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 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 [1]:
```
import selenium
import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.by import By

import time

import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(5)
```

In [3]:
```
# Opening naukri page on the automated chrome browser
driver.get("https://www.naukri.com/")
```

In [4]:
```
# entering designation as Data analyst

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

# we have class in the page that's why we are using class name, if it was a xpath we could have used By XPATH

designation.send_keys('Data Analyst') # it will send the data analyst search
```

In [5]:
```
# entering location as banglore using absolute xpath method
location = driver.find_element(By.XPATH,"/html/body/div[1]/div[6]/div/div/div[5]/div/div/div/input")
location.send_keys('Banglore')
```

In [6]:
```
search = driver.find_element(By.CLASS_NAME,"qsbSubmit")
search.click()
time.sleep(5) # making sure the page is loaded properly
```

In [7]:
```
# Creating a lists to append the data according to their title,location, company name and experience
job_title = []
job_location = []
company_name = []
experience = []
```

In [8]:
```
# scraping the first 10 job details

title_tags = driver.find_elements(By.XPATH,'//a[@class="title fw500 ellipsis"]')
for i in title_tags[0:10]: # first ten titles
    title = i.text
    job_title.append(title)

location_tags = driver.find_elements(By.XPATH,'//li[@class="fleft grey-text br2 placeHolderLi location"]')
for i in location_tags[0:10]:
    location = i.text
    job_location.append(location)

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

```
experience_tags = driver.find_elements(By.XPATH,'//li[@class="fleft grey-text br2 placeHolderLi experience"]')
for i in experience_tags[0:10]:
    exp = i.text
    experience.append(exp)
```

In [9]:
```
print(len(job_title),len(job_location),len(company_name),len(experience))
```

```
10 10 10 10
```

In [10]:
```
# Creating a DataFrame to store the data
df = pd.DataFrame({'Designation':job_title,'Location':job_location,'Company_name':company_name,'Experience':exper
df # pardon my column names, It is not the same as given in the question
```

Out[10]:

| | Designation | Location | Company_name | Experience |
|---|---|---|---|---|
| 0 | Lead Data Analyst - 1st LOD | Bangalore/Bengaluru, Delhi / NCR | Silicon Valley Bank | 3-8 Yrs |
| 1 | Data Analyst | Bangalore/Bengaluru | Herox | 0-1 Yrs |
| 2 | Senior Data Analyst | Bangalore/Bengaluru | TEKsystems | 7-10 Yrs |
| 3 | Senior Data Analyst | Bangalore/Bengaluru | nurture.farm | 3-8 Yrs |
| 4 | Senior Data Analyst | Bangalore/Bengaluru | Walmart | 6-10 Yrs |
| 5 | Data Analyst - Python / Artificial Intelligence | Kolkata, Mumbai, Hyderabad/Secunderabad, Luckn... | iMindYourBusiness | 0-2 Yrs |
| 6 | Data Analyst ( ADF, Azure Databricks ) | Bangalore/Bengaluru, Greater Noida | Coforge | 6-10 Yrs |
| 7 | Data Analyst, Digital Business | Bangalore/Bengaluru | Sony Pictures Networks | 2-5 Yrs |
| 8 | Senior Data Analyst | Bangalore/Bengaluru | Mobile Premier League | 3-5 Yrs |
| 9 | Executive Data Analyst | Bangalore/Bengaluru(Dodda Banaswadi) | ITC Limited (ITC Foods & Beverages Division) | 3-7 Yrs |

## Q2) Finding the same for data Scientists jobs

In [11]:
```
driver.get("https://www.naukri.com/") # going back to the home page
```

In [12]:
```
designation = driver.find_element(By.CLASS_NAME,"suggestor-input")

designation.send_keys('Data Scientist')
```

In [13]:
```
location = driver.find_element(By.XPATH,"/html/body/div[1]/div[6]/div/div/div[5]/div/div/div/input")

location.send_keys('Banglore')
```

In [14]:
```
search = driver.find_element(By.CLASS_NAME,"qsbSubmit")
search.click()
time.sleep(5)
```

In [15]:
```
job_title = []
job_location = []
company_name = []
```

In [16]:
```
title_tags = driver.find_elements(By.XPATH,'//a[@class="title fw500 ellipsis"]')
for i in title_tags[0:10]: # first ten titles
    title = i.text
    job_title.append(title)

location_tags = driver.find_elements(By.XPATH,'//li[@class="fleft grey-text br2 placeHolderLi location"]')
for i in location_tags[0:10]:
    location = i.text
    job_location.append(location)

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

In [18]:
```
print(len(job_title),len(job_location),len(company_name))
```

```
10 10 10
```

In [19]:
```
# Creating dataframe for data scientists jobs data
df = pd.DataFrame({'Designation':job_title,'Location':job_location,'Company_name':company_name})
df
```

Out[19]:

| | Designation | Location | Company_name |
|---|---|---|---|
| 0 | Data Science Specialist | Kolkata, Mumbai, Hyderabad/Secunderabad, Pune,... | Accenture |
| 1 | Data Science Manager | Kolkata, Mumbai, Hyderabad/Secunderabad, Pune,... | Accenture |
| 2 | Mongodb Database Administrator, Maria DB or Ca... | Hyderabad/Secunderabad, Pune, Chennai, Bangalo... | Mphasis |
| 3 | Analystics & Modeling Specialist | Kolkata, Mumbai, Hyderabad/Secunderabad, Pune,... | Accenture |
| 4 | Senior Data Scientist | Mumbai, New Delhi, Chennai, Bangalore/Bengaluru | Boston Consulting Group |
| 5 | Assistant Manager - Data Science | Mumbai, Pune, Bangalore/Bengaluru | CitiusTech |
| 6 | Senior Data Scientist | Pune, Chennai, Bangalore/Bengaluru | Wipro |
| 7 | Data Scientist | Hyderabad/Secunderabad, Pune, Chennai, Bangalo... | Tech Mahindra |
| 8 | Opportunity For Senior Data Scientist/ Busines... | Gurgaon/Gurugram, Bangalore/Bengaluru, Delhi /... | PayU |
| 9 | Senior Data Scientist | Bangalore/Bengaluru | Cargill |

## Q3) In this question you have to scrape data using the filters available on the webpage

In [5]:
```
import selenium
import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.by import By

import time

import warnings
warnings.filterwarnings('ignore')
```

In [6]:
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(5)
```

In [7]:
```
driver.get("https://www.naukri.com/") # going to the home page
```

In [8]:
```
# Entering Data Scientist on skill, designation or companies field
designation = driver.find_element(By.CLASS_NAME,"suggestor-input")

designation.send_keys('Data Scientist')
```

In [9]:
```
# Now searching
search = driver.find_element(By.CLASS_NAME,"qsbSubmit")
search.click()
time.sleep(3)
```

In [11]:
```
# Checking Location filter Delhi
location_filter = driver.find_element(By.XPATH,'/html/body/div[1]/div[4]/div/section[1]/div[2]/div[5]/div[2]/div[
location_filter.click()
time.sleep(3)
```

In [12]:
```
# Checking salary filter 3-6 lakhs
salary_filter = driver.find_element(By.XPATH,'/html/body/div[1]/div[4]/div/section[1]/div[2]/div[6]/div[2]/div[2]
salary_filter.click()
time.sleep(3) # using sleep to load
```

In [13]:
```
job_title = []
job_location = []
company_name = []
experience_required = []
```

In [14]:
```
title_tags = driver.find_elements(By.XPATH,'//a[@class="title fw500 ellipsis"]')
```

```
for i in title_tags[0:10]: # first ten titles
    title = i.text
    job_title.append(title)

location_tags = driver.find_elements(By.XPATH,'//li[@class="fleft grey-text br2 placeHolderLi location"]')
for i in location_tags[0:10]:
    location = i.text
    job_location.append(location)

experience_tags = driver.find_elements(By.XPATH,'//li[@class="fleft grey-text br2 placeHolderLi experience"]')
for i in experience_tags[0:10]:
    experience = i.text
    experience_required.append(experience)

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

In [15]:
```
print(len(job_title),len(job_location),len(company_name),len(experience_required))
```

```
10 10 10 10
```

In [16]:
```
# Making the scraped data into a DataFrame
df = pd.DataFrame({'Job_Title':job_title,'Job_Location':job_location,'Company_name':company_name,'Experience_Requ
df
```

Out[16]:

| | Job_Title | Job_Location | Company_name | Experience_Required |
|---|---|---|---|---|
| 0 | Data Scientist - Engine Algorithm | Delhi / NCR, Kolkata, Mumbai, Hyderabad/Secund... | Primo Hiring | 1-3 Yrs |
| 1 | Data Scientist | Noida, Nagpur, Bangalore/Bengaluru | GlobalLogic | 8-10 Yrs |
| 2 | DigitalBCG GAMMA Data Scientist | New Delhi, Bangalore/Bengaluru | Boston Consulting Group | 2-5 Yrs |
| 3 | Data Activation Specialist - Adobe Target | Delhi / NCR, Kolkata, Mumbai, Hyderabad/Secund... | Okda Solutions | 7-10 Yrs |
| 4 | Data Scientist | Gurgaon/Gurugram | IHS Markit | 3-6 Yrs |
| 5 | Lead Data Scientist | Noida(Sector-59 Noida)\n(WFH during Covid) | R Systems International | 7-10 Yrs |
| 6 | Data Scientist | Gurgaon/Gurugram | Optum | 2-7 Yrs |
| 7 | Data Scientist / Chat-bot Developer | New Delhi, Bangalore/Bengaluru, Mumbai (All Ar... | Big Seo Buzz | 3-7 Yrs |
| 8 | Data Scientist - Mumbai - Immediate Joiner Req... | Delhi / NCR, Mumbai, New Delhi | HueCanvas Consulting | 2-7 Yrs |
| 9 | Data Scientist | Noida | NGI Ventures | 0-5 Yrs |

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

1. Brand
2. Product Description
3. Price

In [23]:
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(5)
```

In [24]:
```
driver.get('https://www.flipkart.com/')
time.sleep(3)
```

In [25]:
```
search_sunglasses = driver.find_element(By.CLASS_NAME,"_3704LK")

search_sunglasses.send_keys("sunglasses")
```

In [26]:
```
search_button = driver.find_element(By.CLASS_NAME,"L0Z3Pu")
search_button.click()
```

In [27]:
```
brand = []
```

```
p_description = []
price = []
```

In [28]:
```
brand_tags = driver.find_elements(By.XPATH,'//div[@class="_2WkVRV"]')
for i in brand_tags:
    brand.append(i.text)

descript_tags = driver.find_elements(By.XPATH,"//a[@class='IRpwTa']")
for i in descript_tags:
    p_description.append(i.text)

price_tags = driver.find_elements(By.XPATH,'//div[@class="_30jeq3"]')
for i in price_tags:
    price.append(i.text)
```

In [29]:
```
print(len(brand),len(p_description),len(price)) # we have scraped 40 brands,descriptions and price data
```

```
40 40 40
```

In [30]:
```
# Now let's go to next page
next_page = driver.find_element(By.XPATH,'//a[@class="ge-49M"]')
next_page.click()
time.sleep(2)
```

In [31]:
```
# Using the same method to append the second page data
brand_tags = driver.find_elements(By.XPATH,'//div[@class="_2WkVRV"]')
for i in brand_tags:
    brand.append(i.text)

descript_tags = driver.find_elements(By.XPATH,"//a[@class='IRpwTa']")
for i in descript_tags:
    p_description.append(i.text)

price_tags = driver.find_elements(By.XPATH,'//div[@class="_30jeq3"]')
for i in price_tags:
    price.append(i.text)
```

In [32]:
```
print(len(brand),len(p_description),len(price)) # we got this page data aswell
```

```
80 80 80
```

In [34]:
```
# Going to the 3rd page
third_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div[1]/div[2]/div[12]/div/div/nav/a[3]')
third_page.click()
time.sleep(2)
```

In [35]:
```
# We got 80 sunglasses data, we need 20 more
brand_tags = driver.find_elements(By.XPATH,'//div[@class="_2WkVRV"]')
for i in brand_tags[:20]:
    brand.append(i.text)

descript_tags = driver.find_elements(By.XPATH,"//a[@class='IRpwTa']")
for i in descript_tags[:20]:
    p_description.append(i.text)

price_tags = driver.find_elements(By.XPATH,'//div[@class="_30jeq3"]')
for i in price_tags[:20]:
    price.append(i.text)
```

In [36]:
```
print(len(brand),len(p_description),len(price)) # we got all the 100 sunglasses data
```

```
100 100 100
```

In [37]:
```
# Storing in a DataFrame
sunglasses_df = pd.DataFrame({'Brand':brand,'Product_Description':p_description,'Price':price})
sunglasses_df
```

Out[37]:

| | Brand | Product_Description | Price |
|---|---|---|---|
| 0 | Singco India | UV Protection, Riding Glasses, Others Aviator,... | ₹203 |

| | | | |
|---|---|---|---|
| 1 | Singco India | Gradient, Toughened Glass Lens, UV Protection ... | ₹631 |
| 2 | Fastrack | UV Protection Wayfarer Sunglasses (Free Size) | ₹799 |
| 3 | SHAAH COLLECTIONS | UV Protection Round Sunglasses (54) | ₹179 |
| 4 | LIZA ANGEL | UV Protection Wayfarer Sunglasses (Free Size) | ₹199 |
| ... | ... | ... | ... |
| 95 | VINCENT CHASE | by Lenskart Polarized, UV Protection Wayfarer ... | ₹699 |
| 96 | kingsunglasses | UV Protection, Riding Glasses, Mirrored Wayfar... | ₹213 |
| 97 | Fastrack | UV Protection Wrap-around Sunglasses (63) | ₹710 |
| 98 | VINCENT CHASE | Polarized, UV Protection Round Sunglasses (51) | ₹699 |
| 99 | DEIXELS | UV Protection Aviator Sunglasses (Free Size) | ₹229 |

100 rows × 3 columns

## Q5: Scrape 100 reviews data from flipkart.com for iphone11 phone

In [152…
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(5)
```

In [153…
```
# Going back to the homepage
driver.get('https://www.flipkart.com/')
time.sleep(3)
```

In [154…
```
iphone_search = driver.find_element(By.CLASS_NAME,"_3704LK")
iphone_search.send_keys('iphone 11')
```

In [155…
```
# clicking on the search
srch_btn = driver.find_element(By.CLASS_NAME,"L0Z3Pu")
srch_btn.click()
```

In [156…
```
# clicking on the iphone
iphone_click = driver.find_element(By.CLASS_NAME,"_4rR01T")
iphone_click.click()
```

In [157…
```
driver.get('https://www.flipkart.com/apple-iphone-11-black-128-gb/p/itm8244e8d955aba?pid=MOBFWQ6BKRYBP5X8&lid=LST
```

In [158…
```
all_reviews = driver.find_element(By.XPATH,'//div[@class="_3UAT2v _16PBlm"]')
all_reviews.click()
```

In [159…
```
ratings = []
review_summary = []
review = []
```

In [196…
```
rating_tags = driver.find_elements(By.XPATH,'//div[@class="_3LWZlK _1BLPMq"]')
for i in rating_tags:
    ratings.append(i.text)

summary_tags = driver.find_elements(By.XPATH,'//p[@class="_2-N8zT"]')
for i in summary_tags:
    review_summary.append(i.text)

review_tags = driver.find_elements(By.XPATH,'//div[@class="t-ZTKy"]')
for i in review_tags:
    review.append(i.text)
```

In [197…
```
print(len(ratings),len(review_summary),len(review))
```

```
100 103 103
```

In [162…
```
ratings
```

Out[162…
```
['5', '4', '5', '5', '5', '5', '5', '5', '5', '5']
```

In [163…
```
sec_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[2]')
sec_page.click()
```

In [166…
```
third_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[4]')
third_page.click()
```

In [169…
```
fourth_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[5]'
fourth_page.click()
```

In [172…
```
fifth_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[6]')
fifth_page.click()
```

In [175…
```
sixth_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[7]')
sixth_page.click()
```

In [178…
```
seventh_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[8]
seventh_page.click()
```

In [181…
```
eighth_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[7]'
eighth_page.click()
```

In [184…
```
nineth_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[7]'
nineth_page.click()
```

In [189…
```
tenth_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[13]/div/div/nav/a[7]')
tenth_page.click()
```

In [192…
```
missing_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[6]/div/div/nav/a[7]'
missing_page.click()
```

In [195…
```
another_page = driver.find_element(By.XPATH,'/html/body/div[1]/div/div[3]/div/div/div[2]/div[8]/div/div/nav/a[7]'
another_page.click()
```

In [198…
```
print(len(ratings),len(review_summary),len(review))
```

```
100 103 103
```

## 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 [239…
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(2)
```

In [240…
```
driver.get('https://www.flipkart.com/')
time.sleep(2)
```

In [241…
```
sneakers_search = driver.find_element(By.CLASS_NAME,"_3704LK")
```

```
sneakers_search.send_keys('sneakers')
```

In [242…
```
srch_btn = driver.find_element(By.CLASS_NAME,"L0Z3Pu")
srch_btn.click()
```

In [245…
```
sneak_brands = []
product_description = []
PRice = []
Discount = []
```

In [246…
```
# Iterating through the tags and appending in our lists
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']")
    discount=driver.find_elements(By.XPATH,"//div[@class='_3Ay6Sb']")

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



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


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


    for t in discount:
        Discount.append(t.text)
        Discount[:100]
```

In [248…
```
print(len(sneak_brands),len(product_description),len(PRice),len(Discount))
```

```
120 114 135 120
```

In [251…
```
#storing in dataframe
sneak_df = pd.DataFrame({'Brand':sneak_brands[:100],'Product_description':product_description[:100],'Price':PRice
sneak_df
```

Out[251…

| | Brand | Product_description | Price | Discount |
|---|---|---|---|---|
| **0** | Alquwex | Sneakers For Men | ₹524₹1,19956% off | 56% off |
| **1** | RapidBox | Modern Trendy Shoes Sneakers For Men | ₹580₹99941% off | 41% off |
| **2** | BRUTON | Sneakers For Men | ₹470₹1,29963% off | 63% off |
| **3** | RED TAPE | Sneakers For Men | ₹1,499₹4,99970% off | 70% off |
| **4** | Labbin | Lightweight Pack Of 1 Trendy Sneakers Sneakers... | ₹499₹99950% off | 50% off |
| **...** | ... | ... | ... | ... |
| **95** | Robbie jones | Casuals, Canvas, Partywear Sneakers For Men | ₹259₹59956% off | 45% off |
| **96** | Shozie | Sneakers For Men | ₹298₹49940% off | 52% off |
| **97** | Deals4you | Sneaker Sneakers For Men | ₹590₹99940% off | 61% off |
| **98** | Magnolia | denim fabric shoes men and boys canvas sneaker... | ₹1,499₹4,99970% off | 55% off |
| **99** | Shozie | Sneakers Sneakers For Men | ₹499₹99950% off | 50% off |

100 rows × 4 columns

## Q7) Go to the link - https://www.myntra.com/shoes

### Set second Price filter and Color filter to “Black”

In [281…
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(2)
```

In [282…
```
driver.get('https://www.myntra.com/shoes')
time.sleep(2)
```

In [283…
```
check_sec_price = driver.find_element(By.XPATH,'/html/body/div[2]/div/div[1]/main/div[3]/div[1]/section/div/div[5
check_sec_price.click()
```

In [284…
```
black_filter = driver.find_element(By.XPATH,'/html/body/div[2]/div/div[1]/main/div[3]/div[1]/section/div/div[6]/u
black_filter.click()
```

In [285…
```
shoe_brand = []
descr = []
PRIce = []
```

In [289…
```
shoe_tags = driver.find_elements(By.XPATH,'//h3[@class="product-brand"]')
for i in shoe_tags:
    shoe_brand.append(i.text)


descr_tags = driver.find_elements(By.XPATH,'//h4[@class="product-product"]')
for i in descr_tags:
    descr.append(i.text)


prce_tags = driver.find_elements(By.XPATH,'//div[@class="product-price"]')
for i in prce_tags:
    PRIce.append(i.text)
```

In [290…
```
print(len(shoe_brand),len(descr),len(PRIce))
```

```
100 100 100
```

In [288…
```
next_page1 = driver.find_element(By.XPATH,'//li[@class="pagination-number"]')
next_page1.click()
```

In [291…
```
# Making A DATAFRAME
myntra_df = pd.DataFrame({'Brand':shoe_brand,'Description':descr,'Price':PRIce})
myntra_df
```

Out[291…

| | Brand | Description | Price |
|---|---|---|---|
| **0** | Skechers | Men GO WALK - TERRA Shoes | Rs. 8499Rs. 9999(15% OFF) |
| **1** | ADIDAS | Men 4DFWD_Pulse Running Shoes | Rs. 11199Rs. 15999(30% OFF) |
| **2** | Skechers | Men Go Run Hyper Burst Running | Rs. 7224Rs. 8499(15% OFF) |
| **3** | ADIDAS Originals | Men ZX 22 BOOST Sneakers | Rs. 10199Rs. 11999(15% OFF) |
| **4** | UNDER ARMOUR | Men HOVR SonicSE Running Shoes | Rs. 8499Rs. 9999(15% OFF) |
| **...** | ... | ... | ... |
| **95** | ALDO | Textured Block Sandals | Rs. 7999 |
| **96** | Tommy Hilfiger | Men Leather Sneakers | Rs. 6509Rs. 9299(30% OFF) |
| **97** | fitflop | Embellished PU Comfort Pumps | Rs. 7499 |
| **98** | ALDO | Women Leather Horsebit Loafers | Rs. 11999 |
| **99** | ALDO | Men Sneakers | Rs. 10999 |

100 rows × 3 columns

## Q8: 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”

In [457…
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
```

```
time.sleep(2)
```

In [458…
```
driver.get('https://www.amazon.in/')
time.sleep(2)
```

In [294…
```
search_laptop = driver.find_element(By.XPATH,"//input[@type='text']")
search_laptop.send_keys('Laptop')
```

In [296…
```
searchbtn = driver.find_element(By.XPATH,"//input[@id='nav-search-submit-button']")
searchbtn.click()
```

In [298…
```
filter_Cpu = driver.find_element(By.XPATH,'/html/body/div[1]/div[2]/div[1]/div[2]/div/div[3]/span/div[1]/div/div/
filter_Cpu.click()
```

In [301…
```
Title=[]
PrIce=[]
RAting=[]
```

In [302…
```
tit_tags = driver.find_elements(By.XPATH,'//h2[@class="a-size-mini a-spacing-none a-color-base s-line-clamp-2"]')
for i in tit_tags[:10]:
    Title.append(i.text)


pr_tags = driver.find_elements(By.XPATH,'//span[@class="a-price-whole"]')
for i in pr_tags[:10]:
    PrIce.append(i.text)


Rat_tags = driver.find_elements(By.XPATH,'//span[@class="a-icon-alt"]')
for i in Rat_tags[:10]:
    RAting.append(i.text)
```

In [303…
```
print(len(Title),len(PrIce),len(RAting))
```

```
10 10 10
```

## 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.

In [494…
```
driver = webdriver.Chrome(r"C:\Users\deeks\Downloads\chromedriver_win32\chromedriver.exe") # connecting to the dr
time.sleep(2)
```

In [495…
```
driver.get('https://www.ambitionbox.com/')
time.sleep(2)
```

In [496…
```
clickon_salry = driver.find_element(By.XPATH,'/html/body/div/div/div/div[1]/header/nav/ul/li[3]')
```

In [497…
```
from selenium.webdriver import ActionChains # using actionchains module
```

In [498…
```
actions = ActionChains(driver)
```

In [499…
```
actions.move_to_element(clickon_salry).perform() # chaining the tags to hover and click on the salries
time.sleep(2)
driver.find_element(By.XPATH,'/html/body/div/div/div/div[1]/header/nav/ul/li[3]/div/ul/li[1]/div/div[2]/a').click
```

In [500…
```
salar_search =  driver.find_element(By.XPATH,'//input[@id="jobProfileSearchbox"]')
salar_search.send_keys('Data Scientist')
time.sleep(2)
driver.find_element(By.XPATH,'//div[@class="suggestion_wrap tt-suggestion tt-selectable"]').click()
```

In [561…
```
Title_tags = []
```

```
Title_tags = []
EXperience = []
Min_sal = []
Max_sal = []
Avg_sal = []
```

In [562…
```
tittle_taggs = driver.find_elements(By.XPATH,'//div[@class="company-info"]')
for i in tittle_taggs:
    Title_tags.append(i.text.replace('\n','').split())
```

In [563…
```
Title_tags = Title_tags[0:10]
```

In [565…
```
exp_tags =  driver.find_elements(By.XPATH,'//div[@class="sbold-list-header"]')
for i in exp_tags:
    EXperience.append(i.text)
```

In [566…
```
EXperience
```

Out[566…
```
['3-4 yrs experience (based on 24 salaries)',
 '2-4 yrs experience (based on 59 salaries)',
 '2-4 yrs experience (based on 49 salaries)',
 '1-2 yrs experience (based on 35 salaries)',
 '2-4 yrs experience (based on 118 salaries)',
 '1 yr experience (based on 10 salaries)',
 '2-4 yrs experience (based on 70 salaries)',
 '4 yrs experience (based on 11 salaries)',
 '4 yrs experience (based on 10 salaries)',
 '3 yrs experience (based on 14 salaries)']
```

In [578…
```
sal_minmax = []
```

In [579…
```
sal_minmax_tags = driver.find_elements(By.XPATH,'//div[@class="value body-medium"]')
for i in sal_minmax_tags:
    sal_minmax.append(i.text)
```

In [580…
```
sal_minmax
```

Out[580…
```
['₹ 25.0L',
 '₹ 45.0L',
 '₹ 15.0L',
 '₹ 26.0L',
 '₹ 11.0L',
 '₹ 22.6L',
 '₹ 11.0L',
 '₹ 22.0L',
 '₹ 9.0L',
 '₹ 23.0L',
 '₹ 12.7L',
 '₹ 19.7L',
 '₹ 9.0L',
 '₹ 20.0L',
 '₹ 11.0L',
 '₹ 20.0L',
 '₹ 12.0L',
 '₹ 18.0L',
 '₹ 8.8L',
 '₹ 17.5L']
```

In [581…
```
for i in range(0,len(sal_minmax),2):

    Min_sal.append((sal_minmax[i]))

Min_sal # This is the minimum salary
```

Out[581…
```
['₹ 25.0L',
 '₹ 15.0L',
 '₹ 11.0L',
 '₹ 11.0L',
 '₹ 9.0L',
 '₹ 12.7L',
 '₹ 9.0L',
 '₹ 11.0L',
 '₹ 12.0L',
```

'₹ 8.8L']

In [582…
```
for i in range(1,len(sal_minmax),2):

    Max_sal.append((sal_minmax[i]))

Max_sal # This is the minimum salary # This is maximum salary
```

Out[582…
```
['₹ 45.0L',
 '₹ 26.0L',
 '₹ 22.6L',
 '₹ 22.0L',
 '₹ 23.0L',
 '₹ 19.7L',
 '₹ 20.0L',
 '₹ 20.0L',
 '₹ 18.0L',
 '₹ 17.5L']
```

In [583…
```
# Avg salary
avg_tags =  driver.find_elements(By.XPATH,'//p[@class="averageCtc"]')
for i in avg_tags:
    Avg_sal.append(i.text) # Average Salary
```

In [584…
```
Avg_sal
```

Out[584…
```
['₹ 32.2L',
 '₹ 19.8L',
 '₹ 16.4L',
 '₹ 15.9L',
 '₹ 15.5L',
 '₹ 14.7L',
 '₹ 14.6L',
 '₹ 14.5L',
 '₹ 14.0L',
 '₹ 13.9L']
```

In [586…
```
print(len(EXperience),len(Min_sal),len(Max_sal),len(Avg_sal))
```

```
10 10 10 10
```

In [588…
```
# DataFrame
df = pd.DataFrame({'Experience':EXperience,'Minimum_salary':Min_sal,'Maximum_Salary':Max_sal,'Average_Sal':Avg_sa
df
```

Out[588…

| | Experience | Minimum_salary | Maximum_Salary | Average_Sal |
|---|---|---|---|---|
| 0 | 3-4 yrs experience (based on 24 salaries) | ₹ 25.0L | ₹ 45.0L | ₹ 32.2L |
| 1 | 2-4 yrs experience (based on 59 salaries) | ₹ 15.0L | ₹ 26.0L | ₹ 19.8L |
| 2 | 2-4 yrs experience (based on 49 salaries) | ₹ 11.0L | ₹ 22.6L | ₹ 16.4L |
| 3 | 1-2 yrs experience (based on 35 salaries) | ₹ 11.0L | ₹ 22.0L | ₹ 15.9L |
| 4 | 2-4 yrs experience (based on 118 salaries) | ₹ 9.0L | ₹ 23.0L | ₹ 15.5L |
| 5 | 1 yr experience (based on 10 salaries) | ₹ 12.7L | ₹ 19.7L | ₹ 14.7L |
| 6 | 2-4 yrs experience (based on 70 salaries) | ₹ 9.0L | ₹ 20.0L | ₹ 14.6L |
| 7 | 4 yrs experience (based on 11 salaries) | ₹ 11.0L | ₹ 20.0L | ₹ 14.5L |
| 8 | 4 yrs experience (based on 10 salaries) | ₹ 12.0L | ₹ 18.0L | ₹ 14.0L |
| 9 | 3 yrs experience (based on 14 salaries) | ₹ 8.8L | ₹ 17.5L | ₹ 13.9L |

In [ ]:
```
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js