

DATA IMPORTING

Flat files from the web

urllib → Permite cargar datos de la web `urlopen()` → acepta URLs

Para cargar un archivo web a local →

```
from urllib.request import urlopen
url = 'http://... / archivo.csv'
urlopen(url, 'archivo.csv')
```

→ Hace un GET request

Si es un excel y queremos leer todos los hojias →

```
df = pd.read_excel('...', sheet_name=hoja)
# o acceder df['hoja.n'].head()
```

Http request para importar archivos web

Para extraer el html de una pagina →

```
from urllib.request import urlopen, Request
url = '...'
request = Request(url) → hacemos el get
respuesta = urlopen(request) → cargamos la respuesta
html = respuesta.read()
```

Con requests

```
import requests
url = '...'
r = requests.get(url)
texto = r.text
```

Introducción a APIs y JSON

Para cargar un JSON →

```
import json
with open('archivo.json', 'r') as json_file:
    json_data = json.load(json_file)
```

Para hacer un request y cargar un JSON →

```
r = requests.get(url)
json_data = r.json()
```

API de Twitter

Primero hay que registrarse y crear una app.

Retorna los tweets como un json.

Como funciona →

```
import tweepy, json
access_token = ''
consumer_key = ''
consumer_secret = ''

stream = tweepy.Stream(consumer_key, consumer_secret, access_token, access_token_secret)
stream.filter(track=['python tweets']) → Vemos los tweets
```