```
In [ ]:   #Name-Nikhil Lalwani
          #project_name- web scraping assignment 3
          #company_name- FlipRobo
```

```
In [184…  import selenium
          from selenium import webdriver
          import pandas as pd
          from selenium.webdriver.common.by import By
          import warnings
          warnings.filterwarnings('ignore')
          import time
```

```
In [185…  from bs4 import BeautifulSoup
          import pandas as pd
          from tabulate import tabulate
          import requests
          import urllib.request
```

```
In [115…  driver=webdriver.Chrome()
```

1. Write a python program which searches all the product under a particular product from www.amazon.in. The product to be searched will be taken as input from user. For e.g. If user input is 'guitar'. Then search for guitars.

```
In [116…  #opening website on chrome
          driver.get('http://www.amazon.in/')
```

```
In [117…  #seach guitar as input
          input_guitar=driver.find_element(By.XPATH,'/html/body/div[1]/header/div/div[1]/div[2]/
          input_guitar.send_keys('guitars')
```

```
In [118…  #click on submit button
          submit=driver.find_element(By.XPATH,'/html/body/div[1]/header/div/div[1]/div[2]/div/fc
          submit.click()
```

2. In the above question, now scrape the following details of each product listed in first 3 pages of your search results and save it in a data frame and csv. In case if any product has less than 3 pages in search results then scrape all the products available under that product name. Details to be

scraped are: "Brand Name", "Name of the Product", "Price", "Return/Exchange", "Expected Delivery", "Availability" and "Product URL". In case, if any of the details are missing for any of the product then replace it by "-".

In [121...
```python
Brand_Name=[]
name=driver.find_elements(By.XPATH,'//span[@class="a-size-base-plus a-color-base a-tex
for x in name:
    Brand_Name.append(x.text)
```

In [122...
```python
product_price=[]
price=driver.find_elements(By.XPATH,'//div[@class="a-row a-size-base a-color-base"]')
for x in price:
    product_price.append(x.text)
```

In [133...
```python
product_url=[]
url=driver.find_elements(By.XPATH,'//a[@class="a-link-normal s-underline-text s-underl
for x in url:
    url = x.get_attribute('href')
    product_url.append(url)


df=pd.DataFrame({'Brand_Name':Brand_Name,'product_price':product_price,'product_url':p
df
```

Out[133]:

	Brand_Name	product_price	product_url
0	Kadence A281 Professional Acoustic Rosewood gu...	₹6,499 M.R.P:\n₹7,999 (19% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
1	Kadence A281BK Professional Acoustic Rosewood ...	₹6,499 M.R.P:\n₹7,999 (19% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
2	Kadence rosewood Guitar Frontier Series, Elect...	₹4,999 M.R.P:\n₹6,800 (26% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
3	Kadence Slowhand Premium Jumbo Semi Acoustic G...	₹11,999 M.R.P:\n₹16,799 (29% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
4	Juârez Acoustic Guitar, 38 Inch Cutaway, 038C ...	₹1,999 M.R.P:\n₹6,990 (71% off)	https://www.amazon.in/Juarez-Acoustic-Cutaway-...
...	...	...	...
60	Yamaha F280 Acoustic Guitar, Tobacco Brown Sun...	₹7,990	https://www.amazon.in/Yamaha-Acoustic-Guitar-T...
61	Vault Traveller 34 Inch Acoustic Guitar With T...	₹4,084 M.R.P:\n₹9,600 (57% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
62	Imaginea® 41 Inch Guitar 3TS Sunburst Cutaway ...	₹5,499 M.R.P:\n₹10,999 (50% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
63	Medellin Acoustic Guitar + (Free online learni...	₹2,499 M.R.P:\n₹3,999 (38% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...
64	Imaginea® 40 inch Acoustic Guitar for Beginner...	₹5,499 M.R.P:\n₹9,999 (45% off)	https://www.amazon.in/sspa/click? ie=UTF8&spc=M...

65 rows × 3 columns

3. Write a python program to access the search bar and search button on images.google.com and scrape 10 images each for keywords 'fruits', 'cars' and 'Machine Learning', 'Guitar', 'Cakes'.

In [171...
```
driver=webdriver.Chrome()
```

In [174...
```
driver.get('https://www.google.com/search?sca_esv=561848188&sxsrf=AB5stBivgdZqEx8XiKqv
```

In [212...
```
driver.get('https://www.google.com/search?sca_esv=561848188&sxsrf=AB5stBivgdZqEx8XiKqv
fruit_urls = []
image = driver.find_elements(By.XPATH,'//img[@class="rg_i Q4LuWd"]')

for x in image:
    url = x.get_attribute('src')
    fruit_urls.append(url)
```

```
In [216…   df=pd.DataFrame({'fruit_urls':fruit_urls})
           df[:10]
```

Out[216]:

	fruit_urls
0	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
1	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
2	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
3	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
4	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
5	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
6	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
7	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
8	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...
9	data:image/jpeg;base64,/9j/4AAQSkZJRgABAQAAAQA...

4. Write a python program to search for a smartphone(e.g.: Oneplus Nord, pixel 4A, etc.) on www.flipkart.com and scrape following details for all the search results displayed on 1st page. Details to be scraped: "Brand Name", "Smartphone name", "Colour", "RAM", "Storage(ROM)", "Primary Camera", "Secondary Camera", "Display Size", "Battery Capacity", "Price", "Product URL". Incase if any of the details is missing then replace it by "- ". Save your results in a dataframe and CSV.

```
In [151…   url='https://www.flipkart.com/search?q=one%20plus%20nord&otracker=search&otracker1=sea
           r=requests.get(url)
           htmlcontent=r.content
           soup=BeautifulSoup(htmlcontent,'html.parser')

           product_details=[]
           details=soup.find_all('div',class_="col col-7-12")
           for x in details:
               product_details.append(x.text)

           product_price=[]
           price=soup.find_all('div',class_="_30jeq3 _1_WHN1")
```

```
for x in price:
    product_price.append(x.text)

df=pd.DataFrame({'product_details':product_details,'product_price':product_price})
df
```

Out[151]:

	product_details	product_price
0	OnePlus Nord CE 2 Lite 5G (Blue Tide, 128 GB)4...	₹17,737
1	OnePlus Nord CE 2 Lite 5G (Black Dusk, 128 GB)...	₹17,819
2	OnePlus Nord CE 3 Lite 5G (Chromatic Gray, 128...	₹19,999
3	OnePlus Nord CE 3 Lite 5G (Pastel Lime, 256 GB...	₹21,999
4	OnePlus Nord CE 3 Lite 5G (Chromatic Gray, 256...	₹21,999
5	OnePlus Nord 3 5G (Tempest Gray, 256 GB)4.3223...	₹36,566
6	OnePlus Nord 3 5G (Misty Green, 128 GB)4.5573 ...	₹31,962
7	OnePlus Nord 3 5G (Tempest Gray, 128 GB)4.5573...	₹31,141
8	OnePlus Nord CE3 5G (Grey Shimmer, 128 GB)4.41...	₹30,999
9	OnePlus Nord 2T 5G (Gray Shadow, 128 GB)4.411,...	₹28,989
10	OnePlus Nord CE3 5G (Aqua Surge, 128 GB)4.4181...	₹26,995
11	OnePlus Nord CE3 5G (Grey Shimmer, 256 GB)4.41...	₹32,999
12	OnePlus Nord 3 5G (Misty Green, 256 GB)4.3223 ...	₹36,935
13	OnePlus Nord (Gray Onyx, 64 GB)4.247 Ratings &...	₹22,689
14	OnePlus Nord 2T 5G (Jade Fog, 128 GB)4.411,849...	₹28,799
15	OnePlus Nord (Blue Marble, 128 GB)4.1171 Ratin...	₹24,790
16	OnePlus Nord 2T 5G (Gray Shadow, 128 GB)4.411,...	₹28,999
17	OnePlus Nord 2 5G (Blue Haze, 256 GB)4.2304 Ra...	₹35,999
18	OnePlus Nord 2T 5G (Gray Shadow, 128 GB)4.411,...	₹28,999
19	OnePlus Nord 2T 5G (Jade Fog, 256 GB)4.32,536 ...	₹33,990
20	OnePlus Nord 2T 5G (Jade Fog, 128 GB)4.411,849...	₹28,199
21	OnePlus Nord CE 3 Lite 5G (Lime, 256 GB)4.43,8...	₹25,999
22	OnePlus Nord 2 5G (Blue Haze, 128 GB)4.41,992 ...	₹33,999
23	OnePlus Nord CE 5G (Charcoal Ink, 128 GB)3.927...	₹22,989

5. Write a program to scrap geospatial coordinates (latitude, longitude) of a city searched on google maps.

```
In [161…   url='https://www.latlong.net/place/indore-madhya-pradesh-india-2528.html#:~:text=Latit
           r=requests.get(url)
           htmlcontent=r.content
           soup=BeautifulSoup(htmlcontent,'html.parser')

           city_indore=[]
           city=soup.find_all('table',class_="margina")
           for x in city:
               city_indore.append(x.text)

           df = pd.read_html(str(table))[0]
           df
```

Out[161]:

	0	1
0	Country	India
1	Latitude	22.719568
2	Longitude	75.857727
3	DMS Lat	22° 43' 10.4448'' N
4	DMS Long	75° 51' 27.8172'' E
5	UTM Easting	588084.03
6	UTM Northing	2512731.22
7	Category	Cities
8	Country Code	IN
9	Zoom Level	10

6. Write a program to scrap all the available details of best gaming laptops from digit.in.

```
In [12]:   driver=webdriver.Chrome()
```

```
In [13]:   driver.get('https://www.digit.in/top-products/best-gaming-laptops-40.html')
```

```
In [17]:   product_name=[]
           name=driver.find_elements(By.XPATH,'//span[@class="datahreflink"]')[:7]
           for x in name:
               product_name.append(x.text)

           specs_product=[]
           specs=driver.find_elements(By.XPATH,'//div[@class="Specs-Wrap"]')
           for x in specs:
               specs_product.append(x.text)
```

```
In [18]:   df=pd.DataFrame({'product_name':product_name,'specs_product':specs_product})
           df
```

Out[18]:

	product_name	specs_product
0	HP OMEN 17 (2023)	Windows 11 Home\n17.3" (2560 x 1440)\n13th Gen...
1	MSI Titan GT77 12UHS	Windows 11 Home\n17.3" (3840 x 2160)\n12th Gen...
2	Lenovo Legion 5i Pro	Windows 11 Home\n16" (2560 x 1600)\n12th Gen I...
3	ASUS ROG Strix Scar 18 2023	Windows 11 Home\n18" (1920 x 1200)\n13th Gen I...
4	Acer Predator Helios Neo 16	Windows 11 Home\n16" (2560 x 1600)\n13th Gen I...
5	ASUS ROG Zephyrus G14	Windows 11 Home\n14" (1920 x 1200)\nAMD Ryzen ...
6	MSI Cyborg 15	Windows 11 Home\n15.6" (1920 x 1080)\n12th Gen...

7. Write a python program to scrape the details for all billionaires from www.forbes.com. Details to be scrapped: "Rank", "Name", "Net worth", "Age", "Citizenship", "Source", "Industry".

In [30]:
```python
driver=webdriver.Chrome()
```

In [31]:
```python
driver.get('https://www.forbes.com/lists/india-billionaires/?sh=37fd4156109b')
```

In [32]:
```python
person_rank=[]
rank=driver.find_elements(By.XPATH,'//div[@class="rank first table-cell    rank"]')
for x in rank:
    person_rank.append(x.text)
```

In [33]:
```python
person_name=[]
name=driver.find_elements(By.XPATH,'//DIV[@class="personName second table-cell    name
for x in name:
    person_name.append(x.text)
```

In [34]:
```python
person_Net_worth=[]
Net_worth=driver.find_elements(By.XPATH,'//div[@class="finalWorth  table-cell    net w
for x in Net_worth:
    person_Net_worth.append(x.text)
```

In [35]:
```python
person_Industry=[]
Industry=driver.find_elements(By.XPATH,'//div[@class="category  table-cell    industry
for x in Industry:
    person_Industry.append(x.text)
```

In [40]:
```python
df=pd.DataFrame({'person_rank':person_rank,'person_name':person_name,'person_Net_worth
df[:50]
```

Out[40]:

	person_rank	person_name	person_Net_worth	person_Industry
0	1.	Gautam Adani & family	$150 B	Diversified
1	2.	Mukesh Ambani	$88 B	Diversified
2	3.	Radhakishan Damani & family	$27.6 B	Fashion & Retail
3	4.	Cyrus Poonawalla	$21.5 B	Healthcare
4	5.	Shiv Nadar	$21.4 B	Technology
5	6.	Savitri Jindal & family	$16.4 B	Metals & Mining
6	7.	Dilip Shanghvi & family	$15.5 B	Healthcare
7	8.	Hinduja brothers	$15.2 B	Diversified
8	9.	Kumar Birla	$15 B	Diversified
9	10.	Bajaj Family	$14.6 B	Diversified
10	11.	Sunil Mittal & family	$14.5 B	Telecom
11	12.	Uday Kotak	$14.3 B	Finance & Investments
12	13.	Shapoor Mistry & family	$14.2 B	Diversified
13	14.	Godrej family	$13.9 B	Diversified
14	15.	Lakshmi Mittal	$13.8 B	Metals & Mining
15	16.	Madhukar Parekh & family	$12.6 B	Manufacturing
16	17.	Burman family	$9.6 B	Food & Beverage
17	18.	Azim Premji	$9.3 B	Technology
18	19.	Kushal Pal Singh	$8.8 B	Real Estate
19	20.	Ashwin Dani & family	$8.4 B	Manufacturing
20	21.	Ravi Jaipuria	$8.1 B	Food & Beverage
21	22.	Kuldip Singh & Gurbachan Singh Dhingra	$6.8 B	Manufacturing
22	23.	Vikram Lal & family	$6.6 B	Automotive
23	24.	Mahendra Choksi & family	$6.5 B	Manufacturing
24	25.	Murali Divi & family	$6.45 B	Healthcare
25	26.	Sudhir & Samir Mehta	$6.4 B	Healthcare
26	27.	Vinod & Anil Rai Gupta & family	$6.3 B	Manufacturing
27	28.	Hasmukh Chudgar & family	$6.2 B	Healthcare
28	29.	Benu Gopal Bangur	$6 B	Manufacturing
29	30.	Rekha Jhunjhunwala	$5.9 B	Finance & Investments
30	31.	Murugappa Family	$5.8 B	Diversified
31	32.	Harsh Mariwala & family	$5.7 B	Food & Beverage
32	33.	Vijay Chauhan & family	$5.5 B	Food & Beverage

	person_rank	person_name	person_Net_worth	person_Industry
33	34.	Girdhari Lal Bawri, Rajendra Agarwal and Banwa...	$5.45 B	Healthcare
34	35.	M.A. Yusuff Ali	$5.4 B	Fashion & Retail
35	36.	Vakil family	$5.2 B	Manufacturing
36	37.	Mangal Prabhat Lodha	$5.1 B	Real Estate
37	38.	Kapil & Rahul Bhatia	$4.9 B	Service
38	39.	Pavan & Vivek Jain	$4.55 B	Manufacturing
39	40.	Singh family	$4.5 B	Healthcare
40	41.	N.R. Narayana Murthy	$4.3 B	Technology
41	42.	Ramesh Juneja & family	$4.2 B	Healthcare
42	43.	P.P. Reddy & family	$4.1 B	Construction & Engineering
43	44.	Falguni Nayar	$4.08 B	Fashion & Retail
44	45.	Muthoot family	$4.05 B	Finance & Investments
45	46.	Chandru Raheja	$4 B	Real Estate
46	47.	Yusuf Hamied & family	$3.9 B	Healthcare
47	48.	Sridhar Vembu & Siblings	$3.8 B	Technology
48	49.	Pankaj Patel	$3.77 B	Healthcare
49	50.	Ravi Modi	$3.75 B	Fashion & Retail

8. Write a program to extract at least 500 Comments, Comment upvote and time when comment was posted from any YouTube Video

```
In [43]:  driver=webdriver.Chrome()
```

```
In [44]:  driver.get('https://www.youtube.com/watch?v=0oeK3gmtf7o')
```

```
In [56]:  comment_person_name=[]
          name=driver.find_elements(By.XPATH,'//span[@class=" style-scope ytd-comment-renderer s
          for x in name:
              comment_person_name.append(x.text)
```

```
In [58]:  person_comment=[]
          comment=driver.find_elements(By.XPATH,'//div[@class="style-scope ytd-expander"]')[0:56
          for x in comment:
              person_comment.append(x.text)
```

```
In [59]:  post_time=[]
          time=driver.find_elements(By.XPATH,'//a[@class="yt-simple-endpoint style-scope yt-form
          for x in time:
              post_time.append(x.text)
```

```
In [71]:  df=pd.DataFrame({'comment_person_name':comment_person_name,'person_comment':person_com
          index_to_remove = [0,1]
          df = df.drop(index_to_remove)
          df = df.reset_index(drop=True)
          df
```

Out[71]:

	comment_person_name	person_comment	post_time
0	@rameshlaldas4299	sir Netflix Amazon prime want to now your adre…	2 years ago (edited)
1	@nirmalagosha7797	People can say what they want, but you just ca…	1 year ago
2	@rejikrejik	The best part of these series \n1.Full of susp…	2 years ago
3	@takestock569	Gundey needs guns to kill \nGovt need taxes to…	3.Giving
4	@dharmilprajapati4412	I wish i could like a thousand times!! Your st…	2 years ago
...	...	...	...
493	@himynameis5709	Ailla,march se double triple nahi ho raha hai	2 years ago
494	@romitpatel3635	bhaiya make video on kolkata group	2 years ago
495	@DrSaurin	Next video fast like toko sab	4 months ago
496	@PrabirDasiamprabir	Sir jald hii dusra part	1 year ago
497	@forall1390	"Like" scam by AK	2 years ago

498 rows × 3 columns

9. Write a python program to scrape a data for all available Hostels from https://www.hostelworld.com/ in "London" location. You have to scrape hostel name, distance from city centre, ratings, total reviews, overall reviews, privates from price, dorms from price, facilities and property description.

```
In [72]:  driver=webdriver.Chrome()
```

```
In [74]:  driver.get('https://www.hostelworld.com/pwa/wds/s?q=London,%20England&country=London&c
```

```
In [108…  hostel_name=[]
          name=driver.find_elements(By.XPATH,'//div[@class="property-name"]')[:30]
```

```
for x in name:
    hostel_name.append(x.text)
```

In [111…
```
distance_from_city_centre=[]
distance=driver.find_elements(By.XPATH,'//span[@class="distance-description"]')[:30]
for x in distance:
    distance_from_city_centre.append(x.text)
```

In [110…
```
hostel_ratings=[]
ratings=driver.find_elements(By.XPATH,'//span[@class="number"]')[:30]
for x in ratings:
    hostel_ratings.append(x.text)
```

In [113…
```
overall_reviews=[]
review=driver.find_elements(By.XPATH,'//span[@class="keyword"]')[:30]
for x in review:
    overall_reviews.append(x.text)
```

In [114…
```
df=pd.DataFrame({'hostel_name':hostel_name,'distance_from_city_centre':distance_from_c
df[0:30]
```

Out[114]:

	hostel_name	distance_from_city_centre	hostel_ratings	overall_reviews
0	Clink261	- 4.9km from city centre	8.1	Fabulous
1	St Christopher's Village	- 5.1km from city centre	8.2	Fabulous
2	Palmers Lodge - Swiss Cottage	- 1.3km from city centre	8.6	Fabulous
3	YHA London Thameside	- 0.5km from city centre	9.2	Superb
4	YHA London Earls Court	- 2.9km from city centre	6.5	Good
5	Strand Continental	- 10.9km from city centre	7.9	Very Good
6	PubLove @ The Steam Engine, Waterloo	- 3.2km from city centre	7.4	Very Good
7	YHA London Central	- 0.7km from city centre	8.5	Fabulous
8	The Backpackshack	- 1.7km from city centre	7.7	Very Good
9	St Christopher's Inn - Liverpool Street	- 2.1km from city centre	5.8	Very Good
10	London Waterloo Hostel	- 3km from city centre	7.2	Fabulous
11	Safestay London Elephant & Castle	- 5.1km from city centre	8.8	Very Good
12	YHA London Oxford Street	- 5.5km from city centre	7.6	Fabulous
13	Generator London	- 3.2km from city centre	8.8	Fabulous
14	Hostelle - women only hostel London	- 1.8km from city centre	8.6	Fabulous
15	Selina Camden	- 1.8km from city centre	8.1	Fabulous
16	Clink261	- 1.8km from city centre	8.2	Very Good
17	St Christopher's Village	- 1.8km from city centre	7.0	Very Good
18	Astor Victoria	- 1.7km from city centre	7.9	Fabulous
19	St Christopher's Inn - London Bridge	- 3.4km from city centre	8.2	Very Good
20	St Christopher's Oasis-Female Only	- 5.5km from city centre	7.6	Good
21	Dover Castle Hostel and Bar	- 6.1km from city centre	6.4	Superb
22	PubLove @ The Exmouth Arms, Euston	- 5.4km from city centre	9.7	Fabulous
23	Onefam Notting Hill	- 5km from city centre	8.3	Superb
24	NX London Hostel	- 5.5km from city centre	9.5	Very Good
25	Urbany Hostel London	- 5.5km from city centre	7.5	Superb
26	Smart Hyde Park Inn Hostel	- 3.3km from city centre	9.0	Very Good
27	Barmy Badger Backpackers	- 4.2km from city centre	7.7	Fabulous
28	Saint James Backpackers	- 1.9km from city centre	8.7	Very Good
29	Kabannas London St Pancras	- 2.6km from city centre	7.7	Fabulous

In [ ]: