This notebook is for the answers to assignment no.2 of web scrapping using Selenium

Instructions

1. All the questions must be done in a single Jupyter notebook.
2. There should be proper comments in code.

Question 1

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.
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 [1]:
```
# Importing Necessary Libraries
import selenium
from selenium import webdriver as wd
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
```

Question 1:

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.
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.

Answer

In [2]:
```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.naukri.com/'        # Defining the url
driver.get(url)                       # opening the url in our test chrome explorer

search_title= driver.find_element_by_class_name("suggestor-input ") # To find the we
search_title                                                        # calling the we

search_loc= driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[3]/div/div/di
search_loc

search_title.send_keys("Data Analyst") # Entering Values in Search Title
search_loc.send_keys("Bangalore")         # Entering Value in Location field
```

In [3]:
```
# Finding search button using absolute Xpath
search= driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[3]/div/div/div[6]
search
# Clicking the search button
search.click()
```

In [6]:
```
# Extracting Job Titles

job_titles= driver.find_elements_by_xpath('//a[@class="title fw500 ellipsis"]')

job=[]
for i in job_titles:
    job.append(i.text)


# Extracting companies

job_company= driver.find_elements_by_xpath('//a[@class="subTitle ellipsis fleft"]')

company=[]
for i in job_company:
    company.append(i.text)

# Extracting location and experience

job_loc= driver.find_elements_by_xpath('//span[@class="ellipsis fleft fs12 lh16 "]')

col=[]
for i in job_loc:
    col.append(i.text)

loc=col[3:42:3] + col[46:55:3]
exp=col[1:41:3] + col[44:55:3]
```

In [7]:
```
table= pd.DataFrame({"Job Title":job[0:10], "Company":company[0:10], "Location":loc[
print(table)
```

```
                                       Job Title  \
0                                 Data Analyst I
1  Data Science / Data Engineer / Business Analys...
2                          Sr.Business Data Analyst
```

```
3                                Senior Data Analyst
4                                    Sr Data Analyst
5             Data Analyst - IIM/ISB/MDI/FMS/SP Jain
6  Data Analyst / Business analyst - US MNC (anal...
7                      Business Analyst/Data Analyst
8                                Junior Data Analyst
9                       Senior Data Analysis Analyst

                                  Company                   Location  \
0                                  Cerner                   0-3 Yrs
1  NETWORTH DATA PRODUCTS PRIVATE LIMITED              Not disclosed
2                               Collabera         Bangalore/Bengaluru
3               Hudsons bay Company (HBC)         Bangalore/Bengaluru
4                          Thomson Reuters         Bangalore/Bengaluru
5             K12 Techno Services Pvt Ltd              Not disclosed
6                      Aspyra HR Services              Not disclosed
7                    Telamon HR Solutions  10,00,000 - 20,00,000 PA.
8                                ICF Next              Not disclosed
9                                   Capco    1,75,000 - 4,75,000 PA.

                                       Experience
0                                     Not disclosed
1                           4,00,000 - 6,50,000 PA.
2                                           3-4 Yrs
3                                           5-8 Yrs
4                                           4-9 Yrs
5                          10,00,000 - 20,00,000 PA.
6  Bangalore/Bengaluru, Hyderabad/Secunderabad, D...
7                     Bangalore/Bengaluru, New Delhi
8                  Bangalore/Bengaluru, Pune, Chennai
9                               Bangalore/Bengaluru
```

Question 2

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.

Answer

In [8]:

```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.naukri.com/'        # Defining the url
driver.get(url)                       # opening the url in our test chrome explorer

search_title= driver.find_element_by_class_name("suggestor-input ") # To find the we
```

```
search_title                                                    # calling the we

search_loc= driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[3]/div/div/di
search_loc

search_title.send_keys("Data Scientist") # Entering Values in Search Title
search_loc.send_keys("Bangalore")         # Entering Value in Location field
```

In [9]:

```
# Finding search button using absolute Xpath
search= driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[3]/div/div/div[6]
search
# Clicking the search button
search.click()
```

In [10]:

```
# Extracting Job Titles

job_titles= driver.find_elements_by_xpath('//a[@class="title fw500 ellipsis"]')

job=[]
for i in job_titles:
    job.append(i.text)


# Extracting companies

job_company= driver.find_elements_by_xpath('//a[@class="subTitle ellipsis fleft"]')

company=[]
for i in job_company:
    company.append(i.text)

# Extracting location and experience

job_loc= driver.find_elements_by_xpath('//span[@class="ellipsis fleft fs12 lh16 "]')

col=[]
for i in job_loc:
    col.append(i.text)



# Splitting location and Experience in different lists



loc= col[2:9:3]+ col[12:30:3] + col[34:55:3]
exp= col[0:7:3] + col[10:29:3] + col[32:54:3]
```

In [11]:

```
table= pd.DataFrame({"Job Title":job[0:10], "Company":company[0:10], "Location":loc[
print(table)
```

```
                                        Job Title  \
0  Urgent Job Opening For AI Practitioner - Data ...
1                   Hiring For Senior Data Scientist
2                             Senior Data Scientist
3                                Dataiku Consultant
4            Research and Development -AI/ML -(PhD )
5  Opportunity For Data Scientist - Female Candid...
6                       Senior Data Science Engineer
7                 Data Science - Engineering Manager
```

```
8                                    Data Scientist
9  Data & Analytics Tech - Informatica Cloud- Sen...

                          Company  \
0                           Wipro
1  TATA CONSULTANCY SERVICES (TCS)
2                      Spiceworks
3                           Wipro
4                             EXL
5                            PayU
6               Fractal Analytics
7                           Paytm
8               Applied Materials
9                             PwC

                                          Location     Experience
0  Bangalore/Bangaluru, Kochi/Cochin, New Delhi, ...      11-21 Yrs
1                          Bangalore/Bangaluru, Pune       8-13 Yrs
2  Bangalore/Bangaluru, Mumbai, Hyderabad/Secunde...        5-7 Yrs
3                                          4-8 Yrs  Not disclosed
4                                          1-3 Yrs  Not disclosed
5                                          5-9 Yrs  Not disclosed
6                                         9-13 Yrs  Not disclosed
7                                          4-7 Yrs  Not disclosed
8                                         6-10 Yrs  Not disclosed
9                                   Not disclosed  Not disclosed
```

Question 3

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 [12]:

```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.naukri.com/'        # Defining the url
driver.get(url)                       # opening the url in our test chrome explorer
```

```
search_title= driver.find_element_by_class_name("suggestor-input ") # To find the we
search_title                                                        # calling the we

search_loc= driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[3]/div/div/di
search_loc

search_title.send_keys("Data Scientist") # Entering Values in Search Title

# Finding search button using absolute Xpath
search = driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[3]/div/div/div[6
search
# Clicking the search button
search.click()
```

In [13]:

```
# Using Filters
# Checking location Delhi/NCR
loc_chk= driver.find_element_by_xpath("/html/body/div[1]/div[3]/div[2]/section[1]/di
loc_chk.click()
```

In [14]:

```
# Checking salary 3-6 lakhs
sal_chk= driver.find_element_by_xpath('/html/body/div[1]/div[3]/div[2]/section[1]/di
sal_chk.click()
```

In [15]:

```
# Extracting Job Titles

job_titles= driver.find_elements_by_xpath('//a[@class="title fw500 ellipsis"]')

job=[]
for i in job_titles:
    job.append(i.text)


# Extracting companies

job_company= driver.find_elements_by_xpath('//a[@class="subTitle ellipsis fleft"]')

company=[]
for i in job_company:
    company.append(i.text)

# Extracting location and experience

loc_exp= driver.find_elements_by_xpath('//span[@class="ellipsis fleft fs12 lh16 "]')

col=[]
for i in loc_exp:
    col.append(i.text)



# Splitting location and Experience in different lists


loc= col[2:8:3] + col[12:28:3] +  col[31:55:3]
exp= col[0:6:3] + col[10:25:3] + col[29:55:3]

table= pd.DataFrame({"Job Title":job[0:10], "Company":company[0:10], "Location":loc[
print(table)
```

```
                                  Job Title  \
0            DigitalBCG GAMMA Data Scientist
1          Data Scientist - Noida/Bangalore
2           Senior Associate - Data Science
3  Data Scientist For Healthcare Product team
4  Data Scientist For Healthcare Product team
5            Data Scientist - MIND Infotech
6                    Data Science Associate
7          Data Scientist - Engine Algorithm
8                  Knowledge/Data Scientist
9                            Data Scientist

                                    Company                          Location  \
0                  Boston Consulting Group  New Delhi, Bangalore/Bengaluru
1                                      EXL      Noida, Bangalore/Bengaluru
2                             Black Turtle                          2-7 Yrs
3                 SECUREKLOUD TECHNOLOGIES                          4-8 Yrs
4                 SECUREKLOUD TECHNOLOGIES                          2-4 Yrs
5  MOTHERSONSUMI INFOTECH & DESIGNS LIMITED                          1-3 Yrs
6                            Kreate Energy                          3-6 Yrs
7                             Primo Hiring                          2-4 Yrs
8                  BOLD Technology Systems                    Not disclosed
9   Mount Talent Consulting Private Limited                          2-5 Yrs

                              Experience
0                               2-5 Yrs
1                              5-10 Yrs
2                          Not disclosed
3                          Not disclosed
4                          Not disclosed
5                 4,00,000 - 8,00,000 PA.
6                          Not disclosed
7  Delhi / NCR, Pune, Bangalore/Bengaluru
8  Delhi / NCR, Pune, Bangalore/Bengaluru
9                          Not disclosed
```

Question 4

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:

1. Go to Flipkart webpage by url : https://www.flipkart.com/
2. Enter “sunglasses” in the search field where “search for products, brands and more” is written and click the search icon
3. After that you will reach to the page having a lot of sunglasses. From this page you can scrap the required data as usual.
4. After scraping data from the first page, go to the “Next” Button at the bottom other page , then click on it.
5. Now scrape data from this page as usual

6. 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.

Answer

In [16]:
```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.flipkart.com/'        # Defining the url
driver.get(url)                       # opening the url in our test chrome explorer
```

In [17]:
```
# closing the login in pop-up
close= driver.find_element_by_xpath("/html/body/div[2]/div/div/button")
close.click()
```

In [18]:
```
# To find the web element - search title field
search_title= driver.find_element_by_xpath("/html/body/div[1]/div/div[1]/div[1]/div[
search_title                                                    # calling the we

search_title.send_keys("sunglasses") # Entering Values in Search Title
```

In [19]:
```
# Finding search button using absolute Xpath
search = driver.find_element_by_xpath("/html/body/div[1]/div/div[1]/div[1]/div[2]/di
search
# Clicking the search button
search.click()
```

In [20]:
```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')

brand=[]
for i in brand_title:
    brand.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa"]')

prod=[]
for i in prod_rows:
    prod.append(i.text)

# Extracting Price

price_rows= driver.find_elements_by_xpath('//div[@class="_30jeq3"]')

price=[]
for i in price_rows:
    price.append(i.text)


# Extracting Discount
```

```
disc_rows= driver.find_elements_by_xpath('//div[@class="_3Ay6Sb"]')

disc=[]
for i in disc_rows:
    disc.append(i.text)
```

In [21]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath("/html/body/div[1]/div/div[3]/div[1]/div[2]/div
nextb
# Clicking the Next button
nextb.click()
```

In [22]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')

brand2=[]
for i in brand_title:
    brand2.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa"]')

prod2=[]
for i in prod_rows:
    prod2.append(i.text)

# Extracting Price

price_rows= driver.find_elements_by_xpath('//div[@class="_30jeq3"]')

price2=[]
for i in price_rows:
    price2.append(i.text)


# Extracting Discount

disc_rows= driver.find_elements_by_xpath('//div[@class="_3Ay6Sb"]')

disc2=[]
for i in disc_rows:
    disc2.append(i.text)
```

In [23]:

```
# Clicking the Next button
nextb.click()
```

In [24]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')
```

```
brand3=[]
for i in brand_title:
    brand3.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa"]')

prod3=[]
for i in prod_rows:
    prod3.append(i.text)

# Extracting Price

price_rows= driver.find_elements_by_xpath('//div[@class="_30jeq3"]')

price3=[]
for i in price_rows:
    price3.append(i.text)


# Extracting Discount

disc_rows= driver.find_elements_by_xpath('//div[@class="_3Ay6Sb"]')

disc3=[]
for i in disc_rows:
    disc3.append(i.text)
```

In [25]:

```
# Comabining results and limiting records till 100 rows
brands=brand+brand2+brand3[0:20]
prices=price+price2+price3[0:20]
prods=prod+prod2+prod3[0:20]
discs=disc+disc2+disc3[0:20]
```

In [26]:

```
# Creating DataFrame
table=pd.DataFrame({"Brand":brands,"Product Description":prods,"Price":prices,"Disco
print(table)
```

```
           Brand                                Product Description   Price  \
0   VINCENT CHASE  by Lenskart Polarized, UV Protection Rectangul...    ₹749
1   VINCENT CHASE          UV Protection Rectangular Sunglasses (52)    ₹649
2        DAHAAZIL  UV Protection, Night Vision, Riding Glasses Wa...    ₹177
3       New Specs   UV Protection Rectangular Sunglasses (Free Size)    ₹264
4          PIRASO              UV Protection Aviator Sunglasses (54)    ₹249
..            ...                                               ...     ...
95      ROYAL SON    Polarized, UV Protection Sports Sunglasses (68)  ₹1,234
96      ROYAL SON  UV Protection, Gradient Butterfly Sunglasses (62)    ₹699
97   Silver Kartz      UV Protection Wayfarer Sunglasses (Free Size)    ₹288
98         GANSTA              UV Protection Aviator Sunglasses (57)    ₹314
99      ROYAL SON  Polarized, UV Protection Retro Square Sunglass...    ₹664

   Discount
0   62% off
1   67% off
2   82% off
3   89% off
4   84% off
..      ...
```

```
95  50% off
96  65% off
97  80% off
98  84% off
99  55% off

[100 rows x 4 columns]
```

Question 5

Scrape 100 reviews data from flipkart.com for iphone11 phone.

This task will be done in following steps:

1. First get the webpage https://www.flipkart.com/
2. Enter “iphone 11” in “Search” field.
3. Then click the search button.

You will reach to the below shown webpage

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.

Answer

In [27]:
```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.flipkart.com/apple-iphone-11-black-64-gb-includes-earpods-power-ada

driver.get(url)                        # opening the url in our test chrome explore
```

In [28]:
```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings=[]
for i in rating_row:
    ratings.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")
```

```
review=[]
for i in review_sum:
    review.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_xpath('//div[@class="t-ZTKy"]')

full=[]
for i in review_full:
    full.append(i.text)
```

In [30]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [31]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings2=[]
for i in rating_row:
    ratings2.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review2=[]
for i in review_sum:
    review2.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full2=[]
for i in review_full:
    full2.append(i.text)
```

In [33]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [34]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings3=[]
for i in rating_row:
    ratings3.append(i.text)


# Extracting Review Summary
```

```
review_sum= driver.find_elements_by_class_name("_2-N8zT")

review3=[]
for i in review_sum:
    review3.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full3=[]
for i in review_full:
    full3.append(i.text)
```

In [36]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [37]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings4=[]
for i in rating_row:
    ratings4.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review4=[]
for i in review_sum:
    review4.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full4=[]
for i in review_full:
    full4.append(i.text)
```

In [39]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [40]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings5=[]
for i in rating_row:
    ratings5.append(i.text)
```

```
# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review5=[]
for i in review_sum:
    review5.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full5=[]
for i in review_full:
    full5.append(i.text)
```

In [42]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [43]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings6=[]
for i in rating_row:
    ratings6.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review6=[]
for i in review_sum:
    review6.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full6=[]
for i in review_full:
    full6.append(i.text)
```

In [45]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [46]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')
```

```
ratings7=[]
for i in rating_row:
    ratings7.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review7=[]
for i in review_sum:
    review7.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full7=[]
for i in review_full:
    full7.append(i.text)
```

In [48]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [49]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings8=[]
for i in rating_row:
    ratings8.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review8=[]
for i in review_sum:
    review8.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full8=[]
for i in review_full:
    full8.append(i.text)
```

In [51]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [52]:

```
# Extracting Ratings
```

```
rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings9=[]
for i in rating_row:
    ratings9.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review9=[]
for i in review_sum:
    review9.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full9=[]
for i in review_full:
    full9.append(i.text)
```

In [54]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [55]:

```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings10=[]
for i in rating_row:
    ratings10.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review10=[]
for i in review_sum:
    review10.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full10=[]
for i in review_full:
    full10.append(i.text)
```

In [57]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath('/html/body/div/div/div[3]/div/div/div[2]/div[1
nextb
# Clicking the Next button
nextb.click()
```

In [58]:
```
# Extracting Ratings

rating_row= driver.find_elements_by_xpath('//div[@class="_3LWZlK _1BLPMq"]')

ratings11=[]
for i in rating_row:
    ratings11.append(i.text)


# Extracting Review Summary

review_sum= driver.find_elements_by_class_name("_2-N8zT")

review11=[]
for i in review_sum:
    review11.append(i.text)

# Extracting Full Reeview

review_full= driver.find_elements_by_class_name("t-ZTKy")

full11=[]
for i in review_full:
    full11.append(i.text)
```

In [59]:
```
tot_ratings= ratings+ratings2+ratings3+ratings4+ratings5+ratings6+ratings7+ratings8+
tot_reviews= review+review2+review3+review4+review5+review6+review7+review8+review9+
tot_full= full+full2+full3+full4+full5+full6+full7+full8+full9+full10+full11
```

In [61]:
```
# Creating DataFrame
table=pd.DataFrame({"Ratings":tot_ratings[0:100],"Reviews Summary":tot_reviews[0:100
print(table)
```

```
   Ratings        Reviews Summary  \
0        5         Simply awesome
1        5       Perfect product!
2        5    Best in the market!
3        5     Highly recommended
4        5      Worth every penny
..     ...                    ...
95       5       Perfect product!
96       5         Classy product
97       5      Worth every penny
98       5              Fabulous!
99       5  Mind-blowing purchase

                                Reviews Description
0   Really satisfied with the Product I received.....
1   Amazing phone with great cameras and better ba...
2   Great iPhone very snappy experience as apple k...
3   What a camera .....just awesome ..you can feel...
4   Previously I was using one plus 3t it was a gr...
..                                                ...
95  Value for money\n5 star rating\nExcellent came...
96  Gifted my man on his 30th birthday 🎂 He loves ...
97  It is better to buy iPhone 11 over iPhone 12 i...
98  Gift this to your loved ones fabulous product ...
99  awesome Phone Smooth Touch Too good Sexyy look...

[100 rows x 3 columns]
```

Question 6

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.

Note: All the steps required during scraping should be done through code only and not manually.

Answer

In [62]:

```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.flipkart.com/'        # Defining the url
driver.get(url)                       # opening the url in our test chrome explorer

# closing the login in pop-up
close= driver.find_element_by_xpath("/html/body/div[2]/div/div/button")
close.click()


# To find the web element - search title field
search_title= driver.find_element_by_xpath("/html/body/div[1]/div/div[1]/div[1]/div[
search_title                                                    # calling the we

search_title.send_keys("sneakers") # Entering Values in Search Title

# Finding search button using absolute Xpath
search = driver.find_element_by_xpath("/html/body/div[1]/div/div[1]/div[1]/div[2]/di
search
# Clicking the search button
search.click()
```

In [63]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')

brand=[]
for i in brand_title:
    brand.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa"]')

prod=[]
```

```
for i in prod_rows:
    prod.append(i.text)

prod_1=[]
prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa _2-ICcC"]')
for i in prod_rows:
    prod_1.append(i.text)


# Extracting Price

price_rows= driver.find_elements_by_xpath('//div[@class="_30jeq3"]')

price=[]
for i in price_rows:
    price.append(i.text)


# Extracting Discount

disc_rows= driver.find_elements_by_xpath('//div[@class="_3Ay6Sb"]')

disc=[]
for i in disc_rows:
    disc.append(i.text)
```

In [64]:

```
# Finding Next button using absolute Xpath
nextb = driver.find_element_by_xpath("/html/body/div[1]/div/div[3]/div[1]/div[2]/div
nextb
# Clicking the Next button
nextb.click()
```

In [65]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')

brand2=[]
for i in brand_title:
    brand2.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa"]')

prod2=[]
for i in prod_rows:
    prod2.append(i.text)

prod2_1=[]
prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa _2-ICcC"]')
for i in prod_rows:
    prod2_1.append(i.text)

# Extracting Price

price_rows= driver.find_elements_by_xpath('//div[@class="_30jeq3"]')
```

```
price2=[]
for i in price_rows:
    price2.append(i.text)


# Extracting Discount

disc_rows= driver.find_elements_by_xpath('//div[@class="_3Ay6Sb"]')

disc2=[]
for i in disc_rows:
    disc2.append(i.text)
```

In [66]:

```
# Clicking the Next button
nextb.click()
```

In [67]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')

brand3=[]
for i in brand_title:
    brand3.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_xpath('//a[@class="IRpwTa"]')

prod3=[]
for i in prod_rows:
    prod3.append(i.text)


# Extracting Price

price_rows= driver.find_elements_by_xpath('//div[@class="_30jeq3"]')

price3=[]
for i in price_rows:
    price3.append(i.text)


# Extracting Discount

disc_rows= driver.find_elements_by_xpath('//div[@class="_3Ay6Sb"]')

disc3=[]
for i in disc_rows:
    disc3.append(i.text)
```

In [68]:

```
# Comabining results and limiting records till 100 rows
brands=brand+brand2+brand3[0:20]
prices=price+price2+price3[0:20]
```

```
prods=prod+prod_1+prod2+prod2_1+prod3[0:20]
discs=disc+disc2+disc3[0:20]
```

In [69]:

```
# Creating DataFrame
table=pd.DataFrame({"Brand":brands,"Product Description":prods,"Price":prices,"Disco
print(table)
```

```
              Brand                                Product Description Price  \
0            Nilatin                                   Sneakers For Men  ₹549
1        ZF - ALFIYA                                   Sneakers For Men  ₹449
2             BRUTON      Modern Trendy Sneakers Shoes Sneakers For Men  ₹284
3           Magnolia                                   Sneakers For Men  ₹399
4           URBANBOX                    Puma Smash v2 L Sneakers For Men  ₹219
..               ...                                                ...   ...
95  MAST AND HARBOUR                                   Sneakers For Men  ₹999
96          PROVOGUE                                   Sneakers For Men  ₹759
97           Stinson                                   Sneakers For Men  ₹249
98             Eiffel  Comfortable & Ultra Light Weight Sneaker Sneak...  ₹469
99          KWIK FIT                                   Sneakers For Men  ₹499

   Discount
0   45% off
1   55% off
2   78% off
3   60% off
4   78% off
..      ...
95  62% off
96  62% off
97  50% off
98  53% off
99  75% off

[100 rows x 4 columns]
```

Question 7

Go to the link - https://www.myntra.com/shoes

Set second Price filter and 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.

Note: Applying the filter and scraping the data, everything should be done through code only and there should not be any manual step.

In [70]:

```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.myntra.com/shoes'        # Defining the url
driver.get(url)                         # opening the url in our test chrome explorer


# Using Filters
# Checking Black color filters
col_chk= driver.find_element_by_xpath("/html/body/div[2]/div/div[1]/main/div[3]/div[
col_chk.click()
```

In [71]:

```
# Using Filters
```

```
# Checking price filters
col_chk= driver.find_element_by_xpath("/html/body/div[2]/div/div[1]/main/div[3]/div[
col_chk.click()
```

In [72]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_class_name("product-brand")

brand=[]
for i in brand_title:
    brand.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_class_name("product-product")

prod=[]
for i in prod_rows:
    prod.append(i.text)


# Extracting Price

price_rows= driver.find_elements_by_class_name("product-price")

price=[]
for i in price_rows:
    price.append(i.text)
```

In [73]:

```
# Clicking next button
nextb= driver.find_element_by_xpath("/html/body/div[2]/div/div[1]/main/div[3]/div[2]
nextb.click()
```

In [74]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_class_name("product-brand")

brand2=[]
for i in brand_title:
    brand2.append(i.text)


# Extracting Product

prod_rows= driver.find_elements_by_class_name("product-product")

prod2=[]
for i in prod_rows:
    prod2.append(i.text)


# Extracting Price

price_rows= driver.find_elements_by_class_name("product-price")

price2=[]
```

```
for i in price_rows:
    price2.append(i.text)
```

In [76]:

```
# Comabining results and limiting records till 100 rows
brands=brand+brand2
prices=price+price2
prods=prod+prod2
```

In [77]:

```
# Creating DataFrame
table=pd.DataFrame({"Brand":brands,"Product Description":prods,"Price":prices})
print(table)
```

```
        Brand             Product Description                         Price
0        ALDO             Men Leather Loafers   Rs. 7799Rs. 12999(40% OFF)
1        Nike    Men React Infinity 3 Running  Rs. 11196Rs. 13995(20% OFF)
2        ALDO             Men Leather Loafers   Rs. 7999Rs. 15999(50% OFF)
3        Nike      Men KD 15 Basketball Shoes  Rs. 11895Rs. 13995(15% OFF)
4        Nike  Women React MR 3 Running Shoes   Rs. 7871Rs. 10495(25% OFF)
..        ...                              ...                         ...
95     ADIDAS   Women Supernova Running Shoes    Rs. 7999Rs. 9999(20% OFF)
96       Geox     Men Leather Formal Slip-Ons    Rs. 8991Rs. 9990(10% OFF)
97       Xtep               Men Running Shoes                    Rs. 7699
98  J.FONTINI      Men Leather Formal Loafers                    Rs. 7490
99  J.FONTINI       Men Black Leather Loafers                    Rs. 8490

[100 rows x 3 columns]
```

Question 8

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” as shown in the below image:

After setting the filters scrape first 10 laptops data. You have to scrape 3 attributesfor each laptop:

1. Title
2. Ratings
3. Price

Note: All the steps required during scraping should be done through code only and not manually.

In [78]:

```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.amazon.in'        # Defining the url
driver.get(url)                       # opening the url in our test chrome explorer
```

In [79]:

```
# To find the web element - search title field
search_title= driver.find_element_by_xpath("/html/body/div[1]/header/div/div[1]/div[
search_title                                                    # calling the we

search_title.send_keys("Laptop") # Entering Values in Search Title

# Finding search button using absolute Xpath
search = driver.find_element_by_xpath("/html/body/div[1]/header/div/div[1]/div[2]/di
search
# Clicking the search button
search.click()
```

In [80]:

```
# Clicking on Intel Core i7 filter

cpu_chk= driver.find_element_by_xpath("/html/body/div[1]/div[2]/div[1]/div[2]/div/di

cpu_chk.click()
```

In [81]:

```
# Extracting Brand Titles

brand_title= driver.find_elements_by_xpath('//span[@class="a-size-medium a-color-bas
brand=[]
for i in brand_title:
    brand.append(i.text)


# Extracting Price



price=[]
for i in price_rows:
    price.append(i.text)
```

In [82]:

```
# Extracting Ratings
ratings_rows=driver.find_elements_by_xpath('//span[@class="a-icon-alt"]')

ratings=[]
for i in ratings_rows:
    ratings.append(i.text)
```

In [84]:

```
# Creating DataFrame
table=pd.DataFrame({"Brand":brand[0:10],"Product Rating":ratings[0:10],"Price":price
print(table)
```

```
                                               Brand Product Rating  \
0  Acer Extensa 15 Thin & Light Intel Processor P...
1  Acer Aspire 3 Laptop (Made in India) A314-35 3...
2  ASUS VivoBook 14 (2021), 14-inch (35.56 cm) HD...
3  Dell New Inspiron 3521 Laptop, Intel Pqc-N5030...
4  Acer Extensa 15 Thin & Light Intel Processor P...
5  Hp 14S-Intel Pentium Silver N6000- 8Gb Ram/256...
6  Acer Travelmate Business Laptop Intel Pentium ...
7  Microsoft Surface GO 3 8VA-00013 10.5" (26.67 ...
8  Acer Travelmate Intel® Pentium® Gold 7505 Proc...
9  Acer Extensa 15 Thin & Light Laptop Intel Proc...
```

```
                        Price
0   Rs. 8249Rs. 10999(25% OFF)
1   Rs. 7199Rs. 11999(40% OFF)
2                    Rs. 7999
3   Rs. 8505Rs. 10500(19% OFF)
4                   Rs. 11999
5  Rs. 11192Rs. 13990(20% OFF)
6   Rs. 10355Rs. 10900(5% OFF)
7  Rs. 10349Rs. 11499(10% OFF)
8  Rs. 12591Rs. 13990(10% OFF)
9                    Rs. 9990
```

Question 9

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.

Note: All the steps required during scraping should be done through code only and not manually.

In [90]:
```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.ambitionbox.com/'        # Defining the url
driver.get(url)                         # opening the url in our test chrome explorer
```

In [91]:
```
# Clicking on Jobs tab
job_tab= driver.find_element_by_xpath("/html/body/div[1]/nav/nav/a[6]")
job_tab.click()
```

In [92]:
```
search_title= driver.find_element_by_xpath("/html/body/div/div/div/div[2]/div[1]/div
search_title                                                    # calling the we

search_title.send_keys("Data Scientist") # Entering Values in Search Title

# click search button
search_btn= driver.find_element_by_xpath("/html/body/div/div/div/div[2]/div[1]/div[1
search_btn.click()
```

In [94]:
```
# Drop-down search

search_loc= driver.find_element_by_xpath("/html/body/div/div/div/div[2]/div[1]/div[2
search_loc.click()                                                          # calling
```

In [95]:
```
#search_loc.send_keys("Noida") # Entering Values in Search Title

# click search button
search_btn2= driver.find_element_by_xpath("/html/body/div/div/div/div[2]/div[1]/div[
search_btn2.click()
```

In [96]:
```
# Extracting Company Name

company_title= driver.find_elements_by_xpath('//div[@class="company-info"]')
company=[]
for i in company_title:
    company.append(i.text)


# Extracting Posting duration

post_rows= driver.find_elements_by_xpath('//span[@class="body-small-l"]')

post=[]
for i in post_rows:
    post.append(i.text)
```

In [97]:
```
# Creating DataFrame
table=pd.DataFrame({"Company":company[0:10],"Job Posting":post[0:20:2]})
print(table)
```

```
                                             Company Job Posting
0  Optum Global Solutions (India) Private Limited...     11d ago
1  Optum Global Solutions (India) Private Limited...     18d ago
2  GENPACT India Private Limited\n4.0\n(17.9k Rev...      2d ago
3                  Latent bridge\n4.5\n(54 Reviews)      9d ago
4        Dew Solutions Pvt. Ltd.\n4.3\n(80 Reviews)     16d ago
5          InfoEdge India Ltd.\n3.9\n(1.7k Reviews)     17d ago
6      Info Edge India Limited\n3.9\n(1.7k Reviews)     17d ago
7      Info Edge India Limited\n3.9\n(1.7k Reviews)      4d ago
8                    Careerera\n3.8\n(100 Reviews)     12d ago
9                       CRMnext\n4.1\n(74 Reviews)    1mon ago
```

Question 10

Write a python program to scrape the salary data for Data Scientist designation. You have to scrape Company name, Number of salaries, Average salary, Minsalary, Max Salary. The above task will be, done as shown in the below steps:

1. First get the webpage 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.

1. Scrape the data for the first 10 companies. Scrape the company name, total salary record, average salary, minimum salary, maximum salary, experience required.

2. Store the data in a dataframe.

Note: All the steps required during scraping should be done through code only and not manually.

In [98]:
```
# Connecting to a webdriver
driver=wd.Chrome(r'D:\Internship\Web Scraping\Web Scraping 2\chromedriver.exe')

url='https://www.ambitionbox.com/'        # Defining the url
driver.get(url)                          # opening the url in our test chrome explorer
```

In [99]:
```
# Clicking on Salaries tab
sal_tab= driver.find_element_by_xpath("/html/body/div[1]/nav/nav/a[4]")
sal_tab.click()
```

In [100…
```
search_title= driver.find_element_by_xpath("/html/body/div/div/div/main/section[1]/d
search_title                                                    # calling the we

search_title.send_keys("Data Scientist") # Entering Values in Search Title
```

In [101…
```
# clicking Data Scientist from suggestion
sel_btn= driver.find_element_by_xpath("/html/body/div/div/div/main/section[1]/div[2]
sel_btn.click()
```

In [102…
```
# Extracting Company Name

company_title= driver.find_elements_by_tag_name('a')
company=[]
for i in company_title:
    company.append(i.text)


# Extracting Total Salary record

total_rows= driver.find_elements_by_xpath('//span[@class="datapoints"]')

sal_tot=[]
for i in total_rows:
    sal_tot.append(i.text)

# Extracting Minimum Salary record

min_rows= driver.find_elements_by_xpath('//div[@class="value body-medium"]')

sal=[]
for i in min_rows:
    sal.append(i.text)

# Extracting Average Salary

avg_rows= driver.find_elements_by_class_name("averageCtc")
```

```
sal_avg=[]
for i in avg_rows:
    sal_avg.append(i.text)


# Extracting Experience required

exp_reqd= driver.find_elements_by_xpath('//div[@class="sbold-list-header"]')

exp=[]
for i in exp_reqd:
    exp.append(i.text)
```

In [103…

```
min_sal=sal[0:19:2]
max_sal=sal[1::2]
company=company[15:26]
```

In [104…

```
# Creating DataFrame
table=pd.DataFrame({"Company":company[0:10],"Salary Records":sal_tot[0:10],"Minimum
print(table)
```

```
                                  Company          Salary Records  \
0          Ab Inbev\nData Scientist Salary  (based on 12 salaries)
1                ZS\nData Scientist Salary  (based on 33 salaries)
2             Optum\nData Scientist Salary  (based on 15 salaries)
3       Reliance Jio\nData Scientist Salary  (based on 32 salaries)
4  Fractal Analytics\nData Scientist Salary  (based on 21 salaries)
5    Tiger Analytics\nData Scientist Salary  (based on 89 salaries)
6       UnitedHealth\nData Scientist Salary  (based on 51 salaries)
7        EXL Service\nData Scientist Salary  (based on 57 salaries)
8          Deloitte\nData Scientist Salary  (based on 21 salaries)
9                 Software Engineer Salary  (based on 69 salaries)

  Minimum Salary Average Salary Maximum Salary  \
0        ₹ 25.0L        ₹ 25.0L        ₹ 36.0L
1        ₹ 15.0L        ₹ 15.0L        ₹ 26.2L
2        ₹ 11.0L        ₹ 11.0L        ₹ 22.0L
3        ₹ 11.0L        ₹ 11.0L        ₹ 22.5L
4         ₹ 5.6L         ₹ 5.6L        ₹ 26.2L
5        ₹ 10.0L        ₹ 10.0L        ₹ 23.0L
6         ₹ 9.0L         ₹ 9.0L        ₹ 20.0L
7         ₹ 8.3L         ₹ 8.3L        ₹ 21.1L
8         ₹ 7.6L         ₹ 7.6L        ₹ 21.0L
9         ₹ 7.0L         ₹ 7.0L        ₹ 25.0L

                        Experience Required
0    3 yrs experience (based on 12 salaries)
1  3-4 yrs experience (based on 33 salaries)
2    2 yrs experience (based on 15 salaries)
3  3-4 yrs experience (based on 32 salaries)
4  3-4 yrs experience (based on 21 salaries)
5  2-4 yrs experience (based on 89 salaries)
6  2-4 yrs experience (based on 51 salaries)
7  2-4 yrs experience (based on 57 salaries)
8  3-4 yrs experience (based on 21 salaries)
9  2-4 yrs experience (based on 69 salaries)
```

In [ ]: