

WEB SCRAPING ASSIGNMENT-2

Instructions

1. All the questions must be done in a single Jupyter notebook.
2. There should be proper comments in code

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.

Note- All of the above steps have to be done in code. No step is to be done manually.

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, full job-description. 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 search button.
4. Then scrape the data for the first 10 jobs results you get.
5. Finally create a dataframe of the scraped data.

Note- 1. All of the above steps have to be done in code. No step is to be done manually.

WEB SCRAPING ASSIGNMENT-2

2. Please note that you have to scrape full job description. For that you may have to open each job separately as shown below

You will get the search result as shown in the below webpage:

The screenshot shows the Naukri.com website with a search for 'Data Scientist, Machine Learning' in Bangalore. The search results show 1-20 of 4053 jobs. The top result is for Fractal Analytics Ltd, with a rating of 3.8 and 64 reviews. The job title is 'Data Scientist, Machine Learning'. The location is Bangalore. The experience required is 3-7 years. The salary is not disclosed. The job description mentions experience of working on end-to-end data science pipeline: problem scoping, data gathering, supply chain management, Linux, RF, RDBMS, Bfsi, Consulting, Data structures. The job was posted 10 days ago. There is a 'Save' button next to the job listing. On the right, there is a section for 'Get Personalised Job Recommendations' with a 'Register Now' button. Below that, it says 'See 23 jobs in Featured'.

You have to click on each job title. Here in this case if you will click on “Data Scientist, Machine Learning”, you will land up on the below page:

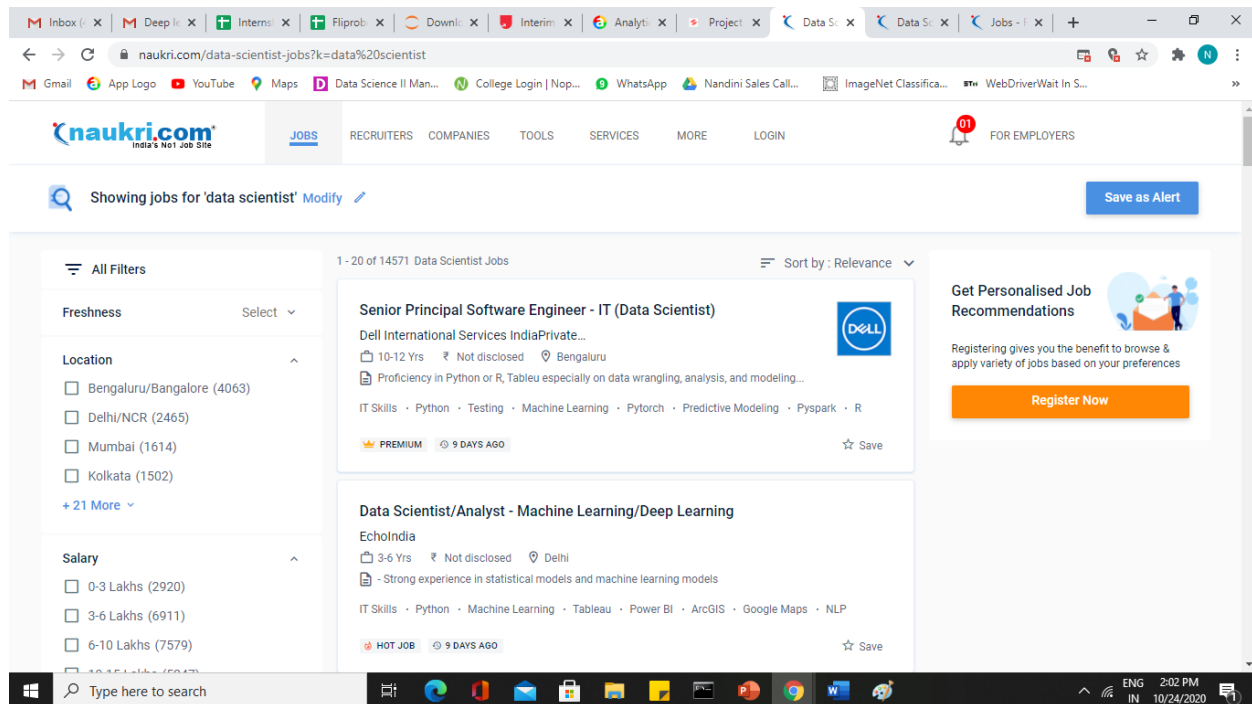
The screenshot shows the detailed job listing for 'Data Scientist, Machine Learning' at Fractal Analytics Ltd. The job title is 'Data Scientist, Machine Learning'. The company name is 'Fractal Analytics Ltd' with a rating of 3.8 and 64 reviews. The location is Bangalore. The experience required is 3-7 years. The salary is not disclosed. The job description mentions experience of working on end-to-end data science pipeline: problem scoping, data gathering, supply chain management, Linux, RF, RDBMS, Bfsi, Consulting, Data structures. The job was posted 11 days ago. There are 10 applicants and 591 job applicants. There are buttons for 'Register to apply' and 'Login to Apply'. Below the job description, there is a section for 'Jobs you might be interested in' with three listings: 'Data Scientist II' at Bundl Technologies Pvt Ltd, 'Advanced Analytics Cons 4' at Wells Fargo India Solutions, and 'Data Scientist, Machine Learning' at Fractal.ai. There is a 'View All' button at the bottom right.

WEB SCRAPING ASSIGNMENT-2

Now you have to scrape data from this page.

Do this for first 10 job search results.

Q3: In this question you have to scrape data using the filters available on the webpage as shown below:



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 the webpage <https://www.naukri.com/>
2. Enter “Data Scientist” in “Skill,Designations,Companies” field .

WEB SCRAPING ASSIGNMENT-2

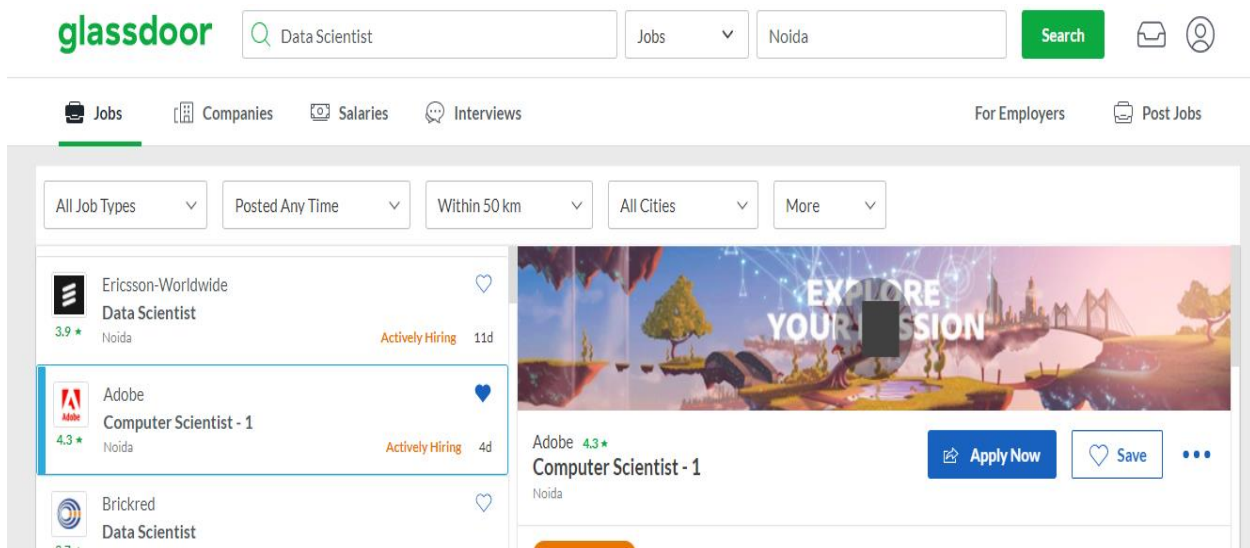
3. Then click the search button.
4. Then apply the location filter and salary filter by checking the respective boxes
4. Then scrape the data for the first 10 jobs results you get.
5. Finally create a dataframe of the scraped data.

Note- All of the above steps have to be done in code. No step is to be done manually.

Q4: Write a python program to scrape data for first 10 job results for Data scientist Designation in Noida location. You have to scrape company_name, No. of days ago when job was posted, Rating of the company.

This task will be done in following steps:

1. first get the webpage <https://www.glassdoor.co.in/index.htm>
2. Enter “Data Scientist” in “Job Title,Keyword,Company” field and enter “Noida” in “location” field.
3. Then click the search button. You will land up in the below page:



4. Then scrape the data for the first 10 jobs results you get in the above shown page.

WEB SCRAPING ASSIGNMENT-2

5. Finally create a dataframe of the scraped data.

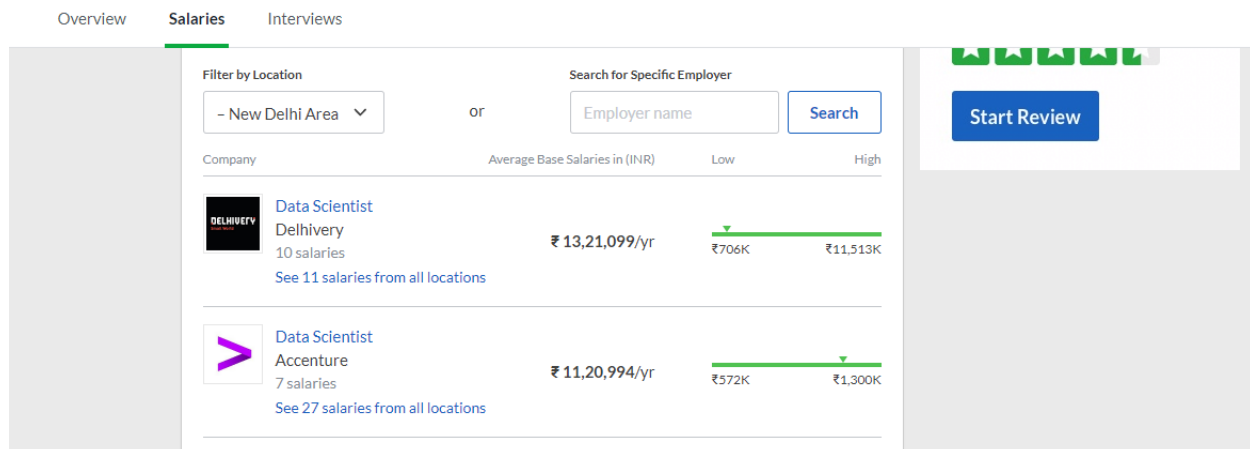
Note- All of the above steps have to be done in code. No step is to be done manually.

Q5: Write a python program to scrape the salary data for Data Scientist designation in Noida location.

You have to scrape Company name, Number of salaries, Average salary, Min salary, Max Salary.

The above task will be, done as shown in the below steps:

1. first get the webpage <https://www.glassdoor.co.in/Salaries/index.htm>
2. Enter “Data Scientist” in Job title field and “Noida” in location field.
3. Click the search button.
4. After that you will land on the below page



You have to scrape whole data from this webpage

5. Scrape data for first 10 companies. Scrape the min salary, max salary, company name, Average salary and rating of the company.

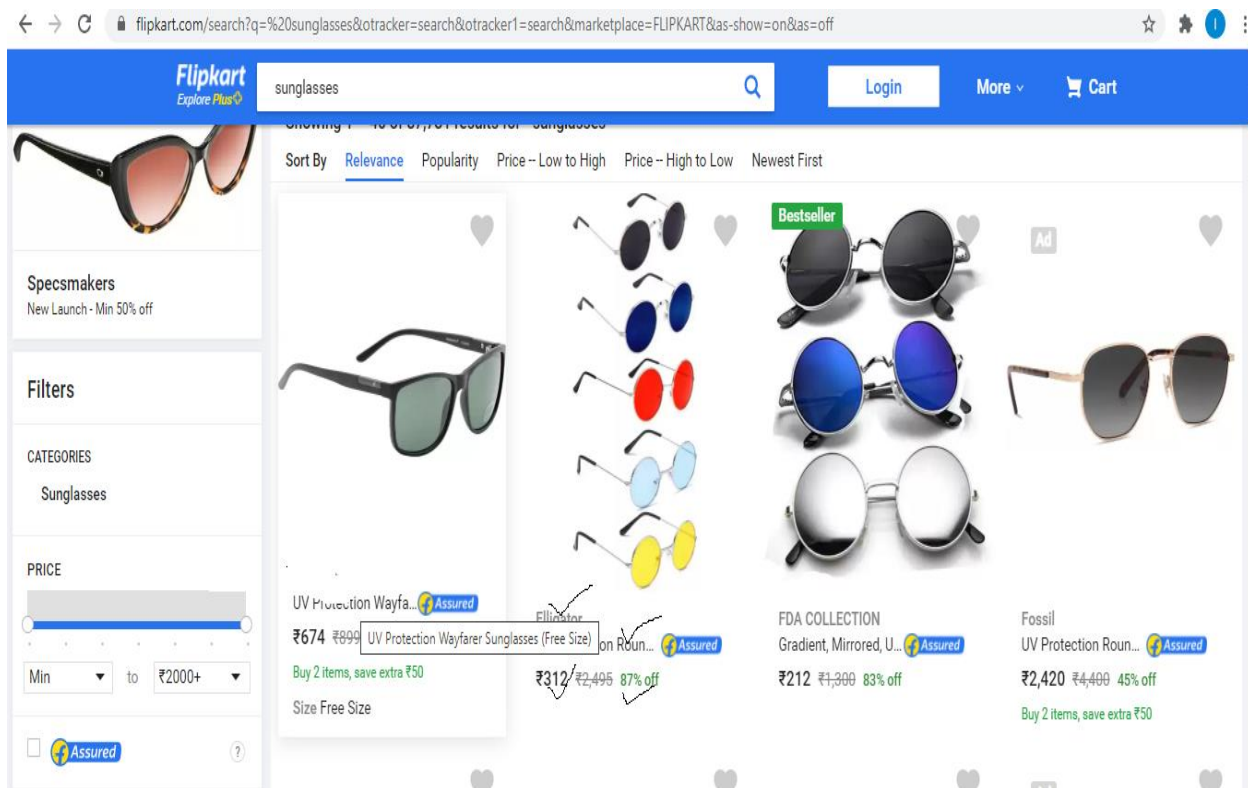
6. Store the data in a dataframe.

Note that all of the above steps have to be done by coding only and not manually.

Q6 : Scrape data of first 100 sunglasses listings on flipkart.com. You have to scrape four attributes:

1. Brand
2. Product Description
3. Price
4. Discount %

The attributes which you have to scrape is ticked marked in the below image.



To scrape the data you have to go through following steps:

1. Go to flipkart webpage by url <https://www.flipkart.com/>
2. Enter “sunglasses” in the search field where “search for products, brands and more” is written and click the search icon
3. after that you will reach to a webpage having a lot of sunglasses. From this page you can scrap the required data as usual.
4. after scraping data from the first page, go to the “Next” Button at the bottom of the page , then click on it

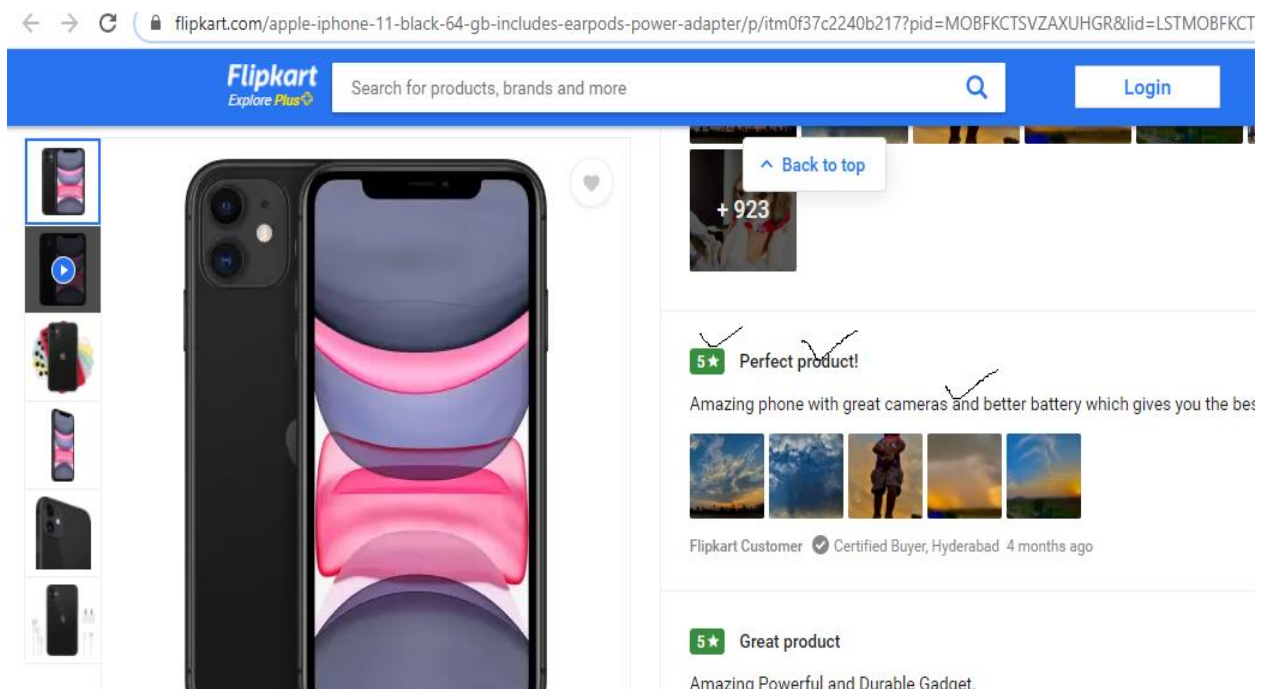
WEB SCRAPING ASSIGNMENT-2

5. Now scrape data from this page as usual
6. repeat this until you get data for 100 sunglasses.

Note that all of the above steps have to be done by coding only and not manually.

Q7: 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-includes-earpods-power-adapter/p/itm0f37c2240b217?pid=MOBFKCTSVZAXUHGR&lid=LSTMOBFKCTSVZAXUHGREPBFGI&marketplace>.

When you will open the above link you will reach to the below shown webpage.



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

These are

1. Rating
2. Review_summary
3. Full review

You have to scrape this data for first 100 reviews.

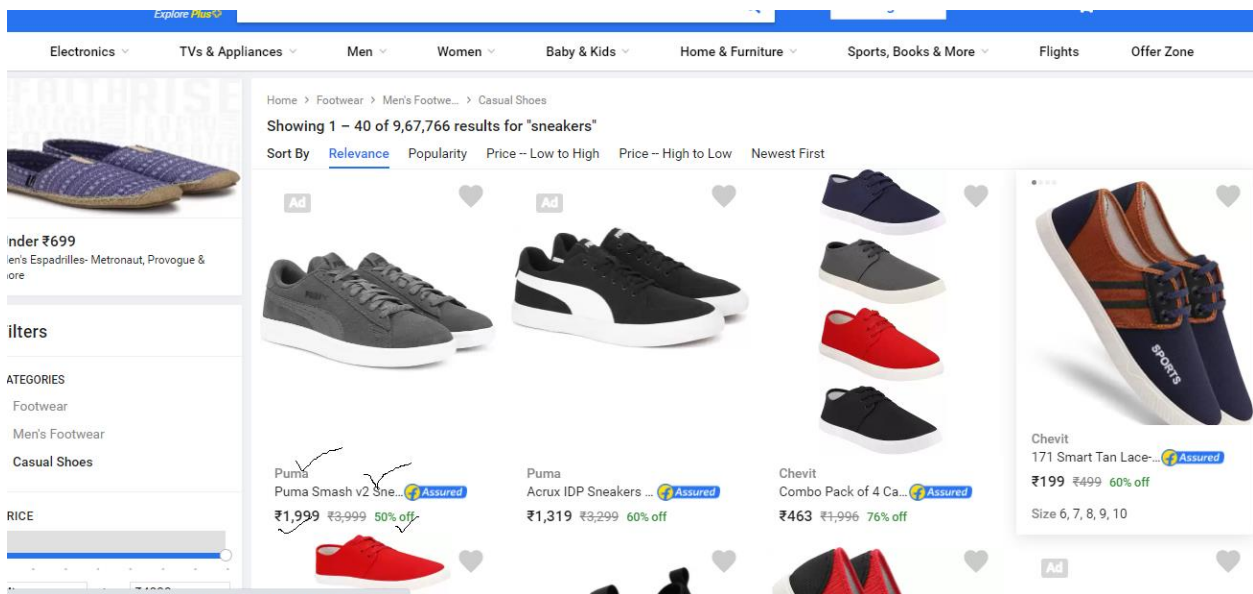
WEB SCRAPING ASSIGNMENT-2

Q8: 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
4. discount %

As shown in the below image, you have to scrape the tick marked attributes.

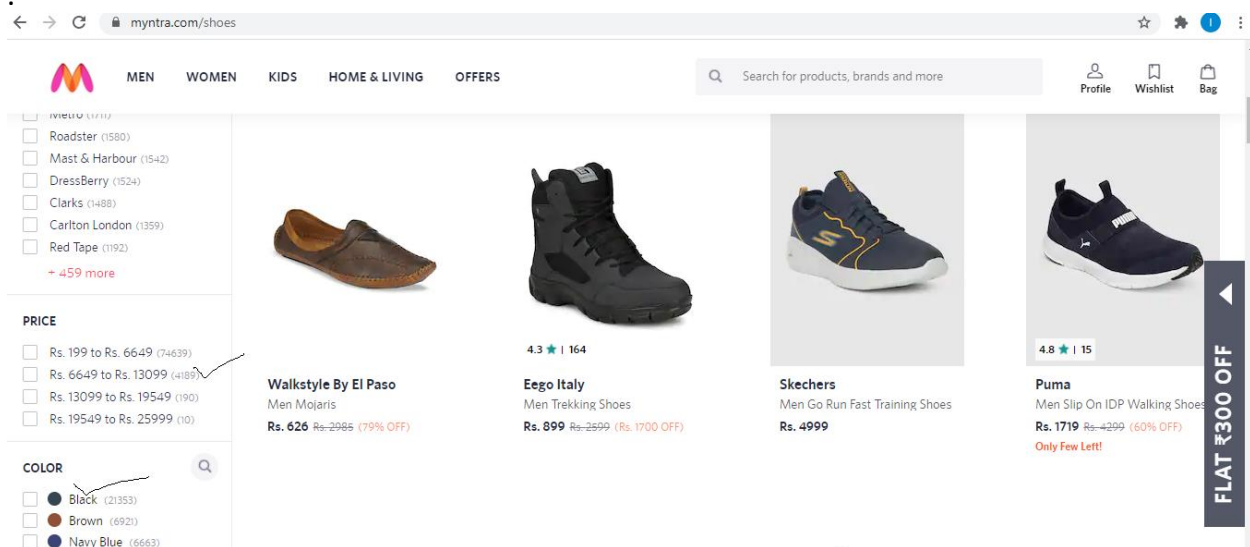


Also note that all the steps required during scraping should be done through code only and not manually.

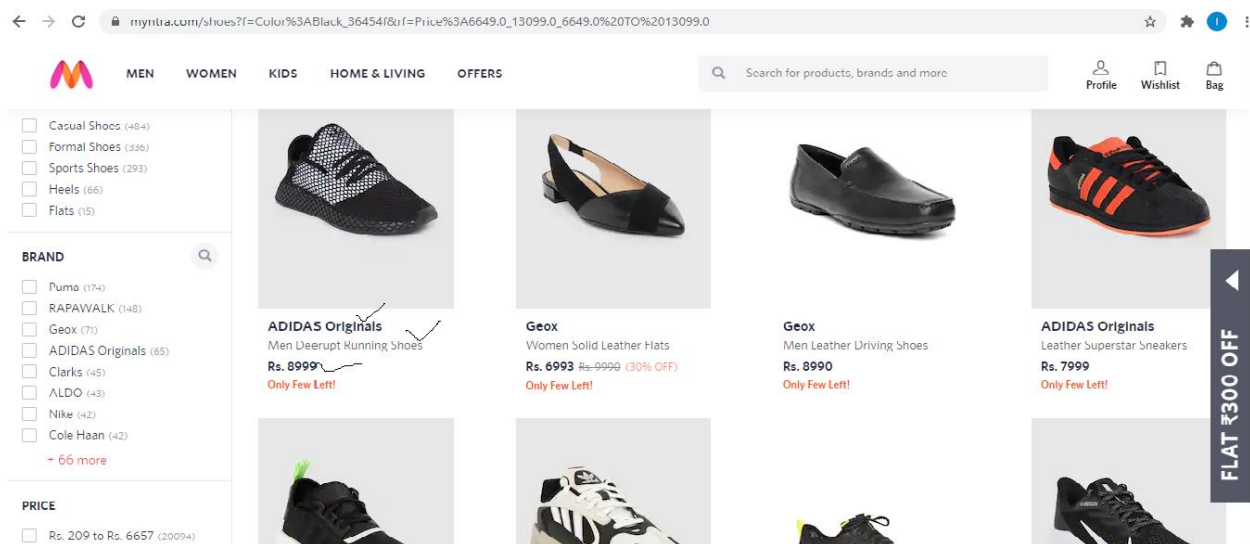
Q9: Go to the link - <https://www.myntra.com/shoes>

Set Price filter to “Rs. 6649 to Rs. 13099” , Color filter to “Black”, as shown in the below image

WEB SCRAPING ASSIGNMENT-2



And then scrape First 100 shoes data you get. The data should include “Brand” of the shoes , Short Shoe description, price of the shoe as shown in the below image.



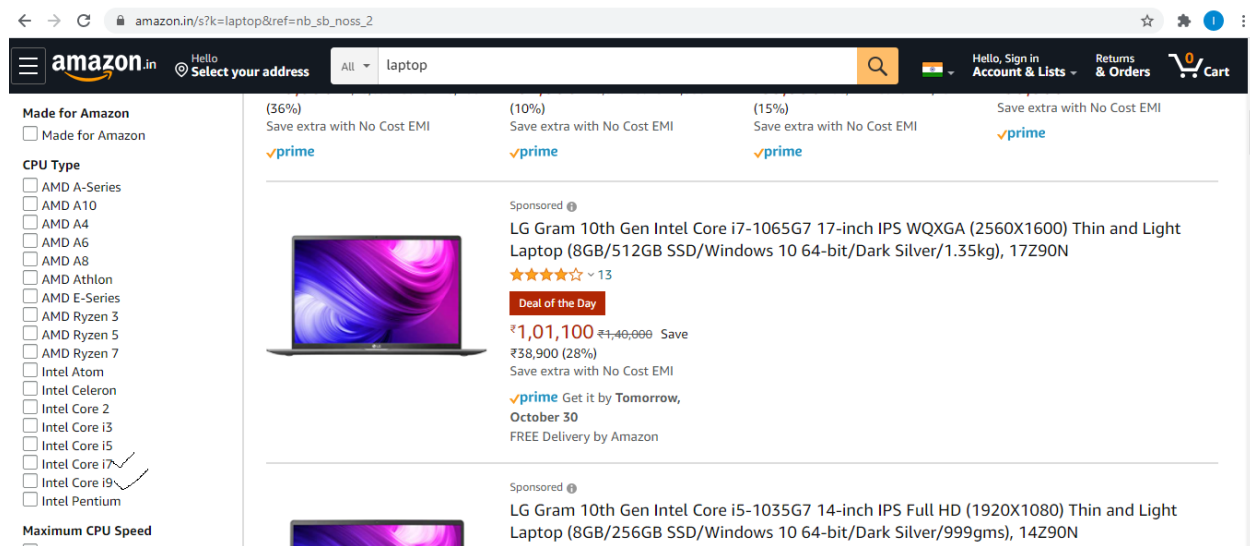
Please note that applying the filter and scraping the data , everything should be done through code only and there should not be any manual step.

Q10: 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” and “Intel Core i9” as shown in the below image:

WEB SCRAPING ASSIGNMENT-2



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

1. title
2. Ratings
3. Price

As shown in the below image as the tick marked attributes.

