In [1]:

```
! pip install selenium
```

```
Requirement already satisfied: selenium in c:\users\91852\anaconda3\lib\site-packages (4.
0.0)
Requirement already satisfied: urllib3[secure]~=1.26 in c:\users\91852\anaconda3\lib\site
-packages (from selenium) (1.26.4)
Requirement already satisfied: trio~=0.17 in c:\users\91852\anaconda3\lib\site-packages (
from selenium) (0.19.0)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\91852\anaconda3\lib\site-p
ackages (from selenium) (0.9.2)
Requirement already satisfied: sniffio in c:\users\91852\anaconda3\lib\site-packages (fro
m trio~=0.17->selenium) (1.2.0)
Requirement already satisfied: sortedcontainers in c:\users\91852\anaconda3\lib\site-pack
ages (from trio~=0.17->selenium) (2.3.0)
Requirement already satisfied: cffi>=1.14 in c:\users\91852\anaconda3\lib\site-packages (
from trio~=0.17->selenium) (1.14.5)
Requirement already satisfied: idna in c:\users\91852\anaconda3\lib\site-packages (from t
rio~=0.17->selenium) (2.10)
Requirement already satisfied: async-generator>=1.9 in c:\users\91852\anaconda3\lib\site-
packages (from trio~=0.17->selenium) (1.10)
Requirement already satisfied: attrs>=19.2.0 in c:\users\91852\anaconda3\lib\site-package
s (from trio~=0.17->selenium) (20.3.0)
Requirement already satisfied: outcome in c:\users\91852\anaconda3\lib\site-packages (fro
m trio~=0.17->selenium) (1.1.0)
Requirement already satisfied: pycparser in c:\users\91852\anaconda3\lib\site-packages (f
rom cffi>=1.14->trio~=0.17->selenium) (2.20)
Requirement already satisfied: wsproto>=0.14 in c:\users\91852\anaconda3\lib\site-package
s (from trio-websocket~=0.9->selenium) (1.0.0)
Requirement already satisfied: certifi in c:\users\91852\anaconda3\lib\site-packages (fro
m urllib3[secure]~=1.26->selenium) (2020.12.5)
Requirement already satisfied: pyOpenSSL>=0.14 in c:\users\91852\anaconda3\lib\site-packa
ges (from urllib3[secure]~=1.26->selenium) (20.0.1)
Requirement already satisfied: cryptography>=1.3.4 in c:\users\91852\anaconda3\lib\site-p
ackages (from urllib3[secure]~=1.26->selenium) (3.4.7)
Requirement already satisfied: six>=1.5.2 in c:\users\91852\anaconda3\lib\site-packages (
from pyOpenSSL>=0.14->urllib3[secure]~=1.26->selenium) (1.15.0)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\91852\anaconda3\lib\site-package
s (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.12.0)
```

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. This task will be done in following steps:

1. **First get the webpage https://www.naukri.com/ 2. Enter “Data Analyst” in “Skill, Designations, Companies” field and enter “Bangalore” in “enter the location” field.**
2. **Then click the search button.**
3. **Then scrape the data for the first 10 jobs results you get.**
4. **Finally create a dataframe of the scraped data.**

In [2]:

```
import warnings
warnings.filterwarnings('ignore')
```

In [3]:

```
#lets now import all the requried libraries
import selenium
import pandas as pd
from selenium import webdriver
```

In [4]:

```
#lets first connect to the web driver
```

```
driver = webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chromedriver_win32/chromedriver.exe")
```

In [5]:

```
# driver.close()
```

In [6]:

```
driver.get('https://www.naukri.com/')
```

In [7]:

```
# finding element for job search bar
search_job = driver.find_element_by_id('qsb-keyword-sugg')
# job search bar
```

In [8]:

```
search_job.send_keys("Data Analyst")
```

In [9]:

```
search_btn = driver.find_element_by_xpath("//div[@class='search-btn']/button")
#location search bar
```

In [10]:

```
#specifing the url of the webpage to be scraped
url="https://www.naukri.com/data-analytics-jobs-in-bangalore-bengaluru?k=data%20analytics&l=bangalore%2Fbengaluru"
```

In [11]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver whether the webpage has opened or not. lets check it

So,Now lets first create 10 empty lists. In these lists the data will be stored while scraping. We have created 10 empty lists for 4 features we have to extract

1.job_titles2.company_names3.locations_list4.experience_list

In [17]:

```
job_titles=[]
company_names=[]
locations=[]
experience_list=[]
```

First,we will extract all the tags where we have the jobs titles.Let me first show you on the webpae in which tags the job titles are put.

In [18]:

```
# so Lets extract all the tags having the jobs-titles
titles_tags=driver.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
titles_tags[0:10]
```

Out[18]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd39523066b", element="0351a6f8-17f1-46cb-b869-4c1294c576b6")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd39523066b", element="94b53808-b4b2-4501-857a-08efe635a6f3")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd39523066b", element="2294f6a0-1da8-4888-9ea2-4fe34b6b27aa")>,
```

```
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="a1a4b8aa-b1c4-4475-9b01-6763c10a0531")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="6c401bce-580b-4d73-9f29-5d87657ca378")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="26403101-d13d-4eeb-81a2-420c8d7aa9bb")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="662f2f0a-9d8c-43d7-9cb8-305d8549b6cc")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="05dad0a6-f883-477f-8693-0a6d9e34e370")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="729ad147-1049-48af-a4a4-ec9fe4853151")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="47653fb9-db9e-4569-8730-1ed90b1ba948")>]
```

Now we have all the tags in which there are the job titles.

Now we will extract the text from these tags one by one by looping over these tags

In [19]:

```
# Now the text of the job title is inside the tags extracted above.
# So we will run a loop to iterate over the tags extracted above and extract the tags

for i in titles_tags:
    title=i.text
    job_titles.append(title)
job_titles[0:10]
```

Out[19]:

```
['Senior Data Scientist- Data & Analytics',
 'Data Analyst - Flipkart Analytics',
 'Cloud Solution Architect - Data & Analytics',
 'Vice President- KYC Data Analytics Operations',
 'Data Analytic Consultant- US based MNC',
 'Data Analytics - For freshers',
 'Data Analytics and Interpretation Application Developer',
 'Manager - Data Analytics',
 'H&M Group Senior Data Engineer - Analytics & Data Platforms',
 'Lead Data Analyst - Claim Analytics']
```

Now we will extract all the html tags where we have the company names. Lets me first show you in which tags the company names are put on the webpage.

In [20]:

```
# so lets extract all the tags having the company names
companies_tags=driver.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
companies_tags[0:10]
```

Out[20]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="f5681c73-c92c-4eed-b230-4c2d32801040")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="0492334a-62da-416a-8dca-56931d759b75")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="ec97f135-5d0d-4b90-960a-d5475395e47e")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="0e5313c9-908d-4a4a-b64c-c83fcdb8fc5a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="a10eaac7-85f6-4d1d-a9f8-0b6adee6131a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="40f9d502-d8d7-438e-9650-0fcb07fd2081")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="f4277c59-766f-4c8d-95f4-2a5f142c48c5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="0a2077be-351b-4c02-995e-04f1924f0ca7")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
```

```
66b", element="80840876-aafe-48e4-99c8-e045b0527d59")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="7f0eca0d-e94b-4068-8c76-759a5ce1e5cd")>]
```

Now we have all the tags in which there are the company names.

Now we will extract the text from these tags one by one by looping over these tags

In [21]:

```
for i in companies_tags:
    company_name=i.text
    company_names.append(company_name)
company_names[0:10]
```

Out[21]:

```
['ExecBoardinAsia',
 'Flipkart',
 'Accenture',
 'Mancer Consulting Services Pvt. Ltd.',
 'RANDSTAD INDIA PVT LTD',
 'IANT',
 'Accenture',
 'LatentView Analytics Private Limited',
 'H and M Hennes and Mauritz (P) Ltd.',
 'InnovAccer']
```

Now we will extract all the html tags where we have the experience requried data. Let me first show you in which tags this data is put on the webpage.

In [22]:

```
# so lets extract all the tags having the experience requried data
experience_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHold
erLi experience']/span[1]")
experience_tags[0:10]
```

Out[22]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="dddd93d1-5d91-4969-84b0-7a34396fb117")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="a6844952-66d5-4659-b665-458742a9e9b4")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="d71946a0-fc8f-49c6-9278-d5dfae3723ba")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="ee0c9d06-db57-40dd-963c-7ba21c6baf68")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="54b888a2-5570-41b8-a2f6-9d4bc61afa05")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="5702dc7f-7a80-41e4-a3e9-c9427255dbb7")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="12f4da58-dea1-4f66-8126-808c0c52e995")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="549a5835-1ffd-4ee5-aa60-4a6a5af62efd")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="140c1c14-09fe-4b29-a05b-0b0d1c7dc994")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="38d28b9b-ef39-4213-88ee-c85098072dcb")>]
```

Now we have all the tags in there is the experience requried data.

Now we will extract the text from these tags one by one by looping over these tags

In [23]:

```
for i in experience_tags:
    experience=i.text
```

```
    experience_list.append(experience)
experience_list[0:10]
```

Out[23]:

```
['5-10 Yrs',
 '0-3 Yrs',
 '12-20 Yrs',
 '10-12 Yrs',
 '4-9 Yrs',
 '0-1 Yrs',
 '4-6 Yrs',
 '8-13 Yrs',
 '6-9 Yrs',
 '7-12 Yrs']
```

Now we will extract all the html tags where we have the location of the job data. Let me first show you in which tags this data is put on yhe webpage.

In [24]:

```
locations_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHolde
rLi location']/span[1]")
locations_tags[0:10]
```

Out[24]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="3b189bf3-dd70-4478-9453-5cedefb04476")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="63e9dd0b-b5fd-494a-a41c-ff9da11c912a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="767b62c0-5028-4108-b387-d122aa14c9de")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="4bbca121-249b-4d27-97ed-907c7b3973b0")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="5e8a57cf-177f-45b5-90fd-f788d0119f31")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="d8f25bbd-2dd1-4480-b177-ddf75569f002")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="c1193f8d-a620-4c4b-8607-6a25fb9df247")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="6f70dc0a-45ec-456a-b345-d5b6b5ea8b26")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="b0df7e4f-ba0a-4181-8945-a9450bc530e6")>,
 <selenium.webdriver.remote.webelement.WebElement (session="ecb8b53331823906c6e4abd395230
66b", element="4a970793-e40b-4cf1-8948-0e6be0ad1298")>]
```

Now we have all the tags in which there is the data about the location of job.

Now we will extract the text(location) from these tags one by one by looping over these tags

In [25]:

```
for i in locations_tags :
    location=i.text
    locations.append(location)
locations[0:10]
```

Out[25]:

```
['Bangalore/Bengaluru',
 'Bangalore/Bengaluru(Bellandur)',
 'Kolkata, Mumbai, Hyderabad/Secunderabad, Pune, Chennai, Bangalore/Bengaluru, Delhi / NC
R',
 'Mumbai, Hyderabad/Secunderabad, Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
```

```
 'Noida, Bangalore/Bengaluru']
```

So,now we have extracted the data requried from the webpage and stored them in the 4 lists mentioned above. Now before creating a dataframe from these lists. Lets first check the length of each of the list. Because if the length of all of the lists are not equal, then a dataframe cannot be formed.

In [26]:

```
print(len(job_titles),len(company_names),len(experience_list),len(locations))
```

```
20 20 20 20
```

In [27]:

```
#import pandas as pd
jobs=pd.DataFrame({})
jobs['title']=job_titles[0:10]
jobs['company']=company_names[0:10]
jobs['experience_requried']=experience_list[0:10]
jobs['location']=locations[0:10]
```

In [28]:

```
jobs
```

Out[28]:

	title	company	experience_requried	location
0	Senior Data Scientist- Data & Analytics	ExecBoardinAsia	5-10 Yrs	Bangalore/Bengaluru
1	Data Analyst - Flipkart Analytics	Flipkart	0-3 Yrs	Bangalore/Bengaluru(Bellandur)
2	Cloud Solution Architect - Data & Analytics	Accenture	12-20 Yrs	Kolkata, Mumbai, Hyderabad/Secunderabad, Pune,...
3	Vice President- KYC Data Analytics Operations	Mancer Consulting Services Pvt. Ltd.	10-12 Yrs	Mumbai, Hyderabad/Secunderabad, Bangalore/Beng...
4	Data Analytic Consultant- US based MNC	RANDSTAD INDIA PVT LTD	4-9 Yrs	Bangalore/Bengaluru
5	Data Analytics - For freshers	IANT	0-1 Yrs	Bangalore/Bengaluru
6	Data Analytics and Interpretation Application ...	Accenture	4-6 Yrs	Bangalore/Bengaluru
7	Manager - Data Analytics	LatentView Analytics Private Limited	8-13 Yrs	Bangalore/Bengaluru
8	H&M Group Senior Data Engineer - Analytics & D...	H and M Hennes and Mauritz (P) Ltd.	6-9 Yrs	Bangalore/Bengaluru
9	Lead Data Analyst - Claim Analytics	InnovAccer	7-12 Yrs	Noida, Bangalore/Bengaluru

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:

1. **First get the webpage https://www.naukri.com/ 2. Enter “Data Scientist” in “Skill, Designations, Companies” field and enter “Bangalore” in “enter the location” field.**
2. **Then click the search button.**
3. **Then scrape the data for the first 10 jobs results you get.**
4. **Finally create a dataframe of the scraped data**

In [29]:

```
#lets now import all the requried libraries
import selenium
import pandas as pd
```

```
from selenium import webdriver

import warnings
warnings.filterwarnings('ignore')
```

In [31]:

```
# Lets first connect to the web driver
driver = webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chro
medriver_win32/chromedriver.exe")
```

In [32]:

```
# driver.close()
```

In [33]:

```
driver.get('https://www.naukri.com/')
```

In [34]:

```
# finding element for job search bar
search_job = driver.find_element_by_id('qsb-keyword-sugg')
# job search bar
```

In [35]:

```
search_job.send_keys("Data Scientist")
#location search bar
```

In [36]:

```
search_btn = driver.find_element_by_xpath("//div[@class='search-btn']/button")
search_btn.click()
```

In [37]:

```
#specifing the url of the webpage to be scraped
url="https://www.naukri.com/data-scientist-jobs-in-bangalore-bengaluru?k=data%20scientist
&l=bangalore%2Fbengaluru"
```

In [38]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver whether the webpage has opened or not. lets check it

So,Now lets first create 10 empty lists. In these lists the data will be stored while scraping.We have created 10 empty lists for 4 features which we have to exract

1.job_titles2.company_names3.locations_list4.experience_list

In [39]:

```
job_titles=[]
company_names=[]
locations=[]
experience_list=[]
```

First, we will extract all the tags where we have the jobs titles.Let me first show you on the webpage in which tags the job titles are put.

In [41]:

```
# so Lets extract all the tags having the jobs-titles
titles_tags=driver.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
titles_tags[0:10]
```

Out[41]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="31cea803-eee3-49e6-9636-5681b2d385af")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="2c3a8ef7-5c48-4759-83e8-a8111a089205")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="6aadb6d6-9270-4937-951e-c21d358e38d9")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="64f5f1d2-17e0-47c6-8849-215b03bf088a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="1a95d2a1-d126-48f5-9101-6987b6cdf022")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="f3fa0870-d929-467f-8d69-0de9850b1b32")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="b6c26a88-c36a-4f67-9681-a8a794749f6f")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="cf5036a2-8f64-4f36-8a3b-e5e243eb91b0")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="8a37a1b2-5a36-4ea0-a0dd-4fe4ad20e10f")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="a0be138c-0059-489a-95cb-739cf6617d51")>]
```

Now we have all the tags in which there are the job titles.

Now we will extract the text from these tags one by one by looping over these tags

In [42]:

```
# Now the text of the job title is inside the tags extracted above.
# so we will run a loop to iterate over the tags extracted above and extract the tags

for i in titles_tags:
    title=i.text
    job_titles.append(title)
job_titles[0:10]
```

Out[42]:

```
['Senior Data Scientist',
 'Data Scientist: Advanced Analytics',
 'Senior Data Scientist | Fortune 500 Supermarket Chain',
 'Senior Data Scientist',
 'Senior Data Scientist- Data & Analytics',
 'Forecasting Analyst/ Data Scientist (US Client)',
 'Data Scientist',
 'Process Innovation Analyst - APAC/Data Scientist - Third Party Role',
 'Data Scientist',
 'Data Scientist']
```

Now we will extract all the html tags where we have the company names.Let me first show you in which tags the company names are put on the webpage.

In [43]:

```
# so lets extract all the tags having the company names
companies_tags=driver.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
companies_tags[0:10]
```

Out[43]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="9e7d9fc4-3c83-4c99-81fd-05fec9db1af1")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="3d26c275-5f5d-4df7-8064-15efba13e864")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="0623c367-b11a-48e0-ada2-cdae0cbb9b68")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="6ee95c36-8338-4e90-9619-e87a07134156")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="c5b4c716-e7c1-4c5b-b6ad-ddd517315720")>,
```

```
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="aa5041c5-802f-4030-9135-0e8b05ea7492")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="76d7fe25-7bc1-4023-9de2-ceb1f2a69eda")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="a72ede06-081a-43dd-8977-13cb79219d09")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="2c1a3699-836c-4d1f-b971-8fa7f6e816db")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="6526bcdd-444f-498d-8d69-19e1eadb537e")>]
```

Now we have all the tags in which there are the company names.

Now we will extract the text from these tags one by one by looping over these tags

In [44]:

```
for i in companies_tags:
    company_name=i.text
    company_names.append(company_name)
company_names[0:10]
```

Out[44]:

```
['IBM India Pvt. Limited',
 'IBM India Pvt. Limited',
 'TALENT500 TECH (INDIA) PRIVATE LIMITED',
 'Hitachi Ltd.',
 'ExecBoardinAsia',
 'Concentrix Daksh Services',
 'Toppr',
 'Bayer',
 'Hitachi Ltd.',
 'Mobiotics IT Solution Pvt Ltd']
```

Now we will extract all the html tags where we have the experience requried data. Let me first show you in which tags this data is put on the webpage.

In [45]:

```
# so Lets extract all the tags having the experience requried data
experience_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHold
erLi experience']/span[1]")
experience_tags[0:10]
```

Out[45]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="895a8eb6-760e-4315-8628-4cd41621e059")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="31011733-2dfd-4d23-86d6-9907bf5205e5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="b17b6d34-e69e-430d-8f4b-b3f52da9b7c5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="31902474-85e0-40c4-af7e-5c1d166e80c4")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="835b9a8e-c66c-469e-918e-2bcc1ceaf515")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="7bb49f2c-1aac-4dcb-a3f6-db1d515f03f3")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="42f783cd-1c49-40b4-a3f2-2595cf1fece6")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="1d7e4893-edf0-40ee-924c-4be1cfe5f946")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="d8fb2e81-a322-411f-99bd-0f5612c5c963")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5
d40", element="0a256c61-aafe-42dd-b94f-6ec910ff98f9")>]
```

Now we have all the tags in there is the experience requried data.

Now we will extract the text from these tags one by one by looping over these tags

Now we will extract the text from these tags one by one by looping over these tags

In [46]:

```
for i in experience_tags :
    experience=i.text
    experience_list.append(experience)
experience_list[0:10]
```

Out[46]:

```
['5-10 Yrs',
 '5-10 Yrs',
 '7-10 Yrs',
 '2-5 Yrs',
 '5-10 Yrs',
 '3-8 Yrs',
 '1-3 Yrs',
 '4-8 Yrs',
 '3-7 Yrs',
 '2-5 Yrs']
```

Now we will extract all the html tags where we have the location of the job data. Let me first show you in which tags this data is put on the webpage.

In [47]:

```
locations_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHolderLi location']/span[1]")
locations_tags[0:10]
```

Out[47]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="3ddfdf63-871d-482b-a43c-e4cbc3e8703b")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="e1cab38c-da61-4229-8f8e-3ad56e3aae1b")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="af4d5e2b-30ec-455e-82b4-2f97b6cbf034")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="80a5ee91-79fc-48b8-8348-7ae90b0f21a7")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="7068def9-1714-4215-a0d8-900ecd26a9e7")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="45c88822-7b31-4e89-8f5f-f0dd95f5bb54")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="6b967ec8-6037-4168-8ce5-9f83b48e63c1")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="33aee364-aa46-414f-85b0-d9b9f24eef85")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="ad33b506-0bea-49f8-ad99-3a4a3108b8eb")>,
 <selenium.webdriver.remote.webelement.WebElement (session="d330f8ffdd008a496d2040c98f9b5d40", element="ccd3e77e-ccb1-433d-a618-7701fa3ac2d4")>]
```

Now we have all the tags in which there is the data about the location of job.

Now we will extract the text(location) from these tags one by one by looping over these tags

In [48]:

```
for i in locations_tags :
    location=i.text
    locations.append(location)
locations[0:10]
```

Out[48]:

```
['Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
```

```
 'Gurgaon/Gurugram, Bangalore/Bengaluru',
 'Hyderabad/Secunderabad, Bangalore/Bengaluru, Mumbai (All Areas)',
 'Bangalore/Bengaluru, Hyderabad',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru(HSR Layout)']
```

So,now we have extracted the data requried from the webpage and stored them in the 4 lists mentioned above. Now before creating a dataframe from these lists. Lets first check the length of each of the list. Because if the length of all of the lists are not equal, then a dataframe cannot be formed.

In [49]:

```
print(len(job_titles),len(company_names),len(experience_list),len(locations))
```

```
20 20 20 20
```

In [50]:

```
#import pandas as pd
jobs=pd.DataFrame({})
jobs['title']=job_titles[0:10]
jobs['company']=company_names[0:10]
jobs['experience_requried']=experience_list[0:10]
jobs['location']=locations[0:10]
```

In [51]:

```
jobs
```

Out[51]:

	title	company	experience_requried	location
0	Senior Data Scientist	IBM India Pvt. Limited	5-10 Yrs	Bangalore/Bengaluru
1	Data Scientist: Advanced Analytics	IBM India Pvt. Limited	5-10 Yrs	Bangalore/Bengaluru
2	Senior Data Scientist | Fortune 500 Supermarke...	TALENT500 TECH (INDIA) PRIVATE LIMITED	7-10 Yrs	Bangalore/Bengaluru
3	Senior Data Scientist	Hitachi Ltd.	2-5 Yrs	Bangalore/Bengaluru
4	Senior Data Scientist- Data & Analytics	ExecBoardinAsia	5-10 Yrs	Bangalore/Bengaluru
5	Forecasting Analyst/ Data Scientist (US Client)	Concentrix Daksh Services	3-8 Yrs	Gurgaon/Gurugram, Bangalore/Bengaluru
6	Data Scientist	Toppr	1-3 Yrs	Hyderabad/Secunderabad, Bangalore/Bengaluru, M...
7	Process Innovation Analyst - APAC/Data Scienti...	Bayer	4-8 Yrs	Bangalore/Bengaluru, Hyderabad
8	Data Scientist	Hitachi Ltd.	3-7 Yrs	Bangalore/Bengaluru
9	Data Scientist	Mobiotics IT Solution Pvt Ltd	2-5 Yrs	Bangalore/Bengaluru(HSR Layout)

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

You have to use the location and salary filter. You have to scrape data for "Data Scientist" designation for first 10 job results. You have to scrape the job-title, job-location, company name, experience required. The location filter to be used is "Delhi/NCR" The salary filter to be used is "3-6" lakhs The task will be done as shown in the below steps:

1. **first get the webpage https://www.naukri.com/**
2. **Enter "Data Scientist" in "Skill, Designations, and Companies" field.**
3. **Then click the search button.**
4. **Then apply the location filter and salary filter by checking the respective boxes**
5. **Then scrape the data for the first 10 jobs results you get.**

6. Finally create a dataframe of the scraped data.

In [70]:

```
import warnings
warnings.filterwarnings('ignore')
```

In [71]:

```
#lets now import all the requried libraries
import selenium
import pandas as pd
from selenium import webdriver
```

In [72]:

```
# Lets first connect to the web driver
driver = webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chro
medriver_win32/chromedriver.exe")
```

In [73]:

```
# driver.close()
```

In [74]:

```
driver.get('https://www.naukri.com/')
```

In [75]:

```
# finding element for job search bar
search_job = driver.find_element_by_id('qsb-keyword-sugg')
```

In [76]:

```
search_job.send_keys("Data Scientist")
```

In [77]:

```
search_btn = driver.find_element_by_xpath("//div[@class='search-btn']/button")
search_btn.click()
```

In [78]:

```
#specifing the url of the webpage to be scraped
url="https://www.naukri.com/data-scientist-jobs?k=data%20scientist&ctcFilter=3to6&cityTyp
eGid=9508"
```

In [79]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver the webpage has opened or not.lets check it

So, Now lets first create 10 empty lists.In these lists the data will be stored while scraping.We have created 10 empty lists for 4 features which we have to extract

1.job_titles2.company_names3.locations_list4.experience_list

In [80]:

```
job_titles=[]
company_names=[]
locations=[]
experience_list=[]
```

First,we will extract all the tags where we have the jobs titles.Let me first show you on the webpage in which

tags the job titles are put.

In [81]:

```
# so Lets extract all the tags having the jobs-titles
titles_tags=driver.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
titles_tags[0:10]
```

Out[81]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="8cf211e8-5bbc-4341-ab8d-1aafaff1a5d5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="ee2305cb-1dcd-4e52-9229-a2b109afd4f1")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="8dad4ad5-6bdd-4366-bd71-5b930bb110fb")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="ed954f38-1397-474d-b733-5832d9313c62")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="bb8be4d7-a289-4d4a-b078-1495da77688c")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="c5e797bc-572b-44aa-93fe-ec4cd23976ea")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="5322394a-172d-43bc-9910-37b2ff5c9b45")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="759bf945-1dc0-4995-880f-3bdbd25728e2")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="71df8d1a-33c6-44ba-b4f6-6fb7ebc09a86")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="6a79f8a7-6d15-4bfa-8248-bd8f14922740")>]
```

Now we have all the tags in which there are the job titles.

Now we will extract the text from these tags one by one by looping over these tags

In [82]:

```
# Now the text of the job title is inside the tags extracted above.
# so we will run a loop to iterate over the tags extracted above and extract the tags

for i in titles_tags:
    title=i.text
    job_titles.append(title)
job_titles[0:10]
```

Out[82]:

```
['Data Scientist - Insurance',
 'Job Opportunity || Data Scientist || HCL Technologies',
 'Data Scientist',
 'Data Scientist',
 'Data Scientist III-2',
 'Data Scientist',
 'Data Scientist',
 'Urgent Hiring || Data Scientist || Delhi',
 'Immediate requirement For Data Scientist',
 'Data Scientist Internship']
```

Now we will extract all the html tags where we have the company names.Let me first show you in which tags the company names are put on the webpage.

In [83]:

```
# so lets extract all the tags having the company names
companies_tags=driver.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
companies_tags[0:10]
```

Out[83]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="972d1e54-1ac5-468a-9018-1c601538fa0f")>,
```

```
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="636e84a7-c974-4b06-8d15-aec0f6b3e056")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="19bdc332-2d02-4550-bdf6-f255e7d974e8")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="d9b1d8aa-42f8-4db1-ae41-ce620ef630fc")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="0473847d-0dcb-4f6e-8825-c884abeedb7c")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="9cd3cb6f-726c-4779-93e1-a281de46e0aa")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="ff8c1f9e-1984-4d5d-88d7-af3acd8053ff")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="a3d75556-1a2e-41dc-b8c9-734c1f7a7f1f")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="937a66aa-d73a-43c5-ac35-17aa7f1bd431")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="a93b8671-af63-4953-aba4-2fddd36210f8")>]
```

Now we have all the tags in which there are the company names.

Now we will extract the text from these tags one by one looping over these tags

In [84]:

```
for i in companies_tags:
    company_name=i.text
    company_names.append(company_name)
company_names[0:10]
```

Out[84]:

```
['Huquo Consulting Pvt. Ltd',
 'HCL Technologies',
 'Think i',
 'ThinkBumblebee Analytics Pvt. Ltd.',
 'Concentrix Daksh Services',
 'MoMagic Technologies Pvt. Ltd.',
 'MoMagic Technologies Pvt. Ltd.',
 'Shriram Automall India Limited',
 'CALIBEHR BUSINESS SUPPORT SERVICES PRIVATE LIMITED',
 'iHackers Inc']
```

Now we will extract all the html tags where we have the experience requried data.Let me first show you in which tags this data is put on the webpage.

In [85]:

```
# so Lets extract all the tags having the experience requried data
experience_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHold
erLi experience']/span[1]")
experience_tags[0:10]
```

Out[85]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="cc460e2d-c832-4168-9d57-0047d324aacf")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="ed55df92-a4b0-417e-984c-0adb44402e7a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="1937334f-5a78-43da-8963-cf8ea5ad70cd")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="15a8ff9d-1bbf-4bd6-977d-8cf155686901")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="a042aede-302f-48f8-8787-0a98cec0a473")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="c5475831-fba6-40ef-bed0-4190220090c5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="1837c486-9d06-4d3f-916a-cddf46463bed")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="9dd8d06c-f768-4610-8dd1-b1b632811381")>,
```

```
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="3891fd6a-f18c-4e47-a2f5-a4fd5dd2c8c5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="7884a99d-d538-4d1b-a3d2-5f2e899a3387")>]
```

Now we have all the tags in there is the experience requried data.

Now we will extract the text from these tags one by one by looping over these tags

In [86]:

```
for i in experience_tags:
    experience=i.text
    experience_list.append(experience)
experience_list[0:10]
```

Out[86]:

```
['2-7 Yrs',
 '4-7 Yrs',
 '0-2 Yrs',
 '2-6 Yrs',
 '3-8 Yrs',
 '4-6 Yrs',
 '4-6 Yrs',
 '2-7 Yrs',
 '2-7 Yrs',
 '0-1 Yrs']
```

Now we will extract all the html tags where we have the location of the job data.Let me first show you in which tags this data is put on the webpage.

In [87]:

```
locations_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHolde
rLi location']/span[1]")
locations_tags[0:10]
```

Out[87]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="c4a2ebdd-e55b-4cc1-a2db-6c6842d50529")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="425f052d-7064-421c-8356-2ed38be7261f")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="3e01bf9a-2cdf-4d65-a70d-3164a4ed7eff")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="5357e4c0-b2b3-4341-9ab8-d1bae4b51630")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="75f94b70-39c4-4015-a340-76d0f4d79feb")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="e3551eac-0e78-445b-9a7b-aa62db1cd304")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="061e7c53-e37a-44bf-8ae8-0cd89c2ab428")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="a2ac582a-6f3e-4a9b-a44e-2fb391eabb9a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="70ec7bb3-6810-4f61-a0d3-56a787e81488")>,
 <selenium.webdriver.remote.webelement.WebElement (session="3670c5ef00fbdfc7420dcbacbbaef
269", element="d8c33a95-8af4-462b-ba0b-f857f063924b")>]
```

Now we have all the tags in which there is the data about the location of job.

Now we will extract the text(location) from these tags one by one by looping over these tags

In [88]:

```
for i in locations_tags:
    location=i.text
    locations.append(location)
```

```
locations[0:10]
```

Out[88]:

```
['Noida, Gurgaon/Gurugram',
 'Delhi / NCR',
 'Kochi/Cochin, Kolkata, Hyderabad/Secunderabad, Pune, Ahmedabad, Chennai, Bangalore/Beng
aluru, Delhi / NCR, Mumbai (All Areas)',
 'Pune, Bangalore/Bengaluru, Delhi / NCR',
 'Gurgaon/Gurugram',
 'Noida(Sector-126 Noida)',
 'Noida(Sector-126 Noida)',
 'Delhi / NCR',
 'Mumbai, Hyderabad/Secunderabad, Pune, Chennai, Tamia, Bangalore/Bengaluru, Delhi / NCR'
,
 'New Delhi']
```

So,now we have extracted the data requried from the webpage and stored them in the 4 lists mentioned above.Now before creating a dataframe from these lists.Lets first check the length of each of the list. Because if the length of all of the lists are not equal, then a dataframe cannot be formed.

In [89]:

```
print(len(job_titles),len(company_names),len(experience_list),len(locations))
```

```
20 20 20 20
```

In [90]:

```
#import pandas as pd
jobs=pd.DataFrame({})
jobs['title']=job_titles[0:10]
jobs['company']=company_names[0:10]
jobs['experience_requried']=experience_list[0:10]
jobs['location']=locations[0:10]
```

In [91]:

```
jobs
```

Out[91]:

	title	company	experience_requried	location
0	Data Scientist - Insurance	Huquo Consulting Pvt. Ltd	2-7 Yrs	Noida, Gurgaon/Gurugram
1	Job Opportunity || Data Scientist || HCL Techn...	HCL Technologies	4-7 Yrs	Delhi / NCR
2	Data Scientist	Think i	0-2 Yrs	Kochi/Cochin, Kolkata, Hyderabad/Secunderabad,...
3	Data Scientist	ThinkBumblebee Analytics Pvt. Ltd.	2-6 Yrs	Pune, Bangalore/Bengaluru, Delhi / NCR
4	Data Scientist III-2	Concentrix Daksh Services	3-8 Yrs	Gurgaon/Gurugram
5	Data Scientist	MoMagic Technologies Pvt. Ltd.	4-6 Yrs	Noida(Sector-126 Noida)
6	Data Scientist	MoMagic Technologies Pvt. Ltd.	4-6 Yrs	Noida(Sector-126 Noida)
7	Urgent Hiring || Data Scientist || Delhi	Shriram Automall India Limited	2-7 Yrs	Delhi / NCR
8	Immediate requirement For Data Scientist	CALIBEHR BUSINESS SUPPORT SERVICES PRIVATE LIM...	2-7 Yrs	Mumbai, Hyderabad/Secunderabad, Pune, Chennai,...
9	Data Scientist Internship	iHackers Inc	0-1 Yrs	New Delhi

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

1. **Brand**
2. **Product Description**

3. **Price The attributes which you have to scrape is ticked marked in the below image.**

In [79]:

```
#lets now import all the requried libraries
import selenium
import pandas as pd
from selenium import webdriver

import warnings
warnings.filterwarnings('ignore')
```

In [92]:

```
#lets first connect to the web driver
driver = webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chro
medriver_win32/chromedriver.exe")
```

In [93]:

```
# driver.close()
```

In [94]:

```
driver.get('https://www.flipkart.com/')
```

In [95]:

```
# finding element for product search bar
product = driver.find_element_by_class_name("_3704LK")
```

In [96]:

```
product.send_keys("sunglasses")
#location search bar
```

In [98]:

```
search_btn = driver.find_element_by_xpath("//button[@class='L0Z3Pu']")
search_btn.click()
```

In [99]:

```
#specifing the url of the webpage to be scraped
url="https://www.flipkart.com/search?q=sunglasses&otracker=search&otracker1=search&market
place=FLIPKART&as-show=on&as=off"
```

In [100]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver whether the webpage has opened or not. lets check it

So,Now lets first create 100 empty lists. In these lists the data will be stored while scraping.We have created 100 empty lists for 3 features which we have to exract

1.Brand2.Description3.Price

In [101]:

```
#creating the empty list
brand=[]
Description=[]
price=[]
```

In [102]:

```
#creating the empty list
brand=[]
description=[]
price=[]

start=0
end=4
for page in range(start,end):
#for loop for scrapping 4 page
    brands=driver.find_elements_by_class_name('_2WkVRV')
#scraping brands name by class name='_2WkVRV'
    for i in brands:
        brand.append(i.text)
#appending the text in Brand list
        prices=driver.find_elements_by_xpath("//div[@class='_30jeq3']")
#scraping the price from the xpath
    for i in prices:
        price.append(i.text)
#appending the text in price list
    desc=driver.find_elements_by_xpath('//a[@class="IRpwTa"]')
#scraping description from the xpath
    for i in desc:
        description.append(i.text)
#appending the text in description list
    nxt_button=driver.find_elements_by_xpath("//a[@class='_1LKTO3']")
#scraping the list of buttons from the page
    try:
        driver.get(nxt_button[1].get_attribute('href'))
#getting the link from the list for next page
    except:
        driver.get(nxt_button[0].get_attribute('href'))
```

In [103]:

```
#creating a dataframe
df=pd.DataFrame({'Brand':brand[:100],
                 'Description':description[:100],
                 'Price':price[:100]})
```

In [105]:

```
#printing dataframe
df
```

Out[105]:

	Brand	Description	Price
0	Singco India	Gradient, Toughened Glass Lens, UV Protection ...	₹699
1	VINCENT CHASE	Polarized, UV Protection Wayfarer Sunglasses (57)	₹799
2	SRPM	UV Protection Wayfarer Sunglasses (56)	₹188
3	SUNBEE	UV Protection, Polarized, Mirrored Round Sungl...	₹276
4	Fastrack	UV Protection Rectangular Sunglasses (Free Size)	₹639
...	...	...	...
95	elegante	UV Protection Round Sunglasses (53)	₹359
96	ROYAL SON	UV Protection Aviator, Wayfarer Sunglasses (Fr...	₹711
97	PHENOMENAL	UV Protection, Others Rectangular Sunglasses (50)	₹249
98	DEIXELS	UV Protection, Gradient Oval Sunglasses (58)	₹229
99	Anemone	Gradient, UV Protection Wayfarer Sunglasses (F...	₹799

100 rows × 3 columns

Q6: Scrape data for first 100 sneakers you find when you visit flipkart.com and search for “sneakers” in the search field. You have to scrape 4 attributes of each sneaker:

1. **Brand**
2. **Product Description**
3. **Price**

In [106]:

```
import warnings
warnings.filterwarnings('ignore')
```

In [107]:

```
#lets now import all the requried libraries
import selenium
import pandas as pd
from selenium import webdriver
```

In [108]:

```
# Lets first connect to the web driver
driver = webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chro
medriver_win32/chromedriver.exe")
```

In [109]:

```
# driver.close()
```

In [110]:

```
driver.get('https://www.flipkart.com/')
```

In [111]:

```
# finding element for product search bar
product = driver.find_element_by_class_name("_3704LK")
```

In [112]:

```
product.send_keys("sneakers")
#location search bar
```

In [114]:

```
search_btn = driver.find_element_by_xpath("//button[@class='L0Z3Pu']")
search_btn.click()
```

In [115]:

```
#specifing the url of the webpage to be scraped
url="https://www.flipkart.com/search?q=sneakers&otracker=search&otracker1=search&marketpl
ace=FLIPKART&as-show=on&as=off"
```

In [116]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver whether the webpage has opened or not. lets check it

So,Now lets first create 100 empty lists. In these lists the data will be stored while scraping.We have created 100 empty lists for 3 features which we have to exract

1.brand2.description3.Price

In [121]:

```
#creating the empty list
```

```
brand=[]
description=[]
price=[]

start=0
end=4
for page in range(start,end):
    #for loop for scrapping 4 page

    brands=driver.find_elements_by_xpath("//div[@class='_2WkVRV']")
    #scraping title tags by xpath
    for i in brands:
        brand.append(i.text)
        #appending the text in brand

        price_tags=driver.find_elements_by_xpath("//div[@class='_30jeq3']")
        #scraping the price from the xpath
        for i in price_tags:
            price.append(i.text)
            #appending the text in price tags

            description=driver.find_elements_by_xpath("//span[@class='B_NuCI']")
            #scraping the description from the xpath
            for i in description:
                description.append(i.text)
                #appending the text in description

                nxt_button=driver.find_elements_by_xpath("//a[@class='_1LKTO3']")
                #scraping the list of buttons from the page
                try:
                    driver.get(nxt_button[1].get_attribute('href'))
                 #getting the link from the list for next page
                except:
                    driver.get(nxt_button[0].get_attribute('href'))
```

In [125]:

```
#creating a dataframe
df=pd.DataFrame({'brand':brand[:100],
                 'description':'description'[:100],
                 'price':price[:100]})
```

In [126]:

```
#printing dataframe
df
```

Out[126]:

	brand	description	price
0	Numenzo	description	₹479
1	SCATCHITE	description	₹398
2	Magnolia	description	₹398
3	CALCADOS	description	₹899
4	D-SNEAKERZ	description	₹348
...	...	...	...
95	Arohi	description	₹458
96	Chevit	description	₹699
97	T-ROCK	description	₹399
98	Labbin	description	₹399
99	Kreverse	description	₹245

100 rows × 3 columns

Q8: Go to webpage https://www.amazon.in Enter "Laptop" in the search field and then click the search icon. Then set CPU Type filter to "Intel Core i7" and "Intel Core i9" After setting the filters scrape first 10 laptops data. You have to scrape 3 attributes for each laptop:

1. **Title**
2. **Ratings**
3. **Price**

In [18]:

```
import warnings
warnings.filterwarnings('ignore')
```

In [19]:

```
# lets now import all the requried librares
import selenium
import pandas as pd
from selenium import webdriver
```

In [20]:

```
# lets first connect the webdriver
driver=webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chrome
driver_win32/chromedriver.exe")
```

In [21]:

```
# driver.close()
```

In [22]:

```
driver.get('https://www.amazon.in/')
```

In [23]:

```
# finding element for job search bar
search_job = driver.find_element_by_id('twotabsearchtextbox')
# job search bar
```

In [24]:

```
search_job.send_keys("laptop")
```

In [25]:

```
search_btn = driver.find_element_by_xpath("//span[@class='nav-search-submit-text nav-spri
te nav-progressive-attribute']")
search_btn.click()
```

In [26]:

```
#specifing the url of the webpage to be scraped
url="https://www.amazon.in/s?k=laptop&i=computers&rh=n%3A1375424031%2Cp_n_feature_thirtee
n_browse-bin%3A12598163031%7C16757432031&dc&qid=1636161846&rnid=12598141031&ref=sr_nr_p_n
_feature_thirteen_browse-bin_27"
```

In [27]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver whether the webpage has opened or not. lets check it

So,Now lets first create 10 empty lists. In these lists the data will be stored while scraping.We have created 10 empty lists for 3 features which we have to exract

1.title nams 2.Ratings 3.Price

In [29]:

```
title_names=[]
Ratings=[]
Price=[]
```

First, we will extract all the tags where we have the title name.Let me first show you on the webpage in which tags the title name are put.

In [31]:

```
# so Lets extract all the tags having the title-tags
title_tags=driver.find_elements_by_xpath("//span[@class='a-size-medium a-color-base a-tex
t-normal']")
title_tags[0:10]
```

Out[31]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="f90c205d-2314-485e-9894-eb7bb20b7387")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="50a0029e-fbb1-44a5-8c51-d3858bb9647a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="ea67204d-42f3-4cd0-abba-a91b7f025ccd")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="95381c35-4321-4a44-aa28-769ca0c16edf")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="b656b64d-de9b-4f86-9b3f-a9ced5a1296d")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="5983a332-f5a5-47e0-a596-c77673ee7ef5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="07322521-18b4-4597-8ee1-154b26d4c448")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="b653ea6a-3aaf-4040-980d-87ae605ac5b6")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="b375d2b3-5c6f-4577-b244-862e777cb75b")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="3f0edaa9-bd79-4de5-a92d-84e9c9d0b31a")>]
```

Now we have all the tags in which there are the title names.

Now we will extract the text from these tags one by one by looping over these tags

In [32]:

```
# Now the text of the title_name is inside the tags extracted above.
# so we will run a loop to iterate over the tags extracted above and extract the tags

for i in title_tags:
    title=i.text
    title_names.append(title)
title_names[0:10]
```

Out[32]:

```
['Mi Notebook Horizon Edition 14 Intel Core i7-10510U 10th Gen Thin and Light Laptop(8GB/
512GB SSD/Windows 10/Nvidia MX350 2GB Graphics/Grey/1.35Kg)(Without Webcam) XMA1904-AF',
 'HP Envy Intel 11th Gen Core i7 Processor 13.3 inches FHD Touchscreen Gaming Laptop (16G
B/1TB SSD/Windows 10/NVIDIA MX450 2GB/Natural Silver/1.3 kg), 13-ba1018TX',
 'ASUS TUF Gaming F17 (2021), 17.3-inch (43.94 cms) FHD 144Hz, Intel Core i7-11800H 11th
Gen, GeForce RTX 3050 Ti 4GB Graphics, Gaming Laptop (16GB/1TB SSD/Windows 10/Eclipse Gra
y/2.6 Kg) FX766HE-HX022T',
 'Mi Notebook Horizon Edition 14 Intel Core i7-10510U 10th Gen Thin and Light Laptop(8GB/
512GB SSD/Windows 10/Nvidia MX350 2GB Graphics/Grey/1.35Kg)(Without Webcam) XMA1904-AF',
 'HP Pavilion (2021) Intel 11th Gen Core i7 14 inches FHD Screen Thin & Light Laptop, 16
GB RAM, 1TB SSD, Iris Xe Graphics, Windows 10, MS Office, Backlit Keyboard, 1.41kg (14-dv
0058TU)',
```

```
 'ASUS TUF Dash F15 (2021), 15.6-inch (39.62 cms) FHD 144Hz, Intel Core i7-11370H 11th Ge
n, RTX 3050 4GB Graphics Gaming Laptop (16GB RAM/512GB SSD/Windows 10/White/2 kg), FX516P
C-HN062T',
 'Dell 14 (2021) i7-1195G7 2in1 Touch Screen Laptop, 16GB, 512GB SSD, Win 10 + MS Office,
14.0" (35.56 cms) FHD Display, Backlit KB, FPR + Active Pen, Platinum Silver (Inspiron 54
10, D560596WIN9S)',
 'Dell 15 (2021) i7-10870H Gaming Laptop, 16GB DDR4, 512GB SSD, Win 10 + MS Office, NVIDI
A RTX 3050 Ti 4GB, 15.6" (39.61 cms) FHD AG 250 nits 120Hz, Backlit KB Orange (G15 5510,
D560534WIN9B)',
 'Lenovo ThinkBook 15 Intel 11th Gen Core i7 15.6"(39.62 cm)FHD IPS 300 nits Antiglare 10
0% sRGB Thin & Light Laptop(16GB/1TB HDD+128GB SSD/Windows 10/MS Office/3 Yr Onsite Warra
nty/1.7 Kg) 20VEA0HBIH',
 'Acer Nitro 5 AN515-57 Gaming Laptop | Intel Core i7-11800H |NVIDIA GeForce RTX 3050 Ti
Laptop Graphics |15.6" FHD 144Hz IPS Display |16GB DDR4 |256GB SSD+1TB HDD |Killer Wi-Fi
6 |RGB Backlit Keyboard']
```

Now we will extract all the html tags where we have the Rating_tags data. Let me first show you in which tags this data is put on the webpage.

In [33]:

```
# so Lets extract all the tags having the Rating
Rating_tags=driver.find_elements_by_xpath("//span[@class='a-icon-alt']")
Rating_tags[0:10]
```

Out[33]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="2d9e0778-c5fc-4f09-8dfb-446fceaa3e3c")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="7106374a-84f7-4c92-ae94-bc0b137bd211")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="a62ac8d4-1db6-4d44-8d0f-12390225c418")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="73a9e590-afa1-4bf7-81d1-c106b0541c4a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="cd688af6-6c49-4487-ada3-377739c6c28f")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="f588b611-bfa5-4019-a05a-4d3bf98ed1e4")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="0674c6d4-e90d-4efb-a20a-0b353317e390")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="3a29c829-1e0b-46a5-bccc-86f9c58674ba")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="9491e481-0afc-4336-8016-5d909a2e44e8")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="47aa341c-d558-4cda-8eee-3c5ff28d81d8")>]
```

Now we have all the tags in there is the experience requried data.

Now we will extract the text from these tags one by one by looping over these tags

In [34]:

```
for i in Rating_tags:
    Rating=i.text
    Ratings.append(i.text)
Ratings[0:10]
```

Out[34]:

```
['', '', '', '', '', '', '', '', '', '']
```

Now we will extract all the tags having the price requried data.Let me first show you in which tags this data is put on the webpage.

In [35]:

```
# so Lets extract all the tags having the price requried data
```

```
price_tags=driver.find_elements_by_xpath("//span[@class='a-price-whole']")
price_tags[0:10]
```

Out[35]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="cd8cd83f-03a9-4e66-8d2b-1a98e934ef32")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="98cd8895-ed9e-4830-9758-fbc7eea91b1e")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="8d21acbe-49ea-487f-b7a7-5b32e25c25db")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="df50654b-3ddd-404a-be8e-6a0431998a8a")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="347cacb8-94a5-4cd3-ba88-ae4b690b0332")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="9c2bec20-ff9b-419e-b5d1-c67cf6a3d14e")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="0f6a241a-5e1b-41ac-8fc9-d1f536f0b6fb")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="c68df098-3c17-453f-9e76-d29e03ced9c5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="4d2bbc67-b483-46b5-8a7d-4ed71f9a9653")>,
 <selenium.webdriver.remote.webelement.WebElement (session="5b208f23d07d8befc87d4de7b89ce
890", element="a4c3ce6f-0786-4db0-bef2-1efb34d4271c")>]
```

Now we have all the tags in there is the price requried data.

Now we will extract the text from these tags one by one by looping over these tags

In [36]:

```
for i in price_tags:
    #price=i.text
    Price.append(i.text)
Price[0:10]
```

Out[36]:

```
['56,990',
 '1,24,000',
 '1,13,990',
 '56,990',
 '84,990',
 '95,990',
 '95,890',
 '96,174',
 '85,990',
 '93,990']
```

So,now we have extracted the data requried from the webpage and stored them in the 3 lists mentioned above.Now before creating a dataframe from these lists.Lets first check the length of each of th list. Because if the length of all of the lists are not equal, then a dataframe cannot be formed.

In [38]:

```
print(len(title_names),len(Ratings),len(Price))
```

```
30 32 30
```

In [41]:

```
#import pandas as pd
products=pd.DataFrame({})
products['title_names']=title_names[0:10]
products['Ratings']=Ratings[0:10]
products['Price']=Price[0:10]
```

In [42]:

```
products
```

Out[42]:

	title_names	Ratings	Price
0	Mi Notebook Horizon Edition 14 Intel Core i7-1...		56,990
1	HP Envy Intel 11th Gen Core i7 Processor 13.3 ...		1,24,000
2	ASUS TUF Gaming F17 (2021), 17.3-inch (43.94 c...		1,13,990
3	Mi Notebook Horizon Edition 14 Intel Core i7-1...		56,990
4	HP Pavilion (2021) Intel 11th Gen Core i7 14 i...		84,990
5	ASUS TUF Dash F15 (2021), 15.6-inch (39.62 cms...		95,990
6	Dell 14 (2021) i7-1195G7 2in1 Touch Screen Lap...		95,890
7	Dell 15 (2021) i7-10870H Gaming Laptop, 16GB D...		96,174
8	Lenovo ThinkBook 15 Intel 11th Gen Core i7 15....		85,990
9	Acer Nitro 5 AN515-57 Gaming Laptop	Intel Co...		93,990

Q9: Write a python program to scrape data for first 10 job results for Data Scientist Designation in Noida location. You have to scrape company name, No. of days ago when job was posted, Rating of the company.

In [43]:

```
import warnings
warnings.filterwarnings('ignore')
```

In [44]:

```
# Lets now import all the requried librares
import selenium
import pandas as pd
from selenium import webdriver
```

In [45]:

```
# lets first connect the webdriver
driver=webdriver.Chrome(r"C:\Users\91852\Dropbox\My PC (LAPTOP-APF4NTSI)\Downloads\chrome
driver_win32/chromedriver.exe")
```

In [46]:

```
# driver.close()
```

In [47]:

```
driver.get('https://www.ambitionbox.com/jobs')
```

In [48]:

```
# finding element for job search bar
search_job = driver.find_element_by_xpath('/html/body/div/div/div/div[2]/div[1]/div/div/d
iv/div/span/input')
```

In [49]:

```
search_job.send_keys("Data Scientist")
#search bar
```

In [50]:

```
search_btn = driver.find_element_by_xpath("//span[@class='ctas-btn-medium']")
search_btn.click()
```

In [51]:

```
#specifing the url of the webpage to be scraped
url="https://www.ambitionbox.com/jobs/search?tag=Data%20Scientist&location=Noida"
```

In [52]:

```
# lets open the webpage through our web driver
driver.get(url)
```

Now we will see in the window opened by webdriver the webpage has opened or not.lets check it

So, Now lets first create 10 empty lists.In these lists the data will be stored while scraping.We have created 10 empty lists for 3 features which we have to extract

1.company_name 2.job_posted 3.rating_company.

In [53]:

```
company_names=[]
job_posted=[]
ratings_company=[]
```

First,we will extract all the tags where we have the company names.Let me first show you on the webpage in which tags the job titles are put.

In [54]:

```
# so Lets extract all the tags having the company-names
companies_tags=driver.find_elements_by_xpath("//a[@class='title noclick']")
companies_tags[0:10]
```

Out[54]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="b1b0f6de-73b1-48a3-9b23-23923c65f3aa")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="c829ab25-1058-47ea-a0ca-0c3bca984fb1")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="31e4dfa2-52a8-487a-9e0c-696ee16594b7")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="a48a507e-b090-475c-b26d-d1ec8fd3c4a1")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="5cadb448-0e78-48ad-b9b0-fff24aa4c6c2")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="6f223a2e-63c8-419e-81b2-d97405765748")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="f54d4163-9f10-4e8f-b416-201667208ff5")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="fa3e7ab7-f13c-4bb5-ac20-af49c9a59698")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="ac30f1a3-8bf8-4f2b-a142-c89f70ade8fa")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="7e78e334-bef3-47a9-b238-9eb4424854b4")>]
```

Now we have all the tags in which there are the company names.

Now we will extract the text from these tags one by one by looping over these tags

In [55]:

```
# Now the text of the company names is inside the tags extracted above.
# so we will run a loop to iterate over the tags extracted above and extract the tags

for i in companies_tags:
    companies=i.text
    company_names.append(i.text)
company_names[0:10]
```

Out[55]:

Out[55]:

```
['Data Scientist',
 'Urgent Requirement || Data Scientist || Noida',
 'Group Lead-Data Scientist',
 'Jubilant FoodWorks - Data Scientist - Deep Learning (2-6 yrs)',
 'Urgent Vacancy || Data Scientist || Noida',
 'Urgent Requirement || Data Scientist || Noida',
 'Senior Data Scientist',
 'Manager - Data Scientist - Retail/BFSI (8-15 yrs)',
 'Data Scientist - Machine Learning (5-14 yrs)',
 'Data Scientist - Data Science/Model Development (0-6 yrs)']
```

Now we will extract all the html tags where we have the job post data. Let me first show you in which tags this data is put on the webpage.

In [56]:

```
# so Lets extract all the tags having the job posted data.
job_post=driver.find_elements_by_xpath("//span[@class='body-small-l']")
job_post[0:10]
```

Out[56]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="5ff888a3-37c5-4405-a1c4-944357991d53")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="7ecd81f0-5e53-4fdd-8a2b-b9977284bf31")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="d41ca21c-1238-440a-a59e-3375cc1a7ea7")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="9c47bcb5-5bcf-4aa7-af43-ee5b10987cc9")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="5600c5b9-b19e-4cb3-957d-443d20f02167")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="37200252-6b0b-4c80-bed4-3157531bd357")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="31da8e07-8035-403e-94bb-888523c4b40d")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="dbdb6f11-15f6-4d75-bcff-edcca7feb1f9")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="61351354-f58e-47e5-87a5-05919bea7595")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="d0fe9120-1bb2-43c1-83e0-79b107b1ff84")>]
```

Now we have all the tags in there is the job posted data.

Now we will extract the text from these tags one by one by looping over these tags

In [57]:

```
for i in job_post:
    job=i.text
    job_posted.append(i.text)
job_posted[0:10]
```

Out[57]:

```
['12d ago',
 'via naukri.com',
 '5d ago',
 'via naukri.com',
 '7d ago',
 'via naukri.com',
 '4d ago',
 'via hirist.com',
 '24d ago',
 'via naukri.com']
```

Now we will extract all the html tags where we have the rating companies data. Let me first show you in which tags this data is put on the webpage

tags this data is put on the webpage.

In [58]:

```
ratings_companies=driver.find_elements_by_xpath("//span[@class='body-medium']")
ratings_companies[0:10]
```

Out[58]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="f95a7339-e705-443e-aa38-21e0e044f1dd")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="59ec7c85-7ce5-4b3a-9b50-211457502e2d")>,
 <selenium.webdriver.remote.webelement.WebElement (session="f5b9e013716779bd1195d899d8e7d
285", element="9bad23f8-1506-4203-80be-3ebd43ce2914")>]
```

Now we have all the tags in which there is the data about the rating_company.

Now we will extract the text from these tags one by one by looping over these tags

In [67]:

```
for i in ratings_companies:
    ratings=i.text
    ratings_company.append(i.text)
ratings_company[0:10]
```

Out[67]:

```
['4.1',
 '₹ 11L',
 '₹ 22L',
 '4.1',
 '₹ 11L',
 '₹ 22L',
 '4.1',
 '₹ 11L',
 '₹ 22L',
 '4.1']
```

In [68]:

```
print(len(company_names),len(job_posted),len(ratings_company))
```

```
10 20 12
```

In [69]:

```
#import pandas as pd
jobs=pd.DataFrame({})
jobs['company_names']=company_names[0:10]
jobs['job_posted']=job_posted[0:10]
jobs['ratings_company']=ratings_company[0:10]
```

In [70]:

```
jobs
```

Out[70]:

	company_names	job_posted	ratings_company
0	Data Scientist	12d ago	4.1
1	Urgent Requirement || Data Scientist || Noida	via naukri.com	₹ 11L
2	Group Lead-Data Scientist	5d ago	₹ 22L
3	Jubilant FoodWorks - Data Scientist - Deep Lea...	via naukri.com	4.1
4	Urgent Vacancy || Data Scientist || Noida	7d ago	₹ 11L
5	Urgent Requirement || Data Scientist || Noida	via naukri.com	₹ 22L

	company_names	job_posted	ratings_company
6	Senior Data Scientist	1d ago	4.1
7	Manager - Data Scientist - Retail/BFSI (8-15 yrs)	via hirist.com	₹ 11L
8	Data Scientist - Machine Learning (5-14 yrs)	24d ago	₹ 22L
9	Data Scientist - Data Science/Model Developmen...	via naukri.com	4.1

In [ ]: