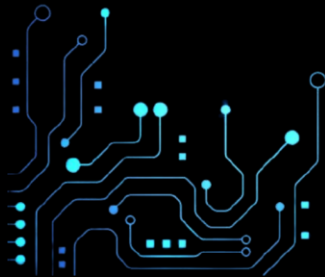


# SQL Job Hunting

**D-LLOYD AGENCIES**



<https://www.linkedin.com/jobs/>

The screenshot displays the LinkedIn Jobs page. At the top, there is a navigation bar with the LinkedIn logo, a search bar labeled "Search by title, skill, or company", and navigation links for Home, My Network, Jobs (which is the active tab), Messaging, and Notifications. On the left side, a sidebar contains links to "My jobs", "Job alerts", "Demonstrate skills", "Interview prep", and "Resume Builder". The main content area is titled "Recent job searches" and includes a "Clear" button. It lists three search entries: "data specialist" with 1,148 new results, "water data science", and another "water data science" search with 25 new results. A "Show more" link is located at the bottom of the list.

in Search by title, skill, or company

Home My Network **Jobs** Messaging Notifications

My jobs

Job alerts

✓ Demonstrate skills

Interview prep

Resume Builder

Recent job searches Clear

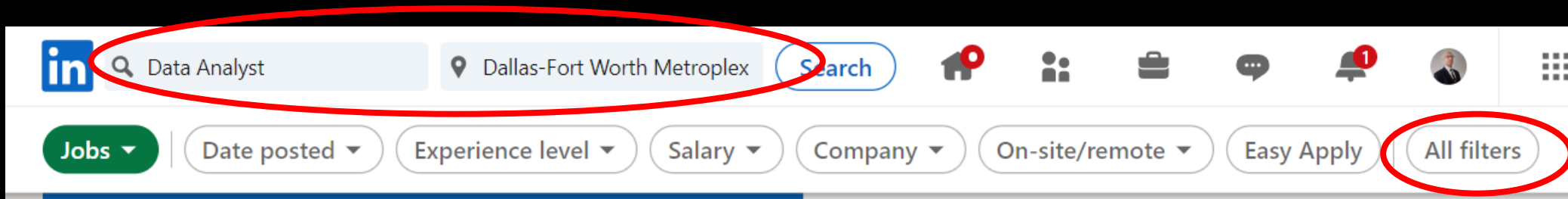
data specialist · 1,148 new  
Alert On · United States · Remote

water data science  
United States · Remote

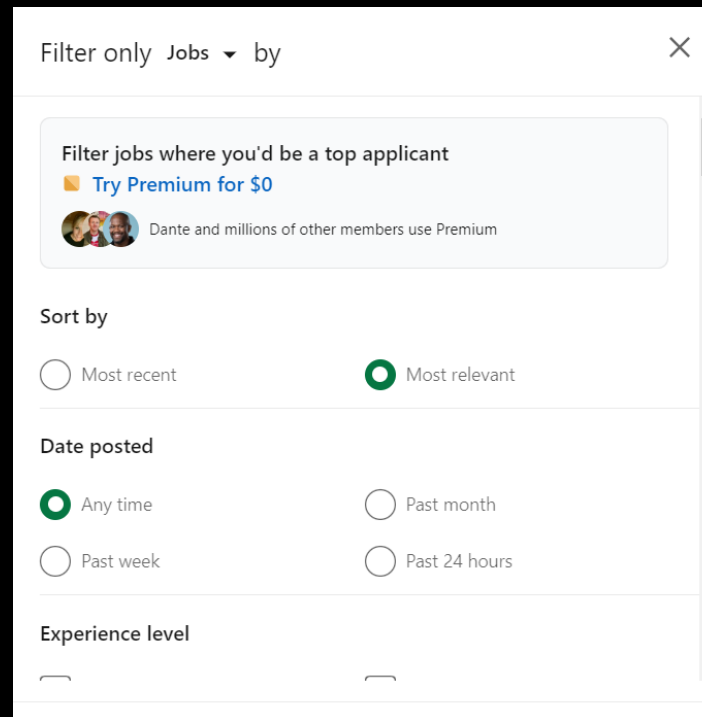
water data science · 25 new  
United States

Show more

# Lookup Your Career, Location, and Filter the Results



The image shows the LinkedIn search interface. The search bar is highlighted with a red oval and contains the text "Data Analyst" and "Dallas-Fort Worth Metroplex". The "Search" button is also highlighted. Below the search bar, there are several filter buttons: "Jobs", "Date posted", "Experience level", "Salary", "Company", "On-site/remote", "Easy Apply", and "All filters". The "All filters" button is highlighted with a red oval. A red arrow points from the "All filters" button to the filter panel below.



The image shows the LinkedIn filter panel. It is titled "Filter only Jobs by" and has a close button (X) in the top right corner. The panel contains several sections:

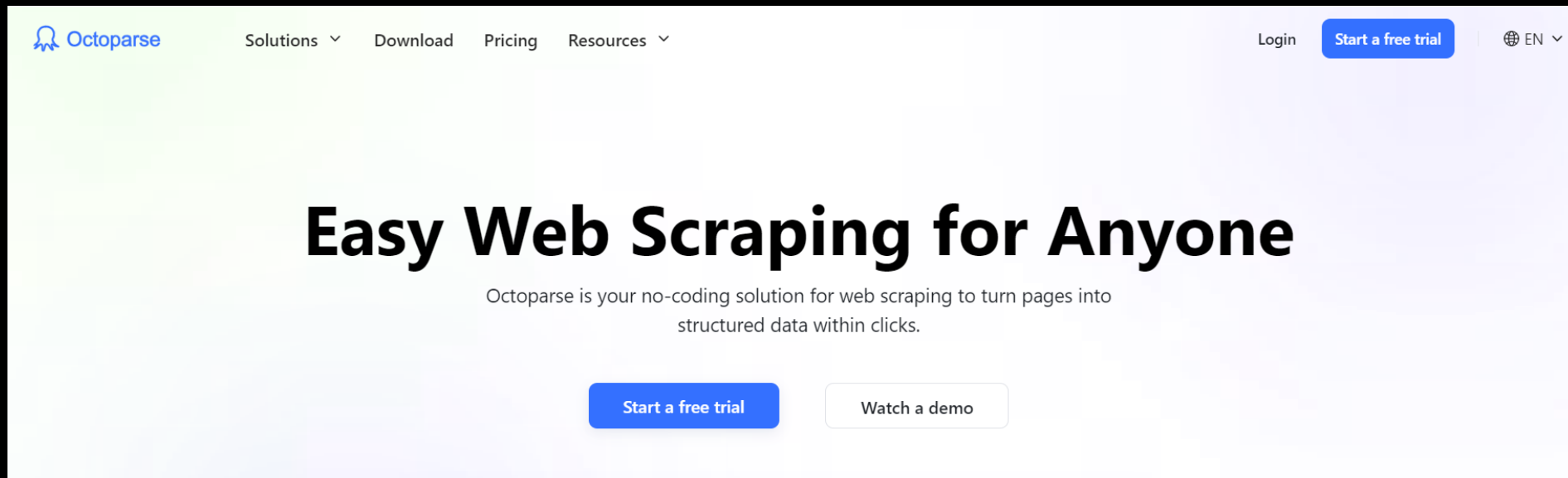
- Filter jobs where you'd be a top applicant**: A section with a "Try Premium for \$0" button and a note that "Dante and millions of other members use Premium".
- Sort by**: Two radio buttons, "Most recent" and "Most relevant". The "Most relevant" button is selected.
- Date posted**: Four radio buttons, "Any time", "Past month", "Past week", and "Past 24 hours". The "Any time" button is selected.
- Experience level**: A section with a dropdown menu.

# Filter Tips

- Aim for more recent posts, I never filter past the past week
- On-site or hybrid work styles are less competitive for getting your foot in the door
- Networking and recruiters are ideal ways to get hired, but webscraping works for career exploration

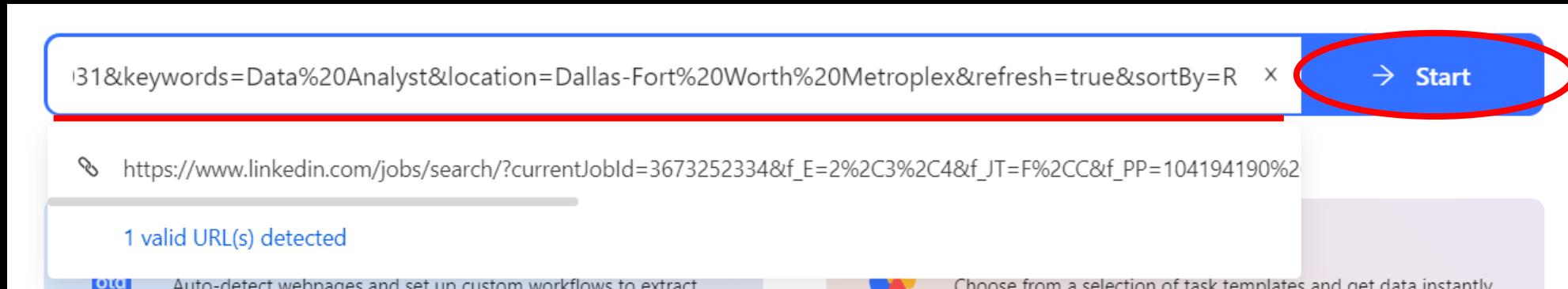
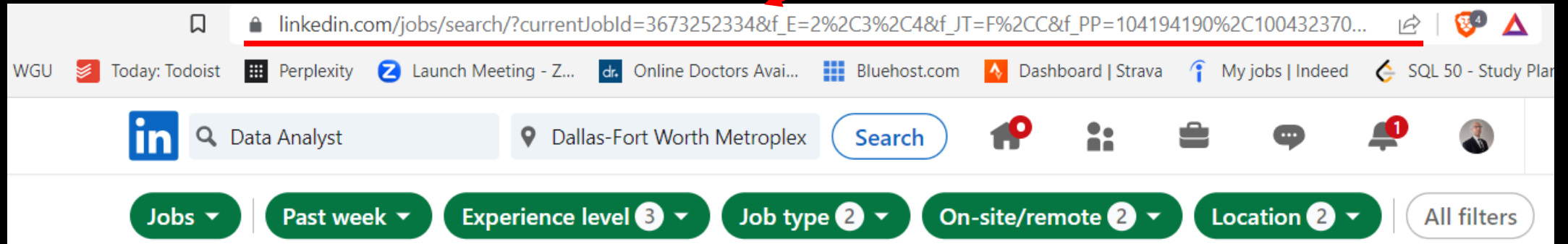
# Octoparse

- <https://www.octoparse.com/>
- A free Web Scraping / Web Crawling Software that runs on the local device (PC).
- Download and watch their tutorial to run web scraping locally



# Web Scrapping: Octoparse

- Copy you're the URL from your search results and paste it into Octoparse



# Web Scrapping Octoparse

- Set up the scraper (with infinite scrolling) and run the extraction

The screenshot shows the Octoparse web scraper interface. The main window displays a job listing page for "97 Data Analyst jobs in Dallas-Fort Worth Metroplex". A notification banner at the top left says "Get notified about new Data Analyst jobs in Dallas-Fort Worth Metroplex. Sign in to create job alert". A blue modal box in the center says "Confirm Detected Results" with the text "Switch to your desired results and click Create workflow to generate a workflow." and a "Got it" button. A configuration modal on the right is open, showing options for "Extract the data in the list", "Add a page scroll", and "Click on a 'Load More' button". The "Run" button in the top right corner is circled in red.

**Data Preview** 8 data fields pre-selected, total 97 lines of data. Preview the first [10] line(s). [Preview more data](#)

Data Fields	No.	Title	Image	basecard_fullll...	hiddennestedl...	hiddennestedl...	Location	resultbenefits_...	Date	Actions
Page1	1	Production Support Da...	https://media.licdn.co...	https://www.linkedin.co...	https://bm.linkedin.co...	Sompo International	Dallas, TX	Actively Hiring	6 days ago	
	2	Data Engineer	https://media.licdn.co...	https://www.linkedin.co...	https://www.linkedin.co...	ScaleneWorks INC	Dallas, TX	Be an early applicant	5 days ago	
	3	Data ops Analyst-.Net	https://media.licdn.co...	https://www.linkedin.co...	https://www.linkedin.co...	ScaleneWorks INC	Dallas, TX	Be an early applicant	5 days ago	

# Web Scrapping Octoparse

- Wait for the extraction before exporting the data

97 Data Analyst jobs in Dallas-Fort Worth Metroplex (5 new) [Show Browser](#)

**869**  
Data Extracted

**Running**  
[Extract Data] Data successfully extracted  
Duplicates: 709 line(s) Time Spent: 3m 24s Avg. Speed: 250 lines/min

**Pause** **Stop**

Task Overview **Data List** Event Log Recent Runs

#	Title	Image	basecard_fulllink_URL	hiddennestedlink_U...	hiddennestedlink	Location	resultbenefits_text	Date
1	Senior Software Engi...	https://media.licdn.c...	https://www.linkedin...	https://www.linkedi...	Wells Fargo	Dallas, TX	Be an early applicant	4 days ago
2	iOS/Android Mobile ...		https://www.linkedin...	https://www.linkedi...	Dice	Dallas, TX	Be an early applicant	5 days ago
3	ServiceNow Senior L...		https://www.linkedin...	https://www.linkedi...	Kyndryl	Dallas, TX	Be an early applicant	1 day ago
4	Android/iOS Develo...		https://www.linkedin...	https://www.linkedi...	Dice	Dallas, TX	Be an early applicant	5 days ago
5	.NET Full Stack/Seni...		https://www.linkedin...	https://www.linkedi...	Dice	Dallas, TX	Be an early applicant	5 days ago
6	Reserving Actuary/R...	https://media.licdn.c...	https://www.linkedin...	https://bm.linkedin.c...	Sompo International	Dallas, TX	Be an early applicant	6 days ago
7	Senior/Lead Service...	https://media.licdn.c...	https://www.linkedin...	https://www.linkedi...	Dice	Dallas, TX	Be an early applicant	4 days ago
8	.NET/Senior Softwar...	https://media.licdn.c...	https://www.linkedin...	https://www.linkedi...	Dice	Dallas, TX	Be an early applicant	5 days ago
9	Senior Mobile Devel...	https://media.licdn.c...	https://www.linkedin...	https://www.linkedi...	Motion Recruitment	Dallas, TX	Be an early applicant	1 day ago

< 1 ... 40 41 42 43 44 > Go to Page



# Web Scraping Octoparse

- Export the web scraping run when complete
- *Note: Example run stopped because I was pulling repeated duplicates*

The screenshot displays the Octoparse web scraping interface. At the top, a task titled "97 Data Analyst jobs in Dallas-Fort Worth Metroplex (5 new)" is shown as "Stopped". A large orange circle highlights the number "873" under "Data Extracted". Below this, it states "Duplicates: 714 line(s)", "Time Spent: 3m 41s", and "Avg. Speed: 237 lines/min". Buttons for "Re-run" and "Export" are visible. A modal dialog titled "Run Stopped" is open in the center, displaying the task name, time spent, and data extracted. It includes a "Rate your experience?" link and two buttons: "Export Later" and "Export". The "Export" button is circled in red. In the background, a table lists job titles and details, and a pagination bar at the bottom shows "Go to Page 44".

97 Data Analyst jobs in Dallas-Fort Worth Metroplex (5 new)

873 Data Extracted

Stopped

Task stopped

Duplicates: 714 line(s) Time Spent: 3m 41s Avg. Speed: 237 lines/min

Re-run Export

Task Overview Data List Event Log Re-run

# Title Image

3 ServiceNow Senior L...

4 Android/iOS Develo...

5 .NET Full Stack/Seni...

6 Reserving Actuary/R...

7 Senior/Lead Service...

8 .NET/Senior Softwar...

9 Senior Mobile Devel...

10 Senior Mobile Devel...

11 Senior Software Engi...

12 Senior iOS Developer

13 Tax Services Senior ...

resultbenefits\_text Date

Be an early applicant 1 day ago

Be an early applicant 5 days ago

Be an early applicant 5 days ago

Be an early applicant 6 days ago

Be an early applicant 4 days ago

Be an early applicant 5 days ago

Be an early applicant 1 day ago

Be an early applicant 5 days ago

Be an early applicant 5 days ago

Be an early applicant 4 days ago

Be an early applicant 1 day ago

Run Stopped

Task: 97 Data Analyst jobs in Dallas-Fort Worth Metroplex (5 new)

Time Spent: 3m 41s

Data extracted: 873 lines 714 duplicates

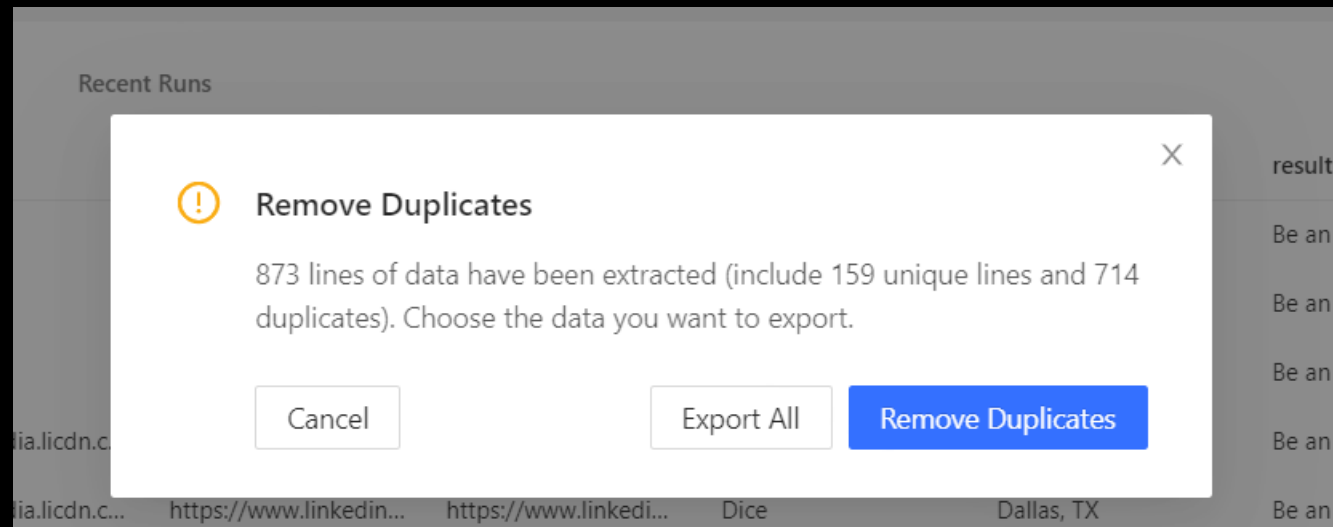
Rate your experience?

Export Later Export

< 1 ... 40 41 42 43 44 > Go to Page

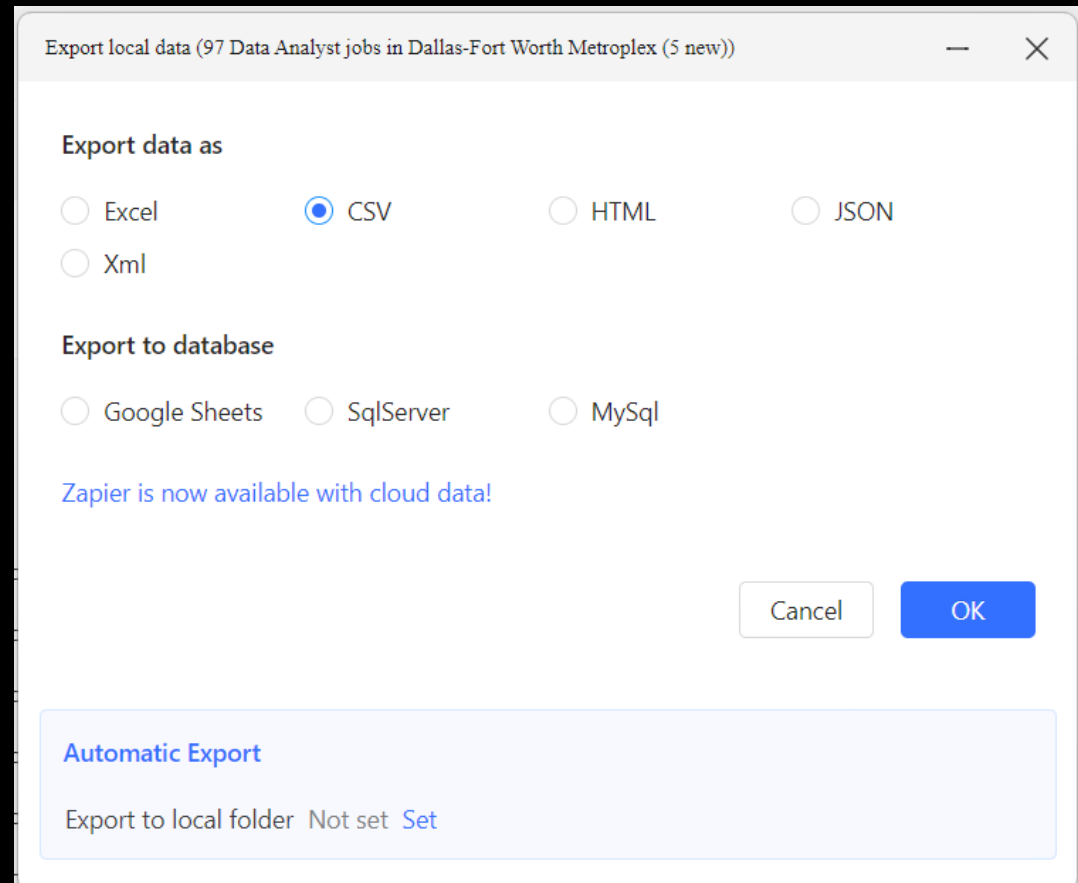
# Web Scrapping Octoparse

- Remove Duplicates if present



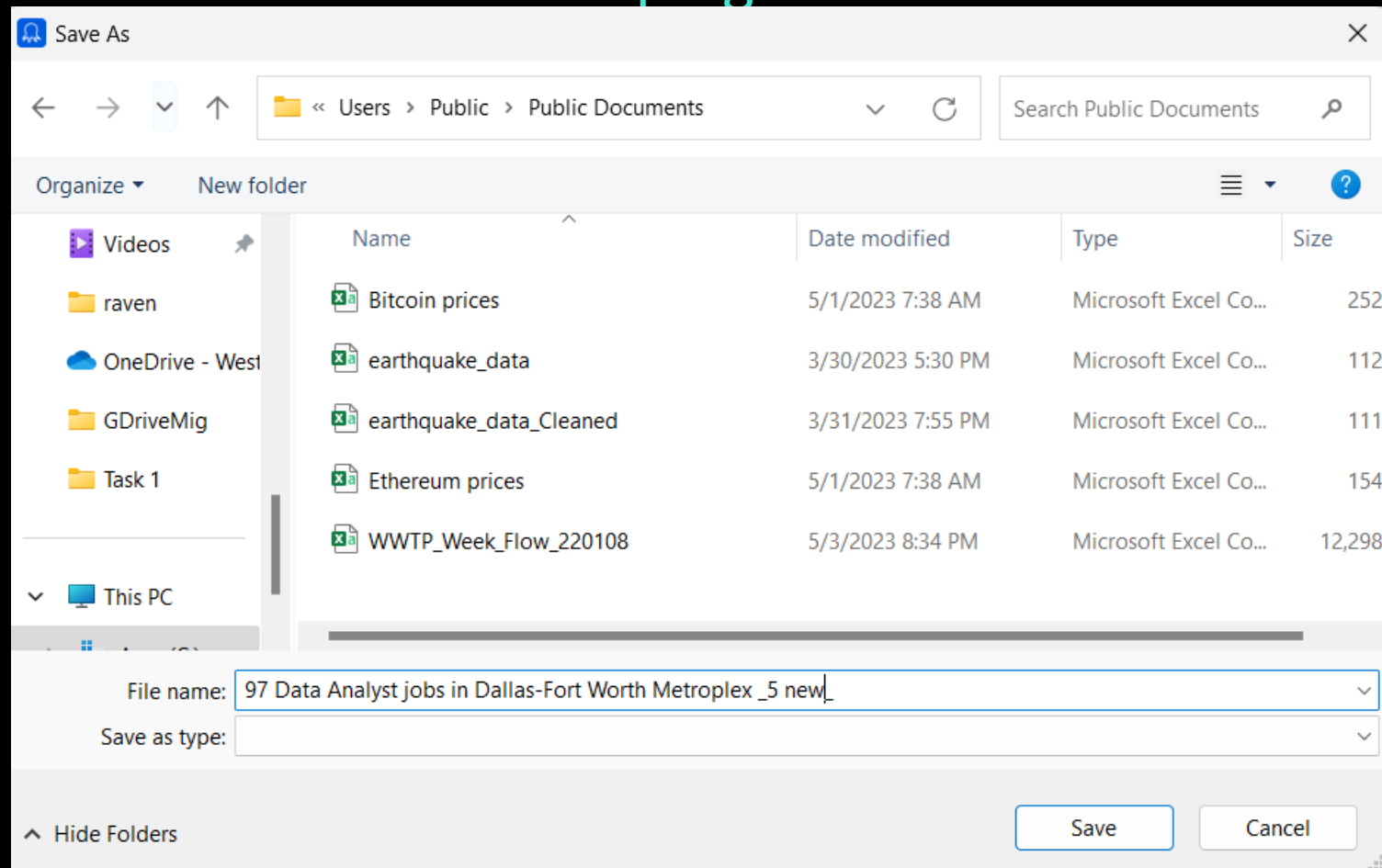
# Web Scrapping Octoparse

- Save as desired data type



# Web Scrapping Octoparse

- Save the CSV file from Web Scrapping



# Update Your Data to SQL

- Clean data headers in a way that makes sense to you (Use Excel or ALTER TABLE).
- No header cleaning took place in the example below

```
2  CREATE TABLE JobSearch(  
3      Title varchar,  
4      Image varchar,  
5      basecard_fulllink_URL varchar,  
6      hiddennestedlink_URL varchar,  
7      hiddennestedlink varchar,  
8      Location varchar,  
9      resultbenefits_text varchar,  
10     Date varchar  
11 );  
12 COPY JobSearch(Title, Image, basecard_fulllink_URL, hiddennestedlink_URL, hiddennestedlink, Location, resultbenefits_text, Date)  
13 FROM 'C:/Users/Public/Documents/97 Data Analyst jobs in Dallas-Fort Worth Metroplex _5 new_.csv' CSV HEADER; 18ms
```

# Query Away!

```
15 SELECT *
16 FROM jobsearch
17 WHERE resultbenefits_text LIKE '%Actively Hiring%'; 3ms
```

jobsearch									
Search results									
		title	image	basecard_fulllink_url	hiddennestedlink_url	hiddennestedlink	location	resultb	
	1	Production Su	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://bm.linkedin.com/co	Sompo International	Dallas, TX	Ac	
	2	Support Analy	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://www.linkedin.com/c	Costco Wholesale	Dallas, TX	Ac	
	3	Data Governar	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://www.linkedin.com/c	Apex Systems	Dallas, TX	Ac	
	4	Senior Data Er	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://www.linkedin.com/c	CoreLogic	Dallas, TX	Ac	
	5	Senior Data Er	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://www.linkedin.com/c	Apixio	Dallas, TX	Ac	
	6	IT Systems Eng	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://www.linkedin.com/c	Motion Recruitment	Dallas, TX	Ac	
	7	Quality Contro	https://media.licdn.com/dm	https://www.linkedin.com/jc	https://www.linkedin.com/c	Randstad Life Scienc	Fort Worth, TX	Ac	
	8	System Integra	(NULL)	https://www.linkedin.com/jc	https://www.linkedin.com/c	Insight Global	Dallas, TX	Ac	