

```
In [ ]: 1
        2 #importing necessary libraries
        3
        4 import pandas as pd
        5 import numpy as np
        6 from selenium import webdriver
        7 from bs4 import BeautifulSoup
        8 from selenium.webdriver.chrome.service import Service
        9 from selenium.webdriver.common.by import By
       10 from selenium.webdriver.common.keys import Keys
       11 from warnings import warn
       12 import time
       13 from selenium.webdriver.support.ui import WebDriverWait
       14 from selenium.webdriver.support import expected_conditions as EC
       15
```

```
In [ ]: #passing required URL for scrapping
        2
        url1='https://www.linkedin.com/jobs/collections/?currentJobId=3378642105'
```

```
In [ ]: 1 driver= webdriver.Chrome(executable_path=r'C:\Users\Administrator\Downloads\chromedriver.exe')
        2 # Maximize Window
        3 driver.maximize_window()
        4 driver.minimize_window()
        5 driver.maximize_window()
        6 driver.switch_to.window(driver.current_window_handle)
        7 driver.implicitly_wait(10)
        8
        9 # Enter to the site
       10 driver.get('https://www.linkedin.com/login');
       11 time.sleep(2)
       12
```

```
In [ ]: 1 #Logging in using keys
        2
        3
        4 # User Credentials
        5 # Reading txt file where we have our user credentials
        6 with open('user_credentials.txt', 'r',encoding="utf-8") as file:
        7     user_credentials = file.readlines()
        8     user_credentials = [line.rstrip() for line in user_credentials]
        9
       10 user_name = lines[0] # First Line
       11 password = lines[1] # Second Line
       12 driver.find_element_by_xpath('//*[@id="username"]').send_keys('aingh.jatin413@gmail.com')
       13 driver.find_element_by_xpath('//*[@id="password"]').send_keys('jetsingh@413')
       14 time.sleep(1)
       15
       16 # Login button
       17 driver.find_element_by_xpath('//*[@id="organic-div"]/form/div[3]/button').click()
       18 driver.implicitly_wait(30)
```

```
In [ ]: 1 #redirecting to desired URL
        2
        3 driver.get("https://www.linkedin.com/jobs/collections/")
```



```

In [ ]: #iterating through page

for i in range(1,41):
    #button path for page numbers
    path = '//button[@aria-label="Page {}"]'.format(i)

    #button clicking
    driver.find_element(By.XPATH, path).click()

    #html data
    src = driver.page_source
    soup = BeautifulSoup(src, 'lxml')

    #main page of one job data
    lk=soup.findAll(class_="disabled ember-view job-card-container__link")

    #link of a single job data
    for i in lk:
        # links
        li=i['href']

        #every page data
        every_page =BeautifulSoup(driver.page_source,'lxml')

        #movig to link using next window_of_ chrome -- alternative of redirecting to original URL
        window_before = driver.window_handles[0]

        window_after = driver.window_handles[0]
        driver.switch_to.window(window_after)

        job_link.append("https://www.linkedin.com{}".format(li))
        driver.get("https://www.linkedin.com{}".format(li))

        # company name
        try:
            c_name = driver.find_elements(By.CLASS_NAME,'jobs-unified-top-card__company-name')
            name.append(c_name[0].text)
        except:
            name.append("N.A.")

        #designation
        try:
            d = driver.find_elements(By.CLASS_NAME,'jobs-unified-top-card__job-title')
            designation.append(d[0].text)
        except:
            designation.append("N.A.")

        #applicants
        try:
            apl= driver.find_elements(By.XPATH, '/html/body/div[5]/div[3]/div/div[1]/div[1]/div/div[1]/div/div/div[1]/div[1]/span[2]/span[2]/span')
            applicants.append(apl[0].text)
        except:
            applicants.append("0")

        #work type
        try:
            w = driver.find_elements(By.CLASS_NAME,'jobs-unified-top-card__workplace-type')
            work_type.append(w[0].text)
        except:
            work_type.append("N.A.")

        #involvement
        try:
            inv = driver.find_elements(By.CLASS_NAME,'jobs-unified-top-card__job-insight')
            involvement.append(inv[0].text)

```

```

except:
    involvement.append("N.A.")

#employee count
try:
    emp = driver.find_elements(By.CLASS_NAME, 'jobs-unified-top-card__job-insight')
    emp_count.append(emp[1].text)
except:
    emp_count.append("N.A.")

#Location
try:
    loc = driver.find_elements(By.CLASS_NAME, 'jobs-unified-top-card__bullet')
    location.append(loc[0].text)
except:
    location.append("N.A.")

#every page data
every_page =BeautifulSoup(driver.page_source, 'lxml')

# details
s = []
src = driver.page_source
soup = BeautifulSoup(src, 'lxml')
detail = soup.findAll(class_='ember-view t-black t-normal')
for z in detail:
    s.append(z)

# selecting new jobs
for i in s:
    pr = i['href']

    #movig to link using next window_of_ chrome
    window_before = driver.window_handles[0]

    window_after = driver.window_handles[0]
    driver.switch_to.window(window_after)
    driver.get("https://www.linkedin.com{}".format(pr))

    time.sleep(6)

    #industry
    try:
        ind = driver.find_elements(By.CLASS_NAME, 'org-top-card-summary-info-list__info-item')
        industry.append(ind[0].text)
    except:
        industry.append("not specify")

    #followers
    try:
        follow = driver.find_elements(By.XPATH, '//*[@id="ember28"]/div[2]/div[1]/div[1]/div[2]/div/div/div[2]/div[2]')
        linkedin_followers.append(follow[0].text)
    except:
        linkedin_followers.append("N/A")

#close current window
driver.close()

#switch to main(starting) tab/window
driver.switch_to.window(driver.window_handles[-1])

# close current window
driver.close()

#switch to main (starting) tab/window
driver.switch_to.window(driver.window_handles[0])

```

