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/

2. Enter "Data Analyst" in "Skill, Designations, Companies" field and enter "Bangalore" in "enter the location" field.

3. Then click the searchbutton.

4. Then scrape the data for the first 10 jobs results youget.

5. Finally create a dataframe of the scraped data

Answer:

```
import selenium

from selenium import webdriver

import pandas as pd

from selenium.webdriver.common.by import By

import warnings

warnings.filterwarnings("ignore")

import time

driver = webdriver.Chrome()

driver.get("https://www.naukri.com/")

designation = driver.find_element(By.CLASS_NAME,"suggestor-input ").send_keys("Data Analyst")

location =
driver.find_element(By.XPATH,"/html/body/div[1]/div[6]/div/div/div[5]/div/div/div/div[1]/div/input")
.send_keys("Bangalore")

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

job_title=[]

job_location=[]

company_name=[]

experience_required=[]
```

```
#craping job title
title_tag = driver.find_elements(By.XPATH,'//a[@class="title ellipsis"]')
for i in title_tag [0:10]:
    title=i.text
    job_title.append(title)

#craping job location

job_location_tag = driver.find_elements(By.XPATH,'//span[@class="ellipsis fleft locWdth"]')
for i in job_location_tag [0:10]:
    location = i.text
    job_location.append(location)

# craping Company Name
company_tag = driver.find_elements (By.XPATH,'//a[@class="subTitle ellipsis fleft"]')
for i in company_tag [0:10]:
    company = i.text
    company_name.append(company)

#Craping Experience
experience_tag = driver.find_elements(By.XPATH,'//span[@class="ellipsis fleft expwdth"]')
for i in experience_tag [0:10]:
    experience = i.text
    experience_required.append(experience)
print (len(job_title), len (job_location), len (company_name), len (experience_required))
df = pd.DataFrame({'Title':job_title,'Location' :job_location,'Company' : company_name,'Experience':
experience_required })
df
```

Q2:Write a python program to scrape data for “Data Scientist” Job position in “Bangalore” location. You have to scrape the job-title, job-location, company_name. You have to scrape first 10 jobs data. This task will be done in following steps:

1. First get the webpage https://www.naukri.com/

2. Enter “Data Scientist” in “Skill, Designations, Companies” field and enter “Bangalore” in “enter the location” field.

3. Then click the searchbutton.

4. Then scrape the data for the first 10 jobs results youget.

5. Finally create a dataframe of the scraped data.

Answer:

```
import selenium

from selenium import webdriver

import pandas as pd

from selenium.webdriver.common.by import By

import warnings

warnings.filterwarnings("ignore")

import time

driver = webdriver.Chrome()

driver.get("https://www.naukri.com/")

designation = driver.find_element(By.CLASS_NAME,"suggestor-input ").send_keys("Data Analyst")

location =
driver.find_element(By.XPATH,"/html/body/div[1]/div[6]/div/div/div[5]/div/div/div/div[1]/div/input")
.send_keys("Bangalore")

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

job_title=[]

job_location=[]

company_name=[]

#craping job title

title_tag = driver.find_elements(By.XPATH,'//a[@class="title ellipsis"]')

for i in title_tag [0:10]:

    title=i.text

    job_title.append(title)
```

```
#craping job location


job_location_tag = driver.find_elements(By.XPATH,'//span[@class="ellipsis fleft locWdth"]')
for i in job_location_tag [0:10]:
    location = i.text
    job_location.append(location)


# craping Company Name
company_tag = driver.find_elements (By.XPATH,'//a[@class="subTitle ellipsis fleft"]')
for i in company_tag [0:10]:
    company = i.text
    company_name.append(company)


print (len(job_title), len (job_location), len (company_name))
df = pd.DataFrame({'Title':job_title,'Location' :job_location,'Company' : company_name })
df
```

Q3: In this question you have to scrape data using the filters available on the webpage as shown below: ASSIGNMENT 2 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 thewebpage https://www.naukri.com/

2. Enter "Data Scientist" in "Skill, Designations, and Companies" field.

3. Then click the searchbutton.

4. Then apply the location filter and salary filter by checking the respectiveboxes

5. Then scrape the data for the first 10 jobs results youget.

6. Finally create a dataframe of the scrapeddata.

Answer:

```
import selenium
from selenium import webdriver
import pandas as pd
from selenium.webdriver.common.by import By
import warnings
warnings.filterwarnings("ignore")
import time
driver = webdriver.Chrome()
driver.get("https://www.naukri.com/")
designation = driver.find_element(By.CLASS_NAME,"suggestor-input ").send_keys("Data Analyst")
button = driver.find_element(By.CLASS_NAME,"qsbSubmit").click()
Salary = driver.find_element (By.XPATH,"//*[contains(@title,'3-6 Lakhs')]").click()
location = driver.find_element (By.XPATH,"//*[contains(@title,'Delhi / NCR')]").click()
job_title=[]
job_location=[]
company_name=[]
experience_required=[]
#craping job title
title_tag = driver.find_elements(By.XPATH,'//a[@class="title ellipsis"]')
for i in title_tag [0:10]:
    title=i.text
    job_title.append(title)

#craping job location

job_location_tag = driver.find_elements(By.XPATH,'//span[@class="ellipsis fleft locWdth"]')
for i in job_location_tag [0:10]:
    location = i.text
    job_location.append(location)

# craping Company Name
```

```
company_tag = driver.find_elements (By.XPATH,'//a[@class="subTitle ellipsis fleft"]')
for i in company_tag [0:10]:
    company = i.text
    company_name.append(company)

#Craping Experience
experience_tag = driver.find_elements(By.XPATH,'//span[@class="ellipsis fleft expwdth"]')
for i in experience_tag [0:10]:
    experience = i.text
    experience_required.append(experience)

print (len(job_title), len (job_location), len (company_name), len (experience_required))
df = pd.DataFrame({'Title':job_title,'Location' :job_location,'Company' : company_name
,'Experience': experience_required})
df
```

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

2. ProductDescription

3. Price

4. Discount percentage

Answer:

```
import selenium
from selenium import webdriver
import pandas as pd
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import warnings
warnings.filterwarnings("ignore")
import time
```

```
driver = webdriver.Chrome()

driver.get("http://www.flipkart.com/")

driver.find_element(By.XPATH,'//button[@class="_2KpZ6l _2doB4z"]').click()

search_box = driver.find_element(By.XPATH,"//*[contains(@title,'Search for Products, Brands and More')and @type='text']").send_keys("sunglasses")

search_box = driver.find_element(By.XPATH,"//*[contains(@title,'Search for Products, Brands and More')and @type='text']").send_keys(Keys.ENTER)

brand_name=[]

product_description=[]

price=[]

discount_percentage=[]

for page in range(0, 3):
    # Scraping brand names
    brand_tag = driver.find_elements(By.XPATH, '//div[@class="_2WkVRV"]')
    for i in brand_tag[0:100]:
        brand = i.text
        brand_name.append(brand)
        #time.sleep(1)


    # Scraping product descriptions
    product_description_tag = driver.find_elements(By.XPATH, '//a[@class="IRpwTa"]')
    for i in product_description_tag[0:100]:
        description = i.text
        product_description.append(description)
        #time.sleep(1)


    # Scraping prices
    price_tag = driver.find_elements(By.XPATH, '//div[@class="_30jeq3"]')
    for i in price_tag[0:100]:
        price1 = i.text
        price.append(price1)
        #time.sleep(3)
```

```
    #scraping for discount percentage
    discount_tag = driver.find_elements(By.XPATH, '//div[@class="_3Ay6Sb"]')
    for i in discount_tag[0:100]:
        discount = i.text
        discount_percentage.append(discount)
        #time.sleep(3)
    # Move to the next page
    next_button = driver.find_element(By.XPATH, "//*[text()='Next']").click()

    time.sleep(3)
```

```
brand_name = brand_name[:100]
product_description = product_description[:100]
price = price[:100]
discount_percentage=discount_percentage[:100]
print (len(brand_name), len (product_description), len (price),len(discount_percentage))
df = pd.DataFrame({'Brand Name':brand_name,'Product Description' :product_description,'Price' : price, 'Discount Percentage' :discount_percentage })
df
```

Q5: Scrape 100 reviews data from flipkart.com for iphone11 phone. You have to go the link: https://www.flipkart.com/apple-iphone-11-black-64-gb/productreviews/itm4e5041ba101fd?pid=MOBFWQ6BXGJCEYNY&lid=LSTMOBFWQ6BXGJCEYNYZXSHRJ&market place=FLIPKART

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

1. Rating

2. Review summary

3. Full review

Answer:

```
import selenium
from selenium import webdriver
import pandas as pd
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import warnings
warnings.filterwarnings("ignore")
import time
driver = webdriver.Chrome()
driver.get("https://www.flipkart.com/apple-iphone-11-black-64-gb/product-reviews/itm4e5041ba101fd?pid=MOBFWQ6BXGJCEYNY&lid=LSTMOBFWQ6BXGJCEYNYZXSHRJ&market")
rating=[]
review_summary=[]
full_review=[]
for page in range(0, 10):
    # Scraping rating
    ratinng_tag = driver.find_elements(By.XPATH, '//div[@class="_3LWZlK _1BLPMq"]')
    for i in ratinng_tag[0:100]:
        rating1 = i.text
        rating.append(rating1)
        #time.sleep(1)

    # Scraping review summary
    review_summary_tag = driver.find_elements(By.XPATH, '//p[@class="_2-N8zT"]')
    for i in review_summary_tag[0:100]:
        review_summary1 = i.text
        review_summary.append(review_summary1)
```

```
        #time.sleep(1)

    # Scraping full review
    full_review_tag = driver.find_elements(By.XPATH, '//div[@class="t-ZTKy"]')
    for i in full_review_tag[0:100]:
        full_review1 = i.text
        full_review.append(full_review1)
        #time.sleep(3)



    # Move to the next page
    next_button = driver.find_element(By.XPATH, "//*[text()='Next']").click()

    time.sleep(3)




rating = rating[:100]
review_summary = review_summary[:100]
full_review = full_review[:100]
print (len(rating), len (review_summary), len (full_review))
df = pd.DataFrame({'Rating':rating,'Review Summary' :review_summary,'Full Review' : full_review})
df
```

Q6: Scrape data forfirst 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. ProductDescription

3. Price

Answer:

```
import selenium
from selenium import webdriver
import pandas as pd
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import warnings
warnings.filterwarnings("ignore")
import time
driver = webdriver.Chrome()
driver.get("http://www.flipkart.com/")
driver.find_element(By.XPATH,'//button[@class="_2KpZ6l _2doB4z"]').click()
search_box = driver.find_element(By.XPATH,"//*[contains(@title,'Search for Products, Brands and More')and @type='text']").send_keys("sneakers")
search_box = driver.find_element(By.XPATH,"//*[contains(@title,'Search for Products, Brands and More')and @type='text']").send_keys(Keys.ENTER)
brand_name=[]
product_description=[]
price=[]

for page in range(0, 3):
    # Scraping brand names
    brand_tag = driver.find_elements(By.XPATH, '//div[@class="_2WkVRV"]')
    for i in brand_tag[0:100]:
        brand = i.text
        brand_name.append(brand)
        #time.sleep(1)

    # Scraping product descriptions
    product_description_tag = driver.find_elements(By.XPATH, '//a[@class="IRpwTa"]')
    for i in product_description_tag[0:100]:
        description = i.text
```

```
        product_description.append(description)
        #time.sleep(1)

    # Scraping prices
    price_tag = driver.find_elements(By.XPATH, '//div[@class="_30jeq3"]')
    for i in price_tag[0:100]:
        price1 = i.text
        price.append(price1)
        #time.sleep(3)



    # Move to the next page
    next_button = driver.find_element(By.XPATH, "//*[text()='Next']").click()

    time.sleep(3)




brand_name = brand_name[:100]
product_description = product_description[:100]
price = price[:100]

print (len(brand_name), len (product_description), len (price))
df = pd.DataFrame({'Brand Name':brand_name,'Product Description' :product_description,'Price' : price })
df
```

Q7: Go to webpage https://www.amazon.in/ Enter “Laptop” in the search field and then click the search icon. Then set CPU Type filter to “Intel Core i7”

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

2. Ratings

3. Price

Answer:

```
import selenium
from selenium import webdriver
import pandas as pd
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import warnings
warnings.filterwarnings("ignore")
import time
driver = webdriver.Chrome()
driver.get("https://www.amazon.in/")
driver.find_element(By.XPATH,'//input[@id="twotabsearchtextbox"]').send_keys("Laptop")
driver.find_element(By.XPATH,'//input[@id="twotabsearchtextbox"]').send_keys(Keys.ENTER)
driver.find_element(By.XPATH, '//span[@class="a-size-base a-color-base" and contains(text(), "Intel Core i7")]').click()
title=[]
rating=[]
price=[]

for page in range(0, 5):
    # Scraping brand names
    title_tag = driver.find_elements(By.XPATH, '//span[@class="a-size-medium a-color-base a-text-normal"]')
    for i in title_tag[0:100]:
        title1 = i.text
        title.append(title1)
        #time.sleep(1)

    # Scraping product descriptions
    rating_tag = driver.find_elements(By.XPATH, '//span[@class="a-icon-alt"]')
```

```
    for i in rating_tag[0:100]:
        rating1 = i.text
        rating.append(rating1)
        #time.sleep(1)

    # Scraping prices
    price_tag = driver.find_elements(By.XPATH, '//span[@class="a-price-whole"]')
    for i in price_tag[0:100]:
        price1 = i.text
        price.append(price1)
        #time.sleep(3)



    # Move to the next page
    next_button = driver.find_element(By.XPATH, "//*[text()='Next']").click()

    time.sleep(3)




title = title[:100]
rating = rating[:100]
price = price[:100]
print (len(title), len (rating), len (price))
df = pd.DataFrame({'Title':title,'Rating' :rating,'Price' : price })
df
```