

## 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.**

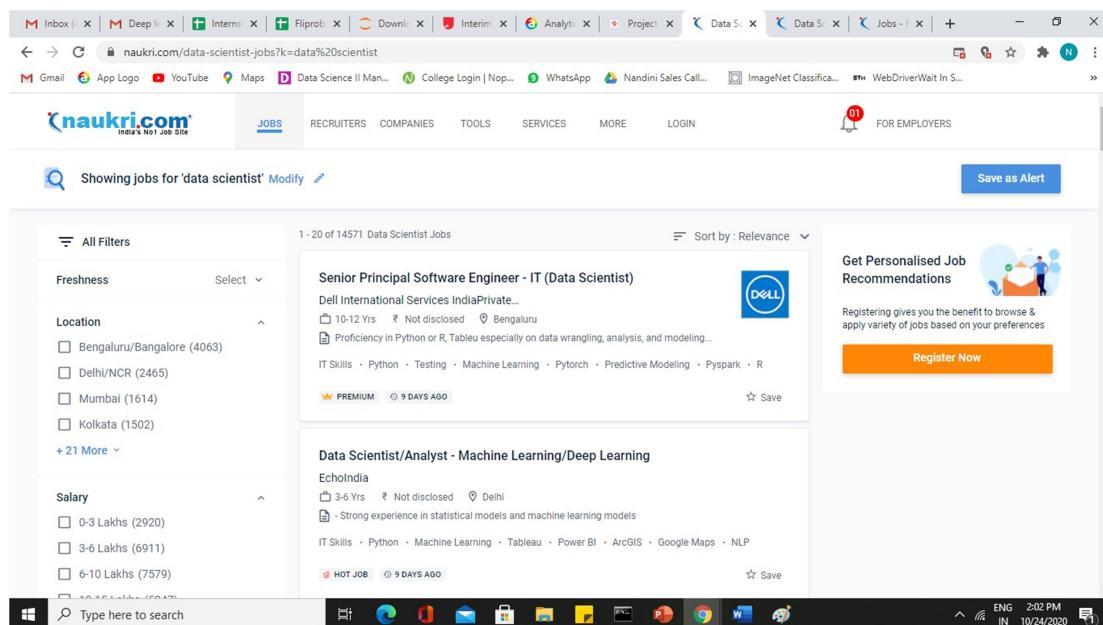
Q2: Write a python program to scrape data for “Data Scientist” Job position in “Bangalore” location. You have to scrape the job-title, job-location, company\_name. You have to scrape first 10 jobs data.

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

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



The screenshot shows a browser window with the naukri.com homepage. The search bar at the top contains the query "Showing jobs for 'data scientist'". Below the search bar, there are several filters on the left side:

- Freshness:** Select dropdown.
- Location:** A dropdown menu showing "Bengaluru/Bangalore (4063)", "Delhi/NCR (2465)", "Mumbai (1614)", "Kolkata (1502)", and "+ 21 More".
- Salary:** A dropdown menu showing "0-3 Lakhs (2920)", "3-6 Lakhs (6911)", and "6-10 Lakhs (7579)".

The main content area displays two job listings:

- Senior Principal Software Engineer - IT (Data Scientist)** at Dell International Services India Private... (3-6 yrs, Not disclosed, Bengaluru). It lists IT Skills: Python, Testing, Machine Learning, Pytorch, Predictive Modeling, Pyspark, and R.
- Data Scientist/Analyst - Machine Learning/Deep Learning** at EchoIndia (3-6 Yrs, Not disclosed, Delhi). It lists IT Skills: Python, Machine Learning, Tableau, Power BI, ArcGIS, Google Maps, and NLP.

On the right side of the page, there is a sidebar with the heading "Get Personalised Job Recommendations" featuring an illustration of a person looking at a screen. Below it is a button labeled "Register Now".

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:

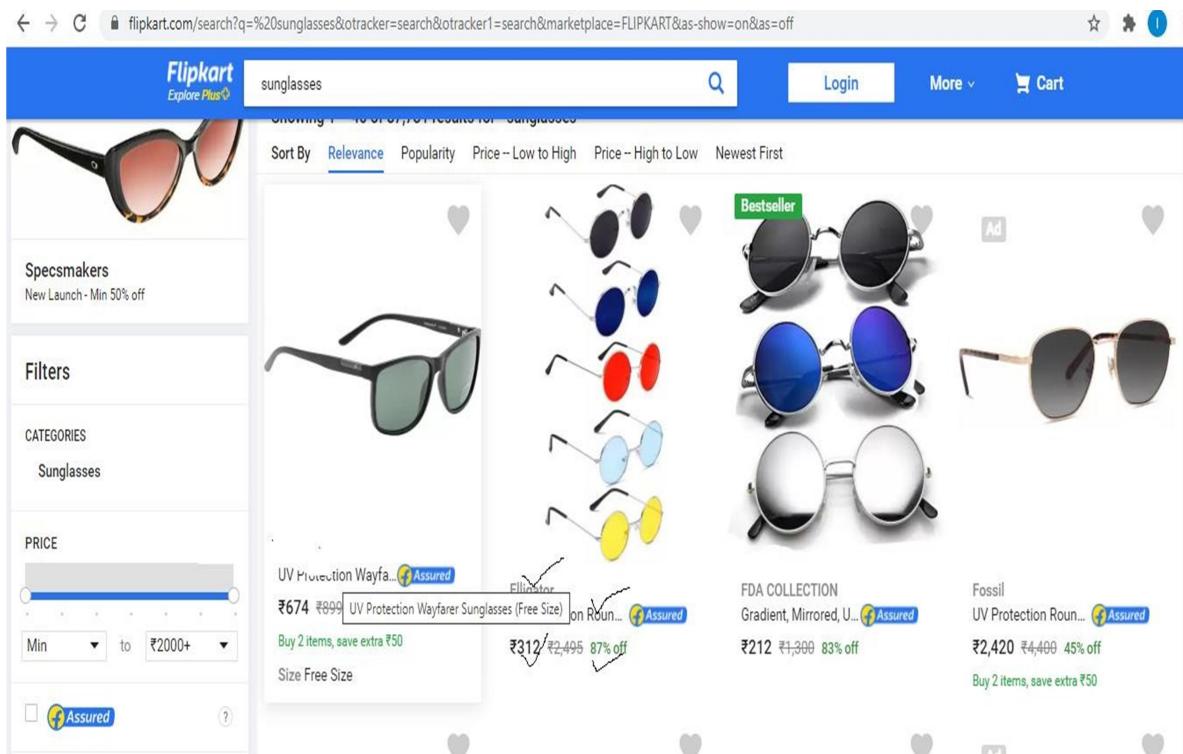
1. first get the webpage <https://www.naukri.com/>
2. Enter “Data Scientist” in “Skill, Designations, and Companies” field.
3. Then click the search button.
4. Then apply the location filter and salary filter by checking the respective boxes
5. Then scrape the data for the first 10 jobs results you get.
6. 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: Scrape data of first 100 sunglasses listings on flipkart.com. You have to scrape four attributes:

1. Brand
2. Product Description
3. Price

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 andmore” is written and click the search icon
3. After that you will reach to the page having a lot of sunglasses. From this page you can scrap the required data as usual.

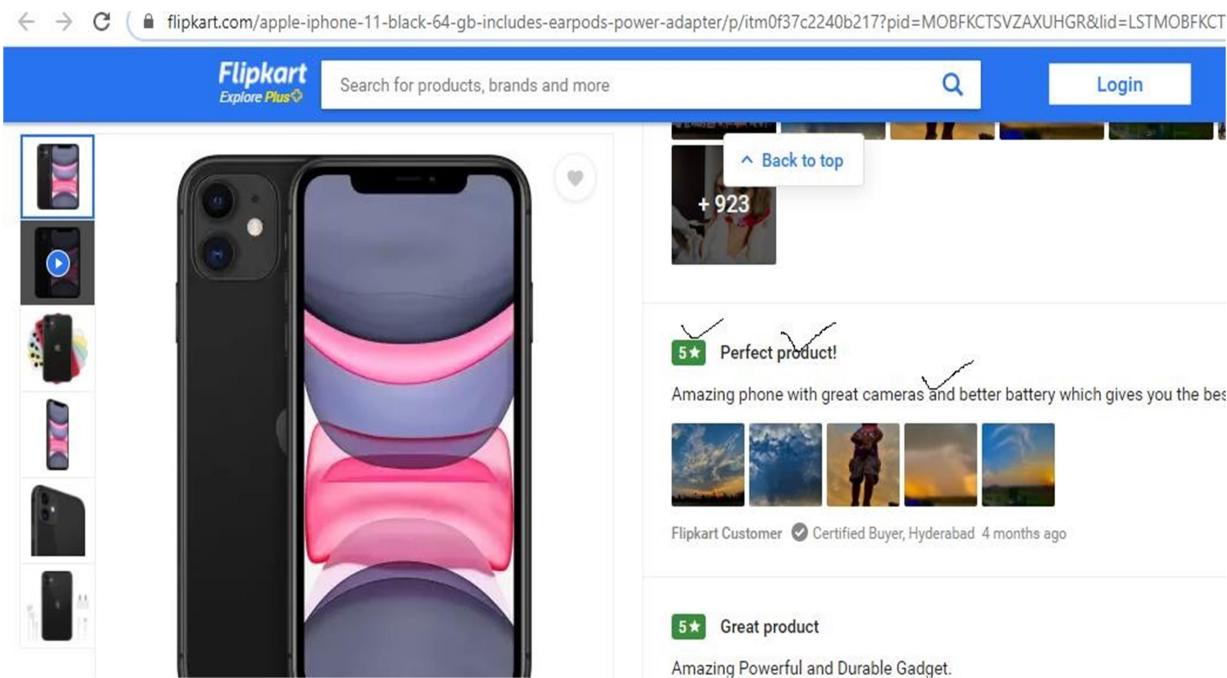
4. After scraping data from the first page, go to the “Next” Button at the bottom of the page , then click on it.
5. Now scrape data from this page as usual
6. Repeat this until you get data for 100 sunglasses.

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

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-includes-earpods-power-adapter/p/itm0f37c2240b217?pid=MOBFKCTSVZAXUHGR&lid=LSTMOBFKCTS梓UHGREGPBFGI&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
4. You have to scrape this data for first 100 reviews.

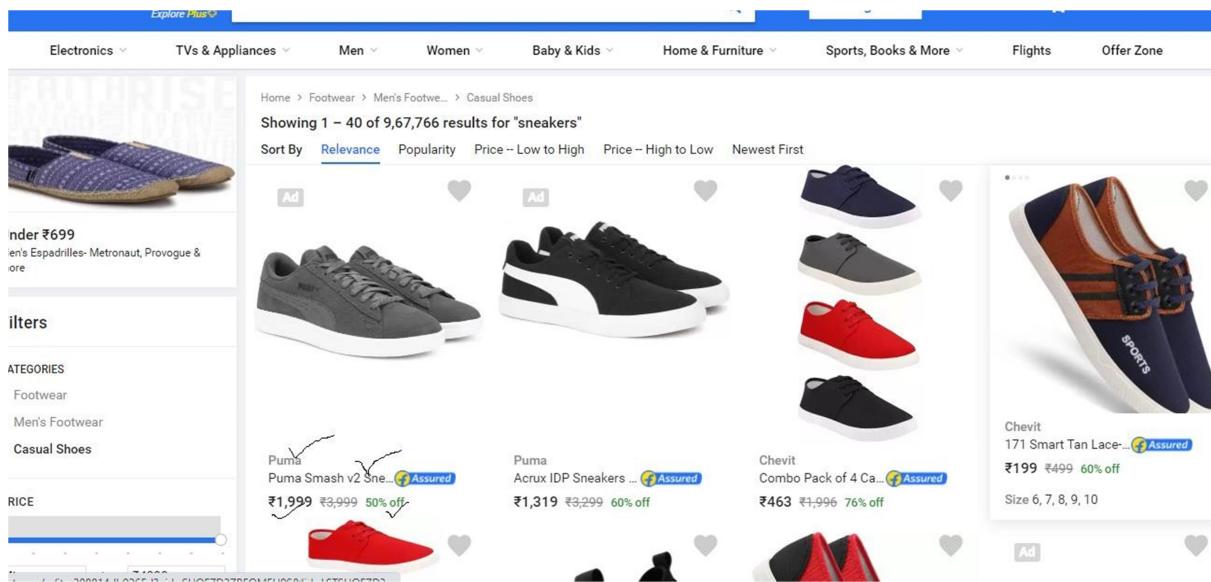
**Note: All the steps required during scraping should be done through code only and not manually.**

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

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



**Showing 1 - 40 of 9,67,766 results for "sneakers"**

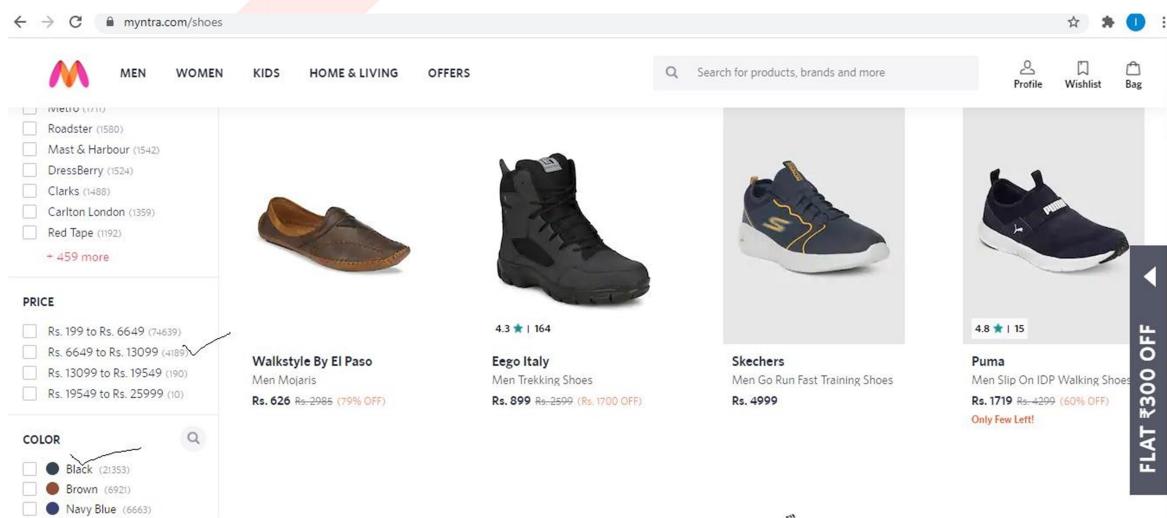
**Sort By:** Relevance, Popularity, Price - Low to High, Price - High to Low, Newest First

Product	Price	Discount
Puma Smash v2 One... Assured	₹1,999	₹3,999 50% off
Puma Acrux IDP Sneakers ... Assured	₹1,319	₹2,299 60% off
Chevit Combo Pack of 4 Ca... Assured	₹463	₹1,996 76% off
Chevit 171 Smart Tan Lace... Assured	₹199	₹499 60% off

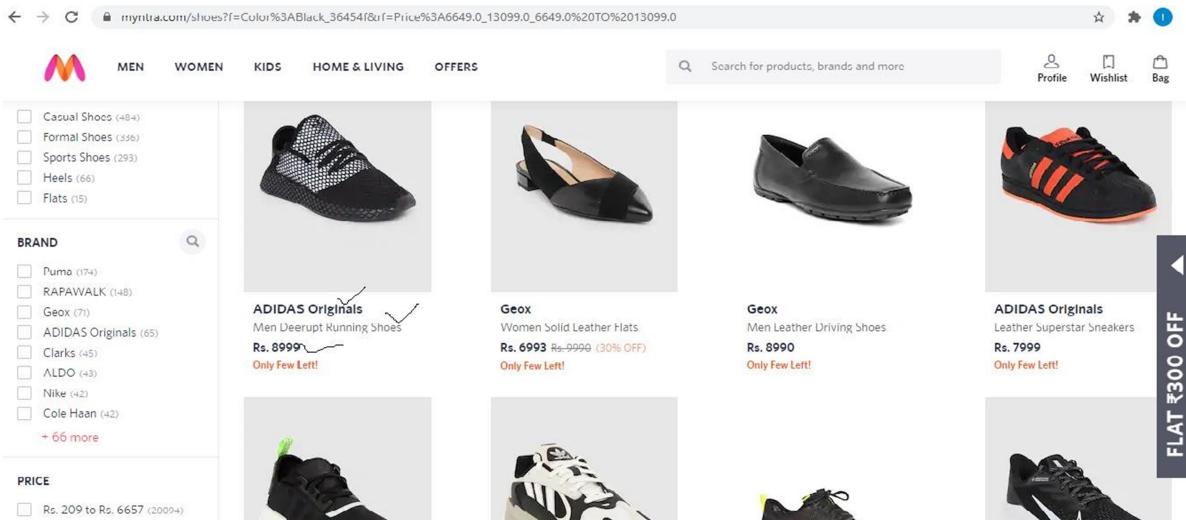
**Note: All the steps required during scraping should be done through code only and not manually.**

Q7: 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.



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.



**Search Results for Men's Shoes:**

- Casual Shoes (484)
- Formal Shoes (556)
- Sports Shoes (293)
- Heels (66)
- Flats (15)

**BRAND**

- Puma (174)
- RAPAWALK (148)
- Geox (71)
- ADIDAS Originals (65)
- Clarks (45)
- ALDO (43)
- Nike (42)
- Cole Haan (42)
- + 66 more

**PRICE**

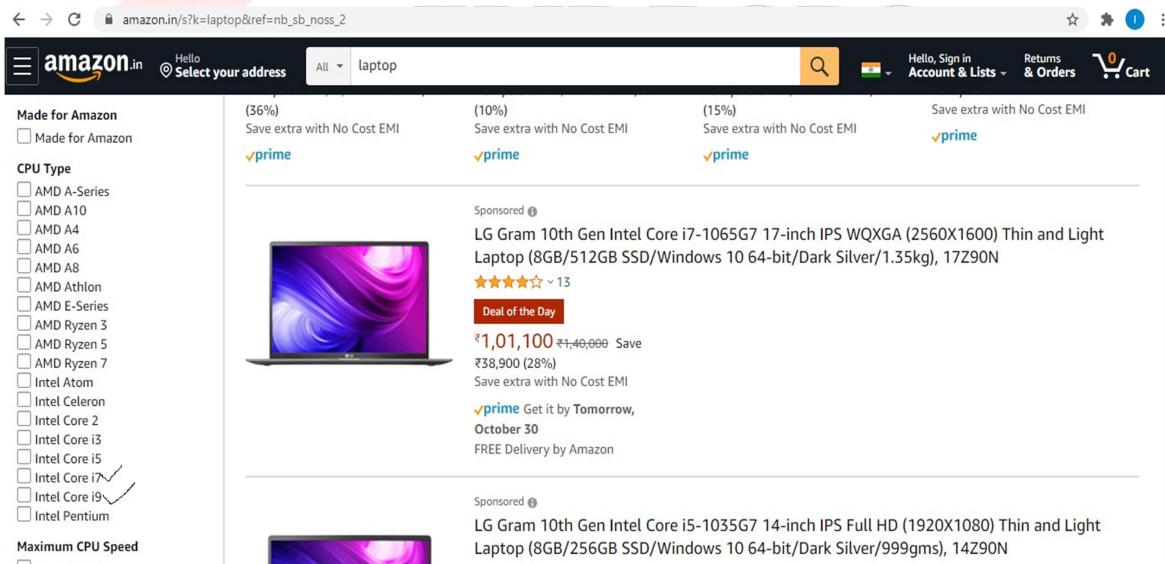
- Rs. 209 to Rs. 6657 (20094)

**Note: Applying the filter and scraping the data, everything should be done through code only and there should not be any manual step.**

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



**Search Results for Laptop:**

Made for Amazon

CPU Type

- AMD A-Series
- AMD A10
- AMD A4
- AMD A6
- AMD A8
- AMD Athlon
- AMD E-Series
- AMD Ryzen 3
- AMD Ryzen 5
- AMD Ryzen 7
- Intel Atom
- Intel Celeron
- Intel Core 2
- Intel Core i3
- Intel Core i5
- Intel Core i7✓
- Intel Core i9✓
- Intel Pentium

Maximum CPU Speed

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.

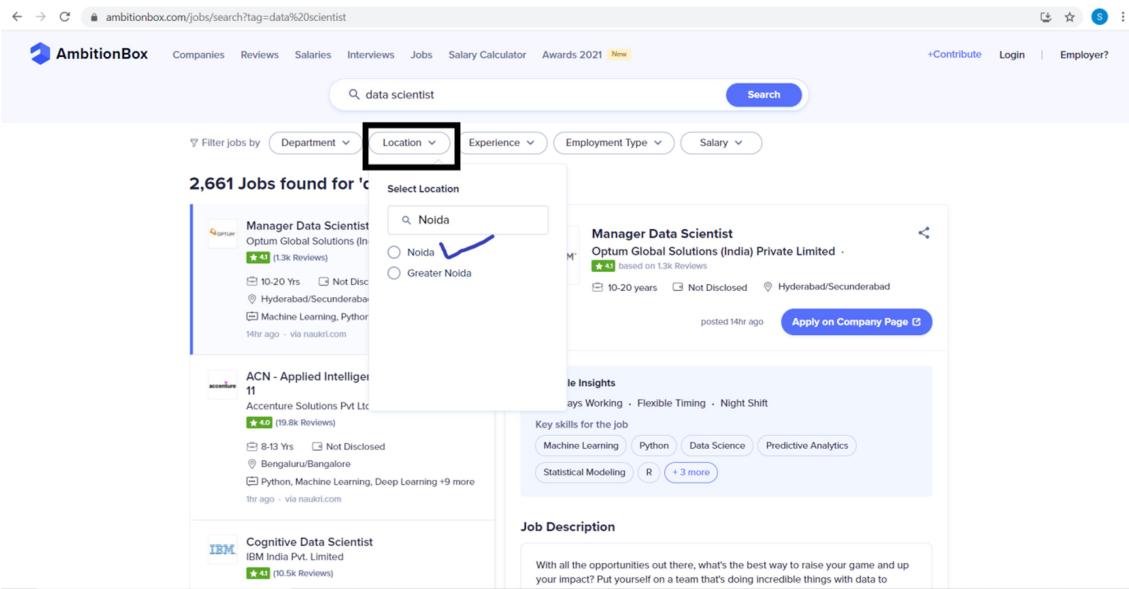
**Note:** All the steps required during scraping should be done through code only and not manually.

**Q9:** 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.ambitionbox.com/>
2. Click on the Job option as shown in the image

3. After reaching to the next webpage, In place of “Search by Designations, Companies, Skills” enter “Data Scientist” and click on search button.

4. You will reach to the following web page click on location and in place of “Search location” enter “Noida” and select location “Noida”.



The screenshot shows a search results page for 'data scientist' on AmbitionBox. A dropdown menu for 'Location' is open, with 'Noida' selected. The results show 2,661 jobs found. One job listing is highlighted: 'Manager Data Scientist' at Optum Global Solutions (India) Private Limited, posted 14hr ago. The job description includes insights like 'ays Working - Flexible Timing - Night Shift' and key skills like 'Machine Learning', 'Python', 'Data Science', 'Predictive Analytics', 'Statistical Modeling', and 'R'. Other companies listed include ACN - Applied Intelligence, Accenture Solutions Pvt Ltd, and IBM.

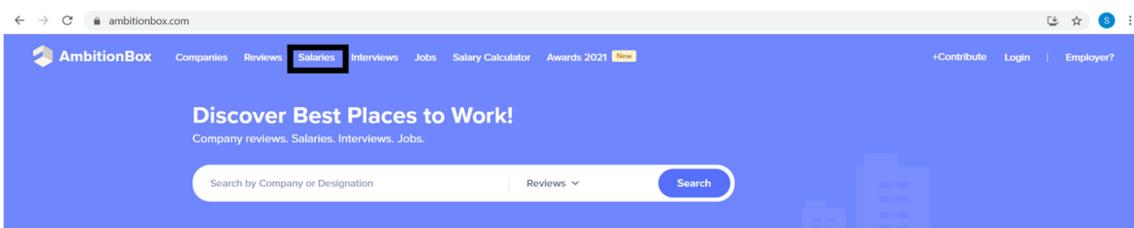
5. Then scrape the data for the first 10 jobs results you get on the above shown page.  
 6. Finally create a data frame of the scraped data.

**Note:** All the steps required during scraping should be done through code only and not manually.

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.

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

1. First get the webpage <https://www.ambitionbox.com/>
2. Click on the salaries option as shown in the image.



The screenshot shows the AmbitionBox homepage with the 'Salaries' tab selected. A banner says 'Discover Best Places to Work!' with subtext 'Company reviews. Salaries. Interviews. Jobs.' Below it is a search bar for 'Search by Company or Designation' and a 'Search' button. The main content area features sections for 'Company reviews' and 'Company salaries'.

#### Company reviews

Real employees share how it is to work at their company. Read reviews and make informed decisions.



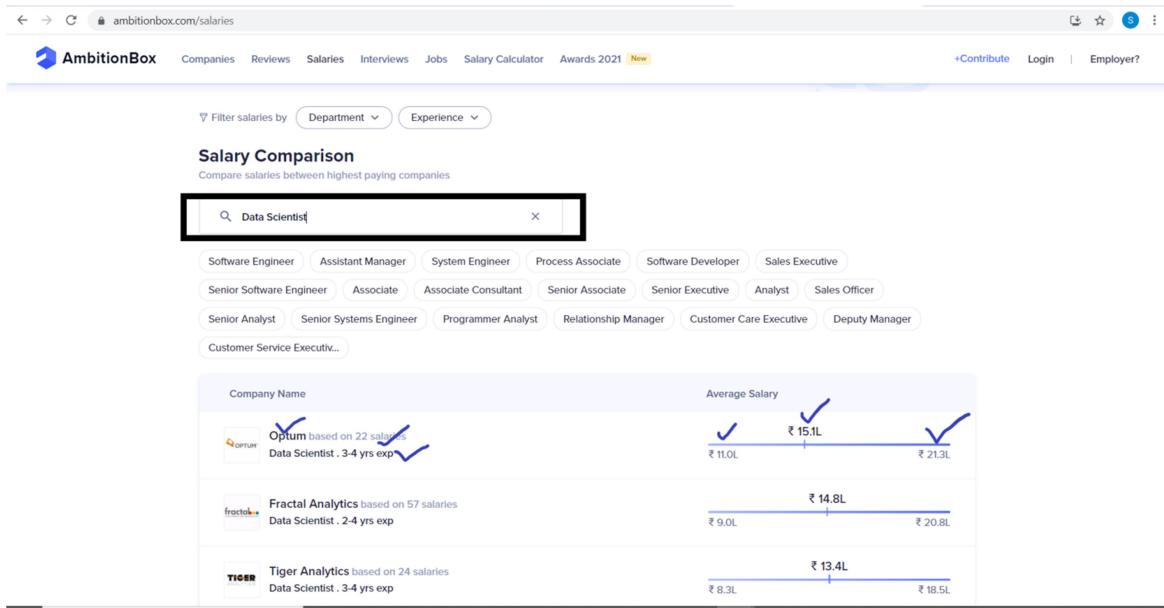
Three review snippets are shown for Amazon, Capgemini, and Reliance Jio. Each snippet includes a profile picture, the employee's name, their current role, and a short summary of their likes and dislikes. Below each snippet is a star rating and a 'View Detailed' link.

#### Company salaries

Search and compare average salaries by company, job title, and experience.

Catalogue Specialist      Team Lead      PPA Developer

3. After reaching to the following webpage, In place of “Search Job Profile” enters “Data Scientist” and then click on “Data Scientist”.



You have to scrape the data ticked in the above image.

4. Scrape the data for the first 10 companies. Scrape the company name, total salary record, average salary, minimum salary, maximum salary, experience required.
5. Store the data in a dataframe.

**Note:** All the steps required during scraping should be done through code only and not manually.