

WEB SCRAPING – ASSIGNMENT 2

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/> (<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.

In [25]:

```
import configparser
import os
import sys
import time
import platform

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.firefox.options import Options
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.support.ui import WebDriverWait
```

```
config = configparser.ConfigParser()
config.read("config.ini")

USERNAME = config['credentials']['username']
PASSWORD = config['credentials']['password']
CVPATH = config['filepaths']['cvpath']

if not os.path.isfile(CVPATH):
    print("Error! File not found: '{}'.format(CVPATH))
    sys.exit(1)

#default to windows
geckodriver_binary = "geckodriver/geckodriver.exe"

if platform.system().lower() == 'linux':
    geckodriver_binary = "geckodriver/geckodriver"
elif platform.system().lower() == 'windows':
    odriver_binary = "geckodriver/geckodriver.exe"
else:
    print("Error! Unable to determine underlying operating system type.")
    sys.exit(1)
```

```
print("Running on: {}".format(platform.system()))
options = Options()
# options.set_headless(headless=True)
## options.add_argument("--headless")

browser = webdriver.Firefox(firefox_options=options, executable_path=geckodriver_binary,
log_path="geckodriver/geckodriver.log")
# assert options.headless # operating in headless mode
print("Firefox Headless Browser Invoked")
# browser.implicitly_wait(10) # seconds

print("Opening: {}".format("https://login.naukri.com/nLogin/Login.php"))
browser.get("https://login.naukri.com/nLogin/Login.php")

try:
    element = WebDriverWait(browser, timeout=10,
poll_frequency=3).until(EC.presence_of_element_located((By.ID, "usernameField")))
    element = WebDriverWait(browser, timeout=10,
poll_frequency=3).until(EC.presence_of_element_located((By.ID, "passwordField")))
except Exception as e:
    print(e)
finally:
    browser.find_element_by_id('usernameField').send_keys(USERNAME)
    # time.sleep(4)
    passwd = browser.find_element_by_id('passwordField')
    passwd.send_keys(PASSWORD)

    time.sleep(3)
    # passwd.submit()
    #
browser.find_element_by_xpath('/html/body/div[2]/div/div/div/div/div/div/div[2]/div/form/
div[5]/div/button').click()
    browser.find_element_by_id("loginForm").submit()

try:
    # wait till you find a search box i.e. page has loaded completely
    element = WebDriverWait(browser, timeout=10,
poll_frequency=2).until(EC.presence_of_element_located((By.ID, "qsb-keyskill-sugg")))
except Exception as e:
    print(e)

print("Opening: {}".format("https://my.naukri.com/Profile/edit?id=&altresid"))
browser.get("https://my.naukri.com/Profile/edit?id=&altresid")
browser.find_element_by_xpath("/html/body/div[2]/div/div/span/div/div/div/div/div[2]/div[
1]/div/div/ul/li[12]/a").click()

# scroll to bottom of the page to load ajax elements
browser.execute_script("window.scrollTo(0, document.body.scrollHeight);")
time.sleep(5)

try:
    element = WebDriverWait(browser, timeout=10,
poll_frequency=2).until(EC.presence_of_element_located((By.ID, "attachCV")))
except Exception as e:
    print(e)
finally:
    # time.sleep(4)
    browser.find_element_by_id("attachCV").send_keys(CVPATH)
    # time.sleep(2)
```

In [11]:

```
#Libraries imported
import requests
from selenium import webdriver
from bs4 import BeautifulSoup
import time
import pandas as pd
from webdriver_manager.chrome import ChromeDriverManager
```

In [26]:

```
!pip install webdriver-manager
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager

driver = webdriver.Chrome(ChromeDriverManager().install())
from webdriver_manager.chrome import ChromeDriverManager
from webdriver_manager.core.utils import ChromeType
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
```

Requirement already satisfied: webdriver-manager in c:\programdata\anaconda3\lib\site-packages (3.7.0)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (from webdriver-manager) (2.26.0)
Requirement already satisfied: python-dotenv in c:\programdata\anaconda3\lib\site-packages (from webdriver-manager) (0.20.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (1.26.7)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (3.2)

```
[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - About to download new driver from https://chromedriver.storage.googleapis.com/102.0.5005.61/chromedriver_win32.zip (https://chromedriver.storage.googleapis.com/102.0.5005.61/chromedriver_win32.zip)
[WDM] - Driver has been saved in cache [C:\Users\hamsa\wdm\drivers\chromedriver\win32\102.0.5005.61]
```

In [13]:

```
url=("https://www.naukri.com/data-analyst-jobs-in-bangalore?k=data%20analyst&l=bangalore")
```

In [14]:

```
page = requests.get(url)
page.text
```

```
animation-duration: 1.8s; -o-animation-duration: 1.8s; -ms-animation-durat
ion: 1.8s; -webkit-animation-duration: 1.8s; -moz-animation-duration: 1.8
s; animation-iteration-count: infinite; -o-animation-iteration-count: infi
nite; -ms-animation-iteration-count: infinite; -webkit-animation-iteration
-count: infinite; -moz-animation-iteration-count: infinite; animation-dire
ction: normal; -o-animation-direction: normal; -ms-animation-direction: no
rmal; -webkit-animation-direction: normal; -moz-animation-direction: norma
l; border-radius: 50%; -o-border-radius: 50%; -ms-border-radius: 50%; -web
kit-border-radius: 50%; -moz-border-radius: 50% } #root, body, html { heig
ht: 100% } #circleG_1 { animation-delay: .45s; -o-animation-delay: .45s; -
ms-animation-delay: .45s; -webkit-animation-delay: .45s; -moz-animation-de
lay: .45s } #circleG_2 { animation-delay: .65s; -o-animation-delay: .65s;
-ms-animation-delay: .65s; -webkit-animation-delay: .65s; -moz-animation-d
elay: .65s } #circleG_3 { animation-delay: .85s; -o-animation-delay: .85s;
-ms-animation-delay: .85s; -webkit-animation-delay: .85s; -moz-animation-d
elay: .85s } #circleG_4 { animation-delay: 1.05s; -o-animation-delay: 1.05
s; -ms-animation-delay: 1.05s; -webkit-animation-delay: 1.05s; -moz-animat
ion-delay: 1.05s } #circleG_5 { animation-delay: 1.25s; -o-animation-dela
y: 1.25s; -ms-animation-delay: 1.25s; -webkit-animation-delay: 1.25s; -moz
-animation-delay: 1.25s } @keyframes bounce_circleG { 50% { width: 10px; h
```

In [15]:

```
driver = webdriver.Chrome(ChromeDriverManager().install())
driver.get(url)
time.sleep(3)
soup = BeautifulSoup(driver.page_source, 'html5lib')
print(soup.prettify())
driver.close()
```

```
<script defer="" src="https://www.naukri.com/akam/13/a137505" type="tex
t/javascript">
</script>
<script async="" src="//static.naukri.com/s/0/1/j/ub_v1.13.min.js">
</script>
<script async="" src="https://securepubads.g.doubleclick.net/gpt/pubads_
impl_2022060701.js">
</script>
<script charset="utf-8" crossorigin="anonymous" src="//static.nauking.co
m/s/7/109/j/vendors~Layout.8b6177ed.min.js">
</script>
<script charset="utf-8" crossorigin="anonymous" src="//static.nauking.co
m/s/7/109/j/Layout~companyPage~searchResult.8ee0ff3a.min.js">
</script>
<link crossorigin="anonymous" href="//static.nauking.com/s/7/109/c/Layou
t~QuickSearchBand~searchResult.b3416188.min.css" rel="stylesheet" type="te
xt/css"/>
<script charset="utf-8" crossorigin="anonymous" src="//static.nauking.co
m/s/7/109/j/Layout~QuickSearchBand~searchResult.bca0899b.min.js">
</script>
```

In [16]:

```
df = pd.DataFrame(columns=['Title', 'Company', 'Ratings', 'Reviews', 'Experience', 'Salary', 'Loc
```

In [17]:

```
results = soup.find(class_='list')
results
```

```

<div class="mt-7"><div class="left grey-text br2 placeholderLi experience"><i class="left icon-16 lh16 mr-4 naukicon naukicon-experience"></i><span class="ellipsis left fs12 lh16" title="5-8 Yrs ">5-8 Yrs</span></li><li class="left grey-text br2 placeholderLi salary"><i class="left icon-16 lh16 mr-4 naukicon naukicon-rupee"></i><span class="ellipsis left fs12 lh16" title="Not disclosed ">Not disclosed</span></li><li class="left grey-text br2 placeholderLi location"><i class="left icon-16 lh16 mr-4 naukicon naukicon-location"></i><span class="ellipsis left fs12 lh16" title="Bangalore/Bengaluru ">Bangalore/Bengaluru</span></li></ul></div></div><div class="job-description fs12 grey-text"><i class="left icon-16 lh16 mr-4 naukicon naukicon-desc"></i> Shift: Flexible timings IST, CET 8 years experience in Data AnalyticsGood knowledge of ...</div><ul class="tags has-description"><li class="left fs12 grey-text lh16 dot">Data Analytics</li><li class="left fs12 grey-text lh16 dot">Google Data Studio</li><li class="left fs12 grey-text lh16 dot">PowerBI</li><li class="left fs12 grey-text lh16 dot">analytical</li><li class="left fs12 grey-text lh16 dot">Tableau</li><li class="left fs12 grey-text lh16 dot">MS Excel</li><li class="left fs12 grey-text lh16 dot">JavaScript</li><li class="left fs12 grey-text lh16 dot">SQL</li></ul><div class="jobTupleFooter mt-20"><div class="type br2 left green"><i class="left mr-4 fs11 naukicon naukicon-history"></i>

```

In [18]:

```
job_elems = results.find_all('article',class_='jobTuple bgWhite br4 mb-8')

for param in range(1,11):
    url = "https://www.naukri.com/data-analyst-jobs-in-bangalore?k=data%20analyst&l=bangalore

    driver = webdriver.Chrome(ChromeDriverManager().install())
    driver.get(url)

    time.sleep(10)

    soup = BeautifulSoup(driver.page_source, 'html5lib')

    # print(soup.prettify())

    driver.close()

    results = soup.find(class_='list')
    job_elems = results.find_all('article',class_='jobTuple bgWhite br4 mb-8')

#Links for applying for the job

for job_elem in job_elems:

    # URL to apply for the job
    URL = job_elem.find('a',class_='title fw500 ellipsis').get('href')
    # print(URL.strip())

    # Post Title
    Title = job_elem.find('a',class_='title fw500 ellipsis')

    # Company Name
    Company = job_elem.find('a',class_='subTitle ellipsis fleft')

    # Ratings
    rating_span = job_elem.find('span',class_='starRating fleft dot')
    if rating_span is None:
        continue
    else:
        Ratings = rating_span.text

    # Reviews Counts
    Review_span = job_elem.find('a',class_='reviewsCount ml-5 fleft blue-text')
    if Review_span is None:
        continue
    else:
        Reviews = Review_span.text

    # Years of experience Required
    Exp = job_elem.find('li',class_='fleft grey-text br2 placeholderLi experience')
    Exp_span = Exp.find('span',class_='ellipsis fleft fs12 lh16')
    if Exp_span is None:
        continue
    else:
        Experience = Exp_span.text

    # Salary offered for the job
    Sal = job_elem.find('li',class_='fleft grey-text br2 placeholderLi salary')
    Sal_span = Sal.find('span',class_='ellipsis fleft fs12 lh16')
```

```

if Sal_span is None:
    continue
else:
    Salary = Sal_span.text

# Location for the job post
Loc = job_elem.find('li',class_='fleft grey-text br2 placeHolderLi location')
Loc_exp = Loc.find('span',class_='ellipsis fleft fs12 lh16')
if Loc_exp is None:
    continue
else:
    Location = Loc_exp.text

# Number of days since job posted
Hist = job_elem.find("div",["type br2 fleft grey","type br2 fleft green"])
Post_Hist = Hist.find('span',class_='fleft fw500')
if Post_Hist is None:
    continue
else:
    Post_History = Post_Hist.text

# Appending data to the DataFrame
df=df.append({'URL':URL,'Title':Title.text,'Company':Company.text,'Ratings':Ratings

```

```

[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\.wdm\drivers\chromedriver\win32\102.0.5005.61
\chromedriver.exe] found in cache
C:\Users\hamsa\AppData\Local\Temp\ipykernel_30556\1554958856.py:6: Deprecati
onWarning: executable_path has been deprecated, please pass in a Service obj
ect
driver = webdriver.Chrome(ChromeDriverManager().install())

```

```

[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\.wdm\drivers\chromedriver\win32\102.0.5005.61
\chromedriver.exe] found in cache

```

```

[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\.wdm\drivers\chromedriver\win32\102.0.5005.61
\chromedriver.exe] found in cache

```

```

[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\.wdm\drivers\chromedriver\win32\102.0.5005.61
\chromedriver.exe] found in cache

```



```
[WDM] - ===== WebDriver manager =====  
[WDM] - Current google-chrome version is 102.0.5005  
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome  
[WDM] - Driver [C:\Users\hamsa\.wdm\drivers\chromedriver\win32\102.0.5005.61  
\chromedriver.exe] found in cache
```

AttributeError Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_30556\1554958856.py in <module>

```
16  
17     results = soup.find(class_='list')  
--> 18     job_elems = results.find_all('article', class_='jobTuple bgWhite  
br4 mb-8')  
19  
20
```

AttributeError: 'NoneType' object has no attribute 'find_all'

In []:

```
df
```

In []:

```
df.columns
```

In []:

```
df.shape
```

In []:

```
df.info
```

In []:

```
Company_loc = df['Location']  
Company_loc
```

In []:

```
Salary_Det= df['Salary']  
Salary_Det
```

In []:

```
Company_name= df['Company']  
Company_name
```

In []:

```
column=df['Company']  
column.max()
```


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/> (<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

In []:

```
import configparser
import os
import sys
import time
import platform

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.firefox.options import Options
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.support.ui import WebDriverWait
```

In []:

```
#Libraries imported
import requests
from selenium import webdriver
from bs4 import BeautifulSoup
import time
import pandas as pd
from webdriver_manager.chrome import ChromeDriverManager
```

In []:

```
!pip install webdriver-manager
```

In []:

```
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager

driver = webdriver.Chrome(ChromeDriverManager().install())
from webdriver_manager.chrome import ChromeDriverManager
from webdriver_manager.core.utils import ChromeType
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
```

In []:

```
url=("https://www.naukri.com/data-scientist-jobs-in-bangalore?k=data%20scientist&l=bangalore")
```

In []:

```
page = requests.get(url)
page.text
```

In []:

```
driver = webdriver.Chrome(ChromeDriverManager().install())
driver.get(url)
time.sleep(3)
soup = BeautifulSoup(driver.page_source, 'html5lib')
print(soup.prettify())
driver.close()
```

In []:

```
df = pd.DataFrame(columns=['Title', 'Company', 'Ratings', 'Reviews', 'Experience', 'Salary', 'Loc
```

In []:

```
results = soup.find(class_='list')
results
```

In []:

```

job_elems = results.find_all('article',class_='jobTuple bgWhite br4 mb-8')

for param in range(1,11):
    url = "https://www.naukri.com/data-scientist-jobs-in-bangalore?k=data%20scientist&l=bangalore"

    driver = webdriver.Chrome(ChromeDriverManager().install())
    driver.get(url)

    time.sleep(10)

    soup = BeautifulSoup(driver.page_source, 'html5lib')

    # print(soup.prettify())

    driver.close()

    results = soup.find(class_='list')
    job_elems = results.find_all('article',class_='jobTuple bgWhite br4 mb-8')

#Links for applying for the job

for job_elem in job_elems:

    # URL to apply for the job
    URL = job_elem.find('a',class_='title fw500 ellipsis').get('href')
    # print(URL.strip())

    # Post Title
    Title = job_elem.find('a',class_='title fw500 ellipsis')

    # Company Name
    Company = job_elem.find('a',class_='subTitle ellipsis fleft')

    # Ratings
    rating_span = job_elem.find('span',class_='starRating fleft dot')
    if rating_span is None:
        continue
    else:
        Ratings = rating_span.text

    # Reviews Counts
    Review_span = job_elem.find('a',class_='reviewsCount ml-5 fleft blue-text')
    if Review_span is None:
        continue
    else:
        Reviews = Review_span.text

    # Years of experience Required
    Exp = job_elem.find('li',class_='fleft grey-text br2 placeholderLi experience')
    Exp_span = Exp.find('span',class_='ellipsis fleft fs12 lh16')
    if Exp_span is None:
        continue
    else:
        Experience = Exp_span.text

    # Salary offered for the job
    Sal = job_elem.find('li',class_='fleft grey-text br2 placeholderLi salary')
    Sal_span = Sal.find('span',class_='ellipsis fleft fs12 lh16')

```

```
if Sal_span is None:
    continue
else:
    Salary = Sal_span.text

# Location for the job post
Loc = job_elem.find('li',class_='fleft grey-text br2 placeHolderLi location')
Loc_exp = Loc.find('span',class_='ellipsis fleft fs12 lh16')
if Loc_exp is None:
    continue
else:
    Location = Loc_exp.text

# Number of days since job posted
Hist = job_elem.find("div",["type br2 fleft grey","type br2 fleft green"])
Post_Hist = Hist.find('span',class_='fleft fw500')
if Post_Hist is None:
    continue
else:
    Post_History = Post_Hist.text

# Appending data to the DataFrame
df=df.append({'URL':URL,'Title':Title.text,'Company':Company.text,'Ratings':Ratings
```

In []:

```
df
```

In []:

```
df.columns
```

In []:

```
df.shape
```

In []:

```
df.info
```

In []:

```
Company_loc = df['Location']
Company_loc
```

In []:

```
Salary_Det= df['Salary']
Salary_Det
```

In []:

```
Company_name= df[ 'Company' ]  
Company_name
```

3) webscrape

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/> (<https://www.naukri.com/>).
2. Enter “Data Scientist” in “Skill, Designations, and Companies” field.
3. Then click the search button.
4. Then apply the location filter and salary filter by checking the respective boxes
5. Then scrape the data for the first 10 jobs results you get.
6. Finally create a dataframe of the scraped data.

In []:

```
from selenium import webdriver  
from webdriver_manager.chrome import ChromeDriverManager  
  
driver = webdriver.Chrome(ChromeDriverManager().install())  
from webdriver_manager.chrome import ChromeDriverManager  
from webdriver_manager.core.utils import ChromeType  
from selenium import webdriver  
from bs4 import BeautifulSoup  
import pandas as pd  
from selenium import webdriver  
from bs4 import BeautifulSoup  
import pandas as pd
```

In []:

```
url=("https://www.naukri.com/data-scientist-jobs-in-delhi?k=data%20scientist&l=delhi")
```

In []:

```
page = requests.get(url)  
page.text
```

In []:

```
driver = webdriver.Chrome(ChromeDriverManager().install())  
driver.get(url)  
time.sleep(3)  
soup = BeautifulSoup(driver.page_source, 'html5lib')  
print(soup.prettify())  
driver.close()
```

In []:

```
df = pd.DataFrame(columns=['Title', 'Company', 'Ratings', 'Reviews', 'Experience', 'Salary', 'Loc
```

In []:

```
results = soup.find(class_='list')  
results
```

In []:

```
job_elems = results.find_all('article',class_='jobTuple bgWhite br4 mb-8')

for param in range(1,11):
    url = "https://www.naukri.com/data-scientist-jobs-in-delhi?k=data%20scientist&l=delhi"

    driver = webdriver.Chrome(ChromeDriverManager().install())
    driver.get(url)

    time.sleep(10)

    soup = BeautifulSoup(driver.page_source, 'html5lib')

    # print(soup.prettify())

    driver.close()

    results = soup.find(class_='list')
    job_elems = results.find_all('article',class_='jobTuple bgWhite br4 mb-8')

#Links for applying for the job

for job_elem in job_elems:

    # URL to apply for the job
    URL = job_elem.find('a',class_='title fw500 ellipsis').get('href')
    # print(URL.strip())

    # Post Title
    Title = job_elem.find('a',class_='title fw500 ellipsis')

    # Company Name
    Company = job_elem.find('a',class_='subTitle ellipsis fleft')

    # Ratings
    rating_span = job_elem.find('span',class_='starRating fleft dot')
    if rating_span is None:
        continue
    else:
        Ratings = rating_span.text

    # Reviews Counts
    Review_span = job_elem.find('a',class_='reviewsCount ml-5 fleft blue-text')
    if Review_span is None:
        continue
    else:
        Reviews = Review_span.text

    # Years of experience Required
    Exp = job_elem.find('li',class_='fleft grey-text br2 placeholderLi experience')
    Exp_span = Exp.find('span',class_='ellipsis fleft fs12 lh16')
    if Exp_span is None:
        continue
    else:
        Experience = Exp_span.text

    # Salary offered for the job
    Sal = job_elem.find('p',class_='grey-text LH20 fleft ml-8 txtLb1')
    Sal_span = Sal.find('span',class_='ellipsis fleft fs12 lh16')
```

```

if Sal_span is None:
    continue
else:
    Salary = Sal_span.text

# Location for the job post
Loc = job_elem.find('li',class_='fleft grey-text br2 placeHolderLi location')
Loc_exp = Loc.find('span',class_='ellipsis fleft fs12 lh16')
if Loc_exp is None:
    continue
else:
    Location = Loc_exp.text

# Number of days since job posted
Hist = job_elem.find("div",["type br2 fleft grey","type br2 fleft green"])
Post_Hist = Hist.find('span',class_='fleft fw500')
if Post_Hist is None:
    continue
else:
    Post_History = Post_Hist.text

# Appending data to the DataFrame
df=df.append({'URL':URL,'Title':Title.text,'Company':Company.text,'Ratings':Ratings

```

In [19]:

df

Out[19]:

Title	Company	Ratings	Reviews	Experience	Salary	Location	Job_Post_History	URL
-------	---------	---------	---------	------------	--------	----------	------------------	-----

In [20]:

df.columns

Out[20]:

```
Index(['Title', 'Company', 'Ratings', 'Reviews', 'Experience', 'Salary',
      'Location', 'Job_Post_History', 'URL'],
      dtype='object')
```

In [21]:

df.shape

Out[21]:

(0, 9)

In [22]:

```
df.info
```

Out[22]:

```
<bound method DataFrame.info of Empty DataFrame  
Columns: [Title, Company, Ratings, Reviews, Experience, Salary, Location, Jo  
b_Post_History, URL]  
Index: []>
```

In [23]:

```
Company_loc = df['Location']  
Company_loc
```

Out[23]:

```
Series([], Name: Location, dtype: object)
```

In [24]:

```
Salary_Det= df['Salary']  
Salary_Det
```

Out[24]:

```
Series([], Name: Salary, dtype: object)
```

In [25]:

```
Company_name= df['Company']  
Company_name
```

Out[25]:

```
Series([], Name: Company, dtype: object)
```

In [26]:

```
column=df['Company']  
column.max()
```

Out[26]:

```
nan
```

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

1. Brand
2. Product Description
3. Price 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 other page , then click on it.
8. Now scrape data from this page as usual
9. Repeat this until you get data for 100 sunglasses.

In [27]:

```
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager

driver = webdriver.Chrome(ChromeDriverManager().install())
from webdriver_manager.chrome import ChromeDriverManager
from webdriver_manager.core.utils import ChromeType
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
```

```
[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\wdm\drivers\chromedriver\win32\102.0.5005.61
\chromedriver.exe] found in cache
C:\Users\hamsa\AppData\Local\Temp\ipykernel_30556\368942963.py:4: Deprecatio
nWarning: executable_path has been deprecated, please pass in a Service obje
ct
driver = webdriver.Chrome(ChromeDriverManager().install())
```

In [28]:

```
import configparser
import os
import sys
import time
import platform

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.firefox.options import Options
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.support.ui import WebDriverWait
import requests
```


In [33]:

```

from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
!pip install bs4
!pip install requests

```

```

Requirement already satisfied: bs4 in c:\programdata\anaconda3\lib\site-pack
ages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\programdata\anaconda3\li
b\site-packages (from bs4) (4.10.0)
Requirement already satisfied: soupsieve>1.2 in c:\programdata\anaconda3\lib
\site-packages (from beautifulsoup4->bs4) (2.2.1)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site
-packages (2.26.0)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests) (1.26.7)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests) (2021.10.8)
Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests) (3.2)

```

In [105]:

```

from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager

driver = webdriver.Chrome(ChromeDriverManager().install())
from webdriver_manager.chrome import ChromeDriverManager
from webdriver_manager.core.utils import ChromeType
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd

```

```

[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\wdm\drivers\chromedriver\win32\102.0.5005.61\chromedriver.exe] found in cache
C:\Users\hamsa\AppData\Local\Temp\ipykernel_30556\859798927.py:4: Deprecatio
nWarning: executable_path has been deprecated, please pass in a Service obje
ct
  driver = webdriver.Chrome(ChromeDriverManager().install())

```

In [106]:

```
driver = webdriver.Chrome(ChromeDriverManager().install())
driver.get(url)
time.sleep(3)
soup = BeautifulSoup(driver.page_source, 'html5lib')
print(soup.prettify())
driver.close()
```

```
[WDM] - ===== WebDriver manager =====
[WDM] - Current google-chrome version is 102.0.5005
[WDM] - Get LATEST chromedriver version for 102.0.5005 google-chrome
[WDM] - Driver [C:\Users\hamsa\wdm\drivers\chromedriver\win32\102.0.5005.61\chromedriver.exe] found in cache
C:\Users\hamsa\AppData\Local\Temp\ipykernel_30556\667705507.py:1: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
  driver = webdriver.Chrome(ChromeDriverManager().install())
```

In [107]:

```
from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrome webdriver\chromedriver.exe")
```

```
C:\Users\hamsa\AppData\Local\Temp\ipykernel_30556\1902732820.py:2: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
  driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chromedriver.exe")
```

In [108]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chromedriver.exe")
time.sleep(2)
```

```
C:\Users\hamsa\AppData\Local\Temp\ipykernel_30556\440257495.py:1: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
  driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chromedriver.exe")
```

In [104]:

```
soup = BeautifulSoup(page.content)
```

In [111]:

```
soup = BeautifulSoup(page.content, 'html.parser')
#it gives us the visual representation of data
print(soup.prettify())
```

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <link href="https://rukminim1.flixcart.com" rel="preconnect"/>
    <link href="//static-assets-web.flixcart.com/www/linchpin/fk-cp-zion/css/app_modules.chunk.94b5e7.css" rel="stylesheet"/>
    <link href="//static-assets-web.flixcart.com/www/linchpin/fk-cp-zion/css/app.chunk.6e7580.css" rel="stylesheet"/>
    <meta content="text/html; charset=utf-8" http-equiv="Content-type"/>
    <meta content="IE=Edge" http-equiv="X-UA-Compatible"/>
    <meta content="102988293558" property="fb:page_id"/>
    <meta content="658873552,624500995,100000233612389" property="fb:admins"/>
    <meta content="noodp" name="robots"/>
    <link href="https://static-assets-web.flixcart.com/www/promos/new/20150528-140547-favicon-retina.ico" rel="shortcut icon"/>
    <link href="/osdd.xml?v=2" rel="search" type="application/opensearchdescription+xml"/>
    <meta content="website" property="og:type"/>
```

In [126]:

```
name = soup.find('div', class_="_2WkVRV")
```

In [127]:

```
name
```

Out[127]:

```
<div class="_2WkVRV">Mi</div>
```

In [128]:

```
name.text
```

Out[128]:

```
'Mi'
```

In [141]:

```
price=soup.find('div', class_="_3I9_wc")
```

In [142]:

```
price
```

Out[142]:

```
<div class="_3I9_wc">₹<!-- -->1,199</div>
```

In [143]:

price.text

Out[143]:

'₹1,199'

In [154]:

product_des = soup.find('div', class="_2B099V")

In [145]:

product_des

Out[145]:

```
<div class="_3E8aIl X3IECw row"><div class="_88-NUv">Reviews for Popular S
unglasses</div><div class="_2nRPpA"><div class="_1Ni40J"><div class="_3hVU
cF"><div class="CXW8mj" style="height:150px;width:150px"></div></div><div class
="_1kLt05"><a href="/mi-wayfarer-sunglasses/p/itmdc239447dbf07?pid=SGLFKFN
2RUBF84HB&marketplace=FLIPKART" rel="noopener noreferrer" target="_bla
nk"><div class="_1W9f5C"><div>1. Mi Wayfarer Sunglasses</div></div><div cl
ass="_3VDxyD"><div class="_3LWZlK">3.7</div><span clas
s="_34hpFu"><span>1,559 Ratings</span><span><span class="_2oY1qB">&lt;/s
pan><span>219 Reviews</span></span></span></div><div class="_2wYYVP"><div
class="_25b18c"><div class="_30jeq3 UMT9wN">₹599</div><div class="_3Ay6Sb
_2FuKQX"><span>40% off</span></div></div></div></a><ul class="_1Sq2Fs"><li
class=" 2OosNL">Frame Size: Free Size (One Size Fits All)</li><li class="
```

In [155]:

product_des.text

Out[155]:

'MiPolarized Aviator Sunglasses (Free Size)₹649₹1,19945% offFree deliveryEOS
S Special'

In [147]:

```
names= []  
  
for i in soup.find_all('div', class="_2WkVRV"):  
    names.append(i.text)
```

names

Out[147]:

```
['Mi',  
'Fastrack',  
'Mi',  
'ROYAL SON',  
'Elligator',  
'SRPM',  
'New Specs',  
'ROYAL SON',  
'PIRASO',  
'Fastrack',  
'ROZZETTA CRAFT',  
'ROYAL SON',  
'PIRASO',  
'PIRASO',  
'PIRASO',  
'SUNBEE',  
'ROYAL SON',  
'SHAAH COLLECTIONS',  
'PIRASO',  
'Lee Topper',  
'ROZZETTA CRAFT',  
'Mi',  
'SHAAH COLLECTIONS',  
'Fastrack',  
'VINCENT CHASE',  
'PIRASO',  
'SUNBEE',  
'ROZZETTA CRAFT',  
'ROYAL SON',  
'United Colors of Benetton',  
'Fastrack',  
'Fastrack',  
'ROZZETTA CRAFT',  
'ROZZETTA CRAFT',  
'New Specs',  
'Fastrack',  
'ROYAL SON',  
'Fastrack',  
'New Specs',  
'Fastrack',  
'ROYAL SON']
```


In [148]:

```
prices=[]  
  
for i in soup.find_all('div', class_="_3I9_wc"):  
    prices.append(i.text)  
  
prices
```

Out[148]:

```
['₹1,199',  
'₹999',  
'₹1,199',  
'₹1,999',  
'₹2,495',  
'₹1,299',  
'₹2,599',  
'₹1,999',  
'₹1,599',  
'₹1,299',  
'₹1,999',  
'₹1,999',  
'₹2,599',  
'₹2,599',  
'₹1,299',  
'₹1,999',  
'₹1,299',  
'₹1,599',  
'₹2,499',  
'₹1,999',  
'₹999',  
'₹1,699',  
'₹799',  
'₹1,999',  
'₹1,599',  
'₹1,299',  
'₹1,999',  
'₹1,999',  
'₹3,700',  
'₹799',  
'₹1,299',  
'₹1,999',  
'₹2,225',  
'₹1,599',  
'₹1,299',  
'₹1,999',  
'₹1,299',  
'₹1,299',  
'₹799',  
'₹1,999']
```

In [182]:

```
product_describes=[]

for i in soup.find_all('div', class="_2B099V"):

    product_describes.append(i.text)

product_describes
```

Out[182]:

```
['MiPolarized Aviator Sunglasses (Free Size)₹649₹1,19945% offFree deliveryEO
SS Special',
 'FastrackUV Protection Wayfarer Sunglasses (Free Size)₹699₹99930% offFree d
elivery',
 'MiPolarized Aviator Sunglasses (Free Size)₹649₹1,19945% offFree deliveryEO
SS Special',
 'ROYAL SONPolarized, UV Protection Wayfarer, Retro Square Sunglas...₹559₹1,
99972% offFree delivery',
 'ElligatorUV Protection Round Sunglasses (54)₹280₹2,49588% offFree delivery
EOSS Special',
 'SRPMUV Protection Wayfarer Sunglasses (50)₹199₹1,29984% offFree deliveryEO
SS Special',
 'New SpecsUV Protection Rectangular Sunglasses (Free Size)₹264₹2,59989% off
Free delivery',
 'ROYAL SONPolarized, UV Protection Wayfarer Sunglasses (61)₹559₹1,99972% of
fFree deliveryLowest price since launch',
 'PIRASOUV Protection Aviator Sunglasses (54)₹249₹1,59984% offFree deliveryE
OSS Special',
 'FastrackUV Protection Aviator Sunglasses (58)₹989₹1,29923% offFree deliver
y',
 'ROZZETTA CRAFTPolarized, Riding Glasses Sports, Wrap-around Sunglasse...₹4
49₹1,99977% offFree deliveryEOSS Special',
 'ROYAL SONUV Protection, Polarized Wayfarer Sunglasses (54)₹639₹1,99968% of
fFree deliveryLowest price since launch',
 'PIRASOUV Protection Aviator Sunglasses (58)₹319₹2,59987% offFree deliveryB
uy 2 items, save extra 10%',
 'PIRASOUV Protection Over-sized Sunglasses (65)₹410₹2,59984% offFree delive
ryEOSS Special',
 'SUNBEEUV Protection, Polarized Wayfarer Sunglasses (Free Size...₹283₹1,299
78% offFree delivery',
 'ROYAL SONPolarized, UV Protection Aviator Sunglasses (57)₹639₹1,99968% off
Free deliveryEOSS Special',
 'SHAAH COLLECTIONSUV Protection Round Sunglasses (Free Size)₹279₹1,29978% o
ffFree delivery',
 'PIRASOUV Protection Aviator Sunglasses (54)₹199₹1,59987% offFree deliveryE
OSS Special',
 'Lee TopperRiding Glasses Wrap-around Sunglasses (Free Size)₹299₹2,49988% o
ffFree delivery',
 'ROZZETTA CRAFTUV Protection Spectacle Sunglasses (Free Size)₹359₹1,99982%
offFree deliveryEOSS Special',
 'MiPolarized Wayfarer Sunglasses (Free Size)₹599₹99940% offFree deliveryEOS
S Special',
 'SHAAH COLLECTIONSUV Protection, Polarized, Mirrored Rectangular Sunglas
s...₹195₹1,69988% offFree delivery',
 'FastrackUV Protection Aviator Sunglasses (Free Size)₹629₹79921% offFree de
livery',
 'VINCENT CHASEUV Protection Rectangular Sunglasses (52)₹616₹1,99969% offFre
e delivery',
```

```
'PIRASOUV Protection Aviator Sunglasses (54)₹219₹1,59986% offFree deliveryEOSS Special',
'SUNBEEUV Protection, Polarized, Mirrored Retro Square Sunglas...₹259₹1,29980% offFree delivery',
'ROZZETTA CRAFTUV Protection, Gradient Retro Square Sunglasses (Free S...₹314₹1,99984% offFree deliveryEOSS Special',
'ROYAL SONPolarized, UV Protection Retro Square Sunglasses (55)₹519₹1,99974% offFree deliveryEOSS Special',
'United Colors of BenettonMirrored, UV Protection Round Sunglasses (57)₹669₹3,70081% offFree deliveryEOSS Special',
'FastrackUV Protection Wayfarer Sunglasses (Free Size)₹639₹79920% offFree delivery',
'FastrackUV Protection Round Sunglasses (52)₹909₹1,29930% offFree deliveryEOSS Special',
'ROZZETTA CRAFTUV Protection, Gradient Round Sunglasses (Free Size)₹359₹1,99982% offFree deliveryEOSS Special',
'ROZZETTA CRAFTUV Protection Retro Square Sunglasses (Free Size)₹449₹2,22579% offFree deliveryEOSS Special',
'New SpecsMirrored, UV Protection, Riding Glasses, Others Round S...₹299₹1,59981% offFree delivery',
'FastrackUV Protection Aviator Sunglasses (58)₹899₹1,29930% offFree deliveryEOSS Special',
'ROYAL SONPolarized, UV Protection Retro Square Sunglasses (62)₹719₹1,99964% offFree deliveryEOSS Special',
'FastrackUV Protection Aviator Sunglasses (58)₹1,025₹1,29921% offFree delivery',
'New SpecsUV Protection Round Sunglasses (Free Size)₹265₹1,29979% offFree delivery',
'FastrackUV Protection Wayfarer Sunglasses (Free Size)₹639₹79920% offFree delivery',
'ROYAL SONGradient Retro Square Sunglasses (60)₹559₹1,99972% offFree delivery']
```

In [183]:

```
print(len(names), len(prices), len(product_describes))
```

40 40 40

In [186]:

```
df = pd.DataFrame({'names':names, 'prices':prices, 'product_describes':product_describes})
```

In [187]:

```
print(len(names),len(prices),len(product_describes))
```

40 40 40

In [188]:

```
# df = pd.DataFrame({'Product Name':products,'Price':prices,'Rating':ratings})
df
```

Out[188]:

	names	prices	product_describes
0	Mi	₹1,199	MiPolarized Aviator Sunglasses (Free Size)₹649...
1	Fastrack	₹999	FastrackUV Protection Wayfarer Sunglasses (Fre...
2	Mi	₹1,199	MiPolarized Aviator Sunglasses (Free Size)₹649...
3	ROYAL SON	₹1,999	ROYAL SONPolarized, UV Protection Wayfarer, Re...
4	Elligator	₹2,495	ElligatorUV Protection Round Sunglasses (54)₹2...
5	SRPM	₹1,299	SRPMUV Protection Wayfarer Sunglasses (50)₹199...
6	New Specs	₹2,599	New SpecsUV Protection Rectangular Sunglasses ...
7	ROYAL SON	₹1,999	ROYAL SONPolarized, UV Protection Wayfarer Sun...
8	PIRASO	₹1,599	PIRASOUV Protection Aviator Sunglasses (54)₹24...
9	Fastrack	₹1,299	FastrackUV Protection Aviator Sunglasses (58)₹...
10	ROZZETTA CRAFT	₹1,999	ROZZETTA CRAFTPolarized, Riding Glasses Sports...
11	ROYAL SON	₹1,999	ROYAL SONUV Protection, Polarized Wayfarer Sun...
12	PIRASO	₹2,599	PIRASOUV Protection Aviator Sunglasses (58)₹31...
13	PIRASO	₹2,599	PIRASOUV Protection Over-sized Sunglasses (65)...
14	SUNBEE	₹1,299	SUNBEEUV Protection, Polarized Wayfarer Sungla...
15	ROYAL SON	₹1,999	ROYAL SONPolarized, UV Protection Aviator Sung...
16	SHAAH COLLECTIONS	₹1,299	SHAAH COLLECTIONSUV Protection Round Sunglasse...
17	PIRASO	₹1,599	PIRASOUV Protection Aviator Sunglasses (54)₹19...
18	Lee Topper	₹2,499	Lee TopperRiding Glasses Wrap-around Sunglasse...
19	ROZZETTA CRAFT	₹1,999	ROZZETTA CRAFTUV Protection Spectacle Sunglas...
20	Mi	₹999	MiPolarized Wayfarer Sunglasses (Free Size)₹59...
21	SHAAH COLLECTIONS	₹1,699	SHAAH COLLECTIONSUV Protection, Polarized, Mir...
22	Fastrack	₹799	FastrackUV Protection Aviator Sunglasses (Free...
23	VINCENT CHASE	₹1,999	VINCENT CHASEUV Protection Rectangular Sunglas...
24	PIRASO	₹1,599	PIRASOUV Protection Aviator Sunglasses (54)₹21...
25	SUNBEE	₹1,299	SUNBEEUV Protection, Polarized, Mirrored Retro...
26	ROZZETTA CRAFT	₹1,999	ROZZETTA CRAFTUV Protection, Gradient Retro Sq...
27	ROYAL SON	₹1,999	ROYAL SONPolarized, UV Protection Retro Square...
28	United Colors of Benetton	₹3,700	United Colors of BenettonMirrored, UV Protecti...
29	Fastrack	₹799	FastrackUV Protection Wayfarer Sunglasses (Fre...
30	Fastrack	₹1,299	FastrackUV Protection Round Sunglasses (52)₹90...
31	ROZZETTA CRAFT	₹1,999	ROZZETTA CRAFTUV Protection, Gradient Round Su...
32	ROZZETTA CRAFT	₹2,225	ROZZETTA CRAFTUV Protection Retro Square Sungl...
33	New Specs	₹1,599	New SpecsMirrored, UV Protection, Riding Glass...

	names	prices	product_describes
34	Fastrack	₹1,299	FastrackUV Protection Aviator Sunglasses (58)₹...
35	ROYAL SON	₹1,999	ROYAL SONPolarized, UV Protection Retro Square...
36	Fastrack	₹1,299	FastrackUV Protection Aviator Sunglasses (58)₹...
37	New Specs	₹1,299	New SpecsUV Protection Round Sunglasses (Free ...
38	Fastrack	₹799	FastrackUV Protection Wayfarer Sunglasses (Fre...
39	ROYAL SON	₹1,999	ROYAL SONGradient Retro Square Sunglasses (60)...

In []:

Q5: 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.

In []:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In []:

```
from bs4 import BeautifulSoup
import urllib
```

In [172]:

```
from bs4 import BeautifulSoup
import requests
```

In []:

```
from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In []:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom
time.sleep(2)
```

In []:

```
# Creating an instance of webdriver for google chrome
driver = webdriver.Chrome()
```

In []:

```
# Using webdriver we'll now open the flipkart website in chrome
url = 'https://flipkart.com'
# We'll use the get method of driver and pass in the URL
driver.get(url)
```

In []:

In []:

```
# Creating a soup object using driver.page_source to retrieve the HTML text and then we'll
# the HTML.
# Creating a soup object using driver.page_source to retrieve the HTML text and then we'll
# the HTML.
soup = BeautifulSoup(driver.page_source, 'html.parser')
```

In []:

```
products=[] #List to store name of the product
prices=[] #List to store price of the product
ratings=[] #List to store rating of the product
review=[]

driver.get("https://www.flipkart.com/search?q=iphone11&otracker=search&otracker1=search&mar
```

In []:

```
page = requests.get(url)
```

In []:

```
page.content
```

In []:

```
import bs4
from bs4 import BeautifulSoup as bs
import requests
```

In []:

```
soup = bs(page.content, 'html.parser')
#it gives us the visual representation of data
print(soup.prettify())
```

In []:

```
!pip install bs4
!pip install requests
```

In []:

```
!pip install requests, pandas, BeautifulSoup
```

In []:

```
import requests
import pandas as pd
from bs4 import BeautifulSoup
```

In []:

```
base_url = 'https://www.flipkart.com/mobiles/pr?sid=tyy%2C4io&otracker=categorytree&p%5B%5D'

def for_all_pages(page_no):
    page_no_url = base_url + str(page_no)
    response = requests.get(page_no_url)

    if response.status_code != 200:
        print('Status code:', response.status_code)
        raise Exception('Failed to fetch web page ' + page_no_url)

    doc = BeautifulSoup(response.text)
    return doc
```

In []:

```
def mobile_name(doc):
    prase = doc_new.find_all('div', class_='_4rR01T')
    names = []
    for tag in prase:
        names.append(tag.text.strip())
    return names

def prices(doc):
    price = doc_new.find_all('div', class_='_30jeq3 _1_WHN1')
    price = []
    for i in price:
        price.append(i.text.strip())
    return price
```

In []:

```
def for_all_pages(page_no):
    page_no_url = base_url + str(page_no)
    response = requests.get(page_no_url)

    if response.status_code != 200:
        print('Status code:', response.status_code)
        raise Exception('Failed to fetch web page ' + page_no_url)

    doc_new = BeautifulSoup(response.text)
    mobile_names_new = mobile_name(doc)
    all_prices_new = prices(doc)
    all_ratings_new = rating(doc)
    first_page_new = mobile_description(doc)
    data = {'Name': mobile_names_new, #getting all the information in a dictionary
            'Prices':all_prices_new,
            'ratings':all_ratings_new,
            'Description':first_page_new}
    df = pd.DataFrame(data)
    csv = df.to_csv('all information.csv',index=None)
    return csv
    return
```

8.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”

In [3]:

```
import bs4
from bs4 import BeautifulSoup as bs
import requests
```

In [32]:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In [4]:

```
from bs4 import BeautifulSoup
import urllib
```

In [5]:

```
from bs4 import BeautifulSoup

import requests
```


In [33]:

```
from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [34]:

```
from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [35]:

```
!pip install webdriver-manager
```

```
Requirement already satisfied: webdriver-manager in c:\programdata\anaconda3\lib\site-packages (3.7.0)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (from webdriver-manager) (2.26.0)
Requirement already satisfied: python-dotenv in c:\programdata\anaconda3\lib\site-packages (from webdriver-manager) (0.20.0)
Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (3.2)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests->webdriver-manager) (1.26.7)
```

In [68]:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In [69]:

```
#lets first connect to webdriver

from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [70]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom
```

In [71]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom
time.sleep(2)
```

In [72]:

```
driver.close()
```

In [123]:

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

In [124]:

```
search_field_product=driver.find_element_by_class_name("twotabsearchtextbox")#job search for
search_field_product.send_keys("laptop")
```

```
-----
NoSuchElementException Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_35884\2638063707.py in <module>
----> 1 search_field_product=driver.find_element_by_class_name("twotabsearch
textbox")#job search for
      2 search_field_product.send_keys("laptop")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_class_name(self, name)
    764         stacklevel=2,
    765     )
--> 766     return self.find_element(by=By.CLASS_NAME, value=name)
    767
    768     def find_elements_by_class_name(self, name) -> List[WebElement]:
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
    1249         value = '[name="%s"]' % value
    1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
    1252         'using': by,
    1253         'value': value})['value']
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428     response = self.command_executor.execute(driver_command, par
ams)
    429     if response:
--> 430         self.error_handler.check_response(response)
    431         response['value'] = self._unwrap_value(
    432             response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245         alert_text = value['alert'].get('text')
    246         raise exception_class(message, screen, stacktrace, alert
_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247     raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:
```

```
NoSuchElementException: Message: no such element: Unable to locate element:
{"method":"css selector","selector":".twotabsearchtextbox"}
(Session info: chrome=102.0.5005.115)
```

Stacktrace:

Backtrace:

```
Ordinal0 [0x00E8D953+2414931]
Ordinal0 [0x00E1F5E1+1963489]
Ordinal0 [0x00D0C6B8+837304]
Ordinal0 [0x00D39500+1021184]
Ordinal0 [0x00D3979B+1021851]
Ordinal0 [0x00D66502+1205506]
Ordinal0 [0x00D544E4+1131748]
Ordinal0 [0x00D64812+1198098]
Ordinal0 [0x00D542B6+1131190]
```

```
Ordinal0 [0x00D2E860+976992]
Ordinal0 [0x00D2F756+980822]
GetHandleVerifier [0x010FCC62+2510274]
GetHandleVerifier [0x010EF760+2455744]
GetHandleVerifier [0x00F1EABA+551962]
GetHandleVerifier [0x00F1D916+547446]
Ordinal0 [0x00E25F3B+1990459]
Ordinal0 [0x00E2A898+2009240]
Ordinal0 [0x00E2A985+2009477]
Ordinal0 [0x00E33AD1+2046673]
BaseThreadInitThunk [0x75CA6739+25]
RtlGetFullPathName_UEx [0x773C8FEF+1215]
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [125]:

```
# Creating an instance of webdriver for google chrome
driver = webdriver.Chrome()
```

In [126]:

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

```
-----
NoSuchElementException Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_35884\1600633362.py in <module>
      1 #clicking on search button
----> 2 search_button=driver.find_element_by_xpath("/html/body/div[1]/heade
r/div/div[1]/div[2]/div/form/div[3]")
      3 search_button.click()

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_xpath(self, xpath)
    524         stacklevel=2,
    525     )
--> 526     return self.find_element(by=By.XPATH, value=xpath)
    527
    528     def find_elements_by_xpath(self, xpath) -> List[WebElement]:

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
    1249         value = '[name="%s"]' % value
    1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
    1252         'using': by,
    1253         'value': value})['value']

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428     response = self.command_executor.execute(driver_command, par
ams)
    429     if response:
--> 430         self.error_handler.check_response(response)
    431         response['value'] = self._unwrap_value(
    432             response.get('value', None))

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245         alert_text = value['alert'].get('text')
    246         raise exception_class(message, screen, stacktrace, alert
_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247         raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:
```

```
NoSuchElementException: Message: no such element: Unable to locate element:
{"method":"xpath","selector":"/html/body/div[1]/header/div/div[1]/div[2]/di
v/form/div[3]"}
(Session info: chrome=102.0.5005.115)
```

Stacktrace:

Backtrace:

```
Ordinal0 [0x00BFD953+2414931]
Ordinal0 [0x00B8F5E1+1963489]
Ordinal0 [0x00A7C6B8+837304]
Ordinal0 [0x00AA9500+1021184]
Ordinal0 [0x00AA979B+1021851]
Ordinal0 [0x00AD6502+1205506]
```

```
Ordinal0 [0x00AC44E4+1131748]
Ordinal0 [0x00AD4812+1198098]
Ordinal0 [0x00AC42B6+1131190]
Ordinal0 [0x00A9E860+976992]
Ordinal0 [0x00A9F756+980822]
GetHandleVerifier [0x00E6CC62+2510274]
GetHandleVerifier [0x00E5F760+2455744]
GetHandleVerifier [0x00C8EABA+551962]
GetHandleVerifier [0x00C8D916+547446]
Ordinal0 [0x00B95F3B+1990459]
Ordinal0 [0x00B9A898+2009240]
Ordinal0 [0x00B9A985+2009477]
Ordinal0 [0x00BA3AD1+2046673]
BaseThreadInitThunk [0x75CA6739+25]
RtlGetFullPathName_UEx [0x773C8FEF+1215]
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In []:

```
search_button=driver.find_element_by_xpath("/html/body/div[1]/header/div/div[1]/div[2]/div/
search_button.click()
```

In []:

```
cpu = driver.find_element_by_xpath('//span[@class="a-size-base a-color-base a-text-bold"]')
```

In []:

```
cpu.text
```

In []:

```
titles_names=[]
ratings=[]
prices=[]
```

In []:

```
titles_tags=driver.find_elements_by_xpath('//span[@class="a-size-medium a-color-base a-text
```

In [127]:

```
titles_tags[0:11]
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_35884\1657808106.py in <module>
----> 1 titles_tags[0:11]
```

NameError: name 'titles_tags' is not defined

In [128]:

```
for i in titles_tags:
    title=i.text
    title_names.append(title)
title_names[0:11]
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_35884\2519145539.py in <module>
----> 1 for i in titles_tags:
      2     title=i.text
      3     title_names.append(title)
      4 title_names[0:11]
      5
```

NameError: name 'titles_tags' is not defined

In [129]:

```
ratings_tags=driver.find_elements_by_xpath('//span[@id="acrCustomerReviewText"]')
```

In [130]:

```
ratings_tags[0:11]
```

Out[130]:

```
[]
```

In [131]:

```
for i in ratings_tags:
    rating=i.text
    ratings.append(title)
ratings[0:11]
```

Out[131]:

```
[]
```

In [132]:

```
prices_tags=driver.find_elements_by_xpath('//span[@class="a-offscreen"]')
```

In [133]:

```
prices_tags
```

Out[133]:

```
[]
```

In [134]:

```
for i in prices_tags:
    price=i.text
    prices.append(title)
prices[0:11]
```

Out[134]:

[]

In [135]:

```
print(len(titles_names), len(ratings), len(prices))
```

0 0 0

In [136]:

```
df=laptops=pd.DataFrame({})
laptops['titles_names']=titles_names
laptops['ratings']=ratings
laptops['prices']=prices
```

In [137]:

laptops

Out[137]:

<u>titles_names</u>	ratings	prices
---------------------	---------	--------

Q6: Scrape data for first 100 sneakers you find when you visit flipkart.com and search for “sneakers” in the search field.

You have to scrape 4 attributes of each sneaker:

1. Brand
2. Product Description
3. Price

In [138]:

```
! pip install selenium
```

```
Requirement already satisfied: selenium in c:\programdata\anaconda3\lib\site-packages (4.2.0)
Requirement already satisfied: trio~=0.17 in c:\programdata\anaconda3\lib\site-packages (from selenium) (0.21.0)
Requirement already satisfied: trio-websocket~=0.9 in c:\programdata\anaconda3\lib\site-packages (from selenium) (0.9.2)
Requirement already satisfied: urllib3[secure,socks]~=1.26 in c:\programdata\anaconda3\lib\site-packages (from selenium) (1.26.7)
Requirement already satisfied: sniffio in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.2.0)
Requirement already satisfied: outcome in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.1.0)
Requirement already satisfied: attrs>=19.2.0 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (21.2.0)
Requirement already satisfied: async-generator>=1.9 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.10)
Requirement already satisfied: cffi>=1.14 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.14.6)
Requirement already satisfied: idna in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (3.2)
Requirement already satisfied: sortedcontainers in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: pycparser in c:\programdata\anaconda3\lib\site-packages (from cffi>=1.14->trio~=0.17->selenium) (2.20)
Requirement already satisfied: wsproto>=0.14 in c:\programdata\anaconda3\lib\site-packages (from trio-websocket~=0.9->selenium) (1.1.0)
Requirement already satisfied: pyOpenSSL>=0.14 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (21.0.0)
Requirement already satisfied: cryptography>=1.3.4 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (3.4.8)
Requirement already satisfied: certifi in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (2021.10.8)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (1.7.1)
Requirement already satisfied: six>=1.5.2 in c:\programdata\anaconda3\lib\site-packages (from pyOpenSSL>=0.14->urllib3[secure,socks]~=1.26->selenium) (1.16.0)
Requirement already satisfied: h11<1,>=0.9.0 in c:\programdata\anaconda3\lib\site-packages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.13.0)
```

In [139]:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In [140]:

```
from selenium import webdriver  
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [141]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom  
time.sleep(2)
```

In [142]:

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

In [155]:

```
search_field_product=driver.find_element_by_class_name("_2M8cLY header-form-search")
search_field_product.send_keys("sneakers")
```

NoSuchElementException

Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_35884\4143858610.py in <module>

```
----> 1 search_field_product=driver.find_element_by_class_name("_2M8cLY header-form-search")
      2 search_field_product.send_keys("sneakers")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element_by_class_name(self, name)
```

```
764         stacklevel=2,
765     )
--> 766     return self.find_element(by=By.CLASS_NAME, value=name)
767
768     def find_elements_by_class_name(self, name) -> List[WebElement]:
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element(self, by, value)
```

```
1249         value = '[name="%s"]' % value
1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
1252         'using': by,
1253         'value': value})['value']
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)
```

```
428     response = self.command_executor.execute(driver_command, params)
429
--> 430     if response:
431         self.error_handler.check_response(response)
432         response['value'] = self._unwrap_value(
433             response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)
```

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247     raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

NoSuchElementException: Message: no such element: Unable to locate element:

{"method": "css selector", "selector": ". _2M8cLY header-form-search"}

(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00E8D953+2414931]
Ordinal0 [0x00E1F5E1+1963489]
Ordinal0 [0x00D0C6B8+837304]
Ordinal0 [0x00D39500+1021184]
Ordinal0 [0x00D3979B+1021851]
Ordinal0 [0x00D66502+1205506]
Ordinal0 [0x00D544E4+1131748]
Ordinal0 [0x00D64812+1198098]
Ordinal0 [0x00D542B6+1131190]
```

```
Ordinal0 [0x00D2E860+976992]  
Ordinal0 [0x00D2F756+980822]  
GetHandleVerifier [0x010FCC62+2510274]  
GetHandleVerifier [0x010EF760+2455744]  
GetHandleVerifier [0x00F1EABA+551962]  
GetHandleVerifier [0x00F1D916+547446]  
Ordinal0 [0x00E25F3B+1990459]  
Ordinal0 [0x00E2A898+2009240]  
Ordinal0 [0x00E2A985+2009477]  
Ordinal0 [0x00E33AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [160]:

```
search_button=driver.find_element_by_xpath("L0Z3Pu")
search_button.click()
```

NoSuchElementException

Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_35884\308818377.py in <module>

```
----> 1 search_button=driver.find_element_by_xpath("L0Z3Pu")
      2 search_button.click()
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element_by_xpath(self, xpath)

```
524         stacklevel=2,
525     )
--> 526     return self.find_element(by=By.XPATH, value=xpath)
527
528     def find_elements_by_xpath(self, xpath) -> List[WebElement]:
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element(self, by, value)

```
1249         value = '[name="%s"]' % value
1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
1252         'using': by,
1253         'value': value})['value']
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)

```
428     response = self.command_executor.execute(driver_command, params)
429
--> 430     if response:
431         self.error_handler.check_response(response)
432         response['value'] = self._unwrap_value(
433             response.get('value', None))
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247     raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

NoSuchElementException: Message: no such element: Unable to locate element:

{ "method": "xpath", "selector": "L0Z3Pu" }

(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00E8D953+2414931]
Ordinal0 [0x00E1F5E1+1963489]
Ordinal0 [0x00D0C6B8+837304]
Ordinal0 [0x00D39500+1021184]
Ordinal0 [0x00D3979B+1021851]
Ordinal0 [0x00D66502+1205506]
Ordinal0 [0x00D544E4+1131748]
Ordinal0 [0x00D64812+1198098]
Ordinal0 [0x00D542B6+1131190]
Ordinal0 [0x00D2E860+976992]
```

```
Ordinal0 [0x00D2F756+980822]
GetHandleVerifier [0x010FCC62+2510274]
GetHandleVerifier [0x010EF760+2455744]
GetHandleVerifier [0x00F1EABA+551962]
GetHandleVerifier [0x00F1D916+547446]
Ordinal0 [0x00E25F3B+1990459]
Ordinal0 [0x00E2A898+2009240]
Ordinal0 [0x00E2A985+2009477]
Ordinal0 [0x00E33AD1+2046673]
BaseThreadInitThunk [0x75CA6739+25]
RtlGetFullPathName_UEx [0x773C8FEF+1215]
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [168]:

```
brands=[]
product_des=[]
prices=[]
```

In [169]:

```
brand_tags=driver.find_elements_by_xpath('//div[@class="_2WkVRV"]')
brand_tags[0:4]
```

Out[169]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="c3d8665c-8b43-4eeb-a3ec-3eff5f4fa01f")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="d9d3f75b-be55-407b-8213-16834cd7cd0e")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="8a2e3f0f-44fd-4645-b7c2-01d1ba70bdc1")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="650524df-da78-4649-97c0-bdcf5bd1a120")>]
```

In [203]:

```
product_tags=driver.find_elements_by_xpath('//div[@class="_2B099V"]')
product_tags[0:4]
```

Out[203]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="295453be-c5d1-422b-91b8-30cacf4d6873")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="d117a6bb-47be-4c36-a9da-312b6c3843d1")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="b1432b86-35da-4550-9730-9eaab3541597")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="81897dee-b388-4ff2-b89b-4d1939e77442")>]
```

In [184]:

```
price_tags=driver.find_elements_by_xpath('//div[@class="_3I9_wc"]')
price_tags[0:4]
```

Out[184]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="0f6b5764-bcb8-4713-983f-c2263e0922c8")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="91dab84c-4ca8-479a-ad8d-814c1e58ab03")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="260cdafc-483c-48c2-96ef-834d64ffe0f4")>,
<selenium.webdriver.remote.webelement.WebElement (session="277e30ed671c5dbd
11158e0c00ce169a", element="f1c971d2-2265-4560-8df2-10d10962fb11")>]
```

In [173]:

```
for i in brand_tags:
    brand=i.text
    brands.append(brand)
brands[0:4]
```

Out[173]:

```
['HIGHLANDER', 'HOTSTYLE', 'Layasa', 'BRUTON']
```

In [177]:

```
for i in product_tags:
    product=i.text
    product_des.append(product)
product_des[0:4]
```

Out[177]:

```
['HIGHLANDER\nSneakers For Men\n₹756₹1,99062% off\nSize 6, 7, 8, 9, 10, 11',
'HOTSTYLE\nSneakers For Men\n₹258₹99874% off\nSize 6, 7, 8, 9, 10',
'Layasa\nSneakers For Men\n₹399₹99960% off\nFree delivery\nLowest price sin
ce launch',
'BRUTON\nModern Trendy Sneakers Shoes Sneakers For Men\n₹299₹1,29976% off\n
Crazy Deal\nSize 6, 7, 8, 9, 10']
```

In [178]:

```
for i in price_tags:
    price=i.text
    prices.append(price)
prices[0:4]
```

Out[178]:

```
['₹1,990', '₹998', '₹999', '₹1,299']
```

In [199]:

```
print(len('brands'), len('product_des'), len('prices'))
```

6 11 6

In [205]:

```
df=sneakers=pd.DataFrame({})
sneakers['brands']=brands
sneakers['product_des']=product_des
sneakers['prices']=prices
```

```
-----
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_35884\2046674830.py in <module>
      1 df=sneakers=pd.DataFrame({})
      2 sneakers['brands']=brands
----> 3 sneakers['product_des']=product_des
      4 sneakers['prices']=prices

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in __setitem__
__(self, key, value)
    3610         else:
    3611             # set column
-> 3612             self._set_item(key, value)
    3613
    3614         def _setitem_slice(self, key: slice, value):

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _set_item
(self, key, value)
    3782         ensure_homogeneity.
    3783         """
-> 3784         value = self._sanitize_column(value)
    3785
    3786         if (

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py in _sanitize
_column(self, value)
    4507
    4508         if is_list_like(value):
-> 4509             com.require_length_match(value, self.index)
    4510             return sanitize_array(value, self.index, copy=True, allow_2d
=True)
    4511

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\common.py in require_
length_match(data, index)
    529         """
    530         if len(data) != len(index):
--> 531             raise ValueError(
    532                 "Length of values "
    533                 f"({len(data)}) "
```

ValueError: Length of values (120) does not match length of index (40)

Q7: Go to the link - <https://www.myntra.com/shoes>
Set second Price filter and Color filter to "Black",

In [206]:

```
! pip install selenium
```

```
Requirement already satisfied: selenium in c:\programdata\anaconda3\lib\site-packages (4.2.0)
Requirement already satisfied: trio~=0.17 in c:\programdata\anaconda3\lib\site-packages (from selenium) (0.21.0)
Requirement already satisfied: urllib3[secure,socks]~=1.26 in c:\programdata\anaconda3\lib\site-packages (from selenium) (1.26.7)
Requirement already satisfied: trio-websocket~=0.9 in c:\programdata\anaconda3\lib\site-packages (from selenium) (0.9.2)
Requirement already satisfied: sniffio in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.2.0)
Requirement already satisfied: async-generator>=1.9 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.10)
Requirement already satisfied: sortedcontainers in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: idna in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (3.2)
Requirement already satisfied: outcome in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.1.0)
Requirement already satisfied: attrs>=19.2.0 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (21.2.0)
Requirement already satisfied: cffi>=1.14 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.14.6)
Requirement already satisfied: pycparser in c:\programdata\anaconda3\lib\site-packages (from cffi>=1.14->trio~=0.17->selenium) (2.20)
Requirement already satisfied: wsproto>=0.14 in c:\programdata\anaconda3\lib\site-packages (from trio-websocket~=0.9->selenium) (1.1.0)
Requirement already satisfied: pyOpenSSL>=0.14 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (21.0.0)
Requirement already satisfied: cryptography>=1.3.4 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (3.4.8)
Requirement already satisfied: certifi in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (2021.10.8)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (1.7.1)
Requirement already satisfied: six>=1.5.2 in c:\programdata\anaconda3\lib\site-packages (from pyOpenSSL>=0.14->urllib3[secure,socks]~=1.26->selenium) (1.16.0)
Requirement already satisfied: h11<1,>=0.9.0 in c:\programdata\anaconda3\lib\site-packages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.13.0)
```

In [2]:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In [3]:

```
from selenium import webdriver  
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [4]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom  
time.sleep(2)
```

In [5]:

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

In [6]:

```
search_field_product=driver.find_element_by_class_name("desktop-searchBar")  
search_field_product.send_keys("black shoes")
```

In [7]:

```
search_button=driver.find_element_by_xpath('//span[@class="myntraweb-sprite desktop-iconSea  
search_button.click()
```

In [8]:

```
brands=[]  
product_des=[]  
prices=[]
```

In [9]:

```
brand_tags=driver.find_elements_by_xpath('//*[id="mountRoot"]/div/div/div/main/div[2]/div[  
brand_tags[0:100]
```

Out[9]:

```
[]
```

In [10]:

```
product_tags=driver.find_elements_by_xpath('//span[@class="myntraweb-sprite pdp-productDeta  
product_tags[0:100]
```

Out[10]:

```
[]
```

In [11]:

```
price_tags=driver.find_elements_by_xpath('//span[@class="price-num"]')
price_tags[0:100]
```

Out[11]:

```
[<selenium.webdriver.remote.webelement.WebElement (session="37e69321267e44c9
654df2f1710504ad", element="8577477c-97b8-4a31-8651-cadcd79a18d6")>,
<selenium.webdriver.remote.webelement.WebElement (session="37e69321267e44c9
654df2f1710504ad", element="77f2f67b-18e0-4e5b-b14e-236a525f2e8f")>,
<selenium.webdriver.remote.webelement.WebElement (session="37e69321267e44c9
654df2f1710504ad", element="60618766-88a7-4d18-9faf-6bb2d2bc7c48")>,
<selenium.webdriver.remote.webelement.WebElement (session="37e69321267e44c9
654df2f1710504ad", element="4c83388b-52ff-4842-a195-5f86b5713065")>]
```

In [12]:

```
for i in brand_tags:
    brand=i.text
    brands.append(brand)
brands[0:100]
```

Out[12]:

[]

In [13]:

```
for i in product_tags:
    product=i.text
    product_des.append(product)
product_des[0:100]
```

Out[13]:

[]

In [14]:

```
for i in price_tags:
    price=i.text
    prices.append(price)
prices[0:100]
```

Out[14]:

['(20087)', '(734)', '(70)', '(1)']

In [15]:

```
df=shoes=pd.DataFrame({})
shoes['brands']=brands
shoes['product_des']=product_des
shoes['prices']=prices
```

9)First get the webpage (<https://www.ambitionbox.com/>)

After reaching to the next webpage, In place of “Search by Designations, Companies, Skills” enter
 “Data Scientist” and click on search button.
 You will reach to the following web page click on location and in place of “Search location” enter
 “Noida” and select location “Noida”.

In [16]:

```
! pip install selenium
```

```
Requirement already satisfied: selenium in c:\programdata\anaconda3\lib\site-packages (4.2.0)
Requirement already satisfied: urllib3[secure,socks]~=1.26 in c:\programdata\anaconda3\lib\site-packages (from selenium) (1.26.7)
Requirement already satisfied: trio-websocket~=0.9 in c:\programdata\anaconda3\lib\site-packages (from selenium) (0.9.2)
Requirement already satisfied: trio~=0.17 in c:\programdata\anaconda3\lib\site-packages (from selenium) (0.21.0)
Requirement already satisfied: sortedcontainers in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (2.4.0)
Requirement already satisfied: attrs>=19.2.0 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (21.2.0)
Requirement already satisfied: async-generator>=1.9 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.10)
Requirement already satisfied: cffi>=1.14 in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.14.6)
Requirement already satisfied: idna in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (3.2)
Requirement already satisfied: sniffio in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.2.0)
Requirement already satisfied: outcome in c:\programdata\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.1.0)
Requirement already satisfied: pycparser in c:\programdata\anaconda3\lib\site-packages (from cffi>=1.14->trio~=0.17->selenium) (2.20)
Requirement already satisfied: wsproto>=0.14 in c:\programdata\anaconda3\lib\site-packages (from trio-websocket~=0.9->selenium) (1.1.0)
Requirement already satisfied: pyOpenSSL>=0.14 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (21.0.0)
Requirement already satisfied: cryptography>=1.3.4 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (3.4.8)
Requirement already satisfied: certifi in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (2021.10.8)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\programdata\anaconda3\lib\site-packages (from urllib3[secure,socks]~=1.26->selenium) (1.7.1)
Requirement already satisfied: six>=1.5.2 in c:\programdata\anaconda3\lib\site-packages (from pyOpenSSL>=0.14->urllib3[secure,socks]~=1.26->selenium) (1.16.0)
Requirement already satisfied: h11<1,>=0.9.0 in c:\programdata\anaconda3\lib\site-packages (from wsproto>=0.14->trio-websocket~=0.9->selenium) (0.13.0)
```

In [24]:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In [18]:

```
from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [19]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom
time.sleep(2)
```

In []:

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

In [28]:

```
search_field_designation=driver.find_element_by_class_name("search_company_input js-menu-search")
search_field_designation.send_keys("data scientist")
```

```
-----
NoSuchElementException Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_30576\3543354709.py in <module>
----> 1 search_field_designation=driver.find_element_by_class_name("search_c
ompany_input js-menu-search tt-input")#job search for
      2 search_field_designation.send_keys("data scientist")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_class_name(self, name)
    764         stacklevel=2,
    765     )
--> 766     return self.find_element(by=By.CLASS_NAME, value=name)
    767
    768     def find_elements_by_class_name(self, name) -> List[WebElement]:
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
    1249         value = '[name="%s"]' % value
    1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
    1252         'using': by,
    1253         'value': value})['value']
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428     response = self.command_executor.execute(driver_command, par
ams)
    429     if response:
--> 430         self.error_handler.check_response(response)
    431         response['value'] = self._unwrap_value(
    432             response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245         alert_text = value['alert'].get('text')
    246         raise exception_class(message, screen, stacktrace, alert
_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247     raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:
```

```
NoSuchElementException: Message: no such element: Unable to locate element:
{"method":"css selector","selector":".search_company_input js-menu-search t
t-input"}
(Session info: chrome=102.0.5005.115)
Stacktrace:
Backtrace:
```

```
Ordinal0 [0x00B1D953+2414931]
Ordinal0 [0x00AAF5E1+1963489]
Ordinal0 [0x0099C6B8+837304]
Ordinal0 [0x009C9500+1021184]
Ordinal0 [0x009C979B+1021851]
Ordinal0 [0x009F6502+1205506]
Ordinal0 [0x009E44E4+1131748]
Ordinal0 [0x009F4812+1198098]
```

```
Ordinal0 [0x009E42B6+1131190]  
Ordinal0 [0x009BE860+976992]  
Ordinal0 [0x009BF756+980822]  
GetHandleVerifier [0x00D8CC62+2510274]  
GetHandleVerifier [0x00D7F760+2455744]  
GetHandleVerifier [0x00BAEABA+551962]  
GetHandleVerifier [0x00BAD916+547446]  
Ordinal0 [0x00AB5F3B+1990459]  
Ordinal0 [0x00ABA898+2009240]  
Ordinal0 [0x00ABA985+2009477]  
Ordinal0 [0x00AC3AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [29]:

```
search_field_loc=driver.find_element_by_class_name("filter-item")
search_field_loc.send_keys("Noida")
```

```
-----
NoSuchElementException Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_30576\2255017339.py in <module>
----> 1 search_field_loc=driver.find_element_by_class_name("filter-item")
      2 search_field_loc.send_keys("Noida")

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_class_name(self, name)
    764         stacklevel=2,
    765     )
--> 766     return self.find_element(by=By.CLASS_NAME, value=name)
    767
    768     def find_elements_by_class_name(self, name) -> List[WebElement]:

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
    1249         value = '[name="%s"]' % value
    1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
    1252         'using': by,
    1253         'value': value})['value']

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428     response = self.command_executor.execute(driver_command, par
ams)
    429     if response:
--> 430         self.error_handler.check_response(response)
    431         response['value'] = self._unwrap_value(
    432             response.get('value', None))

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245         alert_text = value['alert'].get('text')
    246         raise exception_class(message, screen, stacktrace, alert
_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247         raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:
```

NoSuchElementException: Message: no such element: Unable to locate element:
{"method":"css selector","selector":".filter-item"}
(Session info: chrome=102.0.5005.115)
Stacktrace:
Backtrace:

```
Ordinal0 [0x00B1D953+2414931]
Ordinal0 [0x00AAF5E1+1963489]
Ordinal0 [0x0099C6B8+837304]
Ordinal0 [0x009C9500+1021184]
Ordinal0 [0x009C979B+1021851]
Ordinal0 [0x009F6502+1205506]
Ordinal0 [0x009E44E4+1131748]
Ordinal0 [0x009F4812+1198098]
Ordinal0 [0x009E42B6+1131190]
```



```
Ordinal0 [0x009BE860+976992]  
Ordinal0 [0x009BF756+980822]  
GetHandleVerifier [0x00D8CC62+2510274]  
GetHandleVerifier [0x00D7F760+2455744]  
GetHandleVerifier [0x00BAEABA+551962]  
GetHandleVerifier [0x00BAD916+547446]  
Ordinal0 [0x00AB5F3B+1990459]  
Ordinal0 [0x00ABA898+2009240]  
Ordinal0 [0x00ABA985+2009477]  
Ordinal0 [0x00AC3AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [33]:

```
search_button=driver.find_element_by_xpath("ab_btn search-btn round")
search_button.click()
```

InvalidSelectorException Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_30576\4106752796.py in <module>

```
----> 1 search_button=driver.find_element_by_xpath("ab_btn search-btn round")
)
      2 search_button.click()
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element_by_xpath(self, xpath)

```
524         stacklevel=2,
525     )
--> 526     return self.find_element(by=By.XPATH, value=xpath)
527
528     def find_elements_by_xpath(self, xpath) -> List[WebElement]:
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element(self, by, value)

```
1249         value = '[name="%s"]' % value
1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
1252         'using': by,
1253         'value': value})['value']
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)

```
428     response = self.command_executor.execute(driver_command, params)
429
--> 430     if response:
431         self.error_handler.check_response(response)
432         response['value'] = self._unwrap_value(
433             response.get('value', None))
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247     raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

InvalidSelectorException: Message: invalid selector: Unable to locate an element with the xpath expression ab_btn search-btn round because of the following error:

SyntaxError: Failed to execute 'evaluate' on 'Document': The string 'ab_btn search-btn round' is not a valid XPath expression.

(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00B1D953+2414931]
Ordinal0 [0x00AAF5E1+1963489]
Ordinal0 [0x0099C6B8+837304]
Ordinal0 [0x0099F0B4+848052]
Ordinal0 [0x0099EF72+847730]
Ordinal0 [0x0099F200+848384]
```

```
Ordinal0 [0x009C9215+1020437]
Ordinal0 [0x009C979B+1021851]
Ordinal0 [0x009F6502+1205506]
Ordinal0 [0x009E44E4+1131748]
Ordinal0 [0x009F4812+1198098]
Ordinal0 [0x009E42B6+1131190]
Ordinal0 [0x009BE860+976992]
Ordinal0 [0x009BF756+980822]
GetHandleVerifier [0x00D8CC62+2510274]
GetHandleVerifier [0x00D7F760+2455744]
GetHandleVerifier [0x00BAEABA+551962]
GetHandleVerifier [0x00BAD916+547446]
Ordinal0 [0x00AB5F3B+1990459]
Ordinal0 [0x00ABA898+2009240]
Ordinal0 [0x00ABA985+2009477]
Ordinal0 [0x00AC3AD1+2046673]
BaseThreadInitThunk [0x75CA6739+25]
RtlGetFullPathName_UEx [0x773C8FEF+1215]
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [34]:

```
job_titles=[]
company_names=[]
locations_list=[]
experience_list=[]
```

In [38]:

```
titles_tags=driver.find_elements_by_xpath('//p[@class="name bold-display overflow-wrap head
titles_tags[0:4]
```

Out[38]:

```
[]
```

In [39]:

```
#so we will run loop to iterate over the tags extracted above and extract the text inside t

for i in titles_tags:
    title=i.text
    job_titles.append(title)
job_titles[0:4]
```

Out[39]:

```
[]
```


In [47]:

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

0 0 0 0

In [48]:

```
jobs=pd.DataFrame({})
jobs['title']=job_titles
jobs['company']=company_names
jobs['experience_required']=experience_list
jobs['location']=locations_list
```

In [49]:

jobs

Out[49]:

title	company	experience_required	location
-------	---------	---------------------	----------

Q10: 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

In [50]:

```
import selenium
import pandas as pd
from selenium import webdriver
import warnings
warnings.filterwarnings("ignore")
import time
```

In [51]:

```
from selenium import webdriver
driver = webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chr
```

In [52]:

```
driver=webdriver.Chrome(r"C:\Users\hamsa\Desktop\datascience project\chrome webdriver\chrom
time.sleep(2)
```

In [54]:

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

```
-----
WebDriverException                               Traceback (most recent call last)
```

```
~\AppData\Local\Temp\ipykernel_30576\3320636895.py in <module>
```

```
----> 1 driver.get('https://www.ambitionbox.com/')
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in get(self, url)
```

```
440         Loads a web page in the current browser session.
```

```
441         """
```

```
--> 442         self.execute(Command.GET, {'url': url})
```

```
443
```

```
444         @property
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)
```

```
428         response = self.command_executor.execute(driver_command, params)
```

```
429         if response:
```

```
--> 430         self.error_handler.check_response(response)
```

```
431         response['value'] = self._unwrap_value(
```

```
432             response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)
```

```
245         alert_text = value['alert'].get('text')
```

```
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
```

```
--> 247         raise exception_class(message, screen, stacktrace)
```

```
248
```

```
249         def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

```
WebDriverException: Message: chrome not reachable
```

```
(Session info: chrome=102.0.5005.115)
```

```
Stacktrace:
```

```
Backtrace:
```

```
Ordinal0 [0x00CAD953+2414931]
```

```
Ordinal0 [0x00C3F5E1+1963489]
```

```
Ordinal0 [0x00B2C580+836992]
```

```
Ordinal0 [0x00B201E2+786914]
```

```
Ordinal0 [0x00B20A18+789016]
```

```
Ordinal0 [0x00B222F2+795378]
```

```
Ordinal0 [0x00B1BCA9+769193]
```

```
Ordinal0 [0x00B2DA30+842288]
```

```
Ordinal0 [0x00B83FF2+1196018]
```

```
Ordinal0 [0x00B742B6+1131190]
```

```
Ordinal0 [0x00B4E860+976992]
```

```
Ordinal0 [0x00B4F756+980822]
```

```
GetHandleVerifier [0x00F1CC62+2510274]
```

```
GetHandleVerifier [0x00F0F760+2455744]
```

```
GetHandleVerifier [0x00D3EABA+551962]
```

```
GetHandleVerifier [0x00D3D916+547446]
```

```
Ordinal0 [0x00C45F3B+1990459]
```

```
Ordinal0 [0x00C4A898+2009240]
```

```
Ordinal0 [0x00C4A985+2009477]
```

```
Ordinal0 [0x00C53AD1+2046673]
```

```
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [55]:

```
search_field_designation=driver.find_element_by_class_name("search_company_input js-menu-se
search_field_designation.send_keys("data scientist")
```

```
-----
WebDriverException                               Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_30576\3543354709.py in <module>
----> 1 search_field_designation=driver.find_element_by_class_name("search_c
ompany_input js-menu-search tt-input")#job search for
      2 search_field_designation.send_keys("data scientist")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_class_name(self, name)
    764         stacklevel=2,
    765     )
--> 766     return self.find_element(by=By.CLASS_NAME, value=name)
    767
    768     def find_elements_by_class_name(self, name) -> List[WebElement]:
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
    1249         value = '[name="%s"]' % value
    1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
    1252         'using': by,
    1253         'value': value})['value']
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428         response = self.command_executor.execute(driver_command, par
ams)
    429         if response:
--> 430             self.error_handler.check_response(response)
    431             response['value'] = self._unwrap_value(
    432                 response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245         alert_text = value['alert'].get('text')
    246         raise exception_class(message, screen, stacktrace, alert
_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247         raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:
```

WebDriverException: Message: chrome not reachable
(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
Ordinal0 [0x00B742B6+1131190]
```



```
Ordinal0 [0x00B4E860+976992]  
Ordinal0 [0x00B4F756+980822]  
GetHandleVerifier [0x00F1CC62+2510274]  
GetHandleVerifier [0x00F0F760+2455744]  
GetHandleVerifier [0x00D3EABA+551962]  
GetHandleVerifier [0x00D3D916+547446]  
Ordinal0 [0x00C45F3B+1990459]  
Ordinal0 [0x00C4A898+2009240]  
Ordinal0 [0x00C4A985+2009477]  
Ordinal0 [0x00C53AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [56]:

```
search_field_loc=driver.find_element_by_class_name("filter-item")
search_field_loc.send_keys("Noida")
```

WebDriverException

Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_30576\3375496557.py in <module>

```
----> 1 search_field_loc=driver.find_element_by_class_name("filter-item")
      2 search_field_loc.send_keys("Noida")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element_by_class_name(self, name)
```

```
764         stacklevel=2,
765     )
--> 766     return self.find_element(by=By.CLASS_NAME, value=name)
767
768     def find_elements_by_class_name(self, name) -> List[WebElement]:
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_element(self, by, value)
```

```
1249         value = '[name="%s"]' % value
1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
1252         'using': by,
1253         'value': value})['value']
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)
```

```
428     response = self.command_executor.execute(driver_command, params)
429
--> 430     if response:
431         self.error_handler.check_response(response)
432         response['value'] = self._unwrap_value(
433             response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)
```

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247     raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

WebDriverException: Message: chrome not reachable
(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
Ordinal0 [0x00B742B6+1131190]
Ordinal0 [0x00B4E860+976992]
```

```
Ordinal0 [0x00B4F756+980822]  
GetHandleVerifier [0x00F1CC62+2510274]  
GetHandleVerifier [0x00F0F760+2455744]  
GetHandleVerifier [0x00D3EABA+551962]  
GetHandleVerifier [0x00D3D916+547446]  
Ordinal0 [0x00C45F3B+1990459]  
Ordinal0 [0x00C4A898+2009240]  
Ordinal0 [0x00C4A985+2009477]  
Ordinal0 [0x00C53AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [57]:

```
search_button=driver.find_element_by_xpath("ab_btn search-btn round")
search_button.click()
```

```
-----
WebDriverException                               Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_30576\4106752796.py in <module>
----> 1 search_button=driver.find_element_by_xpath("ab_btn search-btn round"
)
      2 search_button.click()

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element_by_xpath(self, xpath)
    524         stacklevel=2,
    525     )
--> 526     return self.find_element(by=By.XPATH, value=xpath)
    527
    528     def find_elements_by_xpath(self, xpath) -> List[WebElement]:

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in find_element(self, by, value)
    1249         value = '[name="%s"]' % value
    1250
-> 1251     return self.execute(Command.FIND_ELEMENT, {
    1252         'using': by,
    1253         'value': value})['value']

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriv
er.py in execute(self, driver_command, params)
    428     response = self.command_executor.execute(driver_command, par
ams)
    429     if response:
--> 430         self.error_handler.check_response(response)
    431         response['value'] = self._unwrap_value(
    432             response.get('value', None))

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorha
ndler.py in check_response(self, response)
    245         alert_text = value['alert'].get('text')
    246         raise exception_class(message, screen, stacktrace, alert
_text) # type: ignore[call-arg] # mypy is not smart enough here
--> 247         raise exception_class(message, screen, stacktrace)
    248
    249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, de
fault: _VT) -> _VT:
```

WebDriverException: Message: chrome not reachable
(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
Ordinal0 [0x00B742B6+1131190]
```

```
Ordinal0 [0x00B4E860+976992]  
Ordinal0 [0x00B4F756+980822]  
GetHandleVerifier [0x00F1CC62+2510274]  
GetHandleVerifier [0x00F0F760+2455744]  
GetHandleVerifier [0x00D3EABA+551962]  
GetHandleVerifier [0x00D3D916+547446]  
Ordinal0 [0x00C45F3B+1990459]  
Ordinal0 [0x00C4A898+2009240]  
Ordinal0 [0x00C4A985+2009477]  
Ordinal0 [0x00C53AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [62]:

```
job_titles=[]  
company_names=[]  
locations_list=[]  
salary_list=[]
```

In [63]:

```
titles_tags=driver.find_elements_by_xpath('//p[@class="name bold-display overflow-wrap head
titles_tags[0:4]
```

```
-----
WebDriverException                                Traceback (most recent call last)
```

```
~\AppData\Local\Temp\ipykernel_30576\1457052961.py in <module>
----> 1 titles_tags=driver.find_elements_by_xpath('//p[@class="name bold-d
isplay overflow-wrap headerTextWrapper marginFix mg-btm-4"]') #locating we
b element of title
      2 titles_tags[0:4]
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdr
iver.py in find_elements_by_xpath(self, xpath)
```

```
547         stacklevel=2,
548     )
--> 549     return self.find_elements(by=By.XPATH, value=xpath)
550
551     def find_element_by_link_text(self, link_text) -> WebElement:
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdr
iver.py in find_elements(self, by, value)
```

```
1282         # Return empty list if driver returns null
1283         # See https://github.com/SeleniumHQ/selenium/issues/4555
1284         (https://github.com/SeleniumHQ/selenium/issues/4555)
-> 1284         return self.execute(Command.FIND_ELEMENTS, {
1285             'using': by,
1286             'value': value})['value'] or []
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdr
iver.py in execute(self, driver_command, params)
```

```
428         response = self.command_executor.execute(driver_command, p
arams)
429         if response:
--> 430             self.error_handler.check_response(response)
431             response['value'] = self._unwrap_value(
432                 response.get('value', None))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\error
handler.py in check_response(self, response)
```

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, ale
rt_text) # type: ignore[call-arg] # mypy is not smart enough here
-> 247         raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT,
default: _VT) -> _VT:
```

```
WebDriverException: Message: chrome not reachable
(Session info: chrome=102.0.5005.115)
```

```
Stacktrace:
```

```
Backtrace:
```

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
```

```
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
Ordinal0 [0x00B742B6+1131190]
Ordinal0 [0x00B4E860+976992]
Ordinal0 [0x00B4F756+980822]
GetHandleVerifier [0x00F1CC62+2510274]
GetHandleVerifier [0x00F0F760+2455744]
GetHandleVerifier [0x00D3EABA+551962]
GetHandleVerifier [0x00D3D916+547446]
Ordinal0 [0x00C45F3B+1990459]
Ordinal0 [0x00C4A898+2009240]
Ordinal0 [0x00C4A985+2009477]
Ordinal0 [0x00C53AD1+2046673]
BaseThreadInitThunk [0x75CA6739+25]
RtlGetFullPathName_UEx [0x773C8FEF+1215]
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [60]:

```
for i in titles_tags:
    title=i.text
    job_titles.append(title)
job_titles[0:4]
```

Out[60]:

```
[]
```

```
companies_tags=driver.find_elements_by_xpath('/html/body/div/div/div/div[2]/div[3]/div[1]/div')  
companies_tags[0:4]
```

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
```



```
Ordinal0 [0x00B742B6+1131190]  
Ordinal0 [0x00B4E860+976992]  
Ordinal0 [0x00B4F756+980822]  
GetHandleVerifier [0x00F1CC62+2510274]  
GetHandleVerifier [0x00F0F760+2455744]  
GetHandleVerifier [0x00D3EABA+551962]  
GetHandleVerifier [0x00D3D916+547446]  
Ordinal0 [0x00C45F3B+1990459]  
Ordinal0 [0x00C4A898+2009240]  
Ordinal0 [0x00C4A985+2009477]  
Ordinal0 [0x00C53AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [64]:

```
locations_tags=driver.find_elements_by_xpath("/html/body/div/div/div/div[2]/div[3]/div[1]/div/div/div/div/div[5]/div")
locations_tags[0:4]
```

WebDriverException Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_30576\3875427482.py in <module>

```
----> 1 locations_tags=driver.find_elements_by_xpath("/html/body/div/div/div/div[2]/div[3]/div[1]/div/div/div/div/div[5]/div")
      2 locations_tags[0:4]
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_elements_by_xpath(self, xpath)

```
547         stacklevel=2,
548     )
--> 549     return self.find_elements(by=By.XPATH, value=xpath)
550
551     def find_element_by_link_text(self, link_text) -> WebElement:
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_elements(self, by, value)

```
1282         # Return empty list if driver returns null
1283         # See https://github.com/SeleniumHQ/selenium/issues/4555 (https://github.com/SeleniumHQ/selenium/issues/4555)
-> 1284         return self.execute(Command.FIND_ELEMENTS, {
1285             'using': by,
1286             'value': value})['value'] or []
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)

```
428         response = self.command_executor.execute(driver_command, params)
429         if response:
--> 430             self.error_handler.check_response(response)
431             response['value'] = self._unwrap_value(
432                 response.get('value', None))
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
-> 247         raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

WebDriverException: Message: chrome not reachable

(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
```

```
Ordinal0 [0x00B742B6+1131190]  
Ordinal0 [0x00B4E860+976992]  
Ordinal0 [0x00B4F756+980822]  
GetHandleVerifier [0x00F1CC62+2510274]  
GetHandleVerifier [0x00F0F760+2455744]  
GetHandleVerifier [0x00D3EABA+551962]  
GetHandleVerifier [0x00D3D916+547446]  
Ordinal0 [0x00C45F3B+1990459]  
Ordinal0 [0x00C4A898+2009240]  
Ordinal0 [0x00C4A985+2009477]  
Ordinal0 [0x00C53AD1+2046673]  
BaseThreadInitThunk [0x75CA6739+25]  
RtlGetFullPathName_UEx [0x773C8FEF+1215]  
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [65]:

```
for i in locations_tags :  
    location=i.text  
    locations_list.append(location)  
locations_list[0:4]
```

Out[65]:

[]

In [80]:

```
salary_tags=driver.find_elements_by_xpath('//div[@class="row_wrap"]')
salary_tags[0:4]
```

WebDriverException

Traceback (most recent call last)

~\AppData\Local\Temp\ipykernel_30576\2385890205.py in <module>

```
----> 1 salary_tags=driver.find_elements_by_xpath('//div[@class="row_wrap"]')
)
      2 salary_tags[0:4]
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_elements_by_xpath(self, xpath)

```
547         stacklevel=2,
548     )
--> 549     return self.find_elements(by=By.XPATH, value=xpath)
550
551     def find_element_by_link_text(self, link_text) -> WebElement:
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in find_elements(self, by, value)

```
1282         # Return empty list if driver returns null
1283         # See https://github.com/SeleniumHQ/selenium/issues/4555 (https://github.com/SeleniumHQ/selenium/issues/4555)
-> 1284         return self.execute(Command.FIND_ELEMENTS, {
1285             'using': by,
1286             'value': value})['value'] or []
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\webdriver.py in execute(self, driver_command, params)

```
428         response = self.command_executor.execute(driver_command, params)
429         if response:
--> 430             self.error_handler.check_response(response)
431             response['value'] = self._unwrap_value(
432                 response.get('value', None))
```

C:\ProgramData\Anaconda3\lib\site-packages\selenium\webdriver\remote\errorhandler.py in check_response(self, response)

```
245         alert_text = value['alert'].get('text')
246         raise exception_class(message, screen, stacktrace, alert_text) # type: ignore[call-arg] # mypy is not smart enough here
-> 247         raise exception_class(message, screen, stacktrace)
248
249     def _value_or_default(self, obj: Mapping[_KT, _VT], key: _KT, default: _VT) -> _VT:
```

WebDriverException: Message: chrome not reachable

(Session info: chrome=102.0.5005.115)

Stacktrace:

Backtrace:

```
Ordinal0 [0x00CAD953+2414931]
Ordinal0 [0x00C3F5E1+1963489]
Ordinal0 [0x00B2C580+836992]
Ordinal0 [0x00B201E2+786914]
Ordinal0 [0x00B20A18+789016]
Ordinal0 [0x00B222F2+795378]
Ordinal0 [0x00B1BCA9+769193]
Ordinal0 [0x00B2DA30+842288]
Ordinal0 [0x00B83FF2+1196018]
```

```
Ordinal0 [0x00B742B6+1131190]
Ordinal0 [0x00B4E860+976992]
Ordinal0 [0x00B4F756+980822]
GetHandleVerifier [0x00F1CC62+2510274]
GetHandleVerifier [0x00F0F760+2455744]
GetHandleVerifier [0x00D3EABA+551962]
GetHandleVerifier [0x00D3D916+547446]
Ordinal0 [0x00C45F3B+1990459]
Ordinal0 [0x00C4A898+2009240]
Ordinal0 [0x00C4A985+2009477]
Ordinal0 [0x00C53AD1+2046673]
BaseThreadInitThunk [0x75CA6739+25]
RtlGetFullPathName_UEx [0x773C8FEF+1215]
RtlGetFullPathName_UEx [0x773C8FBD+1165]
```

In [81]:

```
for i in salary_tags :
    salary=i.text
    salary_list.append(salary)
    salary_list
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_30576\2625213348.py in <module>
----> 1 for i in salary_tags :
      2     salary=i.text
      3     salary_list.append(salary)
      4     salary_list
```

NameError: name 'salary_tags' is not defined

In [82]:

```
print(len(job_titles), len(company_names), len(locations_list), len(salary_list))
```

0 0 0 0

In [83]:

```
jobs=pd.DataFrame({})
jobs['title']=job_titles
jobs['company']=company_names
jobs['salary']=salary_list
jobs['location']=locations_list
```

In [84]:

```
jobs
```

Out[84]:

```
title  company  salary  location
```

In []: