Asignment 2 Webscraping

In [2]:
```
!pip install selenium
```

```
Requirement already satisfied: selenium in c:\users\99minds-1\appdata\local\p
rograms\python\python311\lib\site-packages (4.14.0)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in c:\users\99minds-1
\appdata\local\programs\python\python311\lib\site-packages (from selenium)
(1.26.15)
Requirement already satisfied: trio~=0.17 in c:\users\99minds-1\appdata\local
\programs\python\python311\lib\site-packages (from selenium) (0.22.2)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\99minds-1\appd
ata\local\programs\python\python311\lib\site-packages (from selenium) (0.11.
1)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\99minds-1\appda
ta\local\programs\python\python311\lib\site-packages (from selenium) (2022.1
2.7)
Requirement already satisfied: attrs>=20.1.0 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from trio~=0.17->selenium)
(22.2.0)
Requirement already satisfied: sortedcontainers in c:\users\99minds-1\appdata
\local\programs\python\python311\lib\site-packages (from trio~=0.17->seleniu
m) (2.4.0)
Requirement already satisfied: idna in c:\users\99minds-1\appdata\local\progr
ams\python\python311\lib\site-packages (from trio~=0.17->selenium) (3.4)
Requirement already satisfied: outcome in c:\users\99minds-1\appdata\local\pr
ograms\python\python311\lib\site-packages (from trio~=0.17->selenium) (1.3.0.
post0)
Requirement already satisfied: sniffio in c:\users\99minds-1\appdata\local\pr
ograms\python\python311\lib\site-packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: cffi>=1.14 in c:\users\99minds-1\appdata\local
\programs\python\python311\lib\site-packages (from trio~=0.17->selenium) (1.1
5.1)
Requirement already satisfied: wsproto>=0.14 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from trio-websocket~=0.9->se
lenium) (1.2.0)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\users\99mind
s-1\appdata\local\programs\python\python311\lib\site-packages (from urllib3[s
ocks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: pycparser in c:\users\99minds-1\appdata\local
\programs\python\python311\lib\site-packages (from cffi>=1.14->trio~=0.17->se
lenium) (2.21)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from wsproto>=0.14->trio-web
socket~=0.9->selenium) (0.14.0)


[notice] A new release of pip is available: 23.0.1 -> 24.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [71]:
```
pip install selenium pandas
```

```
Requirement already satisfied: selenium in c:\users\99minds-1\appdata\local\p
rograms\python\python311\lib\site-packages (4.14.0)Note: you may need to rest
art the kernel to use updated packages.


[notice] A new release of pip is available: 23.0.1 -> 24.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
Requirement already satisfied: pandas in c:\users\99minds-1\appdata\local\pro
grams\python\python311\lib\site-packages (1.5.3)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in c:\users\99minds-1
\appdata\local\programs\python\python311\lib\site-packages (from selenium)
(1.26.15)
Requirement already satisfied: trio~=0.17 in c:\users\99minds-1\appdata\local
\programs\python\python311\lib\site-packages (from selenium) (0.22.2)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\99minds-1\appd
ata\local\programs\python\python311\lib\site-packages (from selenium) (0.11.
1)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\99minds-1\appda
ta\local\programs\python\python311\lib\site-packages (from selenium) (2022.1
2.7)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\99minds-1\a
ppdata\local\programs\python\python311\lib\site-packages (from pandas) (2.8.
2)
Requirement already satisfied: pytz>=2020.1 in c:\users\99minds-1\appdata\loc
al\programs\python\python311\lib\site-packages (from pandas) (2022.7.1)
Requirement already satisfied: numpy>=1.21.0 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from pandas) (1.24.2)
Requirement already satisfied: six>=1.5 in c:\users\99minds-1\appdata\local\p
rograms\python\python311\lib\site-packages (from python-dateutil>=2.8.1->pand
as) (1.16.0)
Requirement already satisfied: attrs>=20.1.0 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from trio~=0.17->selenium)
(22.2.0)
Requirement already satisfied: sortedcontainers in c:\users\99minds-1\appdata
\local\programs\python\python311\lib\site-packages (from trio~=0.17->seleniu
m) (2.4.0)
Requirement already satisfied: idna in c:\users\99minds-1\appdata\local\progr
ams\python\python311\lib\site-packages (from trio~=0.17->selenium) (3.4)
Requirement already satisfied: outcome in c:\users\99minds-1\appdata\local\pr
ograms\python\python311\lib\site-packages (from trio~=0.17->selenium) (1.3.0.
post0)
Requirement already satisfied: sniffio in c:\users\99minds-1\appdata\local\pr
ograms\python\python311\lib\site-packages (from trio~=0.17->selenium) (1.3.0)
Requirement already satisfied: cffi>=1.14 in c:\users\99minds-1\appdata\local
\programs\python\python311\lib\site-packages (from trio~=0.17->selenium) (1.1
5.1)
Requirement already satisfied: wsproto>=0.14 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from trio-websocket~=0.9->se
lenium) (1.2.0)
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in c:\users\99mind
s-1\appdata\local\programs\python\python311\lib\site-packages (from urllib3[s
ocks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: pycparser in c:\users\99minds-1\appdata\local
\programs\python\python311\lib\site-packages (from cffi>=1.14->trio~=0.17->se
lenium) (2.21)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\99minds-1\appdata\lo
cal\programs\python\python311\lib\site-packages (from wsproto>=0.14->trio-web
socket~=0.9->selenium) (0.14.0)
```

In [70]:
```
import selenium
import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import warnings
warnings.filterwarnings('ignore')
from selenium.webdriver.common.by import By
import time
```

In [67]:
```
driver=webdriver.Chrome()
# opening the naukri page on automated chrome browser
driver.get("https://www.naukri.com/")

# finding element for job search bar
search_job = driver.find_element(By.CSS_SELECTOR, 'input.suggestor-input')
search_job
search_job.send_keys("Data Scientist")

search_btn= driver.find_element(By.CLASS_NAME, "qsbSubmit")
search_btn.click()

location_tag=driver.find_elements(By.XPATH,'.//*[@class="styles_ellipsis__cvWP
location_tag
location_tag.send_keys("Delhi / NCR")
salary=driver.find_elements(By.CSS_SELECTOR, "styles_chkLbl__n2x09")
salary.send_keys('3-6 Lakhs')

job_title=[]
job_location=[]
company_name=[]
experience_Required=[]

# scraping job_title from the given page
title_tag=driver.find_elements(By.XPATH,'//div[@class="cust-job-tuple layout-wr
for i in title_tag:
    title=i.text
    job_title.append(title)

# scraping job_location from the given page
location_tag=driver.find_elements(By.XPATH,'//span[@class="locWdth"]')
for i in location_tag:
    location=i.text
    job_location.append(location)


# scraping company_name from the given page
company_tag=driver.find_elements(By.XPATH,'//div[@class=" row2"]/span/a[1]')
for i in company_tag:
    company=i.text
    company_name.append(company)

# scraping company_name from the given page
experience_tag=driver.find_elements(By.XPATH,'//span[@class="ni-job-tuple-icon
for i in experience_tag:
    experience=i.text
    experience_Required.append(experience)

df=pd.DataFrame({'Title':job_title, 'Location':job_location, 'Company_Name':co
df
```

```
---------------------------------------------------------------------------
NoSuchElementException                    Traceback (most recent call last)
Cell In[67], line 13
     10 search_btn= driver.find_element(By.CLASS_NAME, "qsbSubmit")
     11 search_btn.click()
---> 13 location_tag=driver.find_element(By.XPATH,'//span[@class="styles_elli
psis__cvWP1 styles_filterLabel__jRP04"]')
     14 location_tag
     15 location_tag.send_keys("Delhi / NCR")

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\webdriver.py:738, in WebDriver.find_element(self, by, value)
    735     by = By.CSS_SELECTOR
    736     value = f'[name="{value}"]'
--> 738 return self.execute(Command.FIND_ELEMENT, {"using": by, "value": valu
e})["value"]

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\webdriver.py:344, in WebDriver.execute(self, driver_command, pa
rams)
    342 response = self.command_executor.execute(driver_command, params)
    343 if response:
--> 344     self.error_handler.check_response(response)
    345     response["value"] = self._unwrap_value(response.get("value", Non
e))
    346     return response

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\errorhandler.py:229, in ErrorHandler.check_response(self, respo
nse)
    227         alert_text = value["alert"].get("text")
    228     raise exception_class(message, screen, stacktrace, alert_text)  #
type: ignore[call-arg]  # mypy is not smart enough here
--> 229 raise exception_class(message, screen, stacktrace)

NoSuchElementException: Message: no such element: Unable to locate element:
{"method":"xpath","selector":"//span[@class="styles_ellipsis__cvWP1 styles_fi
lterLabel__jRP04"]"}
  (Session info: chrome=126.0.6478.127); For documentation on this error, ple
ase visit: https://www.selenium.dev/documentation/webdriver/troubleshooting/e
rrors#no-such-element-exception (https://www.selenium.dev/documentation/webdr
iver/troubleshooting/errors#no-such-element-exception)
Stacktrace:
        GetHandleVerifier [0x00007FF6C91DEEA2+31554]
        (No symbol) [0x00007FF6C9157ED9]
        (No symbol) [0x00007FF6C901872A]
        (No symbol) [0x00007FF6C9068434]
        (No symbol) [0x00007FF6C906853C]
        (No symbol) [0x00007FF6C90AF6A7]
        (No symbol) [0x00007FF6C908D06F]
        (No symbol) [0x00007FF6C90AC977]
        (No symbol) [0x00007FF6C908CDD3]
        (No symbol) [0x00007FF6C905A33B]
        (No symbol) [0x00007FF6C905AED1]
        GetHandleVerifier [0x00007FF6C94E8B1D+3217341]
        GetHandleVerifier [0x00007FF6C9535AE3+3532675]
        GetHandleVerifier [0x00007FF6C952B0E0+3489152]
```

```
GetHandleVerifier [0x00007FF6C928E776+750614]
(No symbol) [0x00007FF6C916375F]
(No symbol) [0x00007FF6C915EB14]
(No symbol) [0x00007FF6C915ECA2]
(No symbol) [0x00007FF6C914E16F]
BaseThreadInitThunk [0x00007FFD41107C24+20]
RtlUserThreadStart [0x00007FFD42BCD4D1+33]
```

Q.2 Write a python program to scrape data for “Data Scientist” Job position in “Bangalore” location. You have to scrape the job-title, job-location, company_name, full job-description. You have to scrape first 10 jobs data.

In [68]:
```
driver=webdriver.Chrome()
# opening the naukri page on automated chrome browser
driver.get("https://www.shine.com/")

# finding element for job search bar
search_job = driver.find_element(By.ID, "id_q")
search_job.send_keys('Data Analyst')

location_field = driver.find_element(By.ID, 'id_loc')
location_field.send_keys("Bangalore")

search_button = driver.find_element(By.CSS_SELECTOR, 'button.search-btn')
search_button.click()

job_title=[]
job_location=[]
company_name=[]
experience_Required=[]

# scraping job_title from the given page
title_tag=driver.find_elements(By.XPATH,'//strong[@class="jobCard_pReplaceH2__
for i in title_tag:
    title=i.text
    job_title.append(title)

# scraping job_location from the given page
location_tag=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_lists
for i in location_tag:
    location=i.text
    job_location.append(location)

# scraping company_name from the given page
company_tag=driver.find_elements(By.XPATH,'//div[@class="jobCard_jobCard_cName_
for i in company_tag:
    company=i.text
    company_name.append(company)

# scraping company_name from the given page
experience_tag=driver.find_elements(By.XPATH,'//div[@class=" jobCard_jobCard_l
for i in experience_tag:
    experience=i.text
    experience_Required.append(experience)

df=pd.DataFrame({'Title':job_title, 'Location':job_location, 'Company_Name':co
df
```

```
---------------------------------------------------------------------------
NoSuchElementException                    Traceback (most recent call last)
Cell In[68], line 6
      3 driver.get("https://www.shine.com/")
      5 # finding element for job search bar
----> 6 search_job = driver.find_element(By.ID, "id_q")
      7 search_job.send_keys('Data Analyst')
      9 location_field = driver.find_element(By.ID, 'id_loc')

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\webdriver.py:738, in WebDriver.find_element(self, by, value)
    735     by = By.CSS_SELECTOR
    736     value = f'[name="{value}"]'
--> 738 return self.execute(Command.FIND_ELEMENT, {"using": by, "value": valu
e})["value"]

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\webdriver.py:344, in WebDriver.execute(self, driver_command, pa
rams)
    342 response = self.command_executor.execute(driver_command, params)
    343 if response:
--> 344     self.error_handler.check_response(response)
    345     response["value"] = self._unwrap_value(response.get("value", Non
e))
    346     return response

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\errorhandler.py:229, in ErrorHandler.check_response(self, respo
nse)
    227         alert_text = value["alert"].get("text")
    228     raise exception_class(message, screen, stacktrace, alert_text)  #
type: ignore[call-arg]  # mypy is not smart enough here
--> 229 raise exception_class(message, screen, stacktrace)

NoSuchElementException: Message: no such element: Unable to locate element:
{"method":"css selector","selector":"[id="id_q"]"}
  (Session info: chrome=126.0.6478.127); For documentation on this error, ple
ase visit: https://www.selenium.dev/documentation/webdriver/troubleshooting/e
rrors#no-such-element-exception (https://www.selenium.dev/documentation/webdr
iver/troubleshooting/errors#no-such-element-exception)
Stacktrace:
        GetHandleVerifier [0x00007FF6C91DEEA2+31554]
        (No symbol) [0x00007FF6C9157ED9]
        (No symbol) [0x00007FF6C901872A]
        (No symbol) [0x00007FF6C9068434]
        (No symbol) [0x00007FF6C906853C]
        (No symbol) [0x00007FF6C90AF6A7]
        (No symbol) [0x00007FF6C908D06F]
        (No symbol) [0x00007FF6C90AC977]
        (No symbol) [0x00007FF6C908CDD3]
        (No symbol) [0x00007FF6C905A33B]
        (No symbol) [0x00007FF6C905AED1]
        GetHandleVerifier [0x00007FF6C94E8B1D+3217341]
        GetHandleVerifier [0x00007FF6C9535AE3+3532675]
        GetHandleVerifier [0x00007FF6C952B0E0+3489152]
        GetHandleVerifier [0x00007FF6C928E776+750614]
        (No symbol) [0x00007FF6C916375F]
```

```
(No symbol) [0x00007FF6C915EB14]
(No symbol) [0x00007FF6C915ECA2]
(No symbol) [0x00007FF6C914E16F]
BaseThreadInitThunk [0x00007FFD41107C24+20]
RtlUserThreadStart [0x00007FFD42BCD4D1+33]
```

Q.3 Scrape 100 reviews data from flipkart.com for iphone11 phone. You have to go the link: https://www.flipkart.com/apple-iphone-11-black-64-gb/product-reviews/itm4e5041ba101fd?pid=MOBFWQ6BXGJCEYNY&lid=LSTMOBFWQ6BXGJCEY (https://www.flipkart.com/apple-iphone-11-black-64-gb/product-reviews/itm4e5041ba101fd?pid=MOBFWQ6BXGJCEYNY&lid=LSTMOBFWQ6BXGJCEY

In [69]:
```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import pandas as pd
import time

# Initialize the webdriver
driver = webdriver.Chrome()

# Open the specific product review page on Flipkart
driver.get("https://www.flipkart.com/apple-iphone-11-black-64-gb/product-reviev

# Wait for the page to load
time.sleep(5)

# Scrape the data for 100 reviews
ratings = []
review_summaries = []
full_reviews = []

while len(ratings) < 100:
    # Get all the reviews on the current page
    review_elements = driver.find_elements(By.CSS_SELECTOR, 'div._1AtVbE')

    for element in review_elements:
        if len(ratings) >= 100:
            break
        try:
            rating = element.find_element(By.CSS_SELECTOR, 'div._3LWZlK._1BLPM
            review_summary = element.find_element(By.CSS_SELECTOR, 'p._2-N8zT'
            full_review = element.find_element(By.CSS_SELECTOR, 'div.t-ZTKy di

            ratings.append(rating)
            review_summaries.append(review_summary)
            full_reviews.append(full_review)
        except Exception as e:
            print(f"Error occurred while scraping review: {e}")

    # Click the "Next" button to go to the next page of reviews
    try:
        next_button = driver.find_element(By.CSS_SELECTOR, 'a._1LKTO3')
        next_button.click()
        time.sleep(5)
    except Exception as e:
        print(f"Error occurred while clicking next button: {e}")
        break

# Create a DataFrame
df = pd.DataFrame({
    'Rating': ratings,
    'Review Summary': review_summaries,
    'Full Review': full_reviews
})

# Print the DataFrame
print(df)
```

```
# Save the DataFrame to a CSV file
df.to_csv('flipkart_iphone11_reviews.csv', index=False)

# Close the webdriver
driver.quit()
```

```
Error occurred while clicking next button: Message: no such element: Unable t
o locate element: {"method":"css selector","selector":"a._1LKTO3"}
  (Session info: chrome=126.0.6478.127); For documentation on this error, ple
ase visit: https://www.selenium.dev/documentation/webdriver/troubleshooting/e
rrors#no-such-element-exception (https://www.selenium.dev/documentation/webdr
iver/troubleshooting/errors#no-such-element-exception)
Stacktrace:
        GetHandleVerifier [0x00007FF6C91DEEA2+31554]
        (No symbol) [0x00007FF6C9157ED9]
        (No symbol) [0x00007FF6C901872A]
        (No symbol) [0x00007FF6C9068434]
        (No symbol) [0x00007FF6C906853C]
        (No symbol) [0x00007FF6C90AF6A7]
        (No symbol) [0x00007FF6C908D06F]
        (No symbol) [0x00007FF6C90AC977]
        (No symbol) [0x00007FF6C908CDD3]
        (No symbol) [0x00007FF6C905A33B]
        (No symbol) [0x00007FF6C905AED1]
        GetHandleVerifier [0x00007FF6C94E8B1D+3217341]
        GetHandleVerifier [0x00007FF6C9535AE3+3532675]
        GetHandleVerifier [0x00007FF6C952B0E0+3489152]
        GetHandleVerifier [0x00007FF6C928E776+750614]
        (No symbol) [0x00007FF6C916375F]
        (No symbol) [0x00007FF6C915EB14]
        (No symbol) [0x00007FF6C915ECA2]
        (No symbol) [0x00007FF6C914E16F]
        BaseThreadInitThunk [0x00007FFD41107C24+20]
        RtlUserThreadStart [0x00007FFD42BCD4D1+33]

Empty DataFrame
Columns: [Rating, Review Summary, Full Review]
Index: []
```

Q.4 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 3 attributes of each sneaker:

1. Brand
2. Product Description
3. Price As shown in the below image, you have to scrape the above attributes.

In [73]:
```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import pandas as pd
import time



# Initialize the webdriver
driver = webdriver.Chrome()

# Open Flipkart
driver.get("https://www.flipkart.com")

# Close the login popup if it appears
try:
    close_login_popup = WebDriverWait(driver, 10).until(
        EC.element_to_be_clickable((By.CSS_SELECTOR, 'button._2KpZ6l._2doB4z')
    )
    close_login_popup.click()
except:
    pass

# Enter “sneakers” in the search field
search_field = WebDriverWait(driver, 10).until(
    EC.presence_of_element_located((By.NAME, 'q'))
)
search_field.send_keys("sneakers")
search_field.send_keys(Keys.RETURN)

# Wait for the results to load
time.sleep(5)

# Scrape the data for the first 100 sneakers
brands = []
product_descriptions = []
prices = []

while len(brands) < 100:
    # Get all the sneaker elements on the current page
    sneaker_elements = driver.find_elements(By.CSS_SELECTOR, 'div._1AtVbE')

    for element in sneaker_elements:
        if len(brands) >= 100:
            break
        try:
            brand = element.find_element(By.CSS_SELECTOR, 'div._2WkVRV').text
            product_description = element.find_element(By.CSS_SELECTOR, 'a.IRpu
            price = element.find_element(By.CSS_SELECTOR, 'div._30jeq3').text

            brands.append(brand)
            product_descriptions.append(product_description)
            prices.append(price)
        except Exception as e:
            print(f"Error occurred while scraping sneaker: {e}")
```

```
    # Click the "Next" button to go to the next page of results
    try:
        next_button = driver.find_element(By.CSS_SELECTOR, 'a._1LKTO3')
        next_button.click()
        time.sleep(5)
    except Exception as e:
        print(f"Error occurred while clicking next button: {e}")
        break

# Create a DataFrame
df = pd.DataFrame({
    'Brand': brands,
    'Product Description': product_descriptions,
    'Price': prices
})

# Print the DataFrame
print(df)

# Save the DataFrame to a CSV file
df.to_csv('flipkart_sneakers.csv', index=False)

# Close the webdriver
driver.quit()
```

```
Error occurred while clicking next button: Message: no such element: Unable t
o locate element: {"method":"css selector","selector":"a._1LKTO3"}
  (Session info: chrome=126.0.6478.127); For documentation on this error, ple
ase visit: https://www.selenium.dev/documentation/webdriver/troubleshooting/e
rrors#no-such-element-exception (https://www.selenium.dev/documentation/webdr
iver/troubleshooting/errors#no-such-element-exception)
Stacktrace:
        GetHandleVerifier [0x00007FF6C91DEEA2+31554]
        (No symbol) [0x00007FF6C9157ED9]
        (No symbol) [0x00007FF6C901872A]
        (No symbol) [0x00007FF6C9068434]
        (No symbol) [0x00007FF6C906853C]
        (No symbol) [0x00007FF6C90AF6A7]
        (No symbol) [0x00007FF6C908D06F]
        (No symbol) [0x00007FF6C90AC977]
        (No symbol) [0x00007FF6C908CDD3]
        (No symbol) [0x00007FF6C905A33B]
        (No symbol) [0x00007FF6C905AED1]
        GetHandleVerifier [0x00007FF6C94E8B1D+3217341]
        GetHandleVerifier [0x00007FF6C9535AE3+3532675]
        GetHandleVerifier [0x00007FF6C952B0E0+3489152]
        GetHandleVerifier [0x00007FF6C928E776+750614]
        (No symbol) [0x00007FF6C916375F]
        (No symbol) [0x00007FF6C915EB14]
        (No symbol) [0x00007FF6C915ECA2]
        (No symbol) [0x00007FF6C914E16F]
        BaseThreadInitThunk [0x00007FFD41107C24+20]
        RtlUserThreadStart [0x00007FFD42BCD4D1+33]

Empty DataFrame
Columns: [Brand, Product Description, Price]
Index: []
```

Q.5 Go to webpage https://www.amazon.in/ (https://www.amazon.in/) Enter “Laptop” in the search field and then click the search icon. Then set CPU Type filter to “Intel Core i7” as shown in the below image:

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

1. Title
2. Ratings
3. Price

In [93]:
```
# Initialize the webdriver
driver = webdriver.Chrome()

# Open Flipkart
driver.get("https://www.amazon.in")

# finding element for job search bar
search_g= driver.find_element(By.XPATH, "//input[@type='text']")
search_g
# <selenium.webdriver.remote.webelement.WebElement (session="a2ad6e875a1461b01
# write on search bar
search_g.send_keys('Laptop')
search_btn=driver.find_element(By.XPATH, "//input[@id='nav-search-submit-butto
search_btn
# <selenium.webdriver.remote.webelement.WebElement (session="a2ad6e875a1461b01
search_btn=driver.find_element(By.XPATH, "//input[@id='nav-search-submit-butto
search_btn.click()
# No core filter are shown on the website

Title=[]
Price=[]
Rating=[]
for i in range(3):
    b_name=driver.find_elements(By.XPATH, "//div[@class='_2WkVRV']")
    p_desc=driver.find_elements(By.XPATH, "//a[@class='IRpwTa']")
    price =driver.find_elements(By.XPATH, "//div[@class='_25b18c']")


    for j  in b_name:
        Title.append(j.text)
    Title[:100]



    for k in p_desc:
        Rating.append(k.text)
    Rating[:100]


    for l in price:
        Price.append(l.text)
    Price[:100]
```

Q.6 Write a python program to scrape data for Top 1000 Quotes of All Time.

The above task will be done in following steps:

1. First get the webpagehttps://www.azquotes.com/ (https://www.azquotes.com/)
2. Click on Top Quote
3. Than scrap a) Quote b) Author c) Type Of Quotes

In [94]:
```
# Initialize the webdriver
driver = webdriver.Chrome()

# Open AZ Quotes
driver.get("https://www.azquotes.com/")

# Click on Top Quotes
top_quotes_link = WebDriverWait(driver, 10).until(
    EC.element_to_be_clickable((By.LINK_TEXT, 'Top Quotes'))
)
top_quotes_link.click()

# Wait for the page to load
time.sleep(5)

# Scrape the data for the top 1000 quotes
quotes = []
authors = []
types = []

while len(quotes) < 1000:
    # Get all the quote elements on the current page
    quote_elements = driver.find_elements(By.CSS_SELECTOR, 'div.quote')

    for element in quote_elements:
        if len(quotes) >= 1000:
            break
        try:
            quote = element.find_element(By.CSS_SELECTOR, 'a.title').text
            author = element.find_element(By.CSS_SELECTOR, 'a.author').text
            type_of_quote = element.find_element(By.CSS_SELECTOR, 'a.tag').tex

            quotes.append(quote)
            authors.append(author)
            types.append(type_of_quote)
        except Exception as e:
            print(f"Error occurred while scraping quote: {e}")

    # Click the "Next" button to go to the next page of results
    try:
        next_button = driver.find_element(By.CSS_SELECTOR, 'li.next a')
        next_button.click()
        time.sleep(5)
    except Exception as e:
        print(f"Error occurred while clicking next button: {e}")
        break

# Create a DataFrame
df = pd.DataFrame({
    'Quote': quotes,
    'Author': authors,
    'Type of Quote': types
})

# Print the DataFrame
print(df)
```

```
# Save the DataFrame to a CSV file
df.to_csv('top_1000_quotes.csv', index=False)

# Close the webdriver
driver.quit()
```

```
---------------------------------------------------------------------------
TimeoutException                              Traceback (most recent call last)
Cell In[94], line 8
      5 driver.get("https://www.azquotes.com/")
      7 # Click on Top Quotes
----> 8 top_quotes_link = WebDriverWait(driver, 10).until(
      9     EC.element_to_be_clickable((By.LINK_TEXT, 'Top Quotes'))
     10 )
     11 top_quotes_link.click()
     13 # Wait for the page to load

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\webdriver\support\wait.py:95, in WebDriverWait.until(self, method, message)
     93     if time.monotonic() > end_time:
     94         break
---> 95 raise TimeoutException(message, screen, stacktrace)

TimeoutException: Message:
Stacktrace:
        GetHandleVerifier [0x00007FF6C91DEEA2+31554]
        (No symbol) [0x00007FF6C9157ED9]
        (No symbol) [0x00007FF6C901872A]
        (No symbol) [0x00007FF6C9068434]
        (No symbol) [0x00007FF6C906853C]
        (No symbol) [0x00007FF6C90AF6A7]
        (No symbol) [0x00007FF6C908D06F]
        (No symbol) [0x00007FF6C90AC977]
        (No symbol) [0x00007FF6C908CDD3]
        (No symbol) [0x00007FF6C905A33B]
        (No symbol) [0x00007FF6C905AED1]
        GetHandleVerifier [0x00007FF6C94E8B1D+3217341]
        GetHandleVerifier [0x00007FF6C9535AE3+3532675]
        GetHandleVerifier [0x00007FF6C952B0E0+3489152]
        GetHandleVerifier [0x00007FF6C928E776+750614]
        (No symbol) [0x00007FF6C916375F]
        (No symbol) [0x00007FF6C915EB14]
        (No symbol) [0x00007FF6C915ECA2]
        (No symbol) [0x00007FF6C914E16F]
        BaseThreadInitThunk [0x00007FFD41107C24+20]
        RtlUserThreadStart [0x00007FFD42BCD4D1+33]
```

Q.7 Write a python program to display list of respected former Prime Ministers of India (i.e. Name, Born-Dead, Term of office, Remarks) from https://www.jagranjosh.com/general-knowledge/list-of-all-prime-

ministers-of-india-1473165149-1 (https://www.jagranjosh.com/general-knowledge/list-of-all-prime-

In [96]:

```
# Initialize the webdriver
driver = webdriver.Chrome()

# Open the webpage
driver.get("https://www.jagranjosh.com/general-knowledge/list-of-all-prime-min

# Wait for the page to load
WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.CSS_SELECTO

# Scrape the data
names = []
born_dead = []
terms_of_office = []
remarks = []

# Find the table containing the list of prime ministers
table = driver.find_element(By.CSS_SELECTOR, 'table')

# Find all rows in the table
rows = table.find_elements(By.TAG_NAME, 'tr')

# Iterate over the rows to extract data
for row in rows[1:]:  # Skip the header row
    columns = row.find_elements(By.TAG_NAME, 'td')
    if len(columns) == 4:
        names.append(columns[0].text)
        born_dead.append(columns[1].text)
        terms_of_office.append(columns[2].text)
        remarks.append(columns[3].text)

# Create a DataFrame
df = pd.DataFrame({
    'Name': names,
    'Born-Dead': born_dead,
    'Term of Office': terms_of_office,
    'Remarks': remarks
})

# Print the DataFrame
print(df)

# Save the DataFrame to a CSV file
df.to_csv('former_prime_ministers_of_india.csv', index=False)

# Close the webdriver
driver.quit()
```

```
Empty DataFrame
Columns: [Name, Born-Dead, Term of Office, Remarks]
Index: []
```

Q8: Write a python program to display list of 50 Most expensive cars in the world (i.e. Car name and Price) from https://www.motor1.com/ (https://www.motor1.com/)

This task will be done in following steps:

1. First get the webpage https://www.motor1.com/ (https://www.motor1.com/)
2. Then You have to type in the search bar ’50 most expensive cars’
3. Then click on 50 most expensive cars in the world..
4. Then scrap the mentioned data and make the dataframe.

In [101]:
```
from selenium import webdriver
import pandas as pd

# Step 1: Get the webpage
driver = webdriver.Chrome()
driver.get('https://www.motor1.com/')

# Step 2: Type in the search bar
search_bar = driver.find_element(By.ID, 'search-input')
search_bar.send_keys('50 most expensive cars')
search_bar.submit()

# Step 3: Click on the link
link = driver.find_element_by_link_text('50 Most Expensive Cars in the World')
link.click()

# Step 4: Scrape the data and create a dataframe
car_names = driver.find_elements(By.XPATH, '//div[@class="article-content"]/h3
car_prices = driver.find_elements(By.XPATH,'//div[@class="article-content"]/p'

data = []
for name, price in zip(car_names, car_prices):
  data.append([name.text, price.text])

df = pd.DataFrame(data, columns=['Car Name', 'Price'])
print(df)

driver.quit()
```

```
---------------------------------------------------------------------------
NoSuchWindowException                     Traceback (most recent call last)
Cell In[101], line 9
      6 driver.get('https://www.motor1.com/')
      8 # Step 2: Type in the search bar
----> 9 search_bar = driver.find_element(By.ID, 'search-input')
     10 search_bar.send_keys('50 most expensive cars')
     11 search_bar.submit()

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\webdriver.py:738, in WebDriver.find_element(self, by, value)
    735     by = By.CSS_SELECTOR
    736     value = f'[name="{value}"]'
--> 738 return self.execute(Command.FIND_ELEMENT, {"using": by, "value": valu
e})["value"]

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\webdriver.py:344, in WebDriver.execute(self, driver_command, pa
rams)
    342 response = self.command_executor.execute(driver_command, params)
    343 if response:
--> 344     self.error_handler.check_response(response)
    345     response["value"] = self._unwrap_value(response.get("value", Non
e))
    346     return response

File ~\AppData\Local\Programs\Python\Python311\Lib\site-packages\selenium\web
driver\remote\errorhandler.py:229, in ErrorHandler.check_response(self, respo
nse)
    227         alert_text = value["alert"].get("text")
    228     raise exception_class(message, screen, stacktrace, alert_text)  #
type: ignore[call-arg]  # mypy is not smart enough here
--> 229 raise exception_class(message, screen, stacktrace)

NoSuchWindowException: Message: no such window: target window already closed
from unknown error: web view not found
  (Session info: chrome=126.0.6478.127)
Stacktrace:
        GetHandleVerifier [0x00007FF6C91DEEA2+31554]
        (No symbol) [0x00007FF6C9157ED9]
        (No symbol) [0x00007FF6C901872A]
        (No symbol) [0x00007FF6C8FED995]
        (No symbol) [0x00007FF6C90944D7]
        (No symbol) [0x00007FF6C90AC051]
        (No symbol) [0x00007FF6C908CDD3]
        (No symbol) [0x00007FF6C905A33B]
        (No symbol) [0x00007FF6C905AED1]
        GetHandleVerifier [0x00007FF6C94E8B1D+3217341]
        GetHandleVerifier [0x00007FF6C9535AE3+3532675]
        GetHandleVerifier [0x00007FF6C952B0E0+3489152]
        GetHandleVerifier [0x00007FF6C928E776+750614]
        (No symbol) [0x00007FF6C916375F]
        (No symbol) [0x00007FF6C915EB14]
        (No symbol) [0x00007FF6C915ECA2]
        (No symbol) [0x00007FF6C914E16F]
        BaseThreadInitThunk [0x00007FFD41107C24+20]
        RtlUserThreadStart [0x00007FFD42BCD4D1+33]
```

In [ ]: