribbean: Salazar's Revenge	2017	6.5	39	(344K)
20	Beauty and the Beast	2017	7.1	65	(333K)
21	Lady Bird	2017	7.4	93	(326K)
22	Call Me by Your Name	2017	7.8	94	(313K)
23	The Greatest Showman	2017	7.5	48	(310K)
24	Alien: Covenant	2017	6.4	65	(302K)
25	Murder on the Orient Express	2017	6.5	52	(295K)
26	War for the Planet of the Apes	2017	7.4	82	(280K)
27	Wind River	2017	7.7	73	(279K)
28	Fast & Furious 8	2017	6.6	56	(253K)
29	Life	2017	6.6	54	(252K)

	movie	year	imdb	metascore	votes
30	Mother!	2017	6.6	76	(249K)
31	The Hitman's Bodyguard	2017	6.9	47	(246K)
32	I, Tonya	2017	7.5	77	(242K)
33	King Arthur: Legend of the Sword	2017	6.7	41	(232K)
34	Ghost in the Shell	2017	6.3	52	(227K)
35	Darkest Hour	2017	7.4	75	(220K)
36	American Made	2017	7.1	65	(207K)
37	Atomic Blonde	2017	6.7	63	(206K)
38	The Mummy	2017	5.4	34	(206K)
39	Baywatch	2017	5.5	37	(201K)
40	Bright	2017	6.3	29	(201K)

The script for multiple pages

In [94]:
```
from time import time
from time import sleep
from random import randint

from IPython.core.display import clear_output
pages = [ '1','2','3','4','5']
years_url = [ '2017', '2018', '2019', '2020']

# Redeclaring the lists to store data in
names = []
years = []
imdb_ratings = []
metascores = []
votes = []

# Preparing the monitoring of the loop
start_time = time()
requests = 0

#For every year in the interval 2000-2017
for year_url in years_url:

        # Make a get request
        url = f'https://www.imdb.com/search/title?release_date={year_url}-01-01,{ye
        agent = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKi
        response = get(url,headers = agent)
        print(response.text[:500])

        # Pause the loop
        sleep(randint(1,5))
```

```
        # Monitor the requests
        requests += 1
        elapsed_time = time() - start_time
        print('Request:{}; Frequency: {} requests/s'.format(requests, requests/elap
        clear_output(wait = True)

        # Throw a warning for non-200 status codes
        if response.status_code != 200:
            print('Request: {}; Status code: {}'.format(requests, response.status_c

        # Break the loop if the number of requests is greater than expected
        if requests > 72:
            print('Number of requests was greater than expected.')
            break
        # Parse the content of the request with BeautifulSoup
        page_html = BeautifulSoup(response.text, 'html.parser')

        # Select all the 50 movie containers from a single page
        mv_containers = page_html.find_all('div', class_ = 'sc-ab6fa25a-3 bVYfLY dl

        # For every movie of these 50
        for container in mv_containers:
            # If the movie has a Metascore, then:
            if container.find('span', class_ = 'sc-b0901df4-0 bcQdDJ metacritic-sco
                # Scrape the name
                name = container.find('h3',class_='ipc-title__text').text[3:]
                names.append(name)

                # Scrape the year
                year = container.find('span', class_ = 'sc-b0691f29-8 ilsLEX dli-ti
                years.append(year)

                # Scrape the IMDB rating
                imdb_rating = container.find('span', class_ = 'ipc-rating-star ipc-
                imdb_ratings.append(imdb_rating)

                # Scrape the Metascore
                metascore = container.find('span', class_ = 'sc-b0901df4-0 bcQdDJ m
                metascores.append(metascore)

                # Scrape the number of votes
                vote = container.find('span', class_ = 'ipc-rating-star--voteCount'
                votes.append(vote)
```

```
<!DOCTYPE html><html lang="en-US" xmlns:og="http://opengraphprotocol.org/schema/" xm
lns:fb="http://www.facebook.com/2008/fbml"><head><meta charSet="utf-8"/><meta name
="viewport" content="width=device-width"/><script>if(typeof uet === 'function'){ uet
('bb', 'LoadTitle', {wb: 1}); }</script><script>window.addEventListener('load', (eve
nt) => {
        if (typeof window.csa !== 'undefined' && typeof window.csa === 'function') {
            var csaLatencyPlugin = window.csa('Content', {

Request:4; Frequency: 0.1575987123833381 requests/s
```

In [95]:
```
movie_ratings = pd.DataFrame({'movie': names,
'year': years,
```

```
'imdb': imdb_ratings,
'metascore': metascores,
'votes': votes
})
print(movie_ratings.info())
movie_ratings.head(10)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 157 entries, 0 to 156
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   movie      157 non-null    object
 1   year       157 non-null    object
 2   imdb       157 non-null    object
 3   metascore  157 non-null    object
 4   votes      157 non-null    object
dtypes: object(5)
memory usage: 6.3+ KB
None
```

Out[95]:

	movie	year	imdb	metascore	votes
0	Logan	2017	8.1	77	(827K)
1	Thor: Ragnarok	2017	7.9	74	(813K)
2	Guardians of the Galaxy Vol. 2	2017	7.6	67	(756K)
3	Dunkirk	2017	7.8	94	(736K)
4	Spider-Man: Homecoming	2017	7.4	73	(716K)
5	Wonder Woman	2017	7.3	76	(698K)
6	Get Out	2017	7.8	85	(691K)
7	Star Wars: Episode VIII - The Last Jedi	2017	6.9	84	(670K)
8	Blade Runner 2049	2017	8.0	81	(658K)
9	Baby Driver	2017	7.5	86	(605K)

