WEB SCRAPING – ASSIGNMENT 2

1.

!pip install selenium

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(r"C:\Users\dell\OneDrive\Desktop\drivers\chromedriver-win64\chromedriver.exe")

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

search\_box = driver.find\_element\_by\_id("rootKeySkills")

search\_box.send\_keys("Data Analyst")

location\_box = driver.find\_element\_by\_id("rootLocation")

location\_box.send\_keys("Bangalore")

search\_button = driver.find\_element\_by\_class\_name("search-btn")

search\_button.click()

job\_data = []

jobs = driver.find\_elements\_by\_xpath('//article[starts-with(@itemtype, "http://schema.org/JobPosting")]')[:10]

for job in jobs:

title = job.find\_element\_by\_class\_name("title").text

location = job.find\_element\_by\_class\_name("location").text

company = job.find\_element\_by\_class\_name("subTitle").text

experience = job.find\_element\_by\_class\_name("experience").text

job\_data.append({

"Job Title": title,

"Location": location,

"Company": company,

"Experience Required": experience

})

df = pd.DataFrame(job\_data)

print(df)

driver.quit()

2.

!pip install selenium

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(r"C:\Users\dell\OneDrive\Desktop\drivers\chromedriver-win64\chromedriver.exe")

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

search\_box = driver.find\_element\_by\_id("rootKeySkills")

search\_box.send\_keys("Data Scientist")

location\_box = driver.find\_element\_by\_id("rootLocation")

location\_box.send\_keys("Bangalore")

search\_button = driver.find\_element\_by\_class\_name("search-btn")

search\_button.click()

job\_data = []

jobs = driver.find\_elements\_by\_xpath('//article[starts-with(@itemtype, "http://schema.org/JobPosting")]')[:10]

for job in jobs:

title = job.find\_element\_by\_class\_name("title").text

location = job.find\_element\_by\_class\_name("location").text

company = job.find\_element\_by\_class\_name("subTitle").text

job\_data.append({

"Job Title": title,

"Location": location,

"Company": company

})

df = pd.DataFrame(job\_data)

print(df)

driver.quit()

3.

!pip install selenium

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(r"C:\Users\dell\OneDrive\Desktop\drivers\chromedriver-win64\chromedriver.exe")

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

url = 'https://www.naukri.com/'

search\_field = driver.find\_element(By.ID, 'rootSearchVal')

search\_field.send\_keys('Data Scientist')

search\_button = driver.find\_element(By.CLASS\_NAME, 'btn')

search\_button.click()

driver.find\_element(By.ID, 'root\_0\_1').click()

driver.find\_element(By.ID, 'root\_0\_3').click()

job\_results = driver.find\_elements(By.CSS\_SELECTOR, "article[itemtype='http://schema.org/JobPosting']")

data = []

for i in range(min(len(job\_results), 10)):

job\_title = job\_results[i].find\_element(By.CSS\_SELECTOR, '.title')

job\_location = job\_results[i].find\_element(By.CSS\_SELECTOR, 'li.location')

company\_name = job\_results[i].find\_element(By.CSS\_SELECTOR, 'li.company')

experience\_required = job\_results[i].find\_element(By.CSS\_SELECTOR, 'li.experience')

data.append({

'Job Title': job\_title.text.strip(),

'Location': job\_location.text.strip(),

'Company Name': company\_name.text.strip(),

'Experience Required': experience\_required.text.strip()

})

df = pd.DataFrame(data)

print(df)

driver.quit()

4.

!pip install selenium

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(r"C:\Users\dell\OneDrive\Desktop\drivers\chromedriver-win64\chromedriver.exe")

url = 'https://www.flipkart.com/'

driver.get(url)

search\_field = driver.find\_element(By.CSS\_SELECTOR, 'input[type=text]')

search\_field.send\_keys('sunglasses')

search\_button = driver.find\_element(By.CSS\_SELECTOR, 'button[type=submit]')

search\_button.click()

data = []

while len(data) < 100:

sunglasses = driver.find\_elements(By.CSS\_SELECTOR, '.IIdQZO.\_1SSAGr')

for sunglass in sunglasses:

brand = sunglass.find\_element(By.CSS\_SELECTOR, 'div.\_2WkVRV').text

description = sunglass.find\_element(By.CSS\_SELECTOR, 'aIRWtM').text

price = sunglass.find\_element(By.CSS\_SELECTOR, '.\_30jeq3.\_1\_WHN1').text

data.append({

'Brand': brand,

'Product Description': description,

'Price': price

})

if len(data) >= 100:

break

next\_button = driver.find\_element(By.CSS\_SELECTOR, '.\_1LKTO3')

next\_button.click()

df = pd.DataFrame(data)

print(df)

driver.quit()

5.

!pip install selenium

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(r"C:\Users\dell\OneDrive\Desktop\drivers\chromedriver-win64\chromedriver.exe")

url = 'https://www.flipkart.com/apple-iphone-11-black-64-gb/product-reviews/itm4e5041ba101fd?pid=MOBFWQ6BXGJCEYNY&lid=LSTMOBFWQ6BXGJCEYNYZXSHRJ&marketplace=FLIPKART'

driver.get(url)

scroll\_pause\_time = 1

scroll\_height = 5

while scroll\_height >= 0:

driver.execute\_script("window.scrollTo(0, document.body.scrollHeight/{})".format(scroll\_height))

time.sleep(scroll\_pause\_time)

scroll\_height -= 1

reviews = driver.find\_elements(By.CSS\_SELECTOR, '.col \_2wzgFH K0kLPL')

data = []

count = 0

for review in reviews:

rating = review.find\_element(By.CSS\_SELECTOR, 'div.E\_Uluw.\_1LZW\_O').text

summary = review.find\_element(By.CSS\_SELECTOR, 'p.\_2-N8zT').text

full\_review = review.find\_element(By.CSS\_SELECTOR, 'div.qwjRop span:nth-child(1)').text

data.append({

'Rating': rating,

'Review Summary': summary,

'Full Review': full\_review

})

count += 1

if count >= 100:

break

df = pd.DataFrame(data)

print(df)

driver.quit()

6.

!pip install selenium

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(r"C:\Users\dell\OneDrive\Desktop\drivers\chromedriver-win64\chromedriver.exe")

url = 'https://www.flipkart.com/'

driver.get(url)

search\_field = driver.find\_element(By.CSS\_SELECTOR, 'input[type=text]')

search\_field.send\_keys('sneakers')

search\_button = driver.find\_element(By.CSS\_SELECTOR, 'button[type=submit]')

search\_button.click()

scroll\_pause\_time = 1

scroll\_height = 5

while scroll\_height >= 0:

driver.execute\_script("window.scrollTo(0, document.body.scrollHeight/{})".format(scroll\_height))

time.sleep(scroll\_pause\_time)

scroll\_height -= 1

sneakers = driver.find\_elements(By.CSS\_SELECTOR, '.\_1AtVbE')

data = []

for sneaker in sneakers:

brand = sneaker.find\_element(By.CSS\_SELECTOR, '.\_2WkVRV').text

description = sneaker.find\_element(By.CSS\_SELECTOR, '.IRpwTa').text

price = sneaker.find\_element(By.CSS\_SELECTOR, '.\_30jeq3.\_1\_WHN1').text

data.append({

'Brand': brand,

'Product Description': description,

'Price': price

})

if len(data) >= 100:

break

df = pd.DataFrame(data)

print(df)

driver.quit()