Nicolás David Espejo Bernal
ID: 637747

Taller Facebook Scraper

Verificación instalación de Python:

```
C:\Users\Nicolas>python -V
Python 3.8.5

C:\Users\Nicolas>
```

Instalación de paquetes:

```
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> pip install xlsxwriter

Collecting xlsxwriter
  Downloading XlsxWriter-3.0.1-py3-none-any.whl (148 kB)
     |                              | 148 kB 2.2 MB/s
Installing collected packages: xlsxwriter
Successfully installed xlsxwriter-3.0.1
WARNING: You are using pip version 21.1.1; however, version 21.2.4 is available.
You should consider upgrading via the 'C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Sc
raper\venv\Scripts\python.exe -m pip install --upgrade pip' command.
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> pip install xlrd
Collecting xlrd
  Downloading xlrd-2.0.1-py2.py3-none-any.whl (96 kB)
     |                              | 96 kB 812 kB/s
Installing collected packages: xlrd
Successfully installed xlrd-2.0.1
WARNING: You are using pip version 21.1.1; however, version 21.2.4 is available.
You should consider upgrading via the 'C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Sc
raper\venv\Scripts\python.exe -m pip install --upgrade pip' command.
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> pip install pandas
```

```
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> pip install matplotlib

Collecting matplotlib
  Downloading matplotlib-3.4.3-cp38-cp38-win32.whl (7.0 MB)
     |                              | 7.0 MB 6.4 MB/s
Requirement already satisfied: numpy>=1.16 in c:\users\nicolas\desktop\universidad\2021-2\mineria de datos\i\scr
```

```
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> pip install facebook-s
craper
Collecting facebook-scraper
  Downloading facebook_scraper-0.2.46-py3-none-any.whl (35 kB)
Collecting requests-html<0.11.0,>=0.10.0
  Downloading requests_html-0.10.0-py3-none-any.whl (13 kB)
Collecting dateparser<2.0.0,>=1.0.0
  Downloading dateparser-1.0.0-py2.py3-none-any.whl (279 kB)
     |                              | 279 kB 6.4 MB/s
```

Nicolás David Espejo Bernal
ID: 637747

Ejecución scraper.py:

```
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> py scraper.py
C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper\venv\lib\site-packages\facebook_s
craper\facebook_scraper.py:437: UserWarning: Locale detected as es_LA - for best results, set to en_US
  warnings.warn(f"Locale detected as {locale} - for best results, set to en_US")
Inauguración Centro de Desarrolladores IT en Medellín
¡Nuestro Aniversario continúa! 🎉 Aún tienes tiempo para comprar con precios increíbles🎊 y celebrar con una piñ
ata hecha en casa 🎨Sigue estos sencillos pasos ¡El diseño lo eliges tú!🎁

Corta los papelitos con los que quieres cubrir tu piñata

Empieza a pegarlos encima del cartón

Acomoda muy bien todos los detalles
VIE, 10 DE SEP.
Inauguración Centro de Desarrolladores IT en Medellín
```
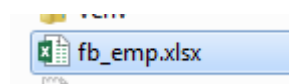
Ejecución scraper2.py:

```
(venv) PS C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper> py scraper2.py
C:\Users\Nicolas\Desktop\Universidad\2021-2\Mineria de Datos\I\Scripts Scraper\venv\lib\site-packages\facebook_s
craper\facebook_scraper.py:437: UserWarning: Locale detected as es_LA - for best results, set to en_US
  warnings.warn(f"Locale detected as {locale} - for best results, set to en_US")
{'post_id': '376386043859531', 'text': 'Inauguración Centro de Desarrolladores IT en Medellín', 'post_text': 'In
auguración Centro de Desarrolladores IT en Medellín', 'shared_text': '', 'time': datetime.datetime(2021, 8, 31,
13, 50, 49), 'timestamp': 1631311212, 'image': None, 'image_lowquality': 'https://scontent.fbog3-2.fna.fbcdn.net
/v/t15.5256-10/cp0/e15/q65/p600x600/241687501_383062766525192_4626146743482535640_n.jpg?_nc_cat=111&ccb=1-5&_nc_
sid=ccf8b3&_nc_ohc=NRnmqXGwJOwAX8j-zf9&_nc_oc=AQlRbV-hYohOhNXarn86m9U3UZFNBOgIgEJsqtk2i5cDVnnAS6ij7cOGT7W3ubPelk
Y&_nc_ht=scontent.fbog3-2.fna&oh=c3cd0ea24c5afc1b86dd5c9e4bdbdac6&oe=61495202', 'images': [], 'images_descriptio
n': [], 'images_lowquality': ['https://scontent.fbog3-2.fna.fbcdn.net/v/t15.5256-10/cp0/e15/q65/p600x600/2416875
01_383062766525192_4626146743482535640_n.jpg?_nc_cat=111&ccb=1-5&_nc_sid=ccf8b3&_nc_ohc=NRnmqXGwJOwAX8j-zf9&_nc_
```

📄 fb_emp.xlsx

post_id	text	post_text	hared_tex	time	timestamp	image	ge_lowqua	images	es_descrip	ges_lowqu	wquality_c	video	duratio
376386043	Inaugurac	Inauguración Centro	########	1,63E+09		https://sc	[]	[]	['https://s	[None]	https://vi	24
383827610	¡Nuestro /	¡Nuestro Aniversario	########	1,63E+09	https://sc	https://sc	['https://s	['Puede se	['https://s	['Puede ser una imagen de t
570306241	VIE., 10 DE SEP.Inaug	VIE., 10 DE	########	1,63E+09	https://sc	https://ex	['https://s	[]	['https://e	[None]
378921900509528			########	1,63E+09	https://sc	https://sc	['https://s	['Puede se	['https://s	['Puede ser una imagen de t
378535417	Si quieres	Si quieres cambiar tu	########	1,63E+09		https://sc	[]	[]	['https://s	[None]	https://vi
374619410	No todas l	No todas l	Suscríbete	########	1,63E+09	https://cd	https://ex	['https://c	[]	['https://e	[None, None, None, None]
372568011	¿Sabes qu	¿Sabes qué es lo que	########	1,63E+09	https://sc	https://sc	['https://s	['Puede se	['https://s	['Puede ser una imagen de r
370894854	Esta es tu	Esta es tu cara cuand	########	1,63E+09		https://sc	[]	[]	['https://s	[None]	https://vi
367076691	Cuando cc	Cuando cc	Ver más e	########	1,63E+09	https://cd	https://ex	['https://c	[]	['https://e	[None, None, None, None,
365612978	Apoyamo:	Apoyamos a los emp	########	1,63E+09		https://sc	[]	[]	['https://s	[None]	https://vi
363432728	¡En el #Día	¡En el #Día	Sorprende	########	1,63E+09	https://cd	https://ex	['https://c	[]	['https://e	[None, None, None, None,
357466582	Para los fa	Para los fa	Opciones	########	1,63E+09	https://cd	https://ex	['https://c	[]	['https://e	[None, None, None, None,
355609149	Nada más	Nada más emociona	########	1,63E+09		https://sc	[]	[]	['https://s	[None]	https://vi