In [1]:
```
#[Q1]Write a python program to scrape data for “Data Analyst” Job position in “Bangalore” location.
!pip install selenium
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: selenium in c:\users\admin\appdata\roaming\python\python39\site-packages (4.11.2
)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in c:\programdata\anaconda3\lib\site-packages (from sele
nium) (1.26.11)
Requirement already satisfied: certifi>=2021.10.8 in c:\programdata\anaconda3\lib\site-packages (from selenium)
(2022.9.14)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\admin\appdata\roaming\python\python39\site-packa
ges (from selenium) (0.10.3)
Requirement already satisfied: trio~=0.17 in c:\users\admin\appdata\roaming\python\python39\site-packages (from
selenium) (0.22.2)
Requirement already satisfied: exceptiongroup>=1.0.0rc9 in c:\users\admin\appdata\roaming\python\python39\site-
packages (from trio~=0.17->selenium) (1.1.3)
Requirement already satisfied: sniffio in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium
) (1.2.0)
Requirement already satisfied: cffi>=1.14 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selen
ium) (1.15.1)
Requirement already satisfied: outcome in c:\users\admin\appdata\roaming\python\python39\site-packages (from tr
io~=0.17->selenium) (1.2.0)
Requirement already satisfied: attrs>=20.1.0 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->se
lenium) (21.4.0)
Requirement already satisfied: idna in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (
3.3)
Requirement already satisfied: sortedcontainers in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17-
>selenium) (2.4.0)
Requirement already satisfied: wsproto>=0.14 in c:\users\admin\appdata\roaming\python\python39\site-packages (f
rom trio-websocket~=0.9->selenium) (1.2.0)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\programdata\anaconda3\lib\site-packages (from
urllib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: pycparser in c:\programdata\anaconda3\lib\site-packages (from cffi>=1.14->trio~=
0.17->selenium) (2.21)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\admin\appdata\roaming\python\python39\site-packages (f
rom wsproto>=0.14->trio-websocket~=0.9->selenium) (0.14.0)
```

In [1]:
```
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 [3]:
```
driver =webdriver.Chrome()
```

In [4]:
```
driver.get("https://www.shine.com/")
```

In [5]:
```
designation =driver.find_element(By.CLASS_NAME,"form-control")
designation.send_keys('Data Analyst')
```

In [6]:
```
location=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2]/div/form/div/div[1]/ul/li[2]/
location.send_keys('Banglore')
```

In [7]:
```
search =driver.find_element(By.CLASS_NAME,"searchForm_btnWrap_advance__VYBHN")
search.click()
```

In [8]:
```
company_name=[]
job_location=[]
experience_required=[]
```

In [9]:
```
company_tags=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName__mYnow"]')
for i in company_tags[0:10]:
    company=i.text
    company_name.append(company)

location_tags=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__YxRkV jobCard_locationI
for i in location_tags[0:10]:
    location=i.text
    job_location.append(location)

experience_tags=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__YxRkV jobCard_jobIcon
for i in experience_tags[0:10]:
    exp=i.text
    experience_required.append(exp)
```

In [10]:
```
print(len(company_name),len(job_location),len(experience_required))
```

```
10 10 10
```

In [11]:
```
import pandas as pd
df=pd.DataFrame({'Company_name':company_name,'Location':job_location,'Experience':experience_required})
df
```

Out[11]:

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

In [12]:
```
#[Q2]Write a python program to scrape data for “Data Scientist” Job position in “Bangalore” location.
```

In [13]:
```
driver =webdriver.Chrome()
```

In [14]:
```
driver.get("https://www.shine.com/")
```

In [15]:
```
designation =driver.find_element(By.CLASS_NAME,"form-control")
designation.send_keys('Data Scientist')
```

In [16]:
```
location=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2]/div/form/div/div[1]/ul/li[2]/
location.send_keys('Banglore')
```

In [17]:
```
search =driver.find_element(By.CLASS_NAME,"searchForm_btnWrap_advance__VYBHN")
search.click()
```

In [18]:
```
company_name=[]
job_location=[]
experience_required=[]
```

In [19]:
```
company_tags=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName__mYnow"]')
for i in company_tags[0:10]:
    company=i.text
    company_name.append(company)

location_tags=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__YxRkV jobCard_locationI
for i in location_tags[0:10]:
    location=i.text
    job_location.append(location)

experience_tags=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__YxRkV jobCard_jobIcon
for i in experience_tags[0:10]:
    exp=i.text
    experience_required.append(exp)
```

In [20]:
```
print(len(company_name),len(job_location),len(experience_required))
```

```
10 10 10
```

In [21]:
```
import pandas as pd
df=pd.DataFrame({'Company_name':company_name,'Location':job_location,'Experience':experience_required})
df
```

Out[21]:

	Company_name	Location	Experience
0	kavya staffing solutions	Bangalore\n+17	0 to 4 Yrs
1	kavya staffing solutions	Bangalore\n+17	0 to 4 Yrs
2	skyleaf consultants	Bangalore	5 to 10 Yrs
3	divya interprises	Bangalore\n+14	0 to 4 Yrs
4	deuglo infosystem private limited	Bangalore	4 to 6 Yrs
5	deuglo infosystem private limited	Bangalore\n+8	4 to 6 Yrs
6	seven geomax consulting private lim...	Bangalore	6 to 9 Yrs
7	employberry consultants hiring for ...	Bangalore	3 to 6 Yrs
8	niharika enterprises	Bangalore\n+15	0 to 4 Yrs
9	deuglo infosystem private limited	Bangalore\n+8	4 to 6 Yrs

In [22]:
```
#Q3Write a python program to scrape data for “Data Scientist” Job position in “Delhi-NCR region” location.
```

In [23]:
```
driver =webdriver.Chrome()
```

In [24]:
```
driver.get("https://www.shine.com/")
```

In [25]:
```
designation =driver.find_element(By.CLASS_NAME,"form-control")
designation.send_keys('Data Scientist')
```

In [26]:
```
location=driver.find_element(By.XPATH,"/html/body/div[1]/div[4]/div/div[2]/div[2]/div/form/div/div[1]/ul/li[2]/
location.send_keys('Delhi-NCR Region')
```

In [27]:
```
salary=driver.find_element(By.CLASS_NAME,"form-control")
salary.send_keys('3-6 lakhs')
```

In [29]:
```
search =driver.find_element(By.CLASS_NAME,"searchForm_btnWrap_advance__VYBHN")
search.click()
```

In [30]:
```
company_name=[]
job_location=[]
experience_required=[]
```

In [31]:
```
company_tags=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName__mYnow"]')
for i in company_tags[0:10]:
    company=i.text
    company_name.append(company)

location_tags=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__YxRkV jobCard_locationI
for i in location_tags[0:10]:
    location=i.text
    job_location.append(location)

experience_tags=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_lists_item__YxRkV jobCard_jobIcon
for i in experience_tags[0:10]:
    exp=i.text
    experience_required.append(exp)
```

In [32]:
```
print(len(company_name),len(job_location),len(experience_required))
```

```
10 10 10
```

In [33]:
```
import pandas as pd
df=pd.DataFrame({'Company_name':company_name,'Location':job_location,'Experience':experience_required})
df
```

Out[33]:

	Company_name	Location	Experience
0	boyen haddin consulting and technol...	Delhi\n+4	5 to 9 Yrs
1	boyen haddin consulting and technol...	Delhi\n+4	5 to 9 Yrs
2	deuglo infosystem private limited	Delhi\n+8	4 to 6 Yrs
3	deuglo infosystem private limited	Delhi\n+8	4 to 6 Yrs
4	quiscon biotech	Delhi\n+6	0 to 3 Yrs
5	skyleaf consultants	Delhi	3 to 6 Yrs
6	quiscon biotech	Delhi\n+6	0 to 1 Yr
7	nina s hr consultancy	Delhi	2 to 3 Yrs
8	nina s hr consultancy	Delhi	2 to 3 Yrs
9	acme services private limited	Delhi\n+4	3 to 5 Yrs

In [34]:
```
#Q4Scrape data of first 100 sunglasses listings on flipkart.com. You have to scrape for following attributes:
#Brand,ProductDescription,Price
```

In [35]:
```
driver = webdriver.Chrome()
```

In [40]:
```
driver.get('https://www.flipkart.com/')
```

In [44]:
```
search_item = driver.find_element(By.CLASS_NAME,"_3704LK")
search_item.send_keys("sunglasses")
time.sleep(3)
```

In [45]:
```
click_search_button = driver.find_element(By.CLASS_NAME,"_34RNph")
click_search_button.click()
time.sleep(3)
```

In [46]:
```
Brand = []
Product_Description = []
Price = []

for i in range(3):

    brand = driver.find_elements(By.XPATH,"//div[@class='_2WkVRV']")
    product_des = driver.find_elements(By.XPATH,"//div[@class='_2B099V']/a[1]")
    price = driver.find_elements(By.XPATH,"//div[@class='_30jeq3']")
```

```
    for i in brand:
        Brand.append(i.text)
    for i in product_des:
        Product_Description.append(i.text)
    for i in price :
        Price.append(i.text)

time.sleep(3)

nxt_button = driver.find_element(By.XPATH,"//a[@class='_1LKTO3']")
nxt_button.click()
```

In [47]:
```
print(len(Brand),len(Product_Description),len(Price))
```

```
120 120 120
```

In [48]:
```
df = pd.DataFrame({'Brand':Brand,'Product_Description':Product_Description,'Price':Price})

df[0:100]
```

Out[48]:

	Brand	Product_Description	Price
0	Singco India	UV Protection Rectangular Sunglasses (55)	₹502
1	ROYAL SON	UV Protection Rectangular, Retro Square Sungla...	₹497
2	SRPM	UV Protection Wayfarer Sunglasses (50)	₹204
3	PIRASO	UV Protection Clubmaster Sunglasses (54)	₹239
4	Elligator	UV Protection Cat-eye, Retro Square, Oval, Rou...	₹179
...	...	...	...
95	Fastrack	UV Protection Wayfarer Sunglasses (58)	₹769
96	Eyewearlabs	Polarized, UV Protection Wayfarer Sunglasses (54)	₹1,799
97	Eyewearlabs	Polarized, UV Protection Wayfarer Sunglasses (56)	₹2,099
98	PIRASO	Others Retro Square Sunglasses (50)	₹256
99	Fastrack	UV Protection Aviator Sunglasses (58)	₹1,037

100 rows × 3 columns

In [49]:
```
#Q6Scrape data for first 100 sneakers listings on flipkart.com. You have to scrape for following attributes:
#Brand,ProductDescription,Price
```

In [50]:
```
driver = webdriver.Chrome()
```

In [51]:
```
driver.get('https://www.flipkart.com/')
```

In [52]:
```
search_item = driver.find_element(By.CLASS_NAME,"_3704LK")
search_item.send_keys("sneakers")
time.sleep(3)
```

In [53]:
```
click_search_button = driver.find_element(By.XPATH,"/html/body/div[1]/div/div[1]/div[1]/div[2]/div[2]/form/div/
click_search_button.click()
time.sleep(3)
```

In [54]:
```
Brand = []
Product_Description = []
Price = []

for i in range(3):

    brand = driver.find_elements(By.XPATH,"//div[@class='_2WkVRV']")
    product_des = driver.find_elements(By.XPATH,"//div[@class='_2B099V']/a[1]")
    price = driver.find_elements(By.XPATH,"//div[@class='_30jeq3']")



    for i in brand:
        Brand.append(i.text)
    for i in product_des:
        Product_Description.append(i.text)
    for i in price :
        Price.append(i.text)

time.sleep(3)

nxt_button = driver.find_element(By.XPATH,"//a[@class='_1LKTO3']")
nxt_button.click()
```

In [55]:
```
print(len(Brand),len(Product_Description),len(Price))
```

```
120 120 120
```

In [56]:
```
df = pd.DataFrame({'Brand':Brand,'Product_Description':Product_Description,'Price':Price})

df[0:100]
```

Out[56]:

	Brand	Product_Description	Price
0	SFR	Exclusive Affordable Collection of Trendy & St...	₹349
1	Layasa	Sneakers For Women	₹499
2	BRUTON	Modern Trendy Shoes Sneakers For Men	₹379
3	Nobelite	Sneakers For Men	₹299
4	PUMA	Rebound LayUp SL Sneakers For Men	₹2,399
...	...	...	...
95	Sparx	Sneakers For Men	₹562
96	RapidBox	Sneakers For Men	₹750
97	PREKANZO	New Stylish Casual Sneakers For Women	₹569
98	Free Kicks	Combo Of 2 Shoes FK-444 & FK-206 Sneakers For Men	₹599
99	Kraasa	INFINITY 2.0 Sneakers For Women	₹399

100 rows × 3 columns

In [57]:
```
#Q7Scrape data for first 10 laptop(Intel core i7) listings on amazon.in You have to scrape for following attrib
#Title,Ratings,Price
```

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

In [59]:
```
driver.get('https://www.amazon.in/')
time.sleep(2)
```

In [60]:
```
search_item = driver.find_element(By.XPATH,"//input[@class='nav-input nav-progressive-attribute']")
search_item.send_keys("laptop")
```

In [61]:
```
click_search_button = driver.find_element(By.XPATH,"//input[@id='nav-search-submit-button']")
click_search_button.click()
time.sleep(3)
```

In [69]:
```
Title = []
Ratings = []
Price = []

title = driver.find_elements(By.XPATH,"//span[@class='a-size-medium a-color-base a-text-normal']")
price = driver.find_elements(By.XPATH,"//span[@class='a-price-whole']")

for i in title:Title.append(i.text)
for i in price:Price.append(i.text)
```

In [70]:
```
Laptop = pd.DataFrame({'Title':Title,'Price':Price})
Laptop[0:10]
```

Out[70]:

	Title	Price
0	MSI GF63 Thin, Intel Core i7-11800H, 40CM FHD ...	70,990
1	ASUS Vivobook 15, Intel Core i7-12650H 12th Ge...	64,990
2	Dell Vostro 5630 13th Gen Laptop,Intel i7-1355...	89,990
3	ASUS TUF Gaming F15, 15.6-inch (39.62 cms) FHD...	1,03,790
4	Dell Inspiron 5430 13th Gen Laptop, Intel i7-1...	86,249
5	MSI Modern 14, Intel 12th Gen. i7-1255U, 36CM ...	52,990
6	Lenovo [SmartChoice] IdeaPad Slim 3 Intel Core...	62,990
7	Dell Inspiron 5630 13th Gen Laptop, Intel Core...	89,990
8	Lenovo IdeaPad Slim 5 Intel Core i7 12th Gen 1...	80,990
9	ASUS Creator Series Vivobook 16X (2023), Intel...	86,990

In [71]:
```
#Q8Write a python programe to scrape data for top 1000 quotes of all time from azquotes.com.
#scrape quote,author,type of quotes
```

In [2]:
```
driver = webdriver.Chrome()
```

In [5]:
```
driver.get('https://www.azquotes.com/')
```

In [7]:
```
top_quotes = driver.find_element(By.XPATH,'//a[@href="/top_quotes.html"]')
```

```
top_quotes.click()
```

In [8]:
```
Quote = []
Author= []
Quote_type= []

for i in range(10):

    quote= driver.find_elements(By.XPATH,'//a[@class="title"]')
    author = driver.find_elements(By.XPATH,'//div[@class="author"]')
    quote_type= driver.find_elements(By.XPATH,'//div[@class="tags"]')



    for i in quote:
        Quote.append(i.text)
    for i in author:
        Author.append(i.text)
    for i in quote_type :
        Quote_type.append(i.text)

time.sleep(3)

nxt_button = driver.find_element(By.XPATH,'//li[@class="next"]')
nxt_button.click()
```

In [9]:
```
print(len(Quote),len(Author),len(Quote_type))
```

```
1000 1000 1000
```

In [10]:
```
df = pd.DataFrame({'Quote':Quote,'Author':Author,'Quote_type':Quote_type})

df[0:1000]
```

Out[10]:

	Quote	Author	Quote_type
0	The essence of strategy is choosing what not t...	Michael Porter	Essence, Deep Thought, Transcendentalism
1	One cannot and must not try to erase the past ...	Golda Meir	Inspiration, Past, Trying
2	Patriotism means to stand by the country. It d...	Theodore Roosevelt	Country, Peace, War
3	Death is something inevitable. When a man has ...	Nelson Mandela	Inspirational, Motivational, Death
4	You have to love a nation that celebrates its ...	Erma Bombeck	4th Of July, Food, Patriotic
...	...	...	...
995	When the going gets weird, the weird turn pro.	Hunter S. Thompson	Music, Sports, Hunting
996	When a train goes through a tunnel and it gets...	Corrie Ten Boom	Trust, Encouraging, Uplifting
997	If you think you are too small to make a diffe...	Dalai Lama	Inspirational, Funny, Change
998	God doesn't require us to succeed, he only req...	Mother Teresa	Success, God, Mother
999	Change your thoughts and you change your world.	Norman Vincent Peale	Inspirational, Motivational, Change

1000 rows × 3 columns

In [2]:
```
#Q9write a python programe to display list of respected former prime ministers of India from jagranjosh.com
```

In [3]:
```
driver = webdriver.Chrome()
```

In [14]:
```
driver.get('https://www.jagranjosh.com/')
```

In [21]:
```
url= driver.find_elements(By.XPATH,'//a[@href="/general-knowledge?ref=nav_dd"]')
url[0:1]
```

Out[21]:
```
[<selenium.webdriver.remote.webelement.WebElement (session="5ef2ed1de6df86011728f88f02baa6da", element="4B40F2B
DF1BCE79B4A6B5729C84F9644_element_148")>]
```

In [22]:
```
for i in url[0:1]:
    print(i.get_attribute('href'))
```

```
https://www.jagranjosh.com/general-knowledge?ref=nav_dd
```

In [23]:
```
url_list= driver.find_elements(By.XPATH,'//a[@href="https://www.jagranjosh.com/general-knowledge/list-of-all-pr
url_list[0:1]
```

Out[23]:
```
[]
```

In [24]:
```
for i in url_list[0:1]:
    print(i.get_attribute('href'))
```

In [25]:
```
Table=[]
for i in range(1):
    table= driver.find_elements(By.XPATH,'//div[@class="table-box"]/table/tbody/tr/td')
    for i in table:
```

```
        Table.append(i.text)
```

In [26]:
```
len(table)
```

Out[26]:
```
97
```

In [27]:
```
df = pd.DataFrame({'Table':Table,})

df
```

Out[27]:

	Table
0	1.
1	Jawahar Lal Nehru
2	(1889–1964)
3	15 August 1947 to 27 May 1964\n16 years, 286 days
4	The first prime minister of India and the long...
...	...
92	(born 1950)
93	30 May 2019- Incumbent
94	First non-congress PM with two consecutive ten...
95	List of all Presidents of India
96	List of Nicknames of Indian Prime Ministers

97 rows × 1 columns

In [11]:
```
#Q10 scrape the data to display 50 most expensive cars in the world from motor1.com
```

In [12]:
```
driver = webdriver.Chrome()
```

In [14]:
```
driver.get('https://www.motor1.com/')
```

In [15]:
```
search_item = driver.find_element(By.XPATH,'//input[@class="m1-search-panel-input m1-search-form-text"]')
search_item.send_keys("50 most expensive cars")
```

In [16]:
```
click_search_button = driver.find_element(By.XPATH,'//button[@class="m1-search-panel-button m1-search-form-butt
click_search_button.click()
time.sleep(3)
```

In [17]:
```
url = driver.find_elements(By.XPATH,'//a[@href="/features/308149/most-expensive-new-cars-ever/"]')
url[0:1]
```

Out[17]:
```
[<selenium.webdriver.remote.webelement.WebElement (session="5ca5bd0469c51fcb1c3983998ae3af40", element="E82115F
4E76D67790582DC0D005DAC10_element_312")>]
```

In [18]:
```
for i in url[0:1]:
    print(i.get_attribute('href'))
```

```
https://www.motor1.com/features/308149/most-expensive-new-cars-ever/
```

In [19]:
```
Car_name=[]

for i in range(1):

    car_name = driver.find_elements(By.XPATH,"//h3[@class='subheader']")

    for i in car_name:
        Car_name.append(i.text)

time.sleep(3)
```

In [20]:
```
print(len(Car_name))
```

```
51
```

In [21]:
```
df = pd.DataFrame({'Car_name':Car_name,})

df[0:51]
```

Out[21]:

	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
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

In [ ]: