Name: Divya Patil
Roll No.: 08
Experiment No. 2
To perform web crawling, scraping and parsing using Instant data scraper, Netlytic and Octoparse
Date of Performance: 08/02/2024
Date of Submission: 15/02/2024

Aim: To perform web crawling, scraping and parsing using Instant data scraper, Netlytic and Octoparse.

Objective: To apply web crawling, scraping, and parsing techniques to extract data from Google reviews using Instant Data Scraper, extract data from YouTube comments using Netlytic, and set up and run web scraping tasks to extract data using Octoparse.

Theory:

Web crawling: Web crawling is the process of automatically browsing the internet and indexing web pages. It is typically done by search engines to discover new content and update their indexes. Web crawlers, also known as spiders or bots, follow links from one page to another and download the content of each page for indexing. While web crawling is not the same as web scraping, web scraping often involves web crawling to navigate through a website and extract data from multiple pages.

Web scraping: This is the process of extracting specific information from websites. It involves using software or programming scripts to access the HTML of web pages and extract the desired data, such as text, images, or links. Web scraping can be done manually or automatically, and it is used for various purposes, including data collection, market research, and price monitoring.

Parsing: Parsing is the process of analyzing the structure of a document or data file to extract meaningful information. In the context of web scraping, parsing is used to extract specific data elements from the HTML or other markup languages used to create web pages. This process involves identifying the patterns and structures of the data and using techniques like regular expressions or HTML parsers to extract the desired information.

Instant Data Scraper: Instant Data Scraper is a Chrome extension that allows scraping data from websites directly in your browser. It provides a simple interface for selecting and extracting data elements, and it can export the data in various formats like CSV or Excel. Instant Data

Scraper is useful for quick and easy web scraping tasks, but it may have limitations compared to more advanced scraping tools.

Netlytic: Netlytic is a cloud-based text and social network analyzer that allows users to collect, analyze, and visualize social media data. It can be used to study online communities, track social media trends, and analyze text data from various sources, including Twitter, Facebook, YouTube, and web forums. Netlytic offers features for data collection, text analysis, and network analysis, making it a versatile tool for social media research and analysis.

Octoparse: Octoparse is a web scraping tool that allows you to extract data from websites without the need for programming. It provides a visual interface for selecting the data to scrape and offers features like scheduled scraping, cloud extraction, and data export options. It's commonly used for tasks such as web data collection, price monitoring, and market research.

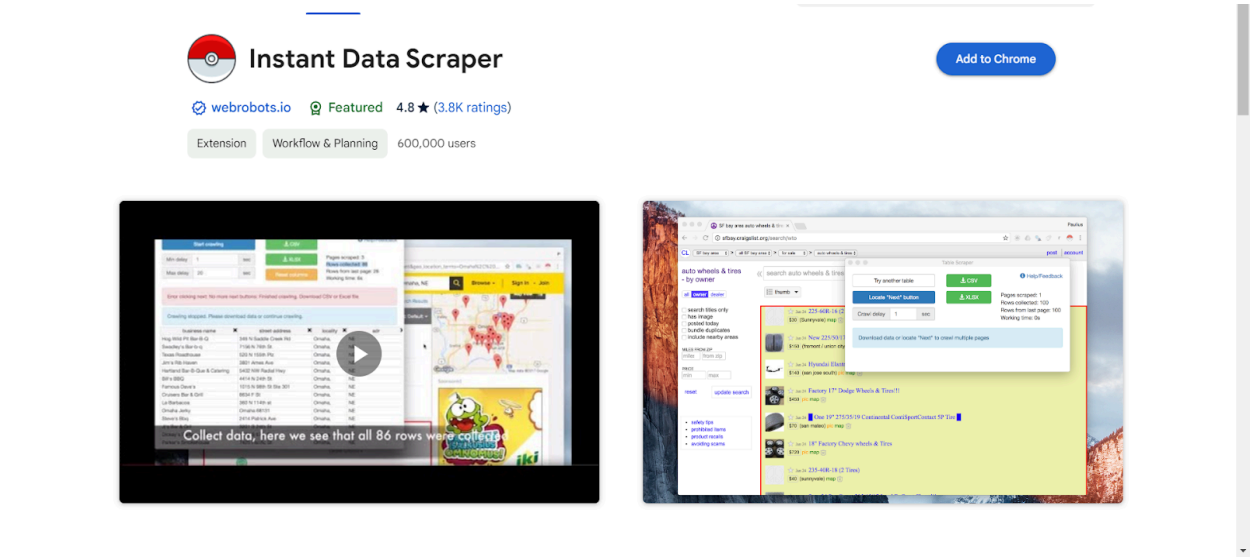
Implementation and Output:

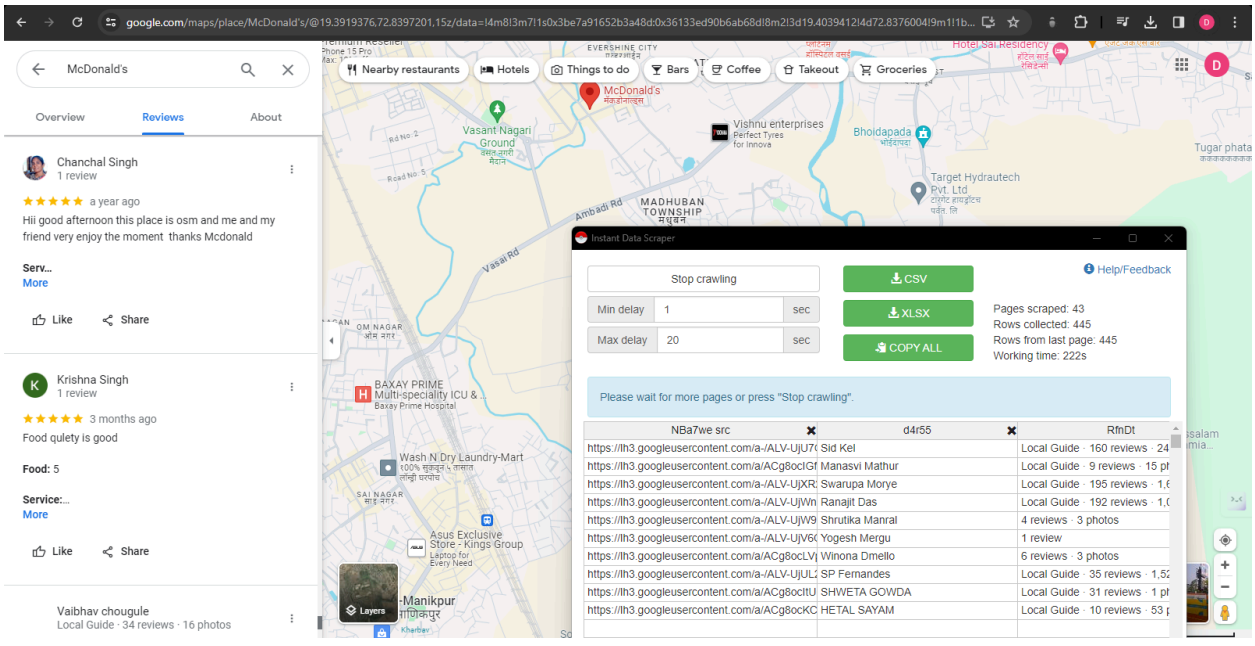
Scrape Google Reviews

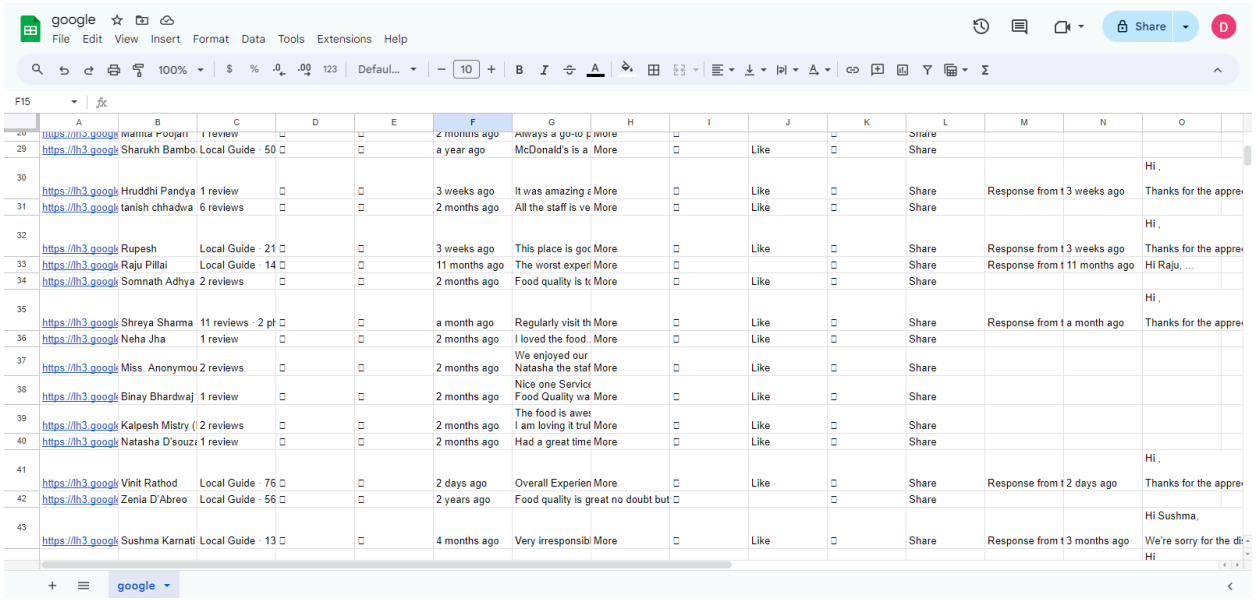
Step 1 : Install the Google Chrome extension Instant Data Scraper to scrape Google reviews for any local business

Step 2 : Go to Google Maps and look for a business that interests you

Step 3 : Choose the reviews and launch Instant Data Scraper to crawl Google reviews. Wait until all reviews have been scraped


Instant Data Scraper
Add to Chrome
webrobots.io
Featured
4.8 ★ (3.8K ratings)
Extension
Workflow & Planning
600,000 users
Collect data, here we see that all 86 rows were collected


McDonald's
Overview
Reviews
About
Chanchal Singh
1 review
a year ago
Hii good afternoon this place is osm and me and my friend very enjoy the moment thanks Mcdonald
Like
Share
Krishna Singh
1 review
3 months ago
Food qulety is good
Food: 5
Service:...
More
Like
Share
Vaibhav chougule
Local Guide · 34 reviews · 16 photos
Nearby restaurants
Hotels
Things to do
Bars
Coffee
Takeout
Groceries
Instant Data Scraper
Stop crawling
CSV
XLSX
COPY ALL
Min delay 1 sec
Max delay 20 sec
Help/Feedback
Pages scraped: 43
Rows collected: 445
Rows from last page: 445
Working time: 222s
Please wait for more pages or press "Stop crawling".
NBa7we src
d4r55
RfnDt



Scrape YouTube Comments using Netlytic

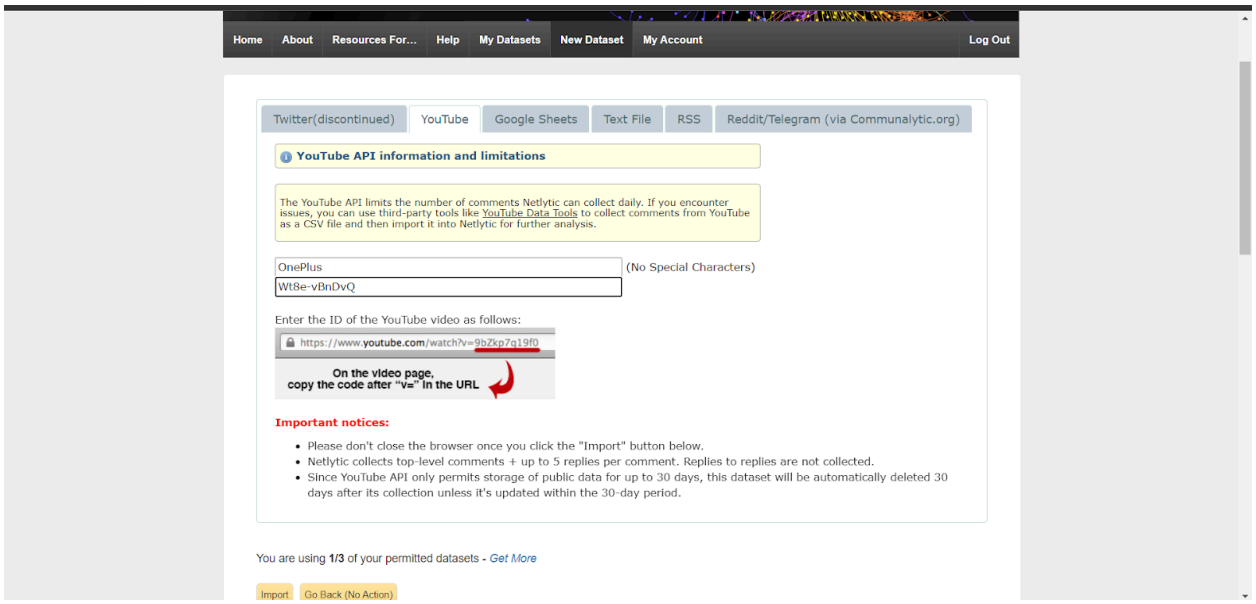
Step 1 : Sign up for Netlytic

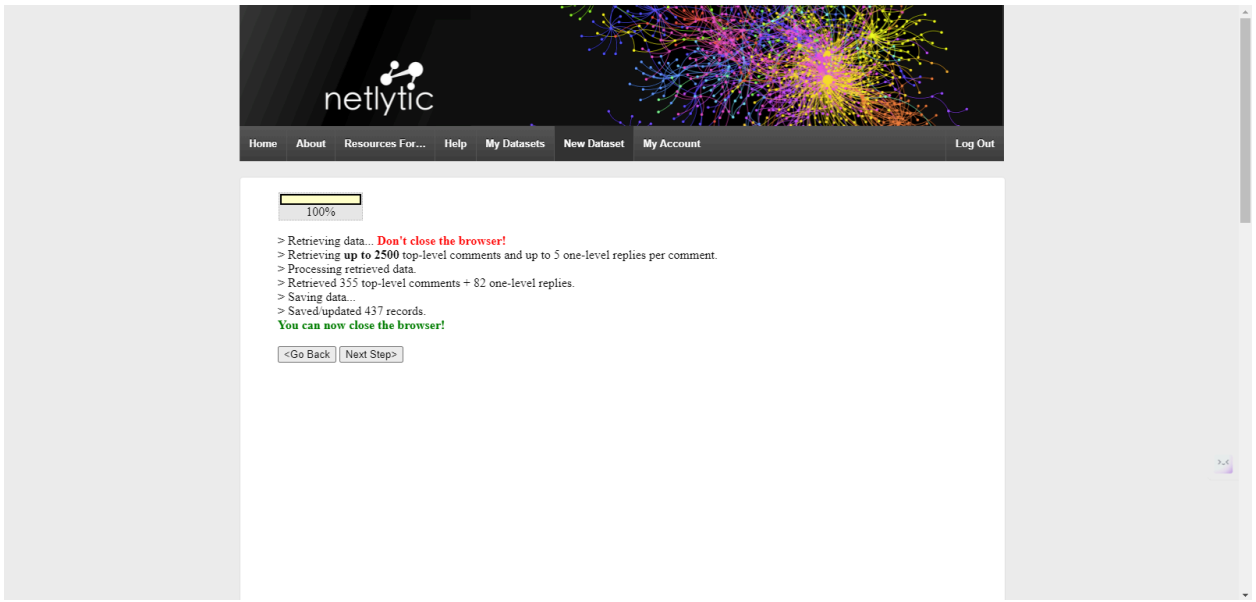
Step 2 : Click "New Dataset"

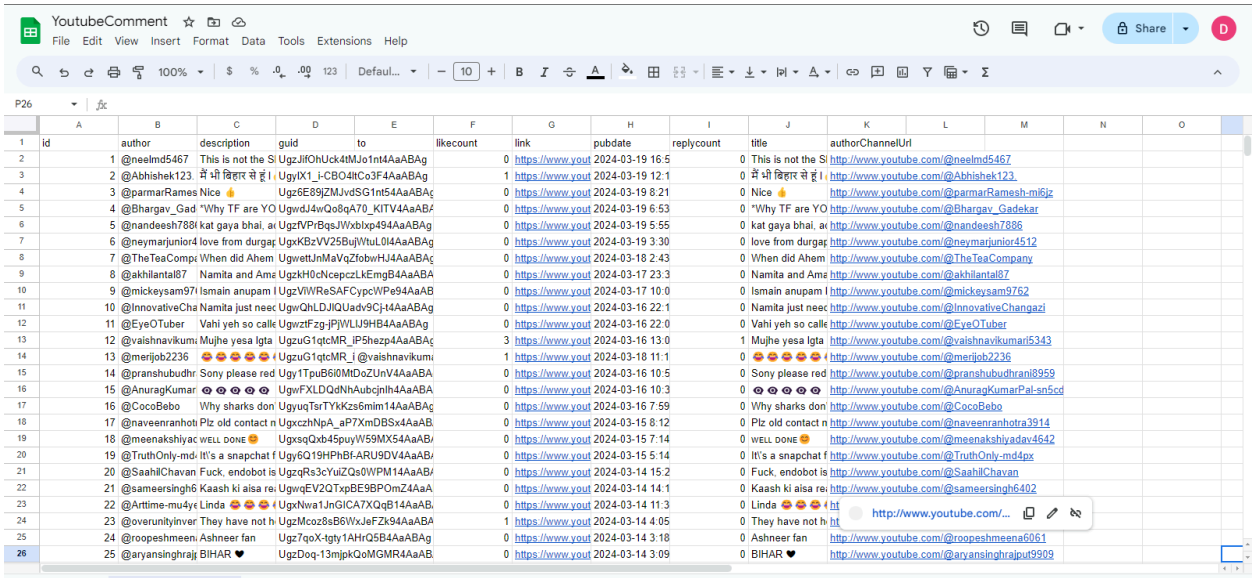
Step 3 : Select "YouTube" as the data source

Step 4 : Copy the YouTube video ID you want to scrape comments from and paste it into Netlytic, also enter Dataset Name and click import

Step 5 : Go to "My Datasets" tab where you can find your dataset


Home
About
Resources For...
Help
My Datasets
New Dataset
My Account
Log Out
Twitter(discontinued)
YouTube
Google Sheets
Text File
RSS
Reddit/Telegram (via Communalytic.org)
YouTube API information and limitations
The YouTube API limits the number of comments Netlytic can collect daily. If you encounter issues, you can use third-party tools like YouTube Data Tools to collect comments from YouTube as a CSV file and then import it into Netlytic for further analysis.
OnePlus
(No Special Characters)
Wt8e-vBnDvQ
Enter the ID of the YouTube video as follows:
https://www.youtube.com/watch?v=9bZkp7q19f0
On the video page, copy the code after "v=" in the URL
Important notices:
Please don't close the browser once you click the "Import" button below.
Netlytic collects top-level comments + up to 5 replies per comment. Replies to replies are not collected.
Since YouTube API only permits storage of public data for up to 30 days, this dataset will be automatically deleted 30 days after its collection unless it's updated within the 30-day period.
You are using 1/3 of your permitted datasets - Get More
Import
Go Back (No Action)


netlytic
Home
About
Resources For...
Help
My Datasets
New Dataset
My Account
Log Out
100%
> Retrieving data... Don't close the browser!
> Retrieving up to 2500 top-level comments and up to 5 one-level replies per comment.
> Processing retrieved data.
> Retrieved 355 top-level comments + 82 one-level replies.
> Saving data...
> Saved/updated 437 records.
You can now close the browser!
<Go Back
Next Step>



Web Scraping using Octoparse

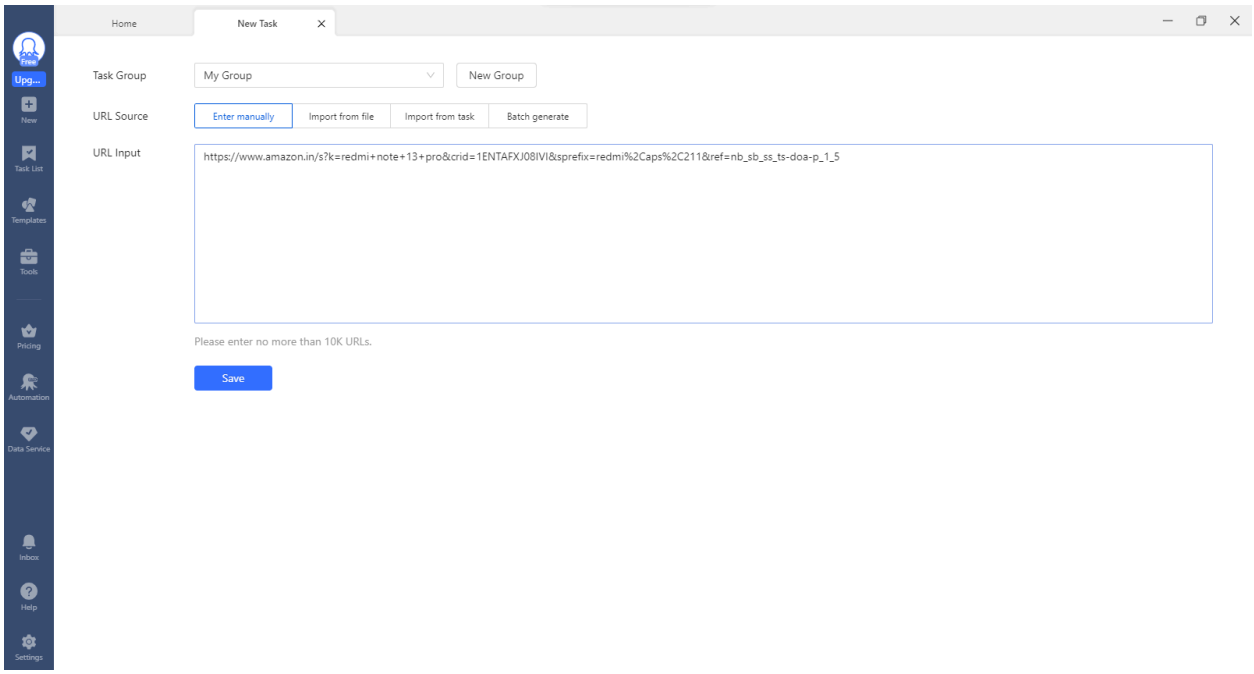
Step 1 : Go to web page

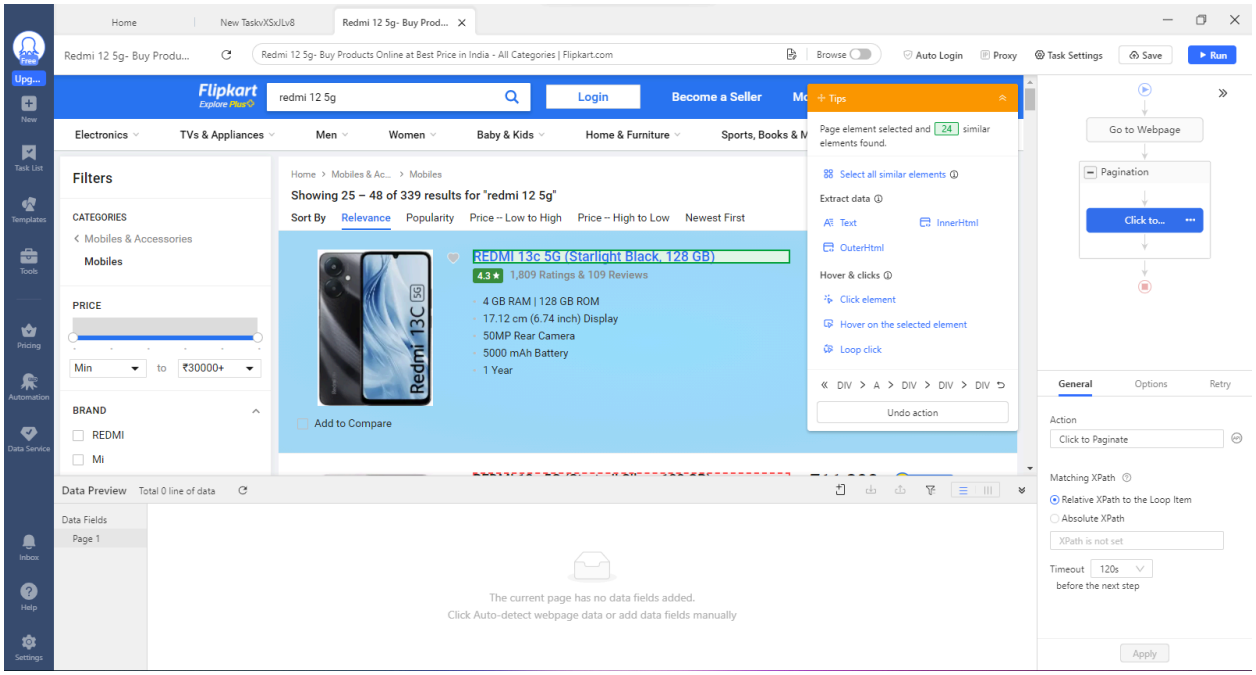
Step 2 : Create pagination

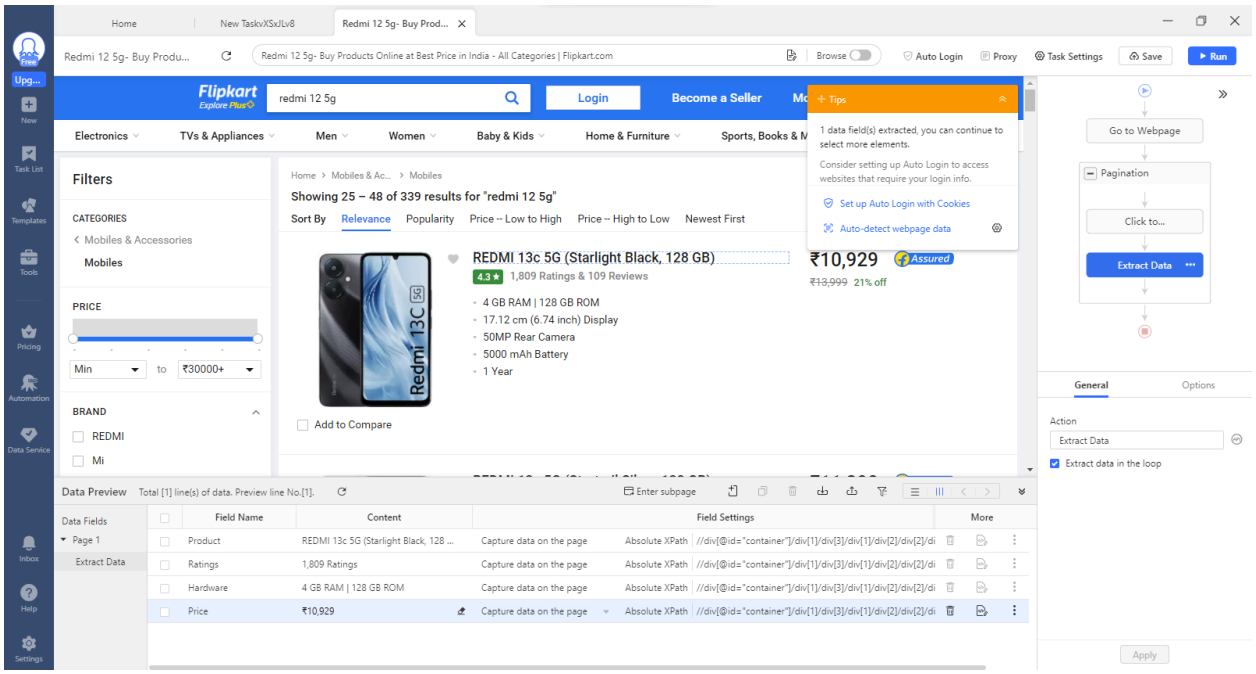
Step 3 : Build a loop item

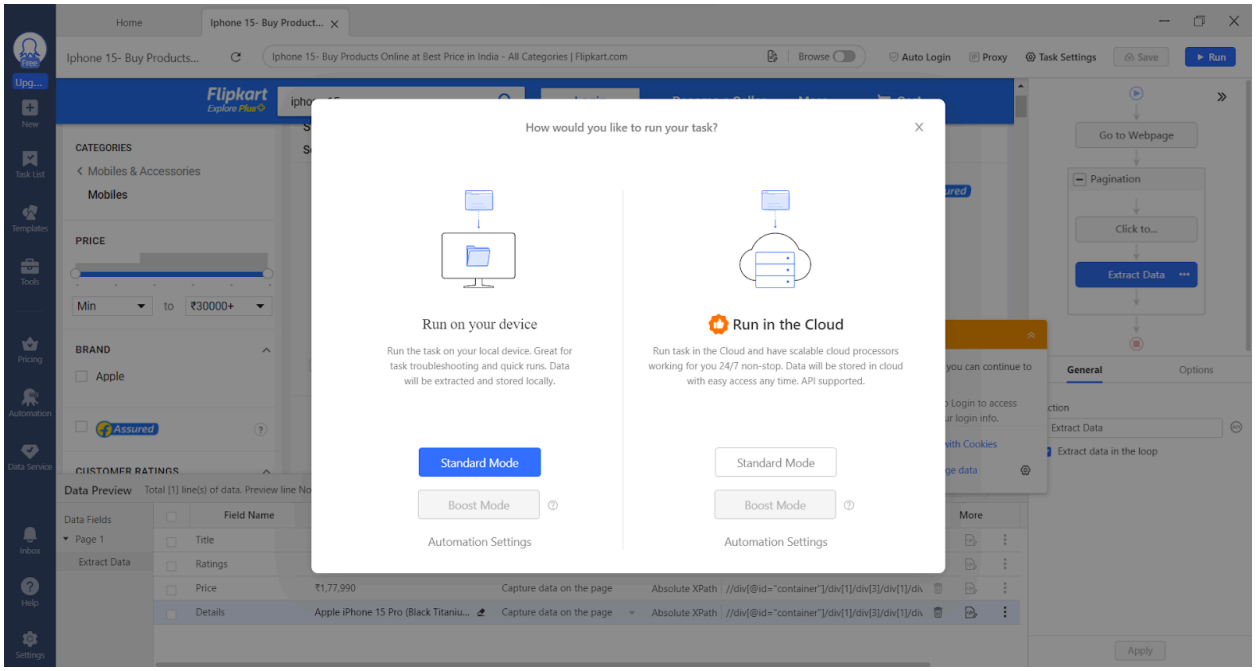
Step 4 : Extract the data

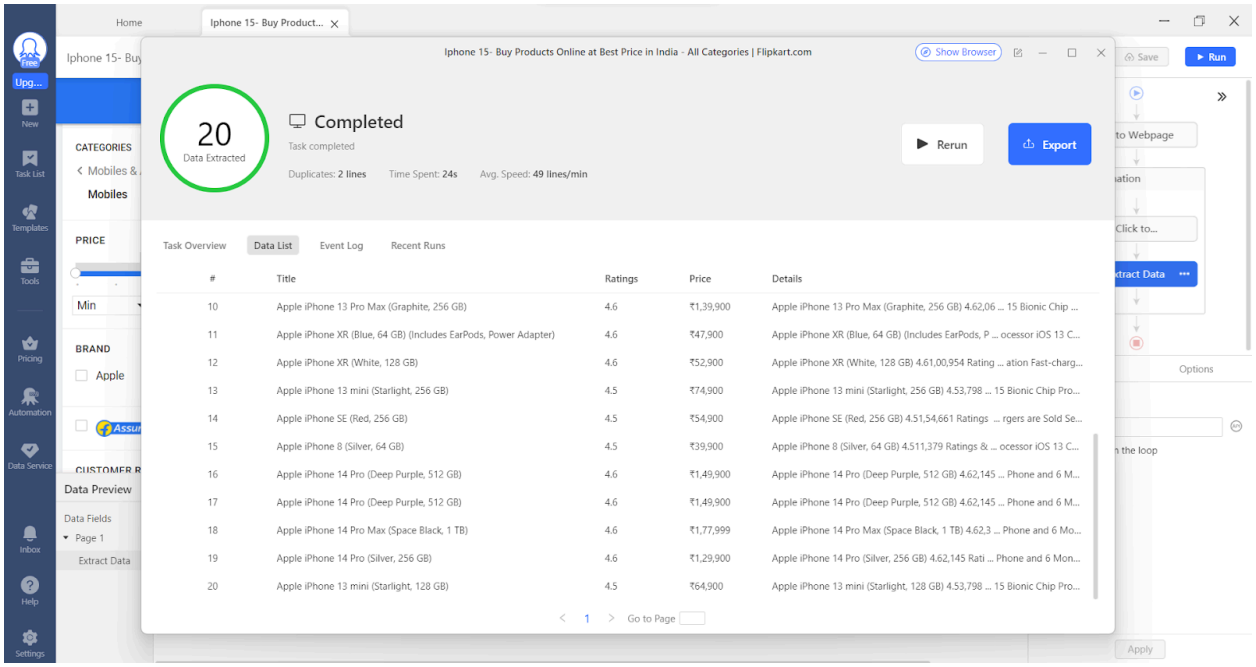
Step 5 : Run the task and get the data

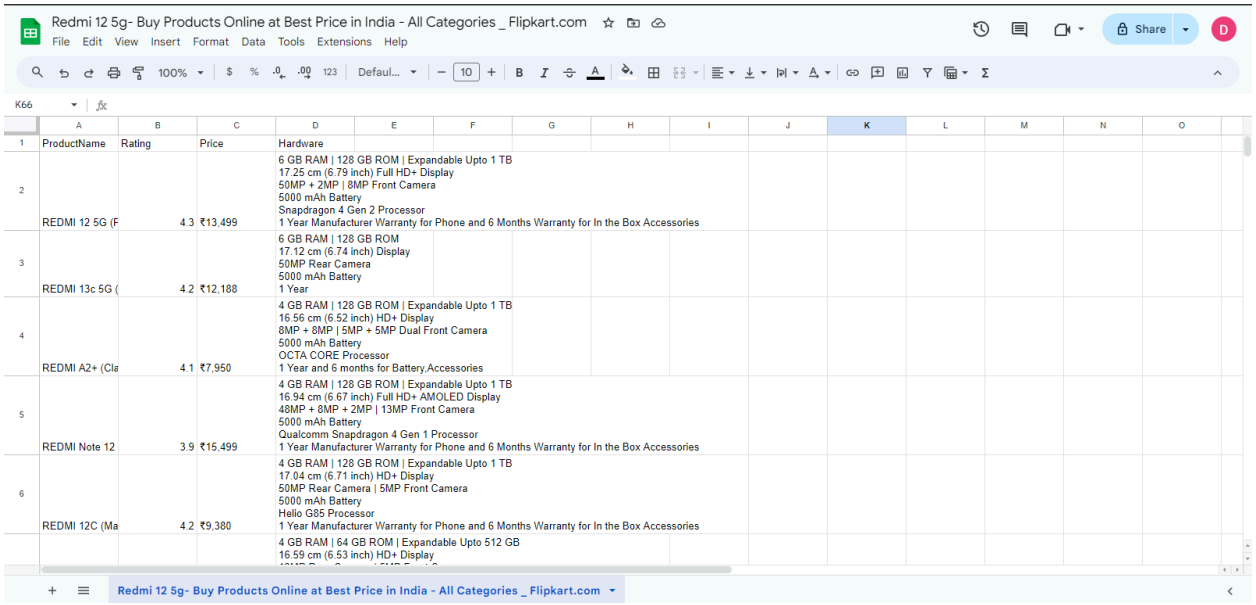



Redmi 12 5g- Buy Produ...
Flipkart
redmi 12 5g
Login
Become a Seller
Filters
CATEGORIES
Mobiles & Accessories
Mobiles
PRICE
BRAND
REDMI
Mi
Showing 25 – 48 of 339 results for "redmi 12 5g"
Sort By
Relevance
Popularity
Price -- Low to High
Price -- High to Low
Newest First
REDMI 13c 5G (Starlight Black, 128 GB)
4.3★ 1,809 Ratings & 109 Reviews
4 GB RAM | 128 GB ROM
17.12 cm (6.74 inch) Display
50MP Rear Camera
5000 mAh Battery
1 Year
Add to Compare
Tips
Page element selected and 24 similar elements found.
Select all similar elements
Extract data
Text
InnerHtml
OuterHtml
Hover & clicks
Click element
Hover on the selected element
Loop click
Undo action
Go to Webpage
Pagination
Click to...
General
Options
Retry
Action
Click to Paginate
Matching XPath
Relative XPath to the Loop Item
Absolute XPath
Timeout
120s
before the next step
Apply
Data Preview
Total 0 line of data
Data Fields
Page 1
The current page has no data fields added.
Click Auto-detect webpage data or add data fields manually


Redmi 12 5g- Buy Produ...
Flipkart
redmi 12 5g
Login
Become a Seller
Filters
CATEGORIES
Mobiles & Accessories
Mobiles
PRICE
BRAND
REDMI
Mi
Showing 25 – 48 of 339 results for "redmi 12 5g"
Sort By
Relevance
Popularity
Price -- Low to High
Price -- High to Low
Newest First
REDMI 13c 5G (Starlight Black, 128 GB)
4.3★ 1,809 Ratings & 109 Reviews
4 GB RAM | 128 GB ROM
17.12 cm (6.74 inch) Display
50MP Rear Camera
5000 mAh Battery
1 Year
Add to Compare
₹10,929
Assured
₹13,999 21% off
Tips
1 data field(s) extracted, you can continue to select more elements.
Consider setting up Auto Login to access websites that require your login info.
Set up Auto Login with Cookies
Auto-detect webpage data
Go to Webpage
Pagination
Click to...
Extract Data
General
Options
Action
Extract Data
Extract data in the loop
Apply
Data Preview
Total [1] line(s) of data. Preview line No.[1].
Data Fields
Page 1
Extract Data
Field Name
Content
Field Settings
More
Product
REDMI 13c 5G (Starlight Black, 128 ...
Capture data on the page
Absolute XPath
Ratings
1,809 Ratings
Hardware
4 GB RAM | 128 GB ROM
Price
₹10,929


How would you like to run your task?
Run on your device
Run the task on your local device. Great for task troubleshooting and quick runs. Data will be extracted and stored locally.
Standard Mode
Boost Mode
Automation Settings
Run in the Cloud
Run task in the Cloud and have scalable cloud processors working for you 24/7 non-stop. Data will be stored in cloud with easy access any time. API supported.
Standard Mode
Boost Mode
Automation Settings


Iphone 15- Buy Products Online at Best Price in India - All Categories | Flipkart.com
20
Data Extracted
Completed
Task completed
Duplicates: 2 lines
Time Spent: 24s
Avg. Speed: 49 lines/min
Rerun
Export
Task Overview
Data List
Event Log
Recent Runs
Title Ratings Price Details
10 Apple iPhone 13 Pro Max (Graphite, 256 GB) 4.6 ₹1,39,900 Apple iPhone 13 Pro Max (Graphite, 256 GB) 4.62,06 ... 15 Bionic Chip ...
11 Apple iPhone XR (Blue, 64 GB) (Includes EarPods, Power Adapter) 4.6 ₹47,900 Apple iPhone XR (Blue, 64 GB) (Includes EarPods, P ... ocessor iOS 13 C...
12 Apple iPhone XR (White, 128 GB) 4.6 ₹52,900 Apple iPhone XR (White, 128 GB) 4.61,00,954 Rating ... ation Fast-charg...
13 Apple iPhone 13 mini (Starlight, 256 GB) 4.5 ₹74,900 Apple iPhone 13 mini (Starlight, 256 GB) 4.53,798 ... 15 Bionic Chip Pro...
14 Apple iPhone SE (Red, 256 GB) 4.5 ₹54,900 Apple iPhone SE (Red, 256 GB) 4.51,54,661 Ratings ... rgers are Sold Se...
15 Apple iPhone 8 (Silver, 64 GB) 4.5 ₹39,900 Apple iPhone 8 (Silver, 64 GB) 4.511,379 Ratings & ... ocessor iOS 13 C...
16 Apple iPhone 14 Pro (Deep Purple, 512 GB) 4.6 ₹1,49,900 Apple iPhone 14 Pro (Deep Purple, 512 GB) 4.62,145 ... Phone and 6 M...
17 Apple iPhone 14 Pro (Deep Purple, 512 GB) 4.6 ₹1,49,900 Apple iPhone 14 Pro (Deep Purple, 512 GB) 4.62,145 ... Phone and 6 M...
18 Apple iPhone 14 Pro Max (Space Black, 1 TB) 4.6 ₹1,77,999 Apple iPhone 14 Pro Max (Space Black, 1 TB) 4.62,3 ... Phone and 6 Mo...
19 Apple iPhone 14 Pro (Silver, 256 GB) 4.6 ₹1,29,900 Apple iPhone 14 Pro (Silver, 256 GB) 4.62,145 Rati ... Phone and 6 Mon...
20 Apple iPhone 13 mini (Starlight, 128 GB) 4.5 ₹64,900 Apple iPhone 13 mini (Starlight, 128 GB) 4.53,798 ... 15 Bionic Chip Pro...
Go to Page
Data Preview
Extract Data


Redmi 12 5g- Buy Products Online at Best Price in India - All Categories _ Flipkart.com
File Edit View Insert Format Data Tools Extensions Help
Share
ProductName Rating Price Hardware
REDMI 12 5G (F 4.3 ₹13,499 6 GB RAM | 128 GB ROM | Expandable Upto 1 TB 17.25 cm (6.79 inch) Full HD+ Display 50MP + 2MP | 8MP Front Camera 5000 mAh Battery Snapdragon 4 Gen 2 Processor 1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories
REDMI 13c 5G (4.2 ₹12,188 6 GB RAM | 128 GB ROM 17.12 cm (6.74 inch) Display 50MP Rear Camera 5000 mAh Battery 1 Year
REDMI A2+ (Cla 4.1 ₹7,950 4 GB RAM | 128 GB ROM | Expandable Upto 1 TB 16.56 cm (6.52 inch) HD+ Display 8MP + 8MP | 5MP + 5MP Dual Front Camera 5000 mAh Battery OCTA CORE Processor 1 Year and 6 months for Battery,Accessories
REDMI Note 12 3.9 ₹15,499 4 GB RAM | 128 GB ROM | Expandable Upto 1 TB 16.94 cm (6.67 inch) Full HD+ AMOLED Display 48MP + 8MP + 2MP | 13MP Front Camera 5000 mAh Battery Qualcomm Snapdragon 4 Gen 1 Processor 1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories
REDMI 12C (Ma 4.2 ₹9,380 4 GB RAM | 128 GB ROM | Expandable Upto 1 TB 17.04 cm (6.71 inch) HD+ Display 50MP Rear Camera | 5MP Front Camera 5000 mAh Battery Hello G85 Processor 1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories
4 GB RAM | 64 GB ROM | Expandable Upto 512 GB 16.59 cm (6.53 inch) HD+ Display
Redmi 12 5g- Buy Products Online at Best Price in India - All Categories _ Flipkart.com

Conclusion: In conclusion, this experiment showcased the practical application of web crawling, scraping, and parsing techniques using Instant Data Scraper, Netlytic, and Octoparse. Instant Data Scraper proved useful for extracting Google reviews with its user-friendly interface, while Netlytic demonstrated its efficiency in analyzing social media data by extracting YouTube comments effectively. Octoparse's flexibility and automation features made it ideal for complex web scraping tasks, including pagination and data extraction from multiple pages. These tools collectively offer a range of capabilities for web data extraction, catering to different needs and skill levels, and can be valuable assets in research, analysis, and data-driven decision-making processes.