

```
In [1]: ! pip install selenium
```

```
Requirement already satisfied: selenium in c:\users\gamy\anaconda3\lib\site-packages (4.1.0)
```

```
WARNING: You are using pip version 21.0.1; however, version 21.3.1 is available.
```

```
Requirement already satisfied: trio-websocket~=0.9 in c:\users\gamy\anaconda3\lib\site-packages (from selenium) (0.9.2)
```

```
Requirement already satisfied: trio~=0.17 in c:\users\gamy\anaconda3\lib\site-packages (from selenium) (0.19.0)
```

```
Requirement already satisfied: urllib3[secure]~=1.26 in c:\users\gamy\anaconda3\lib\site-packages (from selenium) (1.26.7)
```

```
Requirement already satisfied: sortedcontainers in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (2.2.2)
```

```
Requirement already satisfied: idna in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (2.10)
```

```
Requirement already satisfied: cffi>=1.14 in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.14.0)
```

```
Requirement already satisfied: async-generator>=1.9 in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.10)
```

```
Requirement already satisfied: sniffio in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.2.0)
```

```
Requirement already satisfied: outcome in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.1.0)
```

```
Requirement already satisfied: attrs>=19.2.0 in c:\users\gamy\anaconda3\lib\site-packages (from trio~=0.17->selenium) (19.3.0)
```

```
Requirement already satisfied: pycparser in c:\users\gamy\anaconda3\lib\site-packages (from cffi>=1.14->trio~=0.17->selenium) (2.20)
```

```
Requirement already satisfied: wsproto>=0.14 in c:\users\gamy\anaconda3\lib\site-packages (from trio-websocket~=0.9->selenium) (1.0.0)
```

```
Requirement already satisfied: certifi in c:\users\gamy\anaconda3\lib\site-packages (from urllib3[secure]~=1.26->selenium) (2020.6.20)
```

```
Requirement already satisfied: cryptography>=1.3.4 in c:\users\gamy\anaconda3\lib\site-packages (from urllib3[secure]~=1.26->selenium) (2.9.2)
```

```
Requirement already satisfied: pyOpenSSL>=0.14 in c:\users\gamy\anaconda3\lib\site-packages (from urllib3[secure]~=1.26->selenium) (19.1.0)
```

```
You should consider upgrading via the 'c:\users\gamy\anaconda3\python.exe -m pip install --upgrade pip' command.
```

```
In [42]: #DownLoading web driver  
#import all the required Libraries  
import selenium  
import pandas as pd  
from selenium import webdriver  
  
import warnings  
warnings.filterwarnings('ignore')
```

```
In [253]: #Let's first connect to web driver
driver=webdriver.Chrome(r"C:/Users/Gamya/Downloads/chromedriver_win32/chromedriver")
```

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

```
In [62]: search_job=driver.find_element_by_id('qsb-keyword-sugg')
search_job.send_keys("Data Analyst")
search_loc=driver.find_element_by_xpath("//input[@id='qsb-location-sugg']")
search_loc.send_keys("Bangalore")
```

```
In [63]: search_btn=driver.find_element_by_xpath("//div[@class='search-btn']/button")
search_btn.click()
```

```
In [64]: titlelist=[]
companylist=[]
locationlist=[]
expilist=[]
```

```
In [65]: ts=driver.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
ts[0:2]
```

```
Out[65]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="362d7549-b102-42bb-a51c-81f66d1e5a7b")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="582c68c2-f5a4-4f5f-a423-b9f23ed65933")>]
```

```
In [66]: t_new=[]
t_new=ts[0:10]
t_new[0:2]
```

```
Out[66]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="362d7549-b102-42bb-a51c-81f66d1e5a7b")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="582c68c2-f5a4-4f5f-a423-b9f23ed65933")>]
```

```
In [67]: for i in t_new:
    t=i.text
    titlelist.append(t)
titlelist[0:2]
```

```
Out[67]: ['Hiring -Data Analyst, Business Analyst, MIS Analyst -3rd Party Payroll',
          'Business Analyst/ Data Analyst- Capital Market & SQL (Mandatory)']
```

```
In [68]: com_tags=driver.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
com_tags[0:2]
```

```
Out[68]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="8a578b3f-80ac-40c2-931e-cb2e732d1d92")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="014b1f16-9fc1-45cc-a3fb-76081643ec87")>]
```

```
In [69]: c_new=com_tags[0:10]
c_new[0:2]
```

```
Out[69]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="8a578b3f-80ac-40c2-931e-cb2e732d1d92")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="014b1f16-9fc1-45cc-a3fb-76081643ec87")>]
```

```
In [70]: for i in c_new:
    c=i.text
    companylist.append(c)
companylist[0:2]
```

```
Out[70]: ['Flipkart', 'Genpact']
```

```
In [71]: exp_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHo
exp_tags[0:2]
```

```
Out[71]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="3a59cf50-6c36-4841-9662-3799ef342e35")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="c078090c-fdb4-4fd7-995e-c2566e3e7800")>]
```

```
In [72]: e_new=exp_tags[0:10]
e_new[0:2]
```

```
Out[72]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="3a59cf50-6c36-4841-9662-3799ef342e35")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="c078090c-fdb4-4fd7-995e-c2566e3e7800")>]
```

```
In [73]: for i in e_new:  
    e=i.text  
    expilist.append(e)  
  
    expilist[0:2]
```

```
Out[73]: ['1-6 Yrs', '7-12 Yrs']
```

```
In [74]: loc_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHo  
loc_tags[0:2]
```

```
Out[74]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="b4a09400-1304-44e8-8acd-464f26f00e92")>,  
 <selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="053a899c-a5d4-447f-9772-c7ce4bf4bbb5")>]
```

```
In [75]: l_new=loc_tags[0:10]  
l_new[0:2]
```

```
Out[75]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="b4a09400-1304-44e8-8acd-464f26f00e92")>,  
 <selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="053a899c-a5d4-447f-9772-c7ce4bf4bbb5")>]
```

```
In [76]: for j in l_new:  
    lo=j.text  
    locationlist.append(lo)  
locationlist[0:2]
```

```
Out[76]: ['Bangalore/Bengaluru', 'Pune, Bangalore/Bengaluru']
```

```
In [77]: print(len(titlelist),len(companylist),len(expilist),len(locationlist))
```

```
10 10 10 10
```

```
In [78]: jobs=pd.DataFrame({})  
jobs['title']=titlelist  
jobs['company']=companylist  
jobs['experinece_required']=expilist  
jobs['location']=locationlist
```

In [79]: jobs

Out[79]:

	title	company	experience_required	location
0	Hiring -Data Analyst, Business Analyst, MIS An...	Flipkart	1-6 Yrs	Bangalore/Bengaluru
1	Business Analyst/ Data Analyst- Capital Market...	Genpact	7-12 Yrs	Pune, Bangalore/Bengaluru
2	Data Analyst	GlaxoSmithKline Pharmaceuticals Limited	3-8 Yrs	Bangalore/Bengaluru
3	Data Analyst	Schneider Electric	2-5 Yrs	Bangalore/Bengaluru
4	Data Analyst	GlaxoSmithKline Pharmaceuticals Limited	3-8 Yrs	Bangalore/Bengaluru
5	Senior Data Analyst	Flipkart	3-7 Yrs	Bangalore/Bengaluru
6	Senior Data Analyst for HANA Platform	Intel	5-10 Yrs	Bangalore/Bengaluru
7	Data Analyst	NetraDyne Technology India Private Limited	2-5 Yrs	Bangalore/Bengaluru
8	Assistant/deputy Manager - Geo-spatial Data An...	Maruti Suzuki India	3-5 Yrs	Gurgaon/Gurugram, bangalore
9	Data Analyst	Slice	0-2 Yrs	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.
3. Then click the search button.
4. Then scrape the data for the first 10 jobs results you get.
5. Finally create a dataframe of the scraped data. Note: All of the above steps have to be done in code. No step is to be done manually.

In [80]: driver.get('https://www.naukri.com/')

In [81]: search_job=driver.find_element_by_id('qsb-keyword-sugg')
search_job.send_keys("Data Scientist")
search_loc=driver.find_element_by_xpath("//input[@id='qsb-location-sugg']")
search_loc.send_keys("Bangalore")

In [82]: search_btn=driver.find_element_by_xpath("//div[@class='search-btn']/button")
search_btn.click()

```
In [83]: title=[]
company=[]
location=[]
```

```
In [84]: t_t=driver.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
t_t[0:2]
```

```
Out[84]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="3a9e1821-7e4d-4028-b021-1628af82b9d4")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="d5d5aea0-8fef-4e53-8879-e400996e95ec")>]
```

```
In [85]: t_t_new=t_t[0:10]
t_t_new[0:2]
```

```
Out[85]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="3a9e1821-7e4d-4028-b021-1628af82b9d4")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="d5d5aea0-8fef-4e53-8879-e400996e95ec")>]
```

```
In [86]: for i in t_t_new:
    a=i.text
    title.append(a)
title
```

```
Out[86]: ['Data Scientist: Advanced Analytics',
'Deputy Manager - Data Scientist',
'Tcs Hiring For Azure ML Data Scientist',
'Senior Engineer - AIML - Data Scientist',
'Senior data scientist',
'Cognitive/AI Senior Data Scientist',
'Senior Data Scientist',
'Lead/Senior Data Scientist (NLP)',
'Data scientist',
'Data Scientist']
```

```
In [87]: c_t=driver.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
c_t[0:2]
```

```
Out[87]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="8c7e927b-3fbf-4dc2-85e5-895bf9ee6d3c")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="275997bc-dc38-4d0f-a9a5-e9f27d431dff")>]
```

```
In [88]: c_t_new=c_t[0:10]
c_t_new[0:2]
```

```
Out[88]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="8c7e927b-3fbf-4dc2-85e5-895bf9ee6d3c")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="275997bc-dc38-4d0f-a9a5-e9f27d431dff")>]
```

```
In [89]: for i in c_t_new:
    b=i.text
    company.append(b)
company
```

```
Out[89]: ['IBM India Pvt. Limited',
 'HDFC LIFE INSURANCE COMPANY LIMITED',
 'Tata Consultancy Services Ltd.',
 'Unisys',
 'Compile Inc',
 'IBM India Pvt. Limited',
 'Visa',
 'Samya.AI A FRACTAL Entity',
 'Superior Group',
 'Thoucentric Technology Pvt ltd']
```

```
In [90]: l_t=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHolderLi']")
l_t[0:2]
```

```
Out[90]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="63ded5b1-7a6c-4b64-9ab5-c3bd0cf5ba47")>,
 <selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="81e738d1-c959-4dfc-8b80-f0449bc830c8")>]
```

```
In [91]: l_t_new=l_t[0:10]
l_t_new[0:2]
```

```
Out[91]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="63ded5b1-7a6c-4b64-9ab5-c3bd0cf5ba47")>,
 <selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="81e738d1-c959-4dfc-8b80-f0449bc830c8")>]
```

```
In [92]: for i in l_t_new:
    c=i.text
    location.append(c)
location
```

```
Out[92]: ['Bengaluru/Bangalore',
 'Bangalore/Bengaluru',
 'Hyderabad/Secunderabad, Bangalore/Bengaluru, Mumbai (All Areas)',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru',
 'Bengaluru/Bangalore',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru\n(WFH during Covid)',
 'Bangalore/Bengaluru',
 'Bangalore/Bengaluru']
```

```
In [93]: print(len(title),len(company),len(location))
```

10 10 10

```
In [94]: jobs1=pd.DataFrame({})
jobs1['title']=title
jobs1['company']=company
jobs1['location']=location
jobs1
```

Out[94]:

	title	company	location
0	Data Scientist: Advanced Analytics	IBM India Pvt. Limited	Bengaluru/Bangalore
1	Deputy Manager - Data Scientist	HDFC LIFE INSURANCE COMPANY LIMITED	Bangalore/Bengaluru
2	Tcs Hiring For Azure ML Data Scientist	Tata Consultancy Services Ltd.	Hyderabad/Secunderabad, Bengaluru/Bengaluru, M...
3	Senior Engineer - AIML - Data Scientist	Unisys	Bangalore/Bengaluru
4	Senior data scientist	Compile Inc	Bangalore/Bengaluru
5	Cognitive/AI Senior Data Scientist	IBM India Pvt. Limited	Bengaluru/Bangalore
6	Senior Data Scientist	Visa	Bangalore/Bengaluru
7	Lead/Senior Data Scientist (NLP)	Samya.AI A FRACTAL Entity	Bangalore/Bengaluru\n(WFH during Covid)
8	Data scientist	Superior Group	Bangalore/Bengaluru
9	Data Scientist	Thoucentric Technology Pvt Ltd	Bangalore/Bengaluru

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. Note: All of the above steps have to be done in code. No step is to be done manually.

```
In [95]: driver.get('https://www.naukri.com/')
```

```
In [96]: search_job=driver.find_element_by_id('qsb-keyword-sugg')
search_job.send_keys("Data Scientist")
```

```
In [97]: search_btn=driver.find_element_by_xpath("//div[@class='search-btn']/button")
search_btn.click()
```

```
In [98]: delhi=driver.find_element_by_xpath("//span[@title='Delhi / NCR']")
delhi.click()
```

```
In [99]: sal=driver.find_element_by_xpath("//span[@title='3-6 Lakhs']")
sal.click()
```

```
In [100]: job3=[]
loc3=[]
com3=[]
exp3=[]
```

```
In [101]: j_tag=driver.find_elements_by_xpath("//a[@class='title fw500 ellipsis']")
j_tag[0:2]
```

```
Out[101]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="6d031e08-4e1b-4937-aac6-1ce99cd9595a")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="182451d0-eb01-47ee-9543-c56cad9b122d")>]
```

```
In [102]: j_tag_new=j_tag[0:10]
j_tag_new[0:2]
```

```
Out[102]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="6d031e08-4e1b-4937-aac6-1ce99cd9595a")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="182451d0-eb01-47ee-9543-c56cad9b122d")>]
```

```
In [103]: for i in j_tag_new:
    d=i.text
    job3.append(d)
job3
```

```
Out[103]: ['Hiring For Data Scientist',
'Data Scientist',
'NTT DATA_ Hiring For BIG DATA ,DATA Scientist, Devops',
'Project Manager | Team Leader | Senior Data Scientist',
'Data Scientist',
'Hiring For Data Analyst / Data Scientist',
'Immediate requirement For Data Scientist',
'Data Scientist',
'Data Scientist',
'Data Scientist/ Senior Data Scientist']
```

```
In [104]: c_tag=driver.find_elements_by_xpath("//a[@class='subTitle ellipsis fleft']")
c_tag[0:2]
```

```
Out[104]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="af2d1b52-1ffd-4f51-b7a0-9257ea5e5482")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="0a4e1d1e-e05b-48a4-82ac-603318a18766")>]
```

```
In [105]: c_tag_new=c_tag[0:10]
c_tag_new[0:2]
```

```
Out[105]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="af2d1b52-1ffd-4f51-b7a0-9257ea5e5482")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="0a4e1d1e-e05b-48a4-82ac-603318a18766")>]
```

```
In [106]: for i in c_tag_new:
    e=i.text
    com3.append(e)
com3
```

```
Out[106]: ['Tata Consultancy Services Ltd.',
'LG Electronics India Pvt. Ltd.',
'NTT Data Business Solutions Pvt Ltd',
'Tidyquant (OPC) Private Limited',
'Navikenz India Pvt Ltd',
'Careerera',
'CALIBEHR BUSINESS SUPPORT SERVICES PRIVATE LIMITED',
'Whizhack Technologies pvt ltd',
'One Mobikwik Systems Private Limited',
'Newgen Software Technologies']
```

```
In [107]: l_tag=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeHolder']
l_tag[0:2]
```

```
Out[107]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="1953cf89-5835-4051-a239-9d130f20584a")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="dd2d4c1b-0485-4cad-bac6-acfa76de7ea8")>]
```

```
In [108]: l_tag_new=l_tag[0:10]
l_tag_new[0:2]
```

```
Out[108]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="1953cf89-5835-4051-a239-9d130f20584a")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="dd2d4c1b-0485-4cad-bac6-acfa76de7ea8")>]
```

```
In [109]: for i in l_tag_new:
    loco=i.text
    loc3.append(loco)
loc3
```

```
Out[109]: ['Hyderabad/Secunderabad, Pune, Chennai, Bangalore/Bengaluru, Delhi / NCR, Mumbai (All Areas)\n(WFH during Covid)', 'Noida, New Delhi, Faridabad, Gurgaon/Gurugram, Delhi / NCR', 'Noida, Kolkata, Hyderabad/Secunderabad, Ahmedabad, Chennai, Coimbatore, Bangalore/Bengaluru, Delhi / NCR, Mumbai (All Areas)\n(WFH during Covid)', 'Remote', 'Noida, Noida\n(WFH during Covid)', 'Noida(Sector-59 Noida)', 'Mumbai, Hyderabad/Secunderabad, Pune, Chennai, Tamil Nadu, Bangalore/Bengaluru, Delhi / NCR', 'Gurgaon/Gurugram', 'Gurgaon/Gurugram', 'Noida']
```

```
In [110]: e_tags=driver.find_elements_by_xpath("//li[@class='fleft grey-text br2 placeholder']")  
e_tags[0:2]
```

```
Out[110]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91cc656d4afa9bce", element="476802d5-6194-4b23-8644-8f7c3d5e1990")>, <selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91cc656d4afa9bce", element="1de11f81-286e-4a39-8851-7b60bdc1f75f")>]
```

```
In [111]: e_tags_new=e_tags[0:10]  
e_tags_new[0:2]
```

```
Out[111]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91cc656d4afa9bce", element="476802d5-6194-4b23-8644-8f7c3d5e1990")>, <selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91cc656d4afa9bce", element="1de11f81-286e-4a39-8851-7b60bdc1f75f")>]
```

```
In [112]: for i in e_tags_new:
    f=i.text
    exp3.append(f)
exp3
```

```
Out[112]: ['4-9 Yrs', '0-2 Yrs', '3-8 Yrs', '1-5 Yrs', '2-7 Yrs', '1-3 Yrs', '2-7 Yrs', '2-5 Yrs', '2-5 Yrs', '2-5 Yrs']
```

```
In [113]: print(len(job3),len(com3),len(loc3),len(exp3))
```

```
10 10 10 10
```

```
In [114]: jobs2=pd.DataFrame({})
jobs2['title']=job3
jobs2['company']=com3
jobs2['location']=loc3
jobs2['experience']=exp3
jobs2
```

Out[114]:

	title	company	location	experience
0	Hiring For Data Scientist	Tata Consultancy Services Ltd.	Hyderabad/Secunderabad, Pune, Chennai, Bangalo...	4-9 Yrs
1	Data Scientist	LG Electronics India Pvt. Ltd.	Noida, New Delhi, Faridabad, Gurgaon/Gurugram,...	0-2 Yrs
2	NTT DATA_ Hiring For BIG DATA ,DATA Scientist,...	NTT Data Business Solutions Pvt Ltd	Noida, Kolkata, Hyderabad/Secunderabad, Ahmeda...	3-8 Yrs
3	Project Manager Team Leader Senior Data Sc...	Tidyquant (OPC) Private Limited	Remote	1-5 Yrs
4	Data Scientist	Navikenz India Pvt Ltd	Noida, Noida\nWFH during Covid)	2-7 Yrs
5	Hiring For Data Analyst / Data Scientist	Careerera	Noida(Sector-59 Noida)	1-3 Yrs
6	Immediate requirement For Data Scientist	CALIBEHR BUSINESS SUPPORT SERVICES PRIVATE LIM...	Mumbai, Hyderabad/Secunderabad, Pune, Chennai,...	2-7 Yrs
7	Data Scientist	Whizhack Technologies pvt ltd	Gurgaon/Gurugram	2-5 Yrs
8	Data Scientist	One Mobikwik Systems Private Limited	Gurgaon/Gurugram	2-5 Yrs
9	Data Scientist/ Senior Data Scientist	Newgen Software Technologies	Noida	2-5 Yrs

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. To scrape the data you have to go through following steps:
4. Go to Flipkart webpage by url : <https://www.flipkart.com/> (<https://www.flipkart.com/>)
5. Enter "sunglasses" in the search field where "search for products, brands and more" is written and click the search icon
6. After that you will reach to the page having a lot of sunglasses. From this page you can scrap the required data as usual.
7. After scraping data from the first page, go to the "Next" Button at the bottom of the page , then click on it.
8. Now scrape data from this page as usual

9. Repeat this until you get data for 100 sunglasses. Note: That all of the above steps have to be done by coding only and not manually.

```
In [115]: driver.get('https://www.flipkart.com/')
```

```
In [116]: sunglasses=driver.find_element_by_xpath("//input[@placeholder='Search for product sunglasses'].send_keys("sunglasses")
```

```
In [118]: search=driver.find_element_by_xpath("//button[@class='L0Z3Pu']").click()
```

```
In [119]: brand=[]
des=[]
price=[ ]
```

```
In [120]: b_tag=driver.find_elements_by_xpath("//div[@class='_2WkVRV']").b_tag[0:2]
```

```
Out[120]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="c6c49349-cb3f-4cc6-87ef-333529d3fb2d">,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="4e222d60-4150-465a-b196-5111716872d6">]
```

```
In [121]: for i in b_tag:
    br=i.text
    brand.append(br)
brand[0:2]
```

```
Out[121]: ['Kanishka', 'Kanishka']
```

```
In [122]: d_tag=driver.find_elements_by_xpath("//a[@class='IRpwTa']").d_tag[0:2]
```

```
Out[122]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="191b08df-44f0-45e5-8a7e-19f5862c3473">,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="e8dd074a-4741-4e04-a600-cf9486adc5f0">]
```

```
In [123]: for i in d_tag:
    de=i.text
    des.append(de)
des[0:2]
```

```
Out[123]: ['UV Protection Aviator Sunglasses (55)', 'Others Round Sunglasses (55)']
```

```
In [124]: des.append('Riding Glasses, UV Protection Oval Sunglasses (52)')
```

```
In [125]: p_tag=driver.find_elements_by_xpath("//div[@class=' _30jeq3 ']")
p_tag[0:2]
```

```
Out[125]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="41e4dc25-5508-4df9-a016-f5c263ac3255")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="6921005a-2da3-4344-9d10-f9af807ff0a9")>]
```

```
In [126]: for i in p_tag:
    pr=i.text
    price.append(pr)
price[0:2]
```

```
Out[126]: ['₹165', '₹151']
```

```
In [127]: nextt=driver.find_element_by_xpath("//a[@class=' _1LKTO3 ']")
nextt.click()
```

```
In [128]: d_tag=driver.find_elements_by_xpath("//a[@class='IRpwTa ']")
d_tag[0:2]
```

```
Out[128]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="97f17728-ade8-4445-9708-62a6b2920dab")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="f92eecccd-ac44-49e4-a858-9e482e17d24c")>]
```

```
In [129]: for i in d_tag:
    de=i.text
    des.append(de)
des[0:2]
```

```
Out[129]: ['UV Protection Aviator Sunglasses (55)', 'Others Round Sunglasses (55)']
```

```
In [130]: des.append('UV Protection Round Sunglasses (Free Size)')
```

```
In [131]: des.append('Polarized, UV Protection Round Sunglasses (53)')
```

```
In [132]: b_tag=driver.find_elements_by_xpath("//div[@class=' _2WkVRV ']")
b_tag[0:2]
```

```
Out[132]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="5e0a0d8b-cf78-4a78-b826-09b9c1b5bc01")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="e571b110-dbb5-4a5b-a76a-1c11089c18f7")>]
```

```
In [133]: for i in b_tag:
    br=i.text
    brand.append(br)
brand[0:2]
```

Out[133]: ['Kanishka', 'Kanishka']

```
In [134]: p_tag=driver.find_elements_by_xpath("//div[@class=' _30jeq3 ']")
p_tag[0:2]
```

Out[134]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="c07888c0-cfa1-4efd-93c2-35d0b2bbdd9b")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="c867fc19-c5cd-43a7-ac0b-433d46735171")>]

```
In [135]: for i in p_tag:
    pr=i.text
    price.append(pr)
price[0:2]
```

Out[135]: ['₹165', '₹151']

```
In [136]: next2=driver.find_element_by_xpath("//a[@class=' _1LKTO3 ']")
next2.click()
```

```
In [137]: b_tag=driver.find_elements_by_xpath("//div[@class=' _2WkVRV ']")
b_tag[0:2]
```

Out[137]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="d1d18833-db1d-49b6-bc4c-61d4d9011e11")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="bab2a919-eb1f-4664-91cb-9728bb00a96d")>]

```
In [138]: b_tag_new=b_tag[0:20]
```

```
In [139]: for i in b_tag_new:
    br=i.text
    brand.append(br)
brand[0:2]
```

Out[139]: ['Kanishka', 'Kanishka']

```
In [140]: d_tag=driver.find_elements_by_xpath("//a[@class='IRpwTa ']")
d_tag[0:2]
```

Out[140]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="b4b86f6f-fa9-4726-9f31-6fafb740e645")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c c656d4afa9bce", element="274500a0-2c2e-447f-a1fe-c18d1514b90b")>]

```
In [141]: d_tag_new=d_tag[0:20]
```

```
In [142]: for i in d_tag_new:
    de=i.text
    des.append(de)
des[0:2]
```

```
Out[142]: ['UV Protection Aviator Sunglasses (55)', 'Others Round Sunglasses (55)']
```

```
In [143]: len(des)
```

```
Out[143]: 100
```

```
In [144]: p_tag=driver.find_elements_by_xpath("//div[@class=' _30jeq3']")
p_tag[0:2]
```

```
Out[144]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="f574eddd-fc05-4fd4-b2b7-c9b6b831ad29")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="3ace064d-6d0d-40ed-9bd5-86d5d5b20cf7")>]
```

```
In [145]: p_tag_new=p_tag[0:20]
```

```
In [146]: for i in p_tag_new:
    pr=i.text
    price.append(pr)
price[0:2]
```

```
Out[146]: ['₹165', '₹151']
```

```
In [147]: print(len(brand),len(des),len(price))
```

```
100 100 100
```

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

<https://www.flipkart.com/apple-iphone-11-black-64-gb-includes-earpods-poweradapter/p/itm0f37c2240b217?pid=MOBFKCTSVZAXUHGR&lid=LSTMOBFKC TSVZAXUHGREPBFGI&marketplace>

As shown in the above page you have to scrape the tick marked attributes. These are:

1. Rating
2. Review_summary
3. Full review
4. You have to scrape this data for first 100 reviews. Note: All the steps required during scraping should be done through code only and not manually.

```
In [148]: driver.get('https://www.flipkart.com/apple-iphone-11-black-64-gb-includes-earpods-poweradapter/p/itm0f37c2240b217?pid=MOBFKCTSVZAXUHGR&lid=LSTMOBFKC TSVZAXUHGREPBFGI&marketplace')
```

```
In [149]: rate_list=[]
```

```
In [150]: rate_tag=driver.find_elements_by_xpath("//div[@class='_3LWZlK _1BLPMq']")  
rate_tag[0:2]
```

```
Out[150]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="4249f32d-c1f3-46cd-ba48-bea208e3c029")>,  
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="35a09e49-1d73-451d-b700-c5e12d47a944")>]
```

```
In [151]: rev_tag=driver.find_elements_by_xpath("//p[@class='_2-N8zT']")  
rev_tag[0:2]
```

```
Out[151]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="0fe18682-fa58-4710-80a3-1718de2a30ef")>,  
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="0df6f013-861d-4c0f-80e1-e9189643f9f9")>]
```

```
In [152]: revv_tag=driver.find_elements_by_xpath("//div[@class='t-ZTKy']")  
revv_tag[0:2]
```

```
Out[152]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="2aee30cd-22d2-4f63-bfc9-e95907a8dd43")>,  
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="c9c2bbd5-e04a-499c-96f3-4bc26e9fb0")>]
```

```
In [153]: ratings=[]  
for i in rate_tag:  
    ratings.append(i.text)  
ratings
```

```
Out[153]: ['5', '5', '5', '5', '5', '5', '5', '5', '4', '5', '5']
```

```
In [154]: review=[]  
for i in rev_tag:  
    review.append(i.text)  
review[0:2]
```

```
Out[154]: ['Brilliant', 'Simply awesome']
```

```
In [155]: b_review=[]  
for i in revv_tag:  
    b_review.append(i.text)  
b_review[0:2]
```

```
Out[155]: ['The Best Phone for the Money\n\nThe iPhone 11 offers superb cameras, a more durable design and excellent battery life for an affordable price.\n\nCompelling ultra-wide camera\nNew Night mode is excellent\nLong battery life',  
'Really satisfied with the Product I received... It's totally genuine and the packaging was also really good so if ur planning to buy just go for it.']}
```

```
In [156]: next=driver.find_element_by_xpath("//a[@class='__1LKTO3']")
next.click()
```

```
In [157]: rate_tag=driver.find_elements_by_xpath("//div[@class='__3LWZlK __1BLPMq']")
rate_tag[0:2]
```

```
Out[157]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="4249f32d-c1f3-46cd-ba48-bea208e3c029")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="35a09e49-1d73-451d-b700-c5e12d47a944")>]
```

```
In [158]: for i in rate_tag:
    ratings.append(i.text)
ratings[0:2]
```

```
Out[158]: ['5', '5']
```

```
In [159]: rev_tag=driver.find_elements_by_xpath("//p[@class='__2-N8zT']")
rev_tag[0:2]
```

```
Out[159]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="0fe18682-fa58-4710-80a3-1718de2a30ef")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="0df6f013-861d-4c0f-80e1-e9189643f9f9")>]
```

```
In [160]: revv_tag=driver.find_elements_by_xpath("//div[@class='t-ZTKy']")
revv_tag[0:2]
```

```
Out[160]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="2aee30cd-22d2-4f63-bfc9-e95907a8dd43")>,
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c
c656d4afa9bce", element="c9c2bb5-e04a-499c-96f3-4bc26e9fbbc0")>]
```

```
In [161]: for i in revv_tag:
    review.append(i.text)
review[0:2]
```

```
Out[161]: ['Brilliant', 'Simply awesome']
```

```
In [162]: for i in revv_tag:
    b_review.append(i.text)
b_review[0:2]
```

```
Out[162]: ['The Best Phone for the Money\n\nThe iPhone 11 offers superb cameras, a more durable design and excellent battery life for an affordable price.\n\nCompelling ultra-wide camera\nNew Night mode is excellent\nLong battery life',
'Really satisfied with the Product I received... It's totally genuine and the packaging was also really good so if ur planning to buy just go for it.']}
```

```
In [163]: next=driver.find_element_by_xpath("//a[@class='ge-49M']")
next.click()
```

```
In [164]: rate_tag=driver.find_elements_by_xpath("//div[@class='_3LWZlK _1BLPMq']")  
rev_tag=driver.find_elements_by_xpath("//p[@class='_2-N8zT']")  
revv_tag=driver.find_elements_by_xpath("//div[@class='t-ZTKy']")
```

```
In [165]: for i in rate_tag:  
    ratings.append(i.text)
```

```
In [166]: for i in rev_tag:  
    review.append(i.text)
```

```
In [167]: for i in revv_tag:  
    b_review.append(i.text)
```

```
In [168]: driver.get('https://www.flipkart.com/apple-iphone-11-black-64-gb-includes-earpods')  
◀ ▶
```

```
In [169]: counter=0  
while counter<7:  
    rate_tag=driver.find_elements_by_xpath("//div[@class='_3LWZlK _1BLPMq']")  
    for i in rate_tag:  
        ratings.append(i.text)  
    rev_tag=driver.find_elements_by_xpath("//p[@class='_2-N8zT']")  
    for i in rev_tag:  
        review.append(i.text)  
    revv_tag=driver.find_elements_by_xpath("//div[@class='t-ZTKy']")  
    for i in revv_tag:  
        b_review.append(i.text)  
    counter+=1  
  
next1=driver.find_element_by_xpath("//nav[@class='yFHi8N']")  
next1.click()
```

```
In [170]: ratings[0:2]
```

```
Out[170]: ['5', '5']
```

```
In [171]: review[0:2]
```

```
Out[171]: ['Brilliant', 'Simply awesome']
```

```
In [172]: b_review[0:2]
```

```
Out[172]: ['The Best Phone for the Money\n\nThe iPhone 11 offers superb cameras, a more durable design and excellent battery life for an affordable price.\n\nCompelling ultra-wide camera\nNew Night mode is excellent\nLong battery life', 'Really satisfied with the Product I received... It's totally genuine and the packaging was also really good so if ur planning to buy just go for it.]
```

```
In [173]: print(len(ratings),len(review),len(b_review))
```

```
100 100 100
```

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 As shown in the below image, you have to scrape the tick marked attributes.

```
In [191]: driver.get('https://www.flipkart.com/')
```

```
In [192]: sne=driver.find_element_by_xpath("//input[@placeholder='Search for products, brands']")
sne.send_keys("sneakers")
```

```
In [193]: search=driver.find_element_by_xpath("//button[@class='L0Z3Pu']")
search.click()
```

```
In [194]: br=driver.find_elements_by_xpath("//div[@class='_2WkVRV']")
des=driver.find_elements_by_xpath("//a[@class='IRpwTa']")
pri=driver.find_elements_by_xpath("//div[@class='_30jeq3']")
o=driver.find_elements_by_xpath("//div[@class='_3Ay6Sb']")
```

```
In [195]: brand=[]
for i in br:
    brand.append(i.text)
desp=[]
for i in des:
    desp.append(i.text)
price=[]
for i in pri:
    price.append(i.text)
off=[]
for i in o:
    off.append(i.text)
```

```
In [196]: next=driver.find_element_by_xpath("//a[@class='_1LKTO3']")
next.click()
```

```
In [197]: br=driver.find_elements_by_xpath("//div[@class=' _2WkVRV ']")  
des=driver.find_elements_by_xpath("//a[@class='IRpwTa ']")  
pri=driver.find_elements_by_xpath("//div[@class=' _30jeq3 ']")  
o=driver.find_elements_by_xpath("//div[@class=' _3Ay6Sb ']")
```

```
In [198]: for i in br:  
    brand.append(i.text)  
for i in des:  
    desp.append(i.text)  
for i in o:  
    off.append(i.text)  
for i in pri:  
    price.append(i.text)
```

```
In [199]: next=driver.find_element_by_xpath("//a[@class=' _1LKTO3 ']")  
next.click()
```

```
In [200]: br=driver.find_elements_by_xpath("//div[@class=' _2WkVRV ']")  
des=driver.find_elements_by_xpath("//a[@class='IRpwTa ']")  
pri=driver.find_elements_by_xpath("//div[@class=' _30jeq3 ']")  
o=driver.find_elements_by_xpath("//div[@class=' _3Ay6Sb ']")
```

```
In [201]: br=br[0:20]  
des=des[0:35]  
o=o[0:20]  
pri=pri[0:21]
```

```
In [202]: for i in br:  
    brand.append(i.text)  
for i in des:  
    desp.append(i.text)  
for i in pri:  
    price.append(i.text)  
for i in o:  
    off.append(i.text)
```

```
In [203]: brand[0:2]
```

```
Out[203]: ['PUMA', 'BIRDE']
```

```
In [204]: desp[0:2]
```

```
Out[204]: ['Serve Pro Lite Sneakers For Men',  
          'Casual Shoes For Men Combo Pack Of 2 Sneakers For Men']
```

```
In [205]: price[0:2]
```

```
Out[205]: ['₹2,699', '₹535']
```

```
In [206]: off[0:2]
```

```
Out[206]: ['40% off', '46% off']
```

```
In [207]: print(len(brand),len(desp),len(price),len(off))
```

```
100 103 101 99
```

Q7: Go to the link - <https://www.myntra.com/shoes> (https://www.myntra.com/shoes) Set Price filter to "Rs. 6649 to Rs. 13099" , Color filter to "Black", as shown in the below image. And then scrape First 100 shoes data you get. The data should include "Brand" of the shoes , Short Shoe description, price of the shoe as shown in the below image.

```
In [208]: driver.get('https://www.myntra.com/shoes')
```

```
In [209]: p=driver.find_element_by_xpath("//span[@class='price-num']")
p.click()
```

```
In [210]: c=driver.find_element_by_xpath("//li[@class='colour-listItem']")
c.click()
```

```
In [211]: br=driver.find_elements_by_xpath("//h3[@class='product-brand']")
des=driver.find_elements_by_xpath("//h4[@class='product-product']")
pri=driver.find_elements_by_xpath("//div[@class='product-price']")
```

```
In [212]: brand=[]
for i in br:
    brand.append(i.text)
desp=[]
for i in des:
    desp.append(i.text)
price=[]
for i in pri:
    price.append(i.text)
```

```
In [213]: next=driver.find_element_by_xpath("//a[@rel='next']")
next.click()
```

```
In [214]: br=driver.find_elements_by_xpath("//h3[@class='product-brand']")
des=driver.find_elements_by_xpath("//h4[@class='product-product']")
pri=driver.find_elements_by_xpath("//div[@class='product-price']")
```

```
In [215]: for i in br:
    brand.append(i.text)
for i in des:
    desp.append(i.text)
for i in pri:
    price.append(i.text)
```

```
In [216]: brand[0:2]
```

```
Out[216]: ['Eego Italy', 'Fentacia']
```

```
In [217]: desp[0:2]
```

```
Out[217]: ['Men Trekking Shoes', 'Men Woven Flat Boots']
```

```
In [218]: price[0:2]
```

```
Out[218]: ['Rs. 899Rs. 2599(Rs. 1700 OFF)', 'Rs. 879Rs. 3999(78% OFF)']
```

```
In [219]: print(len(brand),len(desp),len(price))
```

```
100 100 100
```

Q8: Go to webpage <https://www.amazon.in/> (<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” as shown in the below image: After setting the filters scrape first 10 laptops data. You have to scrape 3 attributes for each laptop:

1. Title
2. Ratings
3. Price As shown in the below image as the tick marked attributes

```
In [297]: driver.get('https://www.amazon.in/')
```

```
In [298]: laptop=driver.find_element_by_xpath("//input[@id='twotabsearchtextbox']")
laptop.send_keys("Laptop")
```

```
In [299]: scr=driver.find_element_by_xpath("//input[@id='nav-search-submit-button']")
scr.click()
```

```
In [223]: i7=driver.find_element_by_xpath("//div[@class='a-checkbox a-checkbox-fancy s-navi
i9=driver.find_element_by_xpath("//i[@class='a-icon a-icon-checkbox']")
i7.click()
#i9.click()
```

```
In [224]: lap_n=driver.find_elements_by_xpath("//h2[@class='a-size-mini a-spacing-none a-color-  
lap_n[0:2]
```

```
Out[224]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="8107f43b-d214-4723-8029-7960d79ad5c0")>,  
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="6411d2c6-fc3f-4ca4-8ca6-fe92ab954f9d")>]
```

```
In [225]: lap_new=lap_n[0:10]  
lap_new[0:2]
```

```
Out[225]: [<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="8107f43b-d214-4723-8029-7960d79ad5c0")>,  
<selenium.webdriver.remote.webelement.WebElement (session="2d0506f8445c51df91c  
c656d4afa9bce", element="6411d2c6-fc3f-4ca4-8ca6-fe92ab954f9d")>]
```

```
In [226]: laptop_names=[]  
for i in lap_new:  
    q=i.text  
    laptop_names.append(q)  
laptop_names[0:2]
```

```
Out[226]: ['Lenovo IdeaPad Slim 3 2021 Intel Core i3 11th Gen 14 FHD IPS Thin & Light Lap  
top (8GB/256GB SSD/Windows 10/MS Office/Backlit Keyboard/Fingerprint Reader/2 Y  
ear Warranty/Arctic Grey/1.41kg), 82H700SVIN',  
"Lenovo IdeaPad 3 Intel Celeron N4020 14'' HD Thin & Light Laptop (4GB/256GB H  
DD/Windows 11/MS Office 2021/Platinum Grey/1.5Kg), 81WH007KIN"]
```

```
In [300]: r_tag=driver.find_elements_by_xpath("//div[@class='a-row a-size-small']")  
r_tag[0:2]
```

```
Out[300]: [<selenium.webdriver.remote.webelement.WebElement (session="04ac73ef4b4abf5447f  
ecaca92df7389", element="30ba61d9-3bf8-434b-aa80-b7ccfb07b766")>,  
<selenium.webdriver.remote.webelement.WebElement (session="04ac73ef4b4abf5447f  
ecaca92df7389", element="1ff3febb-7122-435a-92fb-2a66cdf60a74")>]
```

```
In [301]: r_new=r_tag[0:10]  
r_new[0:2]
```

```
Out[301]: [<selenium.webdriver.remote.webelement.WebElement (session="04ac73ef4b4abf5447f  
ecaca92df7389", element="30ba61d9-3bf8-434b-aa80-b7ccfb07b766")>,  
<selenium.webdriver.remote.webelement.WebElement (session="04ac73ef4b4abf5447f  
ecaca92df7389", element="1ff3febb-7122-435a-92fb-2a66cdf60a74")>]
```

```
In [302]: rates=[]  
for i in r_new:  
    w=i.text  
    rates.append(w)  
rates
```

```
Out[302]: ['39', '2', '97', '146', '8', '80', '151', '84', '', '444']
```

```
In [212]: pr=driver.find_elements_by_xpath("//span[@class='a-price-whole']")  
pr[0:2]
```

```
Out[212]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15  
7854e0700356a", element="7c894227-734d-460c-9337-8655222ad54c">,  
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15  
7854e0700356a", element="89a5a5eb-d807-42fd-a608-cf732db53cbf")>]
```

```
In [213]: pr_n=pr[0:10]
```

```
In [214]: price=[]  
for i in pr_n:  
    e=i.text  
    price.append(e)  
price
```

```
Out[214]: ['26,055',  
'43,490',  
'26,355',  
'19,990',  
'60,490',  
'55,990',  
'36,890',  
'66,990',  
'38,499',  
'39,990']
```

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. This task will be done in following steps:

1. First get the webpage <https://www.ambitionbox.com/>
2. Click on the Job option as shown in the image
3. After reaching to the next webpage, In place of "Search by Designations, Companies, Skills" enter "Data Scientist" and click on search button.
4. You will reach to the following web page click on location and in place of "Search location" enter "Noida" and select location "Noida".
5. Then scrape the data for the first 10 jobs results you get on the above shown page.
6. Finally create a dataframe of the scraped data.

```
In [221]: driver.get('https://www.ambitionbox.com/')
```

```
In [222]: jobs=driver.find_element_by_xpath("//a[@class='link jobs']")  
jobs.click()
```

```
In [223]: type1=driver.find_element_by_xpath("//input[@title='Enter Designation, Company or  
type1.send_keys("Data Scientist")
```

```
In [224]: src=driver.find_element_by_xpath("//button[@class='ab_btn search-btn round']")
src.click()
```

```
In [225]: loc=driver.find_element_by_xpath("//div[@title='Location']")
loc.click()
```

```
In [226]: noi=driver.find_element_by_xpath("//label[@for='location_Noida']")
noi.click()
```

```
In [227]: com_name=driver.find_elements_by_xpath("//div[@class='company-info']")
com_name[0:2]
```

```
Out[227]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="f4ff8065-42ce-4117-95f4-e6411d9726d1">,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="b4c5e75f-5a0b-4ec3-bbfa-1857dbc00955")>]
```

```
In [228]: co_n=com_name[0:10]
```

```
In [229]: company=[]
for i in co_n:
    q=i.text
    company.append(q)
company[0:2]
```

```
Out[229]: ['Ericsson India Global Services Pvt. Ltd.\n · \n4.2\nbased on 3.9k Reviews',
'Ericsson India Global Services Pvt. Ltd.\n4.2\n(3.9k Reviews)']
```

```
In [230]: no=driver.find_elements_by_xpath("//span[@class='body-small-1']")
no[0:2]
```

```
Out[230]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="6b5e9141-39f5-42de-b4e7-467bcf4b004a">,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="0c92e980-fa94-4c77-95f6-040fd1b10737")>]
```

```
In [231]: numberOf=[]
for i in no:
    q=i.text
    numberOf.append(q)
numberOf[0:2]
```

```
Out[231]: ['1d ago', 'via naukri.com']
```

```
In [232]: days=[]
junk=[]
for i in range(0,len(numberOf)-1,2):
    days.append(numberOf[i])
    junk.append(numberOf[i+1])
days[0:2]
```

Out[232]: ['1d ago', '6d ago']

```
In [233]: r_tag=driver.find_elements_by_xpath("//span[@class='body-small']")
r_tag[0:2]
```

Out[233]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b157854e0700356a", element="253421f7-43c0-4d50-89ae-48857a41c93f")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b157854e0700356a", element="827b679e-dc76-4aa2-b0cb-d1f8d58c3b88")>]

```
In [234]: rates=[]
for i in r_tag:
    q=i.text
    rates.append(q)
rates
```

Out[234]: ['4.2', '4.0', '4.0', '3.8', '4.0', '4.0', '4.0', '4.0', '4.1', '4.1']

```
In [235]: print(len(company),len(days),len(rates))
```

10 10 10

```
In [236]: j=pd.DataFrame({})
j['company']=company
j['posted no.of days ago']=days
j['company rating']=rates
j
```

Out[236]:

	company	posted no.of days ago	company rating
0	Ericsson India Global Services Pvt. Ltd.\n · \n...	1d ago	4.2
1	Ericsson India Global Services Pvt. Ltd.\n4.2\...	6d ago	4.0
2	GENPACT India Private Limited\n4.0\n(11.8k Rev...	6d ago	4.0
3	GENPACT India Private Limited\n4.0\n(11.8k Rev...	7d ago	3.8
4	NTT Data Business Solutions Pvt Ltd\n3.8\n(2.9...	8d ago	4.0
5	GENPACT India Private Limited\n4.0\n(11.8k Rev...	9hr ago	4.0
6	GI Group\n4.0\n(70 Reviews)	9hr ago	4.0
7	GI Group\n4.0\n(70 Reviews)	4d ago	4.0
8	GI Group\n4.0\n(70 Reviews)	29d ago	4.1
9	Steria India Ltd\n4.1\n(849 Reviews)	1d ago	4.1

Q10: Write a python program to scrape the salary data for Data Scientist designation. You have to
localhost:8888/notebooks/Datatrained_Internship_jupy/Web Scraping Assignment-2.ipynb

26/30

scrape Company name, Number of salaries, Average salary, Min salary, Max Salary. The above task will be, done as shown in the below steps:

1. First get the webpage <https://www.ambitionbox.com/> (`https://www.ambitionbox.com/`)
2. Click on the salaries option as shown in the image.
3. After reaching to the following webpage, In place of "Search Job Profile" enters "Data Scientist" and then click on "Data Scientist". You have to scrape the data ticked in the above image.
4. Scrape the data for the first 10 companies. Scrape the company name, total salary record, average salary, minimum salary, maximum salary, experience required.
5. Store the data in a dataframe.

```
In [237]: driver.get('https://www.ambitionbox.com/')
```

```
In [238]: sa=driver.find_element_by_xpath("//a[@title='Company Salaries']")
sa.click()
```

```
In [239]: d=driver.find_element_by_xpath("//input[@title='Enter a Company or Designation']")
d.send_keys('Data Scientist')
```

```
In [240]: s=driver.find_element_by_xpath("//button[@class='ab_btn search-btn round compone
```

```
In [241]: name=driver.find_elements_by_xpath("//div[@class='name']")
name
```

```
Out[241]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="bc1f0c86-3cac-49d5-a709-3754981aacae")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="647cf147-32bf-449a-b6ee-01bba4cdadfb")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="70a31d64-af3a-4b8e-8d46-d1c1091bcefcc")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="f7d3b48c-81b0-4e7a-89dc-3a89467b8d10")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="f054f7b2-644e-4118-8f39-f5ce7ccacdb8")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="ba27276f-3fe0-4fcc-9637-0d1b669a4021")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="9573de43-adcf-4a07-b291-61f5e28fbe12")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="743fc09-866e-440f-8f98-3aaaec1d2d84")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="928ca83d-06cb-4629-8bee-5d13bb4a47a1")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="86f5aa1d-871d-4a97-aa60-ee18017e910d")>]
```

```
In [242]: comp=[]
for i in name:
    q=i.text
    comp.append(q)
comp[0:2]
```

```
Out[242]: ['Microsoft Corporation\nbased on 217 salaries',
'Goldman Sachs\nbased on 10 salaries']
```

```
In [243]: e=driver.find_elements_by_xpath("//div[@class='salaries sbold-list-header']")
e[0:2]
```

```
Out[243]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="8524d819-3150-4942-9409-bd9e91e055fd")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="e3957460-6974-417e-92e9-1d5d70f4c013")>]
```

```
In [244]: exp=[]
for i in e:
    exp.append(i.text)
exp[0:2]
```

```
Out[244]: ['Software Engineer\n . \n1-4 yrs exp', 'Software Engineer\n . \n2 yrs exp']
```

```
In [245]: a=driver.find_elements_by_xpath("//p[@class='averageCtc']")
a
```

```
Out[245]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="5e53fa70-1c2c-497e-a69d-60407c9ca5e3")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="fb8f0482-d013-4e28-83ba-6d1154c8b39b")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="22a0e79b-8a8f-44e8-a7dc-25438ca1995a")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="2b8e5b6d-e4ac-4081-9b4d-ca440330a6b2")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="30f8130c-0a4e-4580-9c72-f295d155387b")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="ff2c602d-f47c-4c3c-8d54-0ccdb813cbc5")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="68ba112f-9b6c-4e82-9f3b-9c88a75555a4")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="a0009c37-b301-48e3-b961-01a521a2d39f")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="547b2f35-010f-48ab-a5c4-f7a9a9981982")>,
<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="70302a98-686e-442e-9b2b-0ba76ab85e2e")>]
```

```
In [246]: asal=[]
for i in a:
    asal.append(i.text)
asal[0:2]
```

```
Out[246]: ['₹ 22.4L', '₹ 21.2L']
```

```
In [247]: s=driver.find_elements_by_xpath("//div[@class='salary-values']")
s
```

```
Out[247]: [<selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="c90631b2-0e6d-4b98-8ff5-0fabd6dceb23")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="7972ecd1-8bcf-4e45-891e-c483f8b9b8fa")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="6afcb98d-78cd-40af-9600-3e799ecb2c88")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="b4cab56-da8d-4341-af4f-19495bb81bdc")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="e2d29a0b-645c-4a85-9663-d6b20dc7c602")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="4cdf7c1a-676c-42c9-a3d6-81602bcc4528")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="cf635bd8-e279-46db-9c74-550571ffcd25")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="a9c4b176-f847-4678-a32d-9e7a3948994b")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="7db95524-c170-4ef6-80b6-68aa59a09e4c")>,
 <selenium.webdriver.remote.webelement.WebElement (session="90a2ec7eaeacb372b15
7854e0700356a", element="2c7eaed4-676f-4d72-b149-f6a945613a45")>]
```

```
In [248]: x=[]
for i in s:
    x.append(i.text.replace('\n', ''))
x
```

```
Out[248]: ['₹ 12.0L      ₹ 45.0L',
 '₹ 16.0L      ₹ 30.0L',
 '₹ 7.0L       ₹ 30.0L',
 '₹ 8.0L       ₹ 40.0L',
 '₹ 12.0L      ₹ 30.0L',
 '₹ 10.0L      ₹ 32.0L',
 '₹ 12.0L      ₹ 24.5L',
 '₹ 11.2L      ₹ 23.0L',
 '₹ 12.0L      ₹ 23.0L',
 '₹ 12.0L      ₹ 22.0L']
```

```
In [249]: print(len(comp),len(exp),len(asal),len(x))
```

10 10 10 10

```
In [250]: salary=pd.DataFrame({})
salary['comapany']=comp
salary['expirement']=exp
salary['AvgSalary']=asal
salary['min      max']=x
salary
```

Out[250]:

	comapany	expirement	AvgSalary	min	max
0	Microsoft Corporation\nbased on 217 salaries	Software Engineer\n . \n1-4 yrs exp	₹ 22.4L	₹ 12.0L	₹ 45.0L
1	Goldman Sachs\nbased on 10 salaries	Software Engineer\n . \n2 yrs exp	₹ 21.2L	₹ 16.0L	₹ 30.0L
2	Flipkart\nbased on 52 salaries	Software Engineer\n . \n1-4 yrs exp	₹ 20.6L	₹ 7.0L	₹ 30.0L
3	Amazon\nbased on 77 salaries	Software Engineer\n . \n1-4 yrs exp	₹ 19.3L	₹ 8.0L	₹ 40.0L
4	Arcessium\nbased on 34 salaries	Software Engineer\n . \n1 yr exp	₹ 18.3L	₹ 12.0L	₹ 30.0L
5	Walmart\nbased on 77 salaries	Software Engineer\n . \n1-4 yrs exp	₹ 17.8L	₹ 10.0L	₹ 32.0L
6	Servicenow Software Development India\nbased o...	Software Engineer\n . \n2-3 yrs exp	₹ 17.8L	₹ 12.0L	₹ 24.5L
7	ServiceNow\nbased on 16 salaries	Software Engineer\n . \n3 yrs exp	₹ 17.8L	₹ 11.2L	₹ 23.0L
8	PayPal\nbased on 12 salaries	Software Engineer\n . \n1 yr exp	₹ 17.6L	₹ 12.0L	₹ 23.0L
9	Citrix\nbased on 10 salaries	Software Engineer\n . \n1 yr exp	₹ 17.3L	₹ 12.0L	₹ 22.0L