In [96]:
```
movie_ratings.tail(10)
```

Out[96]:

	movie	year	imdb	metascore	votes
147	The Hunt	2020	6.5	50	(128K)
148	Greyhound	2020	7.0	64	(114K)
149	Hamilton	2020	8.3	88	(112K)
150	Eurovision Song Contest: The Story of Fire Saga	2020	6.5	50	(102K)
151	I'm Thinking of Ending Things	2020	6.6	78	(99K)
152	Project Power	2020	6.0	51	(97K)
153	Spenser Confidential	2020	6.2	49	(97K)
154	Underwater	2020	5.9	48	(97K)
155	Minari	2020	7.4	89	(96K)
156	News of the World	2020	6.8	73	(95K)

In [97]:
```
movie_ratings.to_csv('movie_ratings.csv')
```

Data Preparation

Data preprocessing

Data Processing is a process of cleaning the raw data i.e. the data is collected in the real world and is converted to a clean data set. In other words, whenever the data is gathered from different sources it is collected in a raw format and this data isn't feasible for the analysis. Therefore, certain steps are executed to convert the data into a small clean data set, this part of the process is called as data preprocessing.

Most of the real-world data is messy, some of these types of data are: 1. Missing data: Missing data can be found when it is not continuously created or due to technical issues in the application (IOT system). 2. Noisy Data This type of data is also called outliners, this can occur due to human errors (human manually gathering the data) or some technical problem of the device at the time of collection of data. 3. Inconsistent data: This type of data might be collected due to human errors (mistakes with the name or values) or duplication of data.

These are some of the basic pre processing techniques that can be used to convert raw data. 1. Conversion of data: As we know that Machine Learning models can only handle numeric features, hence categorical and ordinal data must be somehow converted into numeric features. 2. Ignoring the missing values: Whenever we encounter missing data in the data set then we can remove the row or column of data depending on our need. This method is known to be efficient but it shouldn't be performed if there are a lot of missing values in the dataset. 3. Filling the missing values: Whenever we encounter missing data in the data set then we can fill the missing data manually, most commonly the mean, median or highest frequency value is used.

Example of Data Preparation of movie_rating.csv

In [98]:
```
movie_ratings = pd.read_csv('movie_ratings.csv')
```

In [99]:
```
movie_ratings['year'].unique()
```

Out[99]:
```
array([2017, 2018, 2019, 2020], dtype=int64)
```

In [100…
```
movie_ratings.dtypes
```

Out[100…
```
Unnamed: 0      int64
movie          object
year            int64
imdb          float64
metascore       int64
votes          object
dtype: object
```

In [101…
```
movie_ratings['year'] = movie_ratings['year'].astype(int)
```

In [102…
```
movie_ratings['year'].unique()
```

Out[102…
```
array([2017, 2018, 2019, 2020])
```

In [103…
```
movie_ratings.dtypes
```

Out[103…
```
Unnamed: 0      int64
movie          object
year            int32
imdb          float64
metascore       int64
votes          object
dtype: object
```

In [104…
```
movie_ratings.head(10)
```

Out[104…

	Unnamed: 0	movie	year	imdb	metascore	votes
0	0	Logan	2017	8.1	77	(827K)
1	1	Thor: Ragnarok	2017	7.9	74	(813K)
2	2	Guardians of the Galaxy Vol. 2	2017	7.6	67	(756K)
3	3	Dunkirk	2017	7.8	94	(736K)
4	4	Spider-Man: Homecoming	2017	7.4	73	(716K)
5	5	Wonder Woman	2017	7.3	76	(698K)
6	6	Get Out	2017	7.8	85	(691K)
7	7	Star Wars: Episode VIII - The Last Jedi	2017	6.9	84	(670K)
8	8	Blade Runner 2049	2017	8.0	81	(658K)
9	9	Baby Driver	2017	7.5	86	(605K)

In [105…

```
movie_ratings.tail(10)
```

Out[105…

	Unnamed: 0	movie	year	imdb	metascore	votes
147	147	The Hunt	2020	6.5	50	(128K)
148	148	Greyhound	2020	7.0	64	(114K)
149	149	Hamilton	2020	8.3	88	(112K)
150	150	Eurovision Song Contest: The Story of Fire Saga	2020	6.5	50	(102K)
151	151	I'm Thinking of Ending Things	2020	6.6	78	(99K)
152	152	Project Power	2020	6.0	51	(97K)
153	153	Spenser Confidential	2020	6.2	49	(97K)
154	154	Underwater	2020	5.9	48	(97K)
155	155	Minari	2020	7.4	89	(96K)
156	156	News of the World	2020	6.8	73	(95K)

In [106…

```
movie_ratings
```

Out[106...

	Unnamed: 0	movie	year	imdb	metascore	votes
0	0	Logan	2017	8.1	77	(827K)
1	1	Thor: Ragnarok	2017	7.9	74	(813K)
2	2	Guardians of the Galaxy Vol. 2	2017	7.6	67	(756K)
3	3	Dunkirk	2017	7.8	94	(736K)
4	4	Spider-Man: Homecoming	2017	7.4	73	(716K)
...	...	...	...	...	...	...
152	152	Project Power	2020	6.0	51	(97K)
153	153	Spenser Confidential	2020	6.2	49	(97K)
154	154	Underwater	2020	5.9	48	(97K)
155	155	Minari	2020	7.4	89	(96K)
156	156	News of the World	2020	6.8	73	(95K)

157 rows × 6 columns