PROJECT=WEB SCRAPING – ASSIGNMENT 2

NAME=Neha Somvanshi

BATCH=DS2307

In [87]:
```
import selenium
from selenium import webdriver
import pandas as pd
from selenium.webdriver.common.by import By
import warnings
warnings.filterwarnings("ignore")
import time
```

In [26]:
```
driver = webdriver.Chrome()
```

Q1: Write a python program to scrape data for “Data Analyst” Job position in “Bangalore” location. You have to scrape the job-title, job-location, company_name, experience_required. You have to scrape first 10 jobs data.

In [27]:
```
#First get the webpage https://www.shine.com/
driver.get('https://www.shine.com/')
```

In [28]:
```
#2. Enter “Data Analyst” in “Job title
designation=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2
designation.send_keys('Data Analyst')
```

In [32]:
```
#3. enter “Bangalore” in “enter the location” field
Location=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2]/d
Location.send_keys('Banglore')
```

In [33]:
```
# clicked on search button
submit=driver.find_element(By.XPATH,'/html/body/div[1]/div[4]/div/div[2]/div[2]/div
submit.click()
```

In [36]:
```
#scraped job location
job=[]
job_location=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_it
for x in job_location:
    job.append(x.text)

#scraped company name
```

```
company_name=[]
name=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName__mYnow"]')
for x in name:
    company_name.append(x.text)

# scraped experience
experience_name=[]
experience=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_iter
for x in experience:
    experience_name.append(x.text)
```

In [44]:
```
#finally created a dataframe
df=pd.DataFrame({'job':job,'company_name':company_name,'experience_name':experienc
df[:10]
```

Out[44]:

	job	company_name	experience_name
0	Bangalore	futures and careers	2 to 4 Yrs
1	Bangalore	boyen haddin consulting and technol...	3 to 6 Yrs
2	Singapore\n+14	kavya staffing solutions	0 to 4 Yrs
3	Bangalore	ara resources private limited	2 to 5 Yrs
4	Gurugram\n+9	ashutosh sabhashankar chaturvedi hi...	7 to 12 Yrs
5	Oman\n+14	divya interprises	0 to 4 Yrs
6	Bangalore	deuglo infosystem private limited	1 to 2 Yrs
7	Bangalore	deuglo infosystem private limited	1 to 2 Yrs
8	Bangalore	deuglo infosystem private limited	1 to 2 Yrs
9	Bangalore	deuglo infosystem private limited	1 to 2 Yrs

Q2:Write a python program to scrape data for “Data Scientist” Job position in“Bangalore” location. You have to scrape the job-title, job-location, company_name. You have to scrape first 10 jobs data. This task will be done in following steps:

In [42]:
```
driver= webdriver.Chrome()
```

In [43]:
```
driver.get('https://www.shine.com/')
```

In [45]:
```
designation=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[
designation.send_keys('Data Scientist')
```

In [46]:
```
location=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2]/
location.send_keys('Bangalore')
```

In [47]:
```
submit=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2]/di
submit.click()
```

In [55]:
```
job_location=[]
location=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__

for x in location:
    job_location.append(x.text)


company_name=[]
company=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName__mYnow"
for x in company:
    company_name.append(x.text)
```

In [57]:
```
df=pd.DataFrame({'company_name':company_name,'job_location':job_location})
df
```

Out[57]:

	company_name	job_location
0	kavya staffing solutions	Bangalore\n+17
1	kavya staffing solutions	Bangalore\n+17
2	skyleaf consultants	Bangalore
3	gfl recruitment private limited	Bangalore\n+7
4	divya interprises	Bangalore\n+14
5	deuglo infosystem private limited	Bangalore
6	deuglo infosystem private limited	Bangalore\n+8
7	seven geomax consulting private lim...	Bangalore
8	employberry consultants hiring for ...	Bangalore
9	niharika enterprises	Bangalore\n+15

Q3: In this question you have to scrape data using the filters available on the webpage

In [58]:
```
driver=webdriver.Chrome()
```

In [60]:
```
driver.get('https://www.shine.com/')
```

In [64]:
```
designation=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[
designation.send_keys('Data Scientist')
```

In [65]:
```
location=driver.find_element(By.XPATH,'/html/body/div[1]/div[4]/div/div[2]/div[2]/
location.send_keys('Delhi')
```

In [66]:
```
submit=driver.find_element(By.XPATH,'/html/body/div[1]/div[4]/div/div[2]/div[2]/di
submit.click()
```

In [67]:
```
salary=driver.find_element(By.XPATH,'/html/body/div[1]/div[1]/div[4]/div/div[1]/div
salary.click()
```

In [68]:
```
salary_criteria=driver.find_element(By.XPATH,'/html/body/div[1]/div[1]/div[4]/div/
salary_criteria.click()
```

In [69]:
```
show_result=driver.find_element(By.XPATH,'/html/body/div[1]/div[1]/div[4]/div/div[1
show_result.click()
```

In [70]:
```
company_name=[]
name=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName__mYnow"]')
for x in name:
    company_name.append(x.text)
```

In [72]:
```
job_location=[]
location=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__
for x in location:
    job_location.append(x.text)
```

In [73]:
```
experience_required=[]
experience=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_iter
for x in experience:
    experience_required.append(x.text)
```

In [79]:
```
df=pd.DataFrame({'company_name':company_name,'job_location':job_location,'experien
df.tail()
```

Out[79]:

	company_name	job_location	experience_required
15	gharondaa advisors private limited	Gurugram	2 to 5 Yrs
16	gharondaa advisors private limited	Noida	2 to 5 Yrs
17	js placement services	Ghaziabad	0 to 3 Yrs
18	js placement services	Faridabad	0 to 3 Yrs
19	js placement services	Noida	0 to 3 Yrs

Q4: Scrape data of first 100 sunglasses listings on flipkart.com. You have to scrape four attributes:

In [80]:
```
driver=webdriver.Chrome()
```

In [81]:
```
driver.get('https://www.flipkart.com/')
```

In [82]:
```
sunglasses=driver.find_element(By.XPATH,'/html/body/div[1]/div/div[1]/div[1]/div[2
sunglasses.send_keys('sunglasses')
```

In [85]:
```
submit=driver.find_element(By.XPATH,'/html/body/div[1]/div/div[1]/div[1]/div[2]/div
submit.click()
```

In [105…
```
brand_name=[]
item_price=[]
start=0
```

```
end=3
for page in range (start,end):

    brand=driver.find_elements(By.XPATH,'//div[@class="_2WkVRV"]')
    for x in brand:
        brand_name.append(x.text)

    price=driver.find_elements(By.XPATH,'//div[@class="_30jeq3"]')
    for x in price:
        item_price.append(x.text)


    next_page=driver.find_element(By.XPATH,'/html/body/div/div/div[3]/div[1]/div[2
    next_page.click()
    time.sleep(3)
```

In [108…

```
df=pd.DataFrame({'brand_name':brand_name,'item_price':item_price})
df[0:100]
```

Out[108]:

	brand_name	item_price
0	LOUIS KOUROS	₹426
1	LOUIS KOUROS	₹410
2	Fastrack	₹968
3	Ray-Ban	₹3,731
4	Ray-Ban	₹11,891
...	...	...
95	PETER JONES	₹547
96	Rich Club	₹235
97	VINCENT CHASE	₹999
98	QUICKAND FAST	₹149
99	VINCENT CHASE	₹1,199

100 rows × 2 columns

Q5: Scrape 100 reviews data from flipkart.com for iphone11 phone. You have to go the link:

In [109…

```
driver=webdriver.Chrome()
```

In [110…

```
driver.get('https://www.flipkart.com/apple-iphone-11-black-64-gb/product-reviews/it
```

In [119…

```
Rating_product=[]
Review_summary=[]
Full_review=[]

start=0
end=3
for page in range(start,end):
```

```
    Rating=driver.find_elements(By.XPATH,'//div[@class="_3LWZlK _1BLPMq"]')
    for x in Rating:
        Rating_product.append(x.text)

    Review=driver.find_elements(By.XPATH,'//div[@class="t-ZTKy"]')
    for x in Review:
        Review_summary.append(x.text)

    Full=driver.find_elements(By.XPATH,'//p[@class="_2-N8zT"]')
    for x in Full:
        Full_review.append(x.text)

    next_page=driver.find_element(By.XPATH,'/html/body/div/div/div[3]/div/div/div[
    next_page.click()
    time.sleep(3)
```

In [121…

```
df=pd.DataFrame({'Rating_product':Rating_product,'Review_summary':Review_summary,'
df
```

Out[121]:

	Rating_product	Review_summary	Full_review
0	5	Value for money 😍	Terrific purchase
1	5	Photos super	Classy product
2	5	Very very good	Terrific
3	5	Camera is awesome\nBest battery backup\nA perf...	Classy product
4	5	This is amazing at all	Wonderful
5	5	Perfect Product!!	Just wow!
6	5	Feeling awesome after getting the delivery of ...	Worth every penny
7	5	V Good all	Perfect product!
8	5	Good Camera	Best in the market!
9	5	Super 🔥 and good performance 👌 ❤️	Fabulous!
10	5	It's really awesome	Must buy!
11	5	Purple is best	Great product
12	4	Camera is just wow 👌 👌	Worth the money
13	5	Go for iPhone 11 , if confused between iPhone ...	Must buy!
14	5	Excellent Phone.	Brilliant
15	5	Value for money 🖤 🖤	Terrific purchase
16	5	very good camera quality	Brilliant
17	5	It's very good battery life and display and vi...	Fabulous!
18	5	NYC	Excellent
19	5	Damn this phone is a blast . Upgraded from and...	Best in the market!
20	5	Such an awesome experience with iPhone 11 awes...	Best in the market!
21	5	Awesome Phone. Battery backup top-notch...	Awesome
22	5	It is better to buy iPhone 11 over iPhone 12 i...	Worth every penny
23	5	Excellent Fabulous Adorable Iphone 11 Value fo...	Wonderful
24	5	happy ❤️	Must buy!
25	5	Best phone	Brilliant
26	5	Perfect iPhone on this budget!! Camera and the...	Brilliant
27	5	Battery backup is extraordinary, camera is dec...	Perfect product!
28	5	iPhone is delivered on time. Display is great ...	Worth every penny
29	5	Outstanding performance this phone	Classy product

Q6: Scrape data forfirst 100 sneakers you find whenyou visit flipkart.com and search for "sneakers" inthe search field.

In [43]:
```
driver=webdriver.Chrome()
```

In [44]:
```
driver.get('https://www.flipkart.com/')
```

In [45]:
```
sneakers=driver.find_element(By.XPATH,'/html/body/div[1]/div/div[1]/div/div/div/di
sneakers.send_keys('sneakers')
```

In [46]:
```
submit=driver.find_element(By.XPATH,'/html/body/div[1]/div/div[1]/div/div/div/div/
submit.click()
```

In [48]:
```
brand_name=[]
price_product=[]

start=0
end=3
for page in range(start,end):

    brand=driver.find_elements(By.XPATH,'//div[@class="_2WkVRV"]')
    for x in brand:
        brand_name.append(x.text)

    price=driver.find_elements(By.XPATH,'//div[@class="_30jeq3"]')
    for x in price:
        price_product.append(x.text)

    next_page=driver.find_element(By.XPATH,'/html/body/div/div/div[3]/div[1]/div[2
    next_page.click()
    time.sleep(3)
```

In [50]:
```
df=pd.DataFrame({'brand_name':brand_name,'price_product':price_product})
df[0:100]
```

Out[50]:

	brand_name	price_product
0	Elevarse	₹279
1	Elevarse	₹279
2	asian	₹499
3	BRUTON	₹379
4	Nobelite	₹299
...	...	...
95	YUNIKO	₹749
96	New Balance	₹4,679
97	Shozie	₹549
98	BIG FOX	₹799
99	SWIGGY	₹757

100 rows × 2 columns

Q7: Go to webpage https://www.amazon.in/ Enter “Laptop” in

the search field and then click the search icon. Thenset CPU Type filter to “Intel Core i7” as shown in the below image:

In [51]:
```
driver=webdriver.Chrome()
```

In [52]:
```
driver.get('https://www.amazon.in/')
```

In [53]:
```
laptop=driver.find_element(By.XPATH,'/html/body/div[1]/header/div/div[1]/div[2]/div
laptop.send_keys('laptop')
```

In [54]:
```
submit=driver.find_element(By.XPATH,'/html/body/div[1]/header/div/div[1]/div[2]/div
submit.click()
```

In [58]:
```
filters=driver.find_element(By.XPATH,'/html/body/div[1]/div[2]/div[1]/div[2]/div/d
filters.click()
```

In [65]:
```
title_product=[]
title=driver.find_elements(By.XPATH,'//h2[@class="a-size-mini a-spacing-none a-colo
for x in title:
    title_product.append(x.text)
```

In [66]:
```
price_product=[]
price=driver.find_elements(By.XPATH,'//span[@class="a-price-whole"]')[0:10]
for x in price:
    price_product.append(x.text)
```

In [67]:
```
rating_product=[]
rating=driver.find_elements(By.XPATH,'//span[@class="a-size-base puis-bold-weight-
for x in rating:
    rating_product.append(x.text)
```

Q8:Write a python program to scrape data for Top 1000 Quotes of All Time

In [88]:
```
driver=webdriver.Chrome()
```

In [89]:
```
driver.get('https://www.azquotes.com/')
```

In [90]:
```
top_quotes=driver.find_element(By.XPATH,'/html/body/div[1]/div[1]/div[1]/div/div[3
top_quotes.click()
```

In [94]:
```
    quotes_types=[]
    types=driver.find_elements(By.XPATH,'//div[@class="tags"]')
    for x in types:
        quotes_types.append(x.text)

    top_quotes=[]
    quotes=driver.find_elements(By.XPATH,'//a[@class="title"]')
    for x in quotes:
        top_quotes.append(x.text)

    Author_list=[]
    types=driver.find_elements(By.XPATH,'//div[@class="author"]')
```

```
    for x in types:
        Author_list.append(x.text)
```

In [95]:
```
df=pd.DataFrame({'quotes_types':quotes_types,'top_quotes':top_quotes,'Author_list'
df
```

Out[95]:

	quotes_types	top_quotes	Author_list
0	Spring, April, Fragrance	Can words describe the fragrance of the very b...	Neltje Blanchan
1	Inspirational, Faith, Spiritual	Faith is to believe what you do not see; the r...	Saint Augustine
2	Inspirational, Motivational, Positive	When everything seems to be going against you,...	Henry Ford
3	Love, Inspirational, Life	I have found that if you love life, life will ...	Arthur Rubinstein
4	Strength, Peace, Gun	To disarm the people... was the best and most ...	George Mason
...	...	...	...
95	Love, Inspirational, Life	When one door closes, another opens; but we of...	Alexander Graham Bell
96	Inspirational, Motivational, Positive	Don't find fault, find a remedy.	Henry Ford
97	Love, Life, Lonely	I used to think the worst thing in life was to...	Robin Williams
98	Love, Inspirational, Friendship	Friends and good manners will carry you where ...	Margaret Walker
99	Inspirational, Change, Inspiring	If you want to make a permanent change, stop f...	T. Harv Eker

100 rows × 3 columns

Q9.Data cannot be scraped using class name and tag name

Q10:Write a python program to display list of 50 Most expensive cars in the world (i.e.Car name and Price) from https://www.motor1.com/

In [96]:
```
driver=webdriver.Chrome()
```

In [97]:
```
driver.get('https://www.motor1.com/')
```

In [99]:
```
car_name=driver.find_element(By.XPATH,'/html/body/div[10]/div[2]/div/div/div[3]/di
car_name.send_keys('50 most expensive cars')
```

In [100…
```
submit=driver.find_element(By.XPATH,'/html/body/div[10]/div[2]/div/div/div[3]/div/
submit.click()
```

In [101…
```
car_name=[]
car=driver.find_elements(By.XPATH,'//h3[@class="subheader"]')
for x in car:
    car_name.append(x.text)
```

In [102…
```
df=pd.DataFrame({'car_name':car_name})
df
```

Out[102]:

	car_name
0	Aston Martin Valour
1	McLaren Elva
2	Czinger 21C
3	Ferrari Monza
4	Gordon Murray T.33
5	Koenigsegg Gemera
6	Zenvo TSR-S
7	Hennessey Venom F5
8	Bentley Bacalar
9	Hispano Suiza Carmen Boulogne
10	Bentley Mulliner Batur
11	Deus Vayanne
12	SSC Tuatara
13	Lotus Evija
14	Aston Martin Vulcan
15	Delage D12
16	Ferrari Daytona SP3
17	McLaren Speedtail
18	Rimac Nevera
19	Pagani Utopia
20	Pininfarina Battista
21	Gordon Murray T.50
22	Lamborghini Countach
23	Mercedes-AMG Project One
24	Zenvo Aurora
25	Aston Martin Victor
26	Hennessey Venom F5 Roadster
27	Koenigsegg Jesko
28	Aston Martin Valkyrie
29	W Motors Lykan Hypersport
30	McLaren Solus
31	Lamborghini Sian
32	Koenigsegg CC850
33	Bugatti Chiron Super Sport 300+
34	Lamborghini Veneno
35	Bugatti Bolide

	car_name
36	Pininfarina B95 Speedster
37	Bugatti Mistral
38	Pagani Huayra Imola
39	Bugatti Divo
40	SP Automotive Chaos
41	Pagani Codalunga
42	777 Hypercar
43	Mercedes-Maybach Exelero
44	Bugatti Centodieci
45	Bugatti Chiron Profilée
46	Rolls-Royce Sweptail
47	Bugatti La Voiture Noire
48	Rolls-Royce Boat Tail*
49	Rolls-Royce La Rose Noire Droptail
50	Most Expensive Cars In The World

----------END----------

In [ ]: