

GOOGLE IMAGES SCRAPING

Using Anaconda Prompt in Windows OS

- By Prissy Nusaiba Yulisa -



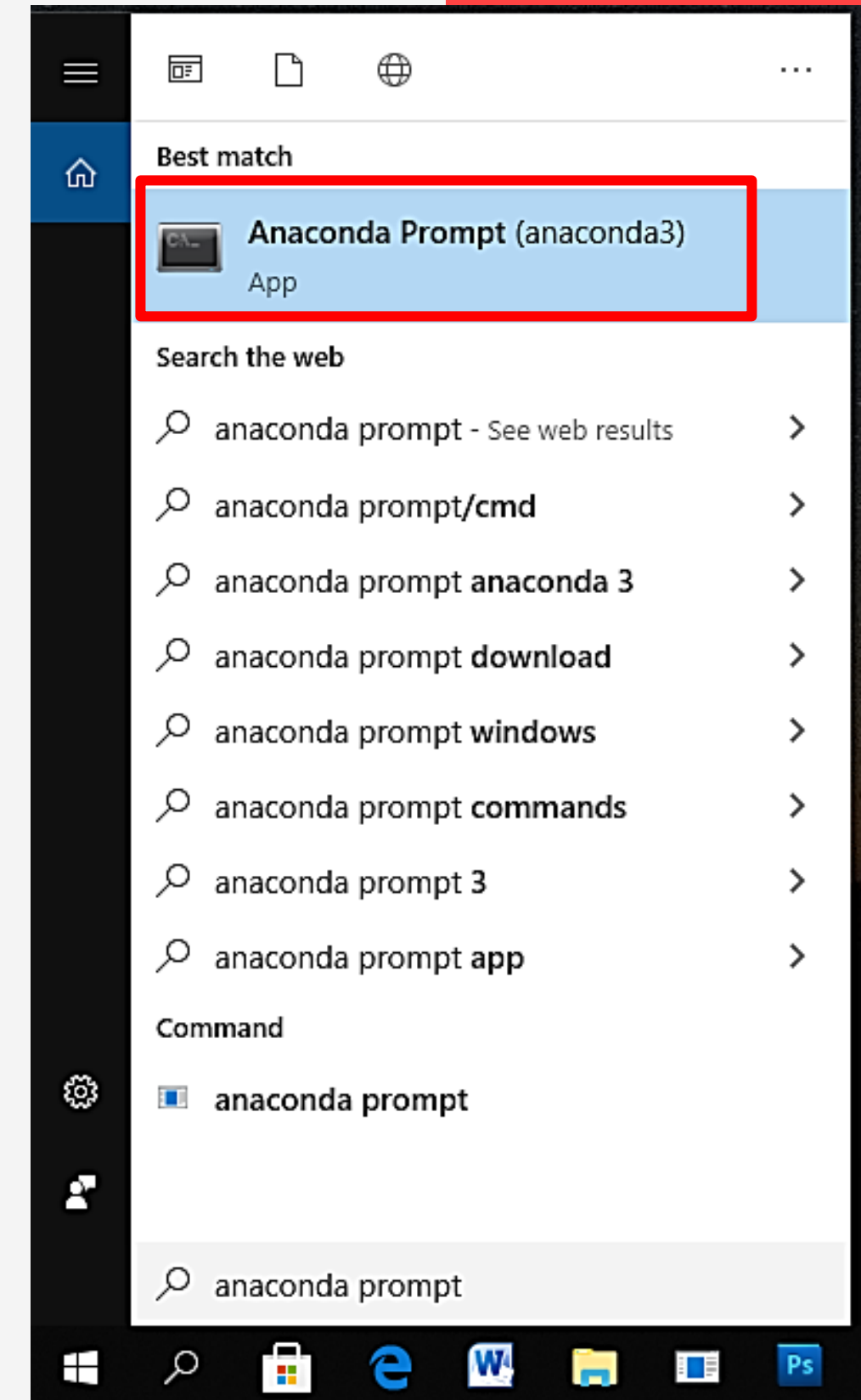
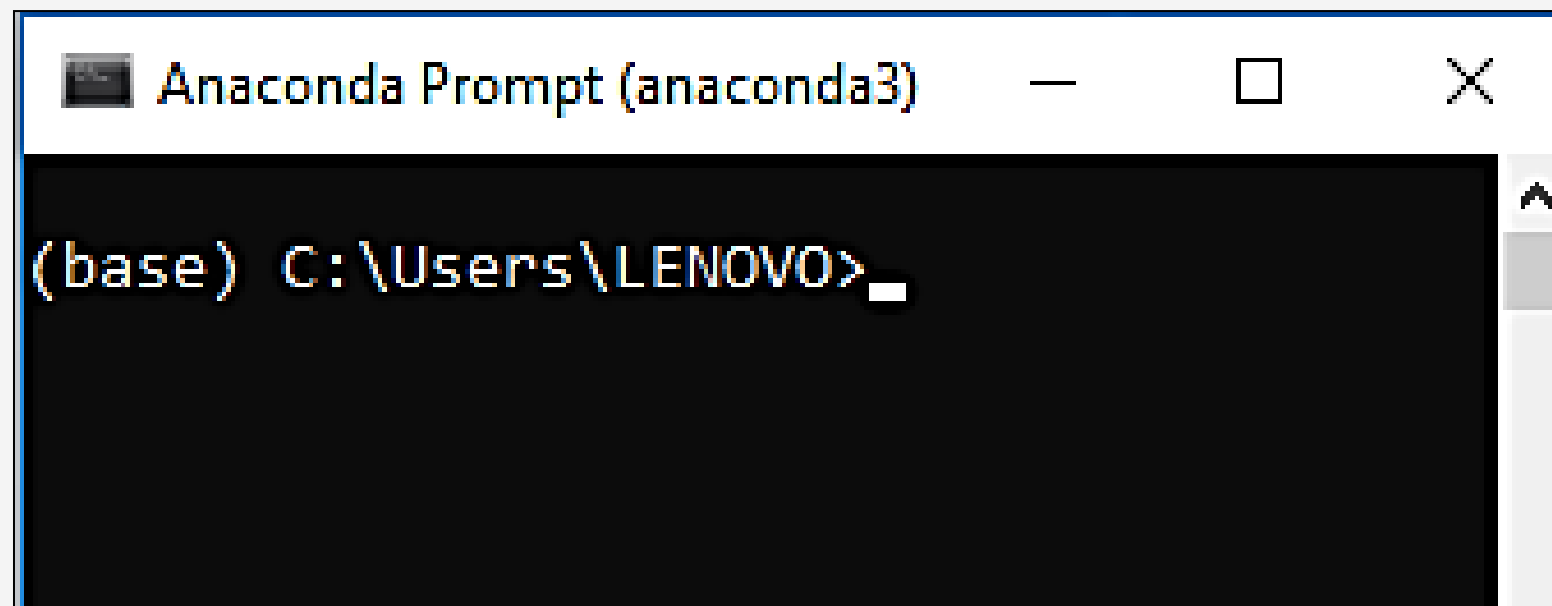
STEP 1

Pastikan anda telah Install Anaconda for Windows melalui link ini : <https://www.anaconda.com/products/individual>



Buka
Anaconda
Prompt

Berikut tampilannya



STEP 2

Ketikkan Perintah “**conda create -n sharing python=3.8**” → klik **Enter** pada anaconda prompt untuk Membuat Virtual Environment

```
Anaconda Prompt (anaconda3) - conda create -n sharing python=3.8

(base) C:\Users\LENOVO> conda create -n sharing python=3.8
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

Note : Nama virtual environment “sharing” itu adalah bebas sesuai keinginan

Ketika selesai, akan muncul pilihan **([y]/n)?** maka harus **pilih y**

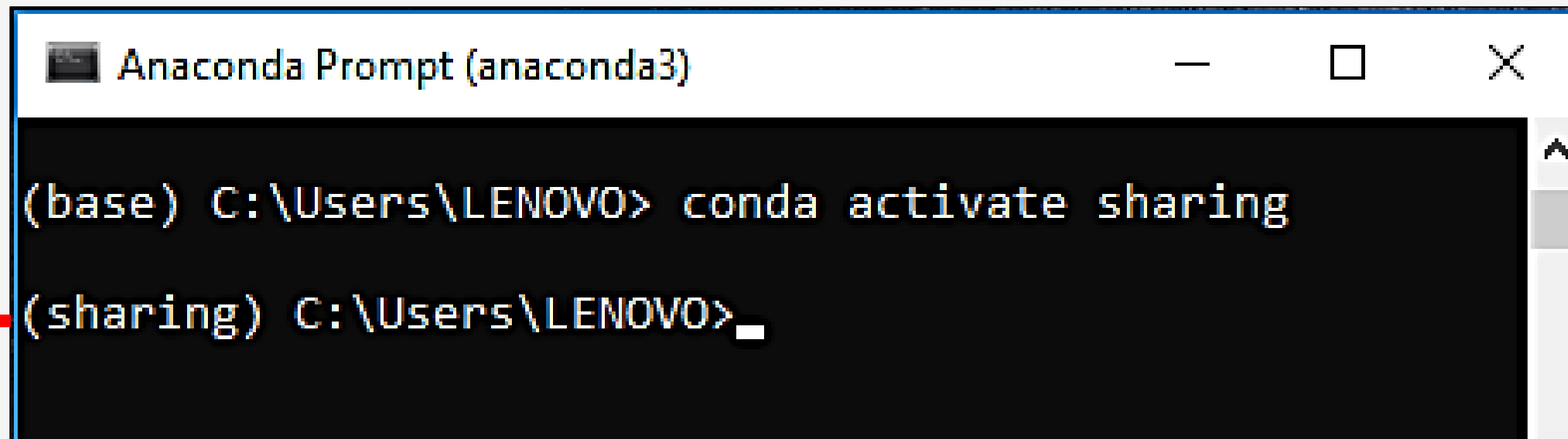
```
Anaconda Prompt (anaconda3)

Proceed ([y]/n)? y

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

STEP 3

Ketikkan perintah "**conda activate sharing**" → klik Enter
untuk mengaktifkan Virtual Environment



```
Anaconda Prompt (anaconda3)

(base) C:\Users\LENOVO> conda activate sharing

(sharing) C:\Users\LENOVO>_
```

Note: Jika virtual environment telah aktif, maka namanya akan muncul menggantikan base.

STEP 4

- 1 Mengubah tempat penyimpanan, yang awalnya di local "C:" diganti dengan mengetikkan perintah "E:" → klik Enter

```
Anaconda Prompt (anaconda3)
(sharing) C:\Users\LENOVO> E:
```

Tahap ini menyesuaikan space penyimpanan yang masih longgar di komputer anda, bisa tidak di ubah, atau di ubah ke local D: dll.

- 2 Membuat folder baru di local E : dengan mengetikkan perintah "mkdir Sharing" → klik Enter

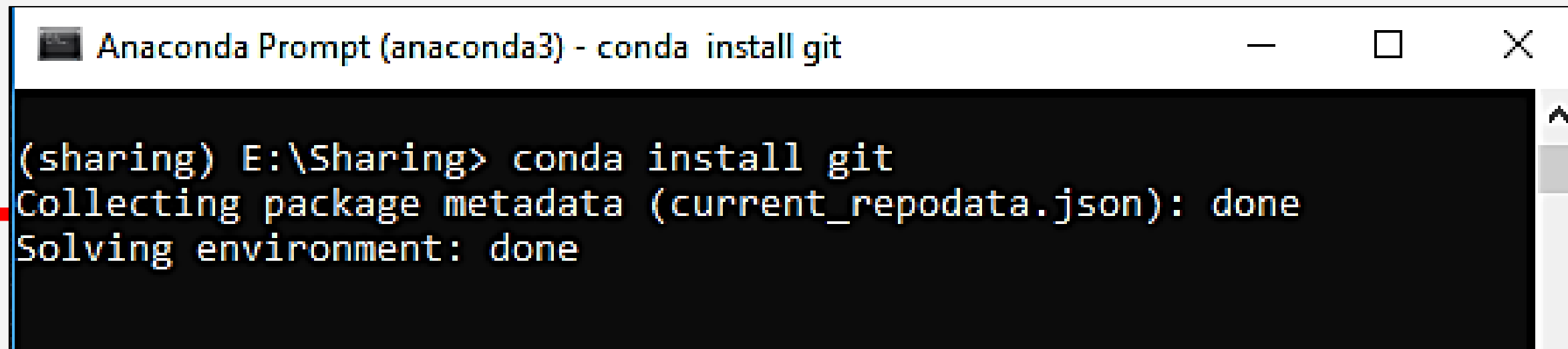
```
Anaconda Prompt (anaconda3)
(sharing) E:\> mkdir Sharing
```

- 3 Selanjutnya ketikkan perintah "cd Sharing" → klik Enter untuk masuk ke dalam folder tersebut

```
Anaconda Prompt (anaconda3)
(sharing) E:\> cd Sharing
(sharing) E:\Sharing>_
```

STEP 5

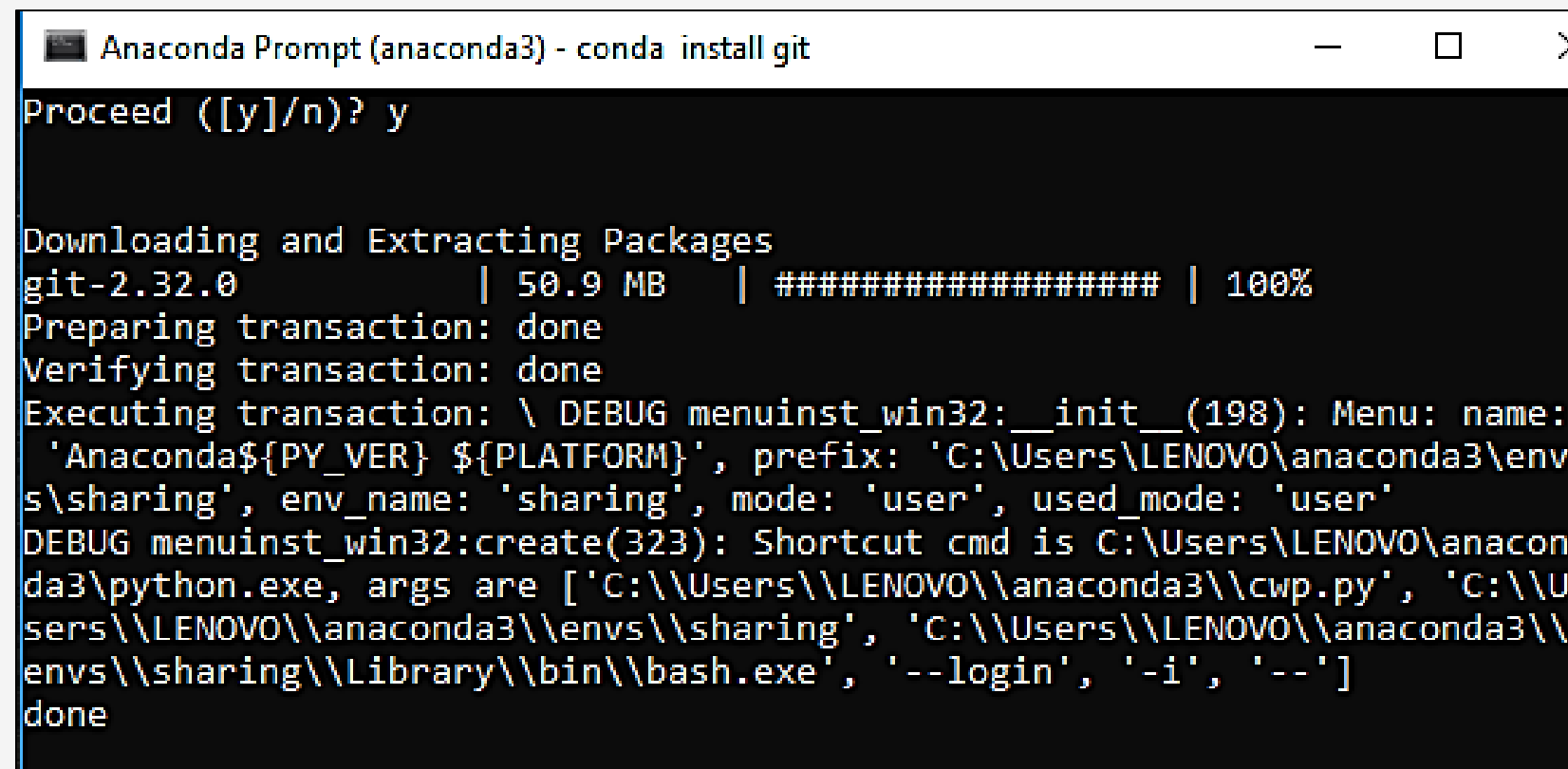
Install git untuk Git Clone dengan mengetikkan perintah "**conda install git**" & klik Enter



```
Anaconda Prompt (anaconda3) - conda install git

(sharing) E:\Sharing> conda install git
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

Ketika selesai, akan muncul pilihan **([y]/n)?** maka harus pilih **y** → klik Enter



```
Anaconda Prompt (anaconda3) - conda install git

Proceed ([y]/n)? y

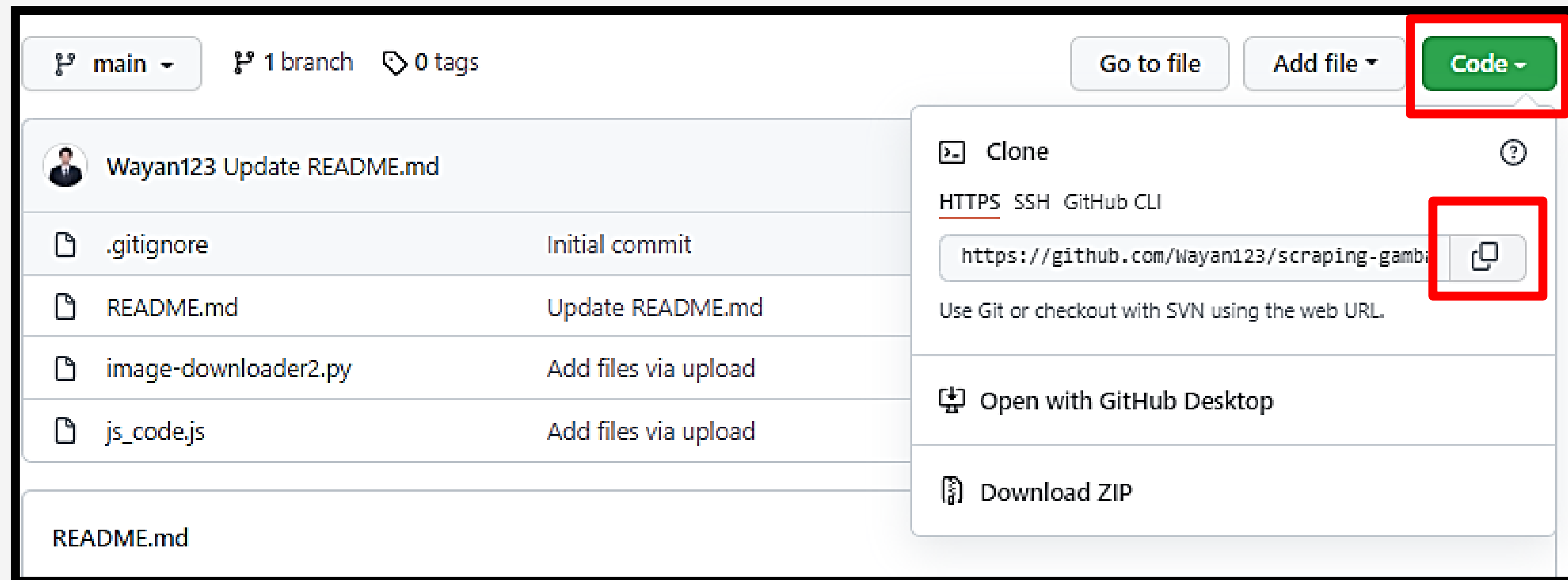
Downloading and Extracting Packages
git-2.32.0          | 50.9 MB | ##### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: \ DEBUG menuinst_win32: __init__(198): Menu: name:
'Anaconda${PY_VER} ${PLATFORM}', prefix: 'C:\Users\LENOVO\anaconda3\envs\sharing', env_name: 'sharing', mode: 'user', used_mode: 'user'
DEBUG menuinst_win32:create(323): Shortcut cmd is C:\Users\LENOVO\anaconda3\python.exe, args are ['C:\\Users\\LENOVO\\anaconda3\\cwp.py', 'C:\\Users\\LENOVO\\anaconda3\\envs\\sharing', 'C:\\Users\\LENOVO\\anaconda3\\envs\\sharing\\Library\\bin\\bash.exe', '--login', '-i', '--']
done
```

Note:

- * Jika anda sudah pernah menginstal git, maka tahap ini tidak perlu dilakukan.
- * Karena saya akan menggunakan syntax untuk scraping yang sudah di upload pada Github, maka git sangat dibutuhkan.

STEP 6

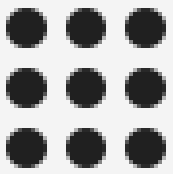
Masuk ke repository github yang berisi syntax untuk scraping, kemudian **pilih Code > klik Copy link**



Link Repository Github → <https://github.com/Wayan123/scraping-gambar-google.git>

STEP 7

Masuk kembali ke anaconda prompt, lalu ketik perintah "**git clone**"
→ paste link yang sebelumnya telah di salin → klik Enter



```
Anaconda Prompt (anaconda3) - conda install git

(sharing) E:\Sharing> git clone https://github.com/Wayan123/scraping-gambar-google.git
Cloning into 'scraping-gambar-google'...
remote: Enumerating objects: 14, done.
remote: Counting objects: 100% (14/14), done.
remote: Compressing objects: 100% (13/13), done.
remote: Total 14 (delta 3), reused 0 (delta 0), pack-reused 0
Receiving objects: 100% (14/14), 6.48 KiB | 1.08 MiB/s, done.
Resolving deltas: 100% (3/3), done.
```

Tunggu hingga done, lalu ketikkan perintah "**cd scraping-gambar-google**" → klik Enter untuk masuk ke direktori tersebut.

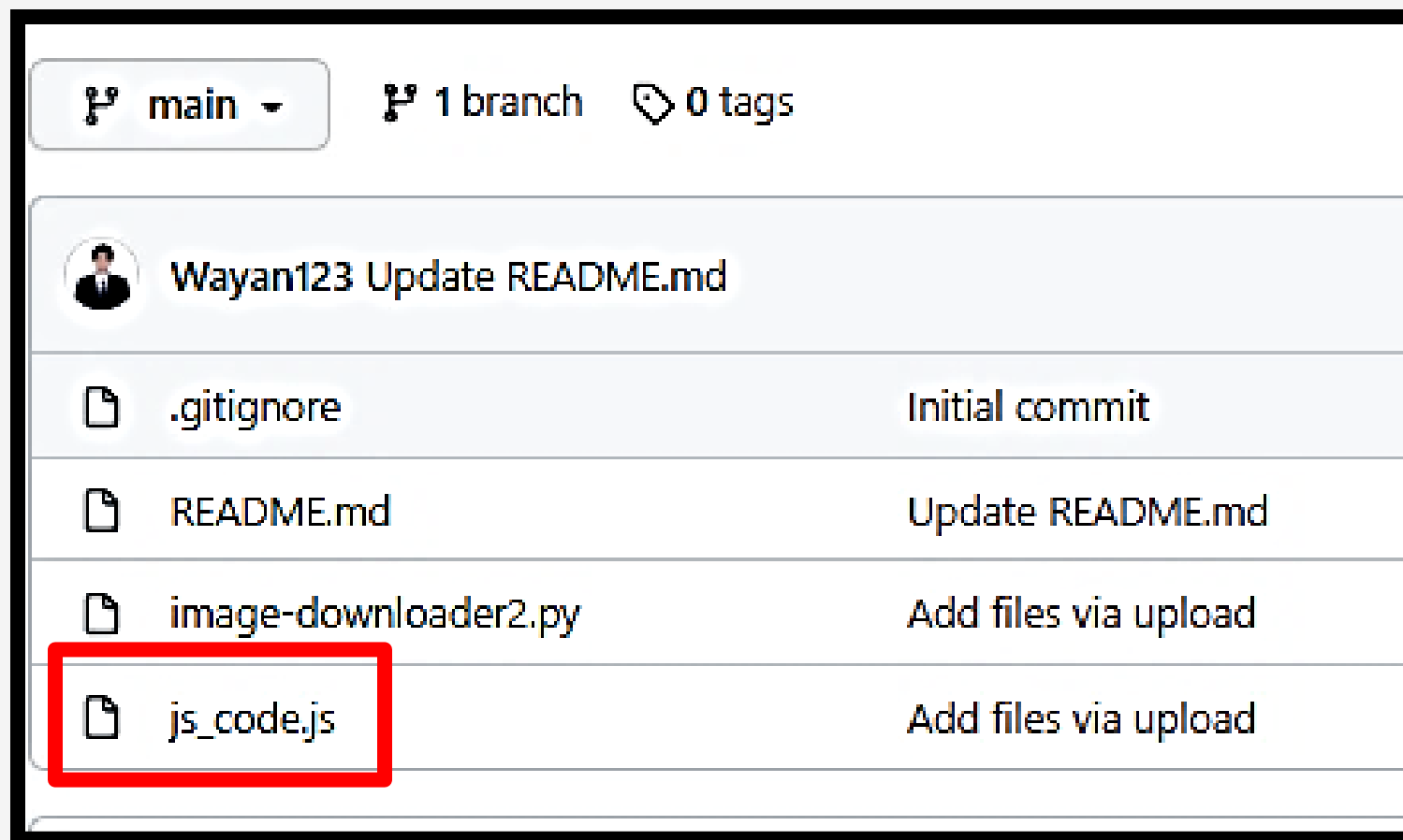
```
Anaconda Prompt (anaconda3) - conda install git

(sharing) E:\Sharing> cd scraping-gambar-google
(sharing) E:\Sharing\scraping-gambar-google>_
```


STEP 8

Scraping Time

Masuk kembali ke repository github sebelumnya, yang berjudul scraping-gambar-google → klik file js_code.js

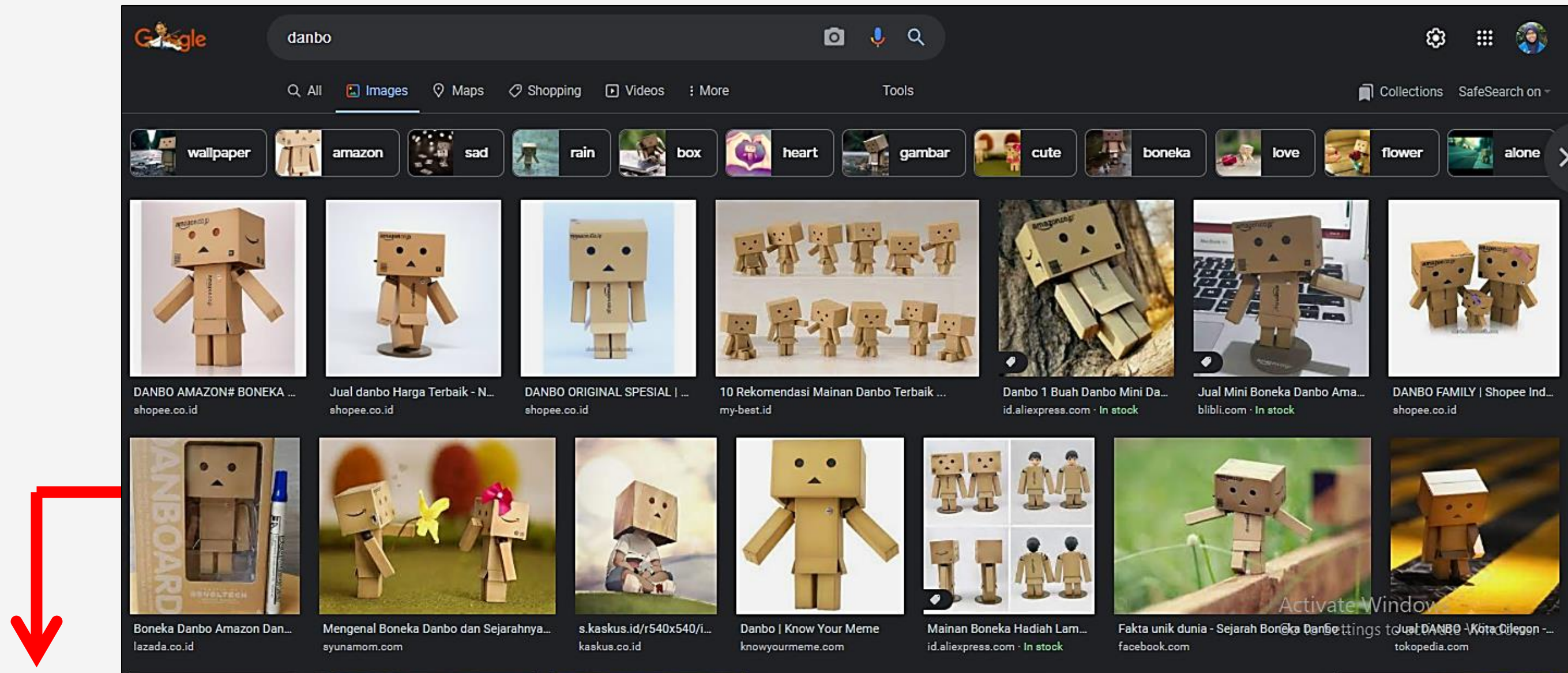


Copy codenya dari line 17-143 atau copy semua code dengan cara klik Raw → Ctrl+A → Ctrl+C

```
143 lines (135 sloc) | 4.61 KB
1
2  /*
3   referensi: https://www.pyimagesearch.com/2017/12/04/how-to-create-a-deep-learning-dataset-using-google-images/
4
5   Cara mendapatkan urls.txt
6   1. Buka google chrome image
7   2. Cari gambar yang ingin di download
8   3. Scroll sampai bawah
9   4. Buka dev tools chrome dengan cara, klik titik 3 di pojok kanan atas, lalu pilih more tools, dan klik
10      Develover tools
11   5. Pilih console
12   6. Lalu paste kode js dibawah ini
13   7. Tunggu sampai teks file urls.txt ter-downloads
14  */
15
16
17  /**
18   * simulate a right-click event so we can grab the image URL using the
19   * context menu alleviating the need to navigate to another page
20   *
21   * attributed to @jmiserez: http://pyimg.co/9qe7y
22   *
23   * @param {object} element DOM Element
24   *
25   * @return {void}
26   */
27  function simulateRightClick( element ) {
28      var event1 = new MouseEvent( 'mousedown', {
29          bubbles: true,
```

STEP 9

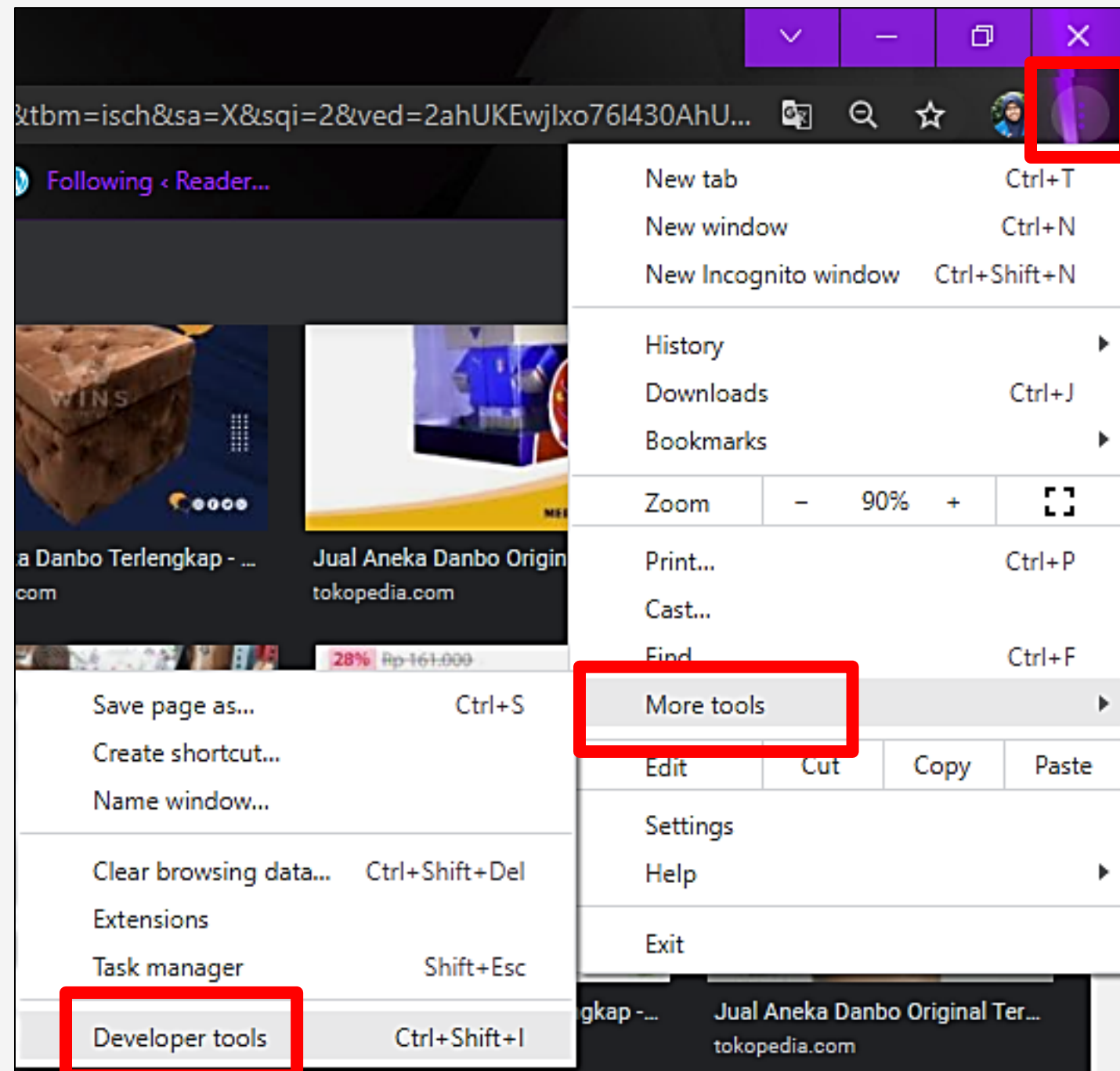
Cari Gambar di Google Images, bebas boleh apapun itu, disini saya mencari “**danbo**”



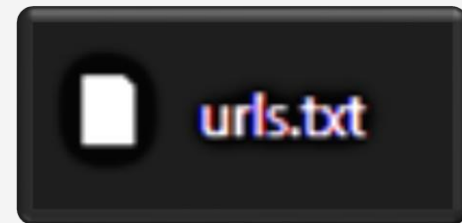
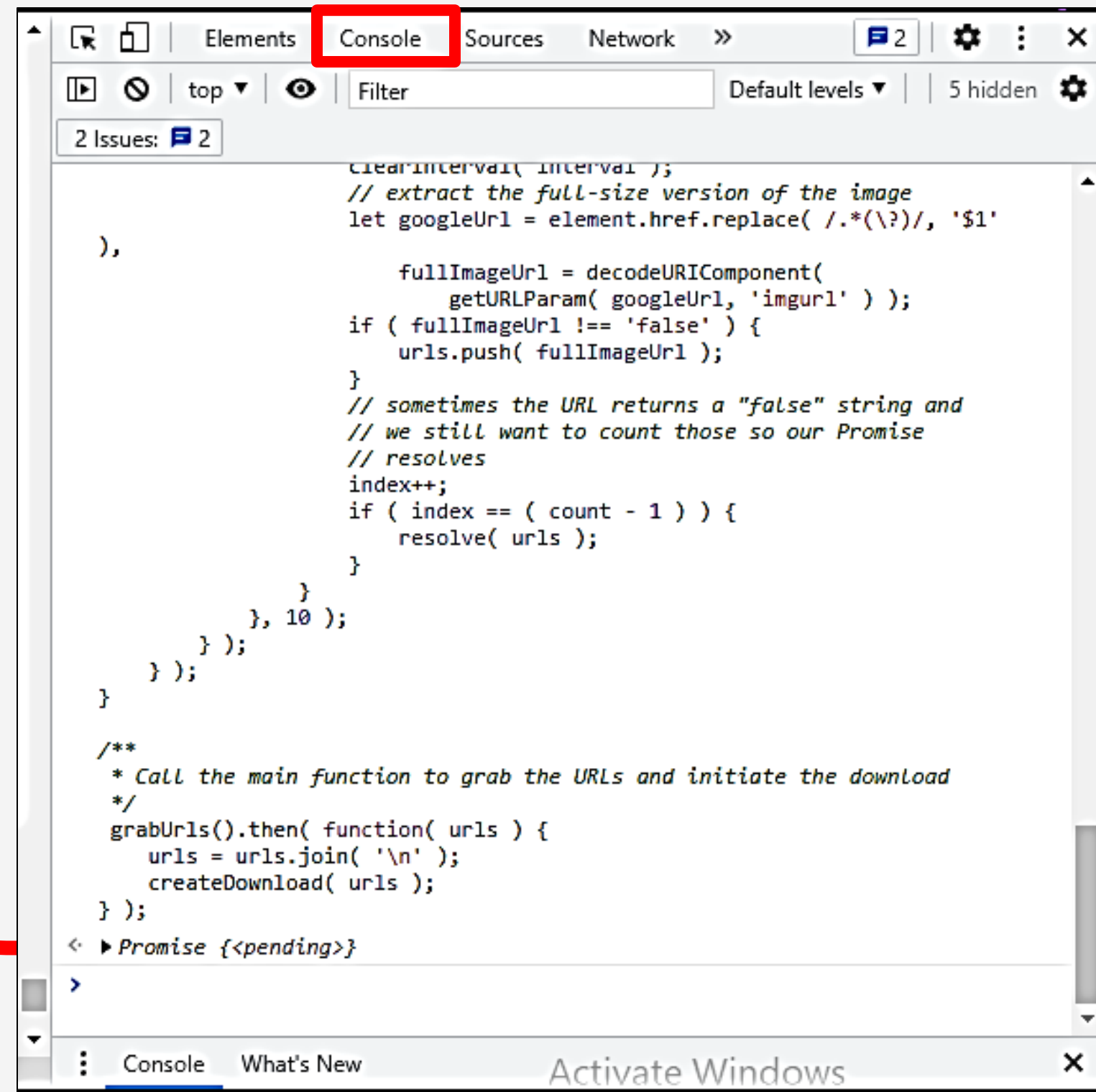
Scroll terus ke bawah & ketika muncul kalimat "**Show more results**" di klik, lanjut scroll hingga muncul kalimat "**Looks like you've reached the end**" ini artinya seluruh gambar dari hasil pencarian telah terdeteksi & mencapai batas akhir.

STEP 10

Klik titik 3 di pojok kanan atas →
more tools → developer tools

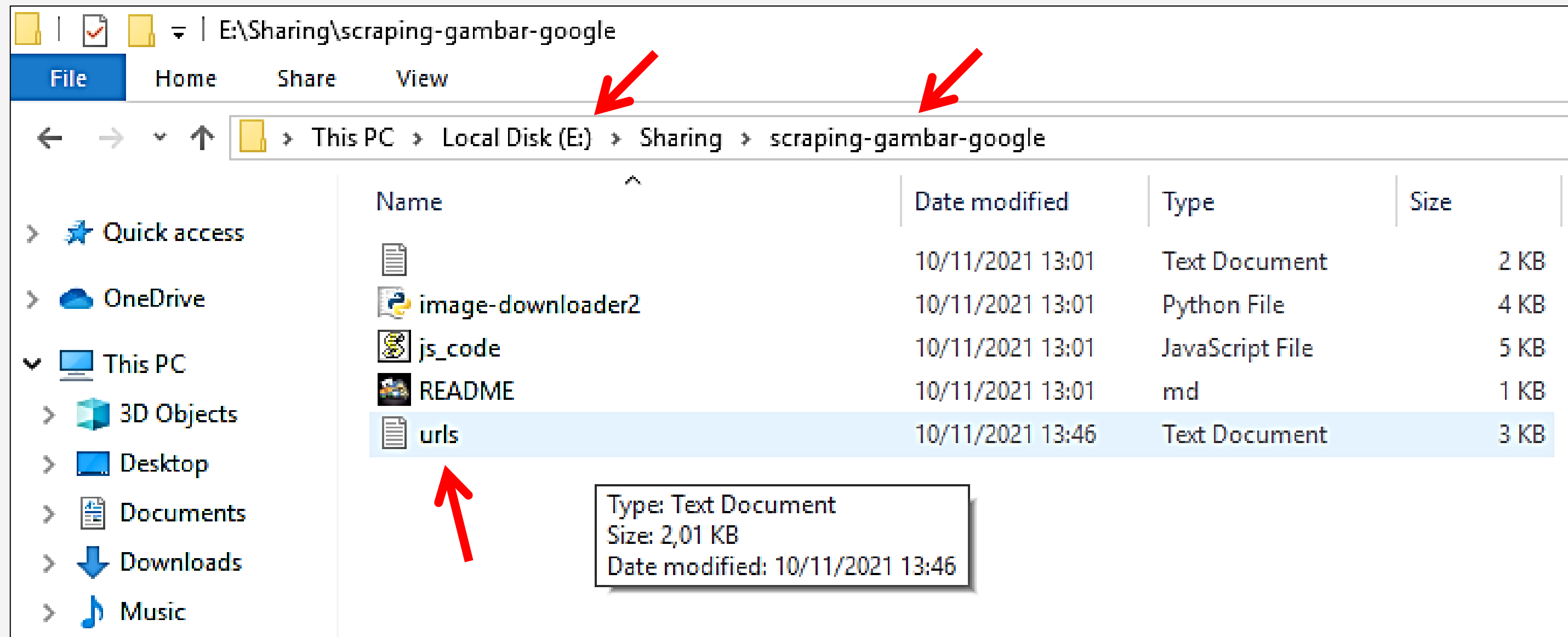


Pilih tab console → paste code js_code.js → klik Enter →
maka file bernama urls.txt akan terdownload

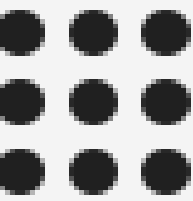


STEP 11

Copy file **urls.txt** ke dalam folder scraping-gambar-google yang terdapat di folder Sharing dalam local disk **E:**



STEP 12



Masuk kembali ke Anaconda Prompt yang masih terbuka, ketik perintah **"mkdir gambar-danbo"** untuk membuat folder baru bernama "gambar-danbo" sebagai tempat menyimpan gambar hasil download.



```
Anaconda Prompt (anaconda3) - conda install git

(sharing) E:\Sharing\scraping-gambar-google> mkdir gambar-danbo
(sharing) E:\Sharing\scraping-gambar-google>
```



Ketikkan perintah **"pip install opencv-python requests imutils"**

```
(sharing) E:\Sharing\scraping-gambar-google> pip install opencv-python requests imutils
Collecting opencv-python
  Using cached opencv_python-4.5.4.58-cp38-cp38-win_amd64.whl (35.1 MB)
Collecting requests
  Using cached requests-2.26.0-py2.py3-none-any.whl (62 kB)
Collecting imutils
  Using cached imutils-0.5.4-py3-none-any.whl
Collecting numpy>=1.17.3
  Using cached numpy-1.21.4-cp38-cp38-win_amd64.whl (14.0 MB)
Collecting idna<4,>=2.5
  Using cached idna-3.3-py3-none-any.whl (61 kB)
Collecting urllib3<1.27,>=1.21.1
  Using cached urllib3-1.26.7-py2.py3-none-any.whl (138 kB)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\lenovo\anaconda3\envs\sharing\lib\site-packages (from requests) (2021.10.8)
Collecting charset-normalizer~=2.0.0
  Using cached charset_normalizer-2.0.7-py3-none-any.whl (38 kB)
Installing collected packages: urllib3, numpy, idna, charset-normalizer, requests, opencv-python, imutils
Successfully installed charset-normalizer-2.0.7 idna-3.3 imutils-0.5.4 numpy-1.21.4 opencv-python-4.5.4.58 requests-2.26.0 urllib3-1.26.7
```

Note : Installing opencv-python, requests, dan imutils dapat dilakukan secara bersamaan seperti screenshot di atas, atau diinstal masing-masing (pip install opencv-python, pip install requests, pip install imutils)

STEP 13

Ketikkan perintah **"dir"** → klik **Enter** untuk melihat isi folder "scraping-gambar-google", ketika folder "gambar-danbo" berhasil dibuat, maka akan muncul di outputnya.

Anaconda Prompt (anaconda3) - conda install git

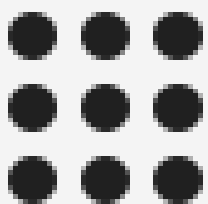
(sharing) E:\Sharing\scraping-gambar-google> dir_

```
(sharing) E:\Sharing\scraping-gambar-google> dir
Volume in drive E has no label.
Volume Serial Number is BEDC-A4F9

Directory of E:\Sharing\scraping-gambar-google

10/11/2021  13:49    <DIR>          .
10/11/2021  13:49    <DIR>          ..
10/11/2021  13:01             1.928 .gitignore
10/11/2021  13:49    <DIR>          gambar-danbo
10/11/2021  13:01             3.395 image-downloader2.py
10/11/2021  13:01             4.718 js_code.js
10/11/2021  13:01              102 README.md
10/11/2021  13:46             2.059 urls.txt
                    5 File(s)              12.202 bytes
                    3 Dir(s)  386.587.201.536 bytes free
```

Output



STEP 14

Ketikkan perintah **"python image-downloader2.py -u urls.txt -o gambar-danbo"** → klik Enter

Anaconda Prompt (anaconda3) - conda install git

```
(sharing) E:\Sharing\scraping-gambar-google> python image-downloader2.py -u urls.txt -o gambar-danbo
```

Anaconda Prompt (anaconda3) - conda install git

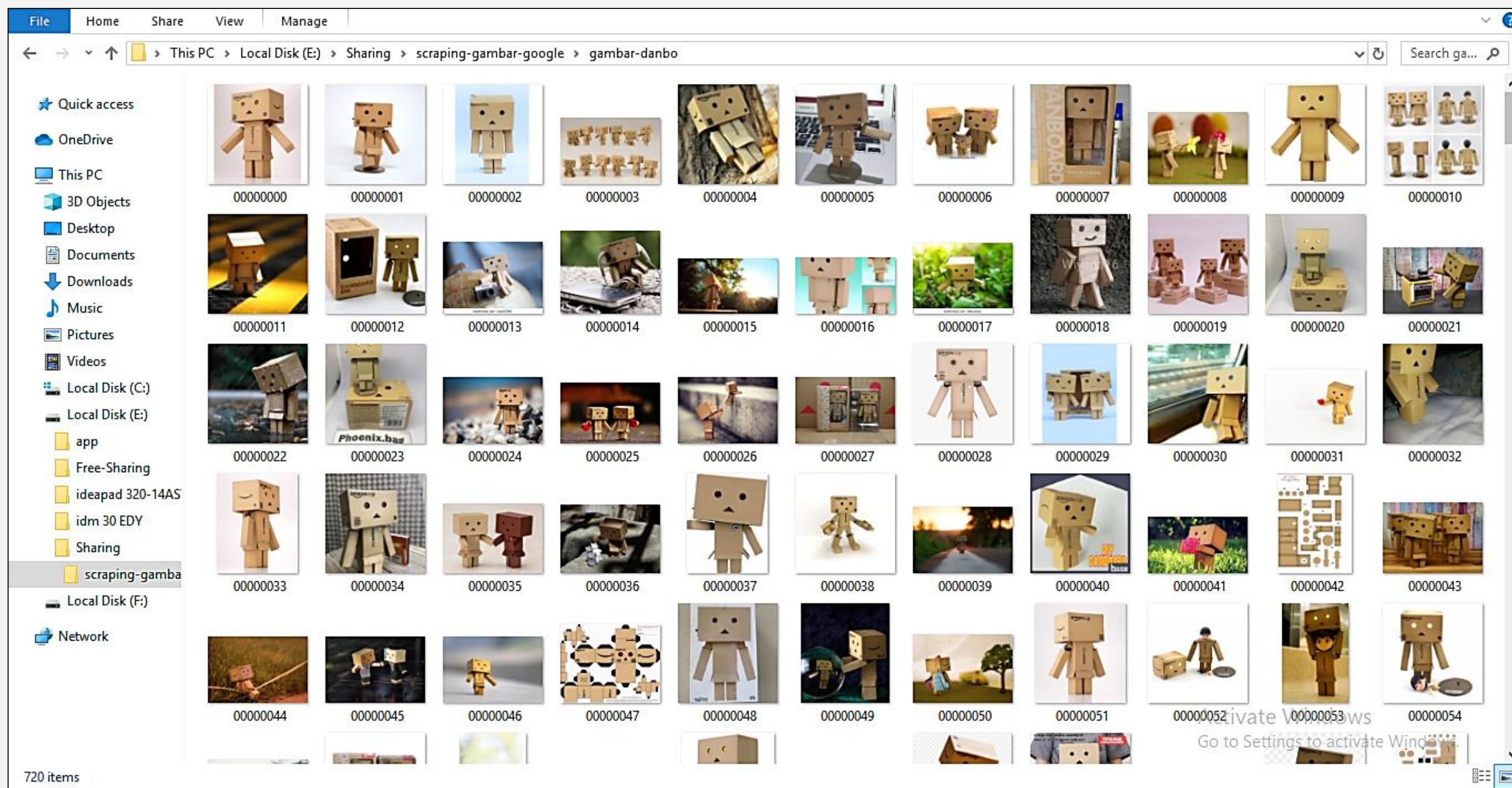
```
Error downloading https://c.wallhere.com/photos/1e/52/1920x1200_px_amazon_autumn_box_danbo-1629515.jpg!s : HTTPSConnectionPool(host='c.wallhere.com', port=443): Max retries exceeded with url: /photos/1e/52/1920x1200_px_amazon_autumn_box_danbo-1629515.jpg!s (Caused by SSLError(SSLCertVerificationError("hostname 'c.wallhere.com' doesn't match either of '*.tri.co.id', 'tri.co.id'")))
libpng warning: iCCP: known incorrect sRGB profile
Error downloading https://mynamepix.com/images/styles/itm_danbo-love2013-05-06_19-28-46_1.jpg : HTTPSConnectionPool(host='mynamepix.com', port=443): Max retries exceeded with url: /images/styles/itm_danbo-love2013-05-06_19-28-46_1.jpg (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x000002F3B7691130>: Failed to establish a new connection: [Errno 11002] getaddrinfo failed'))
Error downloading https://p0.pikist.com/photos/538/346/danbo-figure-longing-friends-fun-figures-sweet-miss-cute-thumbnail.jpg : HTTPSConnectionPool(host='p0.pikist.com', port=443): Max retries exceeded with url: /photos/538/346/danbo-figure-longing-friends-fun-figures-sweet-miss-cute-thumbnail.jpg (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x000002F3B7691100>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed'))
Error downloading https://media-cdn.tripadvisor.com/media/photo-s/1b/ef/19/23/ramen-danbo-australia.jpg : HTTPSConnectionPool(host='media-cdn.tripadvisor.com', port=443): Max retries exceeded with url: /media/photo-s/1b/ef/19/23/ramen-danbo-australia.jpg (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x000002F3B7691430>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed'))
Error downloading https://www.researchgate.net/publication/342104073/figure/fig1/AS:904692917878784@1592707008193/Production-and-analysis-design-for-Danbo-cheese_Q320.jpg : HTTPSConnectionPool(host='www.researchgate.net', port=443): Max retries exceeded with url: /publication/342104073/figure/fig1/AS:904692917878784@1592707008193/Production-and-analysis-design-for-Danbo-cheese_Q320.jpg (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x000002F3B7691AF0>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed'))
Error downloading http://www.drawingskill.com/wp-content/uploads/3/Danbo-Drawing.jpg : HTTPConnectionPool(host='www.drawingskill.com', port=80): Max retries exceeded with url: /wp-content/uploads/3/Danbo-Drawing.jpg (Caused by NewConnectionError('<urllib3.connection.HTTPConnection object at 0x000002F3B76919D0>: Failed to establish a new connection: [Errno 11002] getaddrinfo failed'))
(sharing) E:\Sharing\scraping-gambar-google>_
```

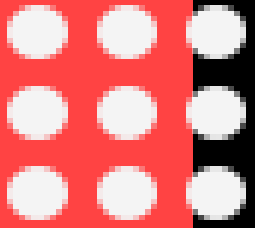
End Processing Code :
Artinya gambar telah
terdownload semua



END

Gambar yang terdownload tersimpan dalam folder "gambar-danbo"





THANK YOU FOR YOUR ATTENTION

KEEP SPIRIT AND HAVE A NICE DAY

