PROJECT TITLE	Indian Human Image Classification Model
GROUP NO.	B8
AGENDA	Tools & Strategy
DATE	07/04/2021
DESCRIPTION	Different Methods for collecting data

-> OUR TEAM HAS COME UP WITH SEVERAL IDEAS TO EXTRACT THE REQUIRED DATASET FOR THIS PROJECT.

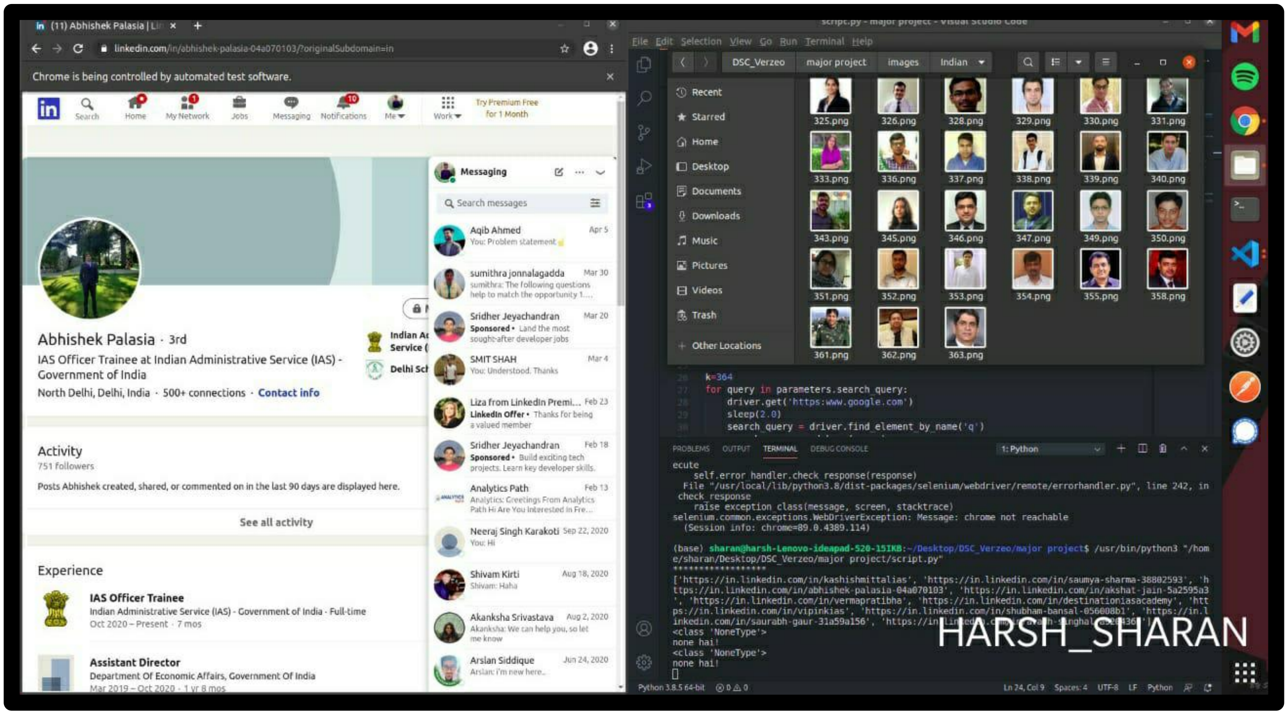
-> ALL METHODS MENTIONED HAVE BEEN TRIED AND TESTED BY ONE OR MORE GROUP MEMBERS WITH THE SCRIPTS RUNNING PERFECTLY(the screenshots have been attached for reference).

1. Manually:

We will search from Google, Youtube, and many different resources and we will be able to get images

2. Selenium (Automation):

By using a selenium webdriver we will be able to get images from a given URL and after applying a keyword we will able to get more images.



Chrome is being controlled by automated test software.

Abhishek Palasia · 3rd
IAS Officer Trainee at Indian Administrative Service (IAS) - Government of India
North Delhi, Delhi, India · 500+ connections · Contact info

Activity
751 followers
Posts Abhishek created, shared, or commented on in the last 90 days are displayed here.

Experience

IAS Officer Trainee
Indian Administrative Service (IAS) - Government of India · Full-time
Oct 2020 – Present · 7 mos

Assistant Director
Department Of Economic Affairs, Government Of India
Mar 2019 – Oct 2020 · 1 yr 8 mos

Aqib Ahmed · Apr 5
You: Problem statement

sumithra jonnalagadda · Mar 30
sumithra: The following questions help to match the opportunity 1....

Sridher Jeyachandran · Mar 20
Sponsored • Land the most sought-after developer jobs

SMIT SHAH · Mar 4
You: Understood. Thanks

Liza from LinkedIn Premi... · Feb 23
LinkedIn Offer • Thanks for being a valued member

Sridher Jeyachandran · Feb 18
Sponsored • Build exciting tech projects. Learn key developer skills.

Analytics Path · Feb 13
Analytics: Greetings From Analytics Path Hi Are You Interested In Fre...

Neeraj Singh Karakoti · Sep 22, 2020
You: Hi

Shivam Kirti · Aug 18, 2020
Shivam: Haha

Akanksha Srivastava · Aug 2, 2020
Akanksha: We can help you, so let me know

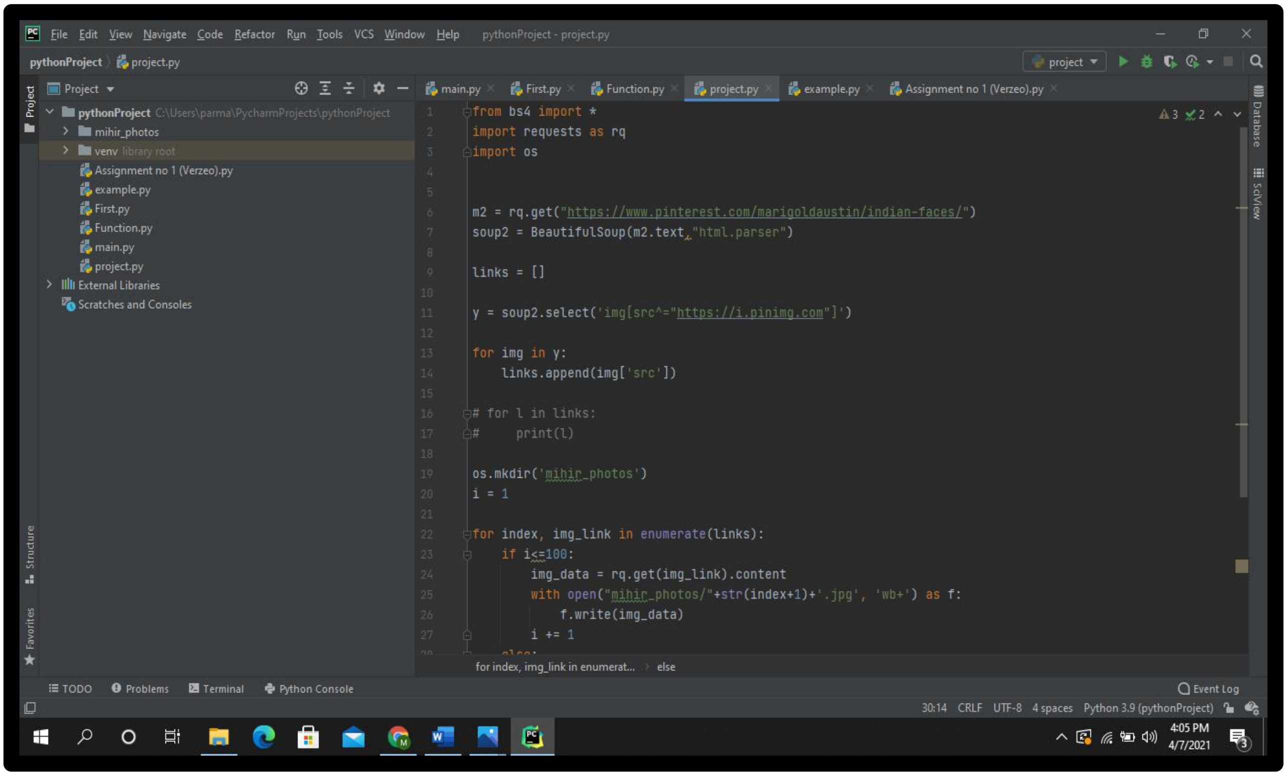
Arslan Siddique · Jun 24, 2020
Arslan: I'm new here...

```
k=364
for query in parameters.search_query:
    driver.get('https:www.google.com')
    sleep(2.0)
    search_query = driver.find_element_by_name('q')
```

```
ecute
    self.error_handler.check_response(response)
  File "/usr/local/lib/python3.8/dist-packages/selenium/webdriver/remote/errorhandler.py", line 242, in
 check_response
    raise exception_class(message, screen, stacktrace)
selenium.common.exceptions.WebDriverException: Message: chrome not reachable
  (Session info: chrome=89.0.4389.114)

(base) sharan@harsh-Lenovo-ideapad-520-15IKB:~/Desktop/DSC_Verzeo/major project$ /usr/bin/python3 "/hom
e/sharan/Desktop/DSC_Verzeo/major project/script.py"
******************
['https://in.linkedin.com/in/kashishmittalias', 'https://in.linkedin.com/in/saumya-sharma-38802593', 'h
ttps://in.linkedin.com/in/abhishek-palasia-04a070103', 'https://in.linkedin.com/in/akshat-jain-5a2595a3
', 'https://in.linkedin.com/in/vermapratibha', 'https://in.linkedin.com/in/destinationiasacademy', 'htt
ps://in.linkedin.com/in/vipinkias', 'https://in.linkedin.com/in/shubham-bansal-056008b1', 'https://in.l
inkedin.com/in/saurabh-gaur-31a59a156', 'https://in[illegible]h-singhal-[illegible]']
<class 'NoneType'>
none hai!
<class 'NoneType'>
none hai!
```

3. Crawler:

Crawler can be used to find links to an image and after obtaining the link we will make a directory to our local system and get all images based on the path which we have given.



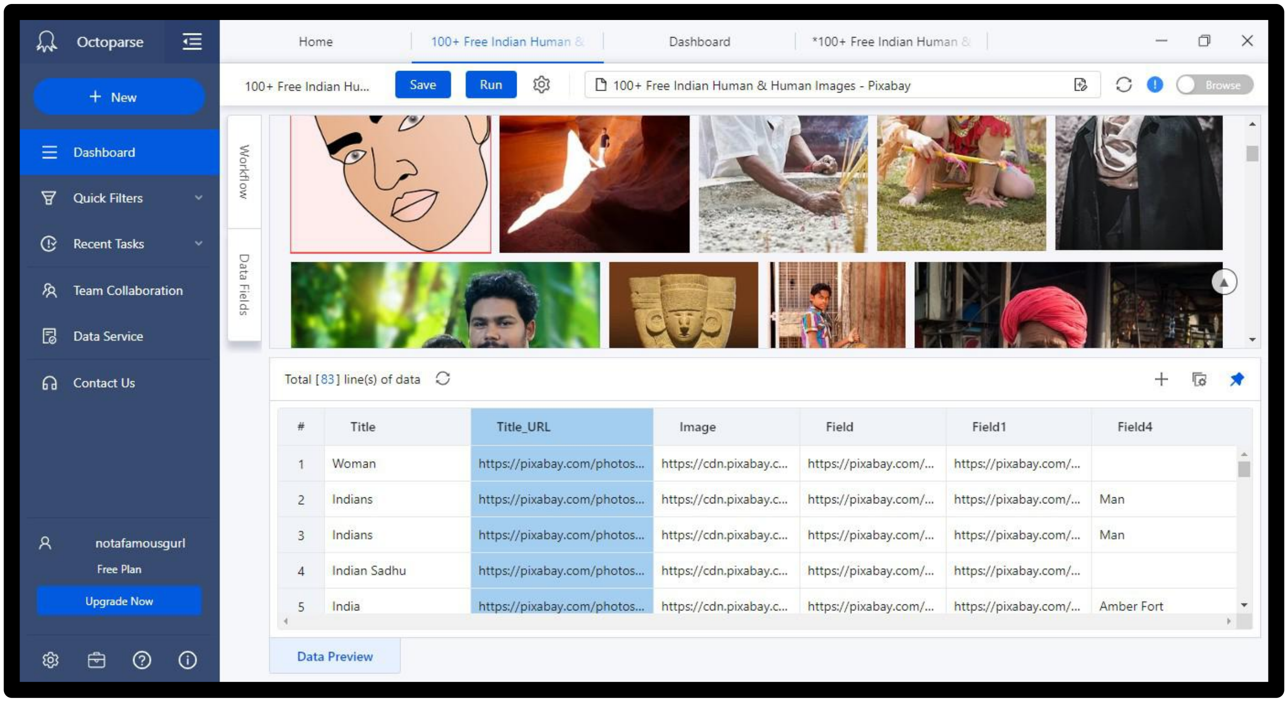
```python
from bs4 import *
import requests as rq
import os


m2 = rq.get("https://www.pinterest.com/marigoldaustin/indian-faces/")
soup2 = BeautifulSoup(m2.text,"html.parser")

links = []

y = soup2.select('img[src^="https://i.pinimg.com"]')

for img in y:
    links.append(img['src'])

# for l in links:
#     print(l)

os.mkdir('mihir_photos')
i = 1

for index, img_link in enumerate(links):
    if i<=100:
        img_data = rq.get(img_link).content
        with open("mihir_photos/"+str(index+1)+'.jpg', 'wb+') as f:
            f.write(img_data)
        i += 1
    else:
```

File Edit View Navigate Code Refactor Run Tools VCS Window Help
pythonProject - project.py
pythonProject 〉 project.py
project
Project
pythonProject C:\Users\parma\PycharmProjects\pythonProject
mihir_photos
venv library root
Assignment no 1 (Verzeo).py
example.py
First.py
Function.py
main.py
project.py
External Libraries
Scratches and Consoles
main.py
First.py
Function.py
project.py
example.py
Assignment no 1 (Verzeo).py
Database
SciView
Structure
Favorites
for index, img_link in enumerat... 〉 else
TODO
Problems
Terminal
Python Console
Event Log
30:14 CRLF UTF-8 4 spaces Python 3.9 (pythonProject)
4:05 PM
4/7/2021

4. From Softwares:

(like Octoparse)

By simply providing the URL of any website and applying some settings it will fetch all the images from the website.

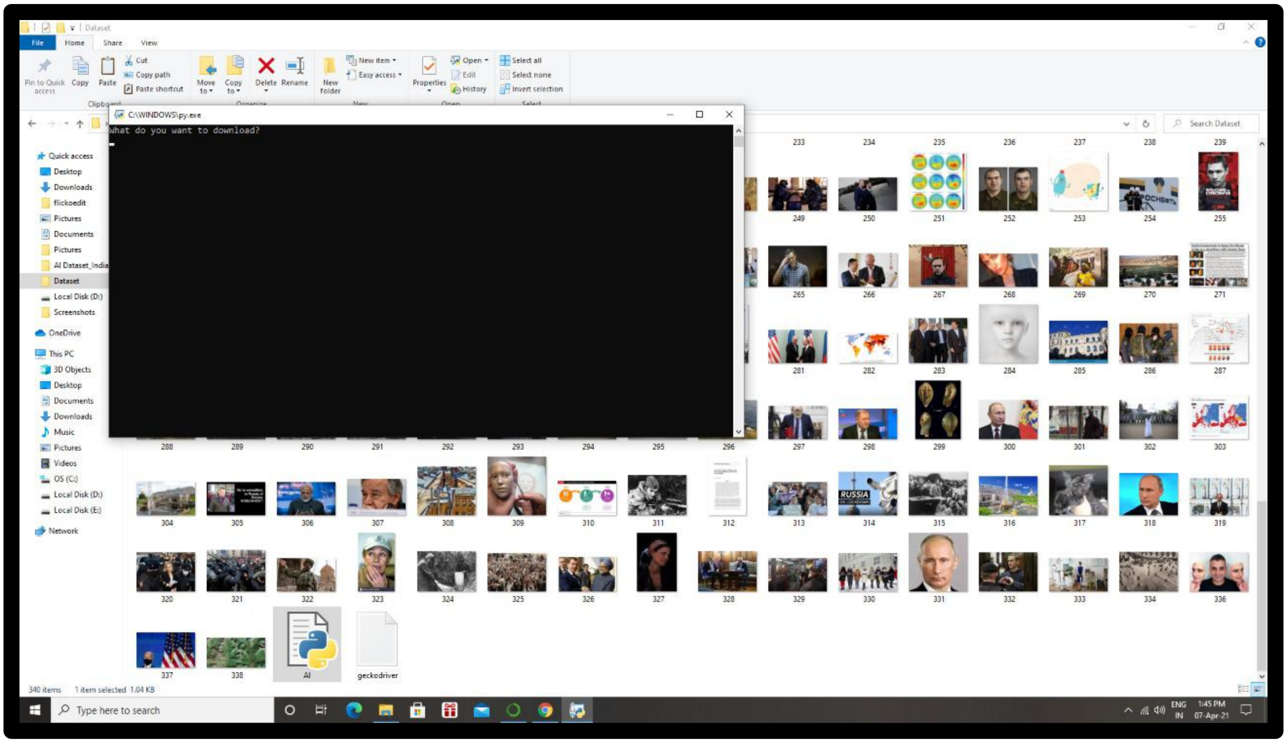


Total [83] line(s) of data

#	Title	Title_URL	Image	Field	Field1	Field4
1	Woman	https://pixabay.com/photos...	https://cdn.pixabay.c...	https://pixabay.com/...	https://pixabay.com/...	
2	Indians	https://pixabay.com/photos...	https://cdn.pixabay.c...	https://pixabay.com/...	https://pixabay.com/...	Man
3	Indians	https://pixabay.com/photos...	https://cdn.pixabay.c...	https://pixabay.com/...	https://pixabay.com/...	Man
4	Indian Sadhu	https://pixabay.com/photos...	https://cdn.pixabay.c...	https://pixabay.com/...	https://pixabay.com/...	
5	India	https://pixabay.com/photos...	https://cdn.pixabay.c...	https://pixabay.com/...	https://pixabay.com/...	Amber Fort

5. ImageScraper:

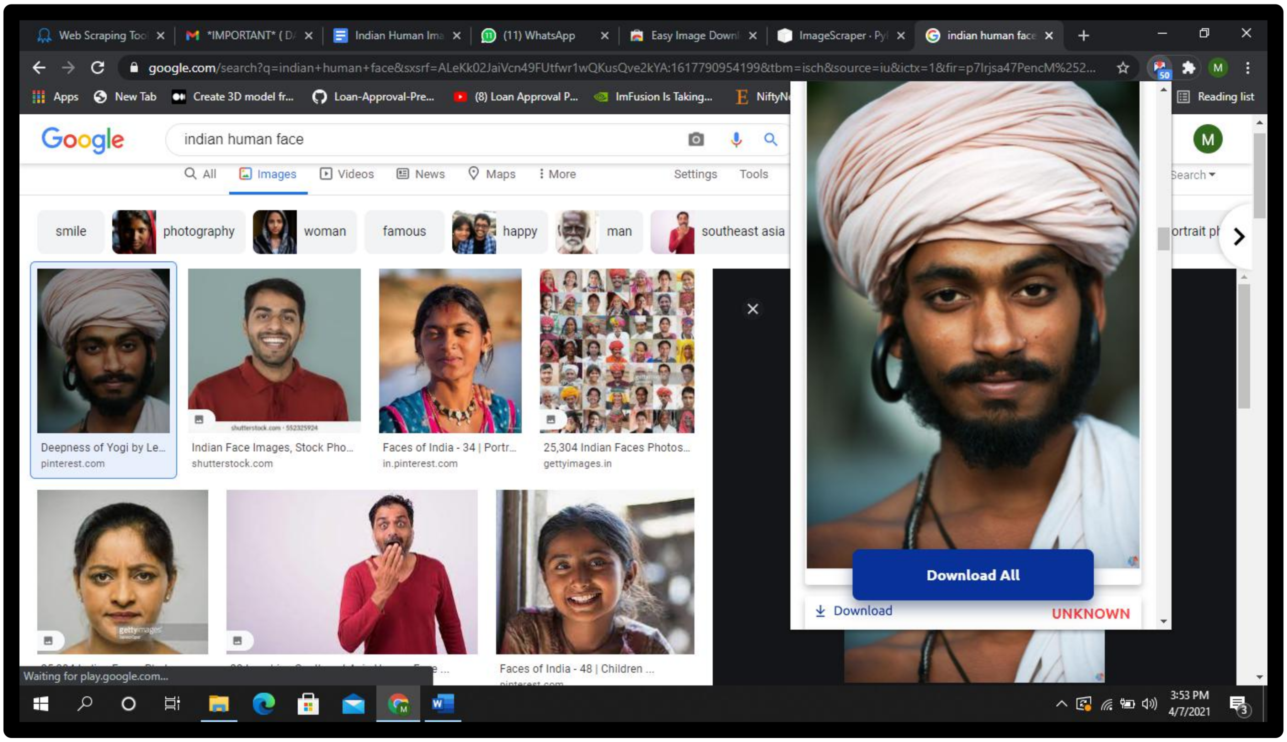
It is a python library. Simply giving a command (let "image-scraper ananth.co.in/test.html") to the command prompt it will give you the desired output.



```
What do you want to download?
```

6. From chrome extensions:

We will be using some chrome extensions in which after giving a particular keyword to any search box it will download all images of that particular keyword.



---------------END---------------